Cell

Systematic Identification of Barriers to Human iPSC Generation

Han Qin,[1,2,9] Aaron Diaz,[1,3,9] Laure Blouin,[1,2] Robert Jan Lebbink,[1,4,7] Weronika Patena,[1,4] Priscilia Tanbun,[1,2] Emily M. LeProust,[6] Michael T. McManus,[1,4,10] Jun S. Song,[1,3,5,8,10,*] and Miguel Ramalho-Santos[1,2,10,*]
[1]Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research
[2]Departments of Ob/Gyn and Pathology, Center for Reproductive Sciences, and Diabetes Center
[3]Institute for Human Genetics
[4]Department of Microbiology and Immunology, Diabetes Center, and the WM Keck Center for Noncoding RNAs
[5]Department of Epidemiology and Biostatistics and Department of Bioengineering and Therapeutic Sciences
University of California, San Francisco, San Francisco, CA 94143, USA
[6]Genomics Solution Unit, Agilent Technologies Inc., Santa Clara, CA 95051, USA
[7]Present address: Department of Medical Microbiology, UMC Utrecht, 3584CX Utrecht, The Netherlands
[8]Present address: Departments of Bioengineering and Physics, University of Illinois, Urbana-Champaign, Urbana, IL 61801, USA
[9]Co-first author
[10]Co-senior author
*Correspondence: songj@illinois.edu (J.S.S.), miguel.ramalho-santos@ucsf.edu (M.R.-S.)
http://dx.doi.org/10.1016/j.cell.2014.05.040

SUMMARY

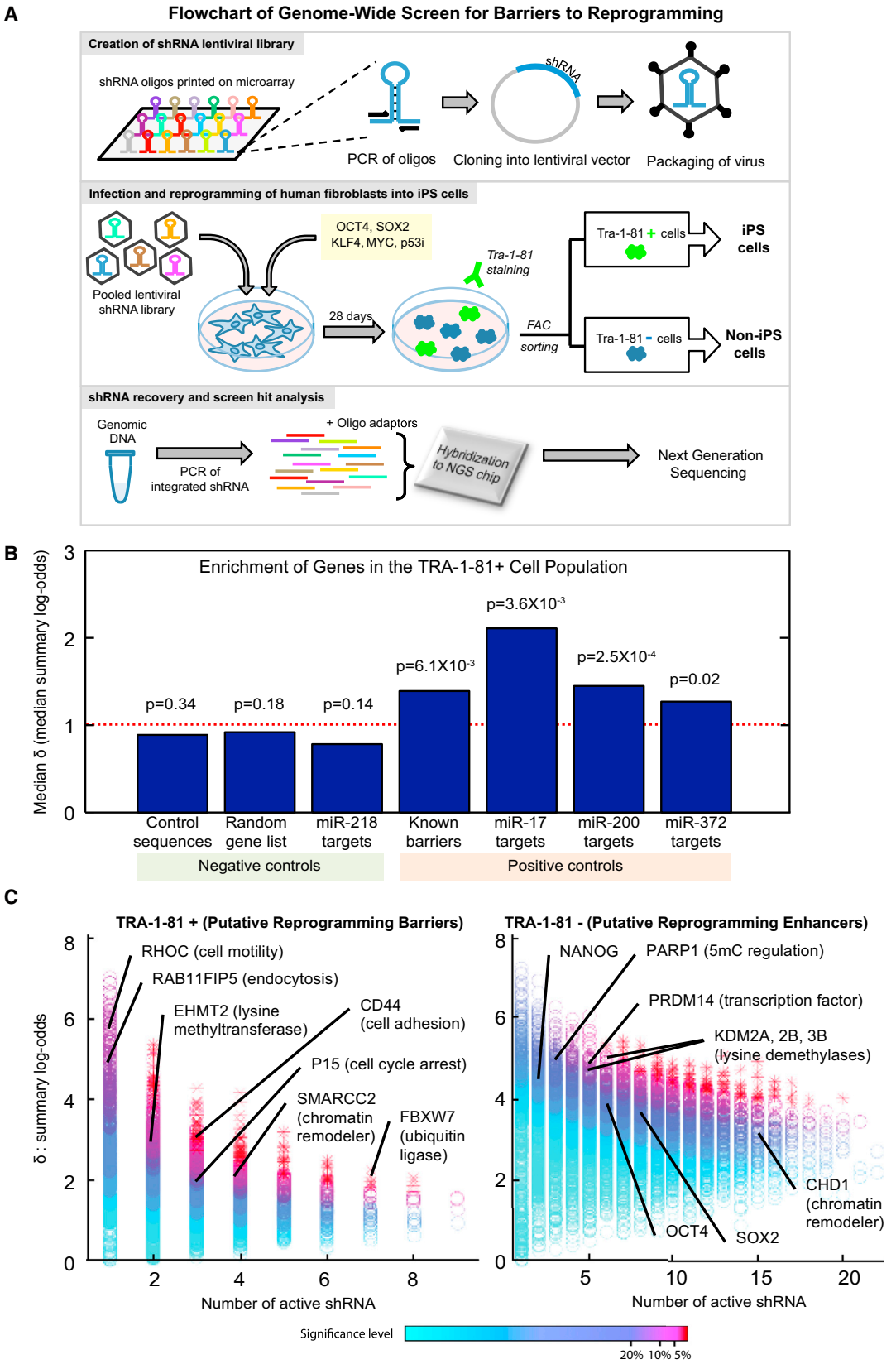
Reprogramming of somatic cells to induced pluripotent stem cells (iPSCs) holds enormous promise for regenerative medicine. To elucidate endogenous barriers limiting this process, we systematically dissected human cellular reprogramming by combining a genome-wide RNAi screen, innovative computational methods, extensive single-hit validation, and mechanistic investigation of relevant pathways and networks. We identify reprogramming barriers, including genes involved in transcription, chromatin regulation, ubiquitination, dephosphorylation, vesicular transport, and cell adhesion. Specific a disintegrin and metalloproteinase (ADAM) proteins inhibit reprogramming, and the disintegrin domain of ADAM29 is necessary and sufficient for this function. Clathrin-mediated endocytosis can be targeted with small molecules and opposes reprogramming by positively regulating TGF-β signaling. Genetic interaction studies of endocytosis or ubiquitination reveal that barrier pathways can act in linear, parallel, or feedforward loop architectures to antagonize reprogramming. These results provide a global view of barriers to human cellular reprogramming.

INTRODUCTION

Cells generally become committed to increasingly differentiated fates during normal development, but experimental paradigms for cellular reprogramming have shown that differentiation is reversible. Initial methods for reprogramming included somatic cell nuclear transfer to enucleated oocytes (Gurdon et al., 1958) or fusion with pluripotent stem cells (Tada et al., 1997), both of which were technically difficult to dissect at the mechanistic level. The finding that overexpression of a key transcription factor leads to fibroblast-to-muscle transdifferentiation (Davis et al., 1987) suggested that perhaps other cell-fate transitions could similarly be induced. The recent discovery that combinations of transcription factors can reprogram somatic cells to induced pluripotent stem cells (iPSCs) has transformed stem cell research and holds enormous promise in regenerative medicine (Takahashi and Yamanaka, 2006; Takahashi et al., 2007; Yu et al., 2007). Understanding the process of reprogramming may also shed light on events that take place during cellular transformation in cancer (Daley, 2008; Ramalho-Santos, 2009). However, due to the low efficiency and protracted nature of the process, the mechanisms that underlie the generation of iPSCs remain poorly understood.

Recent studies have revealed that reprogramming to the iPSC state is a multistep process involving large-scale changes in the transcriptional and epigenetic states of somatic cells along the path to pluripotency (Buganim et al., 2012; Polo et al., 2012). Importantly, the somatic cell is not a *tabula rasa* and expresses genes that antagonize reprogramming, as has been shown for tumor suppressors (p53, INK4a/ARF, LATS2) (Kawamura et al., 2009; Qin et al., 2012; Zhao et al., 2008) and H3K9 methyltransferases (SETDB1, SUV39H, EHMT2) (Chen et al., 2013). In addition, focused RNAi screens have revealed other pathways that act as barriers to reprogramming, such as TGF-β signaling (Samavarchi-Tehrani et al., 2010), H3K79 methylation by DOT1L (Onder et al., 2012), or protein ubiquitination (Buckley et al., 2012). These findings suggest that other critical barriers to reprogramming are likely to exist, but no genome-wide functional screen has yet been carried out in mouse or human iPSC generation.

RNAi provides a powerful technique for exploiting a cell's endogenous machinery for mRNA degradation to obtain selective gene knockdown. Well-based genome-wide RNAi screens,



(legend on next page)

where cells are transfected in separate wells with small pools or individual small interfering RNAs (siRNAs), have been carried out successfully, including for the identification of genes that regulate human embryonic stem cell (hESC) self-renewal and pluripotency (Chia et al., 2010). However, the throughput of this approach is limited, particularly in the context of iPSC generation, because of the low reprogramming efficiency. An alternative to well-based screens yielding much higher throughput is a pooled short hairpin RNA (shRNA)-based screen combined with next-generation sequencing (NGS). This approach has a significantly larger dynamic range and has enabled genome-wide screens at an unprecedented scale (Bassik et al., 2013; 2009). However, the extraction of robust biological information from genome-wide screen data is still challenging: the problems of false-positive hits caused by off-target effects, false-negative hits caused by ineffective RNAi, and variance in sequencing depth can limit reliability.

We use ultracomplex EXPANDed pooled shRNA libraries to report a genome-wide screen for barriers to human cellular reprogramming. We introduce a multiobjective optimization technique for analyzing NGS-based shRNA screen data and combine our method with systems-level meta-analyses and in vitro experiments to discover critical barriers to reprogramming genome wide. Our integrative approach identifies 956 genes predicted to act as barriers to reprogramming, including genes involved in transcription, chromatin regulation, ubiquitination, dephosphorylation, vesicular transport, and cell adhesion. We mechanistically dissect the roles of disintegrin proteins and clathrin-mediated endocytosis as reprogramming barriers and show that barriers from different pathways interact and can have combinatorial effects to antagonize reprogramming. The results are compiled into an online resource (http://song.igb.illinois.edu/ipsScreen/), allowing researchers to browse, query, and visualize the analysis.

RESULTS AND DISCUSSION

Genome-wide shRNA Libraries and NGS Enable an Unbiased Screen for Barriers to Reprogramming

We sought to implement a robust and unbiased screen for barriers to human iPSC generation. We used a recently described method (Bassik et al., 2009) to perform a genome-wide shRNA library screen targeting 19,527 human genes with an average coverage of 30 independent shRNAs per gene. Human BJ fibroblasts were coinfected with lentivirus expressing these shRNAs along with OCT4, SOX2, KLF4, c-MYC (4F), and p53 RNAi (p53i). We chose to add p53i because it has been shown to enhance reprogramming efficiency (Kawamura et al., 2009; Zhao et al., 2008). Moreover, data from a pilot screen demonstrate that p53i increases the sampling rate of fully reprogrammed cells and, hence, improves sensitivity in the detection of reprogramming barriers (Figure S1A available online). Importantly, all of our downstream hit validation was done in the absence of p53i (see below). Following the appearance of colonies with iPSC characteristics on day 28, we fluorescence-activated cell sorting (FACS) purified the transduced cells for TRA-1-81, a marker of fully reprogrammed human iPSCs (International Stem Cell Initiative et al., 2007). Integrated shRNAs were then recovered and identified by PCR amplification from genomic DNA of both the TRA-1-81$^+$ and TRA-1-81$^-$ cell populations and quantified by NGS (Figure 1A).

The relative frequency of reads mapping to a given shRNA in TRA-1-81$^+$ compared to TRA-1-81$^-$, expressed as an odds ratio θ, estimates the positive effect size of that shRNA on reprogramming and, thus by inference, the negative effect size of its targeted gene as a potential reprogramming barrier. We call an shRNA *active* if it has greater odds of being sequenced in TRA-1-81$^+$ compared to TRA-1-81$^-$, i.e., if $\theta > 1$ with sufficient coverage. To assess gene-wise collective shRNA activity levels and to quantify the negative effect size of a given gene on reprogramming, we combined the log-transformed odds ratios for all active shRNAs targeting a gene into a single statistic using a random-effects model. The combined statistic, δ, controls for variance in knockdown efficiency and variance in sequencing depth-of-coverage by estimating them separately from the data. We then developed a gene-ranking algorithm based on multiobjective optimization (Handl et al., 2007). This technique simultaneously maximizes collective shRNA activity and the number of distinct active shRNAs. The result is a gene ranking that controls for off-target effects, as highly ranked genes are precisely the ones that demonstrate a reproducible effect across multiple shRNAs enriched in the TRA-1-81$^+$ population. To assign statistical significance to gene ranks, we computed p values using a permutation test with 5,000 permutations of the read counts. False discovery rates (FDRs) were computed based on library swap to adjust for multiple-hypothesis testing. We compared this multiobjective optimization algorithm, which we call HitSelect, with existing methods, such as RIGER (Luo et al., 2008) and RSA (König et al., 2007), and found that HitSelect has greater sensitivity and specificity and is less prone to off-target effects (Figure S1B). This approach generated 956 TRA-1-81$^+$ screen hits at the 5% significance level, corresponding to an FDR of 7% (Extended Experimental Procedures and Data S1).

A Genome-wide Screen Confirms Known Barriers to Reprogramming

To validate our screen, we used as positive controls previously identified gene barriers, as well as the experimentally validated

Figure 1. A Genome-wide RNAi Screen Identifies Known and Novel Regulators of Human iPSC Generation

(A) Design of the genome-wide RNAi screen in human iPSC generation.

(B) Positive control genes consisting of known barriers identified from the literature and the experimentally validated targets of miR-17, miR-200, and miR-372 (miRNAs known to enhance reprogramming efficiency) are enriched in TRA-1-81$^+$. Shown are the medians of the distributions of collective shRNA activity levels for these genes. The statistical significance was measured by a hypergeometric test.

(C) Each dot on the graph is a gene. The x axis shows the number of distinct active shRNAs targeting the gene. The y axis represents the collective shRNA activity level δ on a log scale; this summarizes the read-count odds ratios of all active shRNAs targeting a given gene. Genes with greater statistical significance are color-coded toward red. The top 1% most significant genes are denoted by stars. Examples of known barriers or enhancers of reprogramming are indicated.

targets of microRNAs (miRNAs) known to enhance reprogramming efficiency. Known barrier genes show significant activity levels in TRA-1-81$^+$ compared to TRA-1-81$^-$ (Figures 1B and 1C). Specifically, we identified 26 barrier genes whose knockdown has been previously shown to enhance the efficiency of reprogramming, including genes associated with tumor suppression (CDKN2B), epithelial-to-mesenchymal transition (TGFBR2), heterochromatin (EHMT2, MECP2, RBL2, SMARCC2), and extracellular matrix and adhesion (ARHGAP26, RHOC, CD44) (Figure 1C and Table S1). These positive controls are significantly enriched in the TRA-1-81$^+$ population (Figure 1B). Targets of miRNAs known to enhance human reprogramming efficiency, such as miR-17, miR-200, and miR-372 (Gregory et al., 2008; Li et al., 2011; Subramanyam et al., 2011), also show significant enrichments in the TRA-1-81$^+$ population (Figure 1B). As negative controls, we used control coding sequences represented in the library (GFP, ampicillin resistance, or luciferase), a random gene list, or validated targets of miR-218, a known tumor suppressor not linked to reprogramming. These controls all show weak median activity (below $\delta = 1$) and statistically insignificant enrichment in the TRA-1-81$^+$ population (Figure 1B). Overall, these data indicate that our approach is well validated by the successful identification of previously reported barriers to reprogramming. This work focuses on genes that act as barriers to reprogramming for two main reasons: (1) they can be easily knocked down to improve the efficiency of iPSC generation, providing straightforward single hit validation; (2) a search for shRNAs enriched in the TRA-1-81$^-$ population is complicated by the high heterogeneity of that population, which is not the case in the TRA-1-81$^+$ population. Nevertheless, we also identified shRNAs for multiple known positive regulators of reprogramming enriched in the TRA-1-81$^-$ population, including OCT4, SOX2, NANOG, PRDM14, CHD1, PARP1, KDM2A, KDM2B, and KDM3B (Figure 1C).

Analysis of Screen Hits Identifies Ubiquitination, Cell Adhesion/Motility, and Endocytosis Pathways as Putative Barriers to Reprogramming

We took an integrated approach to identifying barrier pathways and prioritizing genes for downstream validation. We clustered the 956 TRA-1-81$^+$ screen hits by Interpro protein domain, Gene Ontology (GO) cellular component, and GO biological process annotations via DAVID (Huang et al., 2009a). The most frequently annotated protein domain is "WD40 repeat conserved site" (Figure 2A). This domain often serves as a substrate for protein-protein interactions, and the WD40 annotation cluster contains multiple ubiquitin-conjugating enzymes and ligases. The importance of the ubiquitination pathway is highlighted by the recent finding that the E3 ligase F box and WD40 domain protein 7 (FBXW7) functions as a barrier to reprogramming in mouse (Buckley et al., 2012). FBXW7 is also a hit in our screen and is part of the WD40 annotation cluster. Using GeneMANIA (Zuberi et al., 2013) to aggregate published genetic interaction data, we identified multiple screen hits documented to interact with FBXW7 (Figure 2B). Thus, the role of ubiquitination as a barrier to reprogramming may be conserved (see below, Figure S3).

Also enriched among screen hits are annotations associated with cell adhesion and secretion, as well as adhesion-related extracellular protein domains (Figure 2). The most statistically significant protein domain annotation cluster is "Protocadherin." Protocadherins are the largest mammalian subfamily of cadherins and have been implicated in cell adhesion and signal transduction. Gene network analysis reveals that there is a dense colocalization subnetwork of protocadherins among screen hits (Figure 2B). There are also a large number of proteins containing a "fibronectin type III" domain, as well as a significant number of metallopeptidases. Several members of the metallopeptidase cluster are ADAM proteins. These molecular functions are likely related to the fact that the most enriched biological process term is "Cell adhesion." Thus, pathways mediating cell adhesion and extracellular matrix interactions are likely to contain barriers to reprogramming (see below, Figure 4).

Lastly, screen hits showed a significant enrichment for functional annotations related to endocytosis (Figure 2). This enrichment occurs across all three annotation categories. The second most frequently annotated biological process is "Vesicle-mediated transport." Moreover, the largest cellular component annotation clusters include multiple compartments of the endocytic pathway, such as "Cytoplasmic vesicle," "Lysosome," and "Vacuole." Together, this analysis suggests that endocytosis-mediated trafficking is a barrier to reprogramming (see below, Figure 5).

We also examined published time-course data on gene expression and chromatin marks from reprogramming experiments in mouse. In particular, we considered two recent data sets where secondary mouse embryonic fibroblasts (MEFs) were reprogrammed by dox-inducible expression of 4F (Golipour et al., 2012; Polo et al., 2012). Both groups analyzed the gene-expression profiles of intermediate populations during reprogramming, whereas Polo et al. additionally profiled H3K4me3 and H3K27me3 genome wide. A key feature of both experimental designs was a strategy to enrich intermediate populations for cells with the potential to give rise to iPSCs. Interestingly, our re-analysis of these data sets reveals that functional annotations associated with cell adhesion, motility, and extracellular matrix interactions are overrepresented among genes downregulated during the course of reprogramming in the Polo et al. data (Figure S2A). Furthermore, in the data of Golipour et al., there is significant enrichment for annotations associated with endocytosis in genes downregulated as reprogramming proceeds (Figure S2B). Thus, several of the pathways implicated as barriers by our functional analysis in human reprogramming have expression patterns consistent with such a role during the course of mouse reprogramming.

Taken together, our integrative analysis indicates that pathways associated with ubiquitination, cell adhesion/motility, and endocytosis are putative reprogramming barriers. Other potential barriers include genes with roles in transcription, chromatin regulation, and dephosphorylation (Figure 2).

Single-Gene Analyses Validate Reprogramming Barriers from Multiple Pathways

To assess whether genes and pathways enriched in the 956 TRA-1-81$^+$ hits truly represent reprogramming barriers, we subjected 23 genes to rigorous validation. In addition to genes related to ubiquitination (UBE2D3, UBE2E3, RNF40), ADAM family

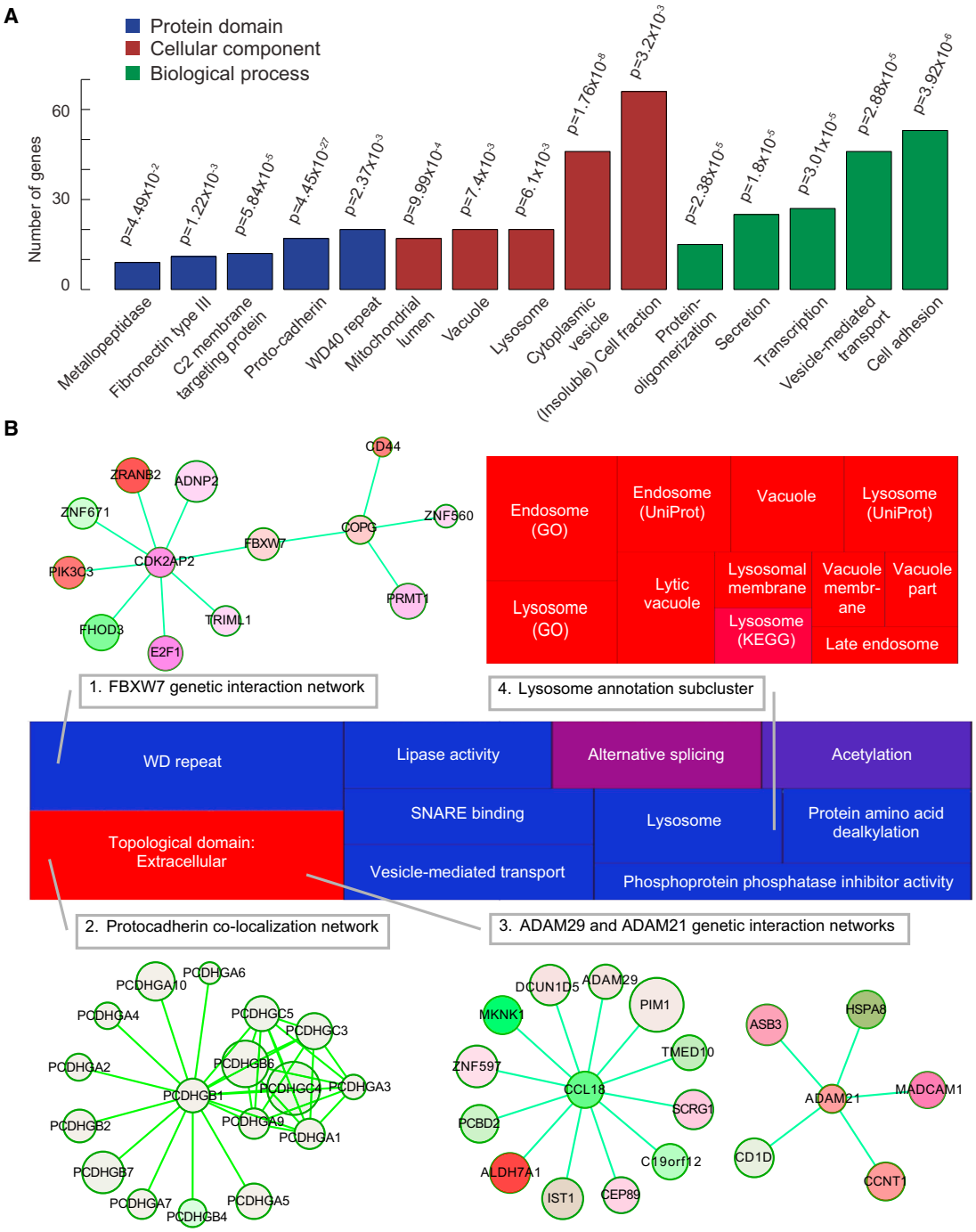


Figure 2. An Integrated Resource Identifies Barrier Pathways to Reprogramming

(A) Screen hits, significant at the 5% level, were clustered by over-represented GO biological process, GO cellular component, and Interpro protein domain annotations using the DAVID functional analysis tool. Shown are the clusters with the five highest enrichment scores. Each cluster is represented by the most frequently occurring annotation term. On the y axis is the number of genes in each cluster. The p value shown was obtained by using Fisher's method to combine DAVID assigned p values for individual annotation terms within a cluster.

(B) Gene functional annotations, aggregated via DAVID, were compiled into an interactive tree-map, available online (http://song.igb.uiuc.edu/ipsScreen/docs/david_treemap.html). All genes shown are screen hits at the 5% significance level. Gene lists for subclusters were analyzed for network structure. This identified several gene networks and subclusters. (1) A genetic interaction network for the ubiquitin ligase FBXW7. This network documents that FBXW7 can directly and indirectly interact with multiple screen hits from a variety of pathways. (2) A dense network of protocadherins. Two genes are connected with an edge if their expression correlates across tissue types. (3) Genetic interaction networks for ADAM family proteins. (4) A subcluster enriched for endocytosis functional annotations.

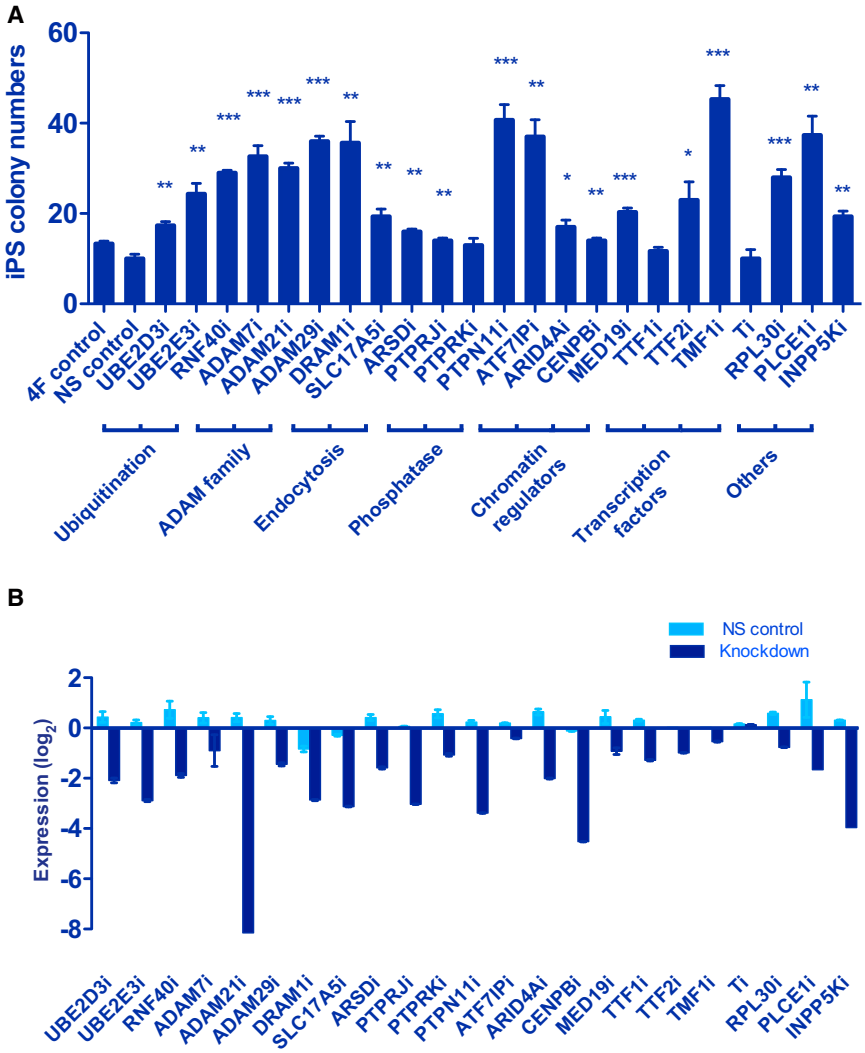


Figure 3. Single-Gene Validation of Reprogramming Barriers

(A) Individual knockdown of screen hits in different pathways increases the efficiency of human iPSC generation. The number of iPSC colonies was counted 25 days after infection of human BJ fibroblasts with 4F control, 4F with nonsense (NS) shRNA control, or 4F with the two most highly active shRNAs targeting the indicated gene. Error bars represent SD compared to 4F + NS. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

(B) Reduction in the levels of gene expression achieved by shRNA constructs was confirmed by qRT-PCR.

genes (ADAM7, ADAM21, ADAM29), and endocytosis (ARSD, SLC17A5, DRAM1), we also tested protein tyrosine phosphatases (PTPRJ, PTPRK, PTPN11), chromatin regulators (ATF7IP, ARID4A, CENPB, MED19), transcription factors (TTF1, TTF2, TMF1, T), and miscellaneous others (RPL30, PLCE1, INPP5K). For each gene, we chose the two most highly active shRNAs from the screen and performed single-gene validation of the knockdown effect on reprogramming. RNAi for 20 out of 23 genes tested (a hit validation of 87%) leads to significant increases in iPS colony count compared with nonsense (NS) shRNA control (Figure 3A). The fold increase in reprogramming efficiency ranges from 1.5 to 4.2, which is comparable with previous studies of knockdown of reprogramming barriers, including p53/p21 (Kawamura et al., 2009; Zhao et al., 2008), LATS2 (Ohi et al., 2011), DOT1L, and SUV39H1 (Onder et al., 2012). Knockdown of genes by shRNAs was confirmed by RT-PCR (Figure 3B).

We next present further validation and mechanistic dissection of the ubiquitin, cell adhesion and motility, and endocytosis pathways. To facilitate exploration of the entire screen data, we implemented our integrative analysis into a publicly available website. This resource combines the RNAi screen data from this study with a large collection of published genetic and epigenetic data, functional annotation analysis, gene interaction network analysis, and gene kinetics during reprogramming (http://song.igb.illinois.edu/ipsScreen/). Researchers are able to query, browse, and visualize this data collection.

The Ubiquitin Pathway Is a Barrier to Reprogramming

Fbxw7, an E3 ubiquitin ligase, was recently described as a barrier to mouse iPSC generation (Buckley et al., 2012). Our systems approach confirms and extends these findings to a much broader network of genes in human reprogramming. A variety of conjugating enzymes (UBE2D3, UBE2E3), ubiquitin ligases (RNF40, FBXW7, NEDD4, MARCH3), and deubiquitination enzymes (USP9X, OTUB2) are targeted in the TRA-1-81$^+$ population (Figure S3A and Data S2). Single-gene validation shows that knocking down UBE2D3, UBE2E3, or RNF40 individually significantly increases reprogramming efficiency compared to NS shRNA control, validating their functional role in reprogramming (Figure 3). We confirmed that iPSCs generated with RNF40 shRNAs express, in addition to TRA-1-81, the pluripotency markers NANOG, SSEA3, and SSEA4 at the protein level (Figure S3B) and express mRNAs of endogenous pluripotency factors at levels similar to those in hESCs (Figure S3C). We further show that RNF40i increases reprogramming efficiency without accelerating cell expansion rates (Figure S3D). Finally, we show that inhibition of multiple members of the ubiquitin pathway increases the protein levels of the reprogramming factor OCT4 (Figure S3E), in support of previous mouse studies (Buckley et al., 2012). These results reveal that the previously reported role of the ubiquitin pathway as a reprogramming barrier in mouse is conserved in human and demonstrate the power of our systems approach to expand upon the pathways identified.

Cell Adhesion and Motility in Reprogramming

Genes involved in the controlled turnover of actin filament needed for cell motility are enriched for screen hits (Figure 4A and Data S2). Actin, coronin, and RHOC, factors needed for filament assembly, branching, and disassembly, are all significant hits in the screen. Moreover, several of the screen hits, including MMP14, ADAM7, ADAM21, and ADAM29, have well-documented interactions with the extracellular matrix. Knockdown of ADAM7, ADAM21, or ADAM29 individually significantly increases human iPSC generation efficiency (Figure 3), with the most notable effect seen for ADAM29. We confirmed by immunofluorescence and qRT-PCR that iPSCs generated with ADAM29 shRNAs express high levels of several pluripotency markers, comparable to levels in hESCs (Figures 4B and 4C). In addition, ADAM29i does not increase iPSC generation efficiency by accelerating cell expansion rates (Figure 4D). ADAM29 knockdown at both mRNA and protein levels and the consequential enhancement of reprogramming efficiency were confirmed with two independent shRNAs (Figures S4A–S4C). Moreover, we found that overexpression of ADAM29 significantly reduces reprogramming efficiency (Figure 4E) and decreases hESC colony formation efficiency (Figure S4D), indicating that ADAM29 impedes both establishment and maintenance of the pluripotent state. To study the mechanism of ADAM29 as a reprogramming barrier, we mutated the metalloprotease and disintegrin domains of ADAM29 and overexpressed the altered proteins during reprogramming. Mutation of the disintegrin domain, but not the metalloprotease domain, abolishes the adverse effect of ADAM29 on reprogramming efficiency (Figures 4E and 4F). Remarkably, we found that adding a short synthetic peptide corresponding to the disintegrin loop of ADAM29 to the culture medium, but not a mutated peptide, significantly reduces iPSC colony formation efficiency (Figure 4G).

These data demonstrate that specific ADAM proteins are potent reprogramming barriers and that the disintegrin domain is necessary and sufficient for their activity during reprogramming. Disintegrin was originally identified from snake venom metalloprotease, which inhibits blood coagulation and is known to regulate integrin-dependent cell adhesion. Multiple reports document the interaction between the ADAM disintegrin domain and integrins, and the specificity of this interaction is mediated by a tripeptide motif in the disintegrin domain (Figure S4E) (Calvete et al., 2005; Igarashi et al., 2007). Interestingly, the screen hits that belong to the ADAM family, ADAM7, ADAM21, and ADAM29, all have an ECD tripeptide motif (Figure S4F), which can inhibit specific integrin dimers including those containing α6. During reprogramming, parental fibroblasts predominantly express integrin α5, which is not targeted by the ECD motif (Calvete et al., 2005), whereas iPSCs/ESCs express high levels of integrin α6 (Figure 4H). Integrin α6 is a component of the laminin receptor, and in turn laminin is important for maintaining self-renewal of hESCs and iPSCs (Rodin et al., 2010). Together, these findings suggest that ECD-containing ADAM proteins may antagonize an integrin switch that occurs during reprogramming to pluripotency, a model that deserves further exploration. Notably, both ADAM7 and ADAM29 are often found mutated in melanoma (Wei et al., 2011), suggesting that ADAM dysfunction may also contribute to cellular transformation in certain cancers.

Endocytosis Is a Barrier to Reprogramming

To obtain a global view of the negative effect size of the endocytosis pathway on reprogramming, we aggregated all screen-hit genes with endocytosis-related GO biological process annotations (Data S2). Remarkably, multiple screen hits were associated with a variety of components of the endocytic pathway: receptor-targeting ubiquitin ligases (MARCH3, RNF40, NEDD4); the mediator of vesicle transport, clathrin (CLTA); plasma-membrane remodeling and early-endosome fusion genes (EHD2, RABEP1); and endosomal/lysosomal surface proteins (MCOLN1, VPS25, ARSD, HSPA8, SCARB2, SLC17A5, DRAM1) (Figure 5A and Data S2). Interestingly, NEDD4 has been reported to ubiquitinate cell-surface receptors for lysosomal degradation (Persaud et al., 2011), suggesting a potential crosstalk between the endocytosis and ubiquitin pathways during reprogramming.

As validation of these findings, we found that knocking down DRAM1, SLC17A5, or ARSD individually with shRNAs leads to a significant increase in iPSC generation relative to NS shRNA control (Figure 3). We also used small molecules to inhibit the endocytosis pathway. We found that two clathrin-specific inhibitors, Pitstop1 and Pitstop2, can dramatically increase reprogramming efficiency, whereas structurally related controls (von Kleist et al., 2011) have no effect (Figure 5B). Furthermore, Pitstop2 can also increase iPSC generation efficiency when episomal vectors are used to reprogram human dermal fibroblasts (HDFs) (Figure S5A). Therefore, the effect of Pitstop is independent of the retroviral infection used in standard reprogramming methods or specific fibroblast lines. We confirmed that iPSCs generated with Pitstop2 express mRNA and protein levels of pluripotency markers similar to those of hESCs (Figure 5D). In addition, Pitstop2 does not affect cell expansion rates during iPSC generation (Figure S5B). These data demonstrate that clathrin-mediated endocytosis is a barrier to reprogramming. Inhibitors of endocytosis may thus be added to the panel of small molecules that enhance reprogramming efficiency.

Endocytosis functions as a master organizer of signaling circuits by modulating the abundance of receptors, their ligands, and downstream effectors at different membrane compartments. We sought to further probe the mechanism by which endocytosis acts as a reprogramming barrier. Treatment of cells with Pitstop2 at different time points post-infection with 4F revealed that endocytosis acts at an early stage of reprogramming (Figure 5E). During that stage, reprogramming cells must undergo a mesenchymal-epithelial transition (MET) (Samavarchi-Tehrani et al., 2010; Subramanyam et al., 2011), and we found that Pitstop2 upregulates the epithelial marker E-cadherin at day 12 of reprogramming (Figures S5C and S5D). Of note, it has been shown that clathrin-mediated endocytosis positively regulates TGF-β signaling by recycling the receptors (Scita and Di Fiore, 2010), and TGF-β signaling is a known barrier to reprogramming that inhibits MET (Samavarchi-Tehrani et al., 2010; Subramanyam et al., 2011). In support of a role for the endocytosis pathway in TGF-β signaling, we found that Pitstop2 inhibits phosphorylation and activation of SMAD2/3, in particular SMAD2, during reprogramming (Figure S5C). Using GeneMANIA, we found that the endocytosis and TGF-β signaling pathways can interact at multiple levels (Figure S5E). Members of the TGF-β pathway, such as MAPK3 and SMAD2, are also

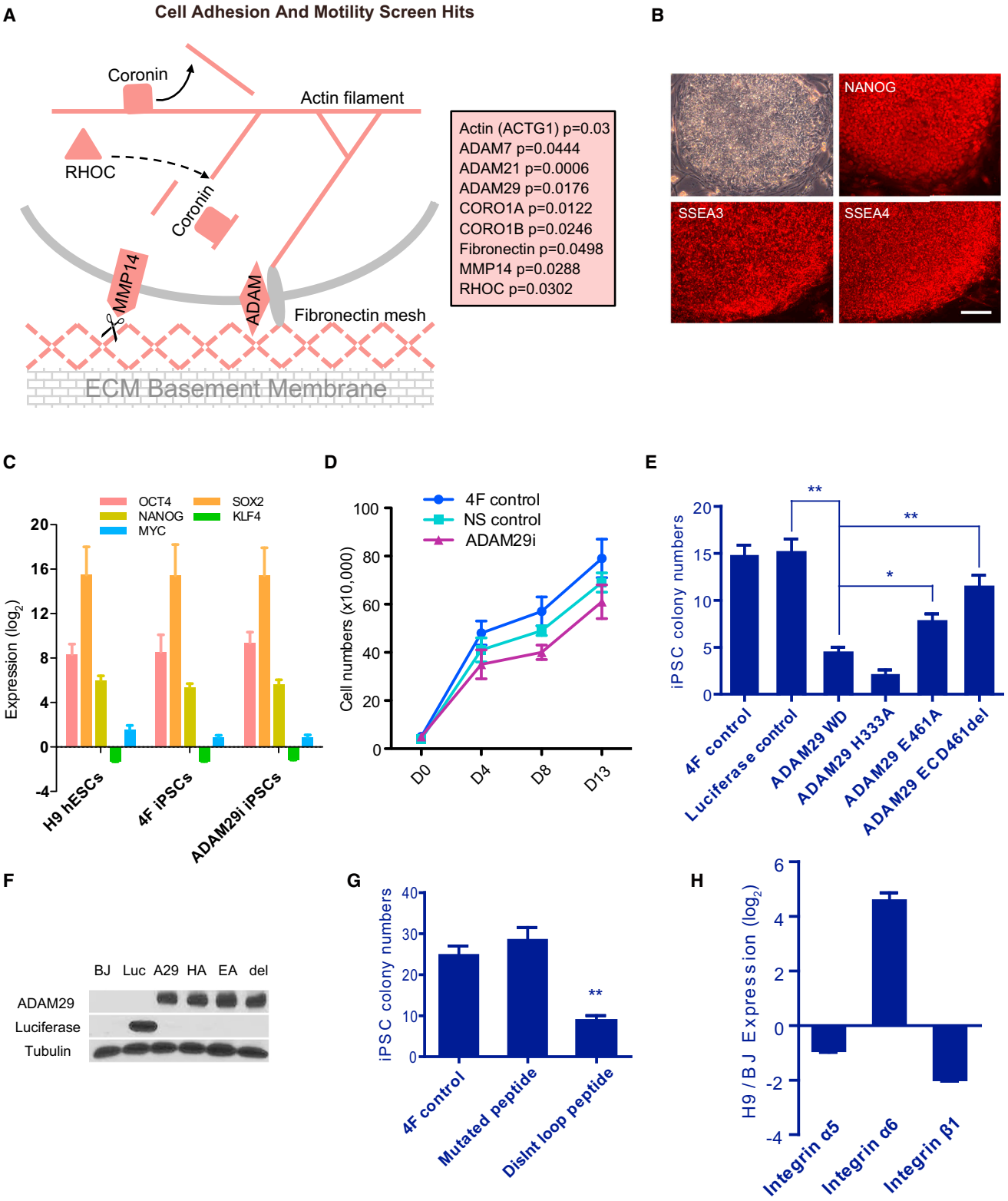


Figure 4. ADAMs Are a Barrier to Reprogramming and Act via Their Disintegrin Loop Domain

(A) Screen hits at the 5% significance level with GO biological process annotations related to cell motility and adhesion are highlighted.

(B) 4F + ADAM29i-induced iPSCs show strong positive immunostaining for pluripotency markers NANOG, SSEA3, and SSEA4. Scale bar, 300 μm.

(legend continued on next page)

screen hits as reprogramming barriers (Figure S5E). Importantly, in the presence of the TGF-β inhibitor SB431542, Pitstop2 does not further decrease the level of phosphorylated SMAD2/3 (Figure 5F), nor does it further increase iPSC generation efficiency (Figure 5G). Taken together, these data indicate that clathrin-mediated endocytosis antagonizes reprogramming via a linear positive interaction with the TGF-β signaling pathway.

Given the observation that inhibition of endocytosis is unable to synergize with inhibition of TGF-β signaling to increase reprogramming efficiency, we tested the effect of combining Pitstop2 with RNAi for members of other barrier pathways. In contrast to TGF-β signaling, we found that Pitstop2 can further enhance the iPSC generation efficiency of cells treated with UBE2E3i, RNF40i, ADAM29i, SLC17A5i, PTPN11i, ATF7IPi, or TTF2i (Figure 5H). The interaction of Pitstop2 with the lysosomal protein SLC17A5 (also known as SIALIN) suggests that the lysosomal subpathway could have additional effects on reprogramming unrelated to TGF-β signaling. Remarkably, when ADAM29i, ATF7IPi, and Pitstop2 are combined together, reprogramming efficiency can be elevated by up to 15-fold (Figure 5H). These data indicate that endocytosis can act in parallel with other barriers, including the ubiquitination and ADAM pathways, to antagonize reprogramming.

Identification of Feedforward Loops between Reprogramming Barriers

To further explore potential interactions between validated reprogramming barriers, we searched the literature for experiments where one of the barrier genes significantly changed at the transcriptional level (t test p value < 0.05) after RNAi against a second barrier gene. Using GeneMANIA, we also aggregated evidence of tissue-specific coexpression of barrier genes and identified common barrier-gene protein domains. In the resulting barrier gene interaction network (Figure 6A, thin edges), RNF40 stands out as a hub (i.e., it has both the highest "betweenness" and "closeness" centralities). We verified these interactions in human BJ fibroblasts and found that RNF40i induces both positive and negative effects on the expression of other barrier genes. RNF40i leads to significant increases in the expression of ADAM29, PTPN11, TTF2, TMF1, and MED19. RNF40i also induces a significant downregulation of SLC17A5, PTPRJ, and CENPB (Figure 6A, thick arrows and 6B). Transcriptional effects of RNF40i likely result from its role in regulating the levels of transcription factors such as OCT4 (Figure S3E) and in ubiquitinating H2B, a histone mark involved in transcriptional elongation and stem cell differentiation (Karpiuk et al., 2012). Thus, inhibition of a reprogramming barrier can alter the expression of other barriers in unexpected ways. We next tested the effect of combining RNF40i with RNAi for other barriers with which RFN40 interacts at the expression level (Figure 6B). RNF40i together with PTPN11i, MED19i, SLC17A5i, or PTPRJi produces combinatorial effects on reprogramming efficiency (Figure 6C). These data suggest that different barrier pathways interact in dynamic feedforward loops (FFLs) that can be classified as either coherent or incoherent and are thought to provide noise filtering and response delay capabilities to stabilize cell types (Mangan and Alon, 2003). As examples of coherent FFLs with an OR-gate integration, RNF40 positively regulates the other barrier genes SLC17A5 and PTPRJ. This kind of network motif can stabilize cells against transient fluctuating signals and can delay the response to removing the source repressor, which corresponds to RNF40 in our case. Thus, even though knockdown of RNF40 attenuates the second barrier's expression, the residual amount of the second barrier can still partially suppress reprogramming without RNF40; as a result, simultaneous RNAi against the second barrier itself drastically relieves the suppression of reprogramming by the FFL as a whole (Figure 6D). Incoherent FFLs with an OR-gate integration have not been studied extensively to date, but we found that such network motifs can also inhibit reprogramming. For example, RNF40 represses the other barrier genes PTPN11 and MED19, so that knockdown of RNF40 alone aberrantly upregulates PTPN11 and MED19, acting as a built-in compensatory suppression mechanism against reprogramming. In this case, a persistent knockdown of RNF40 may only transiently lower the barrier against reprogramming because the subsequent derepression of PTPN11 and MED19 can help raise back the barrier. However, the dampening of the increase in reprogramming efficiency that could result from this derepression of a second barrier can be relieved via knockdown of the second gene (Figure 6D). The FFLs with OR-gate integration thus provide partial redundancy in cell-fate control, and removing the redundancy with multiple RNAi manifests itself in the observed combinatorial knockdown effect (Figure 6C). Barriers can of course interact in multiple ways, not just at the transcriptional level. Nevertheless, the FFLs identified here contribute to revealing some of the complex interactions between barrier pathways. Taken together, our genetic interaction studies reveal that barrier pathways can act in linear, parallel (Figure 5), or FFL (Figure 6) architectures to antagonize reprogramming and show that iPSC generation efficiency can be greatly increased by simultaneous targeting of different reprogramming barriers.

PERSPECTIVE

We use an ultracomplex EXPANDed shRNA library to report the application of a genome-wide screen to the process of human cellular reprogramming to pluripotency. Our integrative study

(C) 4F + ADAM29i-induced iPSCs express endogenous pluripotency markers at similar levels to 4F iPSCs and ESCs.

(D) Growth curves of fibroblasts infected with 4F, 4F + nonsense (NS), and 4F + RNF40i, counted on days 0, 4, 8, and 13 post-infection. RNF40i does not substantially alter total cell numbers during the first 13 days of reprogramming. Error bars represent SD.

(E) ADAM29 overexpression reduces reprogramming efficiency, and the disintegrin domain is critical for the barrier function of ADAM29 in reprogramming. BJ fibroblasts overexpressing wild-type (WD) or mutated ADAM29 (H333A in the active site of the metalloprotease domain; E461A and ECD461 deletion in the disintegrin domain) were induced to pluripotency by 4F.

(F) Overexpression of wild-type or mutated ADAM29 was confirmed by western blotting. Tubulin was used as loading control.

(G) ADAM29 disintegrin loop peptide reduces reprogramming efficiency. Human BJ fibroblasts were infected with 4F. Four days later, a synthetic ADAM29 disintegrin loop peptide (100 μg/ml) was added. A peptide with substitutions of the three critical residues (ECD) was used as controls.

(H) hESCs H9 express higher levels of integrin α6 and lower levels of integrin α5 and β1 compared with BJ fibroblasts.

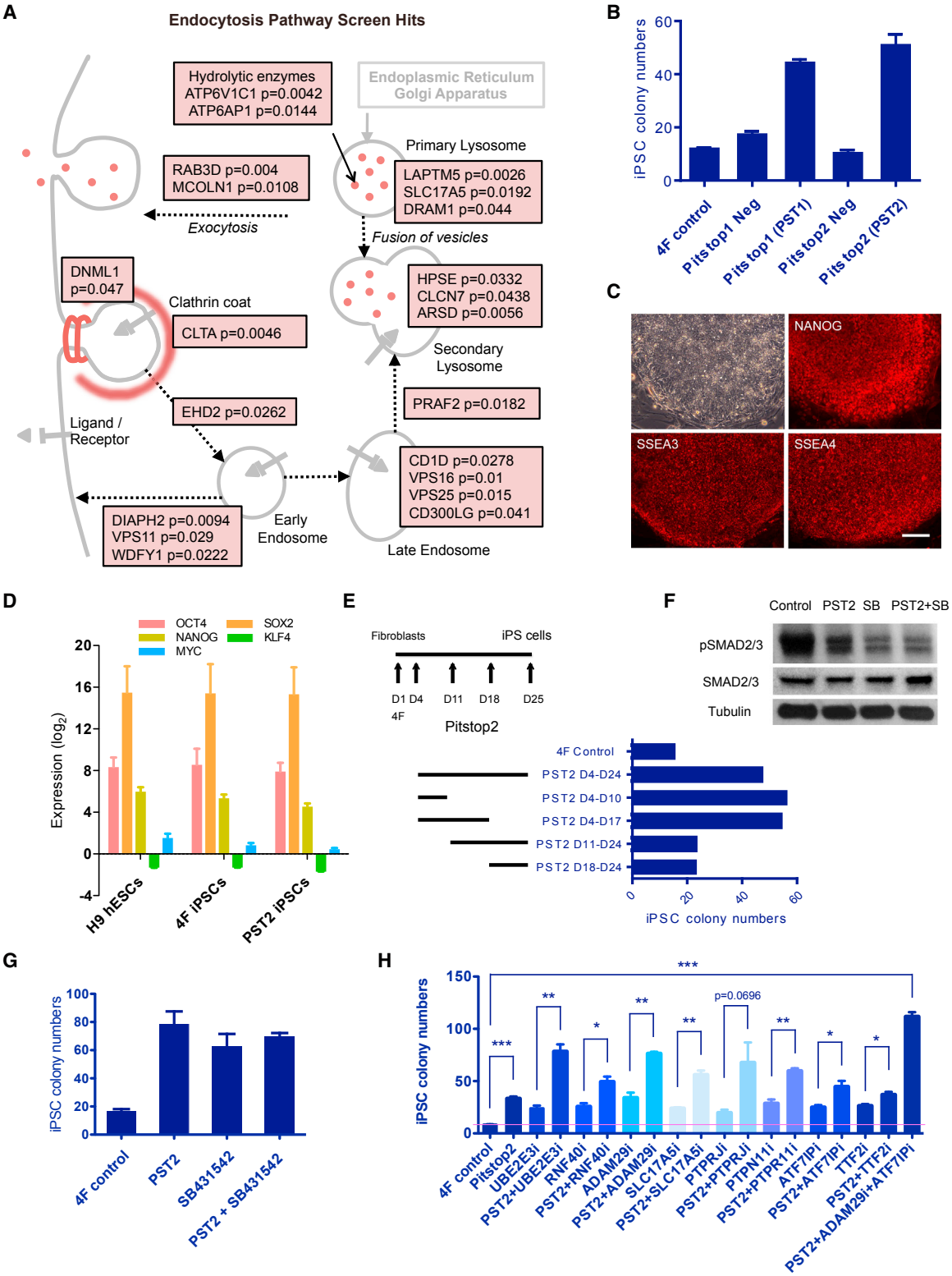


Figure 5. Endocytosis Is an Early Acting Barrier to Reprogramming that Enhances TGF-β Signaling

(A) Screen hits at the 5% significance level with GO biological process annotations related to endocytosis and lysosome are highlighted.

(B) Inhibition of clathrin-dependent endocytosis by small molecules Pitstop1 and Pitstop2 increases human iPSC generation efficiency. Human BJ fibroblasts were infected with 4F. Four days later, clathrin inhibitors (Abcam) were added to medium at a final concentration of 45 μM (Pitstop1) and 30 μM (Pitstop2) until

(legend continued on next page)

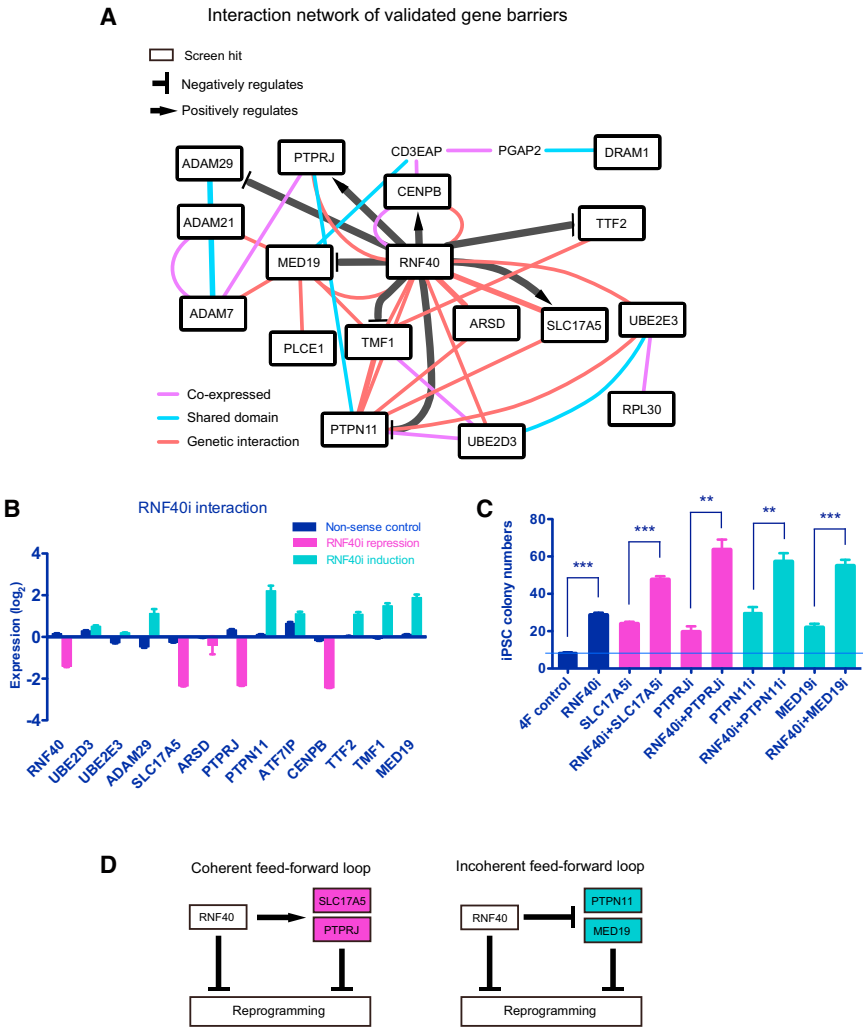


Figure 6. Genetic Interactions between RNA40 and Other Reprogramming Barriers

(A) In this barrier-gene network, nondirectional interactions (colored edges) were identified via a literature search, and genetic interactions were inferred from knockdown experiments. We consider gene A to interact with gene B if B's expression changes significantly (t test p value < 0.05) after the knockdown of A (red edges). Genes sharing common protein domains (blue edges), and genes that are coexpressed in the same tissue type (purple edges), were aggregated via GeneMANIA. Directional interactions (thick black edges) were identified in human fibroblasts from knockdown experiments shown in (B).

(B) RNF40i affects the expression of other barrier genes. qRT-PCR was performed in BJ fibroblasts infected with nonsense (NS) shRNA control or shRNAs against RNF40.

(C) The net effects on reprogramming efficiency of simultaneous knockdown of RNA40 and each of the other barrier genes shown exceed the effect of knocking down either gene alone.

(D) RNF40 forms coherent and incoherent feed-forward loops in the barrier-gene network.

reveals multiple pathways as reprogramming barriers, including ubiquitination, endocytosis, cell adhesion, interaction with extracellular matrix, dephosphorylation, and transcription. Moreover, we found that different barrier pathways interact in complex combinatorial ways to antagonize reprogramming to the iPSC state. The data are implemented as a fully searchable integrative online resource to facilitate other studies of reprogramming.

Our genome-wide RNAi screen provides a systematic approach to interrogate iPSC generation. Several key aspects underlie the successful identification of the reprogramming barriers reported here: (1) the high (30 shRNAs) per-gene coverage of the library reduces off-target effects in hit selection; (2) using NGS for screen readout provides a large dynamic range; (3) a multiobjective optimization algorithm with greater sensitivity and specificity, and less prone to off-target effects than existing methods, helps reduce false-positive hits; 4) systems-level analyses combining large-scale integration of public data with independent functional validation assays elucidate functional pathways. Benefiting from these aspects, our screen has a scale of 600,000 shRNAs tested and a hit validation rate of 87%, whereas previous genome-wide shRNA/siRNA screens tend to have lower scales (average < 40,000) and average < 50% success rates (Table S2).

More generally, our work provides a model for combining a genome-wide screen with large-scale data mining in contexts other than iPSC generation. Our preliminary findings indicate that some of the pathways identified may represent barriers to the induction of stem-like phenotypes in other cellular and disease processes. We applied our deconvolution method to a genome-wide screen for barriers against the proliferation of

colonies were counted. Pitstop1 Neg and Pitstop2 Neg, compounds structurally related to Pitstop1 and Pitstop2, respectively, but lacking the ability to inhibit endocytosis (von Kleist et al., 2011), were used as controls.

(C) 4F + Pitstop2 iPSCs show strong immunostaining for pluripotency markers NANOG, SSEA3, and SSEA4. Scale bar, 300 μm.

(D) 4F + Pitstop2 iPSCs express endogenous pluripotency markers at levels similar to those of 4F iPSCs and ESCs.

(E) Endocytosis acts as a barrier at an early stage of reprogramming. Human BJ fibroblasts were infected with 4F. Pitstop2 was added to the medium at the time points indicated.

(F) Pitstop2 and TGF-β inhibitor SB431542 (SB, 2 μM) lead to a decrease in pSMAD2/3 levels by western blotting. Tubulin was used as loading control.

(G) Both Pitstop2 and SB431542 (days 2–14) increase human iPSC generation efficiency. However, when combined together, no additional increase is observed.

(H) Pitstop2 further enhances reprogramming efficiency upon knockdown of multiple other barriers.

ovarian carcinoma stem cells (Tan et al., 2013). Cross-referencing our iPSC screen hits with the hits in carcinoma stem cells found 35 overlapping genes at the 5% significance level, and DAVID functional analysis identified transcription factors, cell membrane proteins, and ubiquitin pathway genes to be enriched in these hits (Figure S6). Of note, RNF40 was detected in both screens. Also enriched in both screens is arginine methyltransferase PRMT1, previously discussed as part of the FBXW7 interaction network (Figure 2B). Thus, our platform provides an integrative approach for identifying pathways that may act as barriers beyond the setting of reprogramming to pluripotency, including in cancer. We further anticipate that this approach will be useful in the dissection of direct lineage reprogramming and may reveal shared and unique aspects of different reprogramming paradigms.

EXPERIMENTAL PROCEDURES

Pooled shRNA Libraries

The design method for generating ultracomplex EXPANDed shRNA libraries targeting the human genome has been previously described (Bassik et al., 2009, 2013). In brief, the library contains approximately 30 independent shRNAs per gene for all human protein-coding genes (19,527 in total), for a total of ~600,000 shRNAs, hence the term "ultracomplex." The full list of shRNA sequences present in the library is available upon request. Pooled sequences coding for the shRNAs were cloned downstream of a U6 promoter in a modified pSicoR lentiviral vector containing a CMV-Puro-T2A-mCherry cassette. All vectors were pooled to generate one lentiviral library representing the 600,000 shRNAs. The entire pooled library was then used to generate lentiviruses at the UCSF ViraCore.

RNAi Screen in Human iPSC Generation

Human primary newborn foreskin (BJ) fibroblasts were seeded at 630,000 cells per 15 cm dish the day before infection. Sixteen 15 cm dishes were used in total for the whole library screen, to represent 50 cells per shRNA. Cells were infected with concentrated retroviruses (Harvard Gene Therapy Initiative) leading to the overexpression of OCT4, SOX2, and KLF4 (multiplicity of infection [moi] = 10) and c-MYC (moi = 1) (4F), in combination with a lentivirus expressing an shRNA against p53 (moi = 1) (Zhao et al., 2008) and the lentiviral shRNA library (moi = 3). On day 28 after infection, cells were trypsinized, live TRA-1-81 (MAB4381, Millipore) staining was performed, and TRA-1-81$^+$ and TRA-1-81$^-$ populations were isolated by FACS (FACSAria II, BD).

Bioinformatic Analyses

TRA-1-81$^+$/TRA-1-81$^-$ differential shRNA effect size was estimated by read-count odds-ratio. shRNA effect sizes were summarized gene wise using a random effects model. Gene ranking was performed using multiobjective optimization, with collective shRNA effect size and the number of distinct shRNA as criteria. Significance was assessed using a permutation test, and library swap was used to control for false discovery in multiple hypothesis testing (Extended Experimental Procedures). Functional annotations were analyzed via DAVID (Huang et al., 2009a), and gene interaction networks constructed via GeneMANIA (Zuberi et al., 2013) and Cytoscape. Betweenness centrality is defined as the number of shortest paths through the network that pass through a given node. Closeness centrality is defined as the inverse of the sum of all shortest-path distances, from a given node, to all other nodes in the network.

Single-Gene Validation

For gene knockdown, the two most enriched shRNAs against each gene from the screen data were cloned into the lentiviral vector pSicoR downstream of a U6 promoter. Each shRNA was cloned into a separate pSicoR-CMV-Puro-T2A-GFP lentiviral vector. Viruses inducing the expression of two shRNAs against the same gene were combined together with 4F before infecting cells, unless otherwise indicated. Table S3 includes the target sequences of all shRNAs used for follow-up experiments. For gene overexpression, cDNA was cloned into the lentiviral vector pGAMA downstream of an EF1α promoter. Mutations were generated using QuickChange (Agilent) according to the manufacturer's instructions. On days 20–28 after infection, live Tra-1-81 (MAB4381, Millipore) staining was performed, and iPSC colonies were counted. All reprogramming experiments were performed in triplicates, and error bars represent SD. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, four tables, and two data files and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2014.05.040.

AUTHOR CONTRIBUTIONS

M.R.-S. and J.S.S. directed the project. H.Q. and M.R.-S. developed the experimental design with participation from R.J.L. and M.T.M; R.J.L., E.M.L., and M.T.M. developed and produced the RNAi libraries; H.Q., L.B., and P.T. carried out the bench experiments; A.D. and J.S.S. developed and carried out the statistical and bioinformatic analyses, with initial input from W.P.; H.Q., A.D., J.S.S., and M.R.-S. wrote the manuscript.

ACKNOWLEDGMENTS

We are grateful to Robert Blelloch, Keith Mostov, and Zena Werb for critical reading of the manuscript, Jonathan Weissman, Martin Kampann, and Michael Bassick for discussions and help with library generation, Nikki Shariat for input on the experiment design, Xiaoyin Wang for virus production, Fong Ming Koh for providing the pGAMA vector, Charles Nicolet and Henriette O'Geen for support on NGS, and Tara Rambaldo and Michael Kissner for support with FACS. This work was supported by NWO Rubicon grant 825.06.030 and Veni grant 916.10.138 to R.J.L.; the UCSF Program for Breakthrough Biomedical Research and NIH grants 1U01CA168370 and R01 GM80783 to M.T.M.; NIH R01CA163336 and Sontag Foundation Distinguished Scientist Award to J.S.S.; and CIRM grant RB4-06028 and NIH Director's New Innovator Award DP2OD4698 to M.R.-S. E.M.L. is an employee of Agilent Technologies.

Received: September 5, 2013
Revised: March 5, 2014
Accepted: May 27, 2014
Published: July 17, 2014

REFERENCES

Bassik, M.C., Lebbink, R.J., Churchman, L.S., Ingolia, N.T., Patena, W., LeProust, E.M., Schuldiner, M., Weissman, J.S., and McManus, M.T. (2009). Rapid creation and quantitative monitoring of high coverage shRNA libraries. Nat. Methods *6*, 443–445.

Bassik, M.C., Kampmann, M., Lebbink, R.J., Wang, S., Hein, M.Y., Poser, I., Weibezahn, J., Horlbeck, M.A., Chen, S., Mann, M., et al. (2013). A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. Cell *152*, 909–922.

Buckley, S.M., Aranda-Orgilles, B., Strikoudis, A., Apostolou, E., Loizou, E., Moran-Crusio, K., Farnsworth, C.L., Koller, A.A., Dasgupta, R., Silva, J.C., et al. (2012). Regulation of pluripotency and cellular reprogramming by the ubiquitin-proteasome system. Cell Stem Cell *11*, 783–798.

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell *150*, 1209–1222.

Calvete, J.J., Marcinkiewicz, C., Monleón, D., Esteve, V., Celda, B., Juárez, P., and Sanz, L. (2005). Snake venom disintegrins: evolution of structure and function. Toxicon *45*, 1063–1074.

Chen, J., Liu, H., Liu, J., Qi, J., Wei, B., Yang, J., Liang, H., Chen, Y., Chen, J., Wu, Y., et al. (2013). H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. Nat. Genet. *45*, 34–42.

Chia, N.-Y., Chan, Y.-S., Feng, B., Lu, X., Orlov, Y.L., Moreau, D., Kumar, P., Yang, L., Jiang, J., Lau, M.-S., et al. (2010). A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature *468*, 316–320.

Daley, G.Q. (2008). Common themes of dedifferentiation in somatic cell reprogramming and cancer. Cold Spring Harb. Symp. Quant. Biol. *73*, 171–174.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987–1000.

Golipour, A., David, L., Liu, Y., Jayakumaran, G., Hirsch, C.L., Trcka, D., and Wrana, J.L. (2012). A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. Cell Stem Cell *11*, 769–782.

Gregory, P.A., Bert, A.G., Paterson, E.L., Barry, S.C., Tsykin, A., Farshid, G., Vadas, M.A., Khew-Goodall, Y., and Goodall, G.J. (2008). The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nat. Cell Biol. *10*, 593–601.

Gurdon, J.B., Elsdale, T.R., and Fischberg, M. (1958). Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. Nature *182*, 64–65.

Handl, J., Kell, D.B., and Knowles, J. (2007). Multiobjective optimization in bioinformatics and computational biology. IEEE/ACM Trans. Comput. Biol. Bioinformatics *4*, 279–292.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. *37*, 1–13.

Igarashi, T., Araki, S., Mori, H., and Takeda, S. (2007). Crystal structures of catrocollastatin/VAP2B reveal a dynamic, modular architecture of ADAM/adamalysin/reprolysin family proteins. FEBS Lett. *581*, 2416–2422.

International Stem Cell Initiative, Adewumi, O., Aflatoonian, B., Ahrlund-Richter, L., Amit, M., Andrews, P.W., Beighton, G., Bello, P.A., Benvenisty, N., Berry, L.S., Bevan, S., et al. (2007). Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. Nat. Biotechnol. *25*, 803–816.

Karpiuk, O., Najafova, Z., Kramer, F., Hennion, M., Galonska, C., König, A., Snaidero, N., Vogel, T., Shchebet, A., Begus-Nahrmann, Y., et al. (2012). The histone H2B monoubiquitination regulatory pathway is required for differentiation of multipotent stem cells. Mol. Cell *46*, 705–713.

Kawamura, T., Suzuki, J., Wang, Y.V., Menendez, S., Morera, L.B., Raya, Á., Wahl, G.M., and Izpisúa Belmonte, J.C. (2009). Linking the p53 tumour suppressor pathway to somatic cell reprogramming. Nature *460*, 1140–1144.

König, R., Chiang, C.-Y., Tu, B.P., Yan, S.F., DeJesus, P.D., Romero, A., Bergauer, T., Orth, A., Krueger, U., Zhou, Y., and Chanda, S.K. (2007). A probability-based approach for the analysis of large-scale RNAi screens. Nat. Methods *4*, 847–849.

Li, Z., Yang, C.-S., Nakashima, K., and Rana, T.M. (2011). Small RNA-mediated regulation of iPS cell generation. EMBO J. *30*, 823–834.

Luo, B., Cheung, H.W., Subramanian, A., Sharifnia, T., Okamoto, M., Yang, X., Hinkle, G., Boehm, J.S., Beroukhim, R., Weir, B.A., et al. (2008). Highly parallel identification of essential genes in cancer cells. Proc. Natl. Acad. Sci. USA *105*, 20380–20385.

Mangan, S., and Alon, U. (2003). Structure and function of the feed-forward loop network motif. Proc. Natl. Acad. Sci. USA *100*, 11980–11985.

Ohi, Y., Qin, H., Hong, C., Blouin, L., Polo, J.M., Guo, T., Qi, Z., Downey, S.L., Manos, P.D., Rossi, D.J., et al. (2011). Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. Nat. Cell Biol. *13*, 541–549.

Onder, T.T., Kara, N., Cherry, A., Sinha, A.U., Zhu, N., Bernt, K.M., Cahan, P., Marcarci, B.O., Unternaehrer, J., Gupta, P.B., et al. (2012). Chromatin-modifying enzymes as modulators of reprogramming. Nature *483*, 598–602.

Persaud, A., Alberts, P., Hayes, M., Guettler, S., Clarke, I., Sicheri, F., Dirks, P., Ciruna, B., and Rotin, D. (2011). Nedd4-1 binds and ubiquitylates activated FGFR1 to control its endocytosis and function. EMBO J. *30*, 3259–3273.

Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., et al. (2012). A molecular roadmap of reprogramming somatic cells into iPS cells. Cell *151*, 1617–1632.

Qin, H., Blaschke, K., Wei, G., Ohi, Y., Blouin, L., Qi, Z., Yu, J., Yeh, R.F., Hebrok, M., and Ramalho-Santos, M. (2012). Transcriptional analysis of pluripotency reveals the Hippo pathway as a barrier to reprogramming. Hum. Mol. Genet. *21*, 2054–2067.

Ramalho-Santos, M. (2009). iPS cells: insights into basic biology. Cell *138*, 616–618.

Rodin, S., Domogatskaya, A., Ström, S., Hansson, E.M., Chien, K.R., Inzunza, J., Hovatta, O., and Tryggvason, K. (2010). Long-term self-renewal of human pluripotent stem cells on human recombinant laminin-511. Nat. Biotechnol. *28*, 611–615.

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H.-K., Beyer, T.A., Datti, A., Woltjen, K., Nagy, A., and Wrana, J.L. (2010). Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. Cell Stem Cell *7*, 64–77.

Scita, G., and Di Fiore, P.P. (2010). The endocytic matrix. Nature *463*, 464–473.

Subramanyam, D., Lamouille, S., Judson, R.L., Liu, J.Y., Bucay, N., Derynck, R., and Blelloch, R. (2011). Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. Nat. Biotechnol. *29*, 443–448.

Tada, M., Tada, T., Lefebvre, L., Barton, S.C., and Surani, M.A. (1997). Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. EMBO J. *16*, 6510–6520.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

Tan, T.Z., Miow, Q.H., Huang, R.Y.-J., Wong, M.K., Ye, J., Lau, J.A., Wu, M.C., Bin Abdul Hadi, L.H., Soong, R., Choolani, M., et al. (2013). Functional genomics identifies five distinct molecular subtypes with clinical relevance and pathways for growth control in epithelial ovarian cancer. EMBO Mol Med *5*, 983–998.

von Kleist, L., Stahlschmidt, W., Bulut, H., Gromova, K., Puchkov, D., Robertson, M.J., MacGregor, K.A., Tomilin, N., Pechstein, A., Chau, N., et al. (2011). Role of the clathrin terminal domain in regulating coated pit dynamics revealed by small molecule inhibition. Cell *146*, 471–484.

Wei, X., Moncada-Pazos, A., Cal, S., Soria-Valles, C., Gartner, J., Rudloff, U., Lin, J.C., Rosenberg, S.A., López-Otín, C., and Samuels, Y.; NISC Comparative Sequencing Program (2011). Analysis of the disintegrin-metalloproteinases family reveals ADAM29 and ADAM7 are often mutated in melanoma. Hum. Mutat. *32*, E2148–E2175.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917–1920.

Zhao, Y., Yin, X., Qin, H., Zhu, F., Liu, H., Yang, W., Zhang, Q., Xiang, C., Hou, P., Song, Z., et al. (2008). Two supporting factors greatly improve the efficiency of human iPSC generation. Cell Stem Cell *3*, 475–479.

Zuberi, K., Franz, M., Rodriguez, H., Montojo, J., Lopes, C.T., Bader, G.D., and Morris, Q. (2013). GeneMANIA prediction server 2013 update. Nucleic Acids Res. *41* (Web Server issue), W115–W122.