# Systematic Identification of Barriers to Human iPSC Generation

## Inventory of Supplemental Information

**Supplemental Data** – Figure S1-6, including legends; Table S1-4; Extended Experimental Procedures; Supplemental References.

**File S1** – Genes with at least one active shRNA, ranked by enrichment in TRA-1-81+ population, related to Figure 1.

**File S2** – Putative reprogramming barriers with ubiquitination, endocytosis, and cell adhesion/motility gene ontology annotations, related to Figure 2.

**SUPPLEMENTAL FIGURE LEGENDS**

**Figure S1: A Pilot Screen Shows That p53i Increases Signal from the TRA-1-81+ Population, Related to Figure 1**

(A) Since p53i has been shown to significantly enhance reprogramming efficiency (Zhao et al., 2008), we hypothesized that 4F along with p53i would yield a stronger signal from the TRA-1-81+ subpopulation than without p53i, and hence a more robust screen. To test this hypothesis, as well as provide a proof of concept, we first performed two pilot screens using a library of 55,000 shRNAs, targeting 1826 genes. We co-infected human BJ fibroblasts with lentivirus expressing these shRNAs along with 4F. In one of the two screens we included p53i. Using the same procedure as in the genome-wide screen (described in the main text), we compared the pilot screens with and without p53i. We found that p53i greatly increases signal from the TRA-1-81+ population, in the following sense: (1) as described in Methods, we used a random effects model to measure the collective effect size ($\delta$) of knocking down a gene on reprogramming; the use of p53i significantly shifts the distribution of gene effect sizes to the right, indicating that the screening sensitivity is significantly increased in the TRA-1-81+ population with p53i (Wilcoxon rank-sum test $p=9.83 \times 10^{-237}$). (2) The number of genes with zero observed effect size decreased considerably.

(B) When comparing screen hits called by our multi-objective optimization algorithm at the 5% significance level to existing methods, such as RIGER and RSA, we find that HitSelect's sensitivity on validated genes is 90% (data not shown), compared to 85% for RIGER and 0% for RSA. RIGER fails to detect key genes such as ADAM7, which contains the disintegrin loop motif that we demonstrate drastically decreases reprogramming efficiency. While RIGER has decent sensitivity, the false discovery rate is unacceptably high. Consider the left panel: it shows the distribution of gene ranks for positive and negative control genes described in the manuscript. Note that negative control sequences actually show positive enrichment under RIGER, with some even ranked within the top 5% (bottom left panel). In addition, both RIGER and RSA enrich for genes with only 1 active shRNA. Thus, these methods are prone to off-target effects. Consider the right panel: this shows the number of active shRNA for genes ranked in the top 1% by each algorithm. Both RIGER and RSA have large numbers of

"singleton" genes supported by only 1 shRNA, while all of our top genes have at least 2 supporting shRNAs and most have at least 3. Therefore multi-objective optimization has greater sensitivity and specificity, and is less prone to off-target effects than existing methods for screen hit selection.

**Figure S2: The Kinetics of Barrier Genes in iPSC-competent Cells during Mouse Reprogramming, Related to Figure 2**

(A) We examined the mouse orthologs of our TRA-1-81+ screen hits (at the 5% significance level) in the gene expression and epigenetic data of (Polo et al., 2012). Polo *et al.* profiled secondary MEFs at different stages during reprogramming. On days 3, 6, 9, 12, and 15, they FACS-purified cells based on Thy1, SSEA-1, and exogenous Oct4-GFP expression, in order to enrich for potential iPSCs. On those days, they assayed gene expression by microarray. Additionally, the genome-wide patterns of H3K4me3 and H3K27me3 modifications were identified on days 3, 9, 12, and 15 via ChIP-seq. We used their epigenetic data to classify putative gene promoter regions as either H3K4me3-enriched, H3K27me3-enriched, bivalent, or non-enriched. Classification was performed using a 4-state hidden Markov model (Experimental Procedures). We then triaged genes into 6 classes by applying k-means clustering to the gene expression data. (1) In each cluster, the top panel shows a histogram of gene promoter epigenetic states. The percentages of genes that are within a cluster and that have a given epigenetic state, out of all genes having that epigenetic state regardless of cluster, are displayed. This percentage is the probability that a randomly chosen gene belongs to a particular cluster, given its epigenetic state and the day. Epigenetic state probabilities for day 6 were obtained by linearly interpolating the values from day 3 and day 9. We found gene expression changes to be consistent with epigenetic modifications. Repressed clusters (cluster 2, 3 and 4) predominantly contain genes that lose H3K4me3 and gain H3K27me3, and the converse is true for activated clusters (clusters 1, 5 and 6). The bottom panel displays the gene expression profile for all genes in the cluster (grey), that for screen hits (maroon), and the average profile over all screen hits (red). (2) We selected genes from the down-regulated clusters 2, 3 and 4. Then we restricted ourselves to genes whose promoters were being epigenetically silenced, as measured by a loss of H3K4me3 or a gain of H3K27me3 during

reprogramming. We show the functional annotations related to cell adhesion, motility, and extracellular matrix, for these attenuated genes. Full annotation lists for all clusters can be found in our online resource (http://songlab.ucsf.edu/ipsScreen/docs/gexpClust.html)

(B) (Golipour et al., 2012) used a similar mouse secondary system to perform RNA-seq during reprogramming. MEFs were reprogrammed with 4F expressed from doxycycline inducible vectors. On 14, 21, 22, 23, and 27 days post induction, doxycycline was withdrawn and cells expressing markers of pluripotency were identified via AP-staining and Nanog immunofluorescence. Cells expressing these pluripotency markers were labeled “Stabilization competent” (SC). Cells not expressing these markers were labeled “Stabilization incompetent” (SI). RNA-seq was performed on both SC and SI populations, on each of these days. We identified genes which were down regulated in SC cells, from the RNA-seq data, using a single-tailed t-test with a p-value cutoff of p=0.01. (1) A histogram of the number of genes down regulated in SC compared to SI on a given day. The number of genes differentially down regulated in SC is highest on day 23. (2) Functional annotation clusterings of genes down regulated in SC on day 23, performed with DAVID. SC genes enrich for endocytosis and vesicle mediated protein transport annotations.

(C) Gene expression for barrier-genes was measured during the reprogramming of human BJ fibroblasts, using RT-qPCR. Then k-means clustering was applied to the log2 transformed fold-change over human BJ fibroblast expression profiles. We observe patterns of barrier-gene activation similar to those observed in mouse data: 1) a cluster of transiently expressed barrier-genes, similar to mouse cluster 1; 2) a cluster of gradually decreasing barrier-genes, similar to mouse cluster 3; 3) a cluster of barrier-genes activating late in reprogramming, similar to cluster 5; and 4) a cluster of gradually increasing barrier-genes, similar to cluster 6.

**Figure S3: Ubiquitination Is a Barrier to Reprogramming, Related to Figure 3**

(A) Screen hits at the 5% significance level with GO biological process annotations related to ubiquitination. Ubiquitination requires three types of enzymes: a ubiquitin-activating enzyme (E1), a ubiquitin-conjugating enzyme (E2) and a ubiquitin ligase (E3). Deubiquitinating enzyme (DUB) removes

ubiquitin from substrate proteins. Screen hits are highlighted in the text boxes. (B) 4F+RFN40i induced iPSCs show strong positive immunostaining for pluripotency markers NANOG, SSEA3 and SSEA4. Scale bar, 300 μm. (C) 4F+RFN40i induced iPSCs express endogenous pluripotency markers at similar levels to 4F iPSCs and ESCs. (D) Growth curves of fibroblasts infected with 4F, 4F+ non-sense (NS), and 4F+RNF40i, counted on D0, D4, D8, and D13 post-infection. RNF40i did not substantially alter total cell numbers during the first 13 days of reprogramming. Error bars represent standard deviations. (E) UBE2Di, UBE2Ei and RNF40i lead to increases in OCT4 protein levels, as assessed by Western Blotting. BJ fibroblasts were infected with 4F control, 4F with non-sense (NS) shRNA control, or 4F with indicated shRNAs to knockdown genes in the ubiquitin pathway. Cells were collected 7 days after infection. Tubulin was used as loading control. Densitometry analysis of OCT4 protein level standardized to Tubulin is indicated above each lane.

**Figure S4: Additional Data on the Role of ADAM29 as Detrimental to Reprogramming and hESC Expansion, Related to Figure 4**

(A) Knockdown of ADAM29 by two independent shRNAs increases human iPSC generation efficiency. Instead of using two shRNAs against ADAM29 combined together (Figure 5B), BJ fibroblasts were infected with ADAM29 i1 and ADAM29 i2 separately plus 4F. 4F alone and 4F with non-sense (NS) shRNA were used as controls. The number of iPSC colonies was counted 25 days after infection. Infections were performed in triplicates, and error bars represent standard deviations. ***, $p<0.001$. (B) qRT-PCR confirms that two independent shRNAs ADAM29 i1 and ADAM29 i2, can both reduce ADAM29 expression. (C) Western Blot validation of ADAM29 knockdown at protein level by two separate shRNAs. The ADAM29 cDNA with a V5 tag was over-expressed in BJ fibroblasts, and detected by anti-V5 antibody. The first shRNA targets the UTR region of ADAM29 mRNA, so there is no knockdown of the exogenous protein. The second shRNA targets the coding sequence region, and there is a significant down regulation of the exogenous protein. Tubulin was used as loading control. (D) ADAM29 over-expression reduces colony formation efficiency of H9 hESCs. ***, $p<0.001$. (E) ADAM family proteins show strong sequence and structural homology to the type III snake venom

metalloprotease VAP2B. Shown are the VAP2B metalloprotease-disintegrin-cystine-rich (MDC) motif (Igarashi et al., 2007) (top) and the corresponding orthologous domains in human ADAM21 (bottom). (F) Alignment of ADAM family members' disintegrin loop sites. The tripeptide motif critical for the specificity of interaction with Integrins is shown in red.

**Figure S5: Additional Data on the Role of Endocytosis as a Barrier, Related to Figure 5**

(A) Pitstop2 increases human iPSC generation efficiency independently of viral reprogramming methods. Human fibroblasts HDF were electroporated with episomal reprogramming vectors (Okita et al., 2011). Two days later, clathrin-mediated endocytosis inhibitors (Abcam) were added to medium at a final concentration of 30 μM untill colonies were counted. Experiments were performed in triplicates, and error bars represent standard deviation. (B) Growth curves of BJ fibroblasts infected with 4F and 4F+Pitstop2, counted on D0, D4, D8, and D13 post-infection. RNF40i did not substantially alter total cell numbers during the first 13 days of reprogramming. Error bars represent standard deviations. (C) Pitstop2-mediated increase in E-Cadherin and decrease in pSMAD2/3 levels at D12 of reprogramming was confirmed by Western Blotting. Tubulin was used as loading control. (D) Pitstop2-mediated increase in E-Cadherin level was confirmed by qRT-PCR. (E) Interactions with endocytosis pathway and TGFβ signaling pathway. Genetic interactions were inferred from knockdown experiments via a literature search. 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 (purple edges). Genes showing physical interactions, sharing common protein domains, and found to be co-expressed in the same tissue type, were aggregated via GeneMANIA.

**Figure S6: Overlap between This Study and an Ovarian-Carcinoma Stem Cell shRNA Screen, Related to Figure 6**

Using the same multi-objective method used to call hits in the TRA-1-81+ screen, we analyzed the recent pooled shRNA screen data for gene barriers to growth and proliferation in 14 molecular subtypes of ovarian carcinoma (Tan et al., 2013). (A) Using a mesenchymal carcinoma subtype as control

(HeyA8), we called hits in a stem-like carcinoma subtype (StemA). 35 genes are hits at the 5% significance level in both screens. (B) DAVID functional analysis of the overlapping genes identifies transcription factor annotations (with a particular enrichment for homeobox domain proteins), cell membrane receptor annotations, and ubiquitin-mediated proteolysis pathway annotations as over-represented. Each rectangle in the top panel refers to a gene cluster. Over-represented annotation terms in the “Proteolysis” sub-cluster are displayed in the bottom panel. For four annotation clusters, example gene members are displayed, and their p-values in the TRA-1-81+ screen are shown.

## EXTENDED EXPERIMENTAL PROCEDURES

### Amplification of shRNA Pools from Genomic DNA and NGS

Genomic DNA was prepared using the PureGene kit (Qiagen). shRNAs were amplified from genomic DNA in a 50-µl PCR reaction consisting of 30 µl water, 10 µl 5 × Phusion GC buffer, 5 µl of 5 µM primer mix, 1 µl of 10 mM dNTPs, 1.5 µl DMSO, 750 ng genomic DNA and 1 U (0.5 µl) Phusion polymerase (Finnzymes). Cycling parameters were 98 °C for 30 s; then 25 cycles of 98 °C for 30 s, 56 °C for 15 s, 72 °C for 15 s; then 72 °C for 10 min. In some cases several PCR reactions were pooled on a Minelute column (Qiagen) before electrophoresis on 20% PAGE with 0.5 × TBE running buffer, electroelution and concentration on a second column. Library generation and deep-sequencing were performed at the UC Davis Genome Center DNA Technologies Core on a HiSeq machine according to manufacturer's protocols (Illumina).

### Statistical Procedures for Screen De-Convolution

Deep sequencing of shRNAs recovered from TRA-1-81+ and TRA-1-81- populations yielded 59,803,702 and 54,893,134 mapped reads in each population, respectively. Reads were aligned to reference sequences for the shRNA library using novoalign from Novocraft (www.novocraft.com). Read counts from the TRA-1-81+ population were normalized to the TRA-1-81- population by sequencing depth. The activity level for a given shRNA h was assessed by the odds ratio $\theta = \frac{X(N-Y)}{Y(N-X)}$, where X is the number of normalized reads mapping to h in TRA-1-81+ out of N total normalized reads, and Y the number of normalized reads mapping to h in TRA-1-81-. We designate an shRNA as *active* if $\theta$>1 and X is at least one median absolute deviation above 0. The latter constraint removes under-sequenced shRNAs. To aggregate shRNAs gene-wise into a measure of collective effect size, we used a random effects model. These models are often used to combine odds ratios from multiple clinical trials. They account for variation in odds ratio estimates attributable to differences in the numbers of participants between trials. Analogously, different shRNA have been sequenced to different depths and have different knockdown efficiencies, and this will affect the estimation of their odds ratios. We use the DerSimonian-Laird estimator for between-shRNA variance (DerSimonian and Kacker, 2007), and we

define the collective shRNA activity level for a given gene to be $\delta = \sum_{i=1}^{n} w_i \ln\theta_i \Big/ \sum_{i=1}^{n} w_i$ . The sum is over all active shRNAs for the given gene, and the weight $w_i$ is chosen to be inversely proportional to the variance of $\ln\theta_i$ . Since the sample variance for a given shRNA increases as the sequencing depth for that shRNA decreases, this choice has the effect of down-weighting the odds ratios of shRNAs sequenced at insufficient depth. To rank genes, we utilized multi-objective optimization, a technique which simultaneously maximizes two or more metrics (Handl et al., 2007). We chose collective shRNA activity level $\delta$ and the number of distinct active shRNAs as our metrics. This choice preferentially ranks genes with multiple active shRNAs and thus controls for off-target effects. The statistical significance of a gene's rank was estimated by permuting the read counts within each population 5,000 times and re-ranking the genes, based on the permuted counts. The fraction of times the permuted rank improved upon a gene's original rank (the fraction of times the rank ordinal decreased) gives a p-value for the probability that a random gene is ranked equally well or better by chance. We controlled for false discovery rate (FDR) in multiple hypothesis testing by library swap. We swapped the read counts in TRA-1-81+ and TRA-1-81-, re-ran the gene ranking and permutation test, and used the number of genes identified as significant in the negative population at a given p-value cutoff as an estimate of false positives in TRA-1-81+ at that significance level. A detailed discussion of this statistical method, and software source code, will be published elsewhere (A.D., H.Q., M.R-S., J.S.S., in preparation).

**Functional Annotation and Meta-Analyses**

In Figure 1B, the p-values are computed by a hypergeometric test comparing the relative frequencies of genes with $\delta > 2$ between the given target list and the set of all genes in the screen. Validated miRNA targets were obtained from Tarbase, a database of published miRNA-gene interaction experiments (Vergoulis et al., 2012). Identification and clustering of over-represented biological process GO terms, cellular component GO terms and Interpro protein domains were performed with DAVID (Huang et al., 2009) using the default settings. We used Fisher's method to combine the p-values for each annotation term in each cluster into a single p-value for the cluster. For identifying screen hits in

the cell adhesion or motility pathways, we aggregated all screen hits whose biological process GO terms contained any of the following keywords as substrings: “Adhesion”, “Migration”, “Motility”, “Podosome”, or “Lamellipodium”. This yielded 105 genes. Likewise, for ubiquitination, we used the keyword “Ubiquitin”, totaling 32 genes. For endocytosis, we used the keywords “Lysosome”, “Lysosomal”, “Endosome”, “Endocytic”, “Endocytosis”, “Vesicle”, and “Coated pit”, which yielded 110 genes. Quality control for the Polo *et al*. H3K4me3 and H3K27me3 ChIP-seq data (Polo et al., 2012) was performed using CHANCE (Diaz et al., 2012), and downstream segmentation of the genome into active, repressed, bivalent, and not-enriched domains was performed using ChromHMM (Ernst and Kellis, 2012). The promoter region was estimated as a 3kbp window around the transcription start site. In Figure S2, a gene was identified as having changed from active to non-active by a gain of a H3K27me3 mark or the complete loss of all pre-existing H3K4me3 marks. Polo *et al*. microarray data (Polo et al., 2012) was RMA normalized in R and clustered using k-means in MATLAB. We chose the number of clusters (k) by comparing the results of a principal components analysis, silhouette plots and visualizations of the clusters themselves. Gene interaction networks were constructed via GeneMANIA (http://genemania.org) and CytoScape (http://cytoscape.org).

**Fibroblast Culture and Infection**

Human primary newborn foreskin (BJ) fibroblasts were obtained from ATCC (CRL-2522) and cultured in DMEM with 10% FBS, 1x glutamine, 1x non-essential amino acids, 1x sodium pyruvate, 1x penicillin/streptomycin, and 0.06 mM β-mercaptoethanol (fibroblast medium). Cells were seeded at 60,000 cells per well on a 6-well plate the day before infection. Cells were infected with retroviruses leading to the over-expression of OCT4, SOX2, KLF4 and c-MYC alone or in combination with lentivirus for gene knockdown or over-expression. To produce lentivirus, 293T cells at 60-70% confluency were transfected in 10cm plates with 4 µg of the lentiviral vectors together with 1 µg each of the packaging plasmids VSV-G, MDL-RRE and RSVr using Fugene 6 (Roche). After 72 hours viral supernatants were harvested, filtered, titered and stored at -80 °C. Cells were infected in human ESC medium (DMEM/F12 with 20% KSR, 0.5x glutamine, 1x non-essential amino acids, 1x penicillin/streptomycin, 0.1 mM β-

mercaptoethanol, 10 ng/ml bFGF) and 8 µg/ml polybrene. Cells remained in the presence of virus for 48 hours, and fibroblast medium was added on the day after virus addition. 48 hours after infection, virus was removed and cells were cultured in human ESC medium.

**Quantitative Real-Time RT-PCR**

RNA was isolated using the RNeasy Mini RNA Isolation kit (Qiagen) and reverse-transcribed using the High-Capacity cDNA Reverse Transcription kit (Applied BioSystems). The cDNA reaction was diluted 1:5 in TE (10mM Tris-Cl/1mM EDTA, pH 7.6) and used in Sybr Green real-time PCR reactions (BioRad or Applied BioSystems). PCR primers were designed to amplify 100-400 bp fragments spanning exons. Reactions were run in triplicates on a 7900HT machine (Applied BioSystems) according to the manufacturer's instructions. Only samples with single and matching end-point melting curve peaks were used for subsequent analysis. Cycle threshold values were imported into the REST software for fold-change calculations. Values were standardized to *GAPDH* and *UBB*, and then normalized to uninfected BJ fibroblasts. Data are from triplicate PCR reactions, and error bars represent standard deviation. Primer sequences are listed in Table S4.

**Immuno-Staining and Western Blotting**

For Immunofluorescence, cells were fixed directly in tissue culture plates using 4% paraformaldehyde or cold 100% methanol, and permeabilized with 0.1% Triton X-100. Cells were then stained with primary antibodies against NANOG (AF1997, R&D), SSEA-3 (MAB4303, Millipore), SSEA-4 (MAB4304, Millipore), Tra1-81 (MAB4381, Millipore). Respective secondary antibodies were conjugated to either Alexa Fluor 594 or Alexa Fluor 488 (Invitrogen) and used at 1:500. For SDS-PAGE and Western Blotting, primary antibodies against OCT4 (GT486, GeneTex), V5 (ab9116, Abcam), SMAD2/3 (#8685, Cell Signaling), pSMAD2 (Ser 465/467)/SMAD3 (Ser 423/425) (#8828, Cell Signaling), pSMAD2 (Ser465/467, #3108, Cell Signaling), pSMAD3 (S423/425, ab52903, Abcam), E-Cadherin (610181, BD) and alpha-Tubulin (ab18251, Abcam) were used. Respective secondary antibodies conjugated with HRP (Abcam) were used according to standard protocols.

**ADAM29 Disintegrin Loop Peptides**

Peptides corresponding to the disintegrin loop of ADAM29 (CRKEVNECDLPEWC), as well as a mutated sequence as control (CRKEVNAAALPEWC), were synthesized on a peptide synthesizer (GenScript) and purified by high performance liquid chromatography. Peptides were amidated at the COOH terminus and acetylated at the NH2 terminus. The two terminal cysteine residues were protected with acetoamidomethyl groups. Peptides were added to medium at a final concentration of 100 μg/ml for immediate use.

## SUPPLEMENTAL REFERENCES


Brass, A.L., Dykxhoorn, D.M., Benita, Y., Yan, N., Engelman, A., Xavier, R.J., Lieberman, J., and Elledge, S.J. (2008). Identification of host proteins required for HIV infection through a functional genomic screen. Science *319*, 921–926.

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.

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. Nature Genetics *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.

DerSimonian, R., and Kacker, R. (2007). Random-effects model for meta-analysis of clinical trials: an update. Contemp Clin Trials *28*, 105–114.

Diaz, A., Nellore, A., and Song, J.S. (2012). CHANCE: comprehensive software for quality control and validation of ChIP-seq data. Genome Biology *13*, R98.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nature Methods *9*, 215–216.

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. Stem Cell *11*, 769–782.

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

Hu, S., Wilson, K.D., Ghosh, Z., Han, L., Wang, Y., Lan, F., Ransohoff, K.J., Burridge, P., and Wu, J.C. (2013). MicroRNA-302 Increases Reprogramming Efficiency via Repression of NR2F2. Stem Cells *31*, 259–268.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols *4*, 44–57.

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.

Jiao, J., Dang, Y., Yang, Y., Gao, R., Zhang, Y., Kou, Z., Sun, X.-F., and Gao, S. (2013). Promoting Reprogramming by FGF2 Reveals that the Extracellular Matrix Is a Barrier for Reprogramming Fibroblasts to Pluripotency. Stem Cells *31*, 729–740.

Kittler, R., Pelletier, L., Heninger, A.-K., Slabicki, M., Theis, M., Miroslaw, L., Poser, I., Lawo, S., Grabner, H., Kozak, K., et al. (2007). Genome-scale RNAi profiling of cell division in human tissue culture cells. Nature Cell Biology *9*, 1401–1412.

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

Lake, B.B., Fink, J., Klemetsaune, L., Fu, X., Jeffers, J.R., Zambetti, G.P., and Xu, Y. (2012). Context-Dependent Enhancement of Induced Pluripotent Stem Cell Reprogramming by Silencing Puma. Stem Cells *30*, 888–897.

Liang, G., He, J., and Zhang, Y. (2012). Kdm2b promotes induced pluripotent stem cellgeneration by facilitating gene activation earlyin reprogramming. Nature Cell Biology *14*, 457–466.

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. Proceedings of the National Academy of Sciences *105*, 20380–20385.

MacKeigan, J.P., Murphy, L.O., and Blenis, J. (2005). Sensitized RNAi screen of human kinases and phosphatases identifies new regulators of apoptosis and chemoresistance. Nature Cell Biology *7*, 591–600.

Okita, K., Matsumura, Y., Sato, Y., Okada, A., Morizane, A., Okamoto, S., Hong, H., Nakagawa, M., Tanabe, K., Tezuka, K.-I., et al. (2011). A more efficient method to generate integration-free human iPS cells. Nature Methods *8*, 409–412.

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

O'Malley, J., Skylaki, S., Iwabuchi, K.A., Chantzoura, E., Ruetz, T., Johnsson, A., Tomlinson, S.R., Linnarsson, S., and Kaji, K. (2013). High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. Nature *499*, 88–91.

Pfaff, N., Fiedler, J., Holzmann, A., Schambach, A., Moritz, T., Cantz, T., and Thum, T. (2011). miRNA screening reveals a new miRNA family stimulating iPS cell generation via regulation of Meox2. EMBO Reports *12*, 1153–1159.

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.

Possemato, R., Marks, K.M., Shaul, Y.D., Pacold, M.E., Kim, D., Birsoy, K., Sethumadhavan, S., Woo, H.-K., Jang, H.G., Jha, A.K., et al. (2011). Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. Nature *476*, 346–350.

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. Human Molecular Genetics *21*, 2054–2067.

Schlabach, M.R., Luo, J., Solimini, N.L., Hu, G., Xu, Q., Li, M.Z., Zhao, Z., Smogorzewska, A., Sowa, M.E., Ang, X.L., et al. (2008). Cancer Proliferation Gene Discovery Through Functional Genomics. Science *319*, 620–624.

Silva, J.M., Marran, K., Parker, J.S., Silva, J., Golding, M., Schlabach, M.R., Elledge, S.J., Hannon, G.J., and Chang, K. (2008). Profiling essential genes in human mammary cells by multiplex RNAi screening. Science *319*, 617–620.

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. Nature Biotechnology *29*, 443–448.

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*, 1051–1066.

Vergoulis, T., Vlachos, I.S., Alexiou, P., Georgakilas, G., Maragkakis, M., Reczko, M., Gerangelos, S., Koziris, N., Dalamagas, T., and Hatzigeorgiou, A.G. (2012). TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. Nucleic Acids Research *40*, D222–D229.

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.

**Table S1. Genes Known to Be Barriers to Reprogramming, Related to Figure 1**

| Species | Gene barriers | Title | Journal and reference |
|---|---|---|---|
| Mouse | Suv39h1, Suv39h2, Ehmt2, Setdb1 | H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs | Nature Genetics (Chen et al., 2013) |
| Human | DOT1L, YY1 | Chromatin modifying enzymes as modulators of reprogramming | Nature (Onder et al., 2012) |
| Mouse | Col1a1 | Promoting Reprogramming by FGF2 Reveals that the Extracellular Matrix is a Barrier for Reprogramming Fibroblasts to Pluripotency | Stem Cells (Jiao et al., 2013) |
| Mouse | Fbxw7 | Regulation of Pluripotency and Cellular Reprogramming by the Ubiquitin-Proteasome System | Cell Stem Cell (Buckley et al., 2012) |
| Human | NR2F2 | MicroRNA-302 Increases Reprogramming Efficiency via Repression of NR2F2 | Stem Cells (Hu et al., 2013) |
| Mouse | Ink4a/Arf/(CDKN2A) | Kdm2b promotes induced pluripotent stem cell generation by facilitating gene activation early in reprogramming | Nature Cell Biology (Liang et al., 2012) |
| Human | LATS2 | Transcriptional analysis of pluripotency reveals the Hippo pathway as a barrier to reprogramming | Human Molecular Genetics (Qin et al., 2012) |
| Mouse | Puma (BBC3) | Context-Dependent Enhancement of Induced Pluripotent Stem Cell Reprogramming By Silencing Puma | Stem Cells (Lake et al., 2012) |
| Mouse | Meox2 | miRNA screening reveals a new miRNA family stimulating iPS cell generation via regulation of Meox2 | EMBO Reports (Pfaff et al., 2011) |
| Mouse | Cd44 | High-resolution analysis with novel cell-surface markers identifies routes to iPS cells | Nature (O'Malley et al., 2013) |
| Human | CDKN1A, RBL2, CDC2L6 (CDK19), AKT1, ARHGAP26, RHOC, TGFBR2, MECP2, MBD2, SMARCC2, RAB5C, RAB11FIP5 | Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells | Nature Biotechnology (Subramanyam et al., 2011) |

**Table S2. Comparison of This and Other Genome-wide shRNA/siRNA Screen Studies, Related to Experimental Procedures**

| Study and Reference | Validation Rate | Gene # | shRNA/siRNA # |
|---|---|---|---|
| Nat Cell Biol (MacKeigan et al., 2005) | 29% | 650 | 13,000 |
| Nat Cell Biol (Kittler et al., 2007) | 59% | 17,828 | 19,585 |
| Nat Methods (König et al., 2007) (RSA) | 65% | 19, 628 | 53,850 |
| PNAS (Luo et al., 2008) (RIGER) | 45% | 9,500 | 45,000 |
| Science (Schlabach et al., 2008) | 22% | 2,924 | 8,203 |
| Science (Brass et al., 2008) | 71% | 21,121 | 84,484 |
| Science (Silva et al., 2008) | 74% | 7,250 | 20,000 |
| Nature (Chia et al., 2010) | 64% | 21,121 | 84,484 |
| Nature (Possemato et al., 2011) | 31% | 133 | 700 |
| EMBO Mol Med (Tan et al., 2013) | 21% | 16,000 | 80,000 |
| | | | |
| Average | 48% | 11,616 | 40,930 |
| | | | |
| Our work | 87% | 19,527 | 600,000 |

**Table S3. Sequences for shRNA Cloning Used in This Study, Related to Experimental Procedures**

| Gene name | Primer sequences |
|---|---|
| UBE2D3 | TGTTCAGTTAAAGTGCACAGTCTGTTCTAGAGACAGACTGTGCACTTTAACTGAACTTTTTC |
| shRNA-1 | TCGAGAAAAAAGTTCAGTTAAAGTGCACAGTCTGTCTCTAGAACAGACTGTGCACTTTAACTGAACA |
| UBE2D3 | TGATGCCATGTGATGCTACCTTAATTCTAGAGATTAAGGTAGCATCACATGGCATCTTTTTTC |
| shRNA-2 | TCGAGAAAAAAGATGCCATGTGATGCTACCTTAATCTCTAGAATTAAGGTAGCATCACATGGCATCA |
| UBE2E3 | TGTTCACATAATTTGTATGCAGTGTTCTAGAGACACTGCATACAAATTATGTGAACTTTTTTC |
| shRNA-1 | TCGAGAAAAAAGTTCACATAATTTGTATGCAGTGTCTCTAGAACACTGCATACAAATTATGTGAACA |
| UBE2E3 | TGCTGTGTATATGTTATACTGATTTTCTAGAGAAATCAGTATAACATATACACAGCTTTTTTC |
| shRNA-2 | TCGAGAAAAAAGCTGTGTATATGTTATACTGATTTCTCTAGAAAATCAGTATAACATATACACAGCA |
| RNF40 | TGCTGAACCAACATCATCAGTTTCTTCTAGAGAGAAACTGATGATGTTGGTTCAGCTTTTTTC |
| shRNA-1 | TCGAGAAAAAAGCTGAACCAACATCATCAGTTTCTCTCTAGAAGAAACTGATGATGTTGGTTCAGCA |
| RNF40 | TGTGCCGACGAAATCCTCCAGGAGTTCTAGAGACTCCTGGAGGATTTCGTCGGCACTTTTTTC |
| shRNA-2 | TCGAGAAAAAAGTGCCGACGAAATCCTCCAGGAGTCTCTAGAACTCCTGGAGGATTTCGTCGGCACA |
| DRAM1 | TGCTGTGGTGGAATGAATCCATACTTCTAGAGAGTATGGATTCATTCCACCACAGCTTTTTTC |
| shRNA-1 | TCGAGAAAAAAGCTGTGGTGGAATGAATCCATACTCTCTAGAAGTATGGATTCATTCCACCACAGCA |
| DRAM1 | TGTGTACAGTAACTCTAAGTGAATTTCTAGAGAATTCACTTAGAGTTACTGTACACTTTTTTC |
| shRNA-2 | TCGAGAAAAAAGTGTACAGTAACTCTAAGTGAATTCTCTAGAAATTCACTTAGAGTTACTGTACACA |
| SLC17A5 | TGCACTGCTATGGTCTTGATACATTTCTAGAGAATGTATCAAGACCATAGCAGTGCTTTTTTC |
| shRNA-1 | TCGAGAAAAAAGCACTGCTATGGTCTTGATACATTCTCTAGAAATGTATCAAGACCATAGCAGTGCA |
| SLC17A5 | TGTTGACATTATTGCCTACTTATATTCTAGAGATATAAGTAGGCAATAATGTCAACTTTTTTC |
| shRNA-2 | TCGAGAAAAAAGTTGACATTATTGCCTACTTATATCTCTAGAATATAAGTAGGCAATAATGTCAACA |
| ARSD | TGTGCCCATTCATTGATCTGAGAATTCTAGAGATTCTCAGATCAATGAATGGGCACTTTTTTC |
| shRNA-1 | TCGAGAAAAAAGTGCCCATTCATTGATCTGAGAATCTCTAGAATTCTCAGATCAATGAATGGGCACA |
| ARSD | TGATGTGTCTTCCTGATATTGTTATTCTAGAGATAACAATATCAGGAAGACACATCTTTTTTC |
| shRNA-2 | TCGAGAAAAAAGATGTGTCTTCCTGATATTGTTATCTCTAGAATAACAATATCAGGAAGACACATCA |
| ADAM7 | TGGACAGACAATGTTTAAGAGAAATTCTAGAGATTTCTCTTAAACATTGTCTGTCCTTTTTTC |
| shRNA-1 | TCGAGAAAAAAGGACAGACAATGTTTAAGAGAAATCTCTAGAATTTCTCTTAAACATTGTCTGTCCA |
| ADAM7 | TGACGTCTTTACAACCTTACCTAGTTCTAGAGACTAGGTAAGGTTGTAAAGACGTCTTTTTTC |
| shRNA-2 | TCGAGAAAAAAGACGTCTTTACAACCTTACCTAGTCTCTAGAACTAGGTAAGGTTGTAAAGACGTCA |
| ADAM21 | TGCAGTCTGTCTGTCTTGTCACATTTCTAGAGAATGTGACAAGACAGACAGACTGCTTTTTTC |
| shRNA-1 | TCGAGAAAAAAGCAGTCTGTCTGTCTTGTCACATTCTCTAGAAATGTGACAAGACAGACAGACTGCA |
| ADAM21 | TGTACTGTGTGTGGCCCAGGAAAGTTCTAGAGACTTTCCTGGGCCACACACAGTACTTTTTTC |
| shRNA-2 | TCGAGAAAAAAGTACTGTGTGTGGCCCAGGAAAGTCTCTAGAACTTTCCTGGGCCACACACAGTACA |
| ADAM29 | TGGATGCTTAATGTTAGAGTACAATTCTAGAGATTGTACTCTAACATTAAGCATCCTTTTTTC |
| shRNA-1 | TCGAGAAAAAAGGATGCTTAATGTTAGAGTACAATCTCTAGAATTGTACTCTAACATTAAGCATCCA |
| ADAM29 | TGGTGTGAGAATGTGACAGAAATTTTCTAGAGAAATTTCTGTCACATTCTCACACCTTTTTTC |
| shRNA-2 | TCGAGAAAAAAGGTGTGAGAATGTGACAGAAATTTCTCTAGAAAATTTCTGTCACATTCTCACACCA |
| PTPRJ | TGATGTGTTGATGTGGACTCTAAATTCTAGAGATTTAGAGTCCACATCAACACATCTTTTTTC |
| shRNA-1 | TCGAGAAAAAAGATGTGTTGATGTGGACTCTAAATCTCTAGAATTTAGAGTCCACATCAACACATCA |
| PTPRJ | TGTTCCGAGTATGTCTACCATTTATTCTAGAGATAAATGGTAGACATACTCGGAACTTTTTTC |
| shRNA-2 | TCGAGAAAAAAGTTCCGAGTATGTCTACCATTTATCTCTAGAATAAATGGTAGACATACTCGGAACA |
| PTPRK | TGCAGAGACCAAGTCATTACATTGTTCTAGAGACAATGTAATGACTTGGTCTCTGCTTTTTTC |
| shRNA-1 | TCGAGAAAAAAGCAGAGACCAAGTCATTACATTGTCTCTAGAACAATGTAATGACTTGGTCTCTGCA |
| PTPRK | TGCCGTATTACTTTGCTGCAGAACTTCTAGAGAGTTCTGCAGCAAAGTAATACGGCTTTTTTC |
| shRNA-2 | TCGAGAAAAAAGCCGTATTACTTTGCTGCAGAACTCTCTAGAAGTTCTGCAGCAAAGTAATACGGCA |
| PTPN11 | TGGACAGATCTTGTGGAACATTATTTCTAGAGAATAATGTTCCACAAGATCTGTCCTTTTTTC |
| shRNA-1 | TCGAGAAAAAAGGACAGATCTTGTGGAACATTATTCTCTAGAAATAATGTTCCACAAGATCTGTCCA |
| PTPN11 | TGATTTGCTAATGTTCTACATTAATTCTAGAGATTAATGTAGAACATTAGCAAATCTTTTTTC |
| shRNA-2 | TCGAGAAAAAAGATTTGCTAATGTTCTACATTAATCTCTAGAATTAATGTAGAACATTAGCAAATCA |
| ATF7IP | TGTAGCTACCATCTCTATGCTTACTTCTAGAGAGTAAGCATAGAGATGGTAGCTACTTTTTTC |
| shRNA-1 | TCGAGAAAAAAGTAGCTACCATCTCTATGCTTACTCTCTAGAAGTAAGCATAGAGATGGTAGCTACA |
| ATF7IP | TGCAGCAATAATAACATGTCTTACTTCTAGAGAGTAAGACATGTTATTATTGCTGCTTTTTTC |
| shRNA-2 | TCGAGAAAAAAGCAGCAATAATAACATGTCTTACTCTCTAGAAGTAAGACATGTTATTATTGCTGCA |
| ARID4A | TGACGATTGAAGTTGATAGTATTGTTCTAGAGACAATACTATCAACTTCAATCGTCTTTTTTC |
| shRNA-1 | TCGAGAAAAAAGACGATTGAAGTTGATAGTATTGTCTCTAGAACAATACTATCAACTTCAATCGTCA |

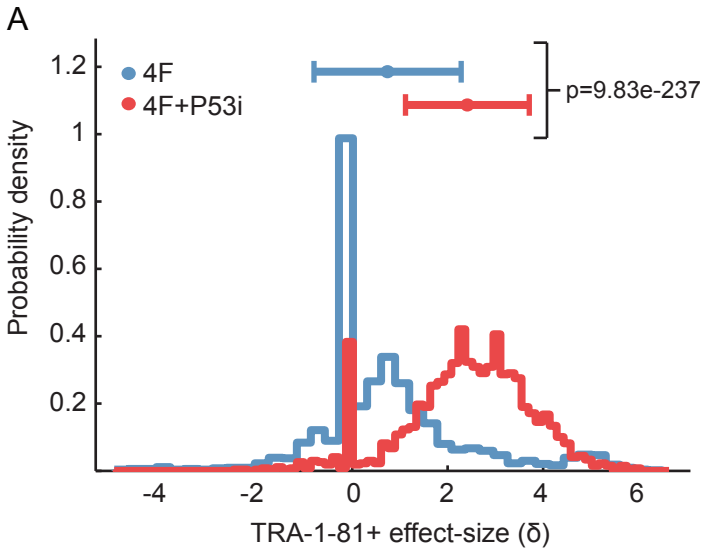
**Table S3. Sequences for shRNA Cloning Used in This Study, Related to Experimental Procedures, Continued**

| | |
|---|---|
| ARID4A shRNA-1 | TGTCTTCTCTTCCTGTTACTGAAGTTCTAGAGACTTCAGTAACAGGAAGAGAAGACTTTTTTC |
| ARID4A shRNA-2 | TCGAGAAAAAAGTCTTCTCTTCCTGTTACTGAAGTCTCTAGAACTTCAGTAACAGGAAGAGAAGACA |
| CENPB shRNA-1 | TGGAGGAAGAGGATGATGAAGATGTTCTAGAGACATCTTCATCATCCTCTTCCTCCTTTTTTC |
| CENPB shRNA-1 | TCGAGAAAAAAGGAGGAAGAGGATGATGAAGATGTCTCTAGAACATCTTCATCATCCTCTTCCTCCA |
| CENPB shRNA-2 | TGTCTGGTTCATGTGACCAGGAAGTTCTAGAGACTTCCTGGTCACATGAACCAGACTTTTTTC |
| CENPB shRNA-2 | TCGAGAAAAAAGTCTGGTTCATGTGACCAGGAAGTCTCTAGAACTTCCTGGTCACATGAACCAGACA |
| TTF1 shRNA-1 | TGTTGTGTTGATATGAGCATAGAATTCTAGAGATTCTATGCTCATATCAACACAACTTTTTTC |
| TTF1 shRNA-1 | TCGAGAAAAAAGTTGTGTTGATATGAGCATAGAATCTCTAGAATTCTATGCTCATATCAACACAACA |
| TTF1 shRNA-2 | TGTCTTGCTAGTGCCATAGGTGATTTCTAGAGAATCACCTATGGCACTAGCAAGACTTTTTTC |
| TTF1 shRNA-2 | TCGAGAAAAAAGTCTTGCTAGTGCCATAGGTGATTCTCTAGAAATCACCTATGGCACTAGCAAGACA |
| TTF2 shRNA-1 | TGGTGACTTTACTTCCCATGGAACTTCTAGAGAGTTCCATGGGAAGTAAAGTCACCTTTTTTC |
| TTF2 shRNA-1 | TCGAGAAAAAAGGTGACTTTACTTCCCATGGAACTCTCTAGAAGTTCCATGGGAAGTAAAGTCACCA |
| TTF2 shRNA-2 | TGATTCCTGTTTCCCATTGCTTATTTCTAGAGAATAAGCAATGGGAAACAGGAATCTTTTTTC |
| TTF2 shRNA-2 | TCGAGAAAAAAGATTCCTGTTTCCCATTGCTTATTCTCTAGAAATAAGCAATGGGAAACAGGAATCA |
| TMF1 shRNA-1 | TGAACTGAGTCTACCATTTCTTACTTCTAGAGAGTAAGAAATGGTAGACTCAGTTCTTTTTTC |
| TMF1 shRNA-1 | TCGAGAAAAAAGAACTGAGTCTACCATTTCTTACTCTCTAGAAGTAAGAAATGGTAGACTCAGTTCA |
| TMF1 shRNA-2 | TGAAGCATGGATCTGATTGAATAATTCTAGAGATTATTCAATCAGATCCATGCTTCTTTTTTC |
| TMF1 shRNA-2 | TCGAGAAAAAAGAAGCATGGATCTGATTGAATAATCTCTAGAATTATTCAATCAGATCCATGCTTCA |
| RPL30 shRNA-1 | TGCAGTGGCAATAATATTGAACTGTTCTAGAGACAGTTCAATATTATTGCCACTGCTTTTTTC |
| RPL30 shRNA-1 | TCGAGAAAAAAGCAGTGGCAATAATATTGAACTGTCTCTAGAACAGTTCAATATTATTGCCACTGCA |
| RPL30 shRNA-2 | TGCATCACTACAGTGGCAATAATATTCTAGAGATATTATTGCCACTGTAGTGATGCTTTTTTC |
| RPL30 shRNA-2 | TCGAGAAAAAAGCATCACTACAGTGGCAATAATATCTCTAGAATATTATTGCCACTGTAGTGATGCA |
| PLCE1 shRNA-1 | TGGGGAGACAGCATCATTTAACAATTCTAGAGATTGTTAAATGATGCTGTCTCCCCTTTTTTC |
| PLCE1 shRNA-1 | TCGAGAAAAAAGGGGAGACAGCATCATTTAACAATCTCTAGAATTGTTAAATGATGCTGTCTCCCCA |
| PLCE1 shRNA-2 | TGTTGTGACACCTTGAATGATAAATTCTAGAGATTTATCATTCAAGGTGTCACAACTTTTTTC |
| PLCE1 shRNA-2 | TCGAGAAAAAAGTTGTGACACCTTGAATGATAAATCTCTAGAATTTATCATTCAAGGTGTCACAACA |
| T shRNA-1 | TGTGTCGCCACCTTCCATGTGAAGTTCTAGAGACTTCACATGGAAGGTGGCGACACTTTTTTC |
| T shRNA-1 | TCGAGAAAAAAGTGTCGCCACCTTCCATGTGAAGTCTCTAGAACTTCACATGGAAGGTGGCGACACA |
| T shRNA-2 | TGCAGTCCTACTTTAGTGAGATAATTCTAGAGATTATCTCACTAAAGTAGGACTGCTTTTTTC |
| T shRNA-2 | TCGAGAAAAAAGCAGTCCTACTTTAGTGAGATAATCTCTAGAATTATCTCACTAAAGTAGGACTGCA |
| MED19 shRNA-1 | TGCAGTAGCTCTTTCAATCCTATCTTCTAGAGAGATAGGATTGAAAGAGCTACTGCTTTTTTC |
| MED19 shRNA-1 | TCGAGAAAAAAGCAGTAGCTCTTTCAATCCTATCTCTCTAGAAGATAGGATTGAAAGAGCTACTGCA |
| MED19 shRNA-2 | TGTTCTCAGTAGCTCTTTCAATCCTTCTAGAGAGGATTGAAAGAGCTACTGAGAACTTTTTTC |
| MED19 shRNA-2 | TCGAGAAAAAAGTTCTCAGTAGCTCTTTCAATCCTCTCTAGAAGGATTGAAAGAGCTACTGAGAACA |
| Non-sense | TGATCTCGCTTGGGCGAGAGTAAGTTCTAGAGACTTACTCTCGCCCAAGCGAGATCTTTTTTC |
| | TCGAGAAAAAAGATCTCGCTTGGGCGAGAGTAAGTCTCTAGAACTTACTCTCGCCCAAGCGAGATCA |

**Table S4. Sequences for qRT-PCR Primers Used in This Study, Related to Experimental Procedures**

| Gene name | Forward primer sequence | Reverse primer sequence |
|---|---|---|
| GAPDH | CAATGACCCCTTCATTGACC | GACAAGCTTCCCGTTCTCAG |
| Ubb | TTGTTGGGTGAGCTTGTTTG | GTCTTGCCGGTAAGGGTTTT |
| UBE2D3 | CGGACGTGGGAAGCAAGCCG | CAGACACAGGCGCCTCTTCACC |
| UBE2E3 | ATGCGGACCAGCGAGACCCA | GGCCCAGCACTGCAATTAGGAGG |
| RNF40 | GCCGGGCTGCTCGTGAGAAA | GCAGGTCAACCGCGCCTTGT |
| DRAM1 | GTGACCTGGTCGTCAGCCGC | TGGCTGCACCAAGAAATGCAGA |
| SLC17A5 | TCGAGACCTGGCCCGGAACG | CAGCACACTGGAGCGGCTTCG |
| ARSD | GGGGATCTCGGTTGCTACGGG | GCTGGCGTCCATGCCTGATCT |
| ADAM7 | TTGGCCTGGTGGCGTCAGGA | TCCGTGGCCATCCACAGGATTT |
| ADAM21 | TGTGGCCTCAGCTACAAACCATTG | GTGAGGCACTGCTGGTCTGGC |
| ADAM29 | GCATCCACAGGCACTGTGTCCA | AGGGGGTGGGCCACTGTCAA |
| PTPRJ | GAAAGATGGCAGCCCCCACCC | GTGACCAGCTTCCCCGGTGGT |
| PTPRK | TGGCTGTACAGGGATGGCTACC | AGCATTTAACCCGGCCAACCTCA |
| PTPN11 | GGGGAGAGCAATGACGGCAAGT | CGAGTCGTGTTAAGGGGCTGCT |
| ATF7IP | CCAGCTCAGGCTCCCTTGCG | GGCGTGGGGGCTCAGTATGC |
| ARID4A | CGCGCACTTCACCCGCAGTT | CAGGCTCATCTGCCGCCTTCAT |
| CENPB | GCCTGCCCTGCGACTACACC | AACAGCAGGACCCGGCGAGA |
| TTF1 | AGCAGCACTTCCGGGTTGGG | CCATTTTATTCCCTCCGAAAGCGCC |
| TTF2 | CGTGCGGGCCACCGACATTC | ACACCAGCGTTTCCCCTCTGC |
| TMF1 | AGGCCCAAGAAGAAGCCCGTC | CTCTGCTTCCTGGAGTCTCTGCT |
| RPL30 | TTAGCGGCTGCTGTTGGTTGGG | TCGTCTTCTTTGCGGCCACC |
| PLCE1 | TGGGCATTTTTGGGGTGGGCA | TCCATCAGAAACCTGGCAAACCCT |
| INPP5K | ACAAGCCTGTCTCCGGCACG | CGCAGCCCCACCTTGTACAGT |
| T | CCGGTCCTCACCCTACCCCAG | TGGGGTACTGACTGGAGCTGGTAG |
| MED19 | CTCCACACTGGCCCGTTGCC | CCTGGGTACGGCTCTGTTTGTGC |
| Endogenous OCT4 | TGTACTCCTCGGTCCCTTTC | TCCAGGTTTTCTTTCCCTAGC |
| Endogenous SOX2 | GCTAGTCTCCAAGCGACGAA | GCAAGAAGCCTCTCCTTGAA |
| Endogenous KLF4 | TATGACCCACACTGCCAGAA | TGGGAACTTGACCATGATTG |
| Endogenous cMYC | CGGAACTCTTGTGCGTAAGG | CTCAGCCAAGGTTGTGAGGT |
| Endogenous NANOG | CAGTCTGGACACTGGCTGAA | CTCGCTGATTAGGCTCCAAC |

Figure S1


A
4F
4F+P53i
p=9.83e-237
Probability density
1.2
1
0.8
0.6
0.4
0.2
-4
-2
0
2
4
6
TRA-1-81+ effect-size (δ)

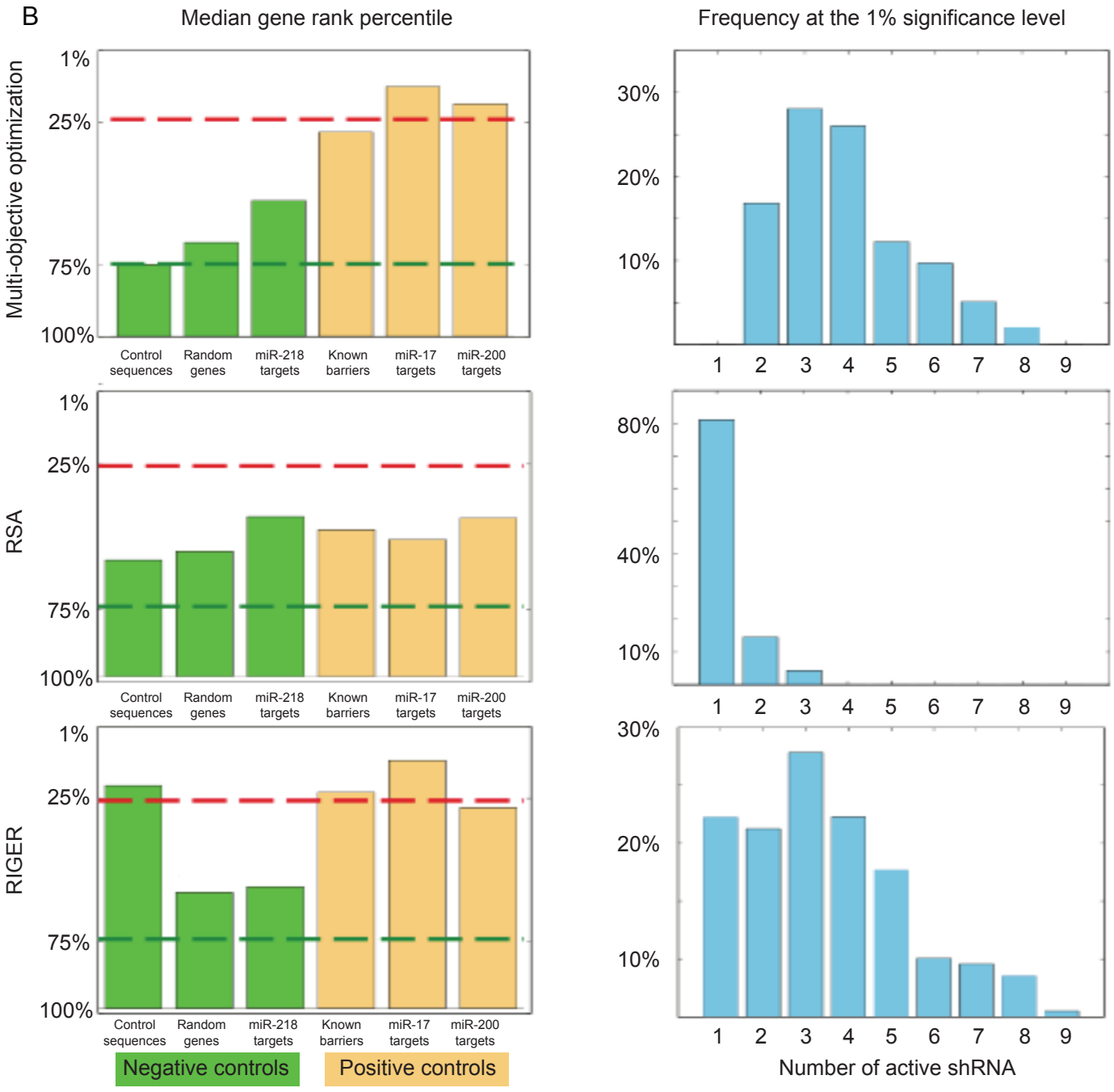

B
Median gene rank percentile
Frequency at the 1% significance level
Multi-objective optimization
RSA
RIGER
1%
25%
75%
100%
Control sequences
Random genes
miR-218 targets
Known barriers
miR-17 targets
miR-200 targets
30%
20%
10%
80%
40%
1
2
3
4
5
6
7
8
9
Number of active shRNA
Negative controls
Positive controls

Figure S2

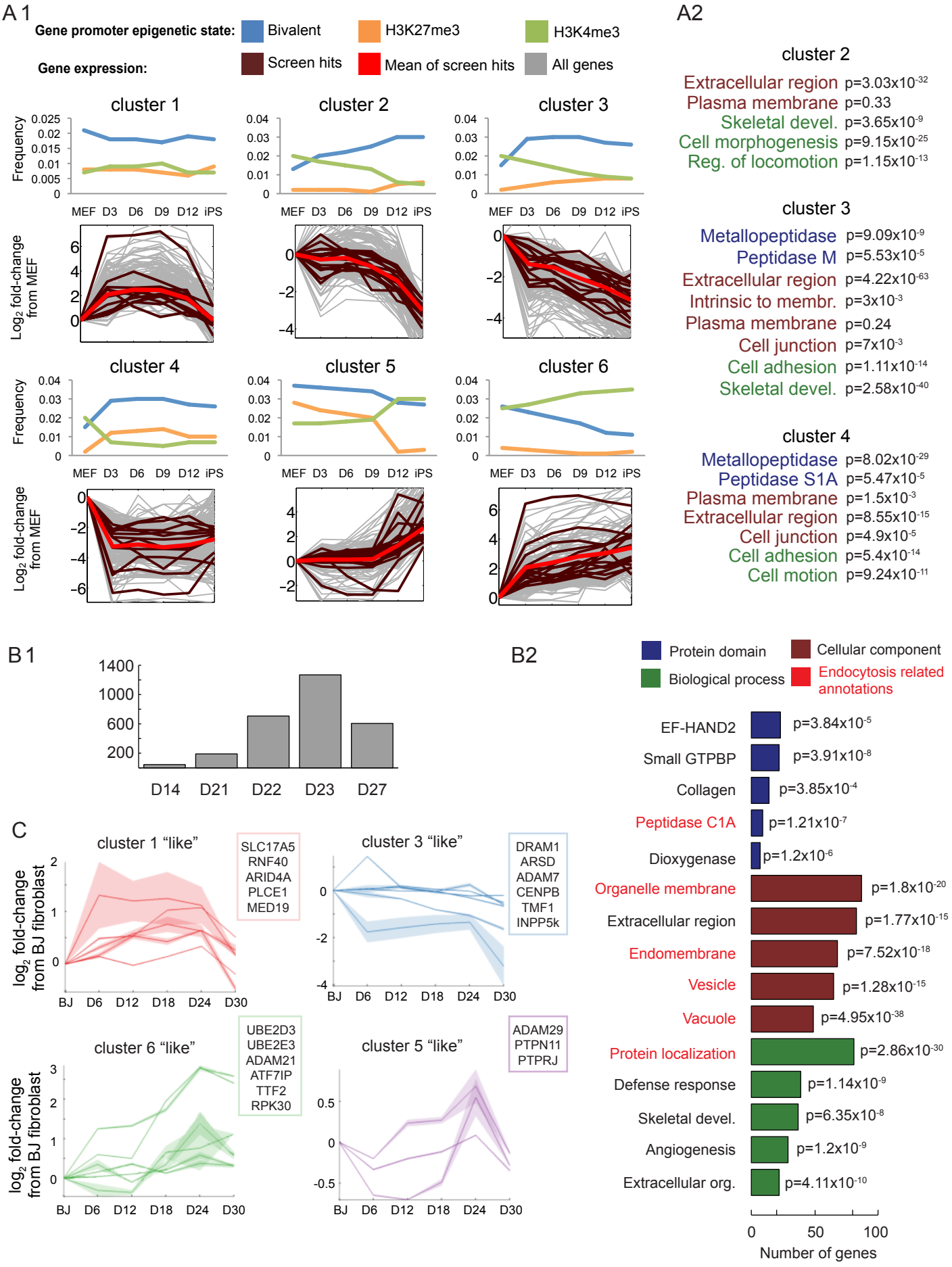

A 1
Gene promoter epigenetic state:
Bivalent
H3K27me3
H3K4me3
Gene expression:
Screen hits
Mean of screen hits
All genes
cluster 1
cluster 2
cluster 3
cluster 4
cluster 5
cluster 6
Frequency
Log2 fold-change from MEF
MEF D3 D6 D9 D12 iPS
A2
cluster 2
Extracellular region p=3.03x10-32
Plasma membrane p=0.33
Skeletal devel. p=3.65x10-9
Cell morphogenesis p=9.15x10-25
Reg. of locomotion p=1.15x10-13
cluster 3
Metallopeptidase p=9.09x10-9
Peptidase M p=5.53x10-5
Extracellular region p=4.22x10-63
Intrinsic to membr. p=3x10-3
Plasma membrane p=0.24
Cell junction p=7x10-3
Cell adhesion p=1.11x10-14
Skeletal devel. p=2.58x10-40
cluster 4
Metallopeptidase p=8.02x10-29
Peptidase S1A p=5.47x10-5
Plasma membrane p=1.5x10-3
Extracellular region p=8.55x10-15
Cell junction p=4.9x10-5
Cell adhesion p=5.4x10-14
Cell motion p=9.24x10-11
B 1
1400
1000
600
200
D14 D21 D22 D23 D27
B2
Protein domain
Cellular component
Biological process
Endocytosis related annotations
EF-HAND2 p=3.84x10-5
Small GTPBP p=3.91x10-8
Collagen p=3.85x10-4
Peptidase C1A p=1.21x10-7
Dioxygenase p=1.2x10-6
Organelle membrane p=1.8x10-20
Extracellular region p=1.77x10-15
Endomembrane p=7.52x10-18
Vesicle p=1.28x10-15
Vacuole p=4.95x10-38
Protein localization p=2.86x10-30
Defense response p=1.14x10-9
Skeletal devel. p=6.35x10-8
Angiogenesis p=1.2x10-9
Extracellular org. p=4.11x10-10
0 50 100
Number of genes
C
cluster 1 “like”
log2 fold-change from BJ fibroblast
BJ D6 D12 D18 D24 D30
SLC17A5
RNF40
ARID4A
PLCE1
MED19
cluster 3 “like”
DRAM1
ARSD
ADAM7
CENPB
TMF1
INPP5k
cluster 6 “like”
UBE2D3
UBE2E3
ADAM21
ATF7IP
TTF2
RPK30
cluster 5 “like”
ADAM29
PTPN11
PTPRJ

Figure S3

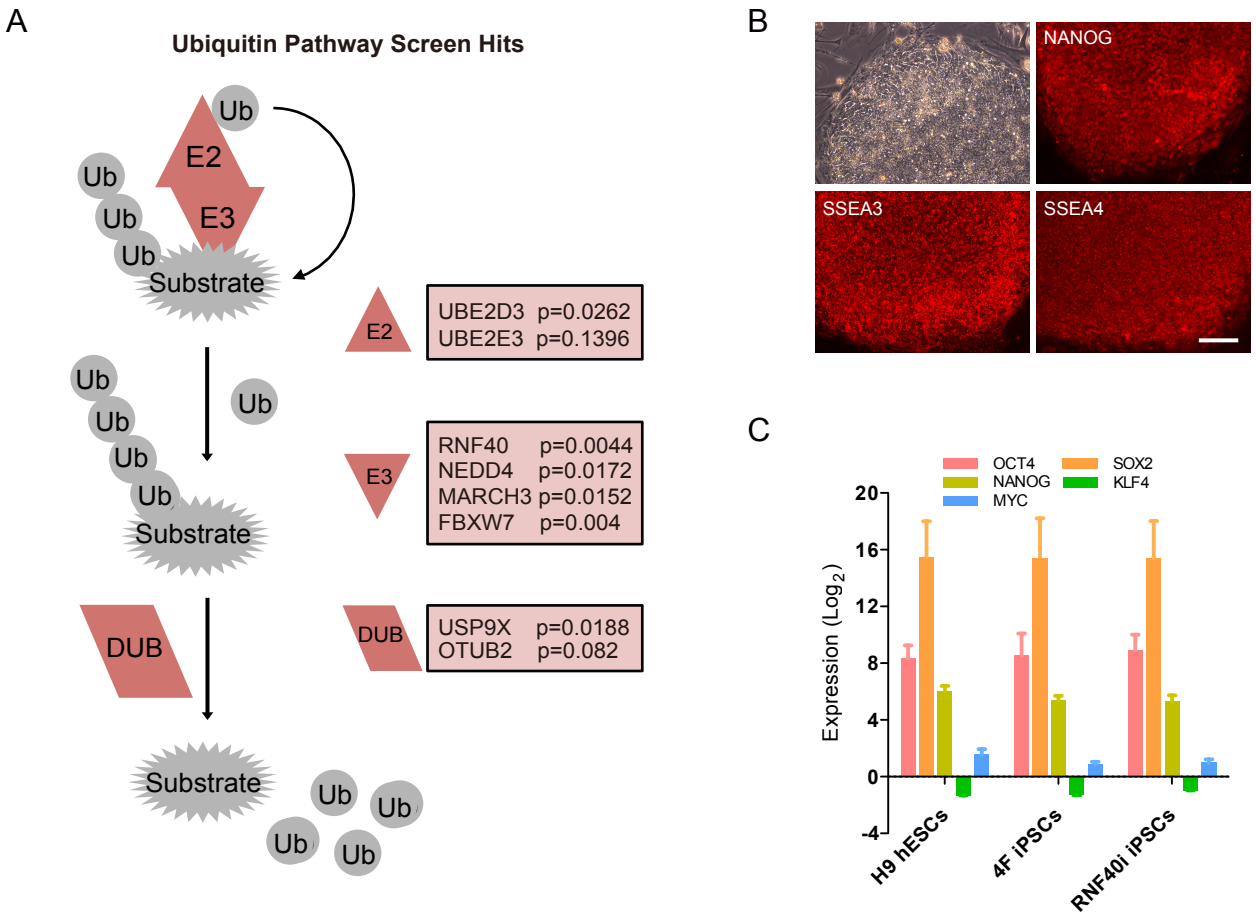
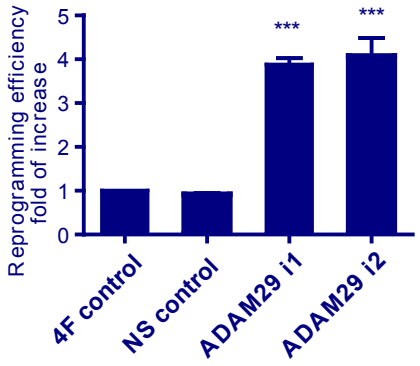
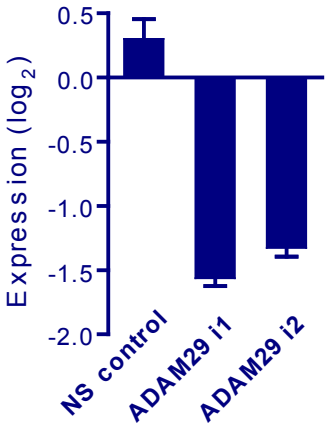

A
Ubiquitin Pathway Screen Hits
Ub
E2
E3
Substrate
UBE2D3 p=0.0262
UBE2E3 p=0.1396
RNF40 p=0.0044
NEDD4 p=0.0172
MARCH3 p=0.0152
FBXW7 p=0.004
DUB
USP9X p=0.0188
OTUB2 p=0.082
B
NANOG
SSEA3
SSEA4
C
OCT4
SOX2
NANOG
KLF4
MYC
Expression (Log2)
H9 hESCs
4F iPSCs
RNF40i iPSCs
D
4F control
NS control
4F RNF40i
Cell numbers (x10,000)
D0
D4
D8
D13
E
4F control
NS control
UBE2D3i
UBE2E3i
RNF40i
1 0.82 1.58 1.63 1.59
OCT4
Tubulin

Figure S4

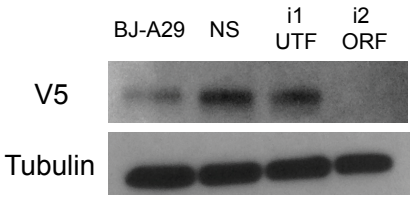
A



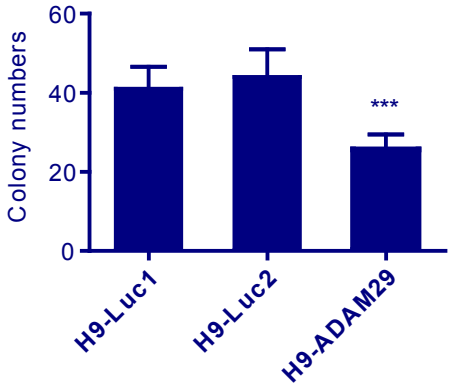

B



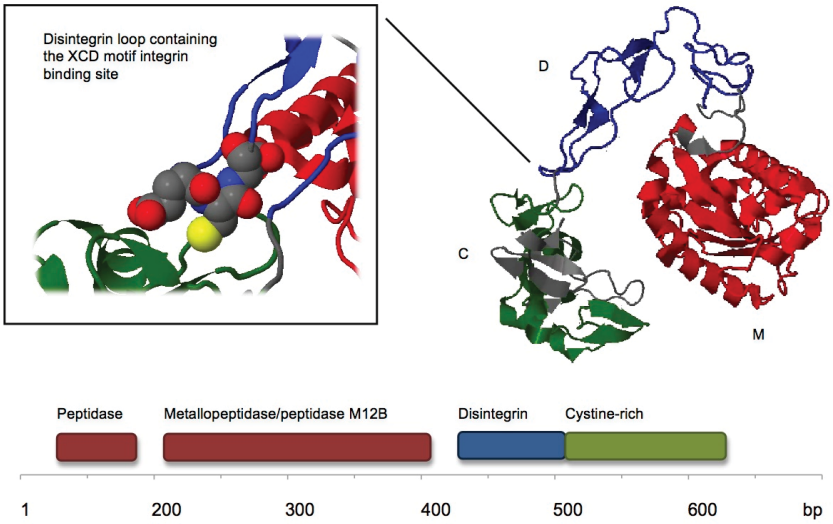

C




D



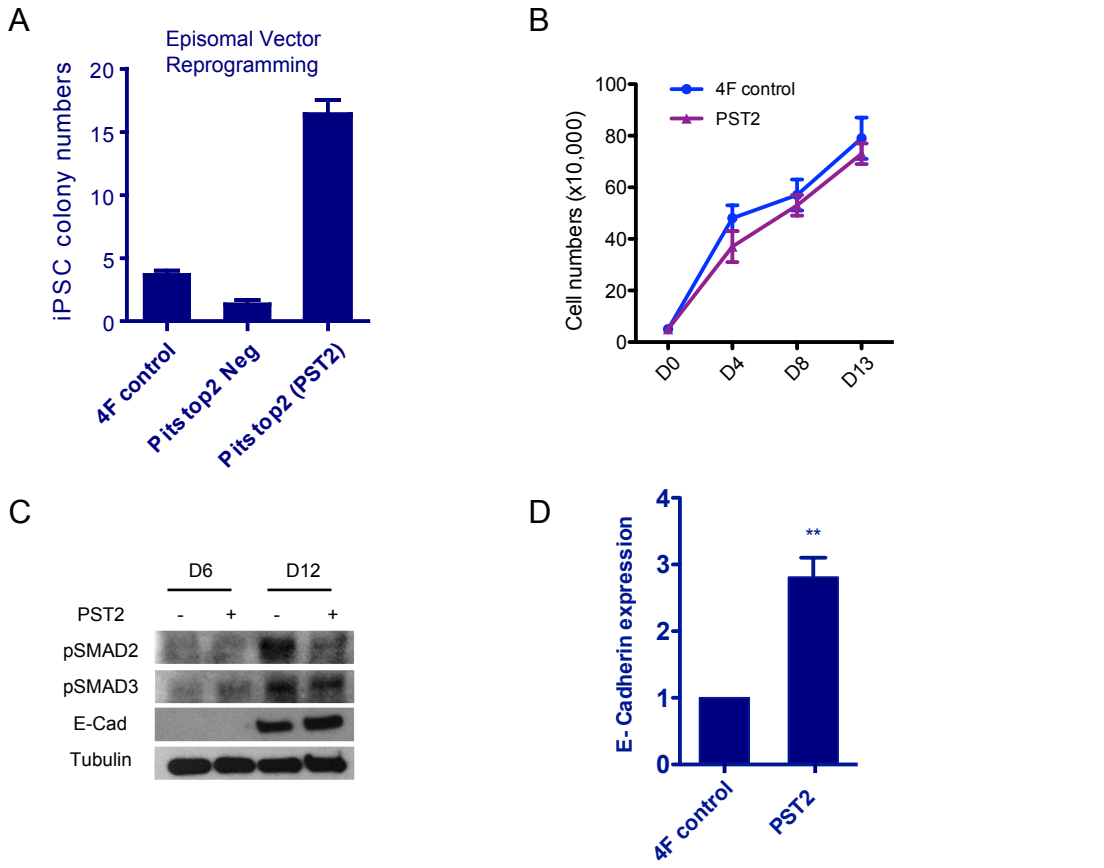

E




F

| Gene | D-loop sequence | Screen p-value |
|---|---|---|
| ADAM2 | CRPSFEECDLPEYC | 0.2706 |
| ADAM7 | CRPAKDECDFPEMC | 0.0444 |
| ADAM8 | CRPKKDMCDLEEFC | 0.3154 |
| ADAM9 | CRGKTSECDVPEYC | 0.1756 |
| ADAM10 | CRDDSDCAREGIC | 0.6802 |
| ADAM11 | CREAVNECDIAETC | 0.6996 |
| ADAM12 | CRDSSNSCDLPEFC | 0.9364 |
| ADAM15 | CRPTRGDCDLPEFC | 0.3122 |
| ADAM17 | CQEAINATCKGVSYC | 0.3012 |
| ADAM18 | CRKSIDPECDFTEYC | 0.886 |
| ADAM19 | CREQARQCDLPEFC | 0.319 |
| ADAM20 | CRQQVGECDLPEWC | 0.1438 |
| ADAM21 | CRQEVNECDLPEWC | 0.0006 |
| ADAM22 | CREAVNDCDIRETC | 0.2966 |
| ADAM23 | CRDAVNECDITEYC | 0.2394 |
| ADAM28 | CRPAKDECDLPEMC | 0.2778 |
| ADAM29 | CRKEVNECDLPEWC | 0.0176 |
| ADAM32 | CRPKAHPECDIAENC | 0.9268 |
| ADAM33 | CRQAMGDCDLPEFC | 0.977 |

Figure S5


A
Episomal Vector Reprogramming
iPSC colony numbers
20
15
10
5
0
4F control
Pits top2 Neg
Pits top2 (PST2)
B
4F control
PST2
Cell numbers (x10,000)
100
80
60
40
20
0
D0
D4
D8
D13
C
D6
D12
PST2
- + - +
pSMAD2
pSMAD3
E-Cad
Tubulin
D
E- Cadherin expression
**
4
3
2
1
0
4F control
PST2

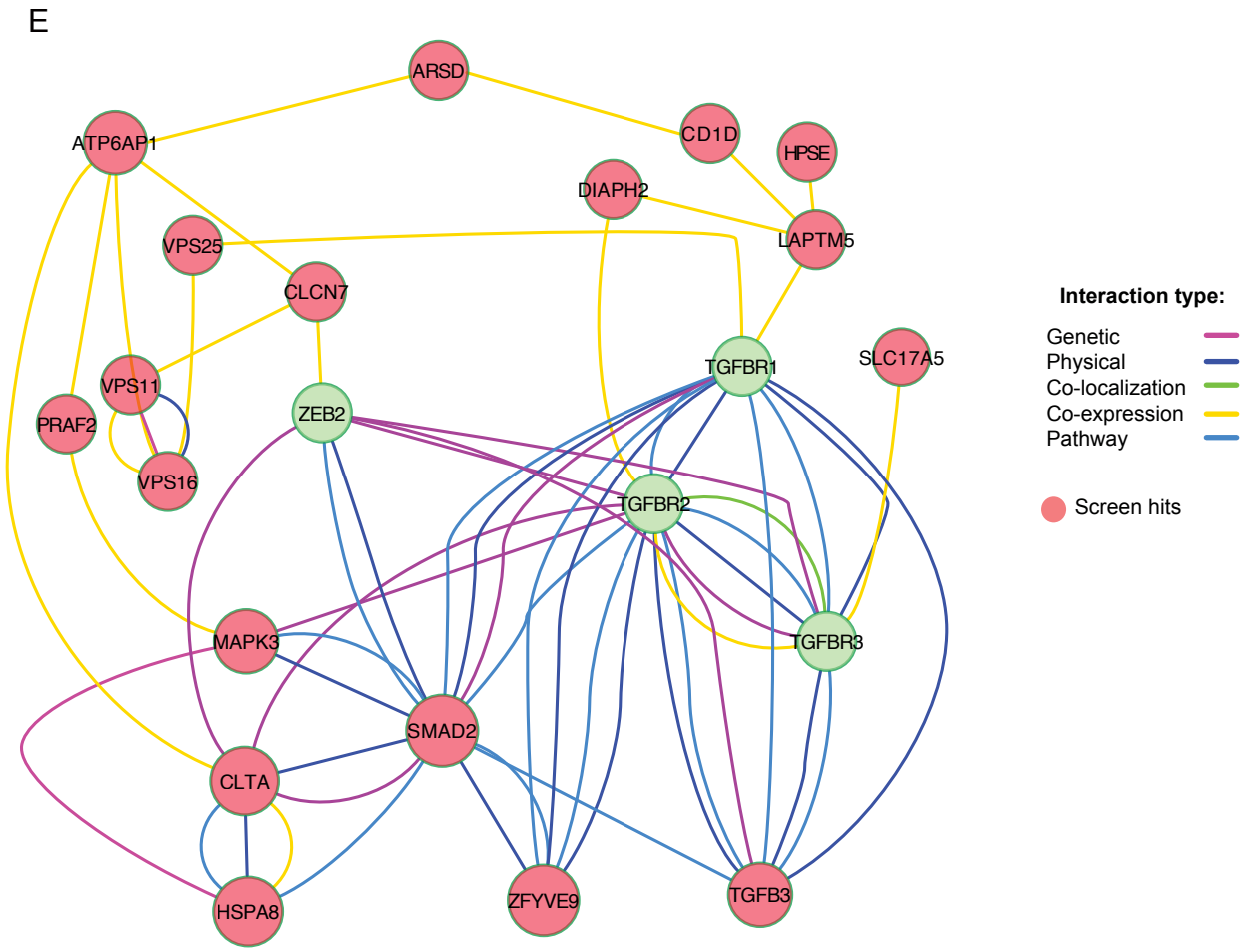
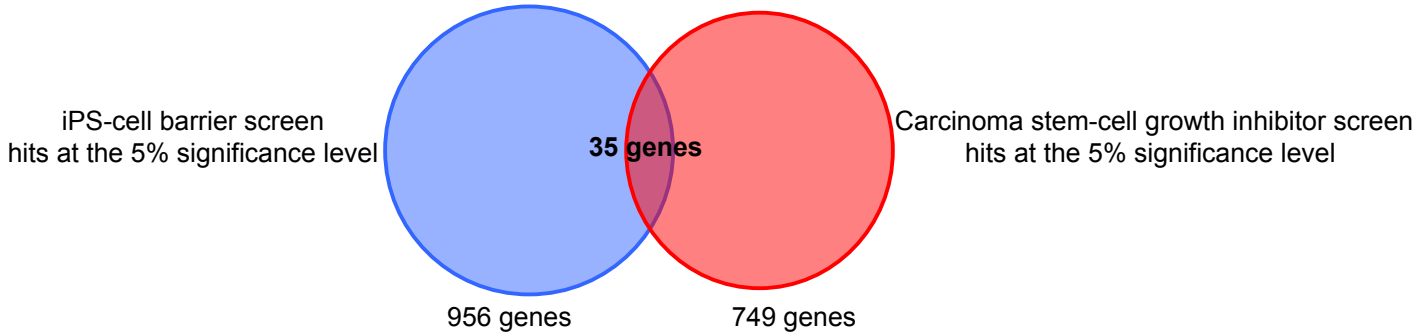

E
ARSD
ATP6AP1
CD1D
HPSE
DIAPH2
LAPTM5
VPS25
CLCN7
SLC17A5
TGFBR1
VPS11
ZEB2
PRAF2
VPS16
TGFBR2
MAPK3
TGFBR3
SMAD2
CLTA
HSPA8
ZFYVE9
TGFB3
Interaction type:
Genetic
Physical
Co-localization
Co-expression
Pathway
Screen hits

Figure S6

A Overlap between a screen for genetic barriers to reprogramming and a screen for genetic barriers to growth in carcinoma stem-cells




B Functional annotations of the overlapping genes