Stem Cell Reports, Volume *18*

Supplemental Information

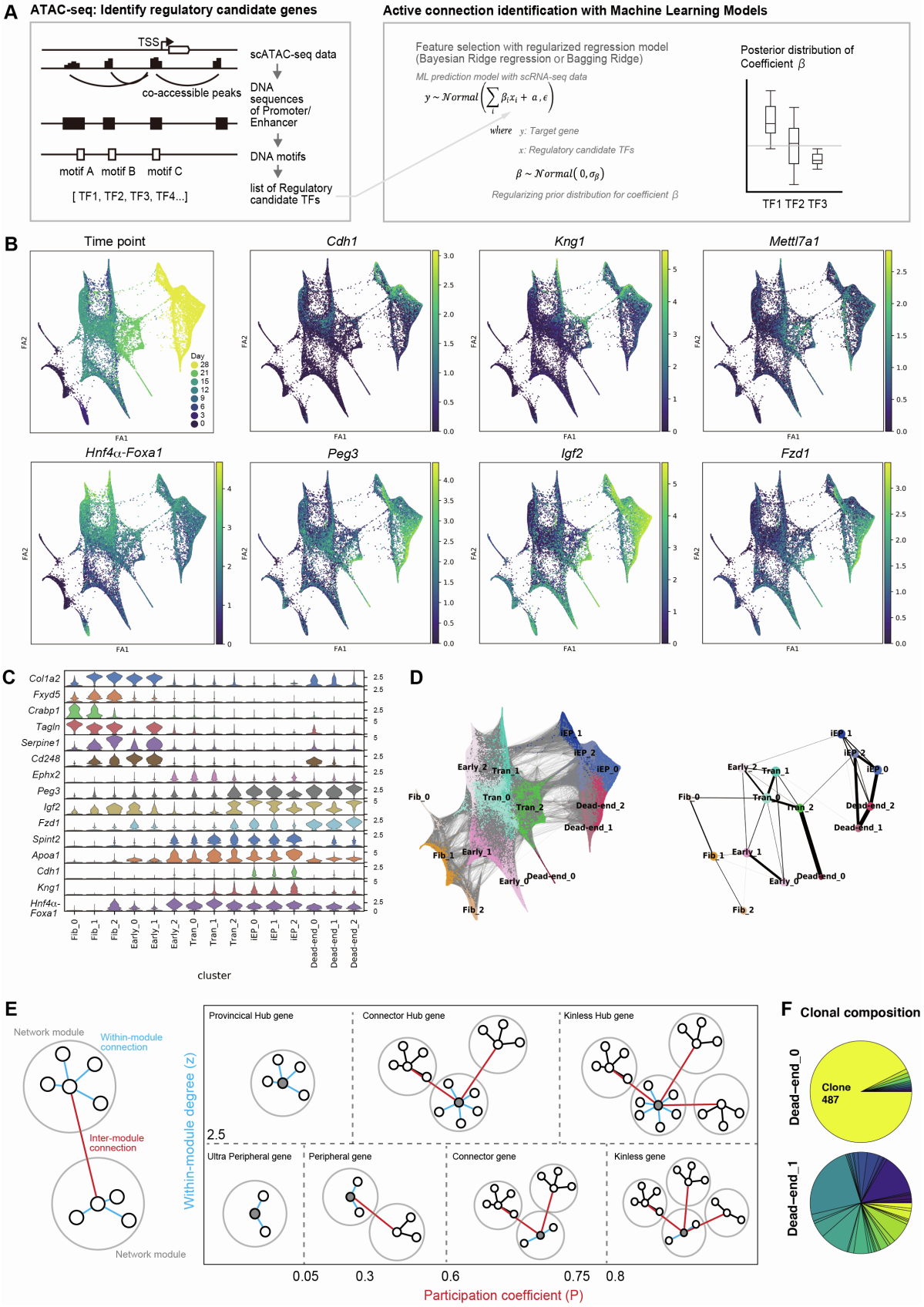
Gene regulatory network reconfiguration in direct lineage reprogramming

Kenji Kamimoto, Mohd Tayyab Adil, Kunal Jindal, Christy M. Hoffmann, Wenjun Kong, Xue Yang, and Samantha A. Morris

Supplemental Figures and Methods

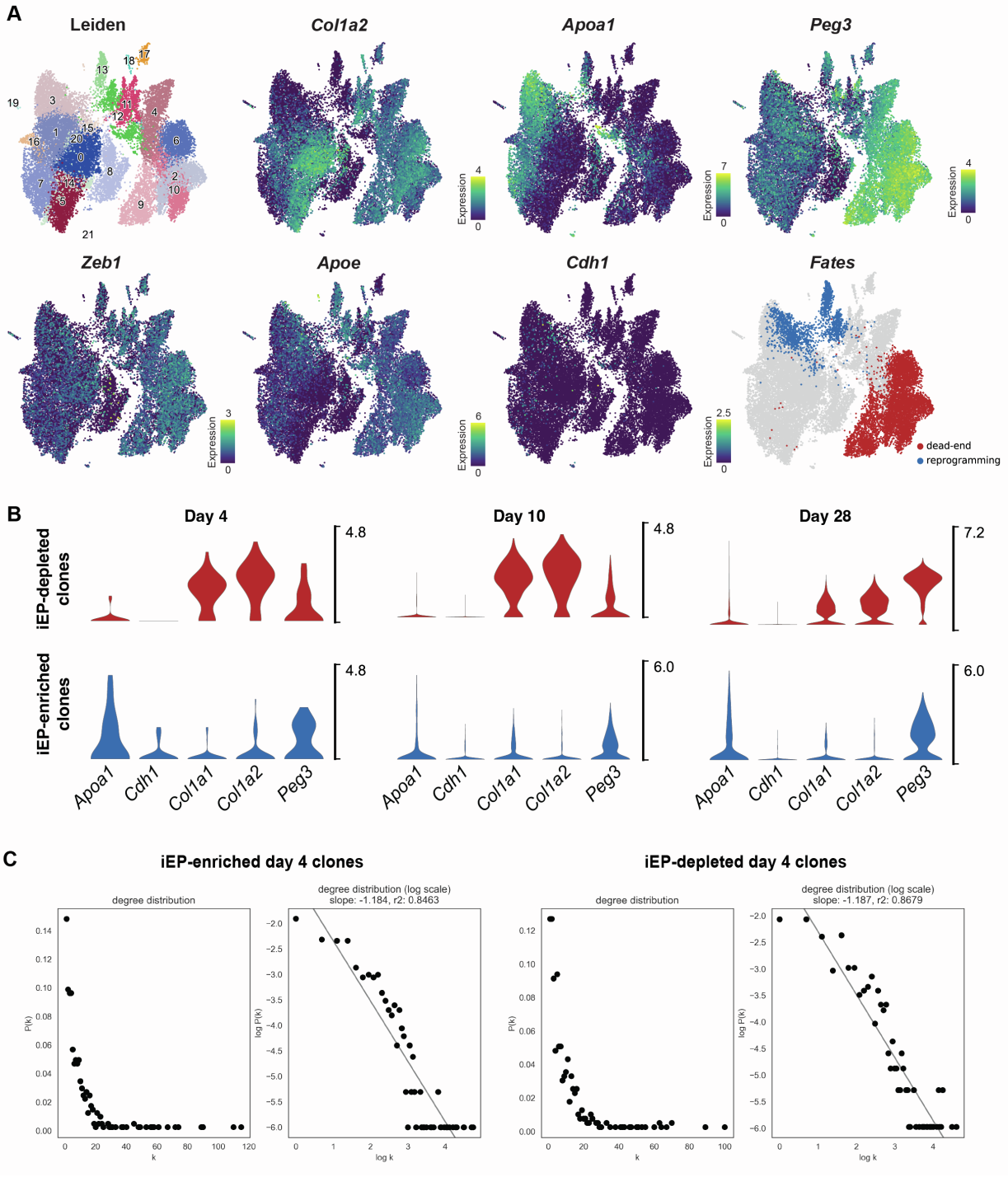
Supplemental Figure 1 (Related to Figure 1). GRN analysis of fibroblast to iEP reprogramming. (A) After base GRN construction (left), single-cell expression data is used to identify active TF-target gene connections for defined cell identities and states. To achieve this, we build a machine learning (ML) model that predicts the relationship between the TF and the target gene. ML model fitting results present the certainty of connection as a distribution, enabling the identification of GRN configurations by removing inactive connections from the base GRN structure. **(B)** Force-directed graph of iEP reprogramming scRNA-seq data (n = 27,663 cells). Reprogramming time point information is projected onto the force-directed graph. There are eight time points; day 0, 3, 6, 9, 12, 15, 21, and 28. *Hnf4α-t2a-Foxa1* (*Hnf4α-Foxa1)* transgene expression levels, and marker gene expression for key iEP states are projected onto the graph. Reprogrammed iEP marker genes: *Cdh1, Apoa1,* and *Kng1*. Fibroblast marker gene: *Col1a2*. Transition marker gene: *Mettl7a1*. Dead-end marker genes: *Peg3, Igf2,* and *Fzd1.* **(C)** Violin plots of marker gene expression in each cluster. **(D)** PAGA connectivity analysis across the reprogramming time course. **(E)** Illustration of the cartography analysis method. The cartography method classifies genes into seven groups according to two network scores: within-module degree and participation coefficient (Guimerà and Amaral, 2005). In complex networks, high-degree nodes (hubs) play the most significant roles in maintaining network structure. **(F)** Pie charts depicting the clonal composition of Dead-end cluster 0 and Dead-end cluster 1. Clone and trajectory information is derived from our previous CellTagging study (Biddy et al., 2018).

Supplemental Figure 1



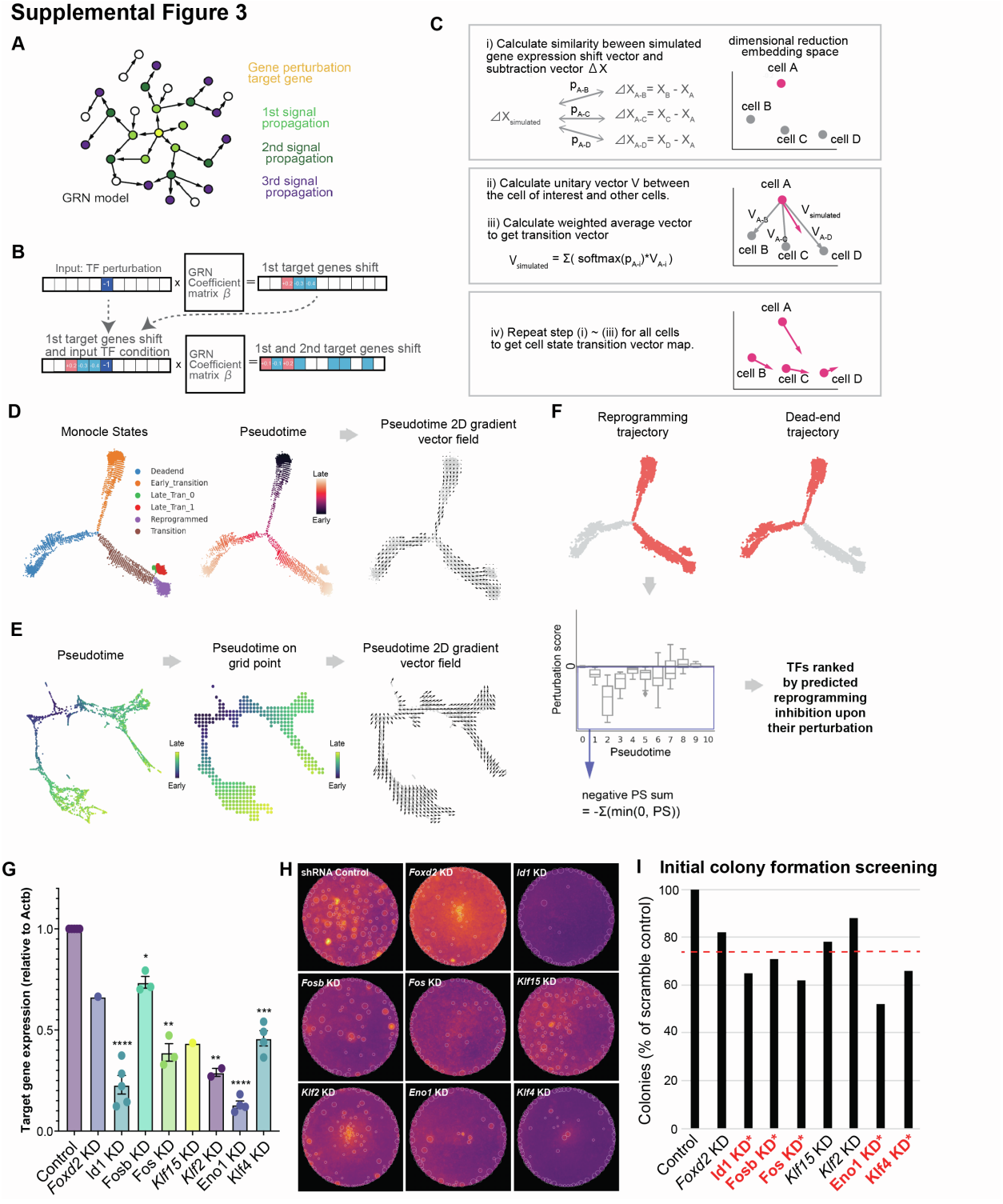
Supplemental Figure 2 (Related to Figure 2). CellOracle network analysis of cells destined to reprogrammed or dead-end fates. (A) Projection of Leiden cluster and gene expression information onto the state-fate UMAP embedding (from **Figure 2C-F**) to identify reprogrammed and dead-end fates. **(B)** Violin plots of reprogrammed (*Apoa1, Cdh1*), fibroblast (*Col1a1, Col1a2*), and dead-end (*Peg3*) marker expression along the iEP-enriched and iEP-depleted trajectories. **(C)** To assess the quality of the inferred networks, we calculate the degree distribution for each GRN configuration after pruning weak network edges based on the p-value and strength. We count the network degree (k), representing the number of network edges for each gene. P(k) is the frequency of network degree k, visualized in scatter plots. We also visualize the relationship between k and P(k) after log transformation, showing that these are scale-free networks, demonstrating successful network inference from these relatively small cell populations.

Supplementary Figure 2



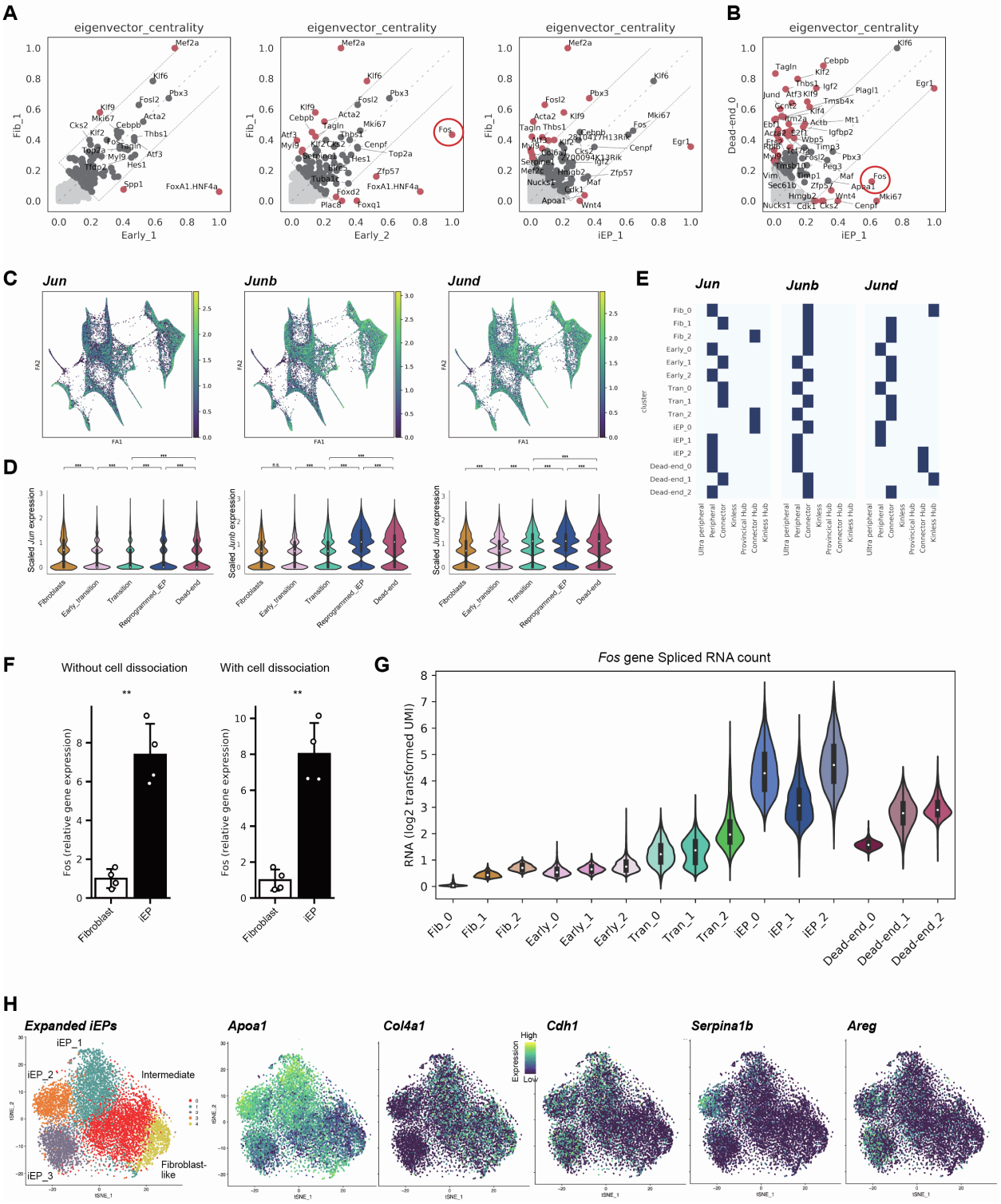
Supplemental Figure 3 (Related to Figure 3). Systematic *in silico* simulation of TF knockout. (A) Overview of signal propagation simulation. CellOracle leverages an inferred GRN model to simulate how target gene expression changes in response to the changes in regulatory gene (TF) expression. The input TF perturbation (yellow) is propagated side-by-side within the network model. **(B)** Leveraging the linear predictive ML algorithm features, CellOracle uses the GRN model as a function to perform the signal propagation calculation. Iterative matrix multiplication steps enable the estimation of indirect and global downstream effects resulting from the perturbation of a single TF. **(C)** After signal propagation, the simulated gene expression shift vector is converted into a 2D vector and projected onto the dimensional reduction space. **(D)** Left: Monocle states identified and used for GRN inference. Right: Calculated pseudotime projected on the Monocle embedding and converted to a 2D gradient vector field. **(E)** Schematic of the method to convert pseudotime to a 2D gradient vector field: First, the pseudotime data is summarized by grid points, then CellOracle calculates a 2D gradient vector of the pseudotime data that represents the directionality of reprogramming pseudotime. **(F)** Outline of reprogramming and dead-end trajectories projected onto the Monocle embedding. The sum of the negative perturbation score was calculated only for reprogramming trajectory clusters in this study. **(G)** Quantitative RT-PCR (qRT-PCR) to validate knockdown efficiency for each shRNA. * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$, **** = $p<0.0001$; unpaired t-test with Welch's correction, two-tailed. **(H)** Colony formation assay (E-cadherin immunohistochemistry) to test iEP reprogramming efficiency following the knockdown of each candidate factor. **(I)** Quantification of colonies formed in the initial screen. Factors marked red and * were selected for further experimental validation.

Supplemental Figure 3



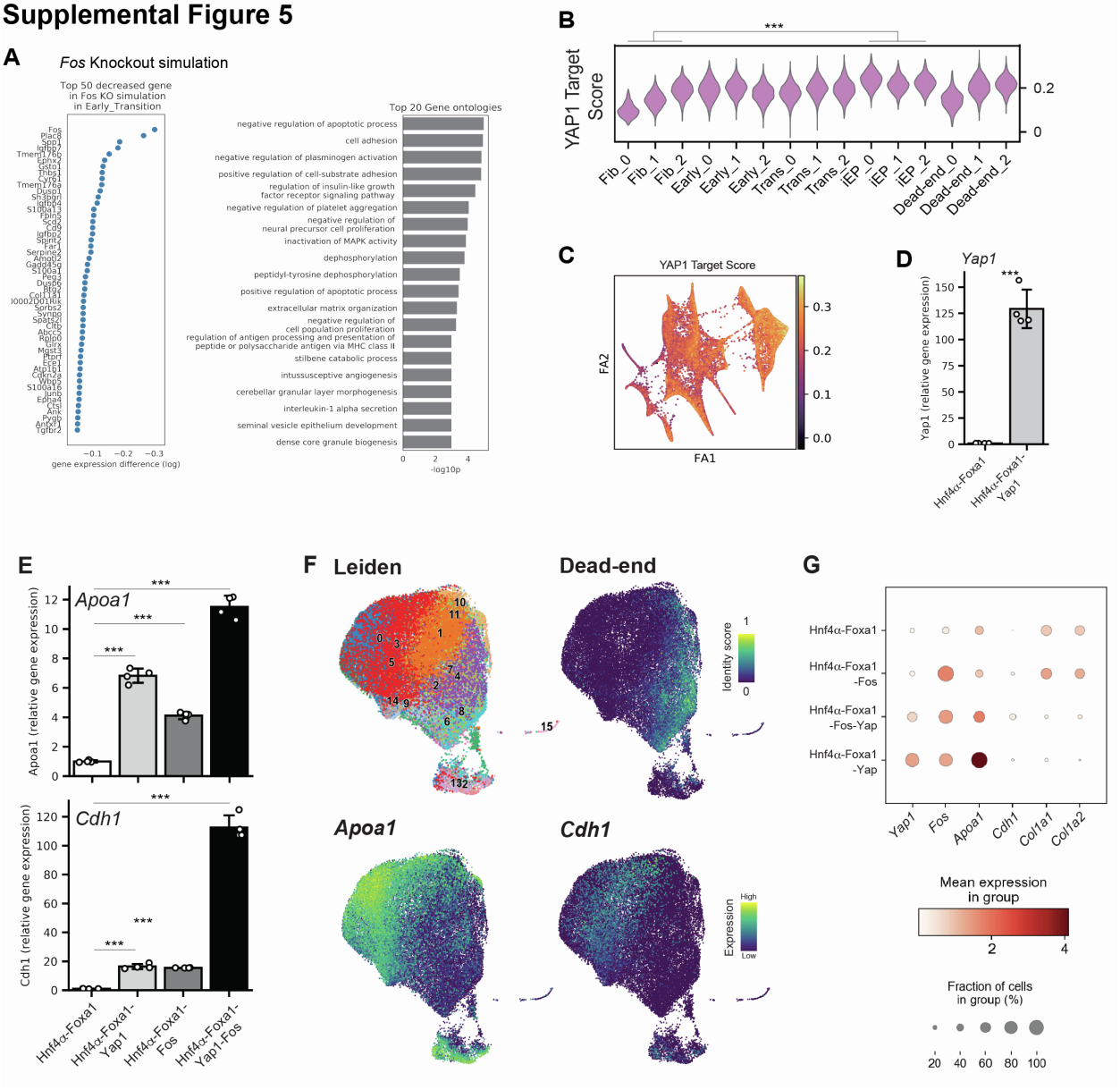
Supplemental Figure 4 (Related to Figure 4). CellOracle analysis of the role of Fos in fibroblast to iEP reprogramming. (A) Comparison of eigenvector centrality scores between the Fib_1 cluster GRN configuration and the GRN configurations of other clusters in relatively early stages of reprogramming. **(B)** Comparison of eigenvector centrality scores between iEP_1 and Dead-end_0 cluster GRN configurations. **(C-E)** Expression and network cartography of Jun family members, *Jun, Junb,* and *Jund*. **(F)** qRT-PCR of *Fos* expression in fibroblasts and iEPs, with and without cell dissociation prior to the assay, ** = $P < 0.01$, *t*-test, one-sided. **(G)** Analysis of *Fos* mRNA splicing state in the scRNA-seq data of iEP reprogramming to investigate the *Fos* mRNA maturation state: Violin plot for spliced *Fos* mRNA counts. **(H)** *t*-SNE plots of 9,914 expanded iEPs, cultured long-term, revealing fibroblast-like, intermediate, and three iEP subpopulations. Expression levels of *Apoa1* (marking typical iEPs)*, Col4a1* (fibroblast-like cells)*, Cdh1, Serpina1b* (hepatic-like iEPs)*,* and *Areg* (intestine-like iEPs) projected onto the *t*-SNE plot.

Supplemental Figure 4



Supplemental Figure 5 (Related to Figure 5). The role of Fos and Yap1 in fibroblast to iEP reprogramming. (A) Top 50 decreased genes in *Fos* knockout simulation in the early reprogramming transition (left) and GO analysis based on these genes (right). **(B)** Violin plot of YAP1 target gene scores across reprogramming, which are significantly enriched as reprogramming progresses (*** = $P < 0.001$, permutation test, one-sided). **(C)** Projection of YAP1 target gene scores onto the force-directed graph of reprogramming. **(D)** qRT-PCR assay for *Yap1* expression following addition of Yap1 and Fos to the Hnf4α-Foxa1 reprogramming cocktail (n = 4 independent biological replicates; *** = $P < 0.001$, ** = $P < 0.01$, *t*-test, one-sided), confirming Yap1 overexpression. **(E)** qRT-PCR assay for iEP marker expression (*Apoa1* and *Cdh1*) following addition of Yap1 and Fos to the Hnf4α-Foxa1 reprogramming cocktail (n = 4 independent biological replicates; *** = $P < 0.001$, ** = $P < 0.01$, *t*-test, one-sided). **(F)** Projection of Leiden cluster, dead-end identity scores, and gene expression information onto the state-fate UMAP embedding (from **Figure 5D, E**). **(G)** Expression of key marker genes for each reprogramming cocktail.

Supplemental Figure 5



Supplemental Table 1. Differentially expressed iEP markers from (Biddy et al., 2018). Top-ranked genes from CellOracle *in silico* perturbation are marked in red.

Supplemental Table 2. Top 50 CellOracle-inferred *Fos* targets across all reprogramming clusters. Confirmed YAP1 targets are highlighted in red.

Supplemental Table 3. Differential expression analysis of day 4 reprogrammed and dead-end destined clones. Genes in bold are also identified by CoSpar analysis. The right column shows TFs prioritized by CellOracle analysis. Genes in bold are also identified by CoSpar analysis.

Supplemental Methods

CellOracle. CellOracle is an integrative tool for GRN inference and network analysis. It consists of several steps: (1) base GRN construction using scATAC-seq data, (2) context-dependent GRN inference using scRNA-seq data, (3) network analysis, and (4) simulation of cell identity after perturbation. We created the algorithm in Python and designed it for use in the Jupyter notebook environment. CellOracle code is open source and available on GitHub (https://github.com/morris-lab/CellOracle), along with detailed function descriptions and tutorials. Further details can be found in the original preprint (Kamimoto et al., 2020).

Alignment and digital gene expression matrix generation. The Cell Ranger v6.0.1 pipeline (https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest) was used to process data generated using the 10x Chromium platform. Cell Ranger processes, filters, and aligns reads generated with the Chromium single-cell RNA sequencing platform. This pipeline was used with a custom reference genome, created by concatenating the sequences corresponding to the *Hnf4α-t2a-Foxa1* transgene as a new chromosome to the mm10 genome. The unique UTRs in the *Hnf4α-t2a-Foxa1* transgene construct allowed us to monitor transgene expression. To create Cell Ranger compatible reference genomes, the references were rebuilt according to instructions from 10x (https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/advanced/references). To achieve this, we first created a custom gene transfer format (GTF) file, containing our transgenes, followed by indexing of the FASTA and GTF files, using Cell Ranger 'mkgtf' and 'mkref' functions. Following this step, the default Cell Ranger pipeline was implemented, then the filtered output data was used for downstream analyses.

CellTag clone calling

Reads containing the CellTag sequence were extracted from the processed and filtered BAM files produced by the 10x Genomics pipeline using our CellTagR pipeline: https://github.com/morris-lab/CellTagR. The resulting filtered CellTag UMI count matrix was then used for all downstream clonal and lineage analyses. The CellTag matrix was initially filtered by removing CellTags that do not appear on the allowlist generated for each CellTag plasmid library. Cells expressing more than 20 CellTags (likely corresponding to cell multiplets) and less than 2 CellTags per cell were filtered out. To identify clonally related cells, Jaccard analysis using the R package Proxy was used to calculate the similarity of CellTag signatures between cells. Clones

were defined as groups of 2 or more related cells. Clones were called on cells pre-filtered for numbers of genes, UMIs, and mitochondrial RNA content.

Cell type classification with Capybara

Cells reprogrammed with Hnf4α-Foxa1, Hnf4α-Foxa1-Fos, Hnf4α-Foxa1-Yap1, and Hnf4α-Foxa1-Fos-Yap1 were classified using Capybara (Kong et al., 2022). Briefly, the single-cell datasets were processed, filtered, and clustered using Seurat, resulting in 35,241 cells (7,414 HF, 8,771 HF-Fos, 8,549 HF-Yap, 10,507 HF-Fos-Yap1). To construct a reference for cell-type classification, we obtained scRNA-seq data of biliary epithelial cells (BECs) and hepatocytes, before and after injury, from GSE125688 (Pepe-Mooney et al., 2019). We built a custom high-resolution reference by incorporating additional tissues from the MCA: fetal liver, MEFs, and embryonic mesenchyme. Following the construction of a high-resolution reference, we performed preprocessing on the reference and the samples, on which we then applied quadratic programming to generate the identity score matrices. Further, we categorized cells into discrete, hybrid, and unknown, calculated the empirical p-value matrices and performed binarization and classification. We calculated the percent composition of each cell type. Cells with hybrid identities were filtered and refined based on their identity scores and representation by more than 0.5% cells of the population. Code and documentation are available at: https://github.com/morris-lab/Capybara.

Differential Gene Expression analysis. Genes differentially expressed between Day 4 reprogramming and dead-end destined cells were identified using Seurat *FindMarkers* command and subsetted to retain hits with an adjusted p-value of less than 0.05 (Bonferroni Correction).

Experimental Methods

Mice and derivation of mouse embryonic fibroblasts. Mouse Embryonic Fibroblasts were derived from E13.5 C57BL/6J embryos. (The Jackson laboratory: 000664). Heads and visceral organs were removed from E13.5 embryos. The remaining tissue was minced with a razor blade and then dissociated in a mixture of 0.05% Trypsin and 0.25% Collagenase IV (Life Technologies) at 37°C for 15 minutes. After passing the cell slurry through a 70μM filter to remove debris, cells were washed and then plated on 0.1% gelatin-coated plates in DMEM supplemented with 10% FBS (Sigma-Aldrich), 2mM L-glutamine and 50mM β-mercaptoethanol (Life Technologies). All animal procedures were based on animal care guidelines approved by the Institutional Animal Care and Use Committee.

Retrovirus Production. Retroviral particles were produced by transfecting 293T-17 cells (ATCC: CRL-11268) with the pGCDN-Sam construct containing Hnf4α-t2a-Foxa1/Fos/Yap1, along with packaging construct pCL-Eco (Imgenex). Virus was harvested 48hr and 72hr after transfection and applied to cells immediately following filtering through a low-protein binding 0.45μM filter.

Lentiviral constructs and lentivirus production. Lentiviral particles were produced by transfecting 293T-17 cells (ATCC: CRL-11268) with the envelope construct pCMV-VSV-G (Addgene plasmid 8454), the packaging construct pCMV-dR8.2 dvpr (Addgene plasmid 8455), and the shRNA expression vector for the respective candidate TF to be knocked down. The shRNA expression vectors (with the TRC2 pLKO.5 backbone) were obtained directly from Millipore-Sigma or cloned into the empty backbone using oligonucleotides (Integrated DNA Technologies). The sequences of shRNA used are SHC202 (non-target shRNA control) CAACAAGATGAAGAGCACCAA; *Eno1* GGCACAGAGAATAAATCTAAA; *Fos* ATCCGAAGGGAACGGAATAAG; *FosB* ATGACGGAAGGACCTCCTTTG; *Foxd2* AGATCATGTCCTCCGAGAGCT *Id1* GAGCTGAACTCGGAGTCTGAA; *Klf2* GACCGATTGTATTTCTATAAG *Klf4* CATGTTCTAACAGCCTAAATG; *Klf15* CTACCCTGGAGGAGATTGAAG. Virus was harvested 48hr and 72hr after transfection and applied to cells following filtering through a low-protein binding 0.45μm filter. For the generation of the complex CellTag library, lentiviral particles were produced by transfecting 293T-17 cells (ATCC: CRL-11268) with the pSMAL-CellTag construct, along with packaging constructs pCMV-dR8.2 dvpr (Addgene plasmid 8455), and pCMV-VSVG (Addgene plasmid 8454), as in (Biddy et al., 2018; Guo et al., 2019; Jindal et al., 2022).

Generation and collection of iEPs. Mouse embryonic fibroblasts (< passage 6) were converted to iEPs as in (Biddy et al., 2018), modified from (Sekiya and Suzuki, 2011). Briefly, we transduced cells every 12hr for two days, with fresh Hnf4α-t2a-Foxa1 retrovirus, in the presence of 4mg/ml Protamine Sulfate (Sigma-Aldrich), followed by culture on 0.1% gelatin-treated plates for one week in hepato-medium (DMEM: F-12, supplemented with 10% FBS, 1 mg/ml insulin (Sigma-Aldrich), dexamethasone (Sigma-Aldrich), 10mM nicotinamide (Sigma-Aldrich), 2mM L-glutamine, 50mM β-mercaptoethanol (Life Technologies), and penicillin/streptomycin, containing 20 ng/ml epidermal growth factor (Sigma-Aldrich). After seven days of culture, the cells were transferred onto plates coated with 5μg/cm^2 Type I rat collagen (Gibco, A1048301). For single-cell processing, 30,000 reprogrammed, expanded iEPs were collected and fixed in methanol, as previously described in (Alles et al., 2017). Briefly, cells were collected and washed in Phosphate

Buffered Saline (PBS), followed by resuspension in ice-cold 80% Methanol in PBS, with gentle vortexing. These cells were stored at -80°C for up to three months and processed on the 10x platform (below). For the state-fate experiments, we followed the above protocol with some slight modifications. We transduced cells every 12hr for two days, with fresh Hnf4α-t2a-Foxa1 retrovirus and added CellTagging lentivirus on the final round of transduction. After 12hr, cells were washed and expanded in hepato-medium for four days, at which point the cells were dissociated and 25% of the population profiled by scRNA-seq. The remaining population was replated, and additional samples were profiled on days 10 and 28.

Colony formation assays. Mouse *Fos* and *Yap1* were cloned from iEPs into the retroviral vector, pGCDNSam (Sekiya and Suzuki, 2011), and retrovirus produced as above. For comparative reprogramming experiments, mouse embryonic fibroblasts ($2x10^5$/well of a 6-well plate) were serially transduced over 72hr (as above). In control experiments, virus produced from an empty vector control expressing only GFP was added to the Hnf4α-Foxa1 reprogramming cocktail. Virus produced from the *Fos* and *Yap1* IRES-GFP constructs was added to the standard Hnf4α and Foxa1 cocktail. Cells underwent reprogramming for two weeks and were processed for colony formation assays: cells were fixed on the plate with 4% PFA, permeabilized in 0.1% Triton-X100 then blocked with the Mouse on Mouse Elite Peroxidase Kit (Vector PK-2200). Primary antibody, mouse anti-E-Cadherin (1:100, BD Biosciences), was applied for 30 min before washing and processing with the VECTOR VIP Peroxidase Substrate Kit (Vector SK-4600). Colonies were visualized on a flatbed scanner, adding heavy cream to each well to increase image contrast. Colonies were counted using our automated colony counting tool: https://github.com/morris-lab/Colony-counter. *Fos and Yap1* overexpression was confirmed by harvesting RNA from Hnf4α-Foxa1 and Hnf4α-Foxa1-Fos/Yap1-transduced cells (RNeasy kit, Qiagen). Following cDNA synthesis (Maxima cDNA synthesis kit, Life Tech), qRT-PCR was performed to quantify *Fos/Yap1* overexpression (TaqMan Probes: Gapdh Mm99999915_g1; *Cdh1* Mm01247357_m1; *Apoa1* Mm00437569_m1; *Fos* Mm00487425_m1; *Yap1* Mm01143263_m1; TaqMan qPCR Mastermix, Applied Biosystems).

Colony formation assays for TF knockdowns were conducted similarly, with the following modifications. To initiate reprogramming, mouse embryonic fibroblasts ($75x10^3$/well of a 6-well plate) were serially transduced over 72hr (as above). Lentivirus produced from the non-target shRNA control and the respective TF knockdown shRNA constructs was then added at 84hr and 96hr (only added at 96hr for the initial screen). At 120hr, cells were seeded for colony formation

assays ($40x10^3$cells/well of a 6-well plate), which were then processed for colony formation on day 14 as above. The remaining cells from each sample were seeded for harvesting RNA for qPCR on day 14, as above. In the initial screen, cells from each sample were split equally and seeded in 6 well plates for colony formation and RNA extraction at 15 days following reprogramming initiation. For *Fos* and *Fosb* knockdowns, mouse embryonic fibroblasts ($120x10^3$ in a 6-cm dish) were transduced with the respective shRNA lentivirus at 24hr and 36hr post-seeding. qPCR confirmation was performed at 72hr post-seeding. TaqMan Probes used: *Actb* Mm02619580_g1; *Eno1* Mm01619597_g1; *Fos* Mm00487425_m1; *Fosb* Mm00500401_m1; *Foxd2* Mm00500529_s1; *Id1* Mm00775963_g1; *Klf2* Mm00500486_g1; *Klf4* Mm00516104_m1; *Klf15* Mm00517792_m1.

CRISPR/Cas9 *Fos* Knockout

The *Fos* knockouts were performed as part of a larger screen, using Perturb-seq as previously described (Dixit et al., 2016). The protocol was modified, as outlined below, to apply the strategy to our experimental system:

(1) Vector backbone and gene barcode pool construction: For Perturb-seq experiments, we used a lentivirus vector to express guide RNAs and gene barcodes (GBC). The lentivirus vector backbone contains an antiparallel cassette containing a guide RNA and GBC. In the original perturb-seq paper, the authors used pPS and pBA439 to construct the guide RNA-GBC vector pool. Here, we modified pPS and pBA439 to generate the pPS2 vector, in which the Blasticidin-t2a-BFP gene replaced the Puromycin-t2a-BFP gene. We constructed the guide RNA-GBC vector using a multi-step cloning strategy: First, we synthesized dsDNA, via PCR, for a random GBC pool. We purified the PCR product with AMPure XP SPRI beads. We inserted the purified GBC pool into the pPS2 vector at the EcoRI site in the 3' UTR of the Blasticidin-t2a-BFP gene. We used the product of Gibson assembly for transformation into DH5α competent cells (NEB: C2987H). Transformed cells were cultured directly in LB. We extracted plasmid DNA to yield the pPS2-GBC pool.

(2) Guide RNA cloning. We designed guide RNAs using https://zlab.bio/guide-design-resources. We synthesized oligo DNA for each guide RNA. Oligo DNA pairs were annealed and inserted into the pPS2-GBC vector following BsmB1 digestion. After isolation and growth of single colonies, plasmid DNA was extracted and sanger DNA sequenced; sequences of the guide RNA inserted site and GBC site were used to construct a gRNA/GBC reference table:

Fos_sg0	CAGCCGACTGAACGCGTTATTC
Fos_sg1	CATATATCAAAGATGAACATTG
Fos_sg2	TCAAGGCTGTAATTTCTTGGGC
empty0	TTGATGAACTGCGCTAGCGAGG
empty1	AAGAGCGGCTCGCAAGGGAAAA
empty2	AGTAGGATACGTGGAGTTAATA

(3) Lentivirus guide RNA pool generation. An equal amount of DNA for each pPS2-guide RNA vector was mixed to generate the plasmid pool. Three control vectors were also mixed with this plasmid vector pool; the weight ratio of each pPS2-guide vector to each control vector was 1:4. We used this mixed DNA pool for lentivirus production. Lentiviral particles were produced by transfecting 293T-17 cells (ATT: CRL-11268) with the pPS-guide RNA-GBC constructs, along with the packaging plasmid, psPAX2 (https://www.addgene.org/12260/), and pMD2.G (https://www.addgene.org/12259/).

(4) Cell culture for Perturb-seq. We transduced reprogrammed iEP cells with retrovirus carrying Cas9 (MSCV-Cas9-Puro). The cells were treated with Puromycin (4 μg/ml) for four days to eliminate non-transduced cells. iEP-Cas9 cells were transduced with the lentivirus guide RNA pool for 24 hours. The concentration of lentivirus was pre-determined to target 10~20% transduction efficiency. After four days of cell culture, we flow sorted BFP-positive cells to purify transduced cells. Cells were cultured for a further 72 hours and fixed with methanol, as previously described (Alles et al., 2017).

(5) GBC amplification and sequencing. Following library preparation on the 10x Chromium platform (below), we PCR amplified the GBC. The amplification was performed according to the original perturb-seq paper (Dixit et al., 2016), but we modified the PCR primer sequence for the Chromium single-cell library v2 kit:

P7_ind_R2_BFP_primer:
CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTGACTGGAGTTCAGACGTGTGCTCTTC
CGATCTTAGCAAACTGGGGCACAAGC
P5_partial_primer: AATGATACGGCGACCACCGA
GBG_Amp_F: GCTGATCAGCGGGTTTAAACGGGCCCTCTAGG

GBG_Amp_R: CGCGTCGTGACTGGGAAAACCCTGGCGAATTG

GBC_Oligo:

TTAAACGGGCCCTCTAGGNNNNNNNNNNNNNNNNNNNNNNCAATTCGCCAGGGTTTTCCC

Following amplification, we purified the PCR product with AMPure XP SPRI beads. The purified sample was sequenced on the Illumina Mi-seq platform.

(6) Alignment of cell barcode/GBC. For preprocessing of Perturb-seq metadata, we used MIMOSCA, a computational pipeline to analyze perturb-seq data (https://github.com/asncd/MIMOSCA). First, the reference table for the cell barcode/GBC pair was generated from Fastq files. The data table was converted into the guide RNA/cell barcode table using the guide RNA-GBC reference table. This metadata was integrated into the scRNA-seq data. The guide metadata was processed with an EM-like algorithm in MIMOSCA to filter out unperturbed cells computationally, as previously described (Dixit et al., 2016).

10x procedure. For single-cell library preparation on the 10x Genomics platform, we used: the Chromium Single Cell 3′ Library & Gel Bead Kit v2 (PN-120237), Chromium Single Cell 3′ Chip kit v2 (PN-120236), and Chromium i7 Multiplex Kit (PN-120262), according to the manufacturer's instructions in the Chromium Single Cell 3′ Reagents Kits V2 User Guide. Prior to cell capture, methanol-fixed cells were placed on ice, then spun at 3000rpm for 5 minutes at 4°C, followed by resuspension and rehydration in PBS, according to (Alles et al., 2017). 17,000 cells were loaded per lane of the chip, aiming to capture 10,000 single-cell transcriptomes. The resulting cDNA libraries were quantified on an Agilent Tapestation and sequenced on an Illumina HiSeq 2500.

References

Alles, J., Karaiskos, N., Praktiknjo, S.D., Grosswendt, S., Wahle, P., Ruffault, P.-L., Ayoub, S., Schreyer, L., Boltengagen, A., Birchmeier, C., et al. (2017). Cell fixation and preservation for droplet-based single-cell transcriptomics. BMC Biol *15*, 44. https://doi.org/10.1186/s12915-017-0383-5.

Biddy, B.A., Kong, W., Kamimoto, K., Guo, C., Waye, S.E., Sun, T., and Morris, S.A. (2018). Single-cell mapping of lineage and identity in direct reprogramming. Nature *564*, 219–224. https://doi.org/10.1038/s41586-018-0744-4.

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with

Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell *167*, 1853-1866.e17. https://doi.org/10.1016/j.cell.2016.11.038.

Guimerà, R., and Amaral, L.A.N. (2005). Cartography of complex networks: modules and universal roles. Journal of Statistical Mechanics: Theory and Experiment *2005*, P02001. https://doi.org/10.1088/1742-5468/2005/02/P02001.

Guo, C., Kong, W., Kamimoto, K., Rivera-Gonzalez, G.C., Yang, X., Kirita, Y., and Morris, S.A. (2019). CellTag Indexing: genetic barcode-based sample multiplexing for single-cell genomics. Genome Biol *20*, 90. https://doi.org/10.1186/s13059-019-1699-y.

Jindal, K., Adil, M.T., Yamaguchi, N., Wang, H.C., Yang, X., Kamimoto, K., Rivera-Gonzalez, G.C., and Morris, S.A. (2022). Multiomic single-cell lineage tracing to dissect fate-specific gene regulatory programs. BioRxiv 2022.10.23.512790. https://doi.org/10.1101/2022.10.23.512790.

Kamimoto, K., Hoffmann, C.M., and Morris, S.A. (2020). CellOracle: Dissecting cell identity via network inference and in silico gene perturbation. BioRxiv 2020.02.17.947416. https://doi.org/10.1101/2020.02.17.947416.

Kong, W., Fu, Y.C., Holloway, E.M., Garipler, G., Yang, X., Mazzoni, E.O., and Morris, S.A. (2022). Capybara: A computational tool to measure cell identity and fate transitions. Cell Stem Cell *29*, 635-649.e11. https://doi.org/10.1016/J.STEM.2022.03.001.

Pepe-Mooney, B.J., Dill, M.T., Alemany, A., Ordovas-Montanes, J., Matsushita, Y., Rao, A., Sen, A., Miyazaki, M., Anakk, S., Dawson, P.A., et al. (2019). Single-Cell Analysis of the Liver Epithelium Reveals Dynamic Heterogeneity and an Essential Role for YAP in Homeostasis and Regeneration. Cell Stem Cell *25*, 23-38.e8. https://doi.org/10.1016/J.STEM.2019.04.004.

Sekiya, S., and Suzuki, A. (2011). Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. Nature *475*, 390–393. https://doi.org/10.1038/nature10263.