

Figure S1: Predicted kinetics and empirical detection of induced reprogramming factors. (A) Half-life estimates for the mRNA (left) and protein (right) products of Yamanaka Factor target genes. Each point is a TF or target gene extracted from the TRRUST or Enrichr TF gene set databases. Median half-lives from public databases are shown as black bars for each program (both panels). (B) We estimated the fraction of Yamanaka Factor program products (mRNA or protein) present at each day of the chase period (3 days total) in our experiments. By the final day of the chase, < 0.05% of any program is still present at the mRNA level and < 2% at the protein level. (C) Detection of the polycistronic reprogramming inducible transgene in MSCs and adipogenic cells. After three days of chase, the inducible transgene is only detectable in a small fraction of cells (6% average) while the constitutive reporter on the same vector (mCherry) is detected in the majority of cells (mean 64%) at much higher levels ($\log_2 \text{Cherry}/\text{Y4TF} > 6$). (D) Detection of the inducible reprogramming transgenes in MSCs and adipogenic cells in the pooled reprogramming system. We cannot reliably distinguish inducible transgene (iTg) expression across vectors with our 3' RNA-seq assay, so we report all reads mapping to an inducible transgene region. Inducible transgenes are detected in a small fraction of cells (mean 6.5% in MSCs, 3.2% in adipogenic cells), while the constitutive reporters and associated barcodes on the same vector are detected with the same aligner in almost all barcode demultiplexed cells (mean > 98.7%). The deviation from 100% constitutive reporter detection in the demultiplexed cells is likely due to alignment failures relative to our more sensitive barcode demultiplexing procedure (Methods).

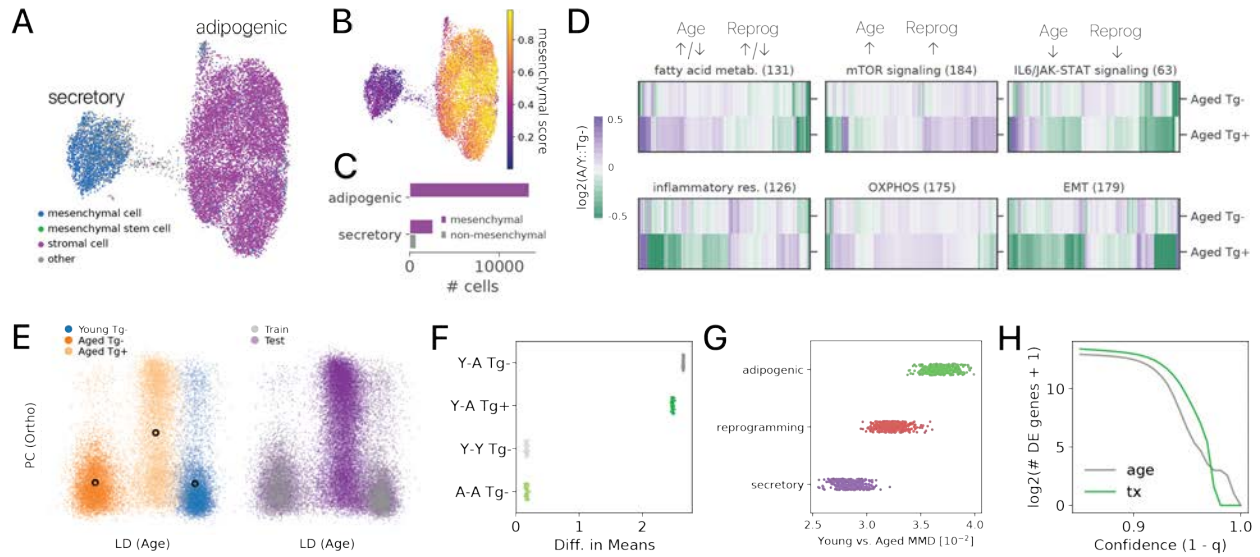


Figure S2: Polycistronic reprogramming in adipogenic cells. (A) Adipogenic cell profiles from the “adipogenic” and “secretory” states embedded with UMAP, labeled with cell type predictions from an scNym classifier trained on the *Tabula Muris*. Almost all cells are predicted to have mesenchymal cell types. (B) scNym confidence scores for each cell profile in the adipogenic cell populations. Secretory cells are less confident than non-secretory adipogenic cells. (C) Quantification of cell type classifications in (A). A large majority of cells in each state are given mesenchymal cell labels. (D) Gene expression levels for all genes in each of six Hallmark gene expression programs. The number of genes in each program is included in the heatmap title. Some programs show opposing effects of aging and reprogramming (left panels), while others show convergent effects of aging and reprogramming (center and right panels). (E) Linear embedding capturing the effects of aging (linear discriminant, LD) and orthogonal reprogramming effects (principal component, PC). Linear discriminant analysis readily separates young and aged control populations, while aged reprogrammed cells interpolate between the two (left). The LD is likewise performant on held-out control cells (test; right). Circled points highlight the shift in mean coordinates for each population. Reprogramming shifts aged cells toward the youthful state, but also induces orthogonal effects along the PC axis. (F) Reprogramming reduces the difference in mean scVI embedding coordinates between aged and young, control cells, matching the results of the MMD statistic. Heterochronic Young-Aged (Y-A) contrasts compare Young Tg- cells to either Aged Tg- or Aged Tg+ populations, while our isochronic contrasts (Y-Y, A-A) used random samples of control Tg- cells from the same age. (G) MMD statistics computed between young and aged cells in each annotated cell state show that reprogramming states are less distinct with age than control adipogenic states. (H) Differential expression analysis shows that reprogramming (tx) induces more differentially expressed genes, but fewer high confidence DE genes than aging (age). The number of reprogramming genes for each contrast is plotted as a function of the false discovery rate (confidence, $1 - q$). Many reprogramming DE genes represent orthogonal effects, as shown in (E).

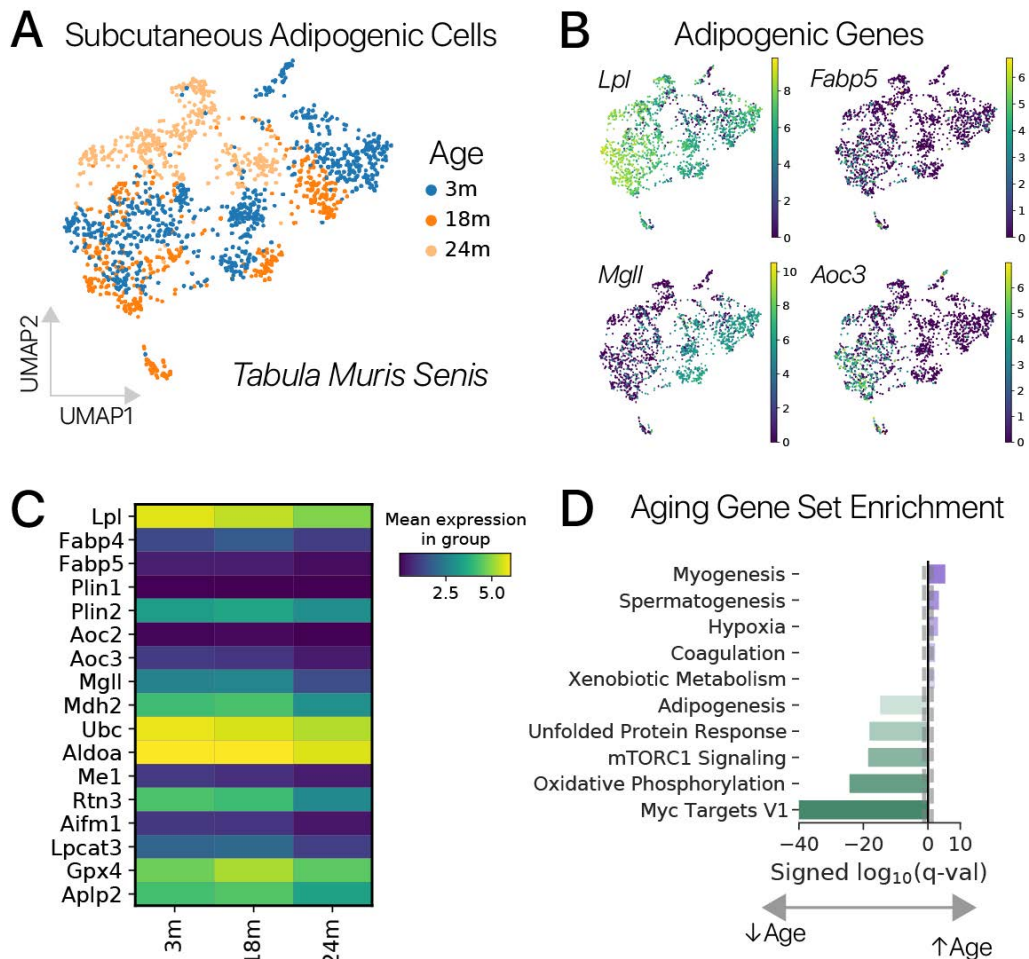


Figure S3: **Adipogenic cells *in vivo* show similar features of aging to adipogenic cells *in vitro*.** (A) Adipogenic cells from the subcutaneous adipose tissue in the *Tabula Muris Senis* [23] embedded with UMAP, colored by age. Young cells (3m, months) and aged cells (24m) clearly separate in transcriptional space, similar to our *in vitro* data. (B) Expression of several adipogenic genes is restricted to young cells. (C) Many adipogenic genes show significantly decreased expression with age, similar to *in vitro* results ($q < 0.05$, t -test). (D) Gene set enrichment analysis for genes significantly changed between 3 months of age (3m) and 24 months of age (24m) reveals strong downregulation of Adipogenesis, Oxidative Phosphorylation, and Unfolded Protein Response programs (Hallmark MSigDB, $q < 0.01$, Fisher's exact test).

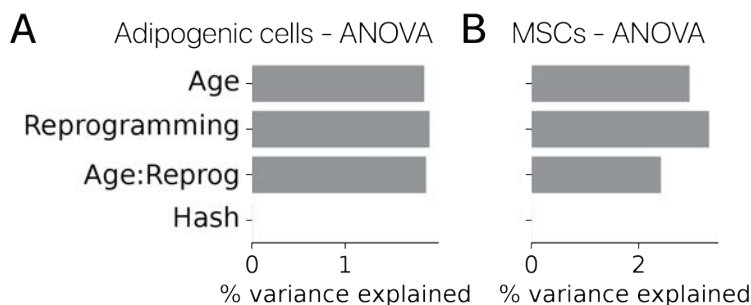


Figure S4: **Age and reprogramming effects are the dominant sources of variation in partially reprogrammed cells.** (A) ANOVA results in partially reprogrammed adipogenic cells, where the remaining variation is unexplained by the covariates shown. (B) ANOVA results in partially reprogrammed MSCs. In both cell types, Age and Reprogramming Treatment are the major source of explained variation, while the animal of origin (Hash) is a minor contributor.

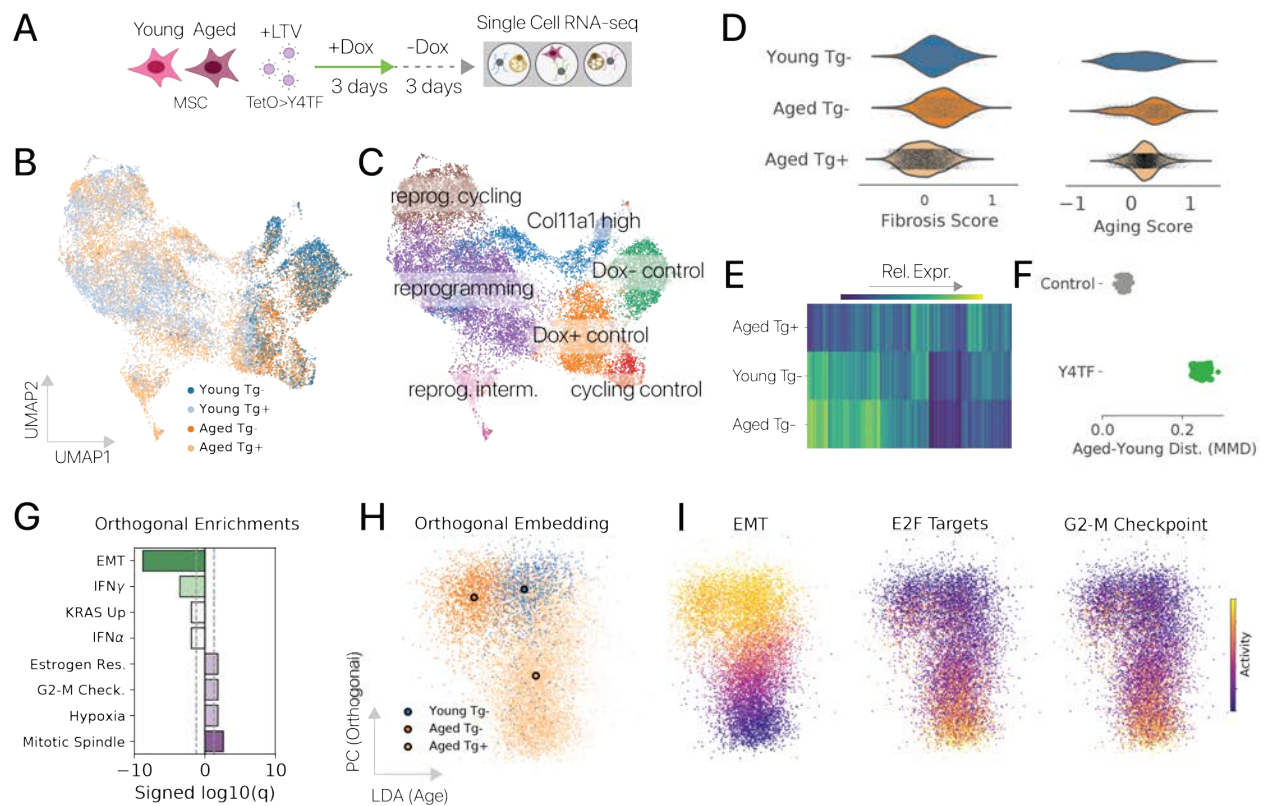


Figure S5: Partial reprogramming in muscle-derived MSCs remodels gene expression. (A) Experimental schematic. Muscle-derived MSCs were transduced with a tetracycline-inducible polycistronic Yamanaka Factor cassette and partially reprogrammed with a 3 Dox day pulse, 3 day Dox chase. (B) Reprogrammed cells (Tg+) clearly segregate from control cells (Tg-) in a latent transcriptional space, while age is smaller source of variation. (C) Single cell profiles resolve distinct cell states induced by reprogramming. (D) Partial reprogramming ameliorates age-related increases in fibrosis gene sets and an aging gene score derived from bulk RNA-seq experiments. (E) Youthful gene expression is restored in hundreds of genes by partial reprogramming, despite the large magnitude of reprogramming effects orthogonal to the axis of aging. (F) Partial reprogramming increases the maximum mean discrepancy between young control cells and aged cells, suggesting that many effects of reprogramming are orthogonal to the axis of aging. (G) We performed gene set analysis on reprogramming genes that were not changed with age, which we term “orthogonal effects.” Significant enrichments for EMT and cell cycle programs suggests that suppression of mesenchymal identity and activation of the cell cycle are dominant effects unrelated to MSC aging. Dashed grey lines indicate significance thresholds ($q < 0.05$). (H) Projection of age and orthogonal variation in MSC single cell data using linear discriminant (LD) analysis on aging genes and the top PC from remaining genes. Circled large points represent group medians in the embedding. (I) Activity of Hallmark gene sets in the orthogonal projection (scaled [0, 1]). Gene programs largely vary along the orthogonal axis, with minimal variability along the aging axis.

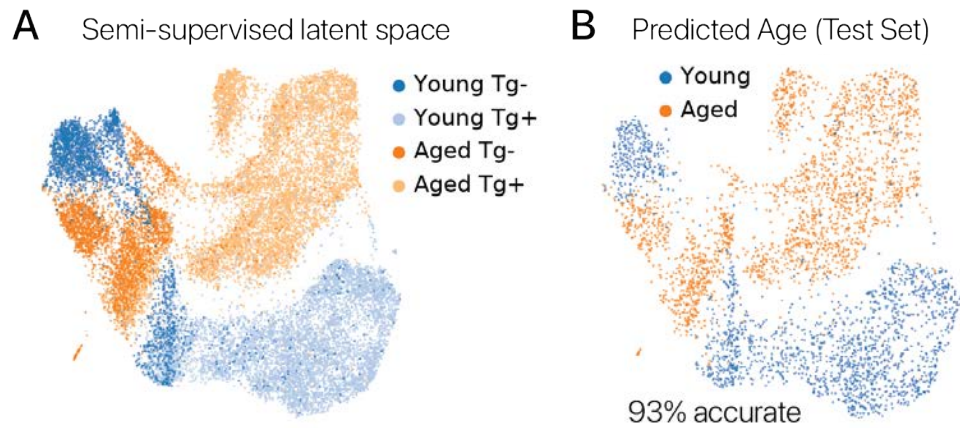


Figure S6: **Semi-supervised aging classifiers demonstrate segregation of MSCs by age in transcriptional space.** (A) MSCs embedded in a latent space using a semi-supervised variational autoencoder model. We trained variational autoencoder to jointly reconstruct transcriptomes and classify cell age, learning a representation that segregates young and aged cells in the latent space. (B) Cells from the test set projected in the semi-supervised latent space colored with the class predictions. We found that age predictions were 93% accurate, confirming that features of aging are captured at the transcriptional level.

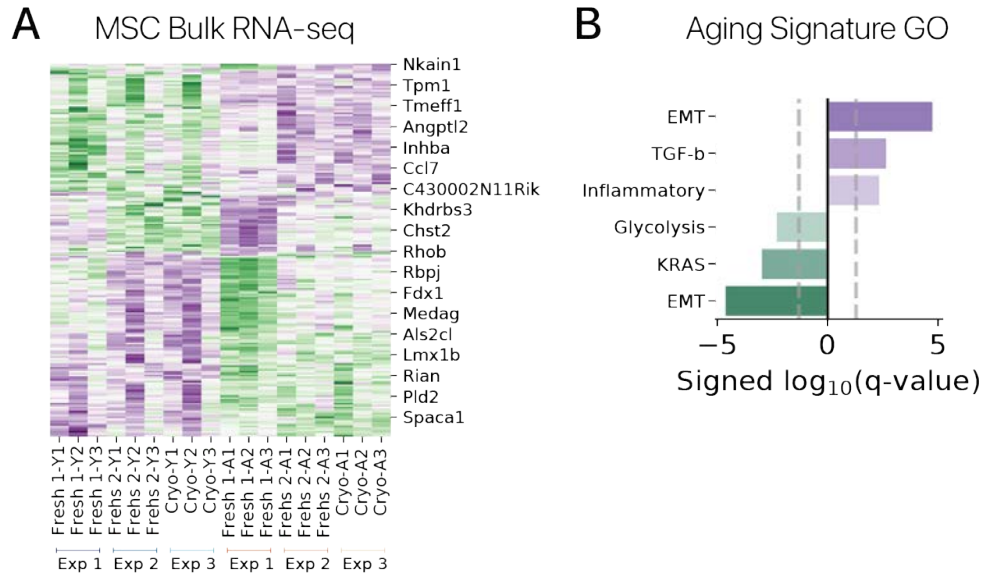


Figure S7: **Bulk RNA-seq reveals a reproducible transcriptional signature of aging in MSCs.** (A) Bulk RNA-seq profiles of young (Y) and aged (A) MSCs collected in three separate experiments. Genes with aging effects in all three experiments are shown. We used these gene sets to derive an “Aging Score” for our transient reprogramming experiment (Fig. S5D). “Fresh” samples were prepared from freshly isolated MSCs after an *in vitro* culture period, while “Cryo” samples were prepared after cryopreservation and subsequent culture. (B) Gene Ontology enrichment analysis for MSigDB Hallmark gene sets. Epithelial-to-mesenchymal transition (EMT) genes are strongly enriched for age-related changes. Both inflammatory genes and the fibrotic signaling factor TGF- β show age-related increases.

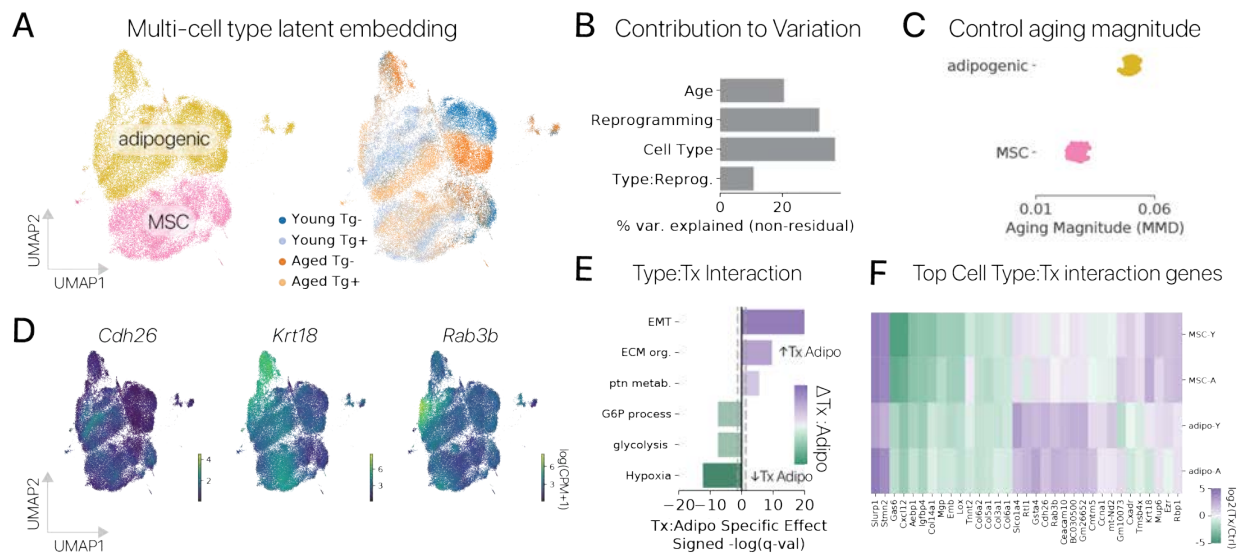


Figure S8: Cell identity dictates the effects of partial reprogramming. (A) Partially reprogrammed adipogenic cells and MSCs from our polyclonic experiments were embedded in a joint latent space with scVI to allow for direct comparison of effect sizes. (B) ANOVA revealed that Cell Type:Reprogramming interactions account for roughly 11% of the total non-residual variation explained by covariates. (C) We re-computed aging magnitudes by maximum mean discrepancy for control adipogenic cells and MSCs in the joint latent space. We found that the magnitude of aging was significantly larger in adipogenic cells ($p < 0.01$, Wilcoxon Rank Sum test). (D) We performed differential expression to identify genes with significant Cell Type:Reprogramming interactions. Qualitative analysis of gene expression in the latent space confirms cell type-specific reprogramming effects. (E) We used Gene Ontology analysis to investigate the most significant Cell Type:Reprogramming interactions. Epithelial-to-mesenchymal transition (EMT) and extracellular matrix (ECM) gene sets were less downregulated in adipogenic cells after reprogramming (positive Adipogenic:Reprogramming interaction coefficients), while glycolytic gene sets were less upregulated (negative interaction coefficients). (F) Clustering a set of genes with significant Cell Type:Reprogramming interactions reveals prominent differences in the effect size of reprogramming across cell types. By contrast, effect sizes across age within the same cell type are fairly consistent.

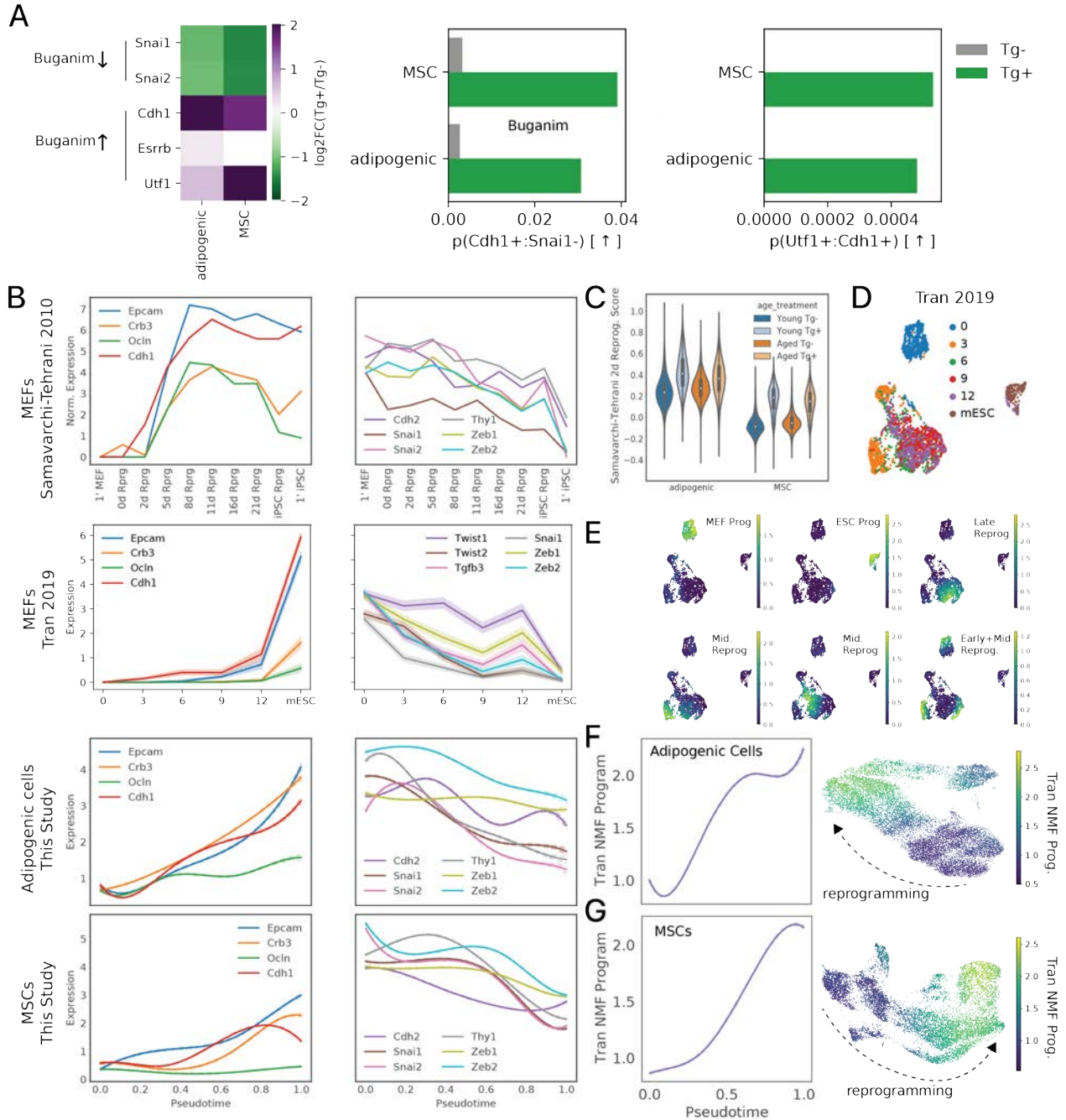


Figure S9: Canonical features of early pluripotent reprogramming are recapitulated by partial reprogramming. To determine if our experiments in adipogenic cells and MSCs are consistent with previous reports of pluripotent reprogramming, we compared our data directly to three previous pluripotent reprogramming studies [28, 29, 19]. **(A)** Heatmap of gene expression changes for mesenchymal-to-epithelial transition (MET) marker genes in our data (left). Arrows indicate the previously reported direction of change [29]. Fraction of cells matching previously reported marker combinations (arrows: previously reported direction). Both marker gene expression and frequency of marker-positive cells match previous reports. **(B)** Expression of reported MET markers across timepoints in two previous studies [28, 19] (upper) and across pseudotime in our data (lower). Marker gene expression dynamics in our data match the early phases of reprogramming in previous reports. **(C)** Activity of a reprogramming score based on the differentially expressed genes at day 2 of reprogramming in a prior microarray study [28]. The reprogramming scores increase in transgene-positive adipogenic cells and MSCs, as expected. **(D)** Single cell RNA-seq data from a previous reprogramming timecourse, labeled with timepoints [19]. **(E)** Activity of latent gene programs derived by non-negative matrix factorization, matching embeddings in (E). **(F)** Activity of an early reprogramming latent gene program derived from the previous study [19] in adipogenic cells and **(G)** MSCs after partial reprogramming. The early reprogramming latent program increases in activity across the reprogramming pseudotime trajectory (GAM fits $\pm 95\%$ CI).

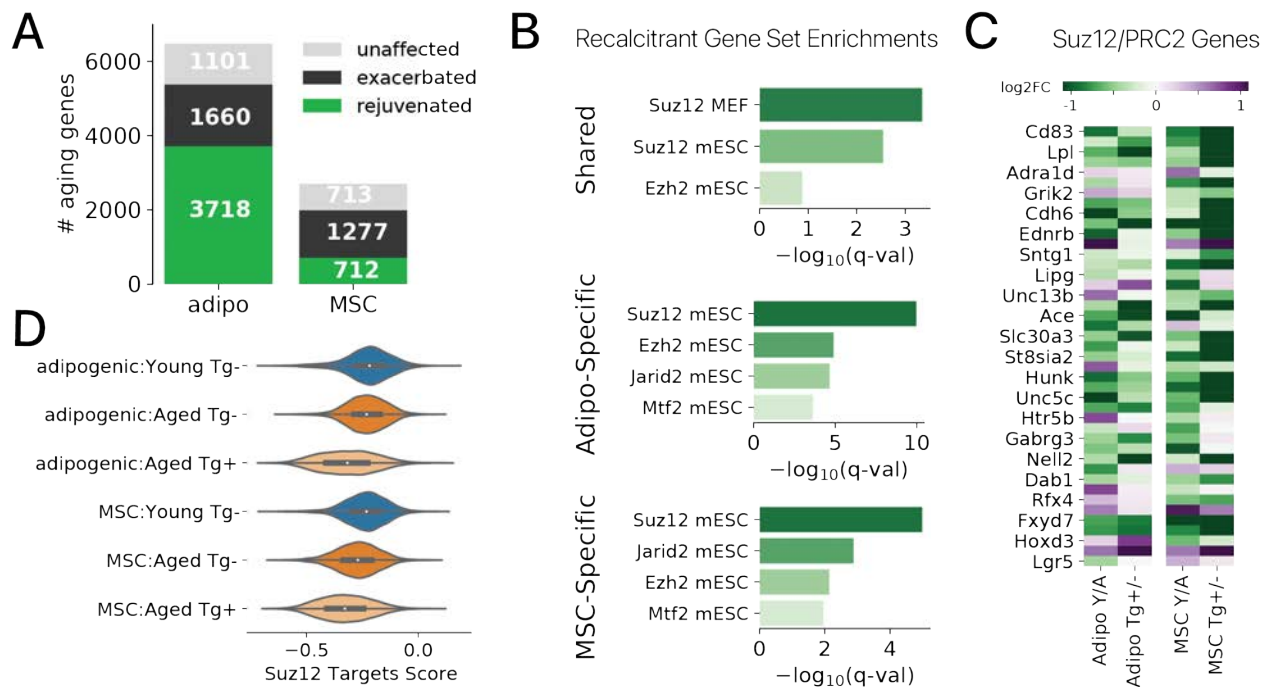


Figure S10: Aging genes recalcitrant to partial reprogramming are enriched for polycomb-repressor complex 2 targets. (A) Number of rejuvenated, unaffected, and exacerbated aging genes in each cell type. We consider unaffected and exacerbated genes recalcitrant to reprogramming. The majority of adipogenic aging genes are rejuvenated, while the majority of MSC aging genes are recalcitrant. (B) Gene set enrichment analysis for recalcitrant genes shared across cell types (462 genes, upper) and recalcitrant genes specific to each cell type (adipogenic, center; MSC, lower). Both shared and cell type-specific recalcitrant genes were strongly enriched for ChIP-seq targets of several polycomb-repressor complex (PRC) members (*Suz12*, *Ezh2*, *Jarid2*, *Mtf2*). (C) *Suz12* ChIP-seq target genes within the shared recalcitrant gene set are downregulated with age (left). Many of these PRC-targets are further downregulated by reprogramming and few are rejuvenated (right). (D) A gene score for annotated *Suz12* ChIP-seq target genes is suppressed with age, and further suppressed by reprogramming in both cell types.

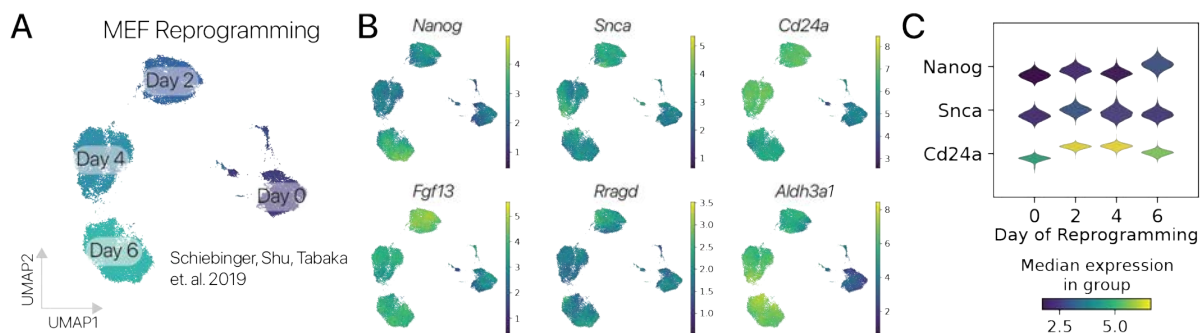


Figure S11: Pluripotency genes are activated in a subset of cells within 2-4 days of OSKM expression in mouse embryonic fibroblasts (MEFs). (A) UMAP projection of single cell mRNA profiles from MEFs undergoing reprogramming using a germline tetracycline-inducible OSKM allele [36]. Reprogramming induces novel cell states after only 2 days, consistent with strong effects we observe after a short OSKM pulse. (B) Expression of pluripotency markers (*Nanog*, *Rragd*) and early reprogramming response genes (others) across the MEF reprogramming timecourse. We observe induction of early reprogramming response genes consistent with our transient reprogramming data. We also observe activation of the *Nanog* pluripotency regulatory early in the timecourse (Day 2, 4), consistent with our transient reprogramming results. (C) Expression of pluripotency and early reprogramming genes across the MEF reprogramming timecourse. *Nanog* is significantly upregulated after only 2 days of reprogramming ($q < 0.1$, Monte Carlo).

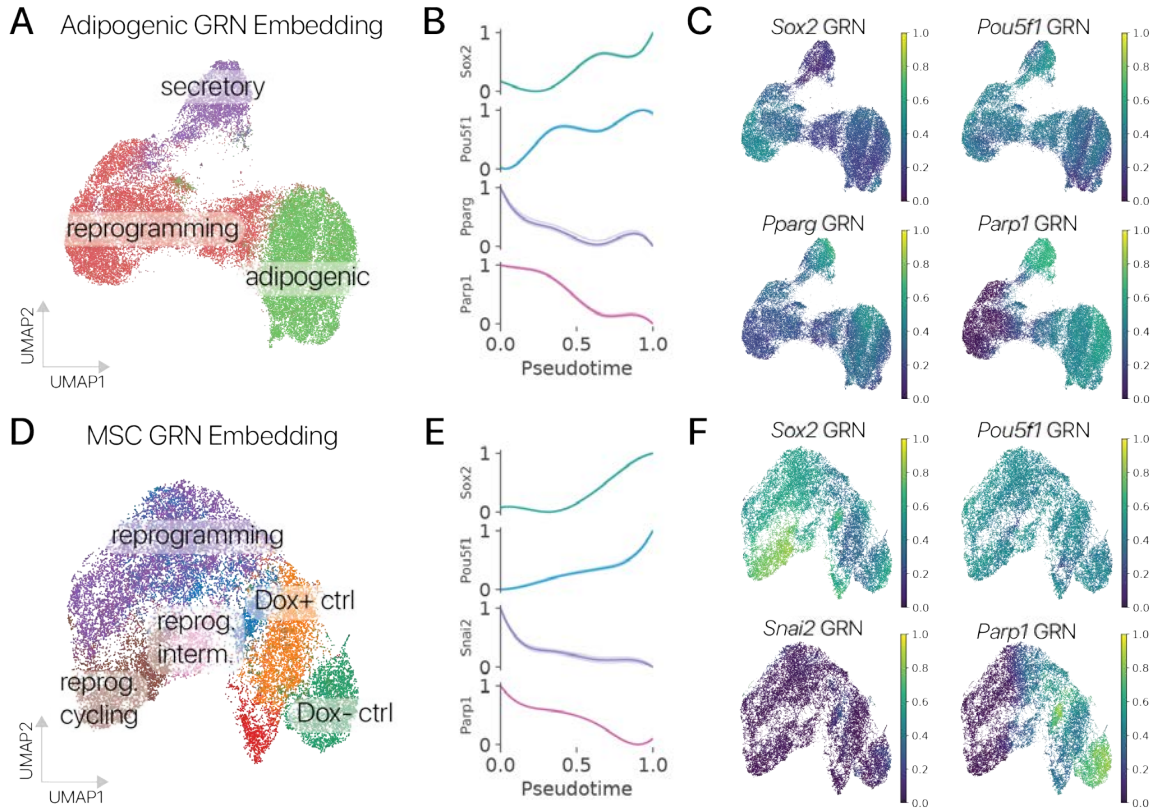


Figure S12: Gene regulatory network (GRN) analysis reveals the suppression of somatic cell identity GRNs during partial reprogramming. We extracted a directed GRN graph from the TRRUST database, linking murine transcription factors to direct target genes [38]. We then scored GRN activity in partially reprogrammed adipogenic cells and MSCs using a rank based integration approach, adopted from SCENIC [37]. **(A)** Adipogenic cells embedded in a transcriptional space derived from GRN activity scores. GRN scores accurately capture the effects of reprogramming and cell state differences within control populations. **(B)** GRN scores across reprogramming pseudotime (higher is more reprogrammed). We found that pluripotency GRNs (*Sox2*, *Pou5f1*) increase as expected with reprogramming, while adipogenic identity GRNs (*Pparg*, *Parp1*) are suppressed (Wald test, Binomial GLM, $p < 0.01$). **(C)** Pluripotency GRNs (top) and adipogenic identity GRNs (bottom) projected in the GRN score latent space clearly distinguish control and partially reprogrammed cells. **(D)** Muscle derived MSCs embedded using GRN activity scores. GRN scores again capture cell state changes induced by partial reprogramming. **(E)** Pluripotency GRNs increase as expected with reprogramming in MSCs, while mesenchymal identity GRNs (*Snai2*, *Parp1*) are suppressed (Wald test, Binomial GLM, $p < 0.01$). **(F)** Pluripotency (top) and mesenchymal identity (bottom) GRN scores in the GRN latent space show clear enrichment for reprogrammed and control populations respectively.

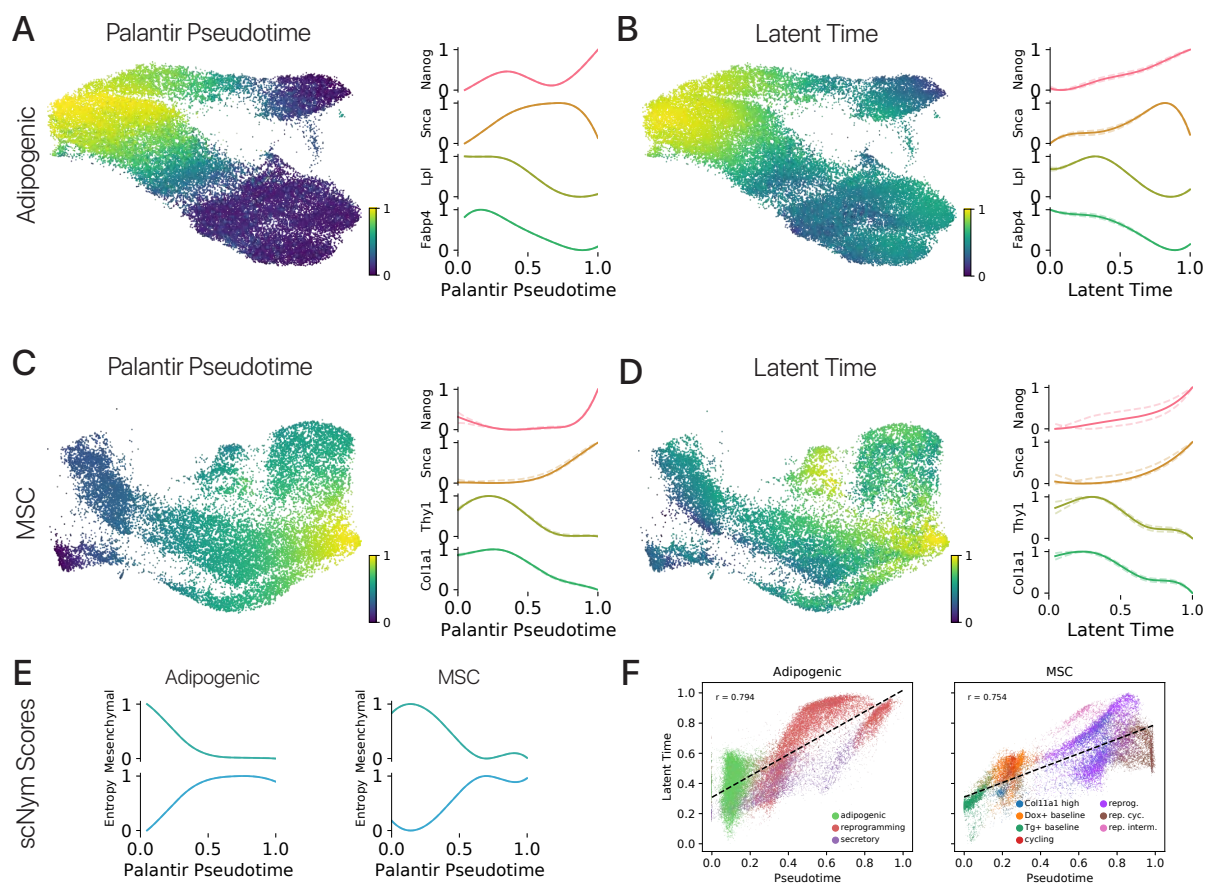


Figure S13: **Reprogramming trajectory analysis results are robust to the choice of pseudotime method.** In addition to velocity-informed diffusion pseudotime (Fig. 2), we performed pseudotime analysis with the profile similarity-based Palantir algorithm and the latent time coordinate recovered purely from RNA velocity dynamics. These two approaches are orthogonal to the diffusion pseudotime method used in Fig. 2. In each of these analyses, higher pseudotime coordinates correspond to more reprogrammed cell states. Pseudotime coordinates and generalized additive model (GAM) gene expression trends across the trajectory are displayed for the adipogenic cell data using (A) Palantir and (B) the scvelo latent time. The same analyses are performed using (C) Palantir and (D) the scvelo latent time in the MSC data. Pluripotency marker genes *Nanog* and *Snca* increased across reprogramming trajectories, while somatic identity genes (*Lpl*, *Fabp4*; *Thy1*, *Col1a1*) were suppressed. These results were consistent across cell types and pseudotime methods. (E) Cell identity scores generated using an scNym model also showed a suppression of the mesenchymal identity across Palantir pseudotime. (F) Velocity-informed diffusion pseudotime (Fig. 2) is highly correlated with latent time in each cell type ($r > 0.75$, $p < 0.001$, linear regression Wald test). Each of these results matches the earlier velocity-informed diffusion pseudotime analysis.

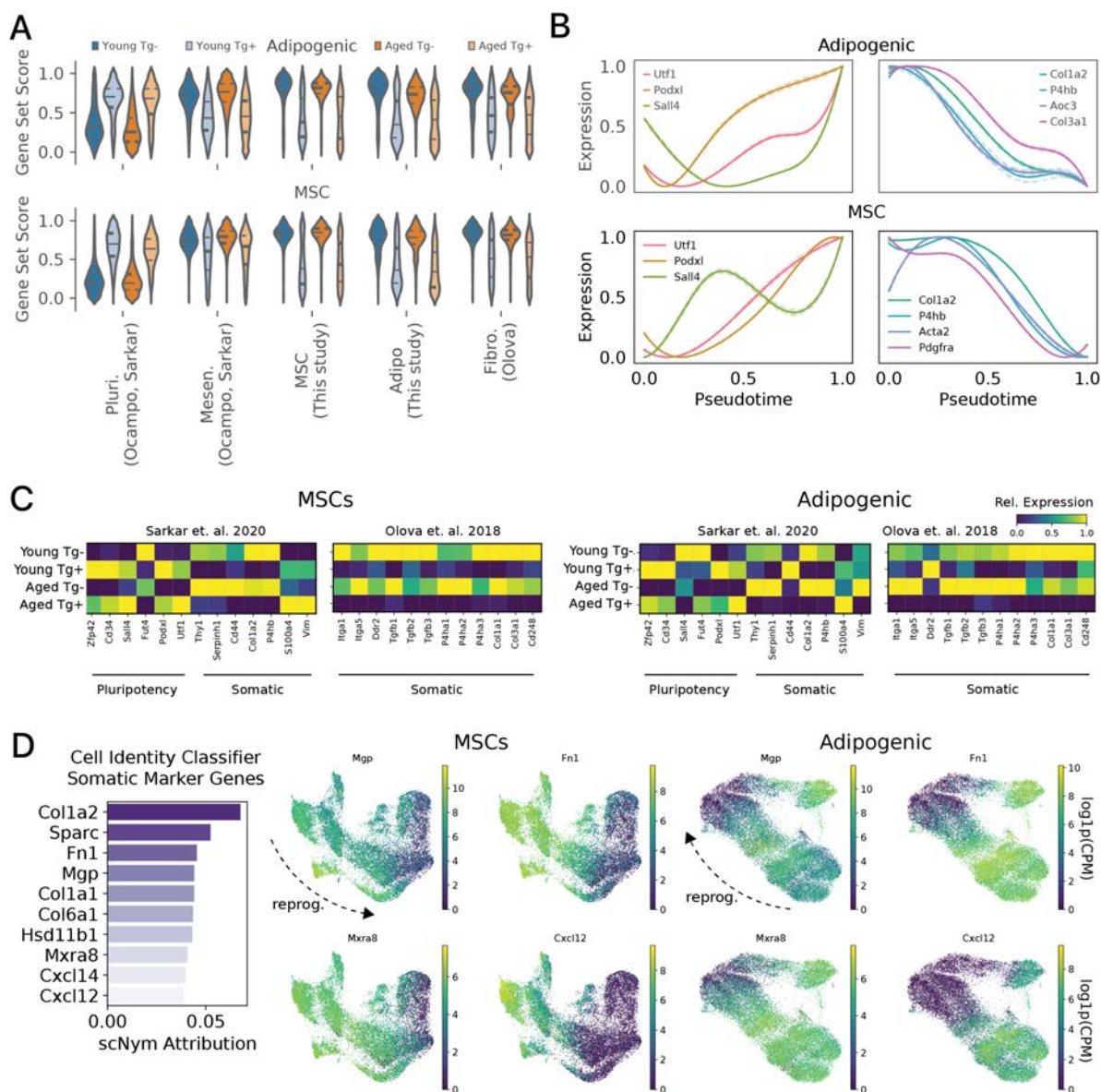


Figure S14: Analysis with previously reported gene sets shows an upregulation of pluripotency genes and suppression of somatic identity genes with reprogramming. We interrogated the expression of pluripotency and somatic identity marker genes used in previous transient reprogramming studies [10, 43]. Two additional studies interrogated only *Nanog* and *Thy1* [11, 9], as shown for our data in (Fig. 2). **(A)** We scored the activity of marker gene sets defined in multiple studies, including two gene sets we defined specifically for adipogenic and MSC somatic identity (Methods). We found that pluripotency marker genes were upregulated with reprogramming, while all somatic identity gene sets were downregulated ($p < 0.001$, Wilcoxon rank sum test). **(B)** The expression of key pluripotency genes increases across adipogenic and MSC reprogramming trajectories, while somatic identity genes decrease. Markers from previous reports [10, 43]. **(C)** Expression of marker genes defined in previous studies are shown before and after reprogramming in adipogenic cells and MSCs. The majority of pluripotency marker genes increase in expression with reprogramming, while somatic identity genes largely decrease. **(D)** Top 10 somatic identity markers extracted from the scNym cell identity classifier. Marker genes were extracted by computing expected gradient values for the “stromal cell” class, effectively asking which genes would lead to stronger predictions for the mesenchymal somatic identity. The scNym classifier recovers several marker genes in pre-existing sets (e.g. *Col1a1*, *Col6a1*), as well as marker genes outside any of the previously reported somatic marker gene sets (e.g. *Fn1*, *Mgp*, *Mxra8*, *Cxcl12*). These new markers specifically mark more somatic cell states in both adipogenic cells and MSCs, highlighting the ability of our cell identity classifier to reveal identity specific cell states that can be used to assess identity suppression. Considering scNym marker genes that were > 4 standard deviations above the mean attribution score, 94/100 markers were not included in previous gene sets.

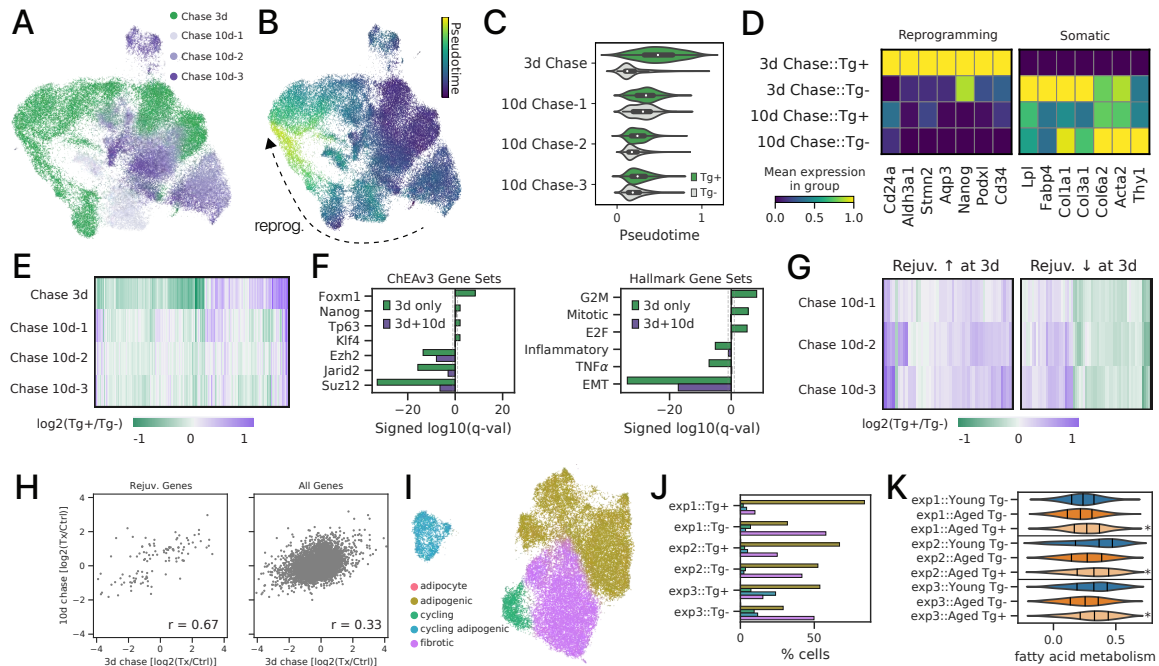


Figure S15: Cell identity is restored and youthful expression is retained after long-term reprogramming factor withdrawal in adipogenic cells. We performed three independent partial reprogramming experiments with a 3d pulse/10d chase scheme and profiled cells by single cell RNA-seq. **(A)** UMAP projection of our 3d pulse/3d chase experiment (Fig. 1) and three independent 3d pulse/10d chase experiments in an integrated latent space. **(B)** Pseudotime coordinates for 10d chase experiment cells inferred using a k-nearest neighbors regressor. **(C)** Distribution of cells across the pseudotime trajectory for 3d chase and 10d chase experiments. All three 10d chase experiments show a significant reduction in the pseudotime coordinate of reprogrammed (Tg+) cells ($p < 0.01$, Wilcoxon rank sum test, 10d Tg+ vs. 3d Tg+). **(D)** Expression of reprogramming and somatic marker genes in 3d chase and 10d chase experiments. Reprogramming marker genes show lower expression in 10d chase cells and somatic marker genes show higher expression, suggesting that cell identity suppression by partial reprogramming is transient. **(E)** Heatmap of highly variable gene expression in each experiment, displayed as reprogrammed/control (Tg+/Tg-) fold-changes. All 10d chase experiments show smaller reprogramming effects than the 3d experiment, suggesting transient activity. **(F)** Gene set enrichments comparing differentially expressed genes induced by reprogramming unique to the 3d chase experiment (3d only) or shared across both timepoints (3d+10d). Pluripotency regulons and cell cycle gene sets are upregulated at the 3d timepoint, but not the 10d timepoint. PRC2 complex members and the EMT program are downregulated at both timepoints, but more strongly at 3d, together suggesting that pluripotency genes are downregulated and identity genes restored after a long-term chase. **(G)** Heatmaps of genes that were rejuvenated in the 3d chase experiment by upregulation (left) or downregulation (right) in aged cells. Most rejuvenated genes continue to show a rejuvenation effect, but a subset of the downregulated genes show upregulation after a 10d chase. **(H)** Correlation of reprogramming-induced fold-changes at 3d and 10d timepoints for significant differentially expressed genes in both experiments (left) or all rejuvenated genes from the 3d chase experiment detected in both experiments (right). Fold-changes show significant correlation across timepoints, but magnitudes are smaller at 10 days, suggesting a restoration of the basal state over time. **(I)** Integrated UMAP projection of 10d chase experiments highlighting distinct cell states. **(J)** Cell state Distribution for each 10d chase experiment. Reprogramming shifts cells toward the adipogenic fate, rather than the fibrotic fate in each instance (χ^2 test $p < 0.001$). **(K)** Fatty acid metabolism gene set scores are significantly upregulated by reprogramming in aged cells (*: $p < 0.001$, Wilcoxon rank sum test).

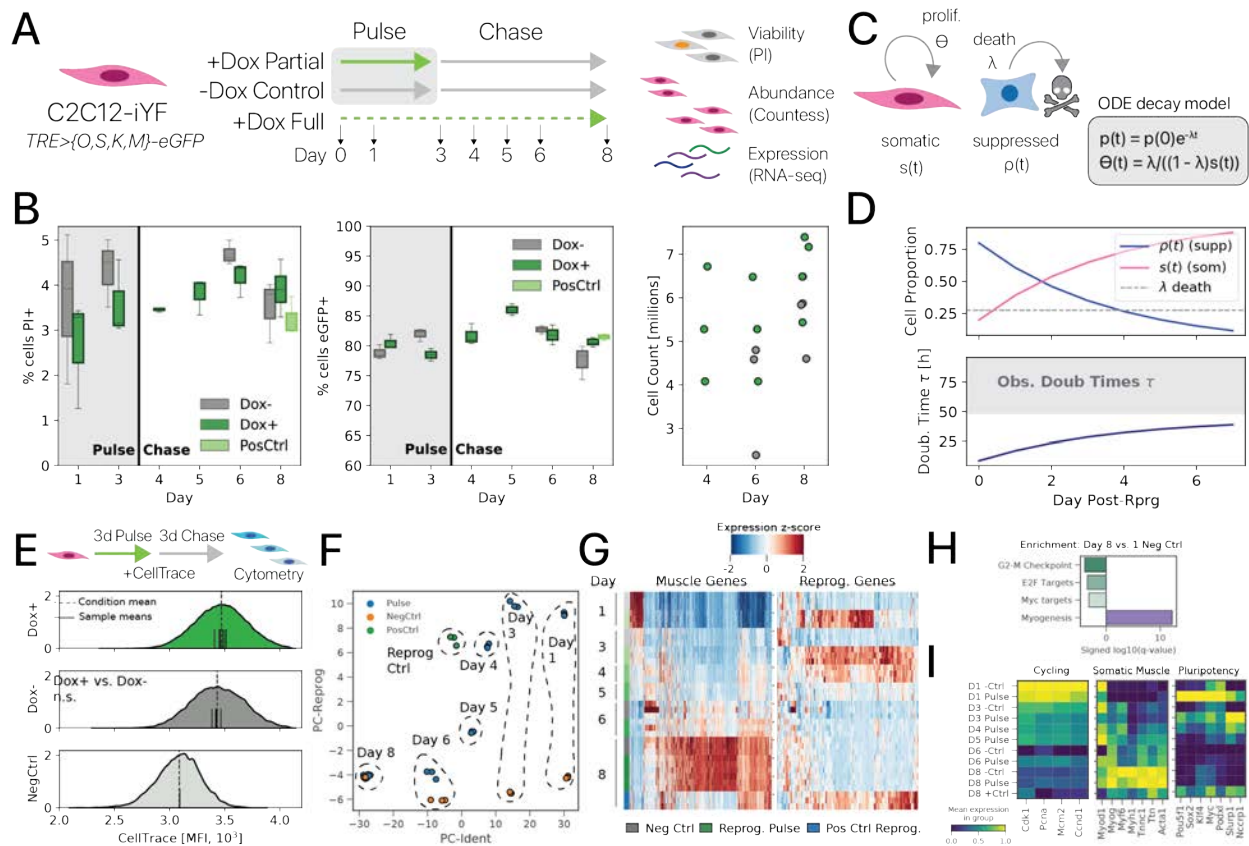


Figure S16: Cell line and mathematical models support a state-transition model of transient identity suppression. We observed that partially reprogrammed cells regain their somatic identities after long-term withdrawal of reprogramming factors. Here, we used a mathematical model to predict the population dynamics required for selective death and expansion of cell subpopulations to explain this observation (Supp. Note 1). We then used a muscle cell line model of partial reprogramming to test these predictions. All lines of evidence are inconsistent with a selective cell death and expansion model, supporting the hypothesis that cells regain somatic identity through state transitions. **(A)** Experimental schematic. C2iYF cells (C2C12 cells harboring Tet3G and Tet-inducible Yamanaka Factor alleles) were partially reprogrammed (3d pulse, 5d chase). We measured cell viability (PI), cell abundance, and gene expression (RNA-seq) across this timecourse. C2C12 cells adopt a somatic muscle identity in culture, allowing us to test the effects of partial reprogramming on somatic identity. **(B)** Flow cytometry measurements for cell death (PI+ indicates cell death), Yamanaka Factor reporters (eGFP), and cell abundance (image cytometry) across the timecourse. **(C)** Schematic representation of an ordinary differential equation model for selective cell death of cells with suppressed identities (blue) and expansion of cells retaining somatic identities (pink). **(D)** We fit the ODE model to observed state proportions from our adipogenic cell experiments. We found that the cell death rates required by the model are far higher than the rates measured in **(B)**. We also found that the doubling rates (lower) required by the model are faster than previously observed for either adipogenic cells or the C2C12 cell line used here (shaded upper box). **(E)** We measured cell proliferation during the chase period of a 3d pulse/3d chase partial reprogramming experiment \pm Dox. CellTrace was not significantly higher in Dox+ cells, suggesting that proliferation rates during the chase period are not influenced by partial reprogramming. Dashed lines indicate condition-wide mean fluorescence, solid half-height lines indicate the means of biological replicates. **(F)** PCA analysis of RNA-seq data from C2iYF reprogramming using PC1 from muscle differentiation genes (control day 8 vs. day 1) and reprogramming genes (day 3 reprogrammed vs. control). The effects of reprogramming are strongest during the SOKM pulse (days 1-3), then fade during the chase (days 4-8). **(G)** Significant differentially expressed genes with somatic differentiation (left, 1436 genes) and partial reprogramming (right, 207 genes). Somatic muscle genes increase in control and pulsed cells across the timecourse, as expected for this cell line. Reprogramming genes are induced on days 1 and 3, then gradually decrease in effect size until day 8. **(H)** Hallmark gene set enrichment analysis confirms that the somatic differentiation genes (day 8 vs. 1 controls) are strongly enriched for myogenic genes. Dashed lines mark $q = 0.1$. **(I)** Relative expression of cell cycle (left), myogenic (center), and pluripotency (right) genes. Cell cycle genes decrease in all conditions over the timecourse, myogenic genes increase in control and pulsed cells, and pluripotency genes peak during the pulse induction then decline. Cell cycle gene expression mirrors our proliferation analysis in **(E)**, suggesting differential proliferation between partially reprogrammed and control cells is unlikely. All of our RNA-seq results suggest that partial reprogramming induces transient somatic identity suppression in C2iYF cells.

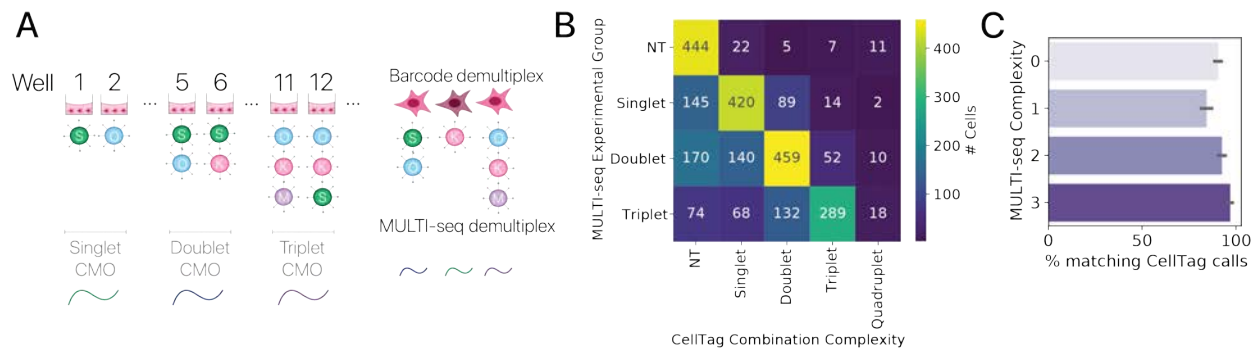


Figure S17: MULTI-seq labeling of combinatorial complexity and expressed barcode based demultiplexing show strong correspondence (A) Experimental schematic. We performed a pilot pooled screen in an arrayed format and labeled the number of unique factors in each treatment condition using MULTI-seq. We then compared the MULTI-seq derived labels to our expressed barcode (CellTag) system for demultiplexing combinatorial perturbations. MULTI-seq label assignments are imperfect and expressed barcodes may suffer detection failures, so we do not expect perfect 1:1 correspondence. **(B)** Heatmap of correspondence between complexity labels derived from MULTI-seq and expressed barcodes for cells with a confident MULTI-seq derived label. We found that the expected complexity based on MULTI-seq barcodes is the dominant mode for expressed barcode classifications. We expect some elements in the lower triangle of the matrix due to (1) failed barcode detections or (2) cells that received less than the target number of unique factors in a given condition due to stochasticity in the transduction. We therefore consider all labels that are in the lower triangular matrix to be accurate. Very few cells fall outside the lower triangle, likely due to improper MULTI-seq barcode calls. **(C)** Correspondence of labels quantified across MULTI-seq treatment condition labels (mean + 95% CI). All complexities show high correspondence (>80% matching calls).

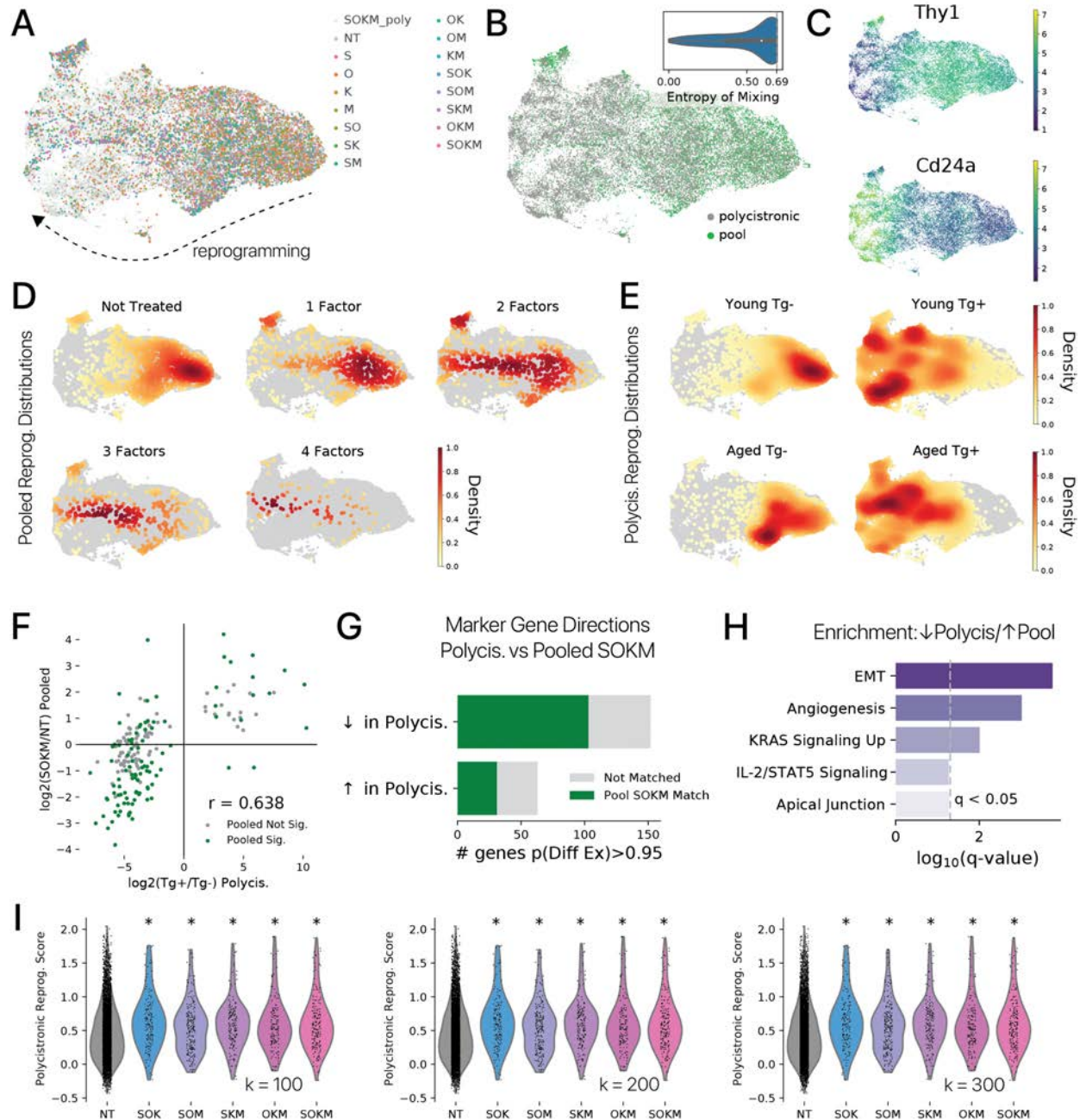


Figure S18: Integration of MSC mRNA profiles confirms that polycistronic and pooled reprogramming factors exert similar effects. (A) UMAP projection of integrated mRNA profiles from MSCs in our polycistronic reprogramming experiment (NT, SOKM_poly) and our pooled screening experiment (NT, others). (B) Polycistronic and pooled screening profiles are well-mixed in the integrated embedding. The entropy of mixing is high across the nearest neighbor graph, approaching the maximum possible value (0.69) in many neighborhoods (inset figure, dotted line is max value). (C) Somatic identity (*Thy1*) and reprogramming (*Cd24a*) marker genes show a clear trajectory of reprogramming in the integrated embedding. (D, E) Reprogramming combinations in the pooled screen show clear enrichment in the embedding for regions co-occupied by polycistronic reprogrammed cells. (F) Comparison of log fold-changes in polycistronic and pooled experiments for highly confident differentially expressed genes in the polycistronic data ($p(\text{DE}) > 0.95$). Fold-changes show strong, significant correlation ($r = 0.638, p < 0.001$). (G) Visualization of the number of reprogramming markers that show a common direction of change in polycistronic SOKM and pooled SOKM cells relative to the respective controls. (H) Gene set enrichment of markers downregulated in polycistronic cells but not pooled SOKM cells reveals an enrichment for EMT and KRAS gene sets. (I) Activity of a polycistronic reprogramming score derived from the top $k = \{100, 200, 300\}$ reprogramming marker genes. All pooled reprogramming combinations with ≥ 3 factors show significant upregulation in the pooled screening experiment and this result is consistent across values of k (*: $p < 0.01$, Wilcoxon Rank Sum test).

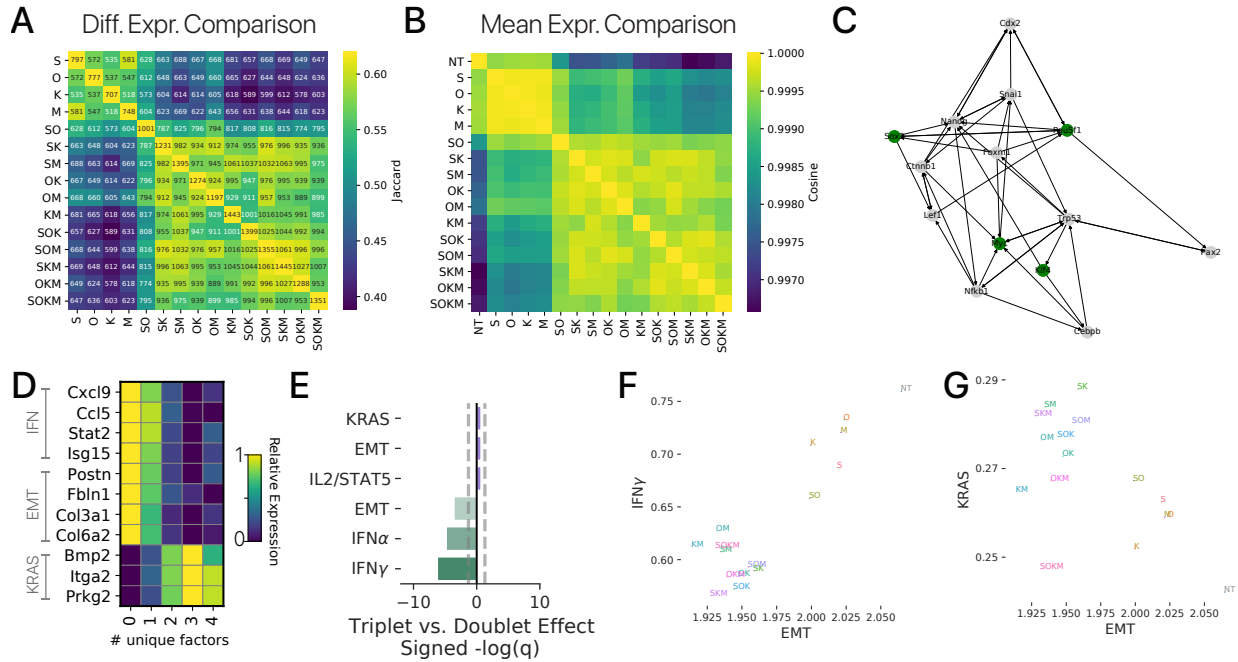


Figure S19: Higher-order combinations of Yamanaka Factors exert similar transcriptional effects. (A) Comparison of differentially expressed gene (DEG) sets relative to the control for each perturbation. The Jaccard index measures the number of shared genes across two sets as a fraction of the total number of unique genes in both sets. Annotation insets are the number of shared DEGs for each comparison. Similar to results extracted from our perturbation classifier, we find that combinations of Yamanaka Factors with similar complexity exert similar transcriptional effects after a transient pulse. (B) Comparison of the mean gene expression vectors for each Yamanaka Factor combination (NT: Not Treated controls) using the cosine similarity (positive values are more similar). We again found that higher order combinations exerted similar transcriptional effects. (C) Force directed layout visualization of connections within the pluripotency gene regulatory network extracted from the TRRUST database. Yamanaka Factors are highlighted in green. All three factor subsets of the Yamanaka Factors have the potential to activate the fourth factor. (D) Comparison of interferon (IFN), epithelial-to-mesenchymal transition (EMT), and KRAS gene signatures across cells with different numbers of unique reprogramming factors. The largest differences between three and two factor combinations are stronger suppression of IFN and EMT programs and stronger activation of the KRAS signaling program. (E) Gene Ontology enrichment analysis with the MSigDB Hallmark gene sets confirms the gene level results in (D). (F) EMT and IFN γ program activity are tightly correlated ($r > 0.9$), suggesting that they are suppressed by a common mechanism. (G) EMT and KRAS program activities are anticorrelated ($r < -0.6$), suggesting these effects likewise might share a mechanism during the reprogramming process.

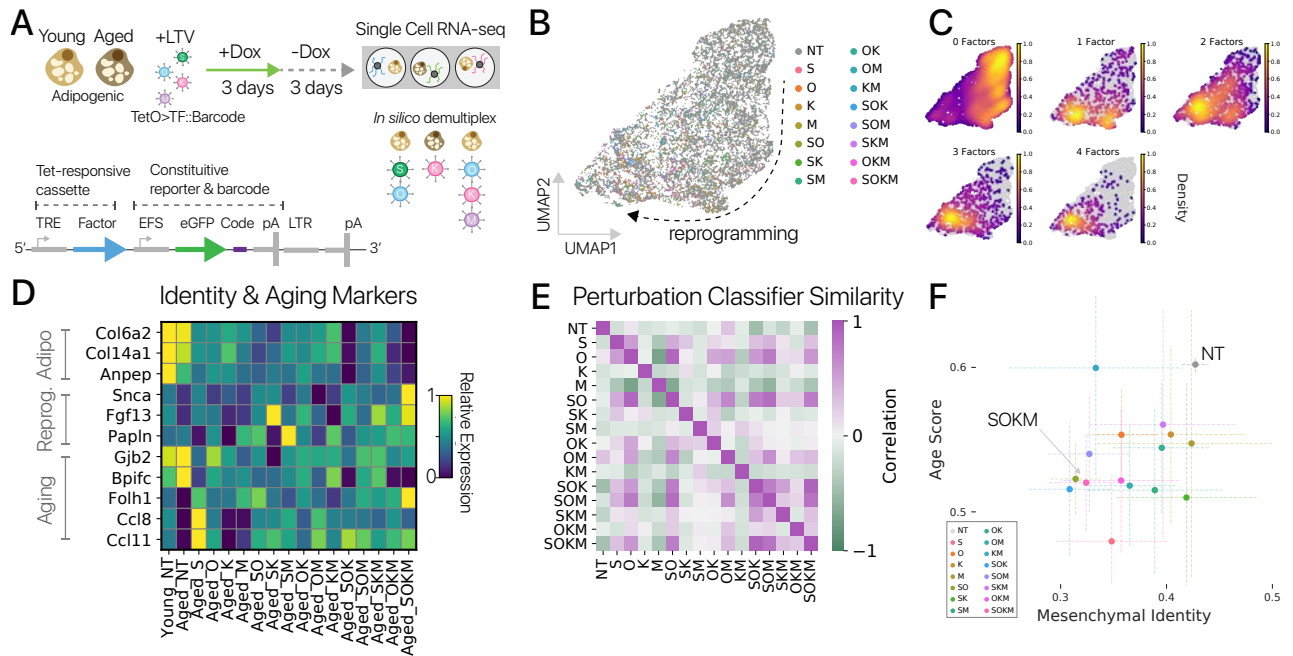


Figure S20: **Pooled screening in adipogenic cells provides similar results to screening in MSCs.** (A) Diagram of Yamanaka Factor pooled screening experiments. Young and aged adipogenic cells were transduced with lentiviruses each harboring a one inducible Yamanaka Factor with expressed barcodes (lower). Reprogramming was induced for a 3 day pulse/3 day chase ($n = 5$ animals per age across two independent experiments). Cells were profiled by single cell RNA-seq and unique combinations of Yamanaka Factors were demultiplexed *in silico* based on expressed barcodes. (B) Cells from the pooled screen embedded using scNym, projected with UMAP, and labeled with the detected reprogramming factors (9,000+ cells). (C) Density of cells perturbed with different numbers of reprogramming factors in the UMAP embedding. Higher order combinations show a larger transcriptional shift relative to control cells. (D) Adipogenic marker genes (top) decrease and reprogramming marker genes (center) increase as the combinatorial complexity (number of unique factors) increases. Aging genes (lower) likewise appear closer to the youthful level with more complex perturbations. (E) Similarity matrix between different Yamanaka Factor combinations extracted from a cell perturbation classification model. Classifier similarity is less dramatic than in MSCs, but the strongest similarities are again observed for higher order combinations of the Yamanaka Factors. (F) Mesenchymal cell identity scores derived from scNym models and aging gene set scores in aged cells reprogrammed with different factor combinations (mean scores \pm 95% CI). Many Yamanaka Factor combinations significantly decreased both the mesenchymal cell identity score and age score relative to aged control cells (NT) (Wald tests, $p < 0.05$). However, age scores and identity scores were not well-correlated, suggesting that cell identity suppression and rejuvenation are not tightly coupled ($\rho = 0.19$, $p > 0.48$, Spearman).

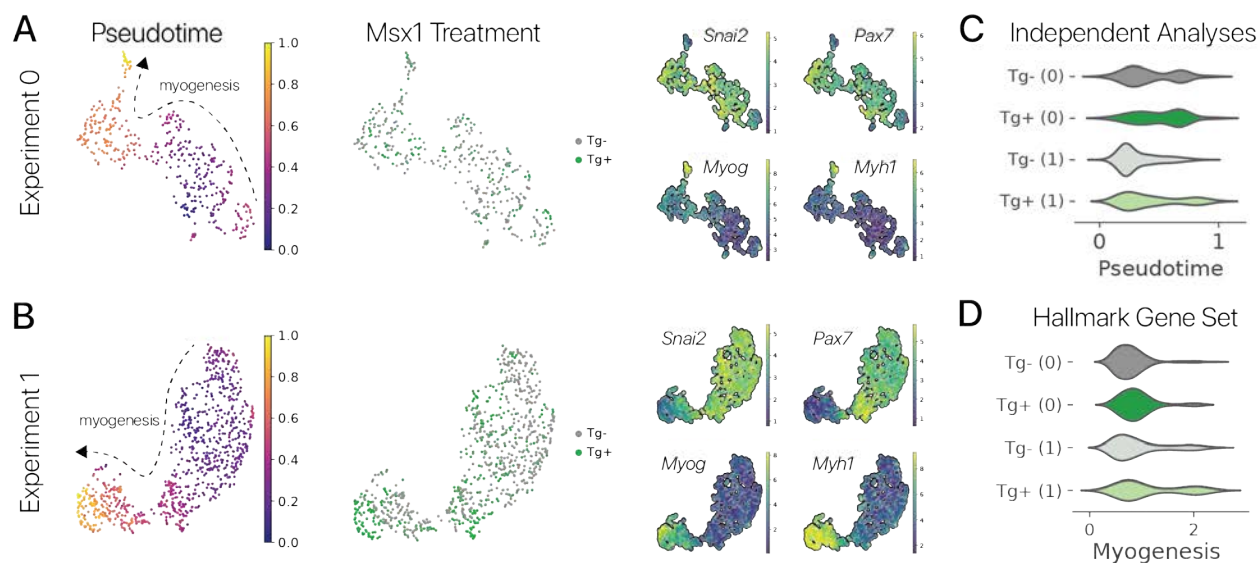


Figure S21: Partial reprogramming with the multipotency factor *Msx1* increases myogenesis in aged myogenic cells. Aged myogenic cells were partially reprogrammed with *Msx1* in two separate experiments. Experiments were analyzed independently below. **(A)** Pseudotime analysis (left) and marker gene analysis (right) reveal a myogenic differentiation trajectory within both the first ($n = 2$ animals, 28 months old) and **(B)** second independent experiment ($n = 3$ animals, 30 months old). **(C)** We performed pseudotime analysis independently in each experimental condition. As in our integrated analysis (Fig. 4), partially reprogrammed cells were significantly more differentiated than controls in both experiments (Wilcoxon Rank Sum test, $p < 0.01$). **(D)** We scored the activity of the Hallmark Myogenesis gene set (MSigDB). Both experiments showed significant increases in myogenic gene set activity after partial reprogramming treatment (Wilcoxon Rank Sum test, $p < 0.01$).

Supplemental Tables

Identity	Gene
Cardiomyocyte	Mef2c
Cardiomyocyte	Tbx5
Cardiomyocyte	Tnni3
Cardiomyocyte	Pln
Cardiomyocyte	Myl2
Cardiomyocyte	Kcnj2
Cardiomyocyte	Atp2a2
Cardiomyocyte	Ppara
Cardiomyocyte	Myh6
Skeletal Muscle	Myod1
Skeletal Muscle	Myh1
Skeletal Muscle	Myl2
Skeletal Muscle	Acta1
Skeletal Muscle	Tnnc1
Skeletal Muscle	Tnnc2
Skeletal Muscle	Myom1
Skeletal Muscle	Myom2
Skeletal Muscle	Des
Skeletal Muscle	Mstn
Skeletal Muscle	Gdf11
Skeletal Muscle	Trim63
B cell	Pax5
B cell	Cd19
B cell	Cd22
B cell	Ighm
B cell	Ighg2b
B cell	Ighg2c
B cell	Rag2
Retina	Fabp3
Retina	Rbpms
Retina	Pou4f1
Retina	Pou4f2
Retina	Isl1
Retina	Sncg
Retina	Elavl2
Retina	Elavl4
Retina	Pkia
Retina	Ebf3
Pluripotency	Nanog
Pluripotency	Fut4
Pluripotency	Cd34
Pluripotency	Sall4
Pluripotency	Utf1
Pluripotency	Zfp42

Table S1: Identity genes used for identity scoring in public partial reprogramming data.