Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/6/37/eaba1190/DC1

Supplementary Materials for

Diversification of reprogramming trajectories revealed by parallel single-cell transcriptome and chromatin accessibility sequencing

Q. R. Xing, C. A. El Farran, P. Gautam, Y. S. Chuah, T. Warrier, C. X. D. Toh, N. Y. Kang, S. Sugii, Y. T. Chang, J. Xu, J. J. Collins, G. Q. Daley, H. Li*, L. F. Zhang*, Y. H. Loh*

*Corresponding author. Email: yhloh@imcb.a-star.edu.sg (Y.H.L.); zhanglf@ntu.edu.sg (L.F.Z.); li.hu@mayo.edu (H.L.)

Published 11 September 2020, *Sci. Adv.* **6**, eaba1190 (2020)
DOI: 10.1126/sciadv.aba1190

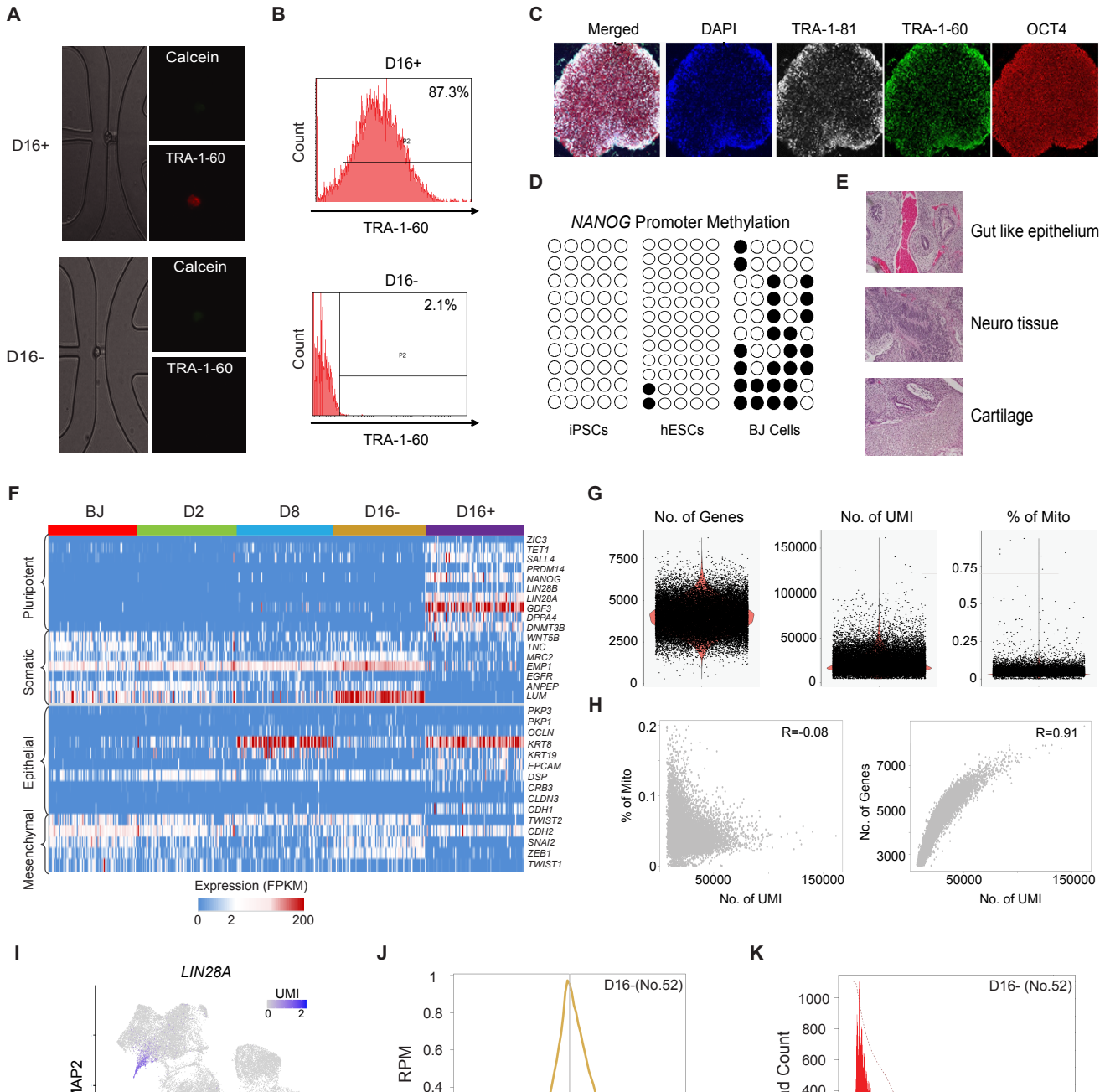
The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/6/37/eaba1190/DC1)

Supplementary Table:

Table S1 includes list of genes that are differentially expressed among RCA subgroups, D8 BDD2-C8 sorted cells, and D8 CD13/CD201 dual antibodies sorted populations; list of transcription factors identified from 10X scRNA-seq library analysis; list of variable transcription factors identified from scATAC-seq libraries across reprogramming; list of FOSL1 and TEAD4 functional targets.



Supplementary Figure 1

Figure S1. QC of scATAC-Seq and scRNA-Seq libraries

(A) Representative bright field images and dye staining images including Calcein (green) and TRA-1-60 (red), in a D16+ (top) and a D16- (bottom) cell.

(B) FACS analysis for MACS sorted D16 TRA-1-60+ (top) and D16 TRA-1-60- (bottom) populations.

(C) Representative staining images of Nucleus (blue), TRA-1-81 (grey), TRA-1-60 (green) and OCT4 (red) in iPSC clones generated using polycistronic O2S, K2M reprogramming system.

(D) Bisulfite DNA methylation analysis on the NANOG promoter in iPSCs (left), hESCs (middle) and BJ cells (right). Black circles denote methylated cytosine residue, whereas the white circles denote unmethylated cytosine residues.

(E) Hematoxylin and eosin staining of teratomas derived from immunodeficient mice injected with iPSCs generated using the polycistronic O2S, K2M reprogramming system.

(F) Heatmaps demonstrating the expression levels of pluripotent and somatic genes (top), and Mesenchymal and Epithelial genes (bottom) throughout reprogramming. Each column represents a single-cell, and each row indicates a gene. The expression ranges from blue (low) to red (high).

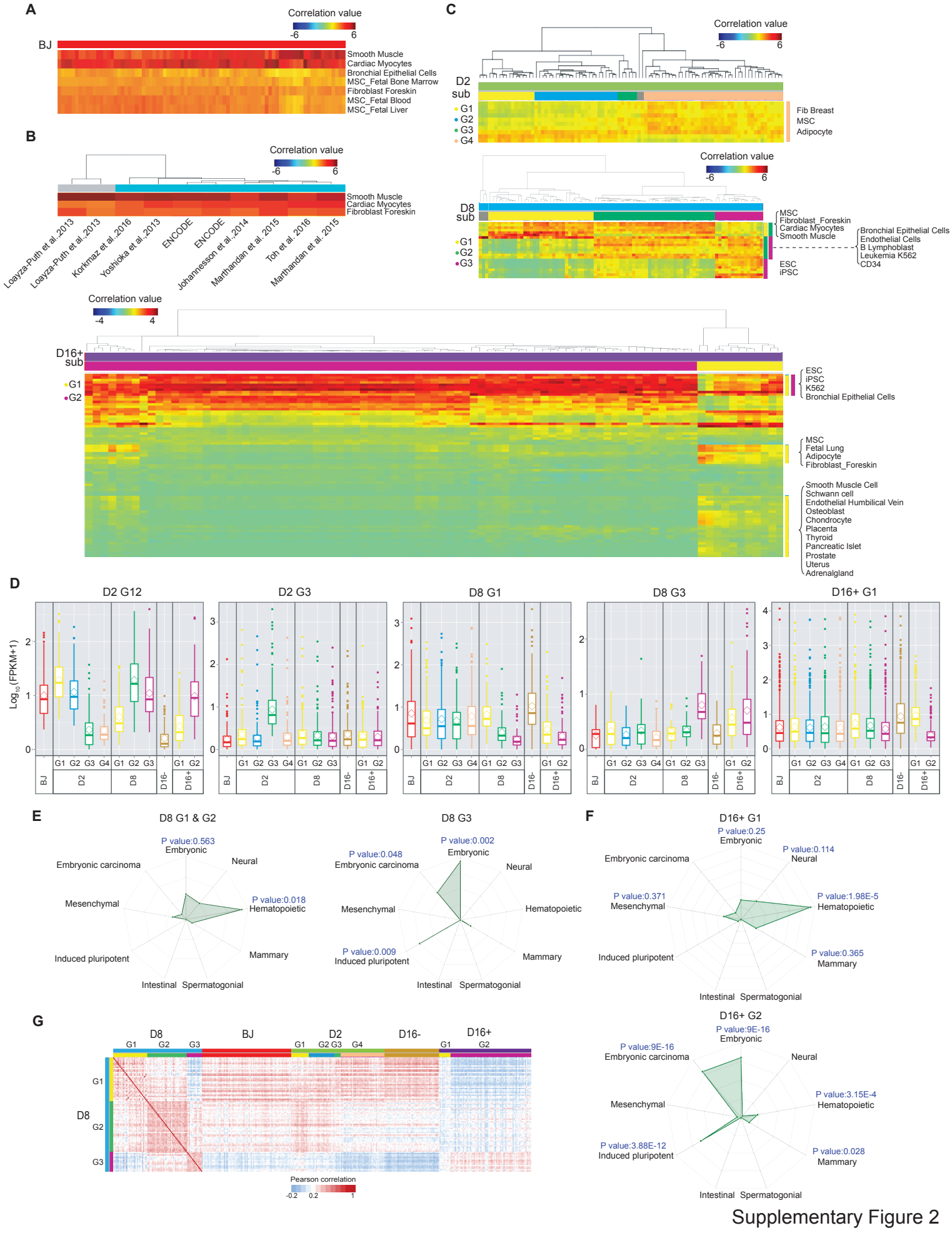
(G) QC of the prepared 10X scRNA-Seq libraries demonstrating the number of genes (left), unique molecular identifiers (UMI) (middle), and percentage of mitochondrial DNA detected (right).

(H) Correlation analysis between mitochondrial DNA and UMI (left) and between number of genes and UMI (right). The analysis showed a negative correlation between UMI and mitochondrial DNA.

(I) Superimposition of the single-cell-expression levels of *LIN28A* on the UMAP plot.

(J) Average enrichment profile of a D16- scATAC-Seq library around Transcription Start Sites (TSS) of the genome with a window of -3K to 3K. Y-axis denotes the average normalized read counts of the library over the indicated region in the genome (x-axis).

(K) Histogram of insert size metrics of a D16- scATAC-Seq library, revealing a nucleosomal pattern, characteristic of a good scATAC-Seq library. The histogram was generated using the "CollectInsertSizeMetrics" of Picard.



Supplementary Figure 2

Figure S2. Identification of subgroups from Fluidigm scRNA-Seq libraries using RCA analysis

(A) RCA heatmap showing the highest correlated lineages for BJ scRNA-Seq libraries.

(B) RCA heatmap showing the highest correlated lineages for the published BJ bulk RNA-Seq libraries. The sources of BJ RNA-Seq libraries are indicated below.

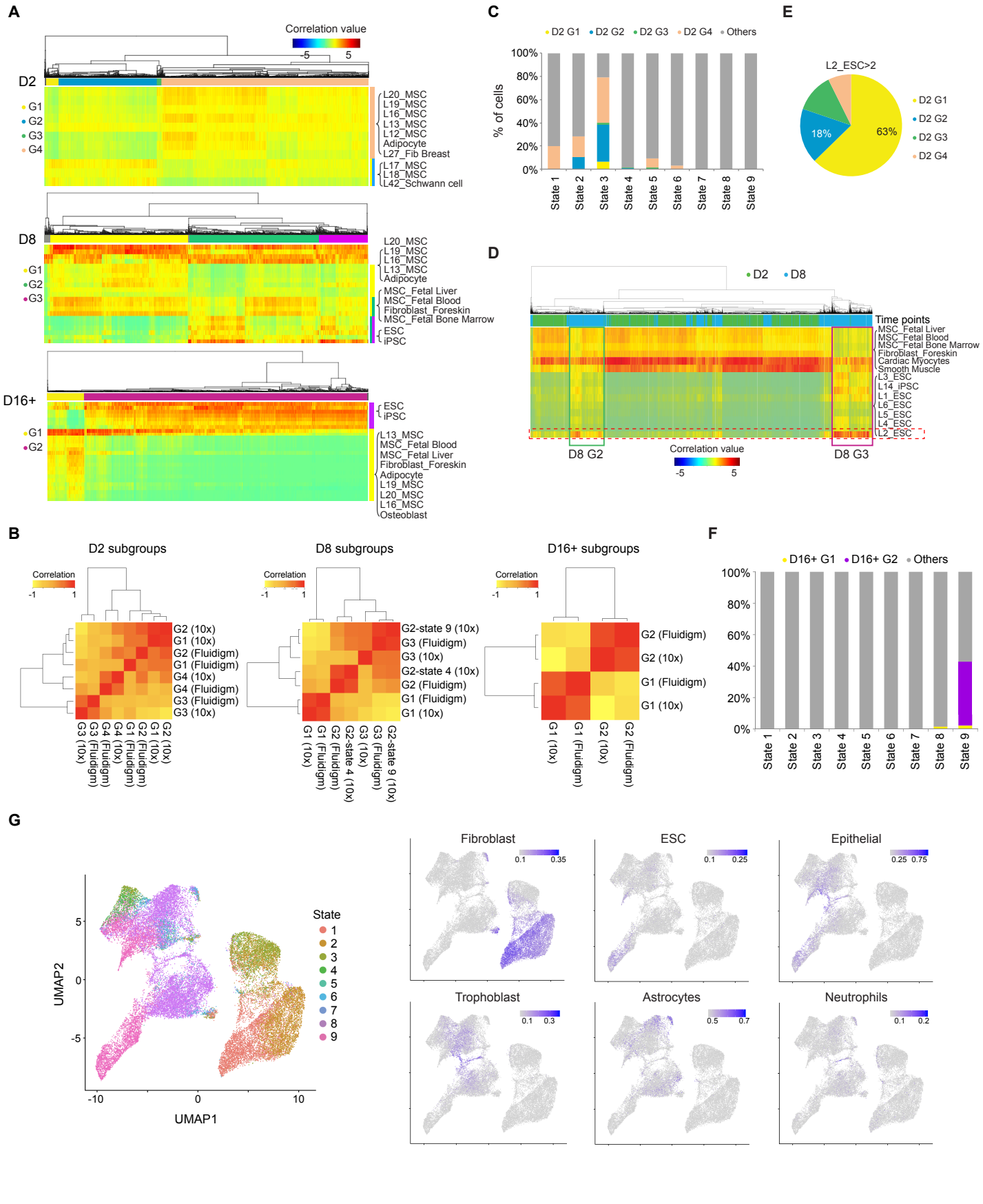
(C) RCA heatmaps showing the subgroups of D2 (top), D8 (middle), and D16+ (bottom) cells clustered based on their correlation to cells of different lineage origin. Each row indicates a lineage. The correlation ranges from blue (low) to red (high). Color bar on top indicates the time-points (above) and the respective subgroups (below). Grey color indicates the minority outlier cells which doesn't belong to the indicated subgroups.

(D) Boxplots demonstrating the expression of the specified differentially expressed genes (DEGs) across the reprogramming time-points and their respective subgroups. Lines in the box represent median expression, whereas diamonds in the box represent mean expression.

(E) Stemness analysis for the genes highly expressed in D8 G1&G2 (left) and D8 G3 (right) cells. Significant adjusted P values are indicated.

(F) Stemness analysis for the genes highly expressed in D16+ G1 (top) and D16+ G2 (bottom) cells. Significant adjusted P values are presented.

(G) Heatmap showing the Pearson correlation calculated between D8 subgroups and the cells of other reprogramming time-points, based on the DEGs of D8 subgroups. The correlation ranges from blue (low) to red (high). Color bar on top indicates the time-points (above) and the respective subgroups (below). Side color bar indicates the subgroups of D8.



Supplementary Figure 3

Figure S3. Analysis of 10X Genomics scRNA-Seq libraries

(A) RCA heatmaps showing the subgroups of D2 (top), D8 (middle), and D16+ (bottom) cells, clustered based on their correlation to various lineages in RCA panel. Each row indicates one lineage. The correlation ranges from blue (low) to red (high). Color bar on top indicates the subgroups. Grey color indicates the minority outlier cells which doesn't belong to the indicated subgroups.

(B) Unsupervised clustering of D2 subgroups (left), D8 subgroups (middle), and D16+ subgroups (right) identified from Fluidigm and 10X Genomics scRNA-Seq libraries, based on the Correlation Coefficient calculated between them. Color represents the Correlation Coefficient value, ranging from yellow (low) to red (high).

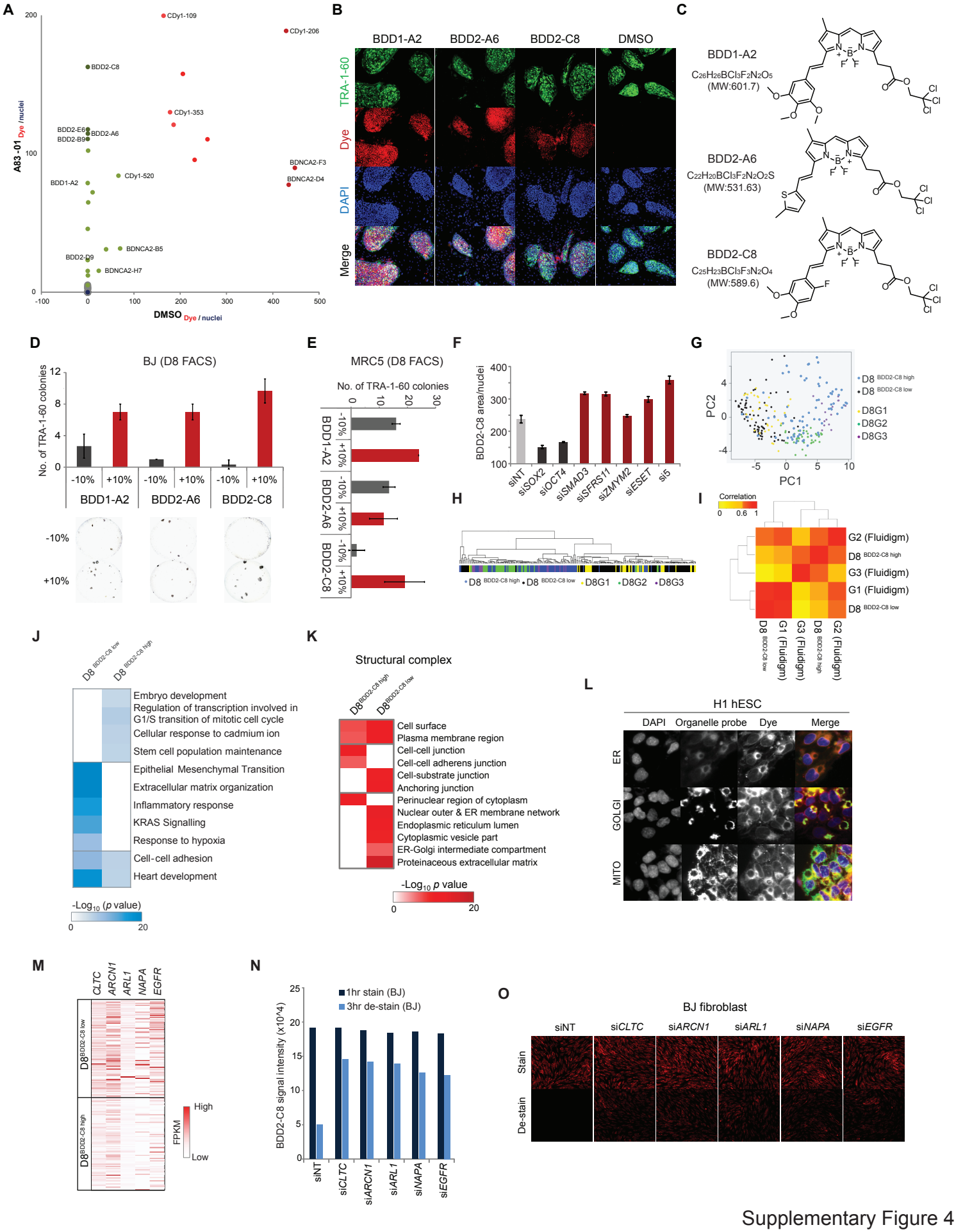
(C) Stacked column revealing the distribution of D2 subgroups across the pseudotemporal states. Colors represent D2 subgroups. Grey color represents cells of other reprogramming time-points.

(D) RCA heatmap showing the combined RCA clustering of D2 and D8 cells based on their correlation to cells of various lineage origins. Each row indicates one lineage. The correlation ranges from blue (low) to red (high). Color bar on top indicates the reprogramming time-points.

(E) Piechart indicating the percentage of D2 subgroups calculated against D2 cells with correlation values of above 2 to the RCA library "L2_ESC".

(F) Stacked column revealing the percentage of the D16+ subgroups across the trajectory states. Grey color indicates cells of the other time-points.

(G) Superimposition of lineage identities on UMAP plot. Color represents pseudotemporal state (left) and lineage enrichment score (right).



Supplementary Figure 4

Figure S4. Identification of fluorescent probes to enrich for the early reprogrammed cells

(A) Distribution plot for the fluorescent probe screen. Positively stained area of the dyes were measured and normalized to the cell number (Dye/nuclei) for D8 cells cultured in A83-01 (y-axis) and DMSO control (x-axis). Red dots represent dyes with unspecific staining, and green dots represent dyes that are capable of distinguishing early reprogrammed cells.

(B) Representative staining images of chemical dyes and TRA-1-60 in D21 reprogramming cells. Cells were stained with the indicated chemical dyes and subsequently fixed for TRA-1-60 immunofluorescence staining.

(C) Chemical structure and molecular formula of the candidate fluorescent probes.

(D) Quantification of TRA-1-60+ colonies yielded from D8 cells sorted with the respective chemical dyes (top). Top 10% and bottom 10% of the stained cells were collected and seeded for TRA-1-60 Immuno-histochemical staining at D21. Representative images are shown below. n=3; error bar indicates SD.

(E) Quantification of TRA-1-60+ colonies yielded from D8 cells sorted with the indicated chemical dyes at day 8 of reprogramming, induced from MRC5. n=2; error bar indicates SD.

(F) Barchart showing the normalized BDD2-C8 staining area over cell number in D12 cells, upon depletion of the reprogramming regulators at day 5 of reprogramming. si5 refers to the combined knockdown of *SMAD3*, *SFRS11*, *ZMYM2*, *ESET* and *SAE1*. n=4; error bar indicates SD.

(G) PCA plot based on the RCA analysis for D8, $D8_{BDD2\text{-}C8\ high}$ and $D8_{BDD2\text{-}C8\ low}$ scRNA-Seq libraries. Color represents cell identity.

(H) Hierarchical dendrogram for the RCA clustering of D8 subgroups, $D8_{BDD2\text{-}C8\ high}$ and $D8_{BDD2\text{-}C8\ low}$ cells. Color represents the cell identity.

(I) Unsupervised clustering of D8 subgroups (Fluidigm) and BDD2-C8 sorted cells, based on the Correlation Coefficient calculated between them. Color represents the Correlation Coefficient value, ranging from yellow (low) to red (high).

(J) Heatmap revealing the GO terms enriched by the genes differentially expressed in $D8_{BDD2\text{-}C8\ high}$ and $D8_{BDD2\text{-}C8\ low}$ cells.

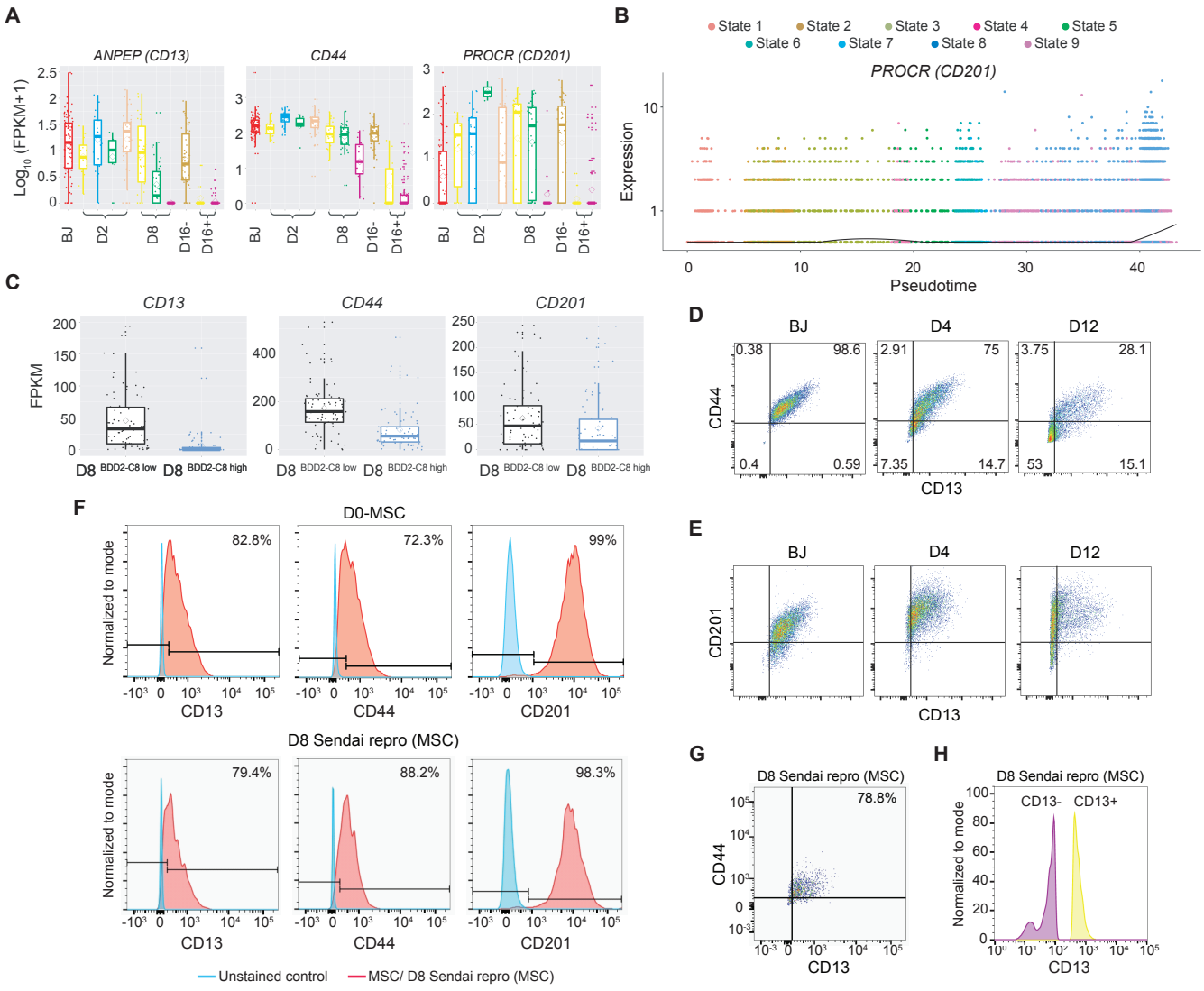
(K) Heatmap revealing the structural complexes enriched by the genes differentially expressed in $D8_{BDD2\text{-}C8\ high}$ and $D8_{BDD2\text{-}C8\ low}$ cells. Enrichment ranges from white (no) to dark red (high).

(L) Confocal images for staining of organelle-specific probes and BDD2-C8 in H1 hESCs. Green indicates organelle probes signal, red indicates dye signal, and blue indicates nuclei staining signal.

(M) Heatmap demonstrating the expression of genes involved in secretory pathway in $D8_{BDD2\text{-}C8\ high}$ and $D8_{BDD2\text{-}C8\ low}$ cells. Expression ranges from white (no) to red (high).

(N) Barchart showing the BDD2-C8 staining signals in BJ fibroblasts, upon depletion of the genes involved in the secretory pathway. BJ cells were stained with BDD2-C8 at 72 hrs post siRNA mediated knockdown for FACS analysis. The knockdown of these genes did not affect the uptake of BDD2-C8 in both cell types. However, the knockdown negatively influenced the secretion and removal of BDD2-C8 from BJ fibroblasts.

(O) Representative images for BDD2-C8 staining in BJ fibroblasts upon depletion of the genes involved in the secretory pathway. BJ cells were stained with BDD2-C8 at 72 hrs post siRNA mediated knockdown. Cells were stained with BDD2-C8 for 1 hr followed by de-staining for 3 hr.



Supplementary Figure 5

Figure S5. Identification of surface markers to enrich for the early reprogrammed cells

(A) Boxplots showing the expression of surface markers across the reprogramming time-points and their respective subgroups. Lines in the box represent the median expression.

(B) Dotplot indicating the expression of *PROCR* (*CD201*) along the pseudotime. Smooth lines represent the mean expression level at each pseudotime, regardless of the state.

(C) Boxplots showing the expression of surface markers in the $D8_{BDD2\text{-}C8\ high}$ and $D8_{BDD2\text{-}C8\ low}$ cells. Lines represent the median expression.

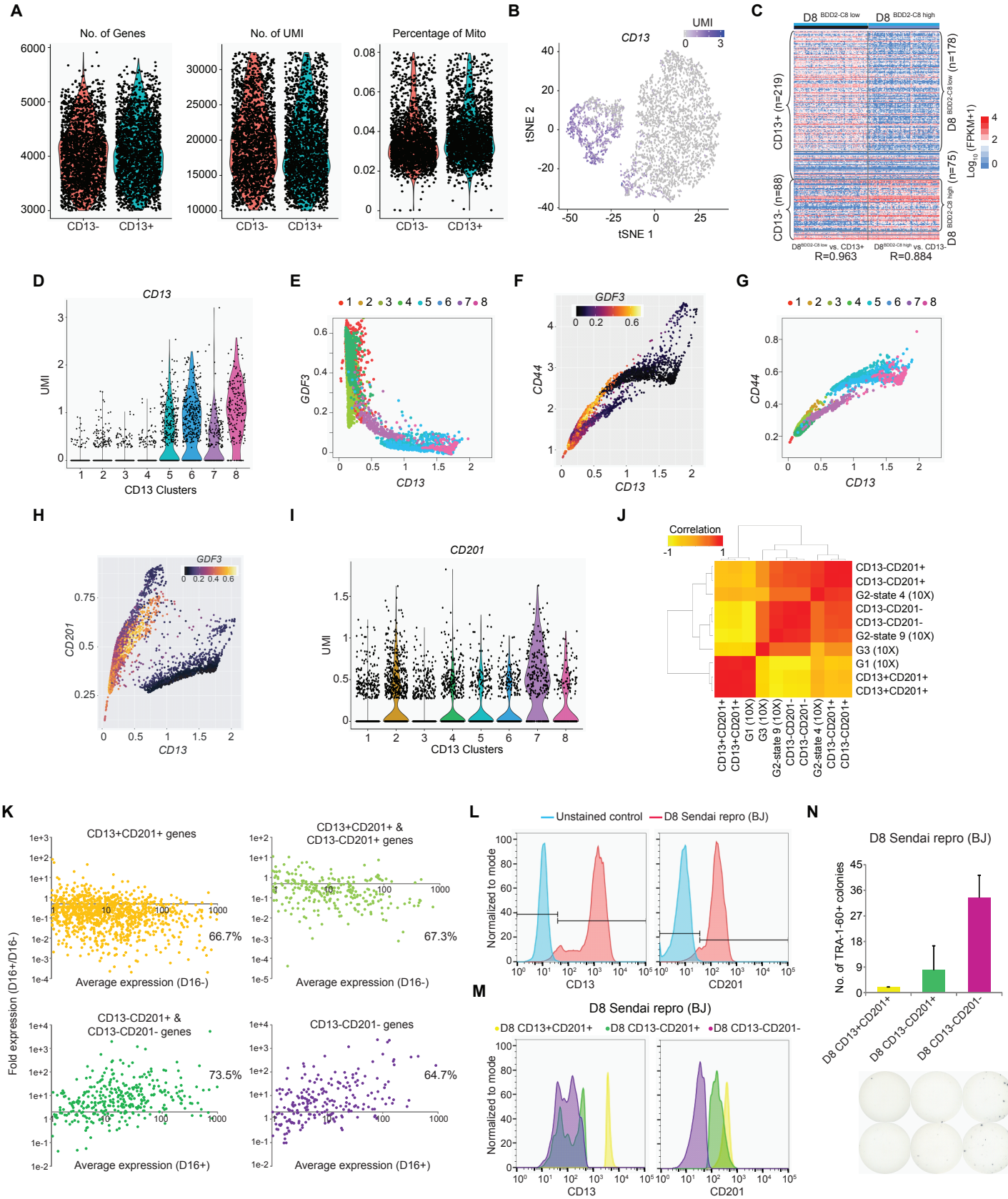
(D) Dotplots showing the co-staining signals of CD13 and CD44 in the reprogramming cells of the indicated time-points. Lines indicate the gating threshold.

(E) Dotplots showing the co-staining signals of CD13 and CD201 in the reprogramming cells of the indicated time-points. Lines indicate the gating threshold.

(F) Overlaid histograms showing the fluorescence intensity (X-axis) of CD13 (left), CD44 (middle), and CD201 (right) in the MSC parental cells (top panel) and D8 Sendai reprogramming cells induced from MSCs (bottom panel). The numbers on top indicate the percentage of positively stained cells with the respective surface markers.

(G) Dotplot showing the co-staining signals of CD13 and CD44 in the D8 Sendai reprogramming cells induced from MSCs. Lines indicate the gating threshold.

(H) Histogram showing the CD13 staining intensities in the sorted CD13+ and CD13- cells.



Supplementary Figure 6

Figure S6. Analysis of CD13 sorted 10X scRNA-Seq libraries

(A) Violin plots demonstrating the number of genes (left), unique molecular identifiers (UMI) (middle), and percentage of mitochondrial DNA detected (right) in the D8 CD13 sorted 10X scRNA-Seq libraries.

(B) Superimposition of *CD13* expression on the tSNE plot of the 10X D8 CD13 sorted scRNA-Seq libraries. Color represents expression level, ranging from grey (low) to purple (high).

(C) Heatmap showing the expression of the differentially expressed genes of CD13- and CD13+ in $D8_{BDD2\text{-}C8\ high}$ and $D8_{BDD2\text{-}C8\ low}$ cells. Pearson correlation is calculated between $D8_{BDD2\text{-}C8\ low}$ and CD13+, and between $D8_{BDD2\text{-}C8\ high}$ and CD13-, which are indicated below. Color represents the expression level, ranging from dark blue (low) to dark red (high).

(D) Violin plot demonstrating the expression of *CD13* across the identified CD13 clusters.

(E) Superimposition of CD13 clusters on the MAGIC plot for *CD13* and *GDF3*. Color indicates the CD13 clusters.

(F) MAGIC plot showing the correlative expression of *CD13* and *CD44* in the D8 CD13 sorted 10X libraries. Color represents the expression level of *GDF3*, ranging from black (no) to yellow (high).

(G) Superimposition of CD13 clusters on the MAGIC plot for *CD13* and *CD44*. Color indicates the CD13 clusters.

(H) MAGIC plot showing the correlative expression of *CD13* and *CD201* in the D8 CD13 sorted 10X libraries. Color represents the expression level of *GDF3*, ranging from black (no) to yellow (high).

(I) Violin plot demonstrating the expression of *CD201* across the identified CD13 clusters.

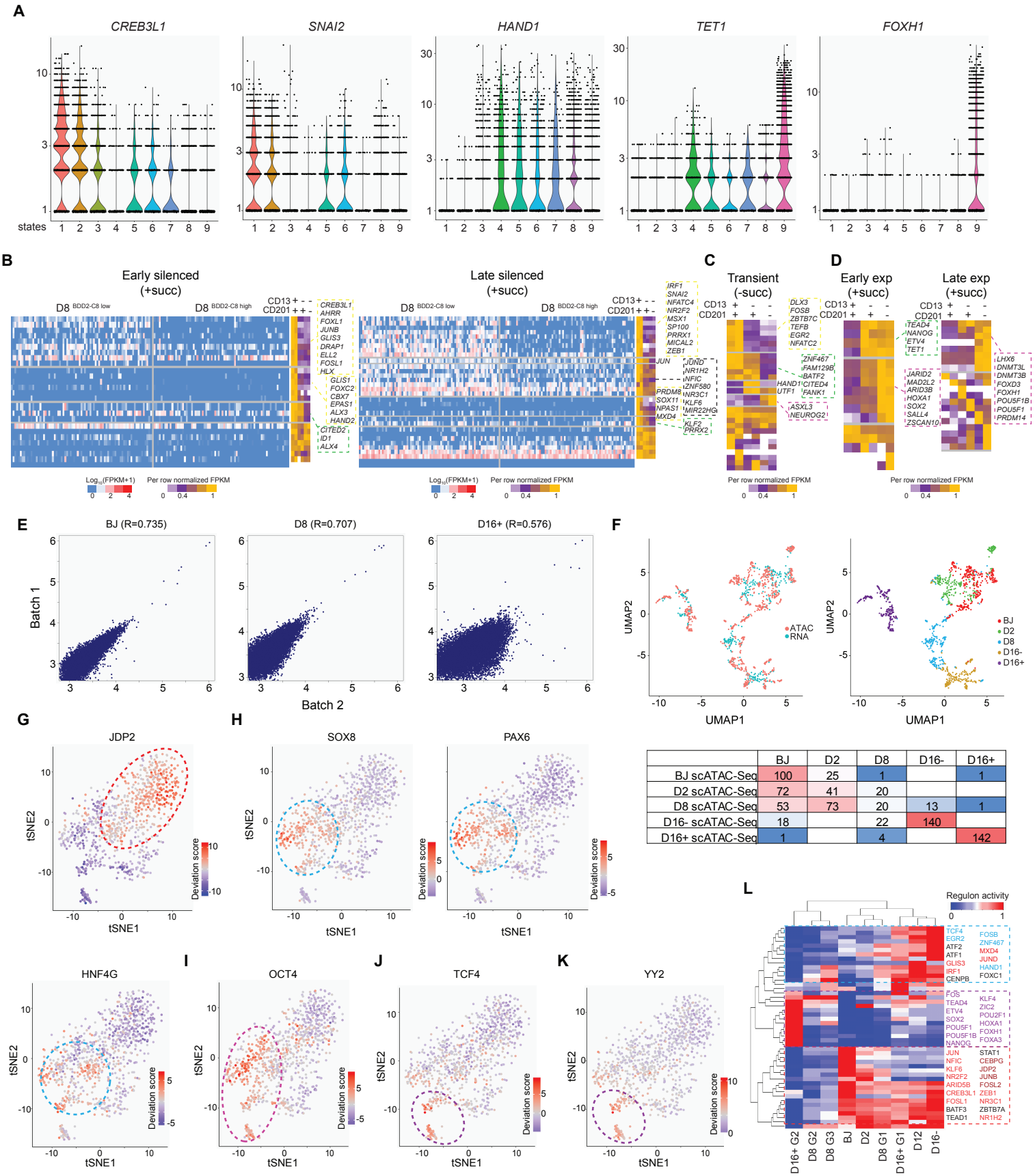
(J) Unsupervised clustering of D8 subgroups (10X) and CD13-CD201 sorted cells, based on the Correlation Coefficient calculated between them. Color represents the Correlation Coefficient value, ranging from yellow (low) to red (high).

(K) Dotplots showing the comparative expression of DEGs of the dual antibodies sorted D8 cells, in D16+ over D16- cells. Each dot represents a gene highly expressed in the respective sorted cells. Y-axis represents the fold change in expression (D16+/D16-), whereas X-axis denotes the mean expression of the respective gene in either D16- (top) or D16+ (bottom) cells. The number besides each plot indicates the percentage of genes expressed higher (more than 2 fold) in D16- (top) or D16+ (bottom) cells.

(L) Overlaid histograms showing the staining intensities of CD13 (left) and CD201 (right) in the D8 Sendai reprogramming cells, induced from BJ cells.

(M) Histograms showing the fluorescence intensities (X-axis) of CD13 (left) and CD201 (right) in the D8 Sendai reprogramming cells with the indicated surface markers profiles. The reprogramming was induced from BJ cells using Sendai virus.

(N) Barchart exhibiting the number of TRA-1-60+ colonies generated from CD13/CD201 dual antibodies sorted D8 cells (top). The reprogramming was induced from BJ cells using Sendai virus. Representative images are shown below. n=2; error bar indicates SD.



	BJ	D2	D8	D16-	D16+
BJ scATAC-Seq	100	25	1		1
D2 scATAC-Seq	72	41	20		
D8 scATAC-Seq	53	73	20	13	1
D16- scATAC-Seq	18		22	140	
D16+ scATAC-Seq	1		4		142

Supplementary Figure 7

Figure S7. Classification of TFs based on the dynamics of their expression and motif accessibility across reprogramming

(A) Violin plots demonstrating the expression of the representative TFs belonging to the categories defined in Figure 5A, across the pseudotemporal states.

(B) Heatmaps showing the expression of Early (left) or Late (right) Silenced TFs in $D8_{BDD2\text{-}C8\ high}$ and $D8_{BDD2\text{-}C8\ low}$ cells. Color represents the expression level and ranges from dark blue (low) to dark red (high expression). Expression of these TFs in the CD13/CD201 dual antibodies sorted D8 cells are indicated beside. TFs that are highly expressed in double positive or single positive cells are labelled on the right. Color represents the per row normalized expression value, which is scaled expression normalized against the maximum value of the respective gene across the samples.

(C) Heatmap showing the expression of the Transiently expressed TFs in the CD13/CD201 dual antibodies sorted D8 cells. Differentially expressed TFs are labelled on the right. Color represents the per row normalized expression value, which is scaled expression normalized against the maximum value of the respective gene across the samples.

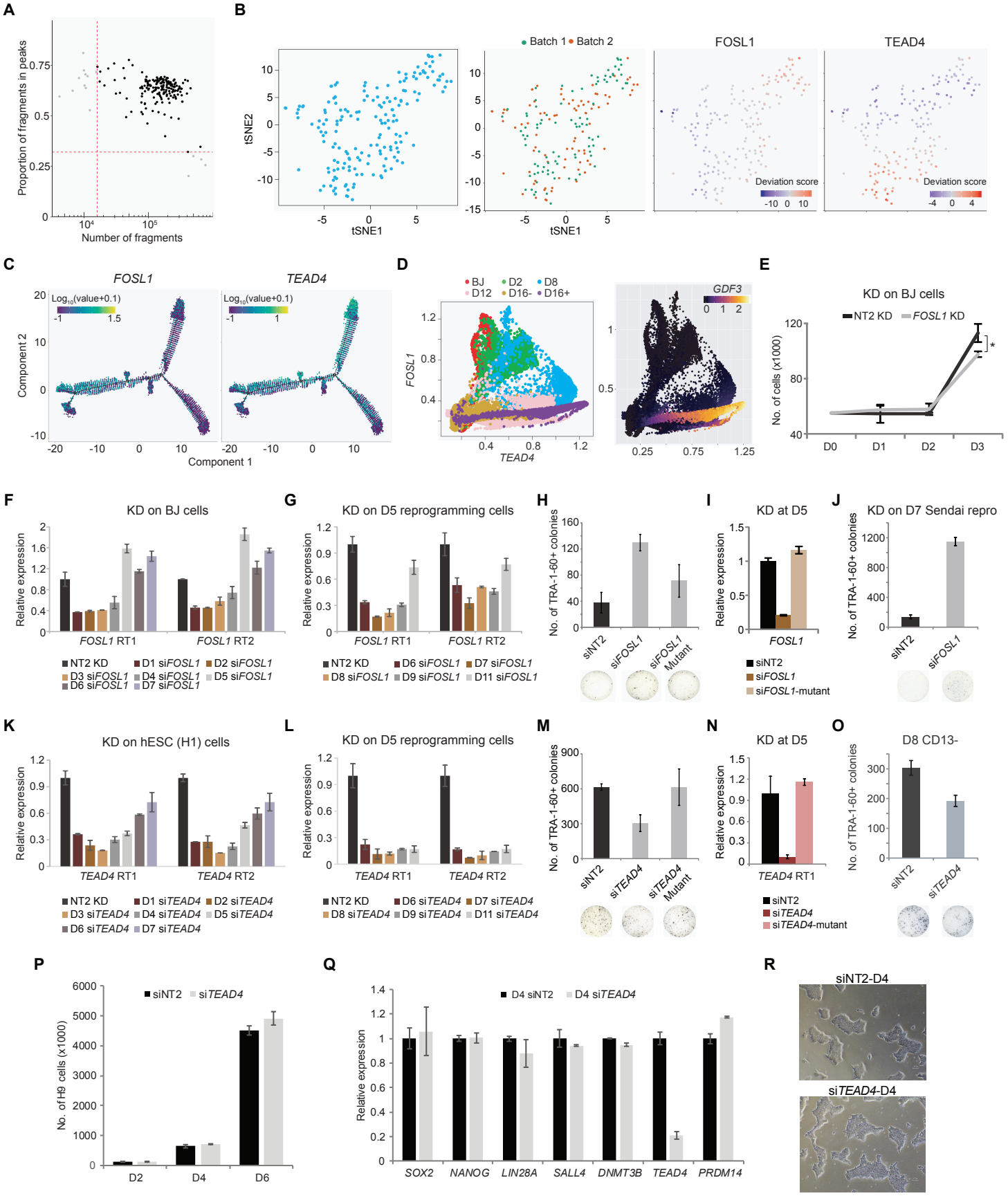
(D) Heatmaps showing the expression of the Early (left) and Late (right) Expressed TFs in the CD13/CD201 dual antibodies sorted D8 cells. TFs that are highly expressed in double negative cells are labelled on the right. Color represents the per row normalized expression value, which is scaled expression normalized against the maximum value of the respective gene across the samples.

(E) Correlation between the duplicate scATAC-Seq libraries (batch 1 and 2) of BJ (left), D8 (middle) and D16+ (right) cells, based on the promoters accessibility levels. The axes represent the log10 of the accessibility level in each replicate.

(F) Top: Integrative UMAP plots for scATAC-Seq and scRNA-Seq libraries (Fluidigm). Library type is labelled on the left plot, and time-points of scRNA-Seq libraries and the predicted time-points of scATAC-Seq are labelled on the right plot. Bottom: Table showing the number of scATAC-Seq libraries (row) predicted to be the scRNA-Seq of the indicated time-points (column).

(G-K) Superimposition of the motif enrichment scores for OC motif- JDP2 (G), Transient motif-SOX8, PAX6, and HNF4G (H), and CO motif: type I-OCT4 (I); type II-TCF4 (J) and YY2 (K) on the scATAC-Seq tSNE plot. Color indicates the accessibility level and ranges from dark blue (no enrichment) to dark red (high enrichment).

(L) Heatmap demonstrating the regulon activity of TFs calculated from the 10X scRNA-Seq libraries by SCENIC, across the reprogramming time-points and their subgroups. Color indicates the regulon activity level, ranging from blue (low) to red (high). Silenced or OC TFs are colored in red and dark red, and Transient TFs are colored in blue, whereas Expressed and CO TFs are colored in purple.



Supplementary Figure 8

Figure S8. Functional roles of FOSL1 and TEAD4 during cellular reprogramming

(A) Dotplot revealing the proportion of fragments in peaks (Y-axis) and library size (X-axis) for each D8 scATAC-Seq library. The red dotted lines represent the threshold for each criterion.

(B) Left: t-SNE plot of D8 scATAC-Seq libraries based on the deviation scores of JASPAR motifs. Right: Superimposition of the batch number and accessibility scores of FOSL1 and TEAD4 motifs on the t-SNE plot.

(C) Superimposition of *FOSL1* and *TEAD4* expression on the reprogramming trajectory.

(D) Left: MAGIC plot showing correlative expression of *FOSL1* and *TEAD4* in the 10X libraries. Right: Superimposition of *GDF3* expression on the MAGIC plot.

(E) Line plot demonstrating the proliferation of BJ cells in which *FOSL1* was depleted (*FOSL1* KD). Non-targeting siRNA construct (NT2 KD) was used as control. X-axis denotes the time-point post siRNA transfection, whereas Y-axis represents the number of cells on the respective time-point. n=3; error bar indicates SD.

(F) Bar chart exhibiting the relative expression of *FOSL1* (Y-axis) at the indicated time-points (X-axis) after depleting *FOSL1* using siRNA in BJ cells. Expression was measured by qRT-PCR. Relative expression of *FOSL1* was calculated by normalizing to the Non-targeting siRNA construct (siNT2) of the respective time-point. n=2. Error bar indicates SD.

(G) Bar chart exhibiting the relative expression of *FOSL1* (Y-axis) in the reprogramming cells of the indicated time-points (X-axis) upon depleting *FOSL1* at D5 of reprogramming. Expression was measured by qRT-PCR. Relative expression of *FOSL1* was calculated by normalizing to the Non-targeting siRNA construct (siNT2) of the respective time-point. n=2. Error bar indicates SD.

(H) Bar chart revealing the number of TRA-1-60+ colonies formed (Y-axis) from wells in which *FOSL1* was depleted (si*FOSL1*) at D5 of reprogramming. Mutant construct (si*FOSL1*-mutant) was used as control along with a non-targeting construct (siNT2). Representative images are shown below. n=3; error bar indicates SD.

(I) Bar chart exhibiting the relative expression of *FOSL1* (Y-axis) upon depleting *FOSL1* using wild-type (si*FOSL1*) or mutant (si*FOSL1*-mutant) siRNA constructs at D5 of reprogramming. Expression was measured by qRT-PCR. Relative expression of *FOSL1* was calculated by normalizing to the Non-targeting siRNA construct (siNT2). n=2. Error bar indicates SD.

(J) Bar chart demonstrating the number of TRA-1-60+ colonies (y-axis) upon knock-down *FOSL1* at D7 of Sendai virus reprogramming, induced from BJ cells. Non-targeting siRNA construct (siNT2) was used as control. Representative images are shown below, n=3. Error bar indicates SD.

(K) Bar chart exhibiting the relative expression of *TEAD4* (Y-axis) at the indicated time-points after depleting *TEAD4* using siRNA in hESCs. Expression was measured by qRT-PCR. Relative

expression of *TEAD4* was calculated by normalizing to the Non-targeting siRNA construct (siNT2) of the respective time-point. n=2. Error bar indicates SD.

(L) Bar chart exhibiting the relative expression of *TEAD4* (Y-axis) in the reprogramming cells of the indicated time-points (X-axis) upon depleting *TEAD4* using siRNA at D5 of reprogramming. Expression was measured by qRT-PCR. Relative expression of *TEAD4* was calculated by normalizing to the Non-targeting siRNA construct (siNT2) of the respective time-point. n=2. Error bar indicates SD.

(M) Bar chart revealing the number of TRA-1-60+ colonies formed (Y-axis) from wells in which *TEAD4* was depleted (si*TEAD4*) at D5 of reprogramming. Mutant construct (si*TEAD4*-mutant) was used as control along with a non-targeting construct (siNT2). Representative images are shown below, n=3. Error bar indicates SD.

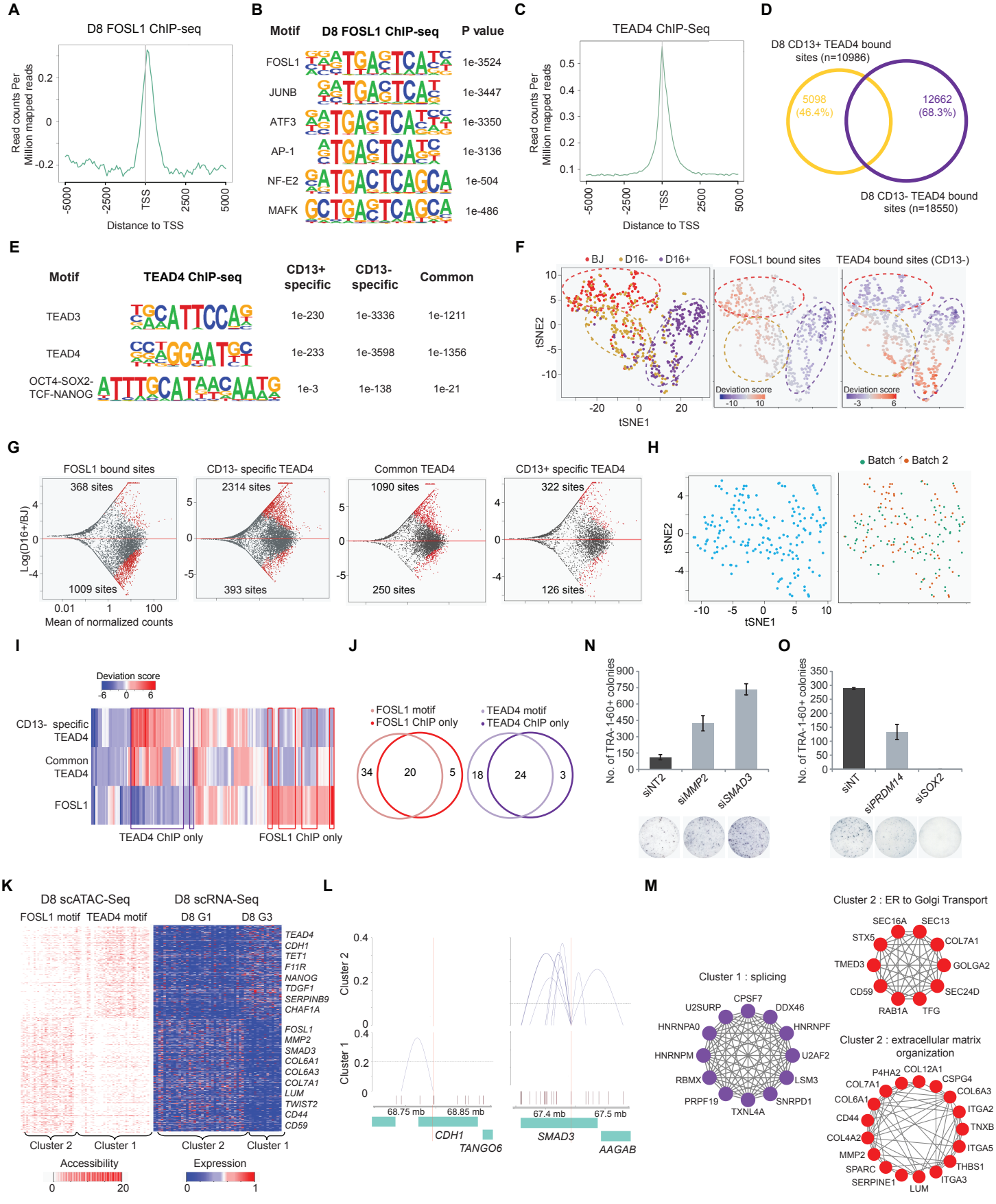
(N) Bar chart exhibiting the relative expression of *TEAD4* (Y-axis) upon depleting *TEAD4* using wild-type (si*TEAD4*) or mutant (si*TEAD4*-mutant) siRNA constructs at D5 of reprogramming. Expression was measured by qRT-PCR. Relative expression of *TEAD4* was calculated by normalizing to the Non-targeting siRNA construct (siNT2). n=2. Error bar indicates SD.

(O) Bar chart demonstrating the number of TRA-1-60+ colonies yielded from D8 CD13- cells with *TEAD4* depletion. Representative images are shown below, n=3. Error bar indicates SD.

(P) Barchart revealing the number of H9 cells at the indicated day after si*TEAD4* transfection. n=3; error bar indicates SD.

(Q) Barchart showing the relative expression of *TEAD4* at day 4 after si*TEAD4* transfection. Expression was measured by qRT-PCR. Relative expression calculated by normalizing to siNT2. n=2; error bar indicates SD.

(R) Representative bright-field images of H9 cells at day 4 post si*TEAD4* transfection.



Supplementary Figure 9

Figure S9. Genomic binding of FOSL1 and TEAD4

(A) Average enrichment profile of D8 FOSL1 ChIP-Seq reads around transcription start sites (TSS) of the genome with a window of -5K to 5K bp.

(B) Motifs significantly enriched for the D8 FOSL1 ChIP-Seq sites.

(C) Average enrichment profile of TEAD4 ChIP-Seq reads around all transcription start sites (TSS) of the genome with a window of -5K to 5K bp.

(D) Venn diagram showing the overlap of loci detected in D8 CD13- TEAD4 and D8 CD13+ TEAD4 ChIP-Seq libraries. The number and percentage of bound sites falling in each category are indicated.

(E) Motifs significantly enriched for the TEAD4 bound sites detected specifically and commonly in the D8 CD13+ TEAD4, and CD13- TEAD4 ChIP-Seq libraries.

(F) Left: t-SNE clustering of BJ, D16+, and D16- scATAC-Seq libraries based on the deviation scores for D8 FOSL1 ChIP-seq, D8 CD13- TEAD4 ChIP-seq, and D8 CD13+ TEAD4 ChIP-seq sites. Right: Superimposition of deviation scores for FOSL1 specific bound sites and TEAD4 CD13- specific bound sites on the t-SNE plot. Color indicates the accessibility level.

(G) MA plot of scATAC-Seq revealing the differentially accessible FOSL1 and TEAD4 bound sites in BJ and D16+ cells. Y-axis represents the Log fold changes (D16+/BJ), whereas X-axis denotes the mean of normalized counts in BJ cells. Red dots denote the sites with significant accessibility changes.

(H) Left: t-SNE plot of D8 scATAC-Seq libraries based on the deviation scores of D8 FOSL1 ChIP-seq, D8 CD13 sorted TEAD4 ChIP-seq. Right: Superimposition of the batch numbers of the scATAC-Seq libraries on the t-SNE plot.

(I) Heatmap of D8 scATAC-Seq libraries based on the deviations scores of D8 FOSL1 specific ChIP-seq sites, CD13- specific and CD13+/- common TEAD4 specific ChIP-seq sites. Color indicates the accessibility level and ranges from dark blue (no enrichment) to dark red (high enrichment). D8 cells, showing high enrichment for FOSL1 specific ChIP-seq sites only, were labelled as “FOSL1 ChIP only” cells, whereas D8 cells, showing high enrichment for CD13- specific or CD13+/- common TEAD4 specific ChIP-seq only, were labelled as “TEAD4 ChIP only” cells.

(J) Left: Venn diagrams showing the overlap between the IDs of cells with strong enrichment of FOSL1 motif and “FOSL1 ChIP only cells”. Right: Venn diagram showing the overlap between the IDs of cells with strong enrichment of TEAD4 motif and “TEAD4 ChIP only cells”.

(K) Left: Heatmaps demonstrating the highly accessible regions (left) and highly expressed genes (right) of each cluster identified by coupled NMF analysis. The expression heatmap (right) demonstrates the corresponding genes whose expression correlates with the accessibility of the regions in the

accessibility heatmap (left). The highly accessible and expressed genes for each cluster are indicated on the right. Colors represent xpression and accessibility levels.

(L) Line plots indicating the Cicero co-accessibility links between the regions highlighted in red and the distal sites in the surrounding region. The height indicates the Cicero co-accessibility score between the connected peaks. The top sets of links are constructed from cells of cluster 2, while the bottom sets are built from cells of cluster 1. *SMAD3* (right) is highly expressed and accessible and has more interaction with the surrounding regions in cluster 2, whereas *CDH1* (left) are highly active and interactive in cluster 2.

(M) The interactome analysis indicating the top pathways enriched by the cluster 1 (left) and cluster 2 (right) specific genes, which were identified from the coupled NMF analysis.

(N-O) Bar charts demonstrating the number of TRA-1-60+ colonies upon knock-down of FOSL1 targets (N) and TEAD4 targets (O) at D5 of reprogramming. Representative images are shown below. n=3. Error bar indicates SD.