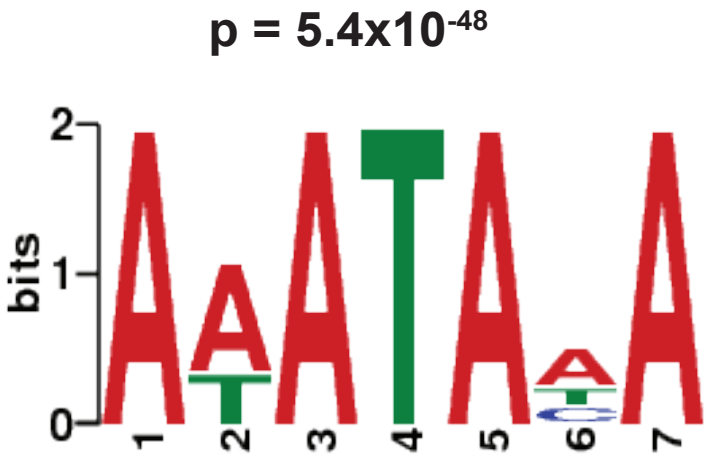
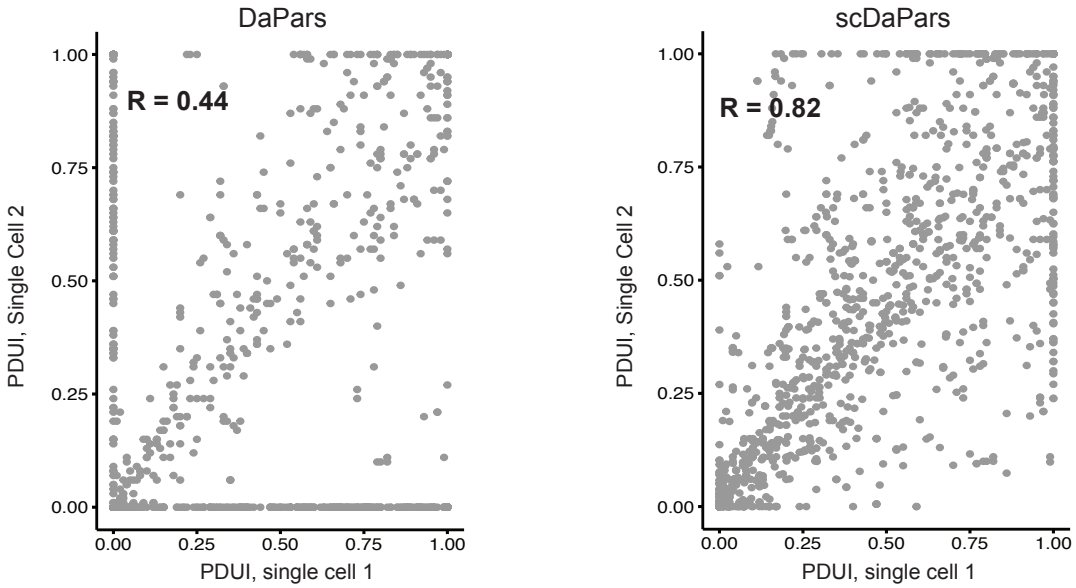
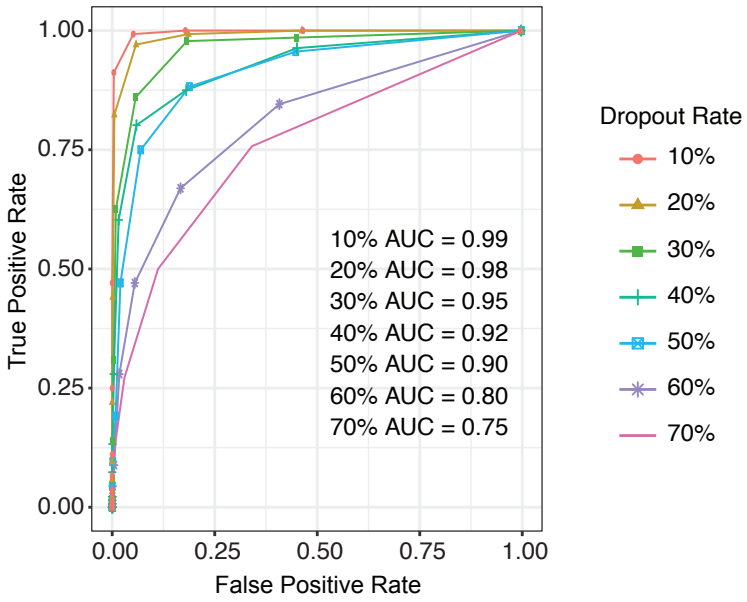


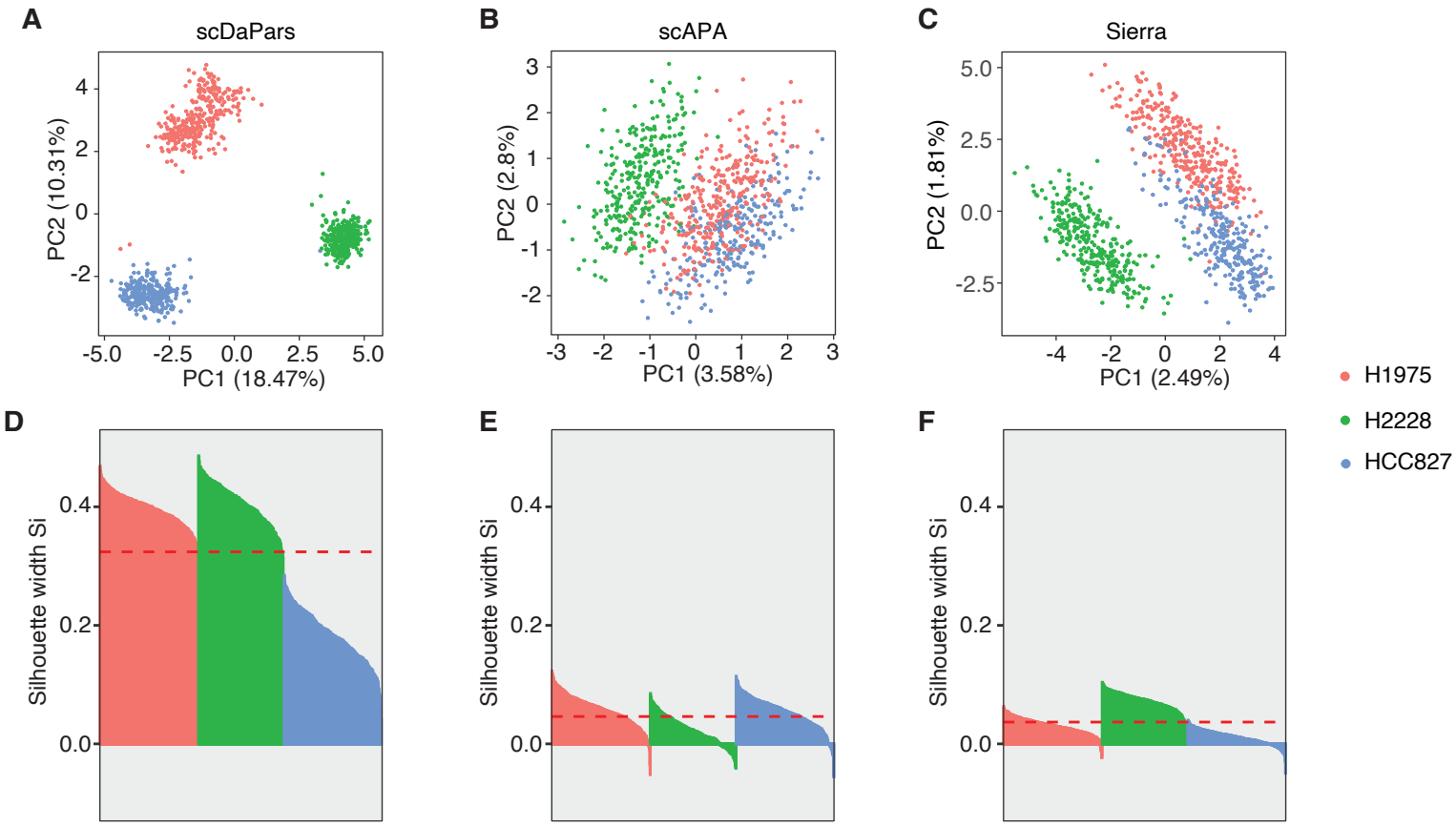
Supplemental Figure S1. The top-scoring signal identified by de novo motif analysis (DREME) among the upstream (-100bp) of predicted poly(A) sites in bulk samples.



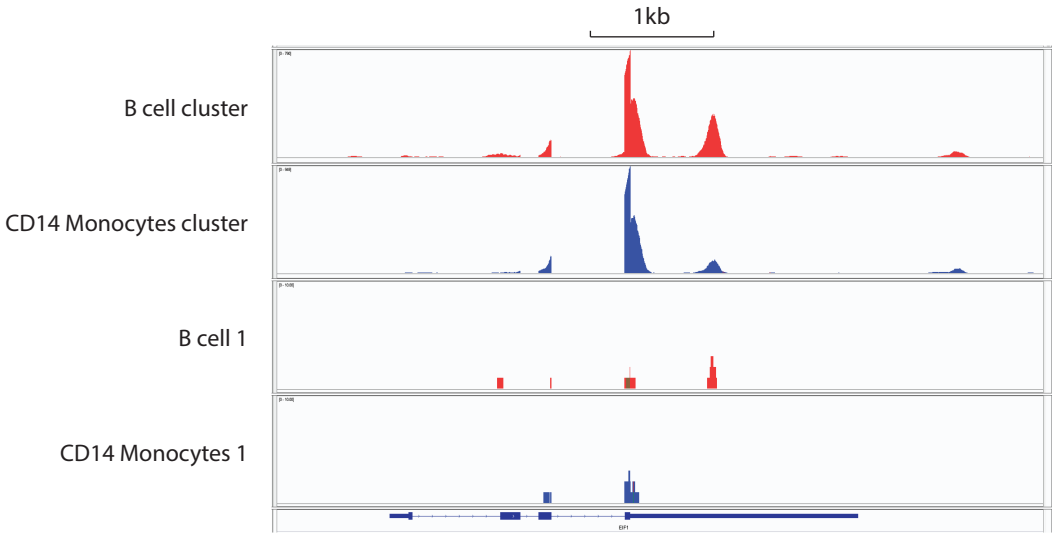
Supplemental Figure S2. Example scatter plots show correlation of PDUI values between B-cell pairs from PBMCs (GSE132044) estimated by bulk methods (i.e. DaPars) (left) and scDaPars (right).



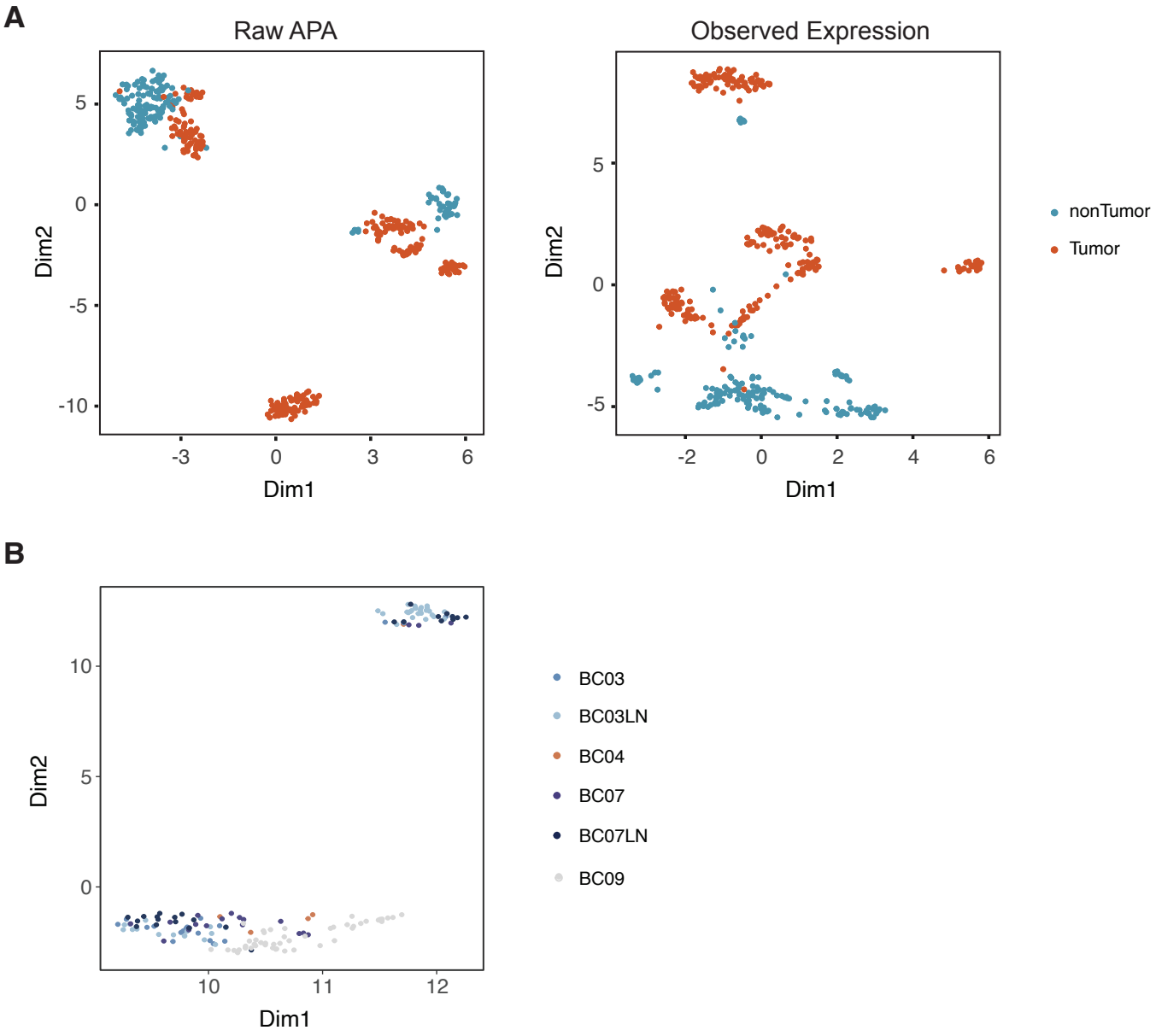
Supplemental Figure S3. ROC curve of dynamic APA events recovered using scDaPars under dropout rates ranging from 10% to 70%.



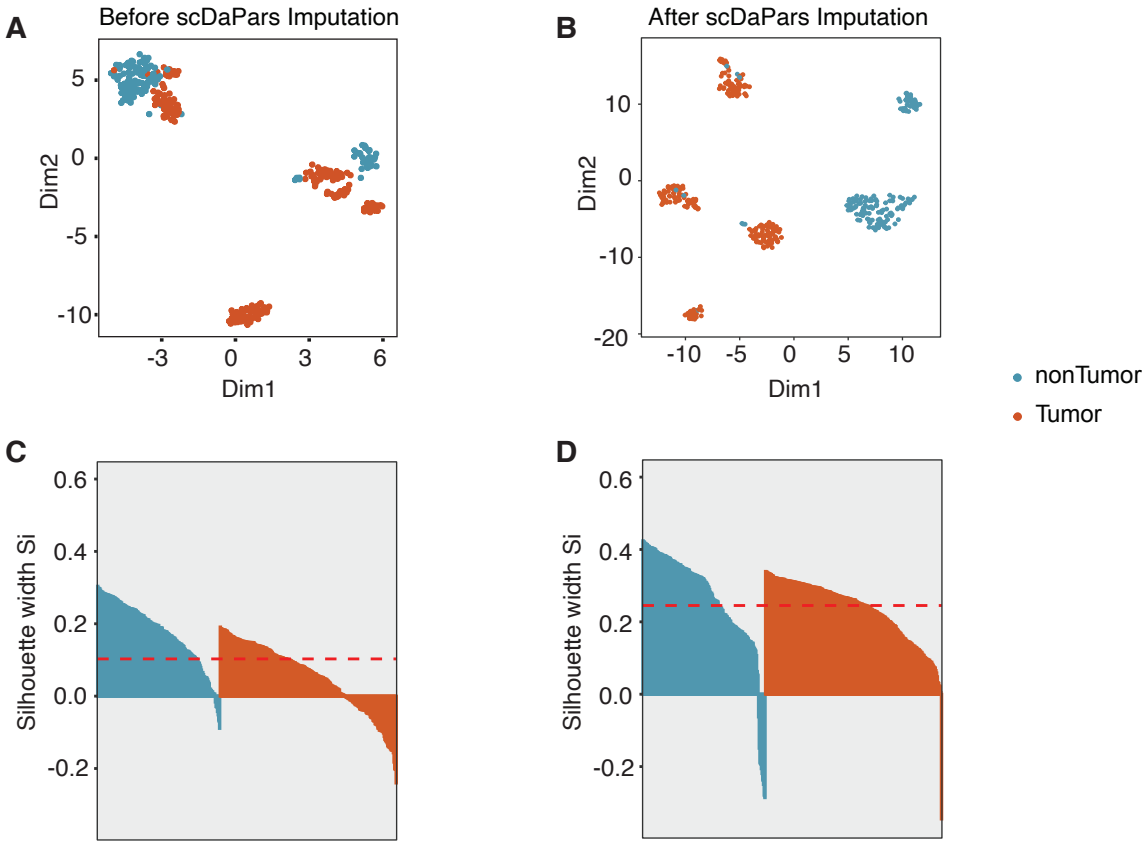
Supplemental Figure S4. Benchmark comparison for scDaPars in lung adenocarcinoma cell lines (GSE118767).
(A) – (C) Scatterplots showing PCA results of 902 lung adenocarcinoma cells based on (A) scDaPars quantified APA usage or (B) scAPA quantified APA usage or (C) Sierra quantified APA usage.
(D) – (F) Silhouette plots for clustering results from (D) scDaPars, (E) scAPA and (F) Sierra. The x-axis represents cells and y-axis is the corresponding silhouette measure Si for each cell. The silhouette coefficient measures how similar a cell is to its own cluster compared to other clusters; therefore, a higher silhouette coefficient indicates better clustering results and a negative coefficient may suggest the cell is assigned to the wrong cluster. The red dashed line is the average Si for all cells.



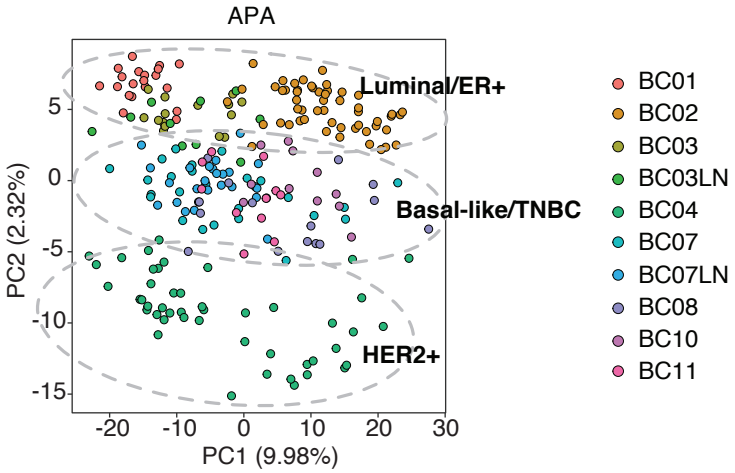
Supplemental Figure S5. Coverage plots of EIF1 gene for B-cell and CD14+-monocyte clusters (extract and combine reads belong to the same cell cluster together) (top two panel) and individual B cell and CD14+ monocyte (bottom two panel).



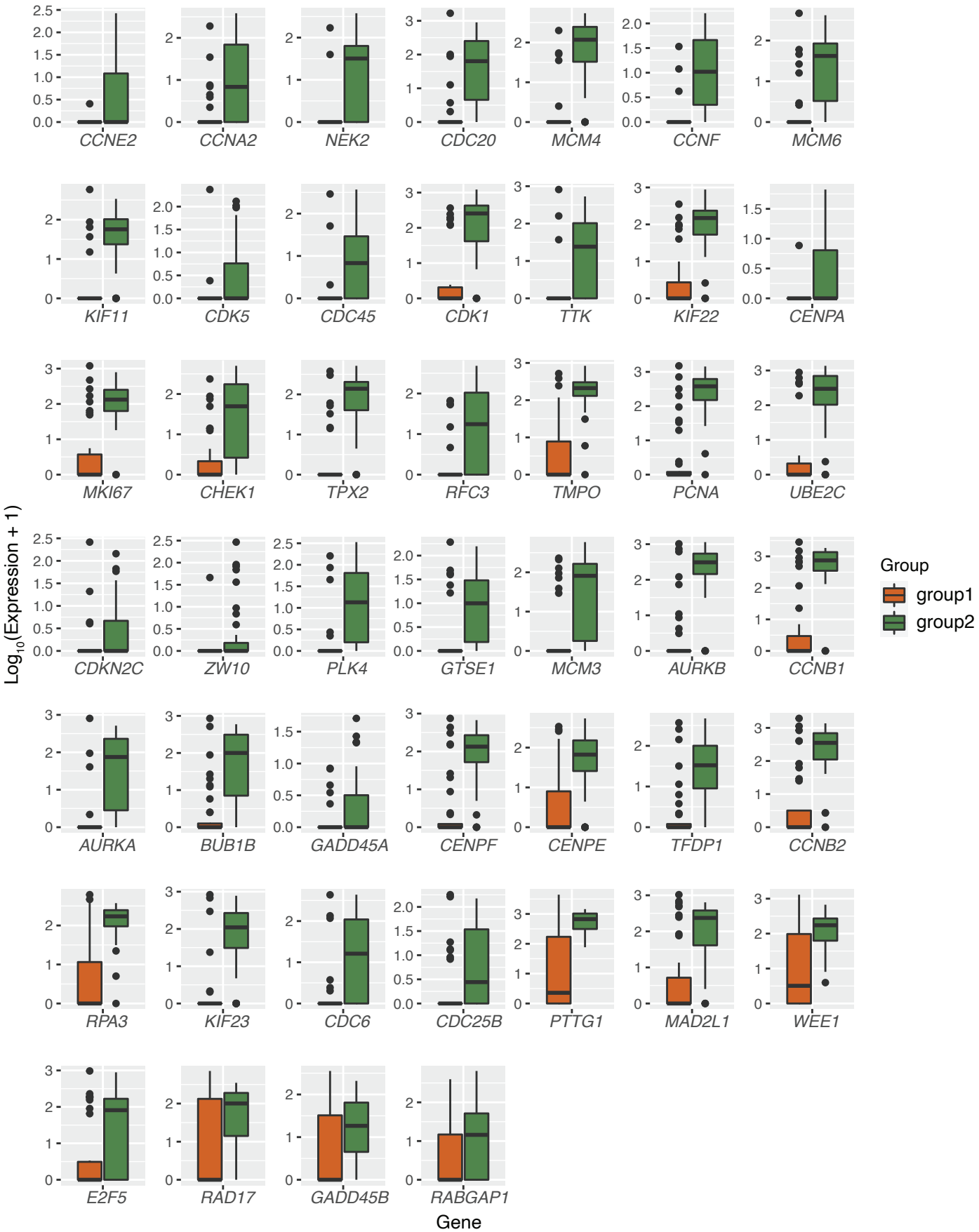
Supplemental Figure S6. Scatter plots of additional UMAP analysis in breast cancer (GSE75688).
(A) Scatter plots shows UMAP results for raw APA profiles (left) and observed/raw gene expression (right). Cells are colored based on cell information provided in the original publication. (B) Scatter plot of UMAP results of Immune cells. Cells are labeled by patient information.



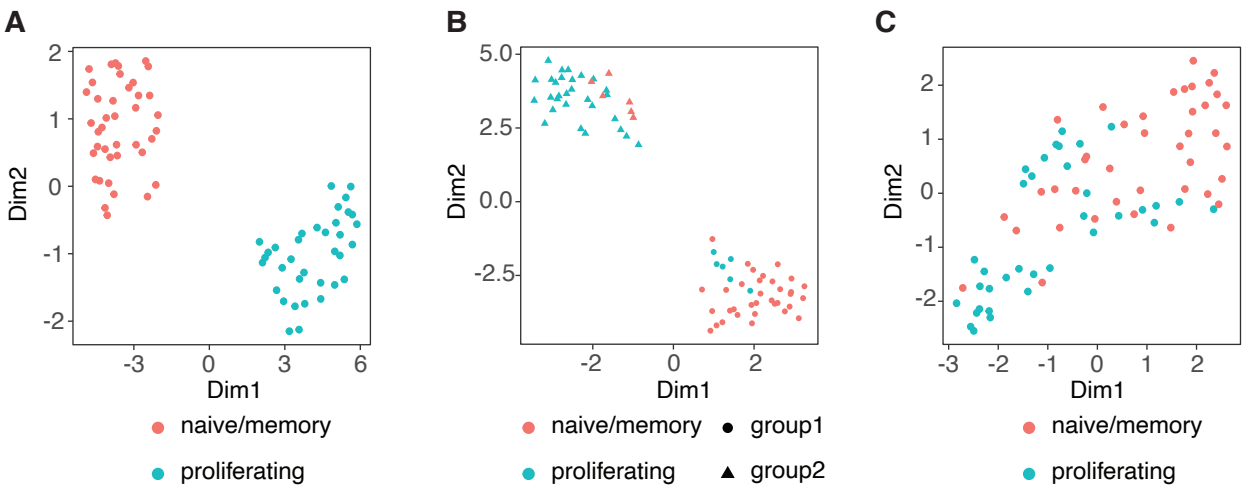
Supplemental Figure S7. scDaPars imputed APA usage better seperate tumors cells from non-tumor cells (GSE75688).
(A) - (B) Scatterplots showing UMAP results of Tumor and Non-tumor cells based on APA usage (A) before scDaPars imputation and (B) after scDaPars imputation.
(C) – (D) Silhouette plots for clustering results from (C) before scDaPars imputation and (D) after scDaPars imputation. The x-axis represents cells and y-axis is the corresponding silhouette measure Si for each cell. The silhouette coefficient measures how similar a cell is to its own cluster compared to other clusters; therefore, a higher silhouette coefficient indicates better clustering results and a negative coefficient may suggest the cell is assigned to the wrong cluster. The red dashed line is the average Si for all cells.



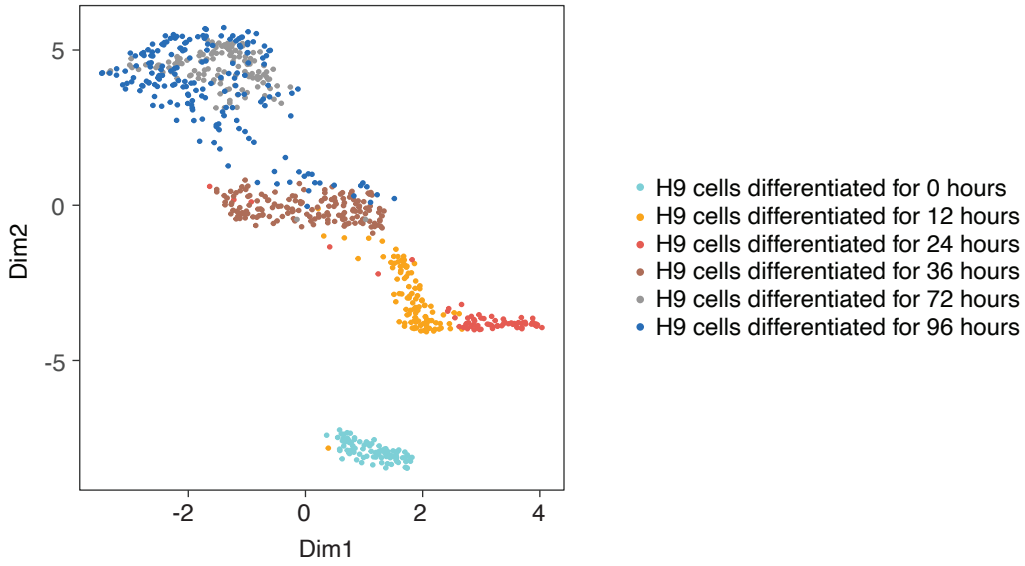
Supplemental Figure S8. The first two Principal Components calculated from the scDaPars-derived APA profile for tumor cells including cells from patients with <20 single cells.



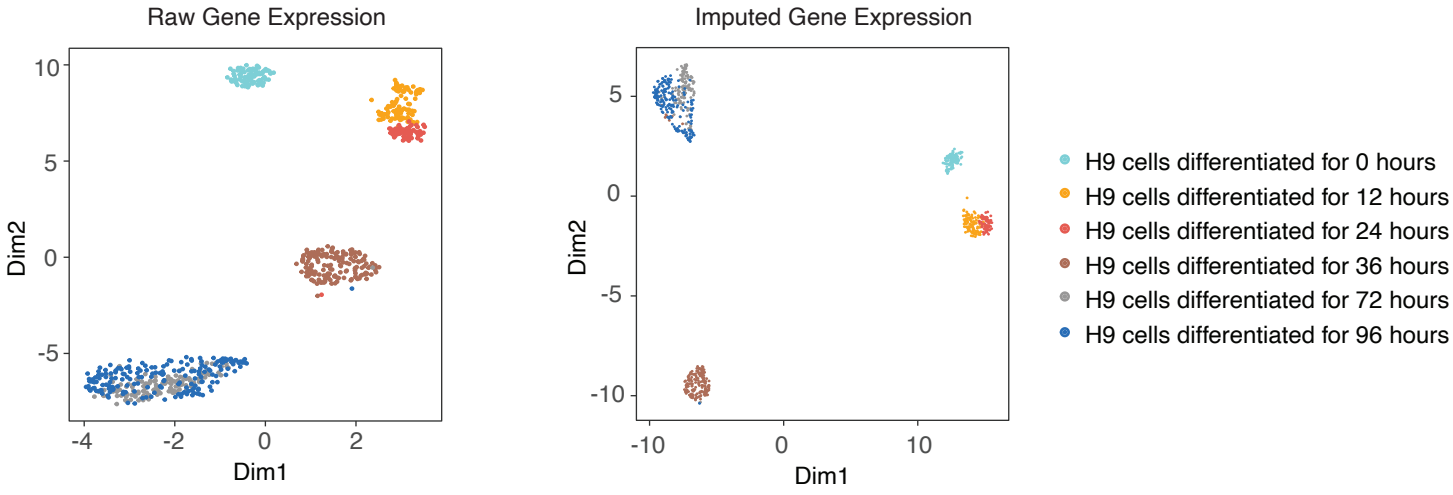
Supplemental Figure S9. Boxplots of expression level of B cells proliferation signature genes between group 1 B cells and group 2 B cells in Fig.4e (GSE75688).



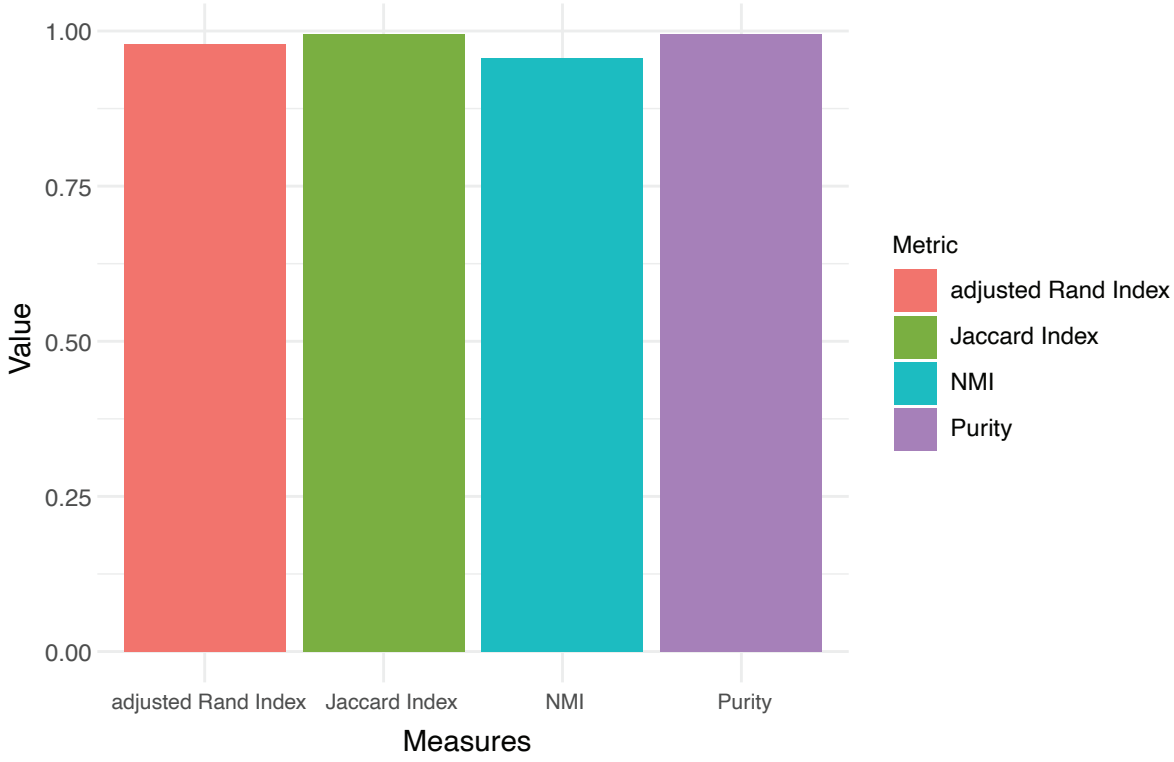
Supplemental Figure S10. Scatter plots of UMAP results in B cells from breast cancer based on (A) Expression of B cell proliferating signature genes, (B) scDaPars APA results, (C) Expression of all genes. cells are colored based on proliferating status.



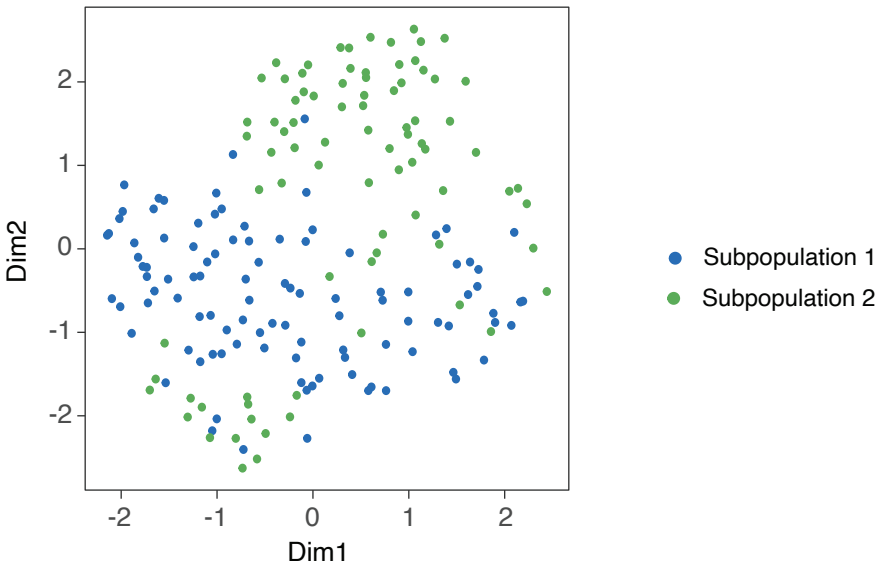
Supplemental Figure S11. The first two dimensions of the UMAP results calculated from raw PDUI values (before scDaPars imputation steps) in definitive endoderm differentiation (GSE75748).



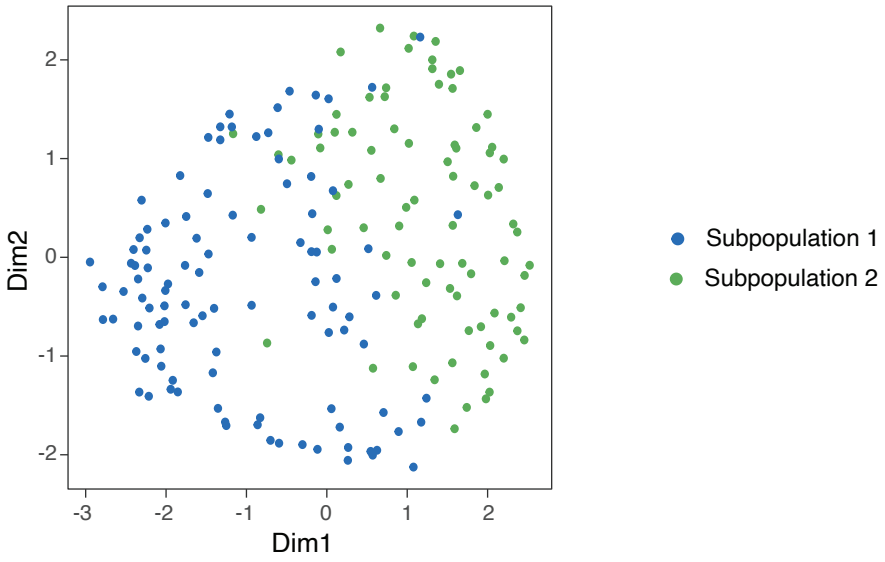
Supplemental Figure S12. The first two dimensions of the UMAP results calculated from raw and Imputed gene expression of time-course definitive endoderm cells (GSE75748).



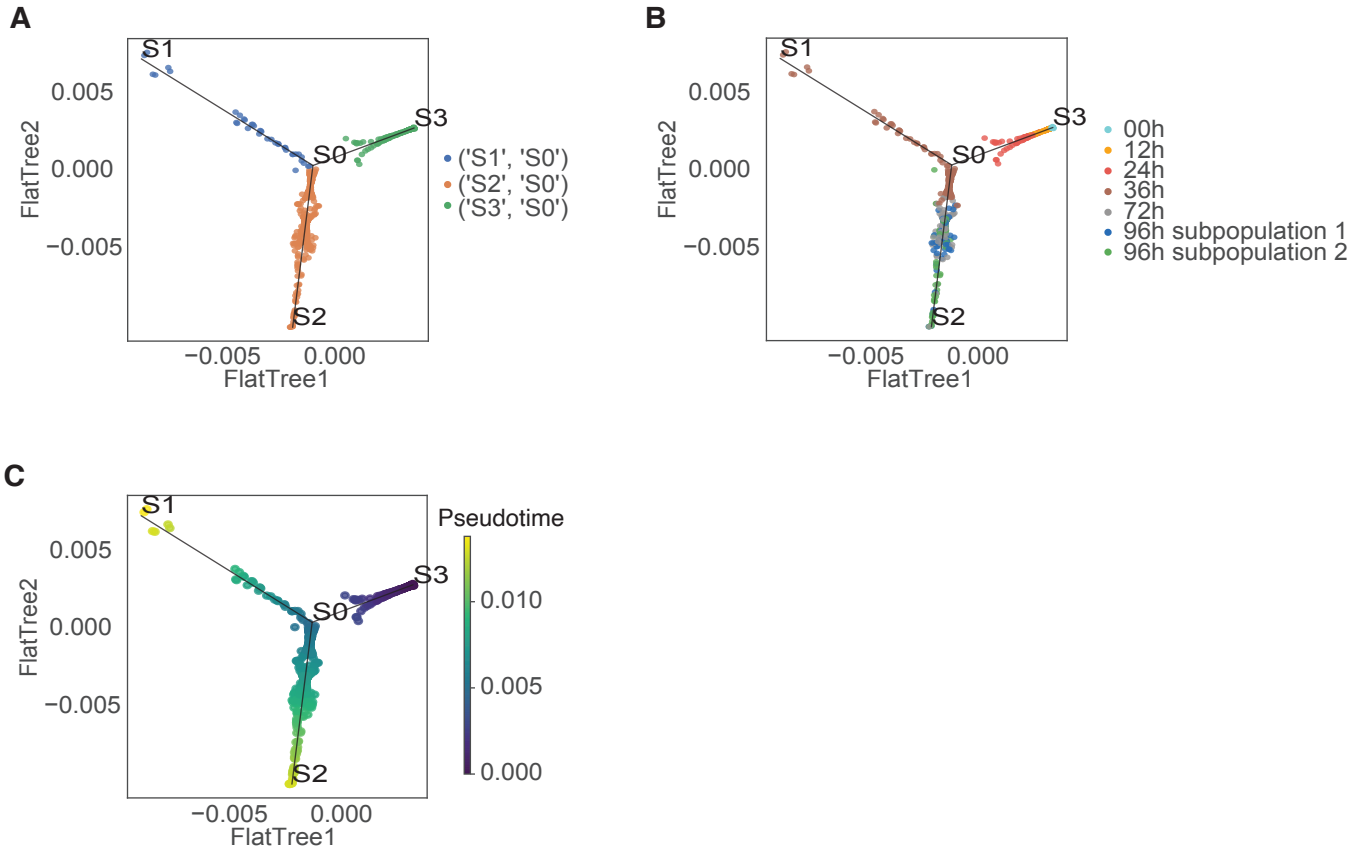
Supplemental Figure S13. The adjusted Rand index, Jaccard index, Normalized Mutual Information (NMI), and Purity scores of clustering results based on APA alone. Clustering was performed by K-means clustering with k = 2 and compared with cell labels in Figure 6B. The four measures are all between 0 and 1, with 1 indicating a perfect match between APA-alone clustering and cell labels in Figure 6B.



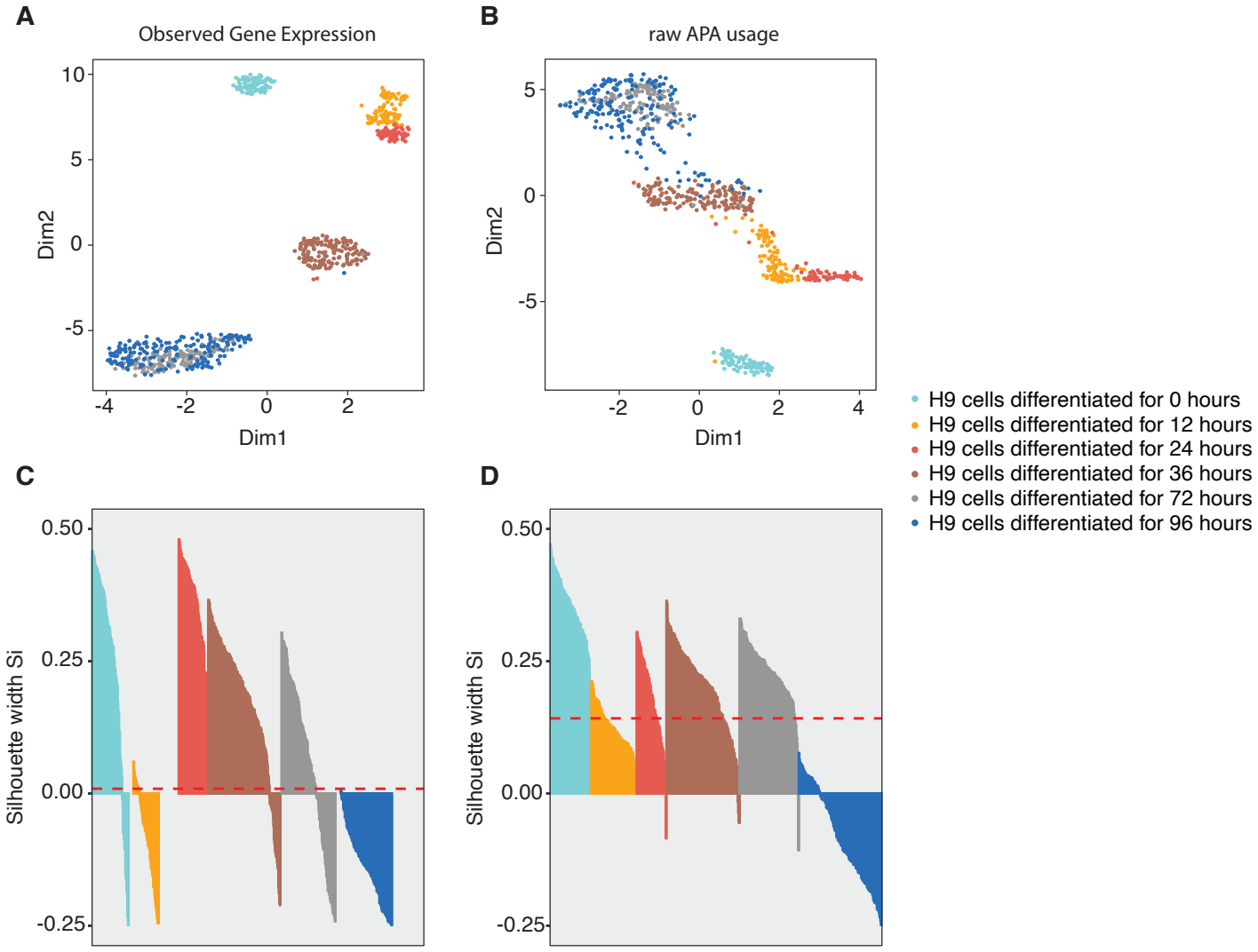
Supplemental Figure S14. The first two dimensions of the UMAP results of definitive endoderm cells at 96h of differentiation calculated from APA dynamics generated by DaPars.



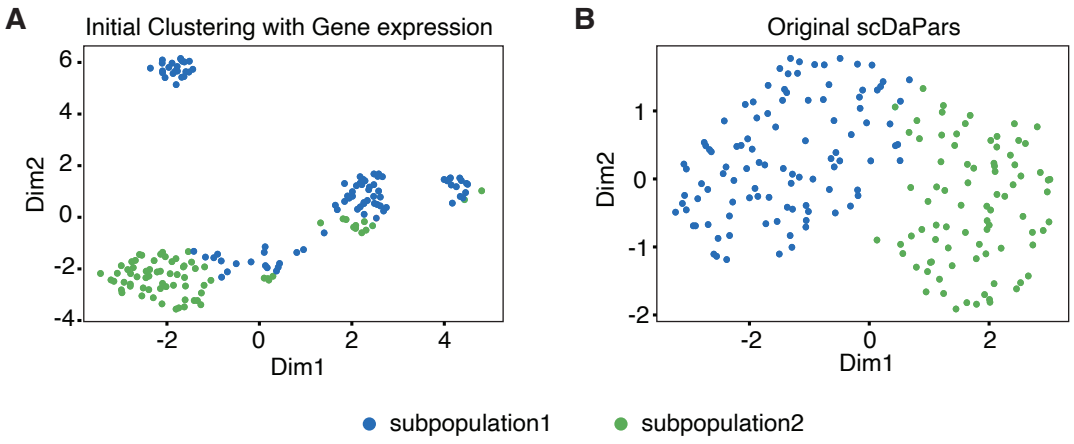
Supplemental Figure S15. The first two dimensions of the UMAP results of definitive endoderm cells at 96h of differentiation calculated from imputed gene expression of differential APA genes.



Supplemental Figure S16. Flat tree plots for single cell trajectory analysis using gene expression data. Cells are colored by (A) STREAM identified branches, (B) cell differentiation time points, (C) STREAM calculated pseudotime.



Supplemental Figure S17. (A) - (B) Scatterplots showing UMAP results of 563 hESCs based on (A) observed gene expression and (B) raw APA usage.
(C) – (D) Silhouette plots for clustering results from (C) observed gene expression and (D) raw APA usage. The ground truth is the annotated cell label in the original literature. The x-axis represents cells and y-axis is the corresponding silhouette measure Si for each cell. The red dashed line is the average Si for all cells.



Supplemental Figure S18. Scatterplots showing UMAP results of 185 hESCs at 96 h of differentiation based on APA usage derived from (A) modified method (use gene expression for initial clustering) and (B) original scDaPars (use raw APA usage for initial clustering).

Method	Single-cell level	Single-gene level	Cell-type-label free	Data Type
scDaPars	√	√	√	3’ end & full-length
scAPA		√		3’ end only
scDAPA		√		3’ end only
Chung et al.	√		√	3’ end & full-length

Supplemental Table S1. Comparison of different scRNA-seq APA quantification methods

Gene Symbol	Log_2 Fold Change (group2/group1)	P-value	Gene Symbol	Log_2 Fold Change (group2/group1)	P-value
CCNE2	8.637907005	4.665×10^{-9}	*PLK4*	2.243609617	2.724×10^{-2}
CCNA2	4.803637943	9.399×10^{-7}	*GTSE1*	2.033800133	2.849×10^{-2}
NEK2	5.043927629	9.463×10^{-7}	*MCM3*	2.041564617	4.137×10^{-2}
CDC20	4.495114905	1.630×10^{-6}	*AURKB*	1.870984406	4.462×10^{-2}
MCM4	3.991918849	2.174×10^{-6}	*CCNB1*	1.660149707	5.296×10^{-2}
CCNF	4.105056458	4.720×10^{-6}	*AURKA*	2.024196574	7.011×10^{-2}
MCM6	3.806024739	6.365×10^{-6}	*BUB1B*	1.690663297	8.141×10^{-2}
KIF11	3.305184966	8.099×10^{-5}	*GADD45A*	2.498681071	8.849×10^{-2}
CDK5	5.729766087	9.758×10^{-5}	*CENPF*	1.397881644	1.135×10^{-1}
CDC45	3.846374532	3.668×10^{-4}	*CENPE*	1.233362375	1.252×10^{-1}
CDK1	2.862335992	8.109×10^{-4}	*TFDP1*	1.387714487	1.271×10^{-1}
TTK	3.664301687	1.292×10^{-3}	*CCNB2*	1.33876417	1.293×10^{-1}
KIF22	2.40064932	2.060×10^{-3}	*RPA3*	0.942772863	2.137×10^{-1}
CENPA	4.606155804	2.292×10^{-3}	*KIF23*	1.253952495	2.247×10^{-1}
MKI67	2.33120824	2.360×10^{-3}	*CDC6*	1.323104442	2.416×10^{-1}
CHEK1	2.61335517	4.228×10^{-3}	*CDC25B*	1.138407673	3.378×10^{-1}
TPX2	2.406900695	5.406×10^{-3}	*PTTG1*	0.637142701	3.651×10^{-1}
RFC3	2.785858262	5.984×10^{-3}	*MAD2L1*	0.618252654	4.851×10^{-1}
TMPO	1.964265986	8.086×10^{-3}	*WEE1*	0.402150684	5.184×10^{-1}
PCNA	2.235343755	8.233×10^{-3}	*E2F5*	0.530712699	6.110×10^{-1}
UBE2C	2.331979017	9.358×10^{-3}	*RAD17*	-0.077271021	9.172×10^{-1}
CDKN2C	3.765539392	1.057×10^{-2}	*GADD45B*	-0.060809059	9.384×10^{-1}
ZW10	5.01976692	1.350×10^{-2}	*RABGAP1*	0.023630566	9.796×10^{-1}

Supplemental Table S2. Differential gene expression summary table for B cell proliferation signature genes between group 1 and group 2 B cells.

Gene Symbol	Log_2 Fold Change (subpopulation 1/subpopulation 2)	P-Value
LHX1	-1.910400379	1.407×10^{-17}
HMGA2	-1.390385557	2.002×10^{-9}
GATA6	-1.022383716	3.526×10^{-5}
RTF1	-0.903637529	1.982×10^{-4}
SMAD4	-1.160681328	3.516×10^{-4}
DUSP4	-1.048825993	3.578×10^{-4}
EOMES	-0.718496251	4.836×10^{-4}
SOX17	-1.308637392	3.216×10^{-3}
GATA4	-0.818421003	4.369×10^{-3}
MIXL1	-0.858496684	6.003×10^{-3}
BPTF	-0.466703334	2.061×10^{-2}
NCKAP1	-0.578379429	2.768×10^{-2}
TGFB1	-1.128799665	4.095×10^{-2}

Supplemental Table S3. Significantly upregulated genes in subpopulation 2 that are annotated with the GO term "endoderm development."

Data	Sequencing Protocol	# of Single cells	GEO accession code
PBMC1	Smart-seq2	384	GSE132044
Breast Cancer	Smart-seq2	563	GSE75688
Definitive Endoderm	Smart-seq2	758	GSE75748
Lung Adenocarcinoma	10x Chromium	902	GSE118767
PBMC2	10x Chromium	3362	GSE132044

Supplemental Table S4. Summary of scRNA-seq data used in this manuscript.