Cell Reports Methods, Volume *3*

Supplemental information

SEQUIN is an R/Shiny framework for rapid and reproducible analysis of RNA-seq data

Claire Weber, Marissa B. Hirst, Ben Ernest, Nicholas J. Schaub, Kelli M. Wilson, Ke Wang, Hannah M. Baskir, Pei-Hsuan Chu, Carlos A. Tristan, and Ilyas Singeç

Supplemental figures and tables

Figure S1: Overview of scRNA metadata analysis

Figure S2: UMAP dimensionality reduction plots and boxplots of gene expression by cluster

Figure S3: tSNE plot by original identity

Figure S4: scRNA tab used to group cells by gene expression

Figure S5: Merging scRNA clusters

Table S1: In-app benchmarking

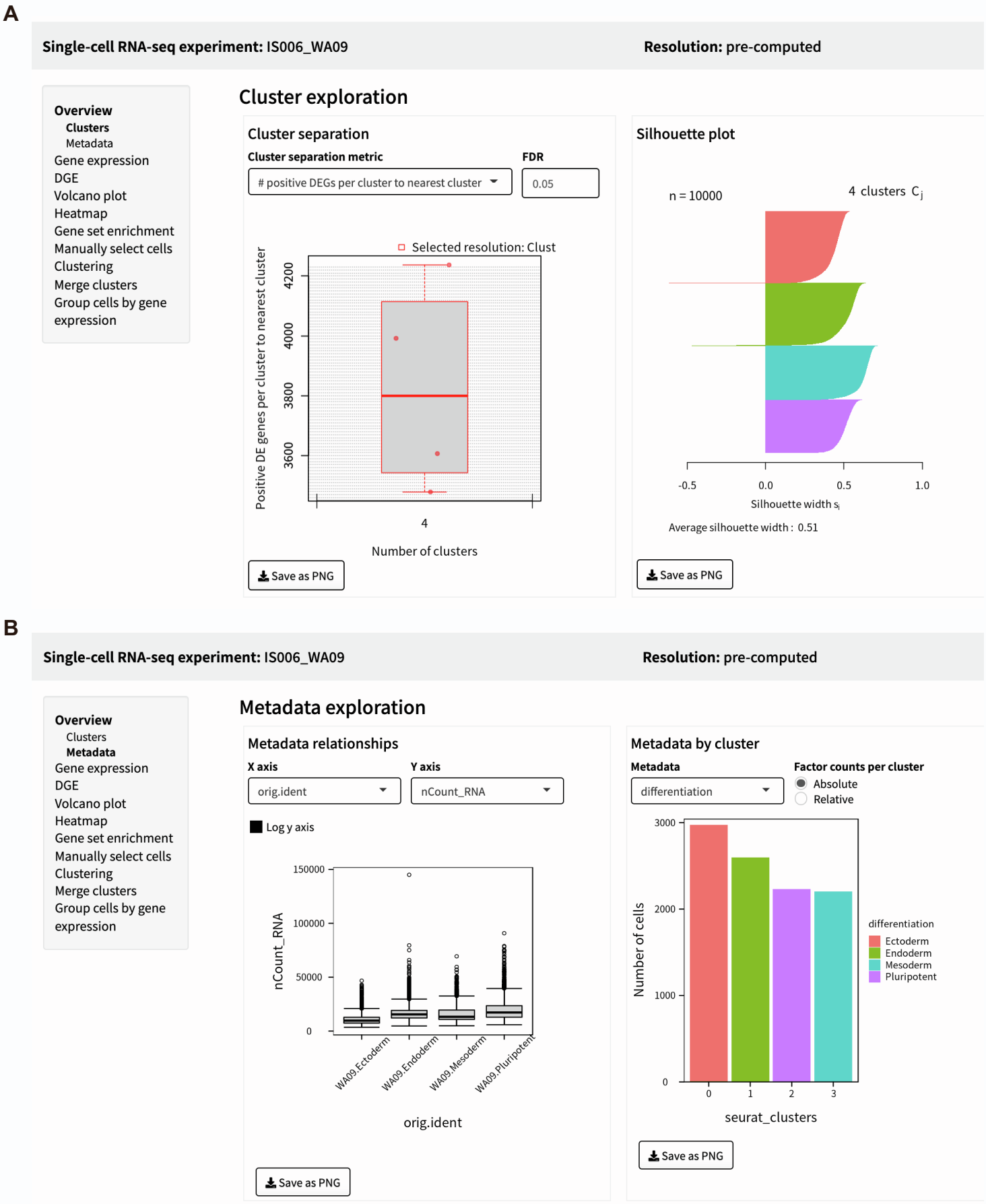

Figure S1. Overview of scRNA metadata analysis, related to figure 2.
A, Exploration of cluster sizes and quality for single cell RNA-Seq analysis of IS006 (WA09). **B**, Identification of relationships between factors in the metadata.

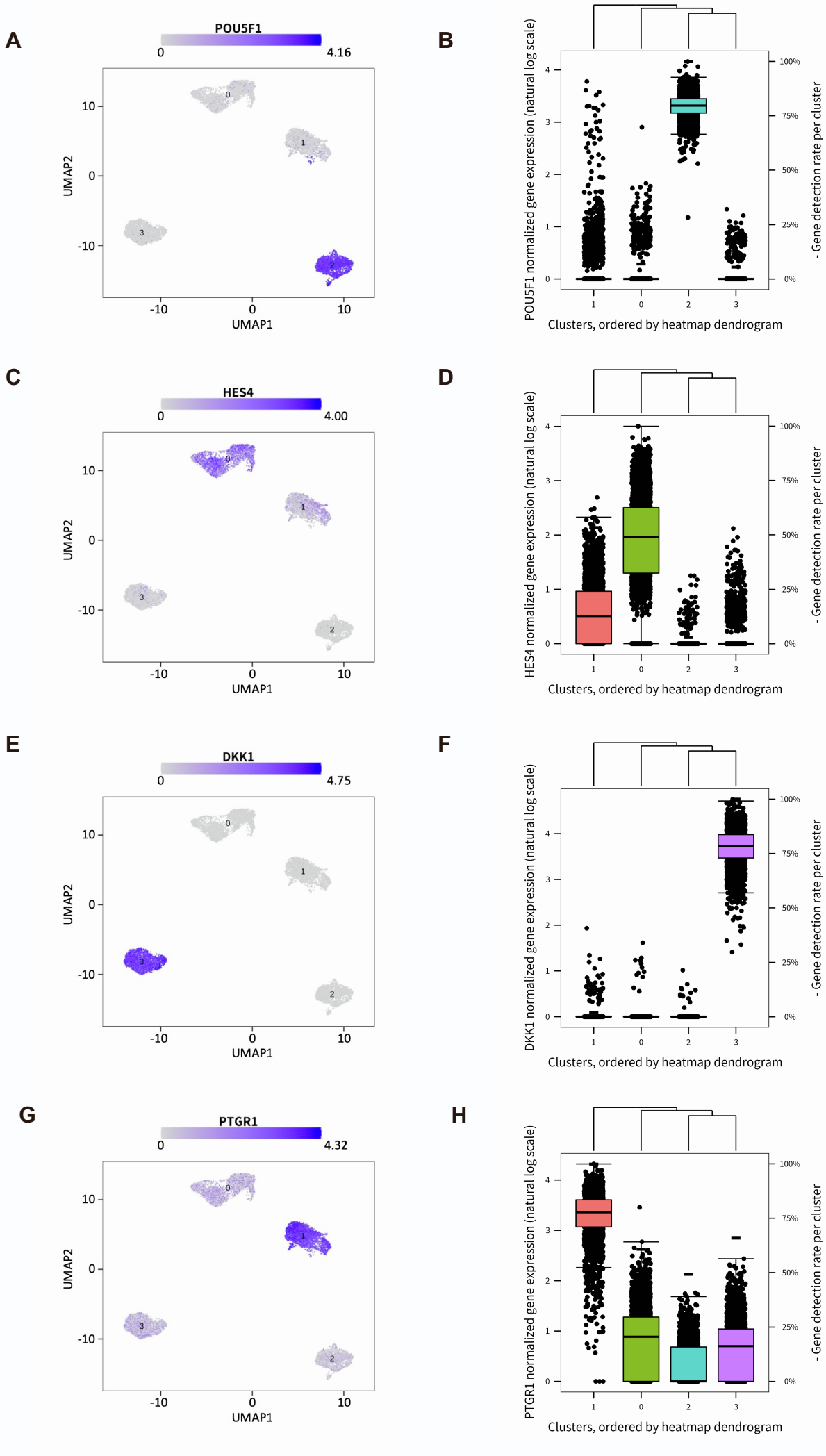
A
POU5F1
0
4.16
UMAP2
UMAP1
B
POU5F1 normalized gene expression (natural log scale)
- Gene detection rate per cluster
Clusters, ordered by heatmap dendrogram
C
HES4
0
4.00
UMAP2
UMAP1
D
HES4 normalized gene expression (natural log scale)
- Gene detection rate per cluster
Clusters, ordered by heatmap dendrogram
E
DKK1
0
4.75
UMAP2
UMAP1
F
DKK1 normalized gene expression (natural log scale)
- Gene detection rate per cluster
Clusters, ordered by heatmap dendrogram
G
PTGR1
0
4.32
UMAP2
UMAP1
H
PTGR1 normalized gene expression (natural log scale)
- Gene detection rate per cluster
Clusters, ordered by heatmap dendrogram

Figure S2. UMAP dimensionality reduction plots and boxplots of gene expression by cluster, related to figure 3.
A, **B**, Pluripotent cells (cluster 2) express POU5F1 (OCT4) in IS006 (WA09) single cell RNA-Seq analysis. **C**, **D**, HES4 is strongly expressed by ectodermal cells (cluster 0). **E**, **F**, Mesodermal cells express DKK1 (cluster 3). **G**, **H**, Expression of PTGR1 by endodermal cells (cluster 1).

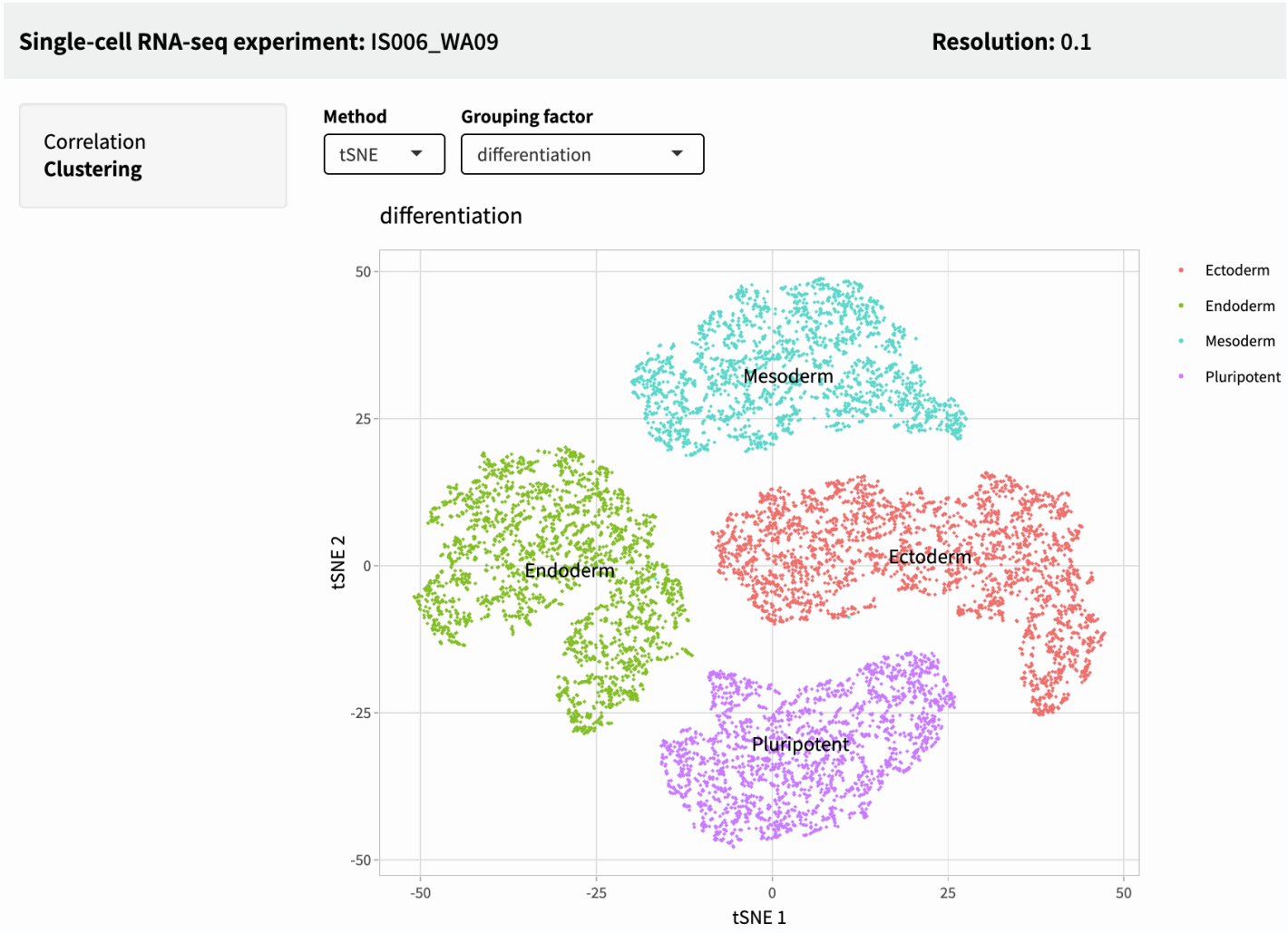

Figure S3. tSNE plot by original identity, related to figure 3.
A resolution of 0.1 reveals four distinct cell clusters with pluripotent, ectodermal, and endodermal, mesodermal signatures in IS006 (WA09) single cell RNA-Seq analysis.

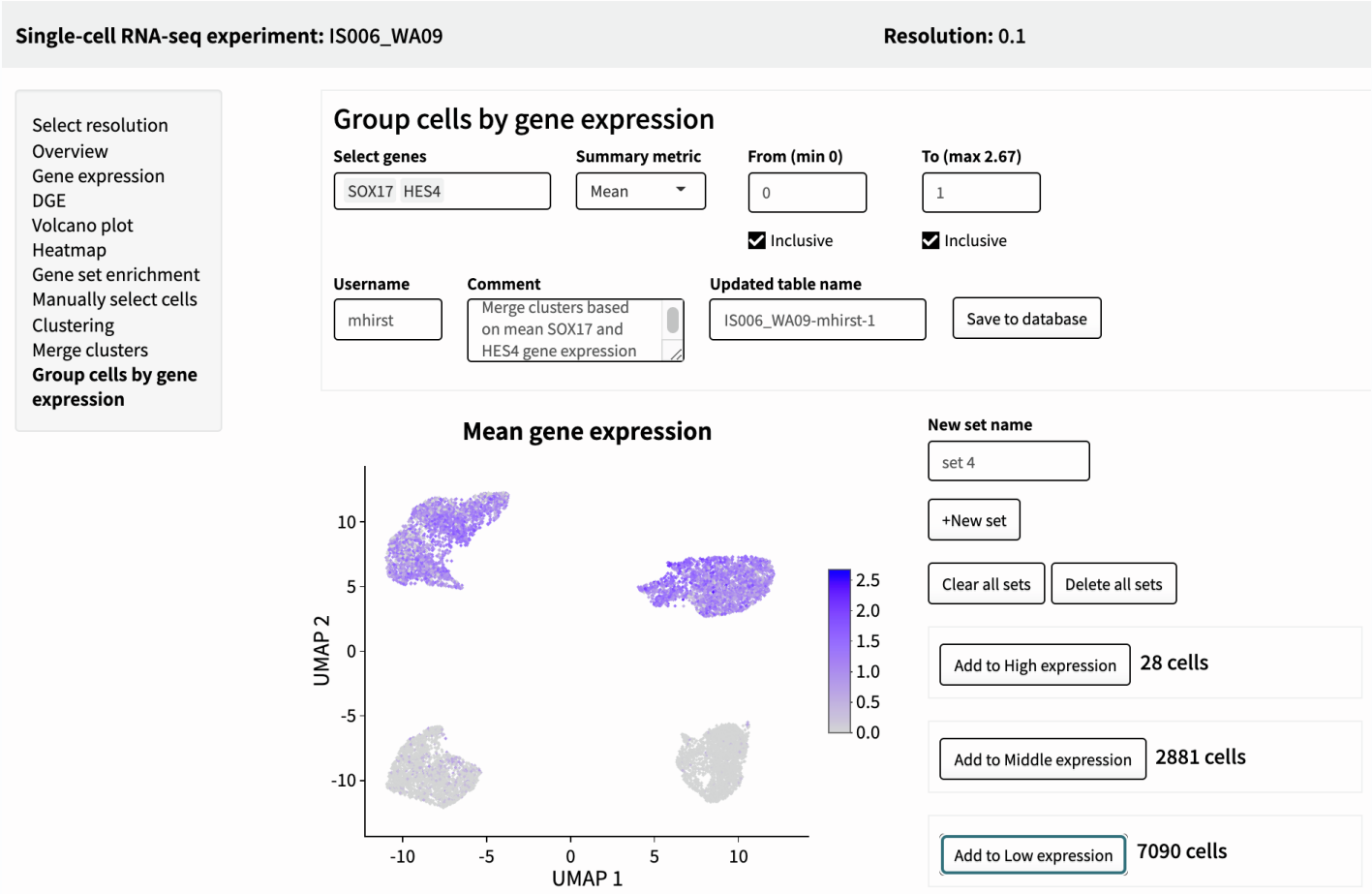

Figure S4. scRNA tab used to group cells by gene expression, related to figures 2 and 4. The mean of two genes per cell was selected here (SOX17 and HES4) and overlaid onto a UMAP of IS006 (WA09). The user can either lasso cells to include them in a given set or define the minimum and maximum of gene expression for the targeted genes to select those cells for a given set. After providing a username and/or description of the type of grouping, the user selects “Save to database” to automatically store the updated metadata in the RDS to be used for updated analysis.

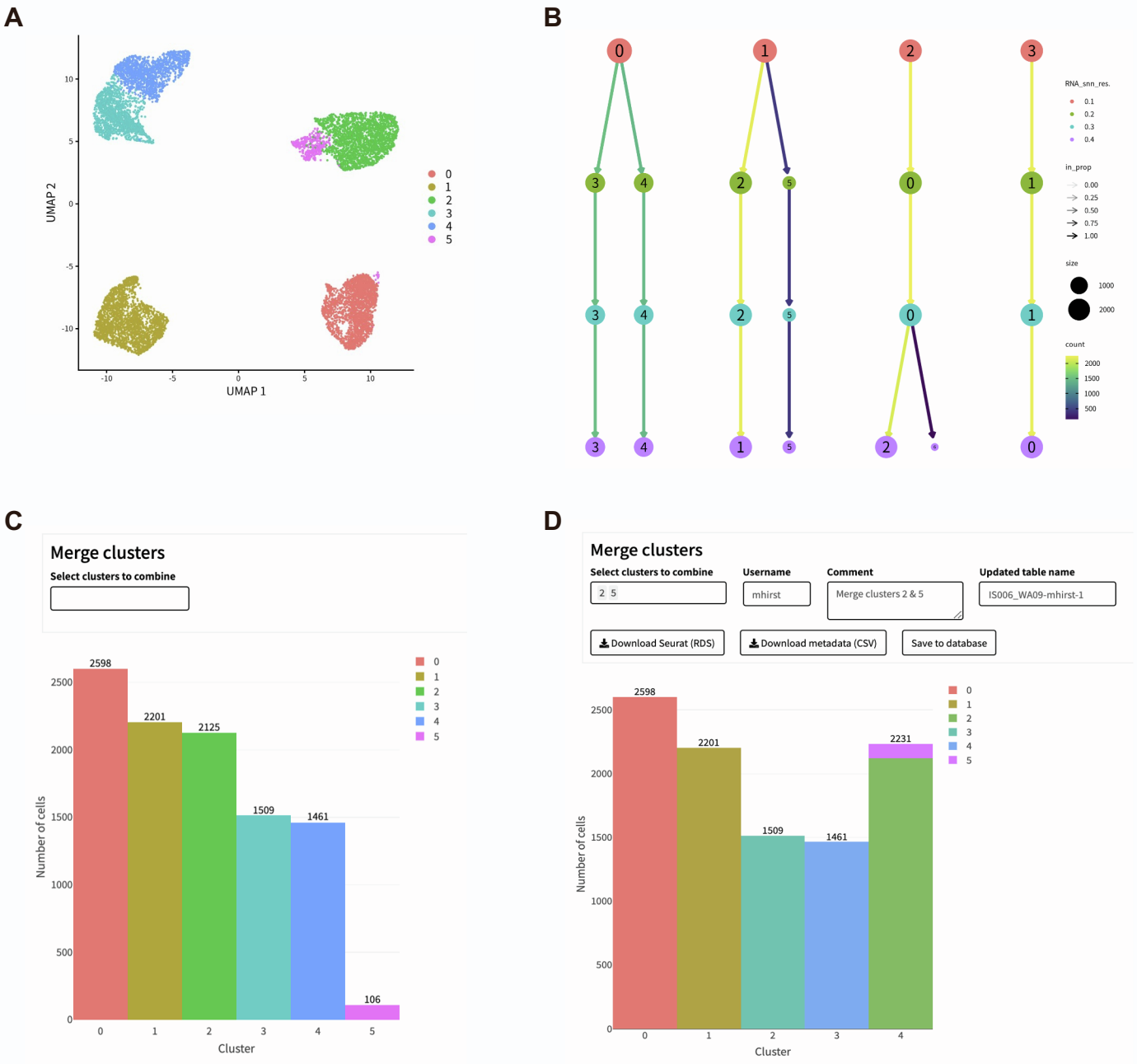

Figure S5. Merging scRNA clusters, related to figures 2 and 4.
A, UMAP of further clustering at 0.3 resolution reveals six clusters, splitting endoderm and ectoderm clusters into two in IS006 (W09). **B**, Clustree flowchart by resolutions 0.1 to 0.4 in rows showing how higher resolutions generate more clusters. **C**, Prior to selecting clusters for merging in IS006 (WA09). **D**, The merging of clusters three and four from IS006 (WA09). After providing a username and/or a comment, the user can select the “Save to database”, which will automatically store the updated metadata in the RDS to be used for updated analysis.

	Dataset name					
Step	**example_sc**	**IS018**	**IS020_WA09**	**IS006_WA09**	**IS006_11**	**Comment**
N cells	90	3875	10058	16582	19759	
Load data	8 s	3 min 50 s	7 min 30 s	8 min 30 s	29 min	1
Load data - resolutions 0.4 to 2.8 (step by 0.4)	13 s	8 min 33 s	24 min 18 s	56 min 46 s	1 hr 30 min	
QC - Box and whisker	1 s	19 s	44 s	1 min 12 s	1 min 29 s	
QC - Histogram	1 s	16 s	39 s	1 min 14 s	1 min 28 s	
QC - Total Reads	1 s	2 s	5 s	9 s	12 s	
Correlation matrix	1 s	4 s	5 s	8 s	10 s	2
Dimensionality reduction plots	1 s	1 s	3 s	4 s	5 s	
Overview	1 s	1 s	1 s	1 s	1 s	3
Gene expression	1 s	3 s	4 s	4 s	5 s	4
DGE by cluster	1 s	1 s	1 s	1 s	1 s	
Custom DGE	2 s	16 s	6 s	3 min 21 s	3 min 45 s	5
Volcano plot	1 s	1 s	1 s	1 s	1 s	
Heatmap	2 s	45 s	1 min 8 s	1 min 12 s	1 min 17 s	
GSE	1 s	1 s	1 s	1 s	1 s	
WGCNA	4 s	9 s	17 s	26 s	27 s	
iPSC Profiler	1 s	7 s	26 s	1 min 3 s	1 min 10 s	

Table S1. In-app benchmarking, related to figure 2.
Discrete analysis steps in SEQUIN were time-stamped with five scRNA datasets of increasing size. Comments: 1. Pre-computed clusters; 2. Default 2000 genes and 500 cells; 3. Clusters and metadata similar times; 4. By cluster and cell distribution similar times; 5. First two-group comparison option: IS020_WA09 had 105 DE genes, IS006_WA09 had 2987, and IS006_11 had 2993.