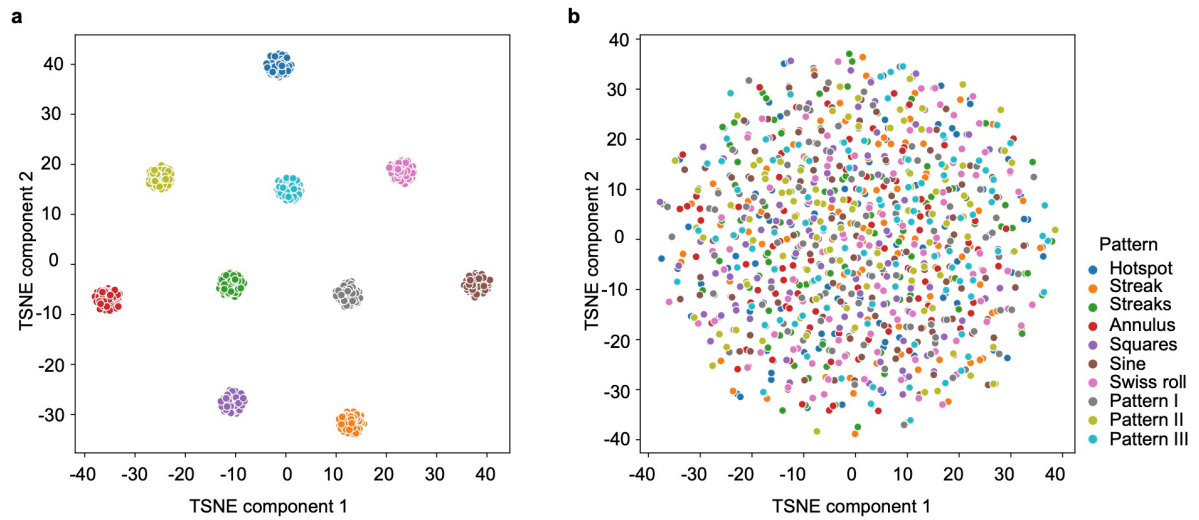




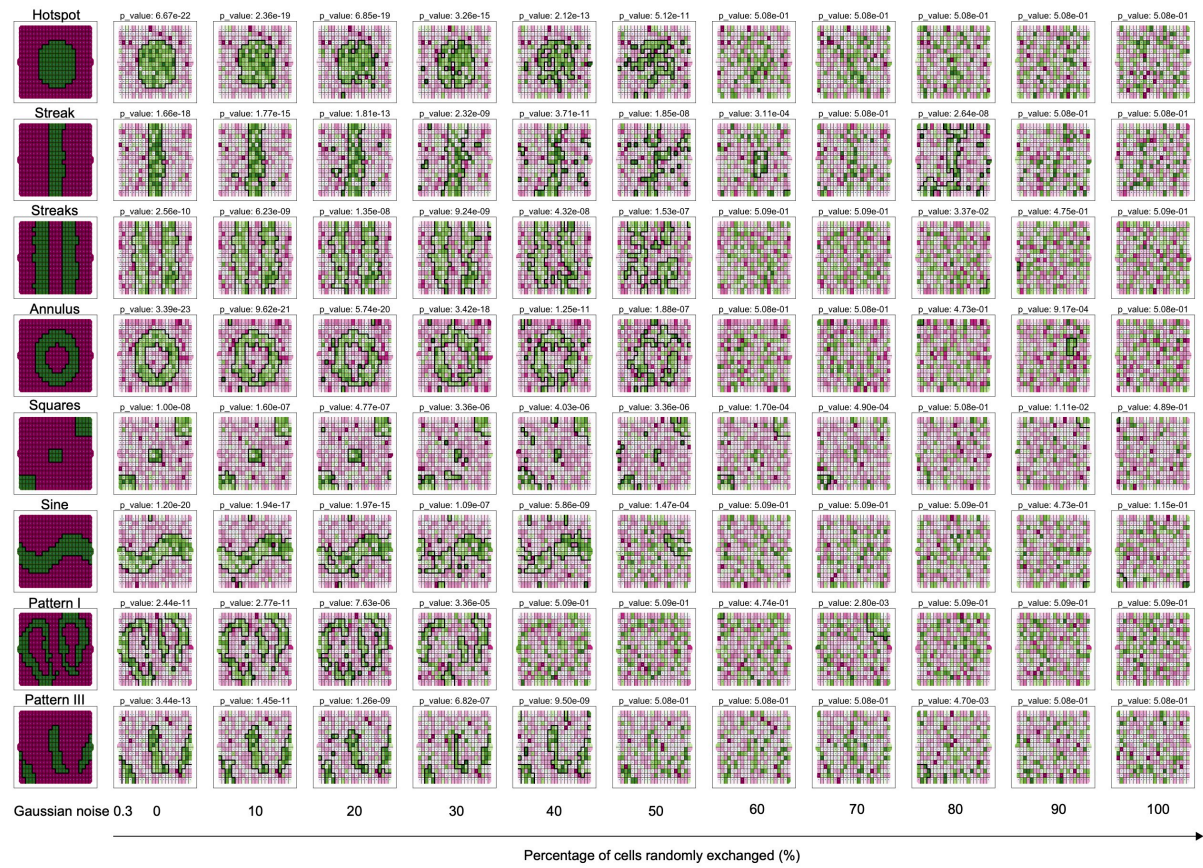
Supplementary Figure 1. T-SNE analysis of simulated data.

(a) t-SNE projections showing that the simulated SV genes formed 10 well-separated clusters. (b) t-SNE projections showing that no clusters were formed among random genes whose gene expression values were randomly shuffled.



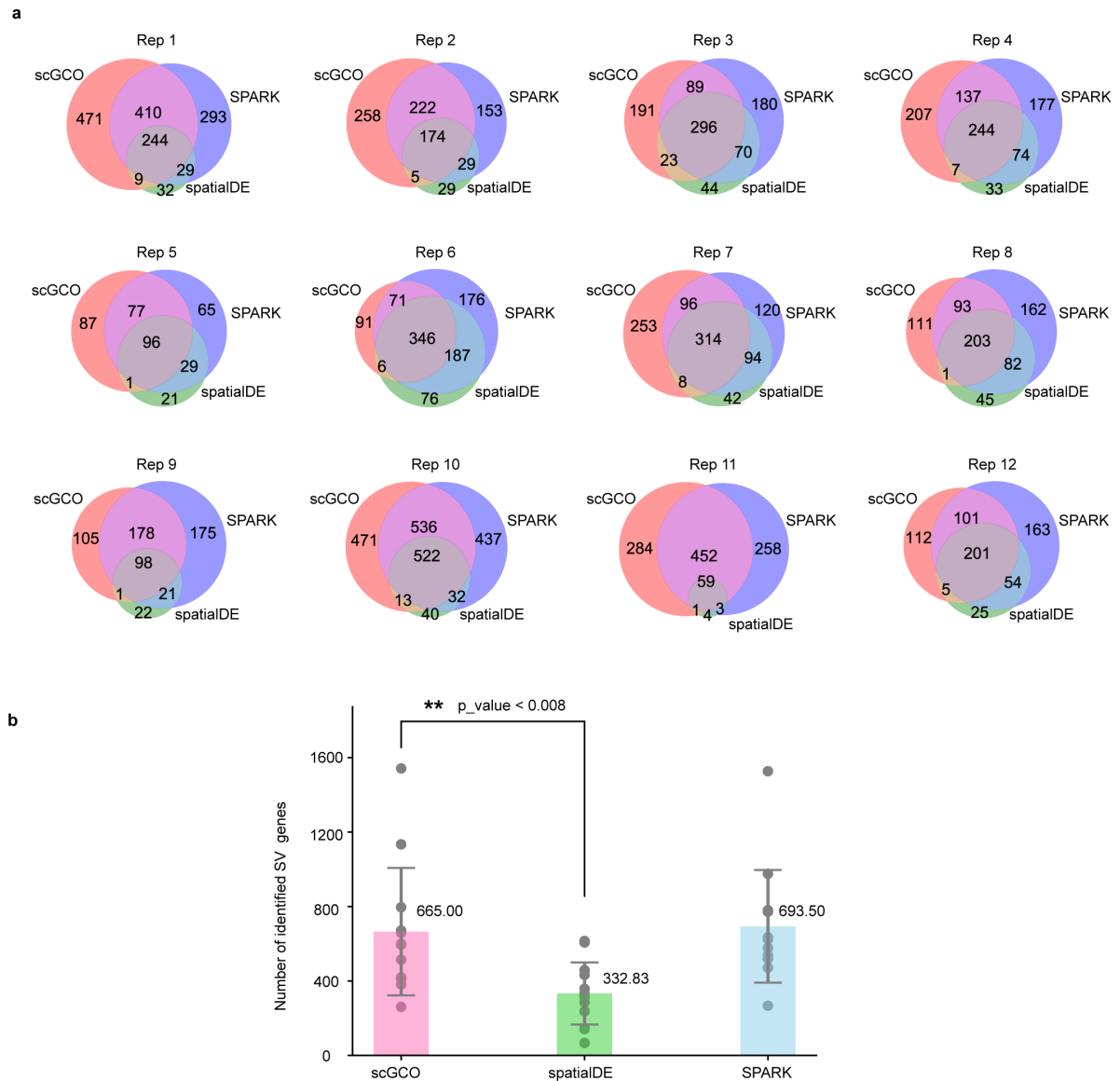
Supplementary Figure 2. Performance of SV gene identification algorithms using simulated data with Gaussian noises.

Representative graph cuts results at increasing Gaussian noise levels for scGCO.



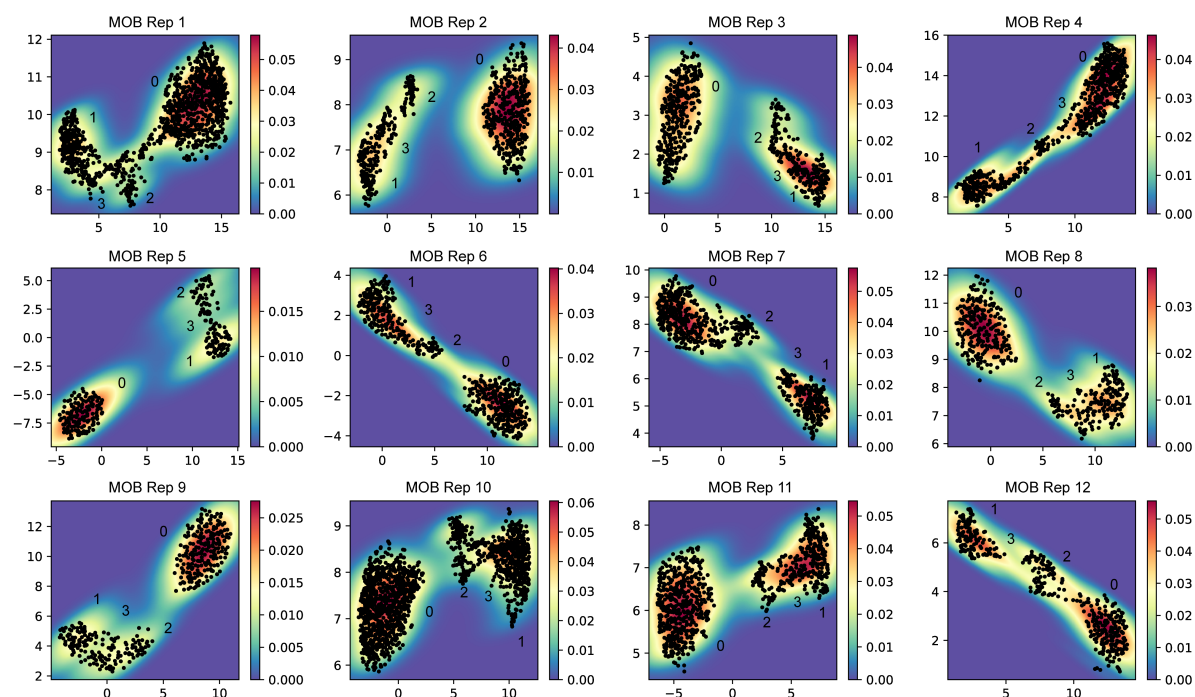
Supplementary Figure 3. Performance of SV gene identification algorithms using simulated data with random exchanges.

Representative graph cuts results at an increasing percentage of randomly exchanged cells for scGCO.



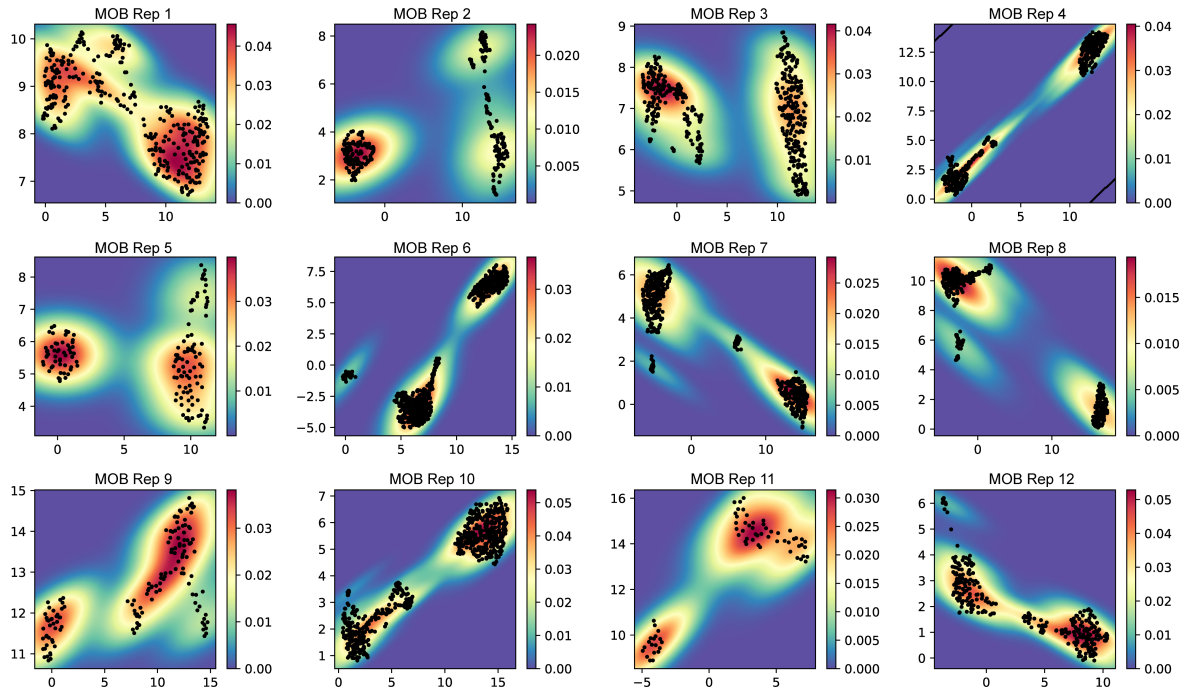
Supplementary Figure 4. ScGCO identifies SV genes from mouse olfactory bulb data.

(a) Venn diagrams showing the set relationship among SV genes identified by scGCO (FDR <0.05), spatialDE (FDR <0.05), and SPARK (FDR <0.05) in mouse olfactory bulb data. **(b)** Bar plots showing the number of identified SV genes from twelve biological replicates by different methods. Error bars indicate means $\pm$ SD.



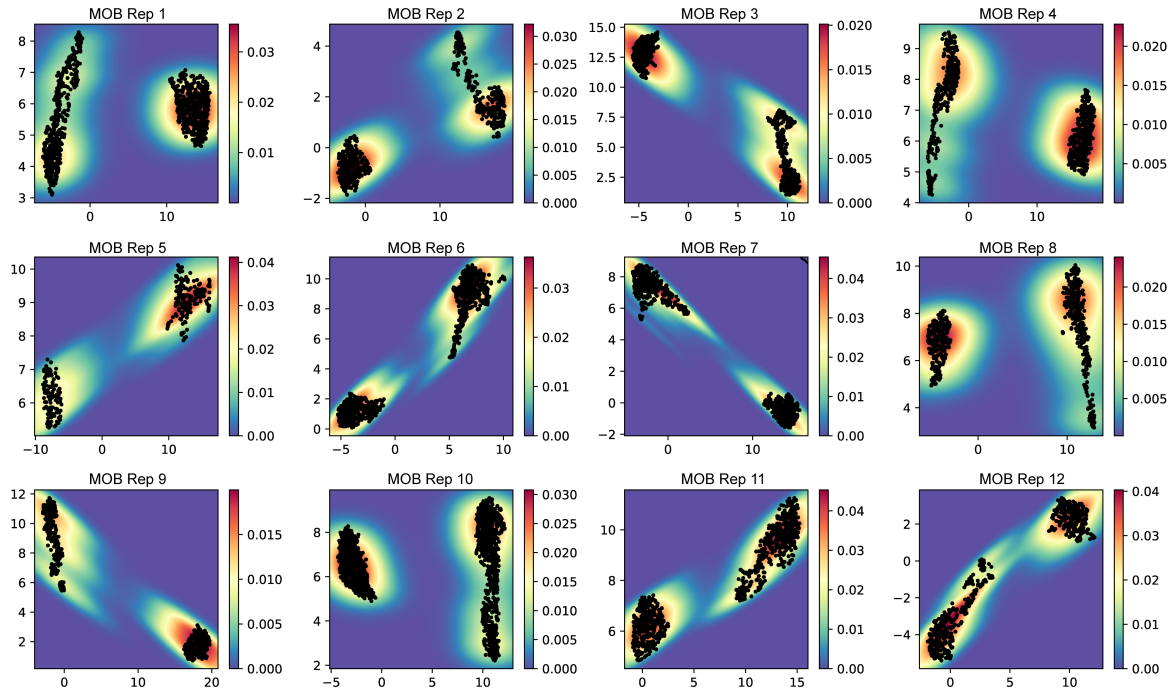
Supplementary Figure 5. Clustering analyses of SV genes identified by scGCO in mouse olfactory bulb data.

UMAP analyses of SV genes identified by scGCO for all 12 replicates of mouse olfactory bulb data. Each point is a gene. Background color indicates density of points determined with kernel density estimation. Numbers indicate identified clusters by HDBSCAN.



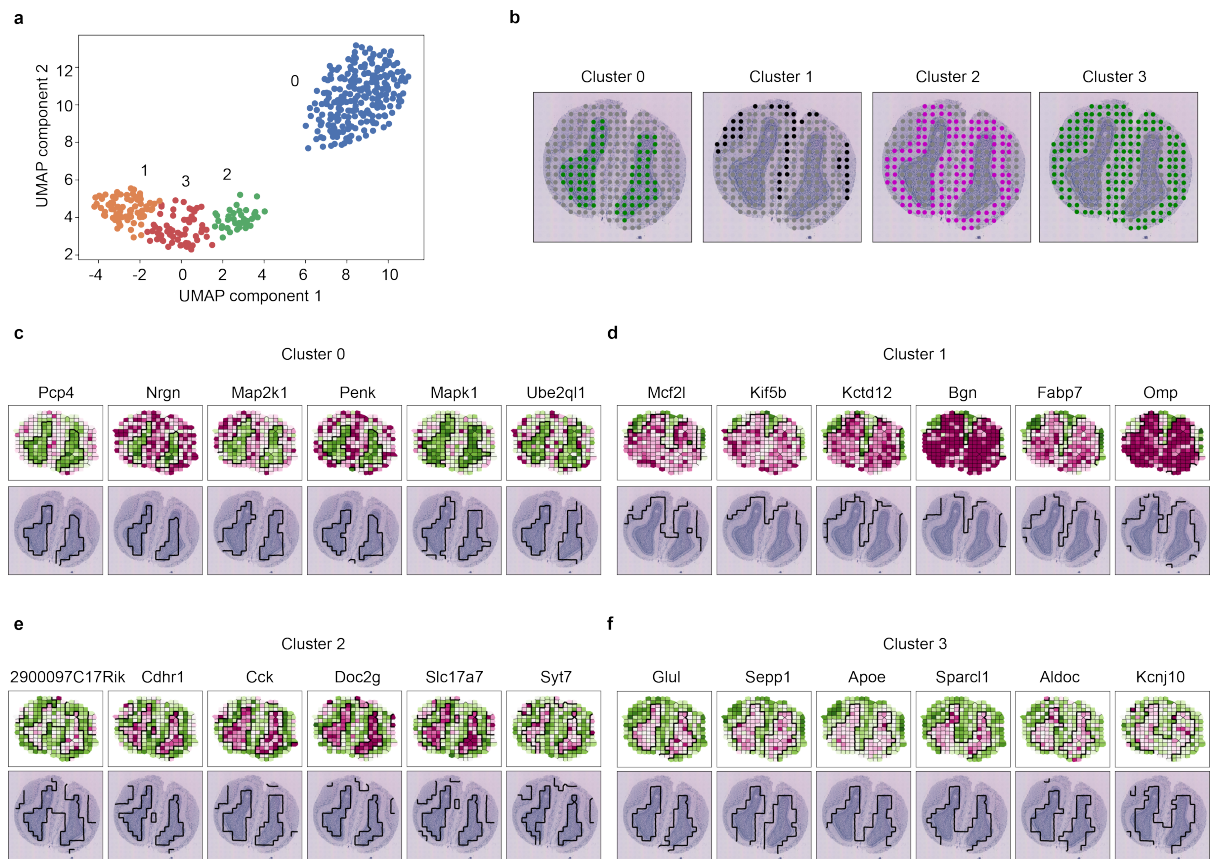
Supplementary Figure 6. Clustering analyses of SV genes identified by spatialDE in mouse olfactory bulb data.

UMAP analyses of SV genes identified by spatialDE for all 12 replicates of mouse olfactory bulb data. Each point is a gene. Background color indicates density of points determined by kernel density estimation.



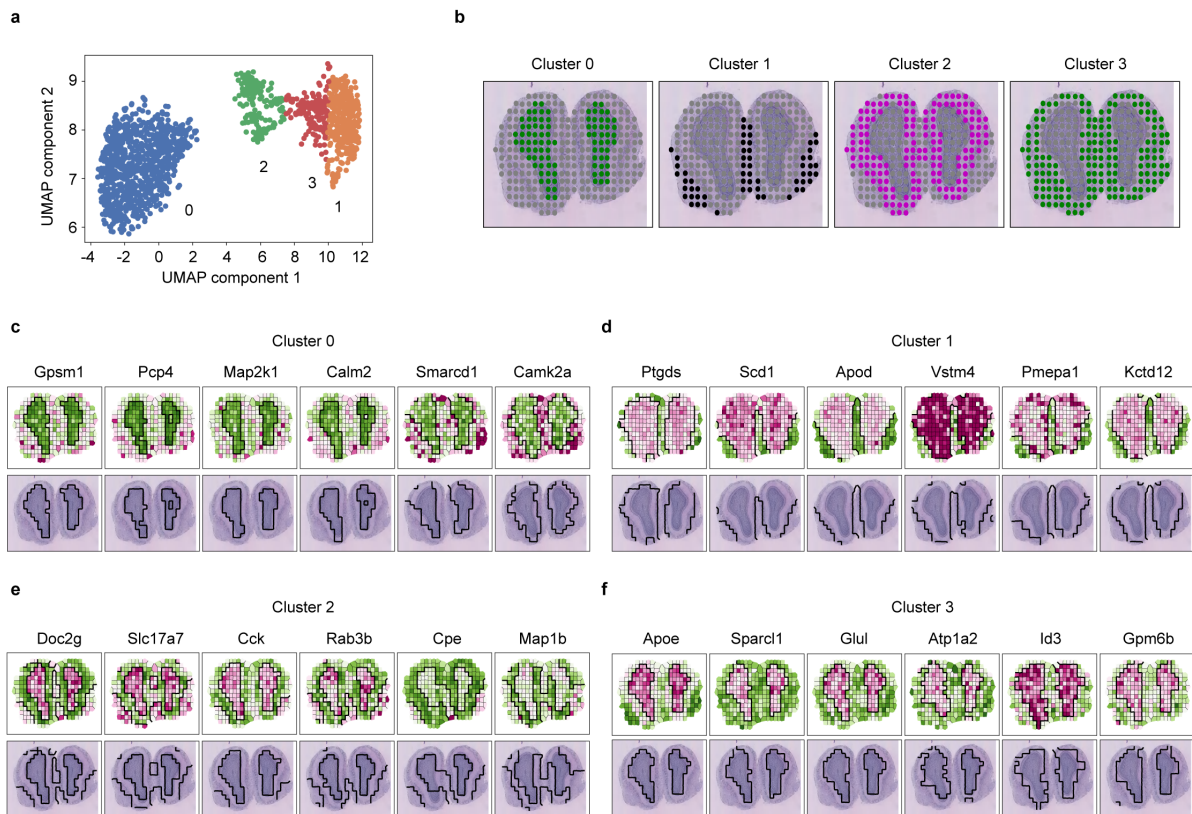
Supplementary Figure 7. Clustering analyses of SV genes identified by SPARK in mouse olfactory bulb data.

UMAP analyses of SV genes identified by SPARK for all 12 replicates of mouse olfactory bulb data. Each point is a gene. Background color indicates density of points determined by kernel density estimation.



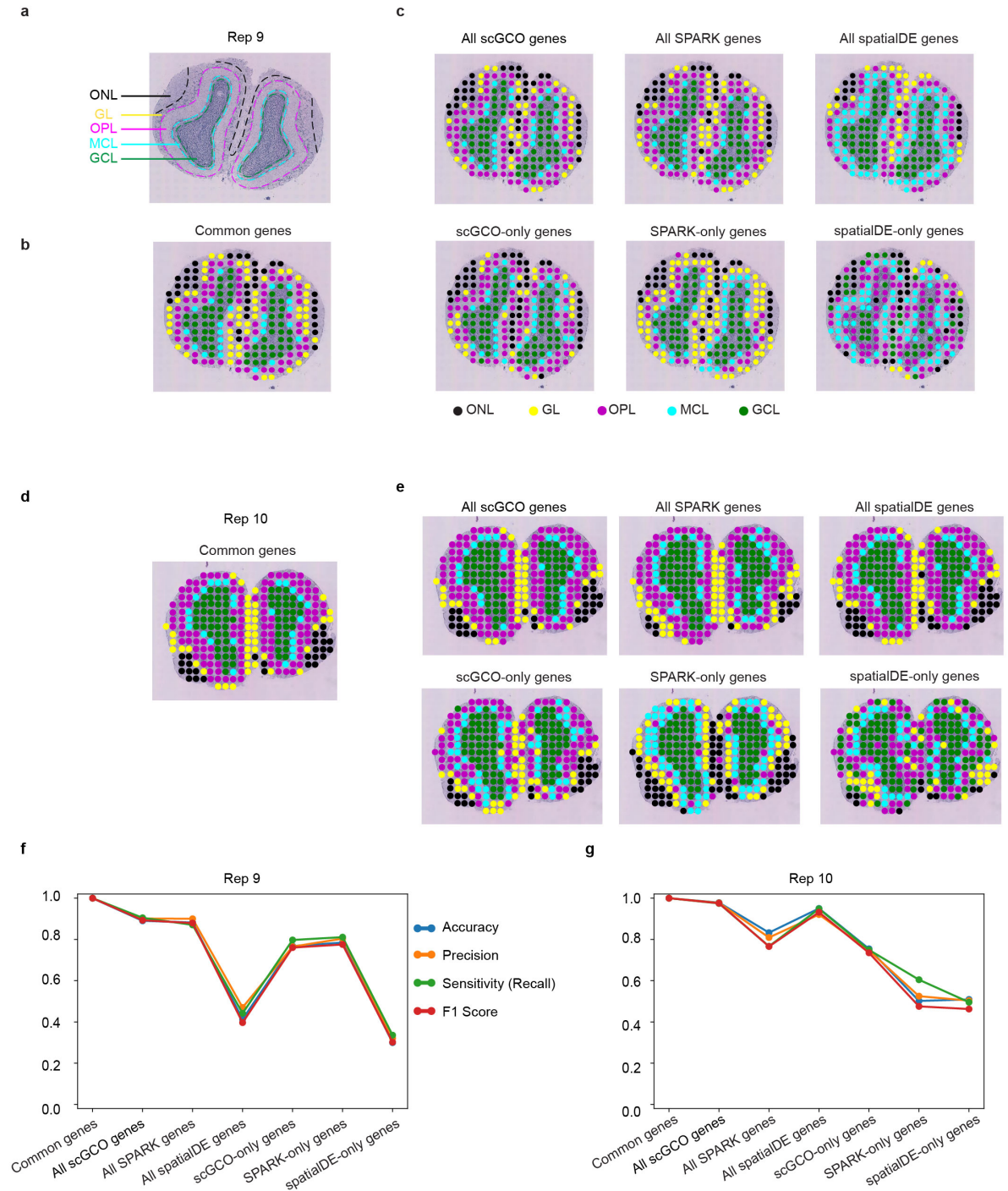
Supplementary Figure 8. Identification of SV genes in replicate 9 of MOB data by scGCO.

(a) UMAP for SV genes identified by scGCO in replicate 9 of MOB data. **(b)** Tissue structures resolved with each cluster of SV genes identified by scGCO for replicate 9. **(c)** Representative Voronoi diagrams showing graph cuts results (top panels), and associated H&E staining overlays (lower panels) for cluster 0. **(d-f)** Same as **(c)** for cluster 1, 2 and 3 respectively. Six examples were shown for each cluster. Thicker black lines denote graph cuts boundaries.



Supplementary Figure 9. Identification of SV genes in replicate 10 of MOB data by scGCO.

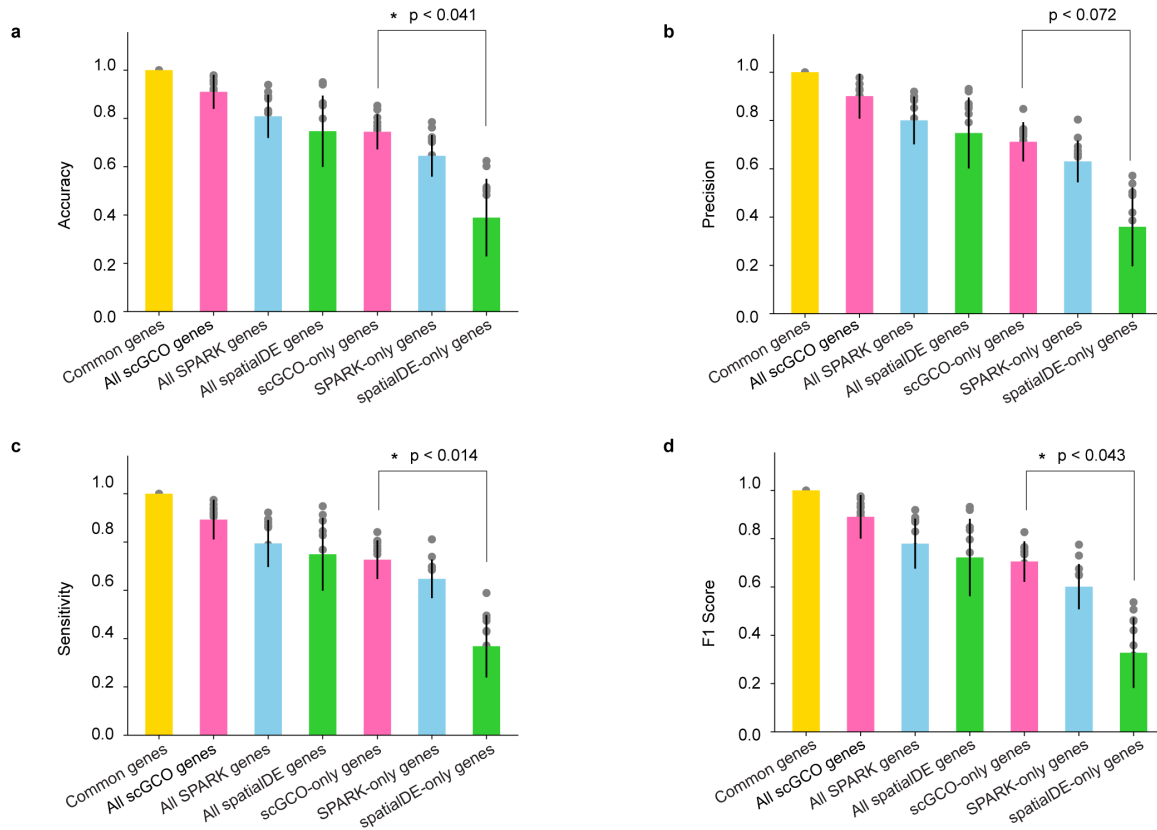
(a) UMAP for SV genes identified by scGCO in replicate 10 of MOB data. **(b)** Tissue structures resolved with each cluster of SV genes identified by scGCO for replicate 10. **(c)** Representative Voronoi diagrams showing graph cuts results (top panels), and associated H&E staining overlays (lower panels) for cluster 0. **(d-f)** Same as **(c)** for cluster 1, 2 and 3 respectively. Six examples were shown for each cluster. Thicker black lines denote graph cuts boundaries.



Supplementary Figure 10. SV genes uniquely identified by scGCO resolved MOB tissue structures.

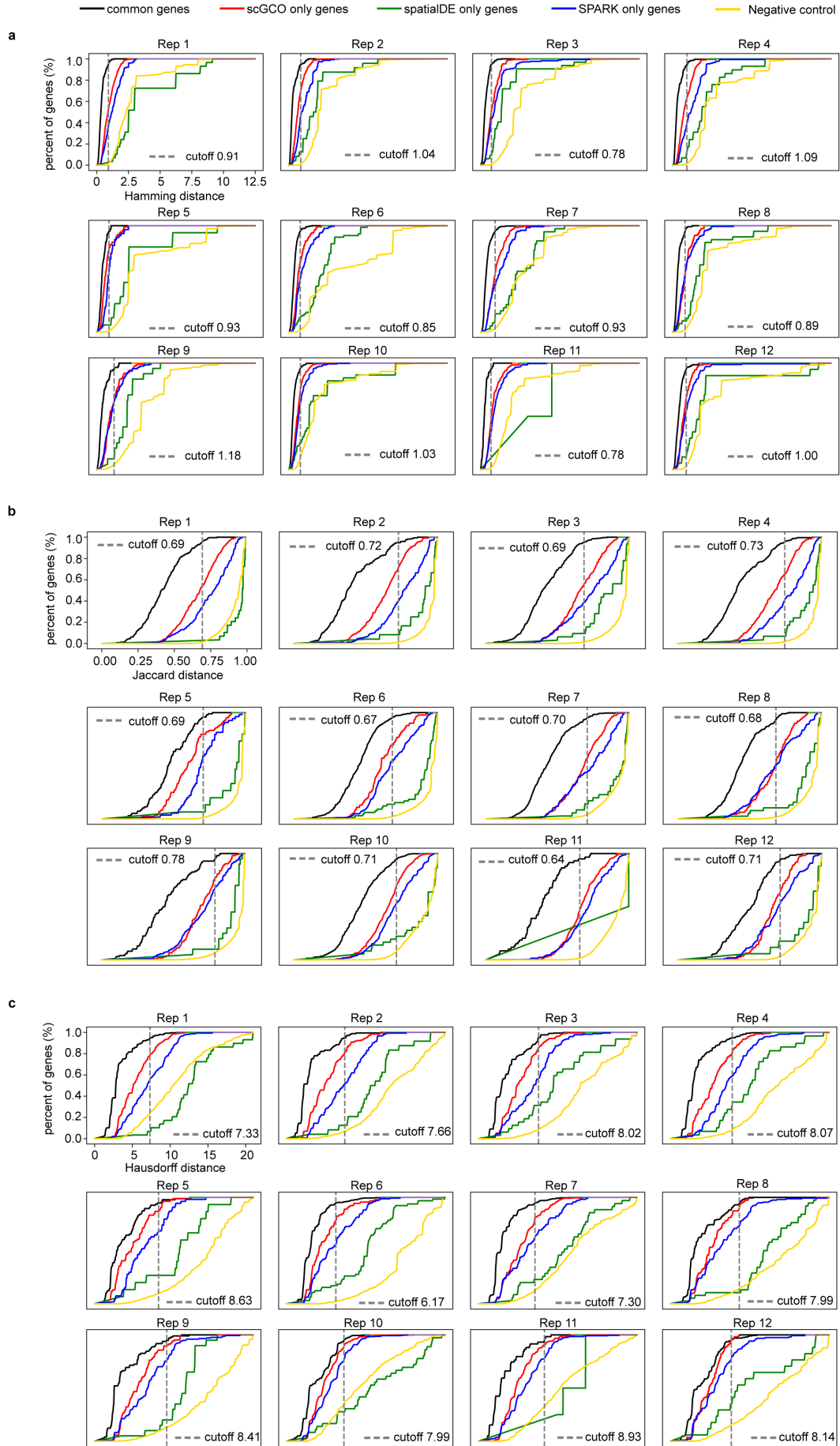
(a) Annotation of the MOB tissue architecture showing the five known layers for replicate 9. ONL: olfactory nerve layer; GL: glomerular layer; OPL: outer plexiform layer; MCL: mitral cell layer; GCL: granular cell layer. **(b)** MOB tissue architecture reconstructed using SV genes jointly identified by the three methods (scGCO, spatialDE and SPARK) for replicate 9.

(c) The MOB tissue structures reconstructed using all scGCO genes (upper column 1), all SPARK genes (upper column 2), all spatialDE genes (upper column 3), scGCO-only genes (lower column 1), SPARK-only genes (lower column 2), and spatialDE-only genes (lower column 3) for replicate 9. **(d)** Same as **(b)** for replicate 10. **(e)** Same as **(c)** for replicate 10. **(f)** Lines plots showing accuracy, precision, sensitivity and F1 score of the reconstructed tissue structures in **(c)** for replicate 9, tissue structure reconstructed with common genes was used as positive reference. **(g)** Same as **(f)** for replicate 10.



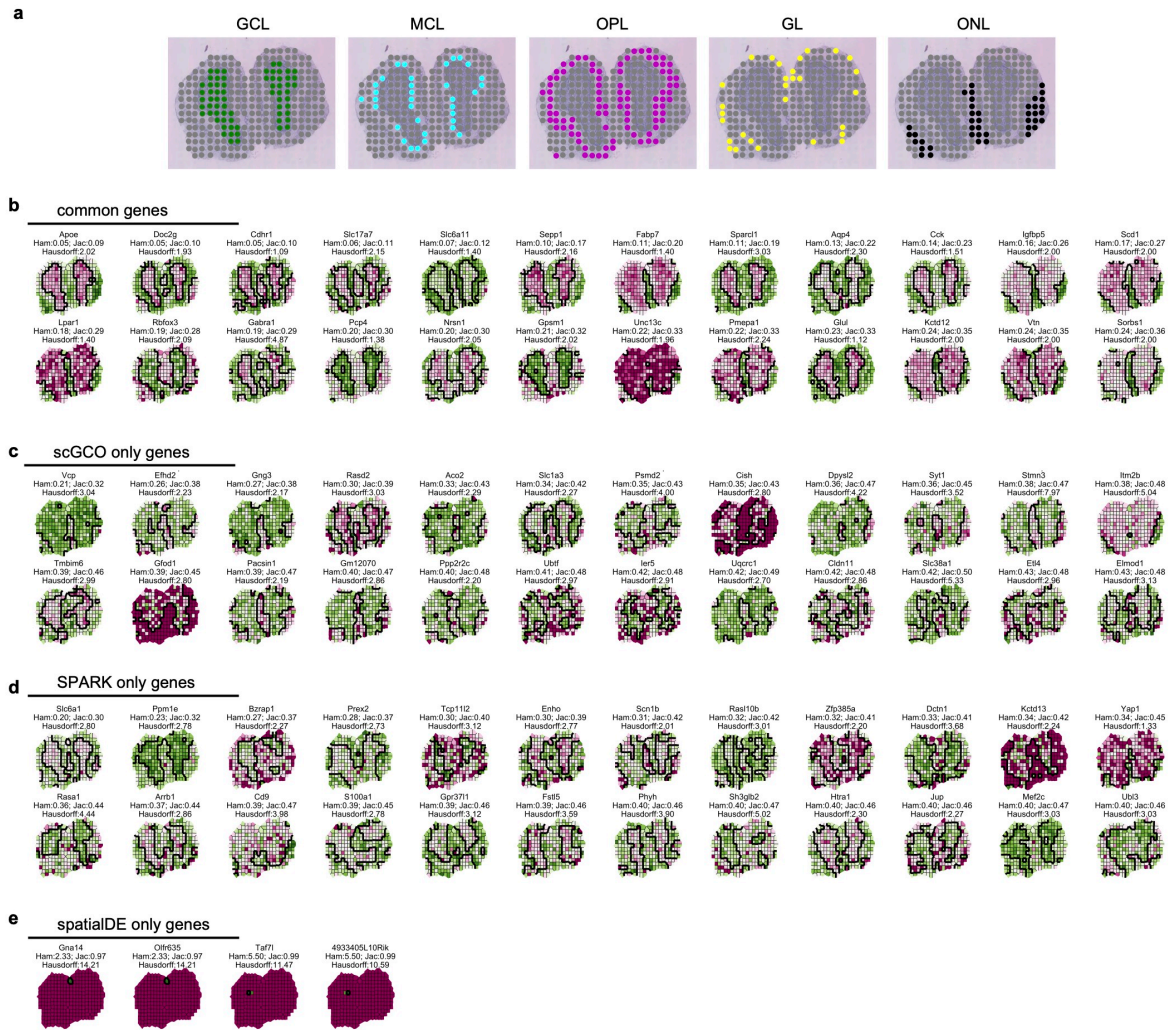
Supplementary Figure 11. Quantifying the performance of identified SV genes to reconstruct MOB tissue structures.

Bar charts showing accuracy (**a**), sensitivity (**b**), precision (**c**) and F1 score (**d**) of reconstructed tissue structures using different sets of SV genes for all 12 replicates. MOB tissue architectures reconstructed using SV genes jointly identified by the three methods were used as positive references. Error bars indicate means $\pm$ SD. The p-values were determined using pairwise t-test.



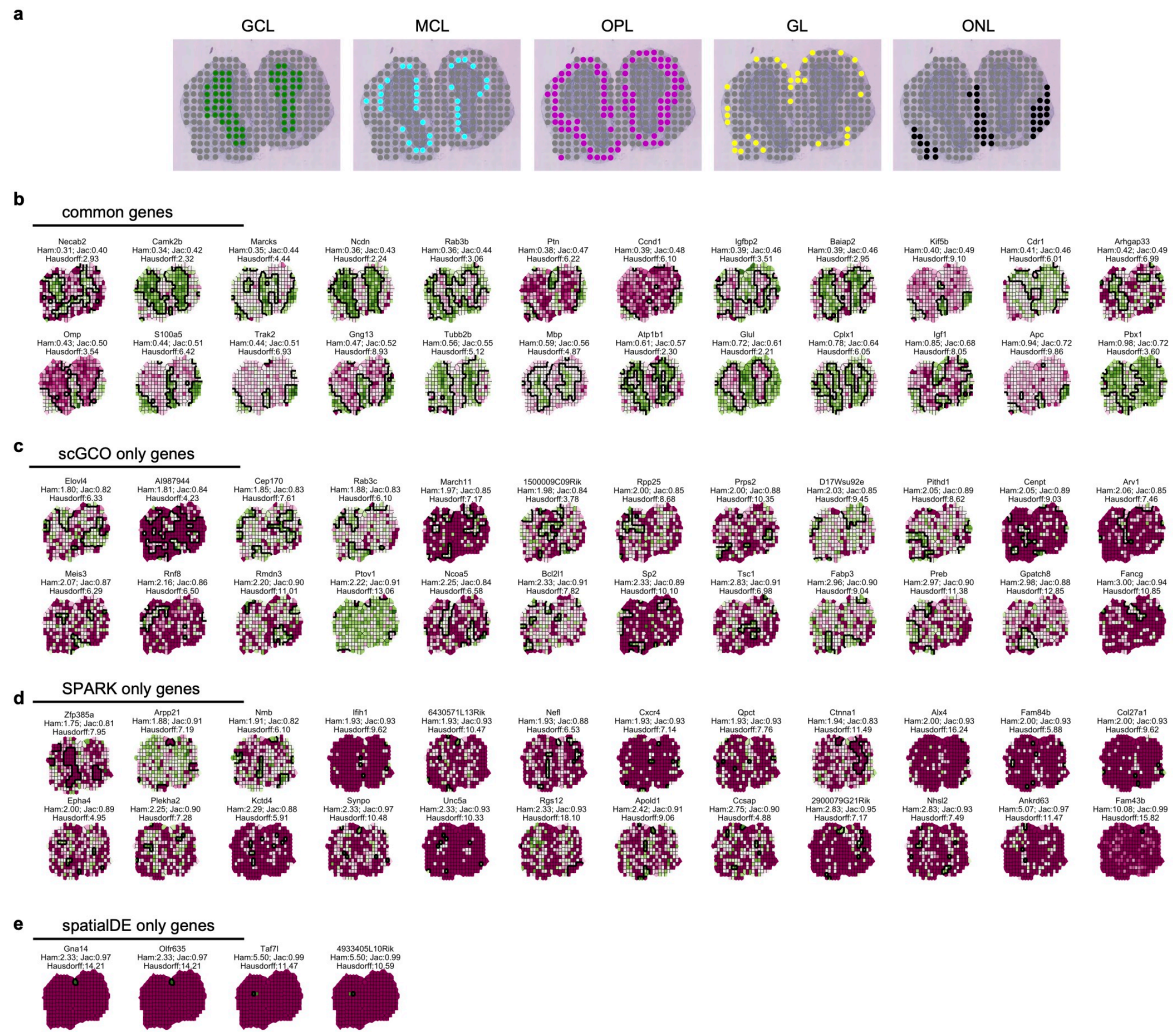
Supplementary Figure 12. Quantifying similarity of identified SV genes to reference tissue structures.

(a) Line plots showing the cumulative distribution of normalized hamming distance of identified SV genes to reference MOB tissue structures reconstructed with common genes for all 12 replicates. Black lines: SV genes jointly identified by all three methods as positive control; red lines: SV genes uniquely identified by scGCO; green lines: SV genes uniquely identified by spatialDE; blue lines: SV genes uniquely identified by SPARK; gold lines: genes with random spatial gene expression patterns as negative control. The dashed vertical lines indicated estimate distance cutoffs corresponding to 95% of genes jointly identified by all three methods. **(b)** Same as **(a)** using Jaccard distance. **(c)** Same as **(a)** using Hausdorff distance.



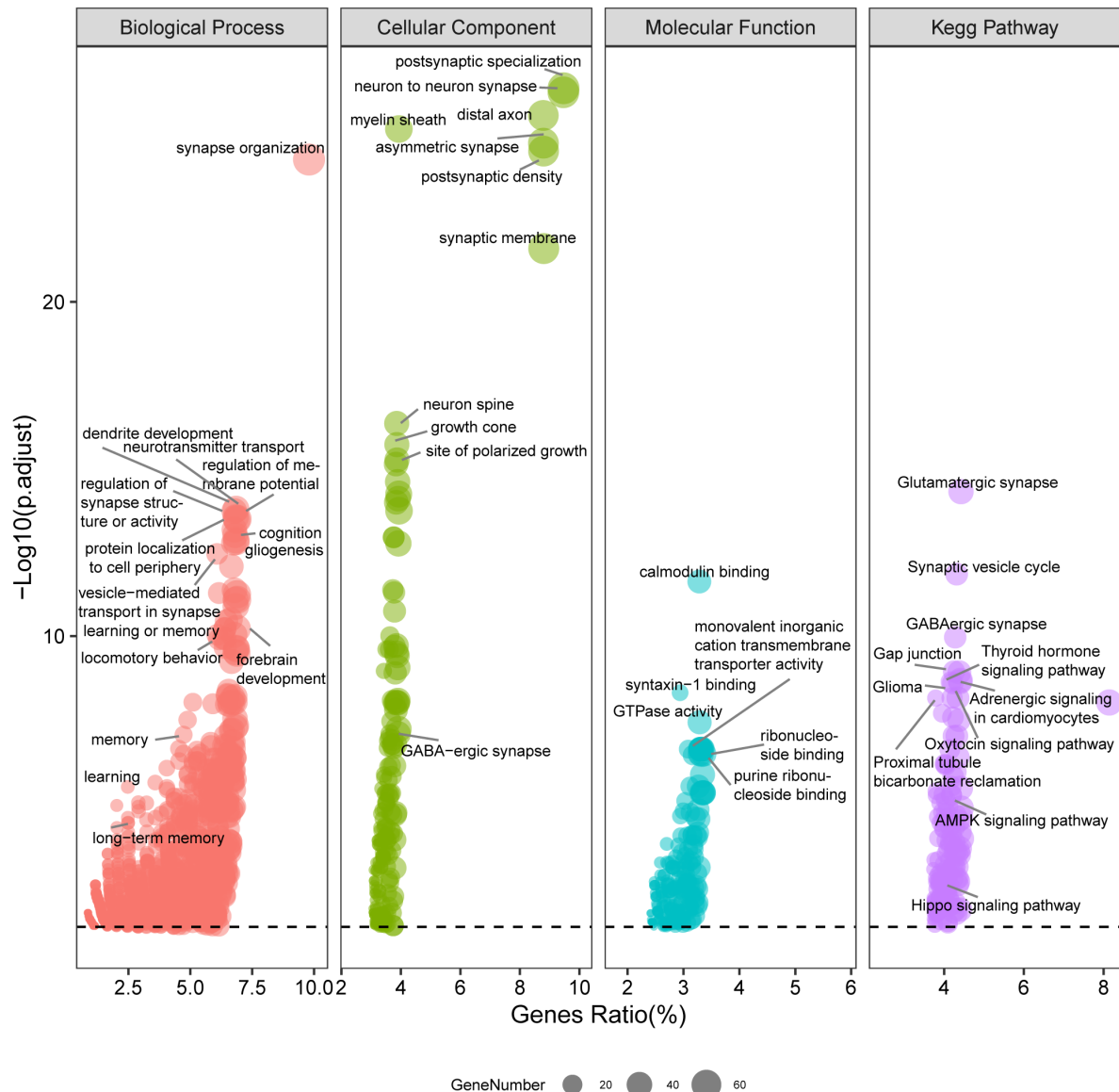
Supplementary Figure 13. Visualizing similarity of identified top ranked SV genes to reference tissue structures.

(a) MOB tissues structures reconstructed using SV genes jointly identified by the three methods as reference for replicate 11. (b) Graph cuts showing the top 24 genes with the smallest hamming distance from SV genes jointly reported by all three methods. (c) Same as (b) for scGCO-only genes. (d) Same as (b) for SPARK-only genes. (e) Same as (b) for spatialDE-only genes (only four genes were uniquely identified by spatialDE).



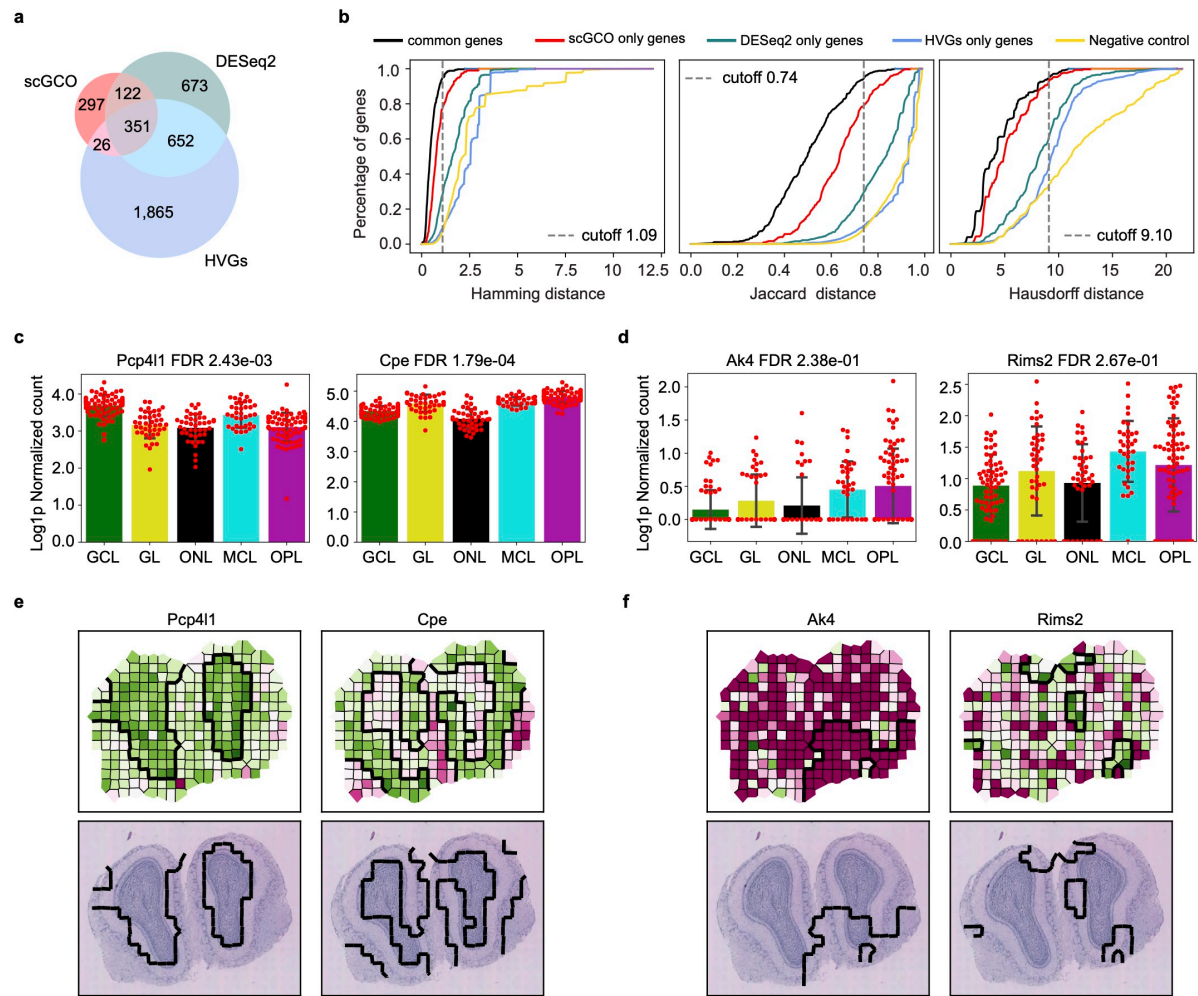
Supplementary Figure 14. Visualizing similarity of bottom ranked SV genes to reference tissue structures.

(a) MOB tissues structures reconstructed using SV genes jointly identified by the three methods as reference for replicate 11. (b) Graph cuts showing the bottom 24 genes with the largest hamming distance from SV genes jointly reported by all three methods. (c) Same as (b) for scGCO-only genes. (d) Same as (b) for SPARK-only genes. (e) Same as (b) for spatialDE-only genes (only four genes were uniquely identified by spatialDE).



Supplementary Figure 15. SV genes identified by scGCO were enriched with biology functions related to neurons.

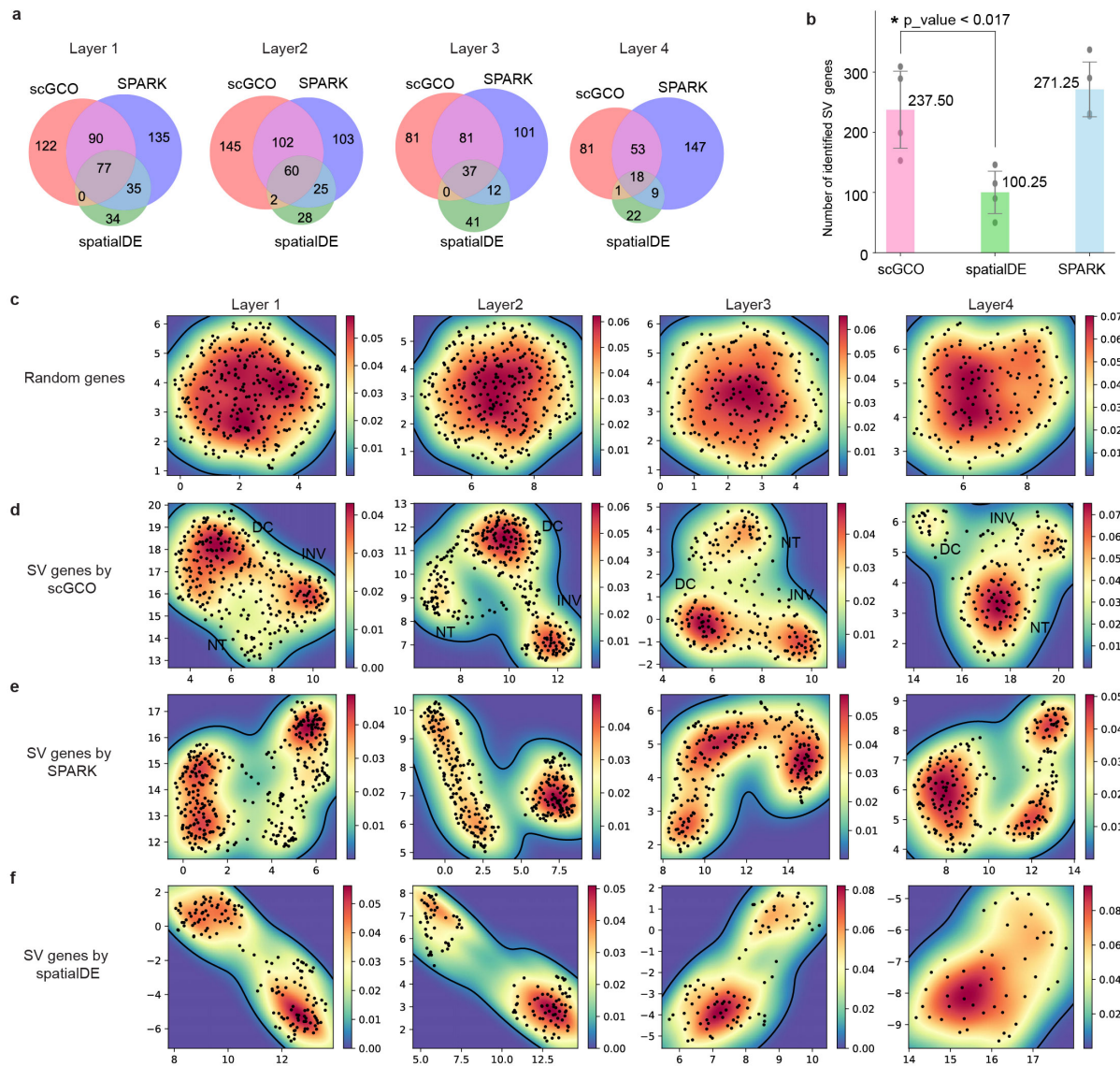
Bubble plots showing significantly enriched gene ontologies and KEGG pathways for SV genes identified by scGCO. The horizontal lines correspond to adjusted p-value of 0.05 with BH method.



Supplementary Figure 16. Comparing genes identified by scGCO and genes identified by DESeq2.

(a) Venn diagram showing the overlap of genes identified by scGCO, HVGs identified with Seurat ignoring spatial context, and genes identified by DESeq2 for replicate 11 of MOB data. DESeq2 genes were identified with $FDR < 0.01$ and absolute fold change > 1 . (b) Line plots showing the cumulative distribution of distance similarity of identified SV genes to reference MOB tissue structure for replicate 11 of MOB data. Common genes are the genes jointly identified by scGCO, DESeq2 and HVGs; red lines: SV genes uniquely identified by scGCO; teal lines: SV genes uniquely identified by DESeq2; cornflower blue lines: SV genes uniquely identified by HVGs; gold lines: genes with random spatial expression patterns as negative control. The vertical dashed lines indicate estimate distance cutoffs corresponding to 95% of common genes. (c) Bar charts showing the gene expression values in the five layers of MOB tissue for representative scGCO-only genes. (d) Same as (c) for representative HVGs-only genes. (e) Heatmaps showing the spatial expression patterns of representative scGCO-only genes. (f) Same as (e) for representative HVGs-only genes.

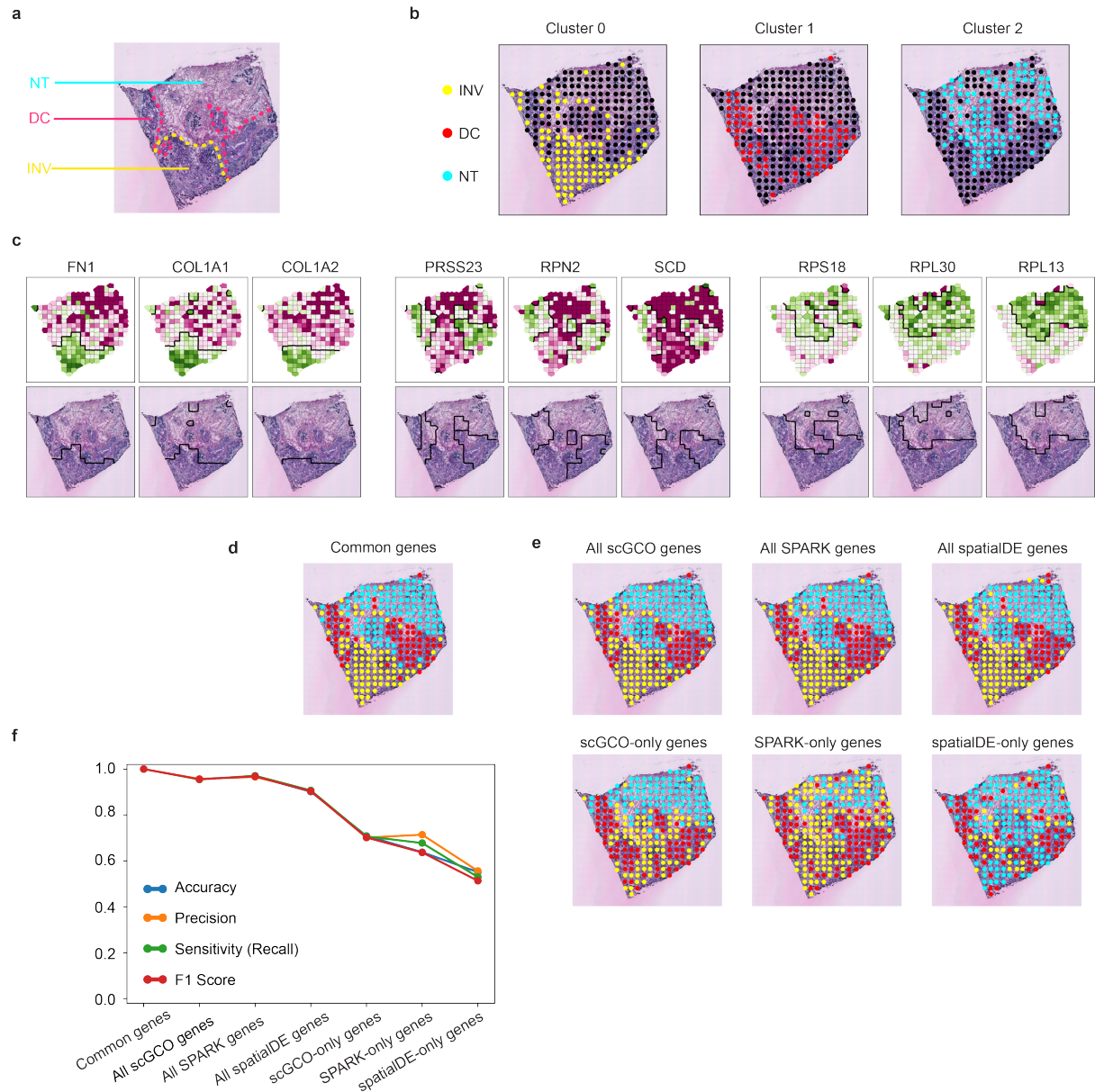
DESeq2-only genes. **(e)** Representative Voronoi diagrams showing graph cuts results (top panels), and associated H&E staining overlays (lower panels) for representative scGCO-only genes. **(f)** Same as **(e)** for representative DESeq2-only genes.



Supplementary Figure 17. Analysis of SV gene identification algorithms using breast cancer biopsies data.

(a) Venn diagram showing the gene set relationship among SV genes identified by scGCO (FDR < 0.05), spatialDE (FDR < 0.05) and SPARK (FDR < 0.05) in breast cancer biopsies. **(b)** Bar plots showing the number of SV genes identified by different methods across all 4 layers, error bars indicate mean $\pm$ SD. **(c)** Kernel density plots of the UMAP projections showing that no clusters were formed among random genes whose expression values were randomly shuffled. **(d)** Kernel density plots of the UMAP projections showing that genes identified by scGCO consistently formed three clusters. **(e)** Kernel density plots of the UMAP projections showing that genes identified by SPARK consistently formed three clusters. **(f)** Kernel

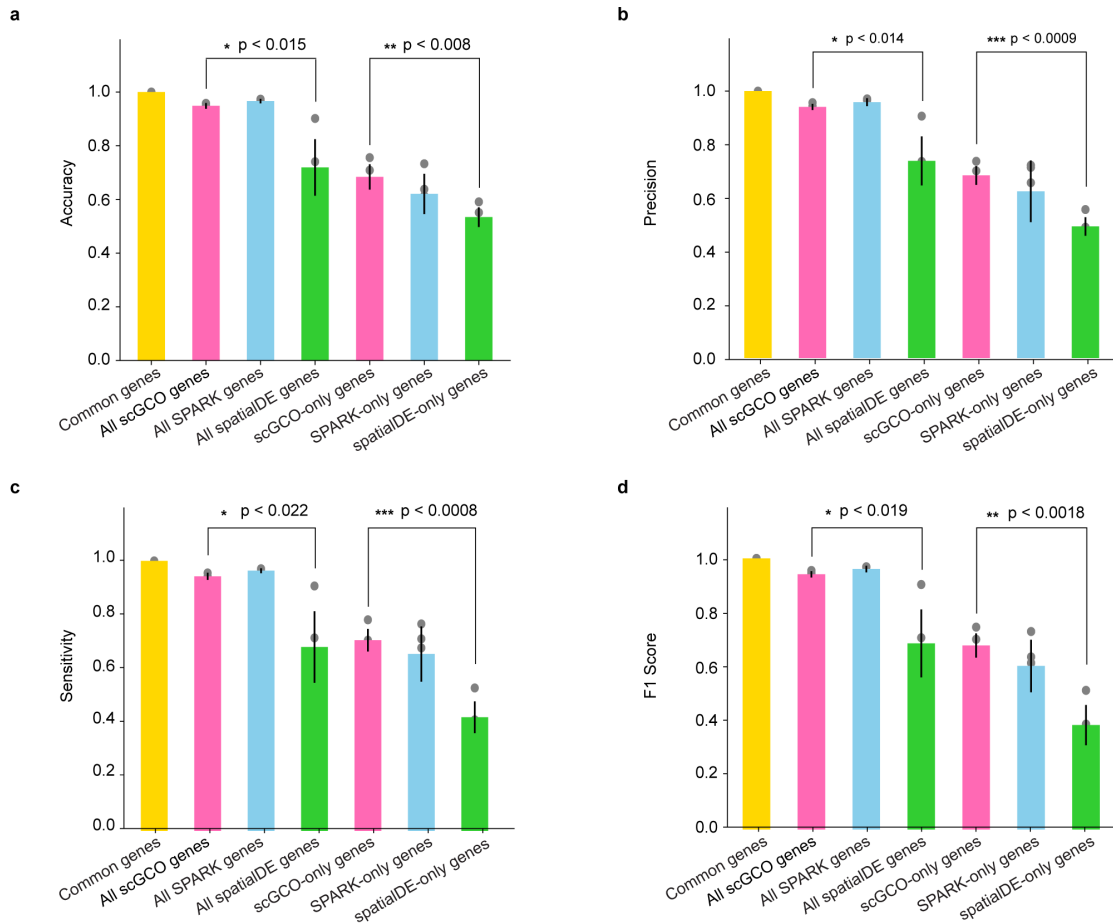
density plots of the UMAP projections showing that genes identified by spatialDE didn't consistently form three clusters.



Supplementary Figure 18. SV genes identified by scGCO reconstruct breast cancer tissue structures.

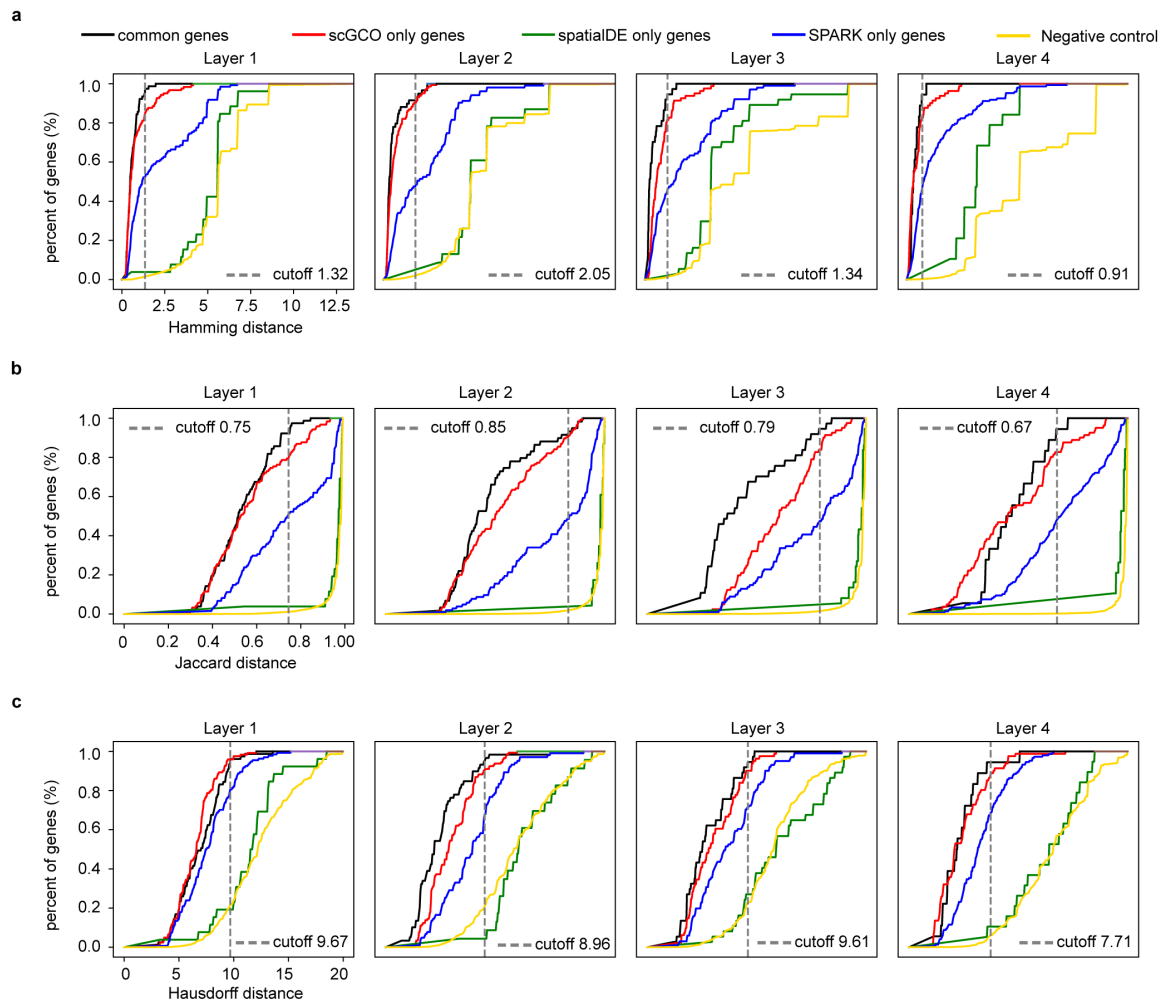
(a) Annotation of the breast cancer tissue architecture showing three regions: INV, DC and NT for layer 1. INV: Invasive Ductal Cancer; DC: Ductal Cancer; NT: Normal Tissue. (b) Tissue domains resolved with each cluster of genes identified by scGCO for layer 1. (c) Representative graph cuts and H&E staining overlays for SV genes identified by scGCO (layer 1). Voronoi diagrams showing graph cuts results (top panels), and associated H&E staining overlays (lower panels). (d) Reconstructed tissue architecture for layer 1 with SV genes jointly identified by all three methods. (e) Breast cancer tissue structures reconstructed

using all scGCO genes (upper column 1), all SPARK genes (upper column 2), all spatialDE genes (upper column 3), scGCO-only genes (lower column 1), SPARK-only genes (lower column 2), and spatialDE-only genes (lower column 3) for layer 1. **(f)** Lines plots showing accuracy, precision, sensitivity and F1 score of the reconstructed cancer tissues in **(e)**, tissue structure reconstructed with common genes was used as reference.



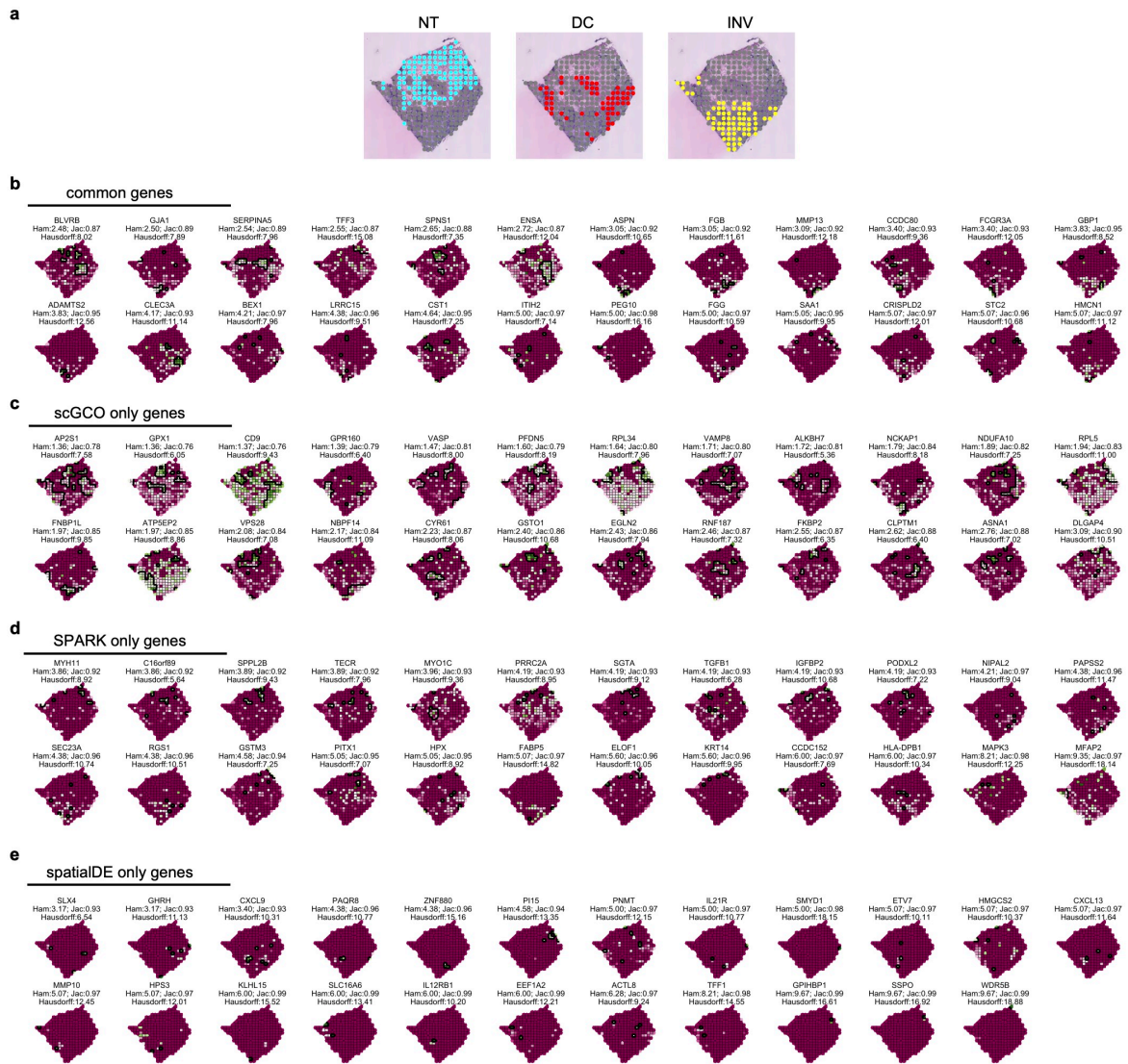
Supplementary Figure 19. Quantifying the performance of identified SV genes to resolve Breast Cancer tissue structures.

Bar plots showing the Accuracy (**a**), Precision (**b**), Sensitivity (Recall) (**c**) and F1 score (**d**) of reconstructed tissue structures using different sets of SV genes for all 4 layers, tissue structure reconstructed with common genes (SV genes jointly identified by all three methods) was used as reference. Error bars indicate means $\pm$ SD.



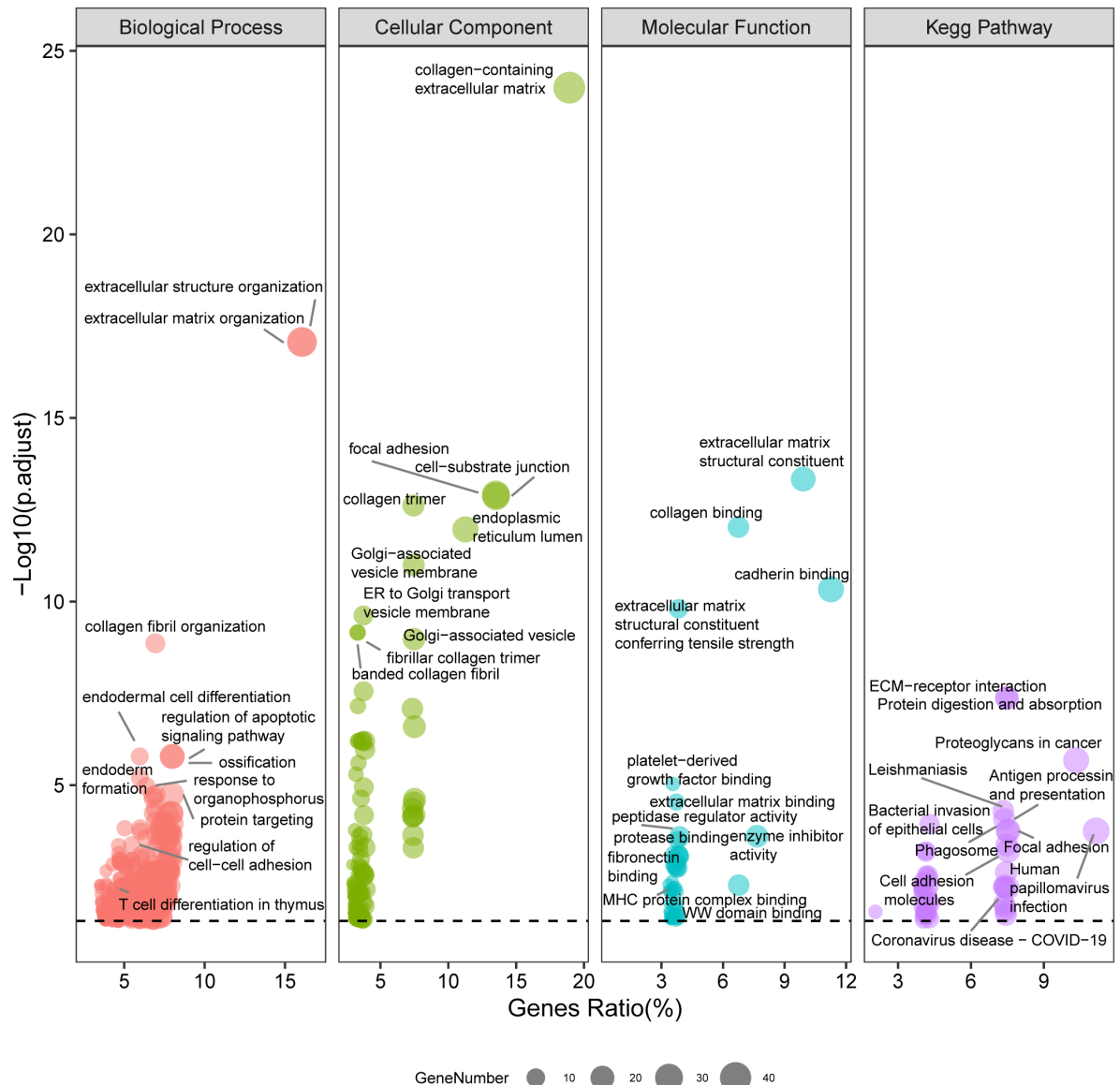
Supplementary Figure 20. Quantifying similarity of identified SV genes to reference cancer tissue structures.

(a) Line plots showing the cumulative distribution of normalized hamming distance of identified SV genes to reference BC tissue structures reconstructed using common genes. Black lines: SV genes jointly identified by all three methods; red lines: SV genes uniquely identified by scGCO; green lines: SV genes uniquely identified by spatialDE; blue lines: SV genes uniquely identified by SPARK; gold lines: genes with random spatial expression patterns as negative control. The dashed vertical lines indicate estimate distance cutoffs corresponding to 95% of common genes. **(b)** Same as **(a)** using Jaccard distance. **(c)** Same as **(a)** using Hausdorff distance.



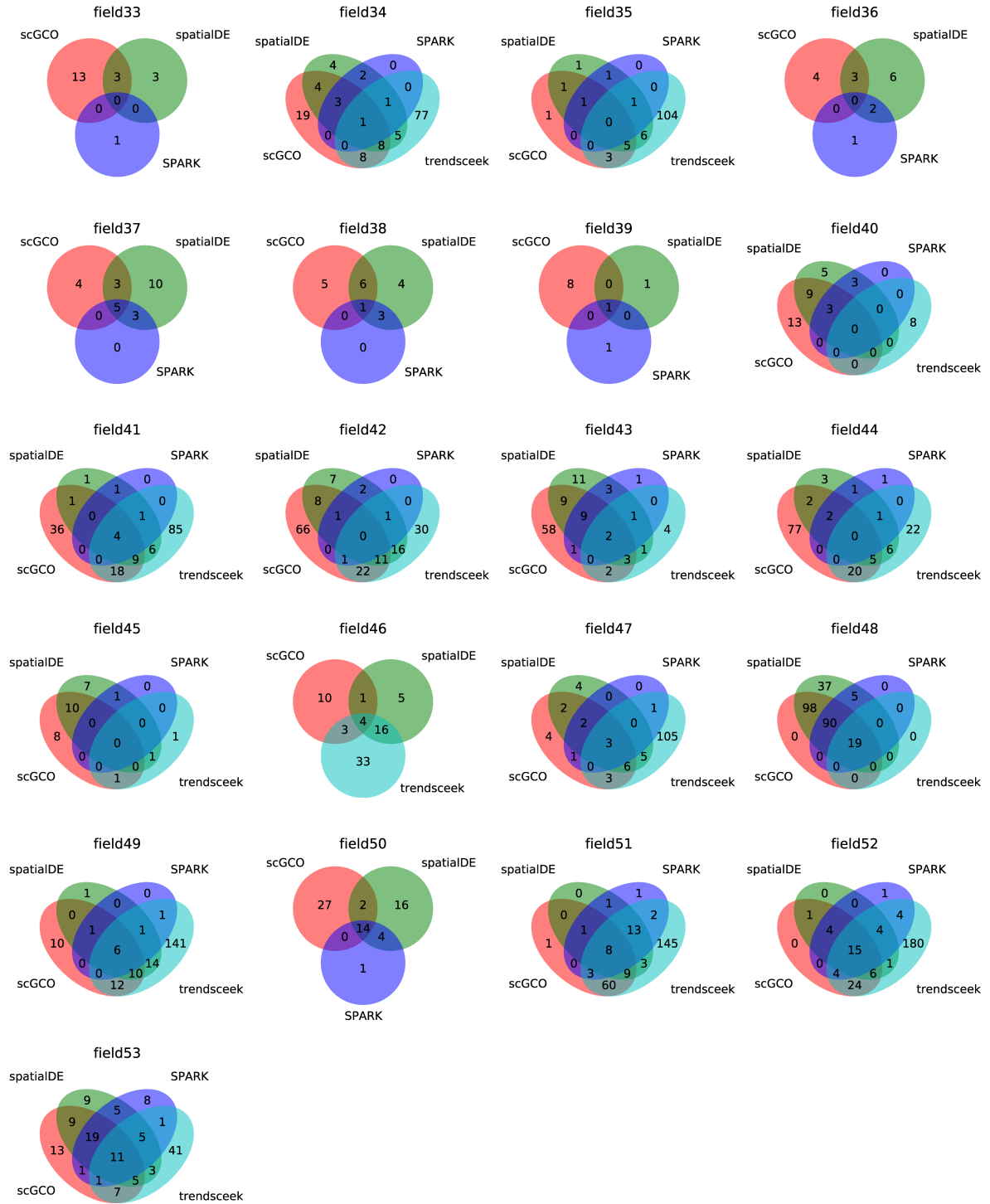
Supplementary Figure 21. Visualizing similarity of identified SV genes to reference breast cancer tissue structures.

(a) Reference cancer tissues structures reconstructed with common genes for layer 2. **(b)** Graph cuts showing bottom 24 genes with the largest hamming distance for common genes. **(c)** Same as **(b)** for scGCO-only genes. **(d)** Same as **(b)** for SPARK-only genes. **(e)** Same as **(b)** for spatialDE-only genes.



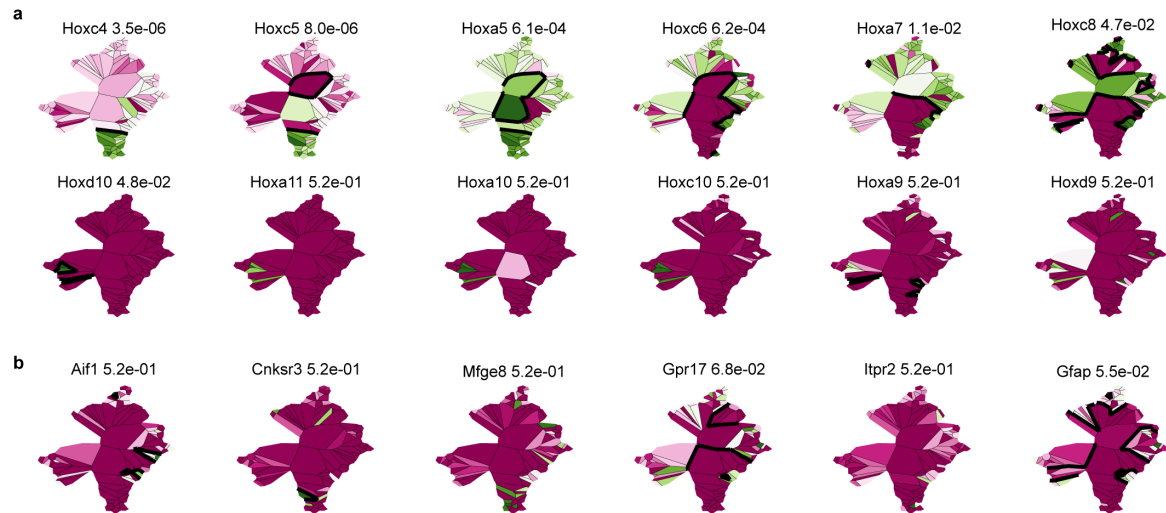
Supplementary Figure 22. SV genes identified by scGCO are enriched with cancer-related function.

Bubble plots showing significantly enriched gene ontologies and KEGG pathways for SV genes identified by scGCO from layer 2. The horizontal lines correspond to adjusted p-value of 0.05 with BH method.



Supplementary Figure 23. Identifying SV genes in mouse hippocampus seqFISH data.

Venn diagram showing the gene set relationship among SV genes identified by scGCO, spatialDE, SPARK and trendSceek from mouse hippocampus seqFISH data (FDR < 0.05).



Supplementary Figure 24. Identifying SV genes in mouse neurons tissues with LCM-seq technology.

(a) Spatial patterns for twelve Hox genes. **(b)** same as **(a)** for six non-spatial expression genes. Green denotes high expression and acacia red denotes low expression. Thicker black lines denote graph cuts boundaries. Numbers on the graph cuts plots represent adjusted p-values by scGCO.