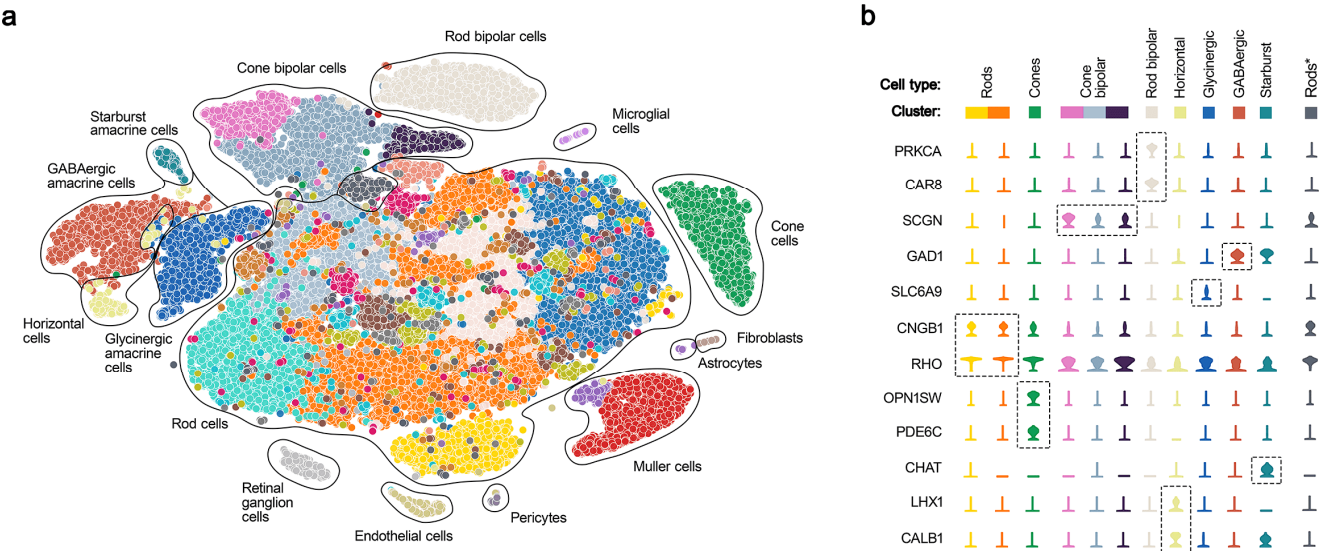
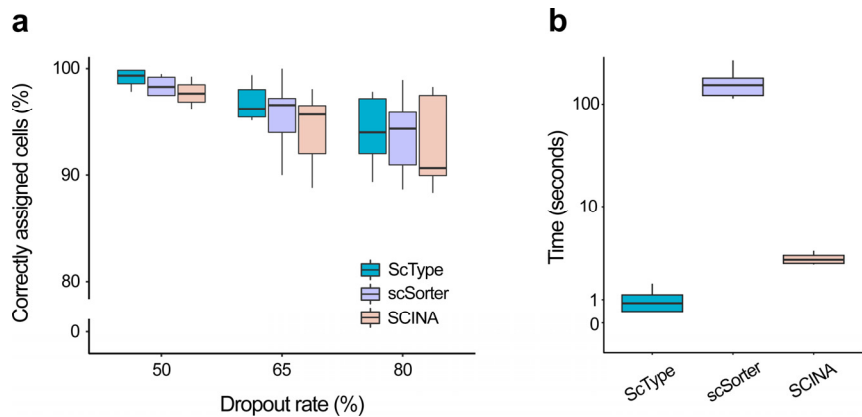


SUPPLEMENTARY FIGURES

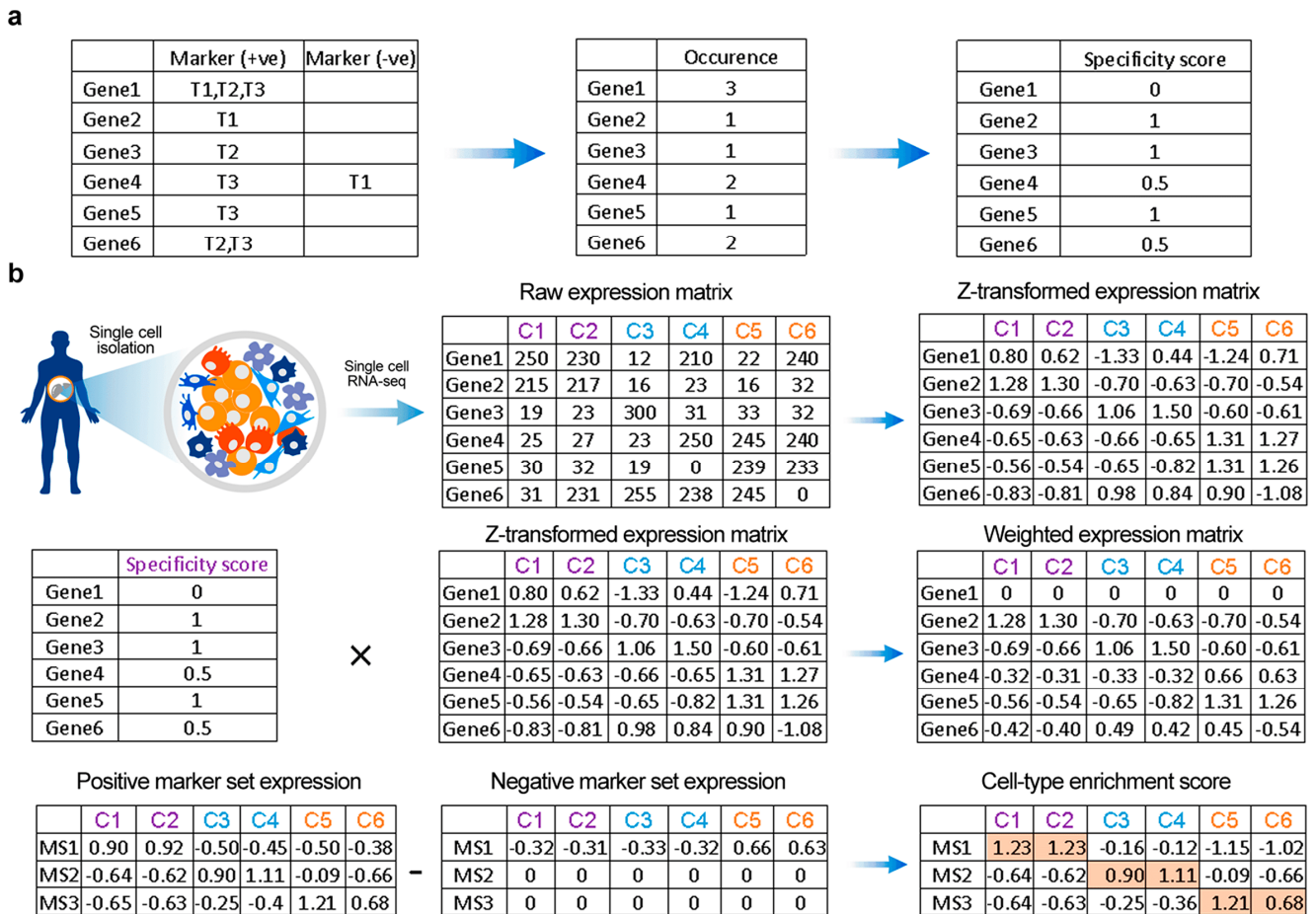


Supplementary Figure 1. ScType cell-type annotation of mouse retina scRNAseq dataset.²³

(a) UMAP plot shows the automated cell-type annotations with ScType. **(b)** Violin plots show the expression levels of the high-specificity marker genes that were used as a validation of correct cell-type assignments.

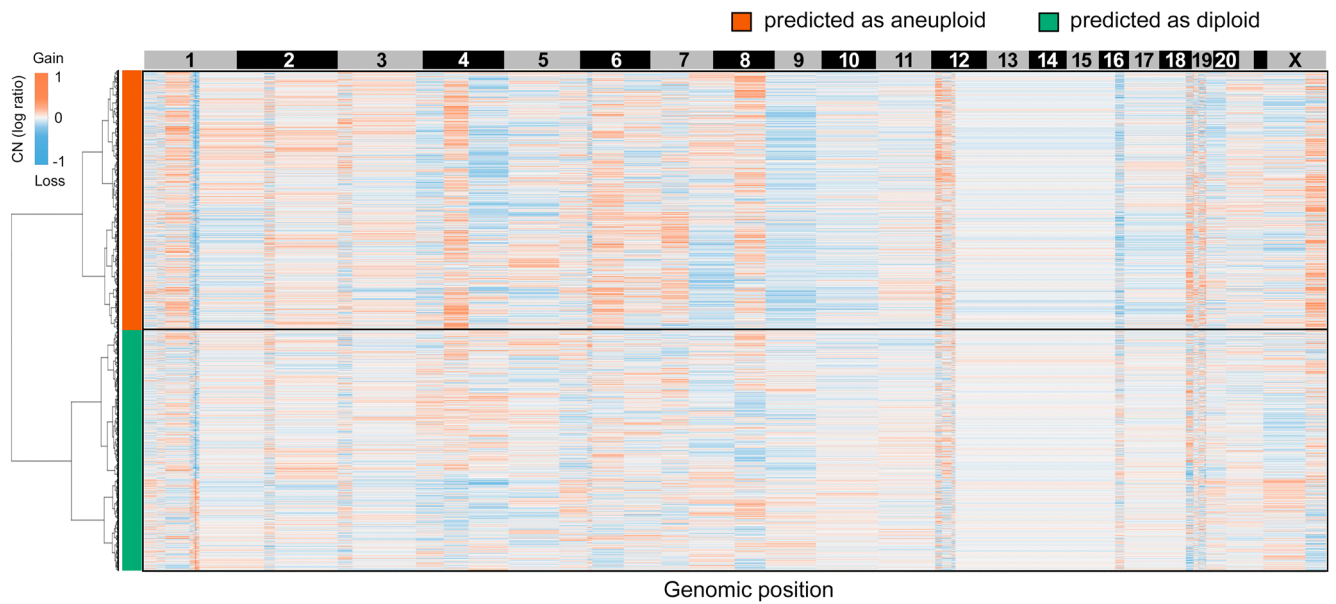


Supplementary Figure 2. Comparison of ScType, scSorter and SCINA accuracy and runtime performance. **(a)** The cell type annotation accuracy was tested in 45 simulated datasets with various dropout rates. **(b)** The ScType running time was faster compared to the other methods across the 45 simulated datasets ($P < 0.0001$, two-sided Wilcoxon test). Note: the time-axis is log-scaled. The horizontal lines of box-plots, in panels a and b, indicate the median, the boxes - interquartile range (IQR), and the whiskers are $Q1 - 1.5 \times IQR$ and $Q3 + 1.5 \times IQR$ (where $Q1$ and $Q3$ are firsts and third quartiles), across the 15 simulated datasets in each condition.

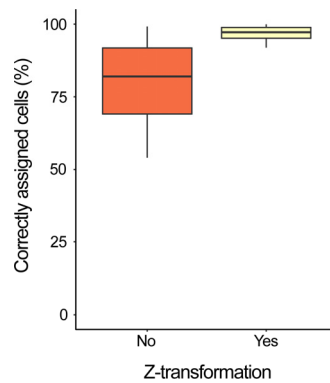


Supplementary Figure 3. An example of calculating marker specificity score and ScType score.

(a) In this toy example, Gene1 is a non-specific marker, and expressed in all 3 cell types (T1,T2,T3), whereas Gene2, Gene3, and Gene5 are specific markers for cell types T1, T2, and T3, respectively. Gene4 is a positive marker for cell type T3 and negative marker for cell type T1. Gene 6 is a positive marker for cell types T2 and T3. In order to calculate the marker specificity score (the right column), the marker occurrence across the cell types (middle column) is counted and scaled between 0 to 1 (0 indicates non-specific markers, maximum occurrence; and 1 for highly specific markers, minimum occurrence). **(b)** To calculate the enrichment score based on the marker expression in ScType, the raw expression data is first normalized and Z-transformed. Next, the normalized matrix is multiplied by the cell-type specificity score. Then, the expression scores of all the positive markers corresponding to a particular cell type are summarized into a single enrichment score by summing them and dividing by square root of their number. The same is done for the negative markers. Finally, the negative marker expression score is subtracted from the positive score to obtain the final enrichment score. Individual cells are assigned to a cell type based on the maximum value for the cell type marker set.



Supplementary Figure 4. Clustered heatmap of single-cell copy number profiles estimated with CopyKAT³⁶ tool in the AML patient sample.



Supplementary Figure 5. The impact of z-score transformation on the ScType annotation accuracy in the 45 simulated datasets with various dropout rates ($P < 0.001$, two-sided Wilcoxon test). The horizontal lines indicate the median, the boxes the interquartile range (IQR), and the whiskers are $Q1 - 1.5 \times IQR$ and $Q3 + 1.5 \times IQR$ (where $Q1$ and $Q3$ are firsts and third quartiles), across the 45 simulated datasets in each condition.