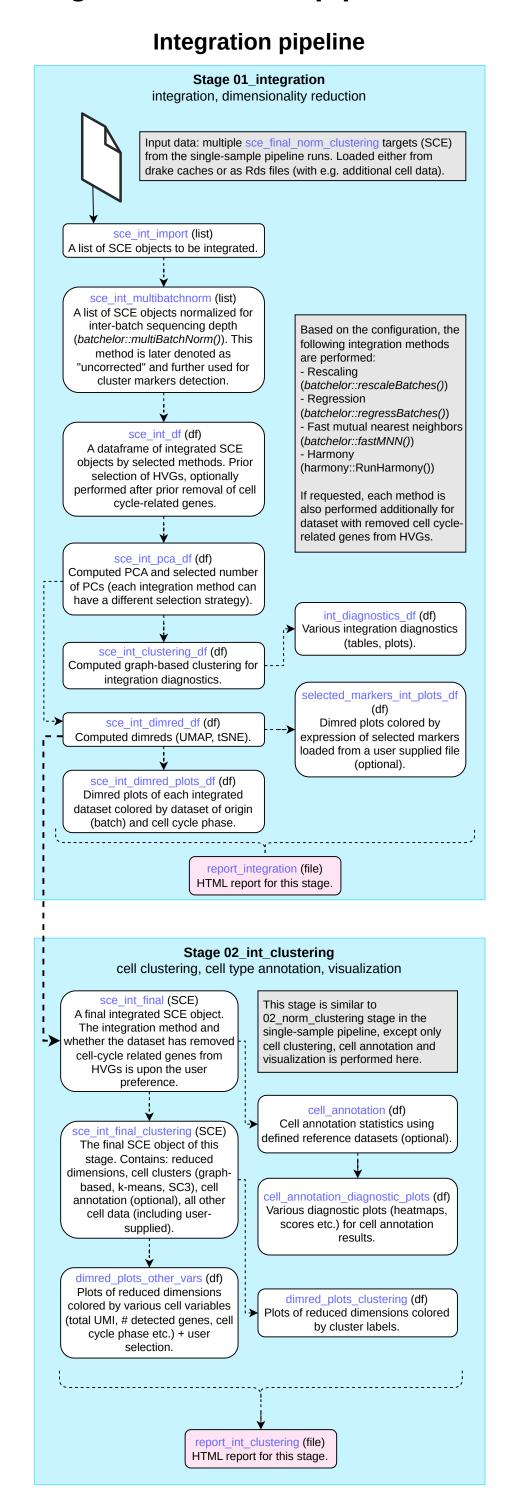
Supplementary Figure 2: Low-level diagram of scdrake pipelines

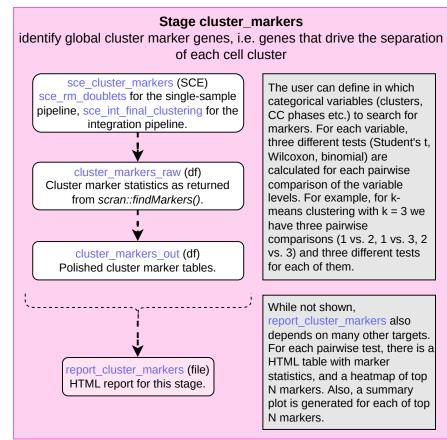
Single-sample pipeline Stage 01_input_qc quality control, cell and gene filtering Input data (one of): cellranger output SingleCellExperiment object (as a Rds file or from a drake cache) Feature-barcode matrix (delimited table) empty_droplets (DF) Statistics from sce_raw (SCE) DropletUtils::emptyDrops() (optional) ce valid cells (SCE) Removed empty droplets if enabled. Calculation of per-cell QC metrics (total number of UMI, % mitochondrial genes, etc.). - Assembly of two cell filters based on QC metrics: by predefined thresholds (custom filters) and by Median Average Deviation (QC filters). sce unfiltered (SCE) sce unfiltered plotlist (list) colData() extended by QC Plots of QC metrics. metrics and boolean filters. sce_custom_filter (SCE) sce_qc_filter (SCE) Filtered by custom filters Filtered by QC filters (MAD). (thresholds). sce_qc_gene_filter (logical) ce_custom_gene_filter (logical) Boolean filter for genes Boolean filter for genes (TRUE = discard). (TRUE = discard). sce_qc_filter_genes (SCE) ce custom filter genes (SCE) Filtered by QC filters (MAD) Filtered by custom filters and gene filter. (thresholds) and gene filter. sce final input qc (SCE) The final SCE object of this stage. Either based on sce_qc_filter_gene or sce_custom_filter_genes (depends on config). Added gene annotation to report_input_qc (file) rowData() HTML report for this stage. Stage 02_norm_clustering normalization, highly variable genes selection, dimensionality reduction, cell clustering, cell type annotation, visualization cc_genes (df) Dataframe of annotated cell cycle genes taken from Seurat::cc.genes.updated.2019 and present in the SCE. sce_norm (SCE) cell annotation diagnostic plots Normalized data by either scran or (df) SCTransform (can be skipped). Added diagnosti cell cycle phases to colData(). (heatmaps, scores etc.) for cell annotation results. sce_norm_hvg (SCE) cell_annotation (df) Added HVGs. Different strategies for Cell annotation statistics heir selection can be used. Optionally, using defined reference cell cycle-related genes can be datasets (optional). removed prior to HVG selection. pca_selected_pcs_plot sce rm doublets (SCE) Removed doublets (optional). Added A plot showing variance doublet score to colData(). explained by principal components and thresholds/strategies to select their optimal number sce_pca (SCE) for downstream usage. PCA computed and added to reducedDims(). elected_markers_plots (df) Dimred plots colored by sce_pca_selected_pcs (SCE) expression of selected PCA subsetted to selected number of markers loaded from a user PCs. Multiple selection strategies are supplied file (optional). available. sce dimred (SCE) dimred_plots_clustering (df) Computed UMAP and tSNE, added to Plots of reduced dimensions reducedDims(). colored by cluster labels. limred_plots_other_vars (df) sce_final_norm_clustering (SCE) Plots of reduced dimensions The final SCE object of this stage. colored by various cell Contains: reduced dimensions, cell variables (total UMI, # clusters (graph-based, k-means, SC3), ◀--detected genes, cell cycle cell annotation (optional), all other cell phase etc.) + user selection. data (including user-supplied). report norm_clustering_simple (file) report_norm_clustering (file) Simplified HTML report for this stage. Full HTML report for this stage. Includes only dimred and cell

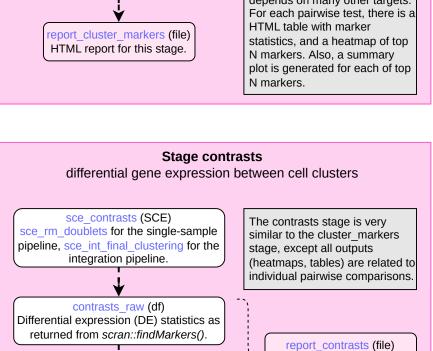
annotation plots.



Common pipelines

stages common to both single-sample and integration pipelines





contrasts_out (df)

Polished DE tables.

HTML report for this stage.

