**Demonstration of SCRAT analysis**

In this example, we demonstrated the major functions of SCRAT using a dataset which contains 230 HEK293T cells and 20 GM12878 cells. The aligned bam files (aligned to hg19) for this example are available at https://github.com/zji90/SCRATdata/tree/master/SCRAT_example_data_bam. In order to help users conveniently test SCRAT, this dataset can be directly loaded into SCRAT by using the "*Load example data*" function in Step 1. For readers' convenience, we also saved the SCRAT summarized features (obtained after performing Step 1 and Step 2 below) of this dataset and provide them at the following web link: https://github.com/zji90/SCRATdata/blob/master/SCRAT_summarized_features_GM12878_HEK293T.txt. If users start the SCRAT analysis from these summarized features, they can skip the procedure described below in Step 1 and Step 2, and use instead the "*Upload Summarization Table*" function in Step 2 to read in the summarized features (**Fig. 1a**). For instance, one can upload the data in the "*Input Summary Table*" section using the "*Choose Files*" button, and read in the data using the "*Read in*" button after the upload is completed (**Fig. 1b**). Users can also use the "*Load example summarized features*" function in Step 2 to load the summary table (**Fig. 1c**). The summarized features can be viewed in the "*Results*" panel (**Fig. 1d**). Then, one can proceed to Step 3 and Step 4.

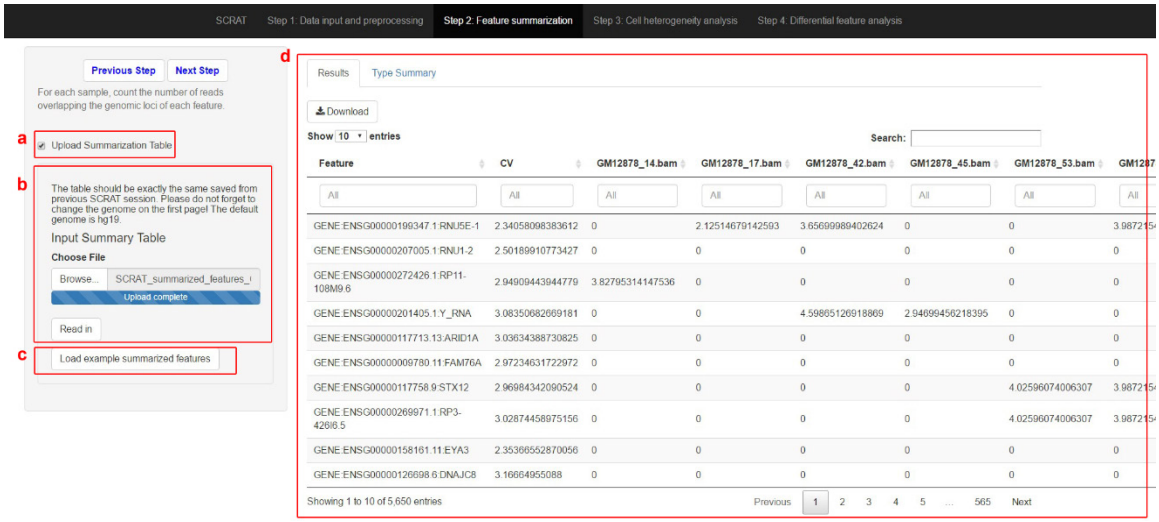


**Figure 1.** User can upload the previously saved summarized features into SCRAT for analysis.

**Step 1: Data input and preprocessing**

The first step is to input the single-cell data (aligned bam files) into SCRAT. First, one has to select the corresponding reference genome from the "*Select Genome*" section (**Fig. 2a**). Then, one can upload bam files in the "*Input Bam Files*" section using the "*Choose Files*" button, and read in the data using the "*Read in*" button after the upload is completed (**Fig. 2b**). To help users test SCRAT, the bam files for this example can be loaded into SCRAT by simply clicking the "Load example data" button (**Fig. 2c**). By default, SCRAT will filter blacklist regions and exclude samples with total number of reads less than 500 (adjustable by the user). Information about the input data will be

shown after they have been read in (**Fig. 2d**). One can proceed to feature summarization using the "*Next step*" button (**Fig. 2e**).

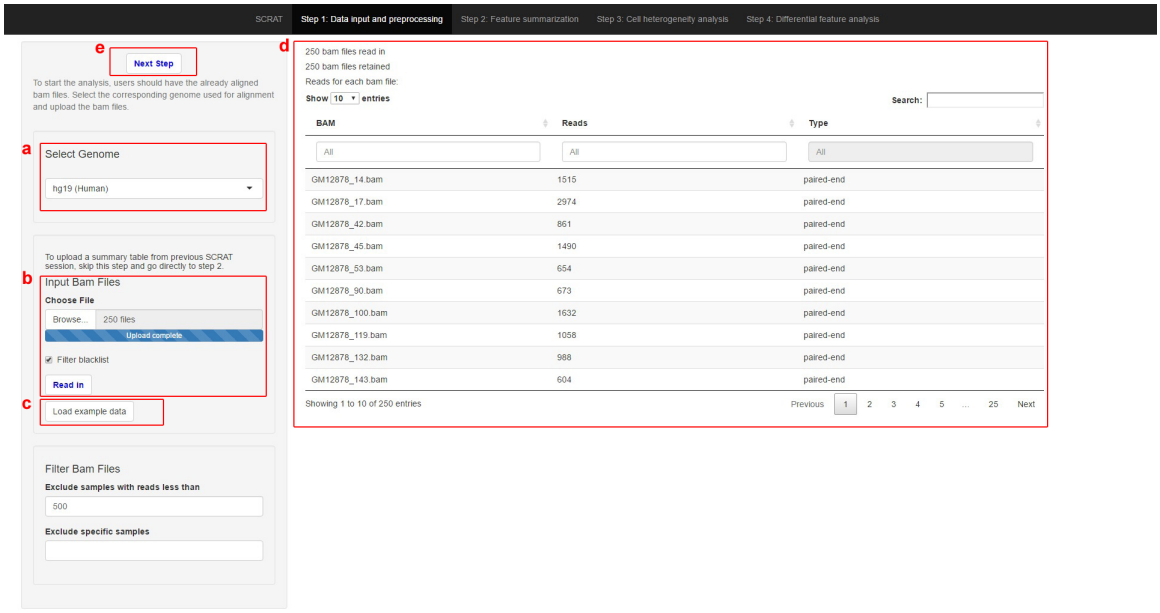


**Figure 2.** SCRAT analysis step 1 -- Data input and preprocessing.

**Step 2: Feature summarization**

The second step is to summarize the input data into different features according to the feature definitions. To demonstrate, we summarized signals in this test dataset using the pre-defined features in SCRAT. In the "*Choose Summarizing Method*" section, we selected all pre-defined SCRAT features for our analysis (i.e., *Motif*, *ENCODE Cluster*, *Gene* and *Gene set*; **Fig. 3a**).

SCRAT provides rich tuning options for each feature type. The parameters of each feature types can be adjusted in the "*Method Details*" section of the user interface (**Fig. 3d**). When the example dataset was summarized based on *Motif*, we asked SCRAT to aggregate reads within 100 base pair (bp) flanking region from both sides of the motif sites. When the data were summarized based on *ENCODE Cluster*, we set the cluster number to be 2000. When the data were summarized based on *Gene*, we asked SCRAT to aggregate reads within the 3000 bp upstream and 1000 bp downstream region from the transcription start site (TSS) of each gene. When the data were summarized based on *Gene Set*, we chose to include only Gene Ontology (GO) gene sets for analysis. For each gene set, we asked SCRAT to aggregate reads within the 3000 bp upstream and 1000 bp downstream regions from TSSs of all genes.

SCRAT allows one to normalize the features and filter them based on the user-provided parameters (**Fig. 3b**). Once all parameters are set, one can start the summarization process using the "*Run Summarization*" button (**Fig. 3c**). After the summarization is done, the summarized features can be viewed and downloaded from the "*Results*" panel (**Fig. 3e**). Then, one can proceed to cell heterogeneity analysis using the "*Next step*" button (**Fig. 3f**).

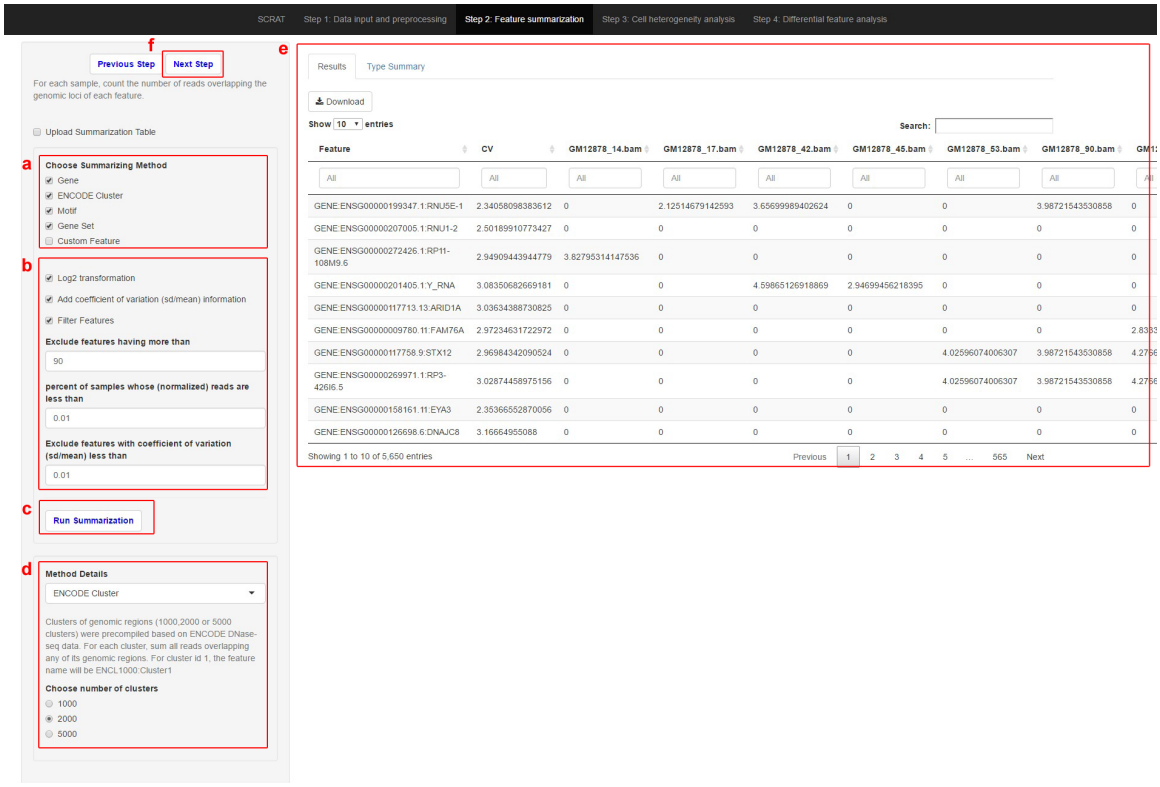


**Figure 3.** SCRAT analysis step 2 -- Feature summarization.

### Step 3: Cell heterogeneity analysis

The third step is to dissect the cell heterogeneity by clustering the cells. First, one can select different types of features for clustering in the "*Select Feature Type*" section (**Fig. 4a**). Second, one can choose what type of methods to use to reduce the dimension of the features in the "*Dimension reduction method*" section (**Fig. 4b**). Third, one can choose the clustering method in the "*Clustering method*" section (**Fig. 4c**). By default, SCRAT selects the *ENCODE Cluster* features and uses the principal components of these features to cluster cells based on model-based clustering. One can start the clustering process using the "*Perform Clustering*" button (**Fig. 4d**). Then, the result will be shown in the "*Clustering Result*" panel (**Fig. 4e**). Applying this procedure to the example data yields two clusters (**Fig. 4e**), corresponding to GM12878 and HEK293T respectively.

After obtaining the cell clustering results, one can further explore the cell identities by comparing the individual cells with the existing cell types in our pre-compiled bulk DNase-seq database. First, one can use the "*Include existing cell types*" function (**Fig. 4f**) to select samples from the existing cell types in the database and project them to the principal component space of the single cells. Second, one can also evaluate the similarity between each cell and the existing cell types using the "*Similarity to existing cell types*" function (**Fig. 4g**) based on the selected features (**Fig. 5a**). One can start the analysis using the "*Calculate Correlations*" button (**Fig. 5b**). The results will be

visualized as a heatmap (**Fig. 5c**). Then, one can proceed to differential feature analysis using the "*Next step*" button (**Fig. 4h**).

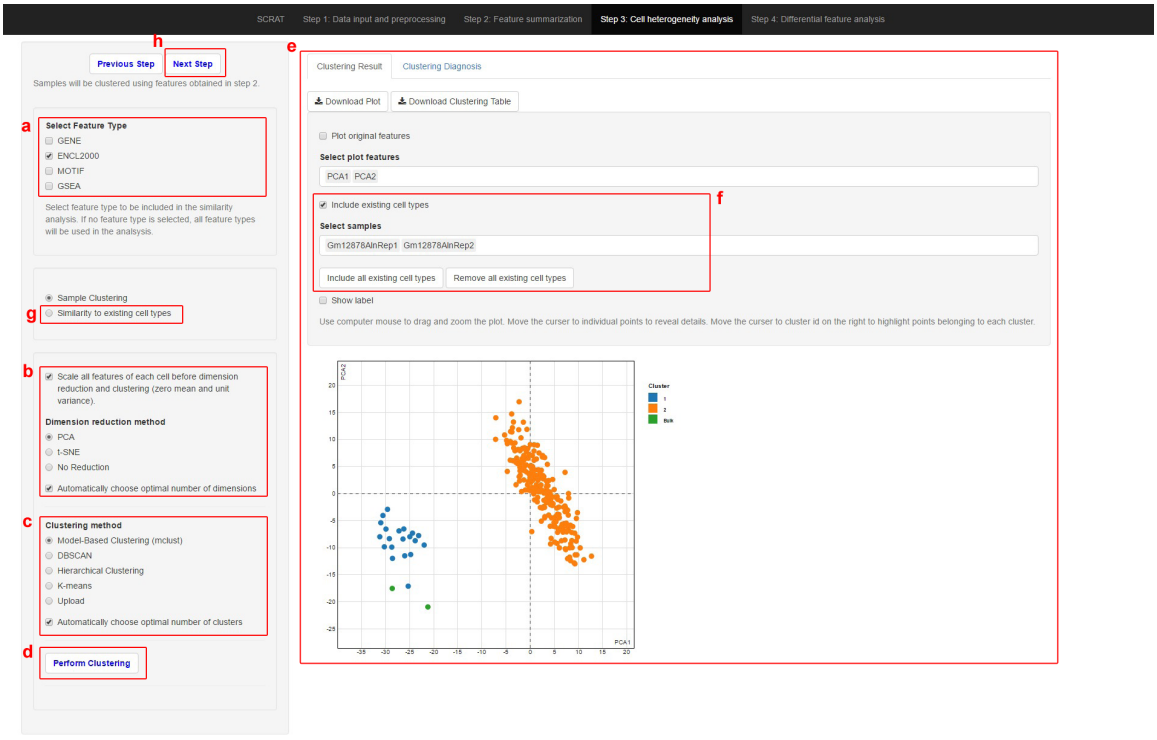


**Figure 4.** SCRAT analysis step 3 -- Cell heterogeneity analysis.

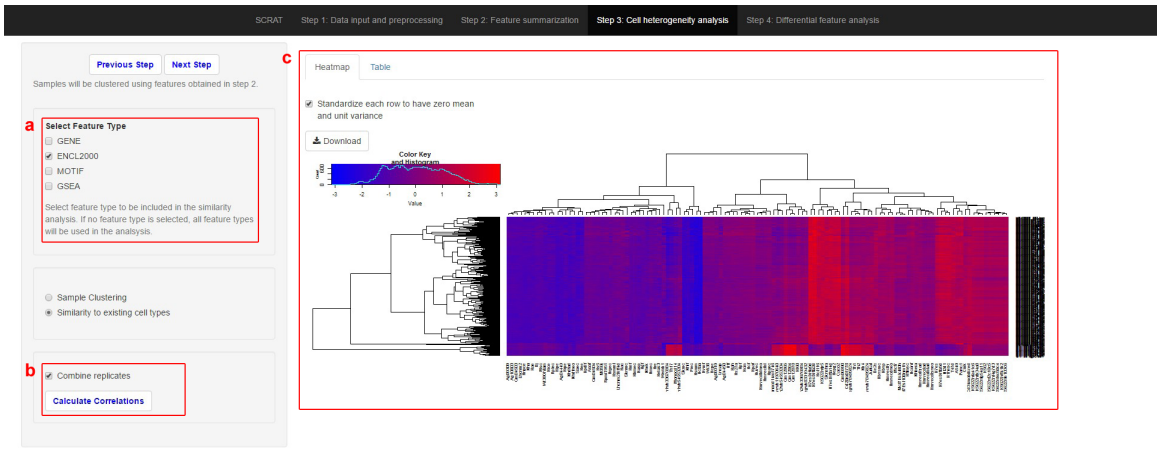


**Figure 5.** SCRAT analysis step 3 -- Cell heterogeneity analysis (cont'd). Evaluating similarity to existing cell types.

**Step 4: Differential feature analysis**

The last step is to identify the differential features among different cell subpopulations. One can perform analysis to all cell clusters obtained from Step 3 or a subset of selected cell clusters (**Fig. 6a**). Then, one can choose a statistical test (**Fig. 6b**). If more than two cell clusters are selected, ANOVA F-test, Kruskal-Wallis test, or permutation test based on F-statistics can be used to identify differential features. If only two cell clusters are selected, t-test, Wilcoxon rank-sum test, or permutation test based on t-statistics can be used. One can click the "Perform Test" button to start the analysis (**Fig. 6c**). The results including the name of the feature, the test statistics and the adjusted p-value (FDR) will be shown in the "*Results*" panel (**Fig. 6d**).

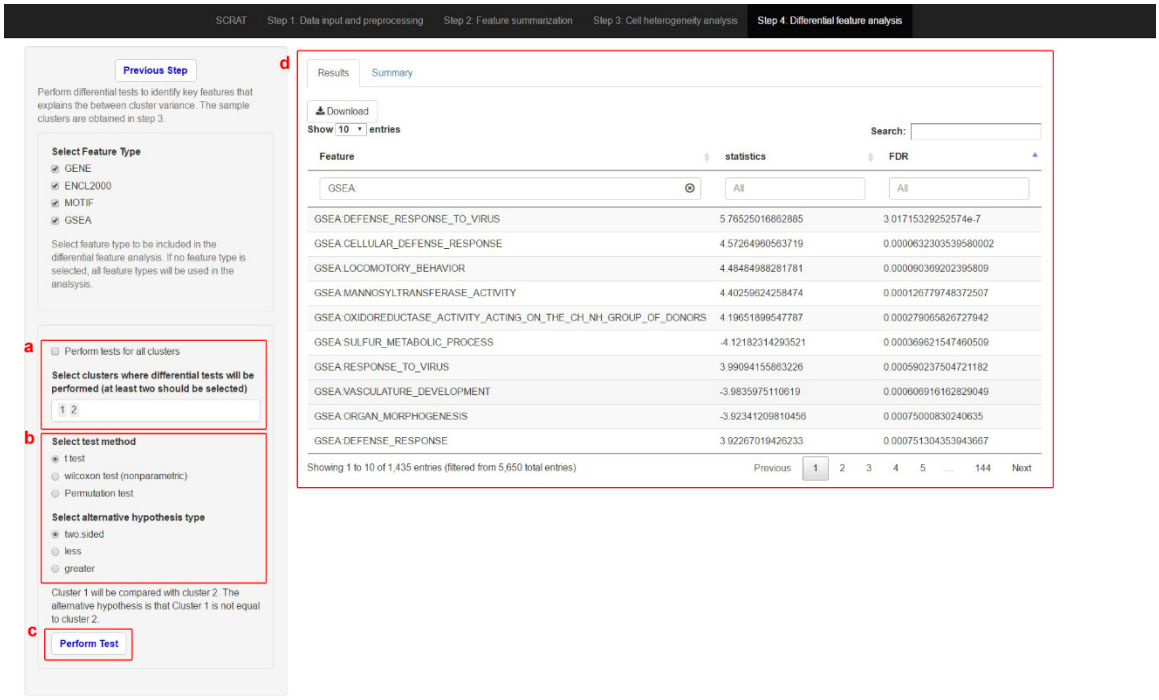


**Figure 6.** SCRAT analysis step 4 -- Differential feature analysis.