
Algorithm 1: Cell Type Label QC From Reference Marker Genes

input : scRNA-seq AnnData obj. \mathcal{A} , dictionary \mathcal{M} of reference cell type markers from GMT file \mathcal{G}
 Hyperparameters flag_threshold
output: Dictionary $\mathcal{D}_{\text{flag}}$ of Flagged Cell Type Labels
 Heatmap \mathcal{H} of Number of Genes

```

1 Initialize dictionary Dictionary  $\mathcal{D}_{\text{flag}}$  to store QC flagged cell types
2 Initialize list  $\text{unique\_celltypes}$  of unique cell types in  $\mathcal{A}$ 
3 Execute Differential Gene Expression analysis on  $\mathcal{A}$ 
4 for  $i \leftarrow 0$  to  $\text{length}(\text{unique\_celltypes})$  do
5   for all key value pairs  $(k, v)$  in  $\mathcal{M}$  do
6     Initialize list  $\text{ref\_genes} = v$ 
7     Initialize list  $\text{genes}$  of differentially expressed genes from  $\text{unique\_celltypes}[i]$  in  $\mathcal{A}$ 
8      $S = (\text{genes} \cap \text{ref\_genes})$ 
9     if  $\text{length } S \geq \text{flag\_threshold}$  then
10      | Append  $S$  and ref celltype  $k$  to  $\mathcal{D}_{\text{flag}}[\text{unique\_celltypes}[i]]$ 
11    end
12  end
13 end
14 Create Heatmap,  $\mathcal{H}$  with y axis as cell type in  $\mathcal{D}$ , x axis as ref cell type in  $\mathcal{M}$ 
15 and cell colors as number of overlapping genes
16 return  $\mathcal{D}_{\text{flag}}, \mathcal{H}$ 

```
