Search engine for genes in
large single-cell sequencing collections
https://scfind.sanger.ac.uk/
(User Guide)

The User interface of Scfind
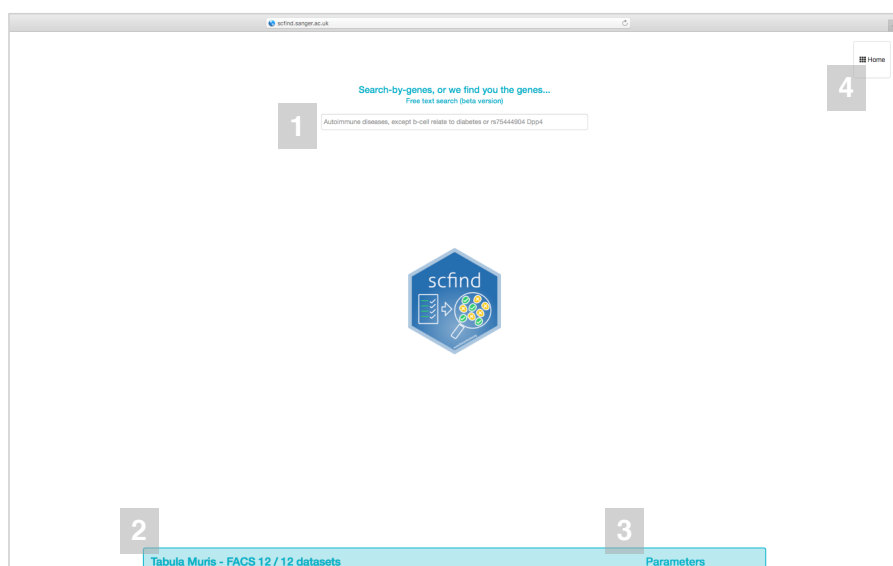
1 **Search box:**
Users can input a list of genes, SNPs ID, MeSH ID or medical/biological/chemical terms

2 **Dataset check box:**
Where the user can find all the datasets in the index and limit the search result by dataset / tissue

3 **Parameters:**
Where the user can adjust the parameters for each search

4 **Home button:**
To the home page of our index collection

4 Steps : From your a query to your target cell type

I

Step *I*: Input your query

- The bar chart **1** is a summary of the number of cells in the index expressing the gene

- User can always amend the initial gene list by clicking on ☑

Step *II*: select one of our recommended sub-query on table **2 according to TF-IDF score**

- subqueries are ranked by TF-IDF score *(Highest = likely the best)*

- click on **TF-IDF ⬍** to sort TF-IDF score in order

II

Step *III*: Select the enriched cell types on table **3**

- Cells expressing all genes in the query are grouped by cell type. The user can rank them based on enrichment by clicking on **p-value ⬍**

- User can limit the dataset by clicking on ☑ in **4**

- User can adjust the degree of greediness in **4** to get the best gene set for each term in the query. The higher the degree, the more gene (but less relevancy) will return.

III

Step *IV*: Check the best representing genes

- Click on any cell type you are interested in **3**

- **5** allows you to view the cells in a *UMAP*, compare *precision / recall / f1 scores* of each gene corresponding to each cell type and generate a *Word Cloud* relevant to your query.

- User can click on **Copy** **CSV** **Excel** to export the tables as raw text, csv file or .xlsx file

IV