CellDepot: A unified repository for scRNAseq data and visual exploration

Dongdong Lin   Yirui Chen   Soumya Negi
Derrick Cheng   Zhengyu Ouyang   David Sexton
Kejie Li   Baohong Zhang

2021-10-01

Contents

Chapter 1

Preface

This is a CellDepot Supplementary book written in **Markdown**.

Chapter 2

Supplementary Tables

2.1 Comparsion matrix of web portal tools

Note: The criteria for query search and data analysis explorer please see Table S2 and S3.

2.2 Criterion for query search

2.3 Criterion for data analysis explorer

2.4 Project metadata captured in CellDepot

Table 2.1: Table S1 - Comparsion matrix of web portal tools

Web.application.repository	CellDepot	Corp
Year	2021	2021
Main Function		
Database Explorer	Y	Y
Query Search	Advanced	NA
Data Analysis Explorer	Advanced	Adva
Data Type supported		
Datasets	divserse	divse
Datatypes	Single-cell	Sing
Visualzation features	Violin, dot, heatmap, bar, QC, scatter, density, embed-ding plots	Bar,
Links		
Source Code link	https://github.com/interactivereport/CellDepot	http
Demo link	http://celldepot.bxgenomics.com/	http

Table 2.2: Table S2 - Criterion for query searchs

Query.Search	Keyword.Search	Multiple.Object.Search	Category.Filters
Basic S2	Y		
Basic II			Y
Intermediate	Y		Y
Advanced	Y	Y	Y

Table 2.3: Table S3 - Criterion for data analysis explorer

Data.Analysis.Explorer	Analyze.scRNAseq.Data	Anaylze.Gene.Expression	Customize.Displays
Basic	Y		
Intermediate	Y	Y	
Advanced	Y	Y	Y

Table 2.4: Table S4 - Project metadata captured in CellDepot

X	General.Category	Expected.Variable.Type	Description
Project Browser/Search			
	Annotation Groups	String	categorical features from h5ad file
	Cell Count	Integer	numbers of cell in study
	Actions	Link	three options: 1) Study summary infor
	Custom Accession	String	Customized accession name for individ
	Description	String	Additional information
	DOI	Link	Digtial Object Identifier
	File Name	String	h5ad file name
	File Size	Integer	size of h5ad file
	Gene Count	Integer	numbers of gene in study
	Name	Link	project name
	Notes	String	study notes
	PMC ID	Link	
	Publication Title		
	PubMed ID	Link	
	Speices	String	
	URL	Link	
	Year	String	

Chapter 3
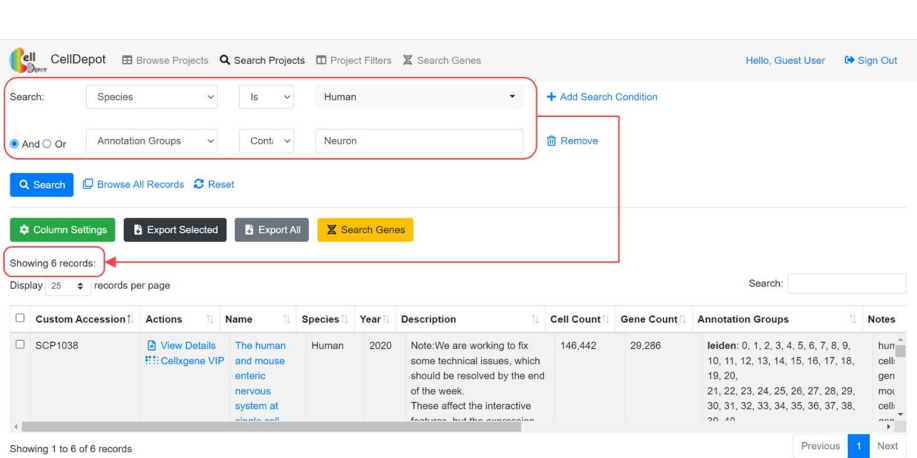
Supplementary Figures

Figure 3.1: Figure S1. Data Filtering. Query search of 'Species is Human' and 'Annotation Groups contains Neuron' brings about nine datasets of interest.
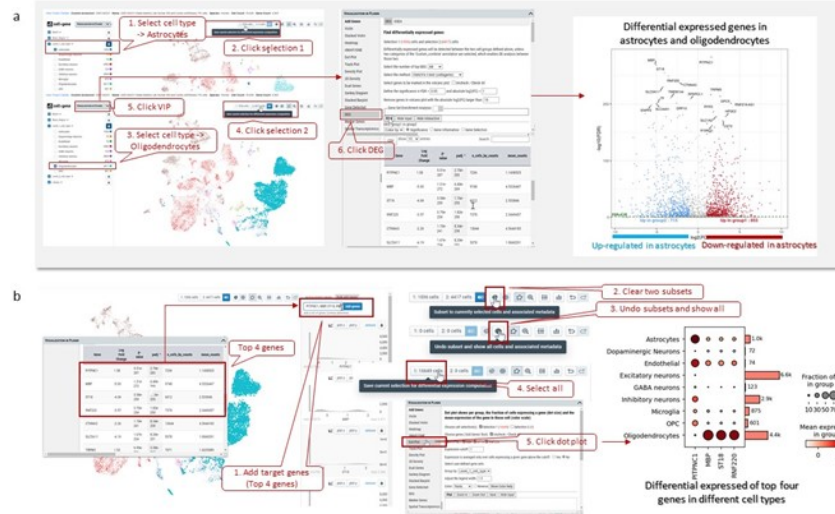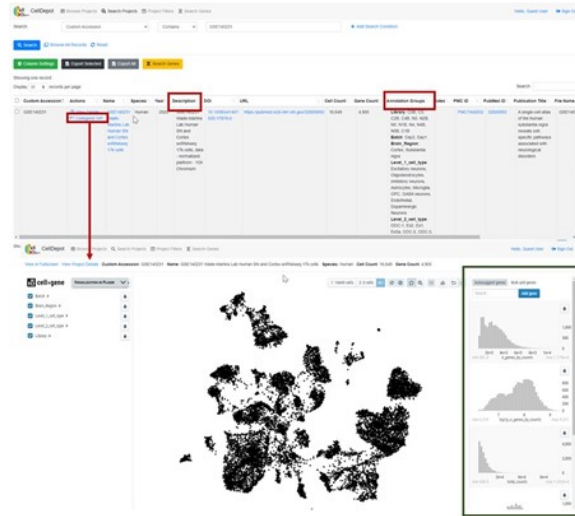
Figure 3.2: Figure S2.Exploration of differential expressed genes in dataset GSE140231 through cellxgene VIP. (a) Differential expressed genes in astrocytes and oligodendrocytes. (b) The expression of top four genes in different cell types.

Figure 3.3: Figure S3. The exploration of observation features of dataset GSE140231. Red-marked categorical features are shown on CellDepot Project page (highlighted by red framed box), the numerical features marked in green color can be visualized as distribution plots on the rightmost panel in cellxgene VIP tool, which is highlighted by green box.

Chapter 4

Supplementary Tutorial

CellDepot is database management system integrated with management system, query searching and data visualization tools [2, 3] for scRNA-seq datasets, which can be accessed by the link http://celldepot.bxgenomics.com. This is a supplemental tutorial provides a detailed guide.

Figure 4.1: Figure S4. CellDepot website

The interface contains multiple tabs, corresponding to import and/or select objects in CellDepot scRNA-seq database, that can be accessed on top panel of the webpage. Users can upload their own dataset or explore the existing datasets for visualization and analysis.

13
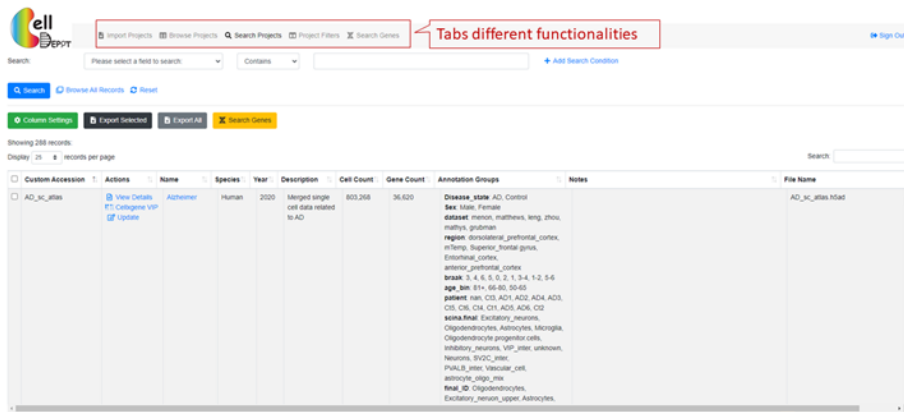
4.1　Upload Projects

To upload new projects in CellDepot database, two files are required: 1) .h5ad files and 2) project information in csv format. Detailed formatting guidance can be found by 'Download Example File' hyperlink on webpage. In addition, two cellxgene VIP launch methods are provided: standard and preload in memory. Standard mode is for the first-time imported datasets, while preload in memory should be chosen when users update the meta information of datasets. After the projects are submitted, CellDepot will automatically analyze the datasets. To explore the detail of uploaded datasets, users can navigate to 'browse projects' page and then search the imported datasets by the customized accession number.
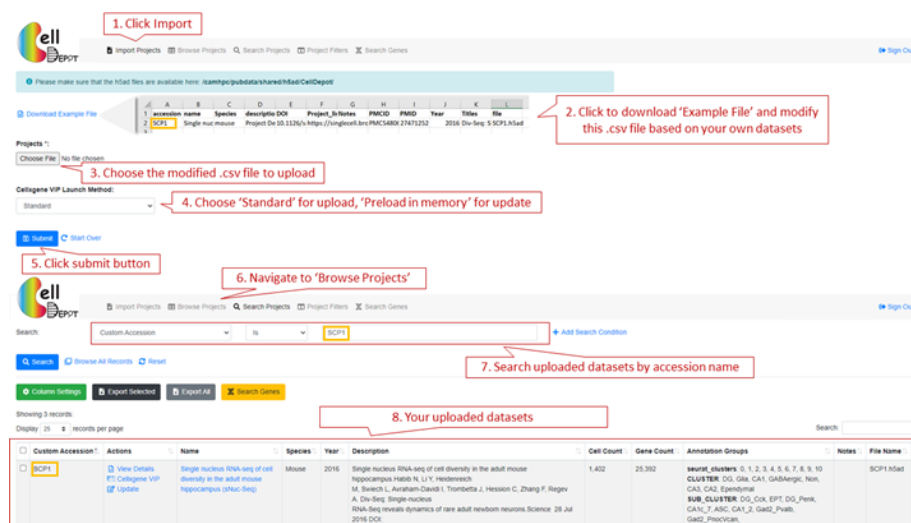

Figure 4.2: Figure S5. Workflow of how to import personal datasets.

4.2　Browse Projects

4.2.1　Search Projects

This function allows the user to search any targeted interests, which also can be accessed through search projects on the top panel of the webpage. Users is allowed to select the projects by 17 attributes: annotation groups, cell count, cellxgene VIP launch method, Custom accession, description, DOI, file name, file size, gene count, name, notes, PMC ID, Publication Title, PubMed ID, Species, URL, Year. These 17 fields can also be (partially) displayed on the webpage through 'column setting' on the webpage. Users can also search projects by the keywords via the search function on the right of the webpage. In addition, by

'column setting', users can set up the customized layout of targeted projects; thereby exporting to csv format.
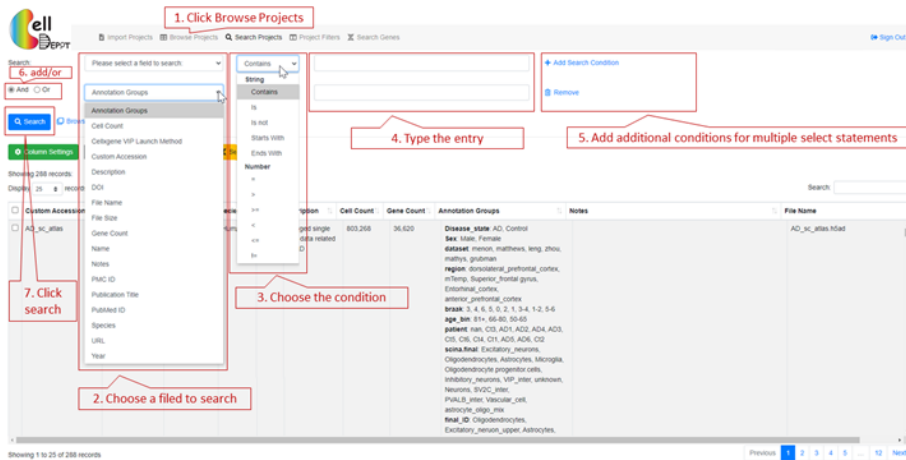

Figure 4.3: Figure S6. Workflow of how to search query on 'Browse Projects' page.

4.2.1.1 Case Study 1

Cross-project comparison of skeletal muscle marker genes PAX3, PAX7, PITX2, MYF5, MYF6, MYOD1, MYOG, NEB, and MYH3 among the datasets whose species is human and cell type is myogenic.

For each project, users can view the datasets information, visualize data analysis, and conduct update through clicking on "View Details", "cellxgene VIP", and "Update" links, respectively.

4.2.2 Project Filters

This page provides the matched datasets by simply clicking the categories. It is a first-time user-friendly functionality as users may not be familiar with the content of the database. The advance search function is the same as that on the 'Browse Projects' page (Details see Figure S4).
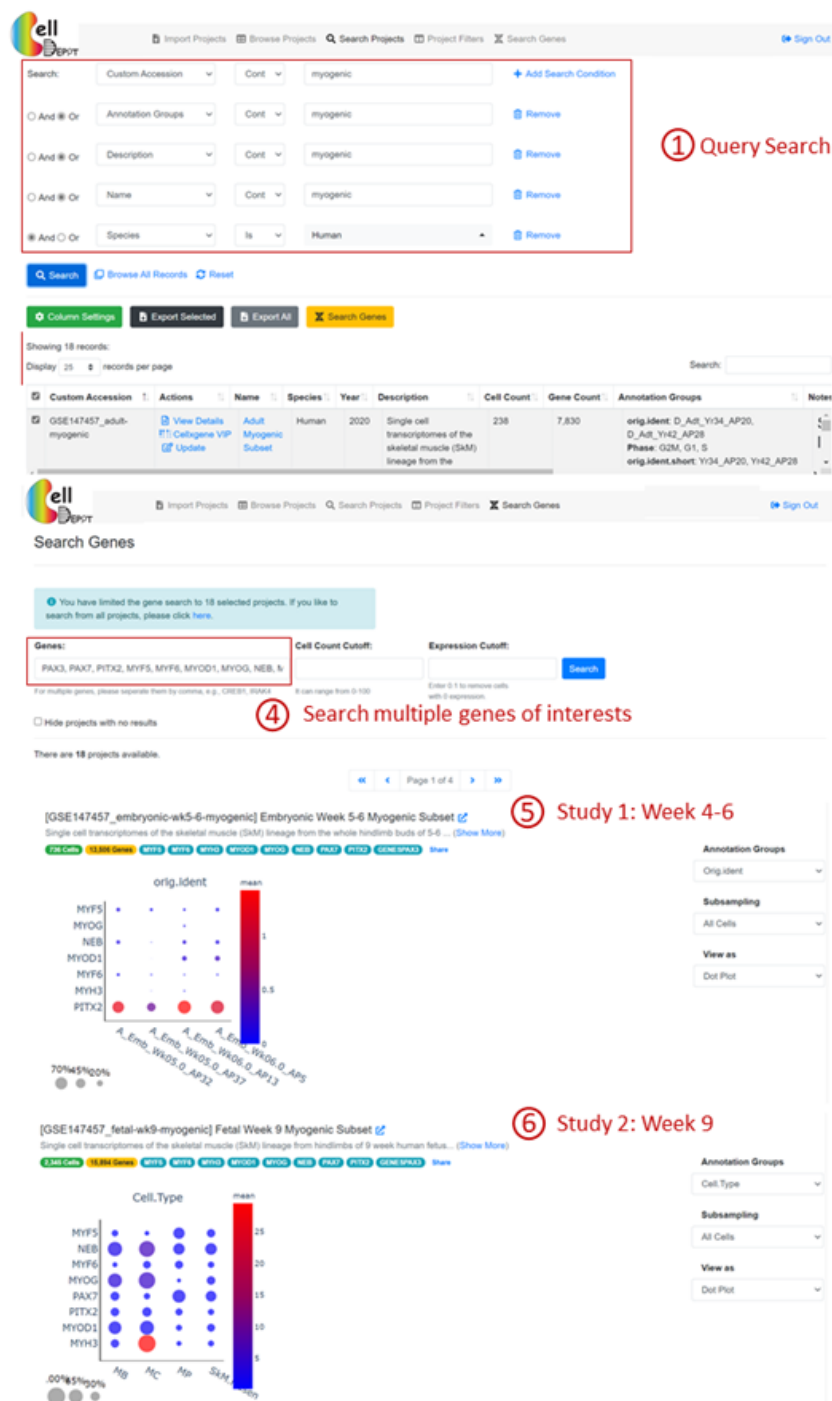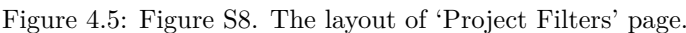
Figure 4.4: Figure S7. Workflow of how to conduct the cross-project comparison of gene sets among the selected datasets.

Figure 4.5: Figure S8. The layout of 'Project Filters' page.

4.3   Visaulize Datasets

4.3.1   View Details

The datasets information consists of project summary and annotation groups. The project summary is provided by each user when uploading projects. The information of annotation groups is retrieved from uploaded .h5ad file.

4.3.2   Update

Project summary information can be updated on 'Browse Project' page with 'Preload in Memory' cellxgene VIP launch method via click 'Update' hyperlink.

4.3.3   Data Visualization and Analysis

CellDepot is not only a database management system, but also a web portal to visualize the scRNA-seq dataset. Here, we embed cellxgene and cellxgene VIP in CellDepot. By clicking 'Cellxgene VIP', data analysis results can be visualized. Detailed guides of cellxgene and cellxgene VIP, please go to https://github.com/chanzuckerberg/cellxgene (cellxgene) and https://github.com/interactivereport/cellxgene_VIP (cellxgene_VIP).
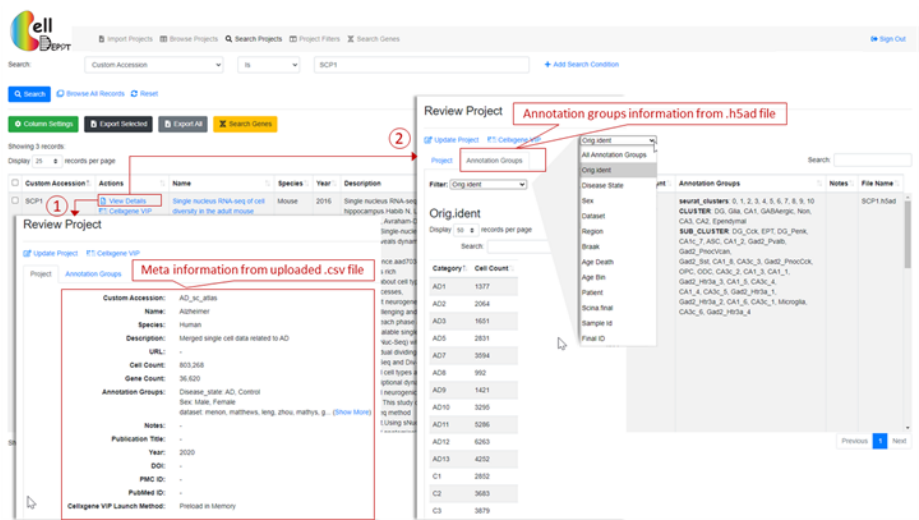
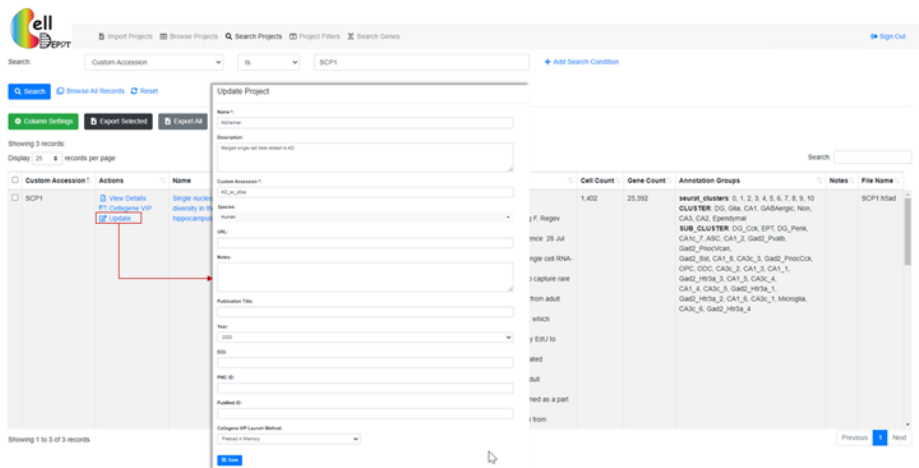Figure 4.6: Figure S9. Visualization of details of datasets.

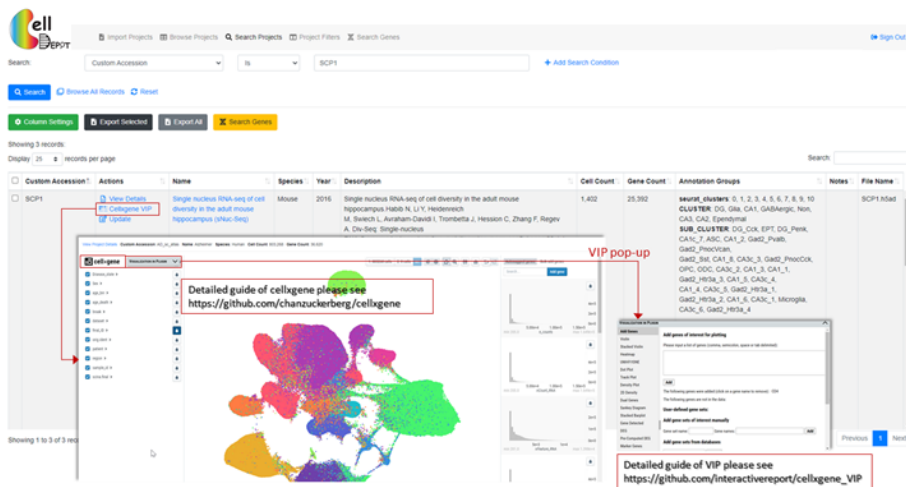Figure 4.7: Figure S10. Update project on 'Browse Project' page.

Figure 4.8: Figure S11. See visualization of selected datasets.

4.4 Search Genes

This tab allows searching on targeted genes with cell count cutoff and expression cutoff. The search outcome provides users every project contains the targeted genes. Each project displays a link to project page and a figure plot if applicable. This plot can be either violin plot or dot plot shows the gene expression level in each annotation groups.

4.5 How to set up cron job?

The following cron job entry is needed to convert h5ad file to CSC format on the background,

```
@hourly <user-name> cd /var/www/html/celldepot/app/core; php ./api_toCSCh5ad.php
```

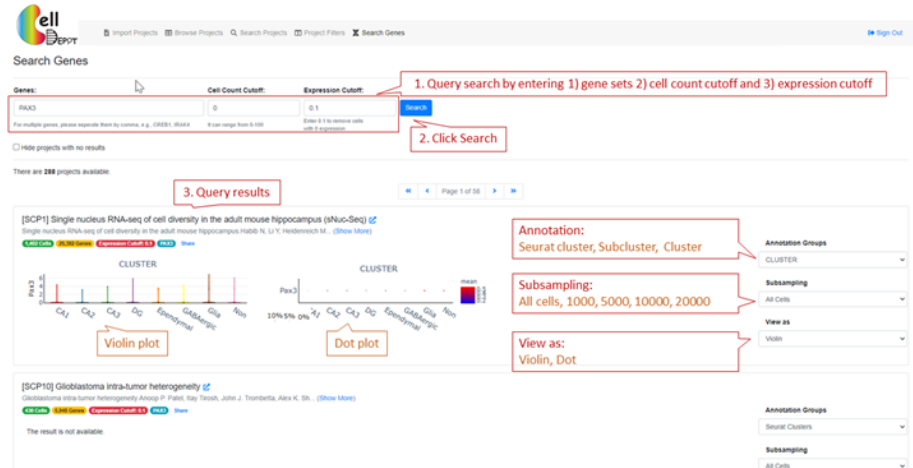Please make sure that the user has the permission to write in the data directory.

Figure 4.9: Figure S12. The layout of 'Search Genes'