

Introduction

CellDepot is database management system integrated with management system, query searching and data visualization tools [1, 2] for scRNAseq datasets, which can be accessed by the link <http://celldepot.bxgenomics.com/>. This is a supplemental tutorial provides a detailed guide.

The screenshot displays the CellDepot web interface. At the top, there is a navigation bar with the CellDepot logo and several tabs: 'Import Projects', 'Browse Projects', 'Search Projects', 'Project Filters', and 'Search Genes'. A red box highlights these tabs with the text 'Tabs different functionalities'. Below the navigation bar, there is a search section with a dropdown menu for 'Please select a field to search:', a 'Contains' dropdown, and a search input field. There are also buttons for 'Search', 'Browse All Records', and 'Reset'. Below the search section, there are buttons for 'Column Settings', 'Export Selected', 'Export All', and 'Search Genes'. The main content area shows 'Showing 288 records' and a table with columns: 'Custom Accession', 'Actions', 'Name', 'Species', 'Year', 'Description', 'Cell Count', 'Gene Count', 'Annotation Groups', 'Notes', and 'File Name'. The first row of the table is for 'AD_sc_atlas' and contains detailed information about the dataset, including 'Disease_state', 'Sex', 'dataset', 'region', 'braak', 'age_bin', 'patient', 'scina.final', and 'final_ID'.

Custom Accession	Actions	Name	Species	Year	Description	Cell Count	Gene Count	Annotation Groups	Notes	File Name
AD_sc_atlas	View Details Cellxgene VIP Update	Alzheimer	Human	2020	Merged single cell data related to AD	803,268	36,620	Disease_state: AD, Control Sex: Male, Female dataset: menon, matthews, leng, zhou, mathys, grubman region: dorsolateral_prefrontal_cortex, mTemp, Superior_frontal_gyrus, Entorhinal_cortex, anterior_prefrontal_cortex braak: 3, 4, 6, 5, 0, 2, 1, 3-4, 1-2, 5-6 age_bin: 81+, 66-80, 50-65 patient: nan, C13, AD1, AD2, AD4, AD3, C15, C16, C14, C11, AD5, AD6, C12 scina.final: Excitatory_neurons, Oligodendrocytes, Astrocytes, Microglia, Oligodendrocyte_progenitor_cells, Inhibitory_neurons, VIP_inter, unknown, Neurons_SV2C_inter, PVALB_inter, Vascular_cell, astrocyte_oligo_mix final_ID: Oligodendrocytes, Excitatory_neuron_upper, Astrocytes,		AD_sc_atlas.h5ad

The interface contains multiple tabs, corresponding to import and/or select objects in CellDepot scRNAseq 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.

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, CellxDepot 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.

1. Click Import

2. Click to download 'Example File' and modify this .csv file based on your own datasets

3. Choose the modified .csv file to upload

4. Choose 'Standard' for upload, 'Preload in memory' for update

5. Click submit button

6. Navigate to 'Browse Projects'

7. Search uploaded datasets by accession name

8. Your uploaded datasets

Custom Accession	Actions	Name	Species	Year	Description	Cell Count	Gene Count	Annotation Groups	Notes	File Name
SCP1	View Details Cellgene VIP Update	Single nucleus RNA-seq of cell diversity in the adult mouse hippocampus (sNuc-Seq)	Mouse	2016	Single nucleus RNA-seq of cell diversity in the adult mouse hippocampus. Habib N, Li Y, Heidenreich M, Swiech L, Avraham-David I, Trombetta J, Hession C, Zhang F, Regev A. Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. Science. 28 Jul 2016 DOI: 10.1126/science.1274712	1,402	25,392	seurat_clusters: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 CLUSTER: DG, Gila, CA1, GABAergic, Non, CA3, CA2, Ependymal SUB_CLUSTER: DG_Cck, EPT, DG_Penik, CA1c_7, ASC, CA1_2, Gad2_Pvalb, Gad2_PnocVcan,		SCP1.h5ad

Figure S2. Workflow of how to import personal datasets.

2. Browse Projects

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 are allowed to select the projects by 17 attributes: annotation groups, cell count, cellgene 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.

The figure consists of two screenshots of the Cell Browser web application, illustrating a search and export workflow. The top screenshot shows the 'Browse Projects' page with a search bar and a list of projects. The bottom screenshot shows the 'Column Settings' dialog box and the 'Export Selected' button.

Top Screenshot: Browse Projects Page

- 1. Click Browse Projects**: Points to the 'Browse Projects' tab in the top navigation bar.
- 2. Choose a field to search**: Points to the 'Please select a field to search:' dropdown menu.
- 3. Choose the condition**: Points to the 'Contains' dropdown menu in the search criteria.
- 4. Type the entry**: Points to the text input field for the search criteria.
- 5. Add additional conditions for multiple select statements**: Points to the '+ Add Search Condition' button.
- 6. add/or**: Points to the 'And' and 'Or' radio buttons for logical operators.
- 7. Click search**: Points to the 'Search' button.

Bottom Screenshot: Column Settings and Export Options

- 1. Click 'column setting' to customize the layout of the query**: Points to the 'Column Settings' button in the top navigation bar.
- 2. Export selected datasets or entire database**: Points to the 'Export Selected' and 'Export All' buttons.
- 3. Further cross-project comparison of gene sets among datasets (Details see Figure S3)**: Points to the 'Search Genes' button.

Figure S2. Workflow of how to search query on 'Browse Projects' page.

2.1.1 Case study 1

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

1 Query Search

2 Checkbox 18 queried datasets

3 Search Genes

4 Search multiple genes of interests

5 Study 1: Week 4-6

6 Study 2: Week 9

The screenshot displays the CellXGene web application. At the top, there's a navigation bar with 'Import Projects', 'Browse Projects', 'Search Projects', 'Project Filters', and 'Search Genes'. The main search area allows users to filter projects by 'Custom Accession', 'Cont', 'myogenic', and 'Add Search Condition'. Below this, a table lists 18 queried datasets, including 'GSE147457_adult-myogenic'. The 'Search Genes' section shows a list of genes: PAX3, PAX7, PITX2, MYF5, MYF6, MYOD1, MYOG, NEB, MYH3, MYH6, MYH7, MYH8, MYH9, MYH10, MYH11, MYH12, MYH13, MYH14, MYH15, MYH16, MYH17, MYH18, MYH19, MYH20, MYH21, MYH22, MYH23, MYH24, MYH25, MYH26, MYH27, MYH28, MYH29, MYH30, MYH31, MYH32, MYH33, MYH34, MYH35, MYH36, MYH37, MYH38, MYH39, MYH40, MYH41, MYH42, MYH43, MYH44, MYH45, MYH46, MYH47, MYH48, MYH49, MYH50, MYH51, MYH52, MYH53, MYH54, MYH55, MYH56, MYH57, MYH58, MYH59, MYH60, MYH61, MYH62, MYH63, MYH64, MYH65, MYH66, MYH67, MYH68, MYH69, MYH70, MYH71, MYH72, MYH73, MYH74, MYH75, MYH76, MYH77, MYH78, MYH79, MYH80, MYH81, MYH82, MYH83, MYH84, MYH85, MYH86, MYH87, MYH88, MYH89, MYH90, MYH91, MYH92, MYH93, MYH94, MYH95, MYH96, MYH97, MYH98, MYH99, MYH100. The 'Study 1: Week 4-6' dot plot shows gene expression across various cell types (MB, MC, Mp, SKM, etc.) and developmental stages (A_Emb_Wk05.0_AP32, A_Emb_Wk06.0_AP13, A_Emb_Wk06.0_AP37, A_Emb_Wk06.0_AP5). The 'Study 2: Week 9' dot plot shows gene expression across various cell types (MB, MC, Mp, SKM, etc.) and developmental stages (A_Emb_Wk09.0_AP32, A_Emb_Wk09.0_AP37, A_Emb_Wk09.0_AP5).

Figure S3. Workflow of how to conduct the cross-project comparison of gene sets among the selected datasets.

For each project, users can view the datasets information, visualize data analysis and conduct update through View Details, cellxgene VIP and Update, separately.

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 S2).

The screenshot shows the 'Project Filters' page in the CellDepot application. The sidebar on the left contains filters for 'Year' and 'Species'. The main content area displays a table of 288 records. The table has columns: Custom Accession, Actions, Name, Species, Year, Description, Cell Count, Gene Count, Annotation Groups, and Notes. A search bar is located at the top right. A dropdown menu is open over the 'Description' column, showing a list of fields to search: DOI, File Name, File Size, Gene Count, Name, Notes, PMCID, Publication Title, PubMed ID, Species, URL, and Year. A red arrow points from the 'Advanced Search' button in the sidebar to the search bar at the top right.

Figure S4. The layout of 'Project Filters' page.

3. Visualize the datasets

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 from uploaded .h5ad file.

Cellxgene EPoT

Search: Custom Accession | Is | SCP1 | [Add Search Condition](#)

[Search](#) | [Browse All Records](#) | [Reset](#)

[Column Settings](#) | [Export Selected](#) | [Export All](#) | [Search Genes](#)

Showing 3 records: Display 25 records per page

Custom Accession	Actions	Name	Species	Year	Description
SCP1	View Details Cellxgene VIP	Single nucleus RNA-seq of cell diversity in the adult mouse	Mouse	2016	Single nucleus RNA-seq of cell diversity in the adult mouse

Review Project

[Update Project](#) | [Cellxgene VIP](#)

Project | Annotation Groups

Filter: Orig.ident

Display 50 records per page

Search:

Orig.ident

Category: Cell Count

Category	Cell Count
AD1	1377
AD2	2064
AD3	1651
AD5	2831
AD7	3594
AD8	992
AD9	1421
AD10	3295
AD11	5286
AD12	6263
AD13	4252
C1	2852
C2	3683
C3	3879

Annotation Groups information from .h5ad file

Orig.ident

All Annotation Groups

Orig.ident

Disease State

Sex

Dataset

Region

Braak

Age Death

Age Bin

Patient

Scina final

Sample Id

Final ID

Meta information from uploaded .csv file

Custom Accession: AD_sc_atlas

Name: Alzheimer

Species: Human

Description: Merged single cell data related to AD

URL: -

Cell Count: 803,268

Gene Count: 36,620

Annotation Groups: Disease_state: AD, Control; Sex: Male, Female; dataset: menon, mathews, leng, zhou, mathys, g... (Show More)

Notes: -

Publication Title: -

Year: 2020

DOI: -

PMC ID: -

PubMed ID: -

Cellxgene VIP Launch Method: Preload in Memory

Figure S5. Visualization of details of datasets.

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.

Cellxgene EPoT

Search: Custom Accession | Is | SCP1 | [Add Search Condition](#)

[Search](#) | [Browse All Records](#) | [Reset](#)

[Column Settings](#) | [Export Selected](#) | [Export All](#) | [Search Genes](#)

Showing 3 records: Display 25 records per page

Custom Accession	Actions	Name
SCP1	View Details Cellxgene VIP Update	Single nucleus RNA-seq of cell diversity in the hippocampus

Update Project

Name: Alzheimer

Description: Merged single cell data related to AD

Custom Accession: AD_sc_atlas

Species: Human

URL: -

Notes: -

Publication Title: -

Year: 2020

DOI: -

PMC ID: -

PubMed ID: -

Cellxgene VIP Launch Method: Preload in Memory

[Save](#)

Table Data:

	Cell Count	Gene Count	Annotation Groups	Notes	File Name
SCP1	1,402	25,392	seurat_clusters: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 CLUSTER: DG, Gila, CA1, GABAergic, Non, CA3, CA2, Ependymal SUB_CLUSTER: DG_Cck, EPT, DG_Penk, CA1c_7, ASC, CA1_2, Gad2_Pvalb, Gad2_PnocVcan, Gad2_Sst, CA1_8, CA3c_3, Gad2_PnocCck, OPC, ODC, CA3c_2, CA1_3, CA1_1, Gad2_Htr3a_3, CA1_5, CA3c_4, CA1_4, CA3c_5, Gad2_Htr3a_1, Gad2_Htr3a_2, CA1_6, CA3c_1, Microglia, CA3c_6, Gad2_Htr3a_4		SCP1.h5ad

Figure S5. Update project on 'Browse Project' page.

3.3 Data visualization and analysis

CellDepot is not only a database management system, but also a web portal to visualize the scRNAseq 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).

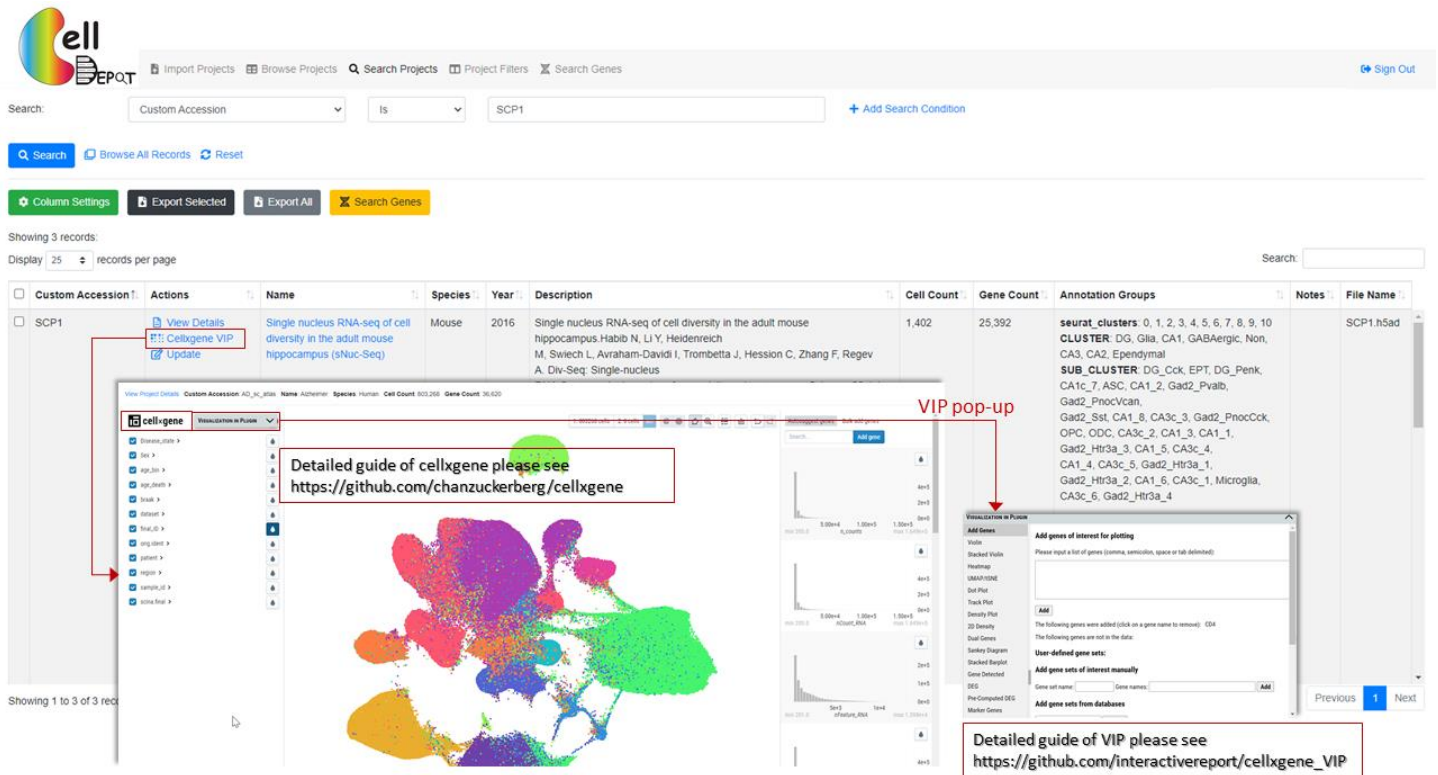


Figure S6. See visualization of selected datasets.

3. Search gene

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.

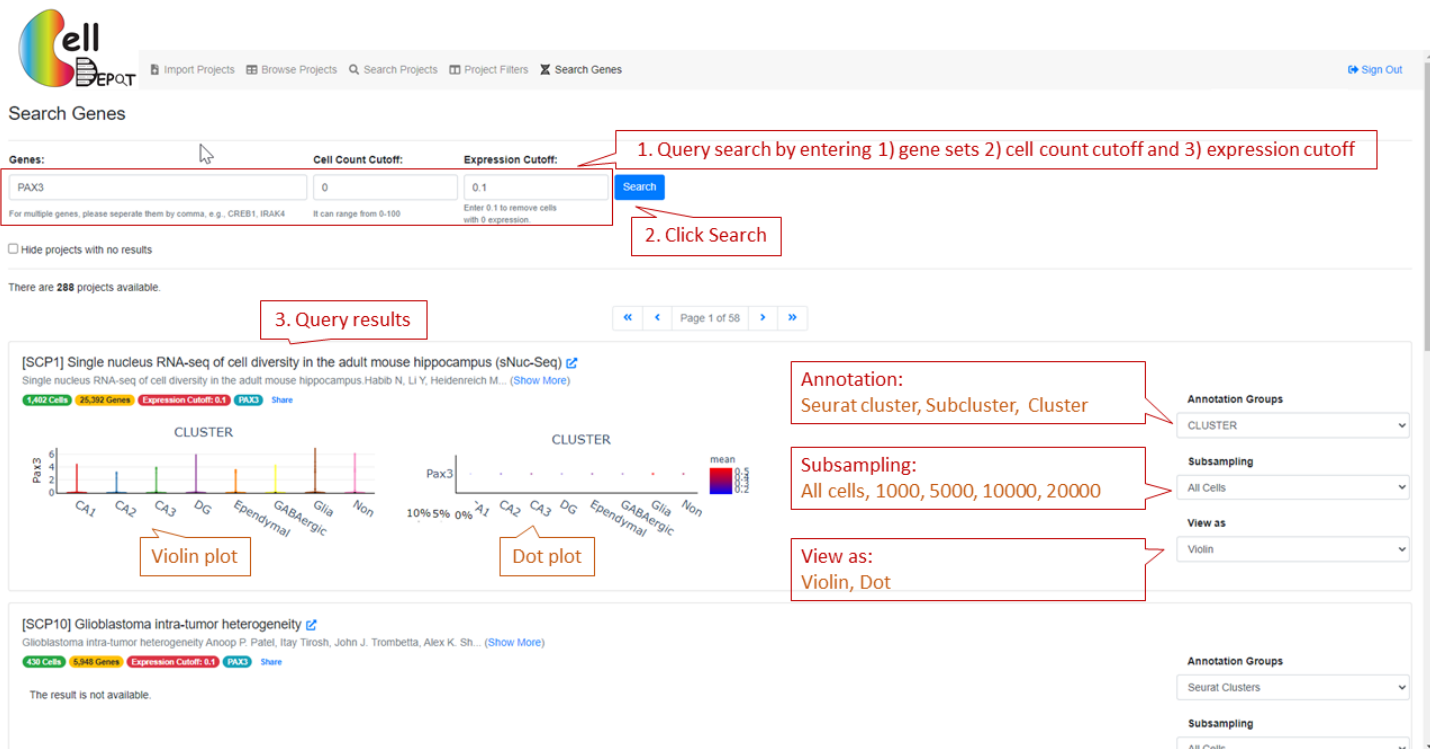


Figure S7. The layout of 'Search Genes'

4. Cron job setting-up

- [1] Li K, Ouyang Z, Lin D, Mingueneau M, Chen W, Sexton D, et al. cellxgene VIP unleashes full power of interactive visualization, plotting and analysis of scRNA-seq data in the scale of millions of cells. 2020:2020.08.28.270652.
- [2] McGill C, Martin B, Weaver C, Bell S, Prins L, Badajoz S, et al. cellxgene: a performant, scalable exploration platform for high dimensional sparse matrices. 2021:2021.04.05.438318.