Supplementary Materials - CellDepot: A unified repository for scRNAseq data and visual exploration

Dongdong Lin[1] Yirui Chen[2] Soumya Negi Derrick Cheng Zhengyu Ouyang
David Sexton Kejie Li*[3] Baohong Zhang*[4]

2021-12-10

---

[1] These two authors contributed equally to this work

[2] These two authors contributed equally to this work

[3] Cooresponding author, email: kejie.li@biogen.com

[4] Cooresponding author, email: baohong.zhang@biogen.com

Contents

Chapter 1

Preface

This is a *supplementary materials* written in Markdown, which provides the detailed guide for CellDepot web portal.

Chapter 2

Getting start with CellDepot

2.1 Sources of annotation and metadata

The original metadata information of each scRNA-seq dataset is retrieved from h5ad file, which is a preferred way of sharing and storing an on-disk representation of anndata object. When importing the dataset to the system, user inputs additional metadata information as shown in (4.6). Both metadata are collected and stored in a MySQL database table that is presented at http://celldepot.bxgenomics.com and Biogen internal instance, http://go.biogen.com/CellDepot.

2.2 Data format, availability, and preparation

CellDepot requires scRNA-seq data in h5ad file where the expression matrix is stored in CSC (compressed sparse column) instead of CSR (compressed sparse row) format to improve the speed of data retrieving. For example, designating genes as columns in the h5ad file creates the interactive plot five times faster than as rows. Just in case, we provide sample scripts to help users generate h5ad files. Having gene expression matrix, metadata, and layout files, users can easily combine and convert their data to h5ad file by following this R script on https://github.com/interactivereport/CellDepot/blob/main/toH5ad.R. In the case of lacking layout file, users can also create h5ad file by following the Jupyter notebook https://github.com/interactivereport/CellDepot/blob/main/raw2h5ad.ipynb with custom python script tailored to their own data. Categorical features extracted from a h5ad file are shown in the 'annotation groups' column of the table on CellDepot home page, while the numerical features are shown as the histograms in the rightmost panel on cellxgene VIP. (4.4.2)

2.3 CellDepot platform and installation

The public version of CellDepot web portal is hosted at the web site, http://celldepot.bxgenomics.com and Biogen internal link http://go.biogen.com/CellDepot. It is implemented with MySQL database, an advanced search engine, and powerful interactive visualizing tools that allow users to explore attributes of datasets as well as scRNA-seq analysis results. Also, users can intentionally select single-cell RNA-seq datasets on the web interface by simply browsing the online dataset table or applying advanced search to perform the cross-dataset comparison. Moreover, CellDepot also provides comprehensive data analysis tools via an embedded interactive visualization plugin. To host private datasets, local instance of CellDepot on Unix server can be installed by following the guide here, https://celldepot.bxgenomics.com/celldepot_manual/install_environment.php.

2.4 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
```

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

2.5 CellDepot API (Application Programming Interface)

The CellDepot API web service provides a direct way to generate figures for users to share or embed in web page. For example, the following URL will generate a gene expression violin plot across cell clusters for IRAK4 gene for the data set with ID equaling one, https://celldepot.bxgenomics.com/celldepot/app/core/api_gene_plot.php?ID=1&Genes=IRAK4&Plot_Type=violin&Subsampling=0&n=0&g=0&Project_Group=CLUSTER. The complete format of the URL and explanation of parameters are detailed in the online documentation, https://celldepot.bxgenomics.com/celldepot_manual/api_gene_plot.php.

2.6 Code availability

The source code, links to tutorials and other supplementary documents are provided at https://github.com/interactivereport/CellDepot. With broad adoption and contribution in mind, CellDepot is released under the MIT open-source license. The detailed instruction of local installation is available at https://celldepot.bxgenomics.com/celldepot_manual.

2.7 Online tutorials

To better assist biologists to use CellDepot and integrated cellxgene VIP visual analytical tool, we created online easy-to-access HTML tutorials with step-by-step guides available at https://interactivereport.github.io/CellDepot/bookdown/docs/SITutorial.html for CellDepot and https://interactivereport.github.io/cellxgene_VIP/tutorial/docs/how-to-use-cellxgene-vip.html for cellxgene VIP, respectively. In addition, a question mark next to the title of each VIP function module provides a direct way to reach the corresponding section of the HTML document for help.

Chapter 3

Supplemental Tables

Table S1 - Comparison matrix of web portal tools

Table link: https://github.com/interactivereport/CellDepot/blob/gh-pages/bookdown/S1.csv

Web application repository	**CellDepot**	**Corpora Data Portal**	**gEAR**	**CHARTS**	**SCANNER**	**Single Cell Portal**	**Sfaria**	**Repro-Genomics**	**PanglaoDB**	**Expression Atlas**	**scRNA-SeqDB**	**conquer**	**Jingle-Bells**	**Human Cell ATLAS**
Year	2021	2021	2021	2020	2020	2020	2020	2019	2019	2019	2019	2018	2017	2017
Main function														
Database Explorer	Y	Y	Y	Y	NA	Y	Y	Y	Y	Y	Y	Y	Y	Y
Query Search	Advanced	NA	Basic I	NA	Intermediate	Advanced	Basic II	Intermediate	Advanced	Intermediate	Basic I	Basic I	Basic I	Advanced
Data Analysis Explorer	Advanced	Advanced	Advanced	Intermediate	Intermediate	Intermediate - Advanced	NA	Basic	Intermediate	Intermediate	Basic - Intermediate	Basic	NA	NA
Visualization features	Violin, dot, heatmap, bar, QC, scatter, density, embedding plots	Bar, scatter, embedding plots	Violin, dot, line, bar, QC, embedding plots, genome browser	Bar, embedding plots	Violin, dot, scatter, heatmap, embedding plots	Violin, scatter, heatmap, embedding plots	NA	Violin, density, scatter plots, genome browser	Bar, QC, scatter, embedding plots	Heatmap, scatter, embedding plots	Scatter, bar plots	Scatter, QC plots	Genome browser	NA
Scalability and Capacity														
Time to interactive	3.5 s	0.9 s	5.5 s	13.2 s	5.5 s	5.6 s	0.8 s	4.7 s	0.8 s	1.4 s	1.3 s	1.9 s	1.1 s	1.1 s
Total blocking time	30 ms	50 ms	470 ms	10,020 ms	60 ms	100 ms	0 ms	300 ms	0 ms	40 ms	0 ms	100 ms	0 ms	0 ms
Cumulative layout shift	0.033	0.596	0.733	0.329	0.004	0.04	0.162	0	0.001	0.281	0	0.127	0.015	0.001
Memory footprint	36,784K	33,252K	42,032K	559,568K	47,800K	73,188K	25,964K	45,696K	32,036K	32,912K	26,536K	34,380K	45,160K	49,472K
Instant MAX CPU	32.7	15.6	54.2	237.3	49.8	144.1	29.6	51.1	7.7	74.0	26.5	35.9	26.5	21.9
Total number of datasets	**270**	**174**	**163**	**21**	**52**	**387 (*)**	**177**	**140**	**237 (1368)**	**229**	**38**	**40**	**120**	**165**
Data type supported														
Datasets	diverse	diverse	hearing/ brain	tumor	diverse	diverse	diverse	reproduction	diverse	diverse	diverse	diverse	immune-related	diverse
Datatypes	Single-cell	Single-cell	Single-cell, Epigenetics	Single-cell	Single-cell	Single-cell	Single-cell	Single-cell, Multi-omics	Single-cell	Single-cell, Protemics	Single-cell	Single-cell	Single-cell	Single-cell, Multi-omics
Links														
Source Code link	https://github.com/interactivereport/CellDepot	https://github.com/chanzuckerberg/corpora-data-portal	https://github.com/IGS/gEAR	https://github.com/stewart-lab/CHARTS	https://github.com/GuoshuaiCai/scanner	https://github.com/broadinstitute/single_cell_portal_core	https://github.com/theislab/sfaira-portal	https://github.com/fchalmel/RGV	https://github.com/oscar-franzen/PanglaoDB	https://github.com/ebi-gene-expression-group/atlas	NA	https://github.com/csoneson/conquer_comparison	https://drive.google.com/drive/folders/0BxSFjdiDhUI1amNoSks0SmpMdE0?resourcekey=0-s7Qw02gR8VhkS7hfdE9nPg	https://github.com/HumanCellAtlas/
Demo link	http://celldepot.bxgenomics.com/	https://cellxgene.cziscience.com/	umgear.org	https://charts.morgridge.org	https://www.thecailab.com/scanner/	https://singlecell.broadinstitute.org/single_cell	https://theislab.github.io/sfaira-portal/	https://rgv.genouest.org/	https://panglaodb.se	https://www.ebi.ac.uk/gxa/sc/home	https://bioinfo.uth.edu/scRNA-seqdb/	http://imlspenticton.uzh.ch:3838/conquer/	https://jinglebells.bgu.ac.il/	https://data.humancellatlas.org/

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

Table S2 - Criterion for query search

Table link: https://github.com/interactivereport/CellDepot/blob/gh-pages/bookdown/S2.csv

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

Table S3 - Criterion for data analysis explorer

Table link: https://github.com/interactivereport/CellDepot/blob/gh-pages/bookdown/S3.csv

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

Table S4 - Project metadata captured in CellDepot

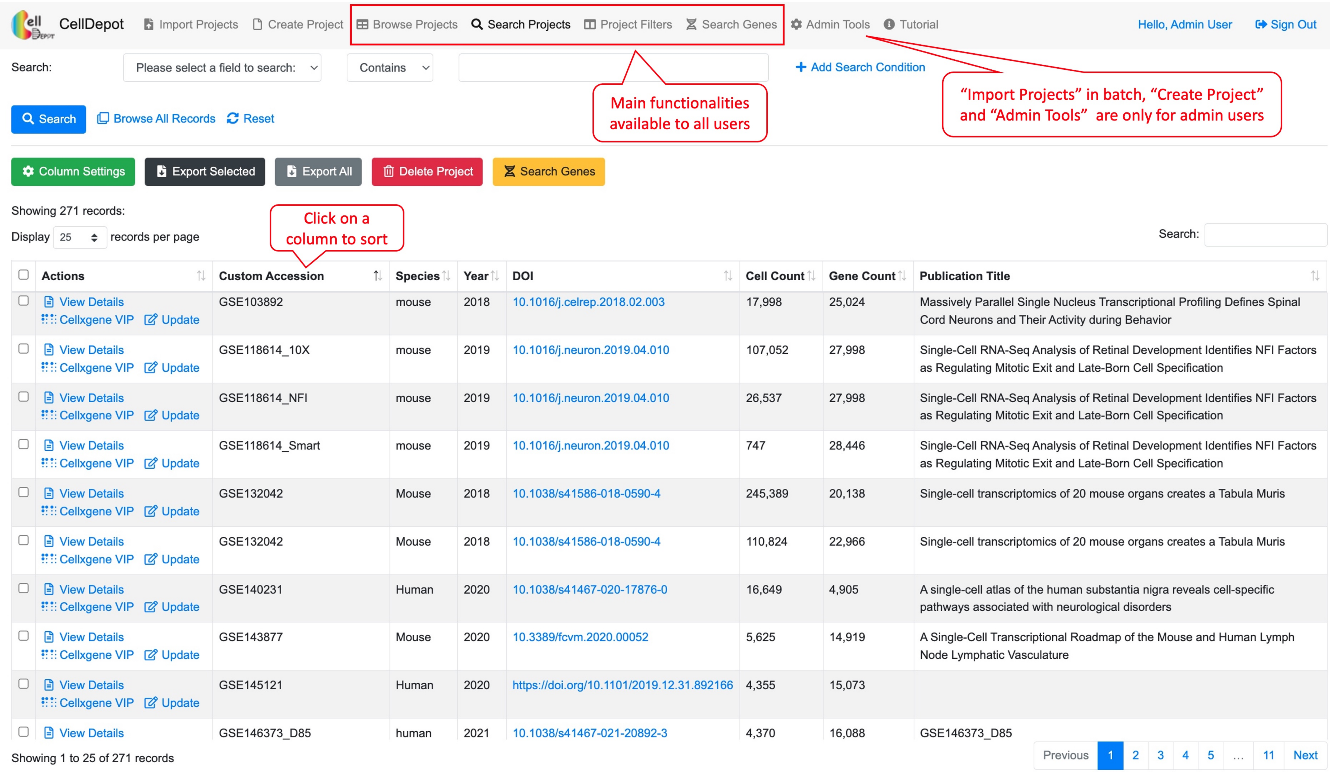
Table link: https://github.com/interactivereport/CellDepot/blob/gh-pages/bookdown/S4.csv

General.Category	Expected.Variable.Type	Description
Annotation Groups	String	categorical features from h5ad file
Cell Count	Integer	numbers of cell in study
Actions	Link	three options: 1) Study summary information; 2) Data visualization and analysis; 3) Update project information
Custom Accession	String	Customized accession name for individual project
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 4

Supplemental Tutorial

CellDepot is a scRNA-seq data portal consisting of a relational database management system, a graphical query builder, and data visualization tools, which can be accessed via the link, http://celldepot.bxgenomics.com for public datasets or a link to private installation, e.g., http://go.biogen.com/CellDepot for Biogen internal data collection. This is the supplemental tutorial providing detailed instructions. Clicking on a figure will bring up the enlarged view.



Main functionalities available to all users

"Import Projects" in batch, "Create Project" and "Admin Tools" are only for admin users

Click on a column to sort

Showing 271 records:

☐	Actions	Custom Accession	Species	Year	DOI	Cell Count	Gene Count	Publication Title
☐	View Details Cellxgene VIP Update	GSE103892	mouse	2018	10.1016/j.celrep.2018.02.003	17,998	25,024	Massively Parallel Single Nucleus Transcriptional Profiling Defines Spinal Cord Neurons and Their Activity during Behavior
☐	View Details Cellxgene VIP Update	GSE118614_10X	mouse	2019	10.1016/j.neuron.2019.04.010	107,052	27,998	Single-Cell RNA-Seq Analysis of Retinal Development Identifies NFI Factors as Regulating Mitotic Exit and Late-Born Cell Specification
☐	View Details Cellxgene VIP Update	GSE118614_NFI	mouse	2019	10.1016/j.neuron.2019.04.010	26,537	27,998	Single-Cell RNA-Seq Analysis of Retinal Development Identifies NFI Factors as Regulating Mitotic Exit and Late-Born Cell Specification
☐	View Details Cellxgene VIP Update	GSE118614_Smart	mouse	2019	10.1016/j.neuron.2019.04.010	747	28,446	Single-Cell RNA-Seq Analysis of Retinal Development Identifies NFI Factors as Regulating Mitotic Exit and Late-Born Cell Specification
☐	View Details Cellxgene VIP Update	GSE132042	Mouse	2018	10.1038/s41586-018-0590-4	245,389	20,138	Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris
☐	View Details Cellxgene VIP Update	GSE132042	Mouse	2018	10.1038/s41586-018-0590-4	110,824	22,966	Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris
☐	View Details Cellxgene VIP Update	GSE140231	Human	2020	10.1038/s41467-020-17876-0	16,649	4,905	A single-cell atlas of the human substantia nigra reveals cell-specific pathways associated with neurological disorders
☐	View Details Cellxgene VIP Update	GSE143877	Mouse	2020	10.3389/fcvm.2020.00052	5,625	14,919	A Single-Cell Transcriptional Roadmap of the Mouse and Human Lymph Node Lymphatic Vasculature
☐	View Details Cellxgene VIP Update	GSE145121	Human	2020	https://doi.org/10.1101/2019.12.31.892166	4,355	15,073	
☐	View Details	GSE146373_D85	human	2021	10.1038/s41467-021-20892-3	4,370	16,088	GSE146373_D85

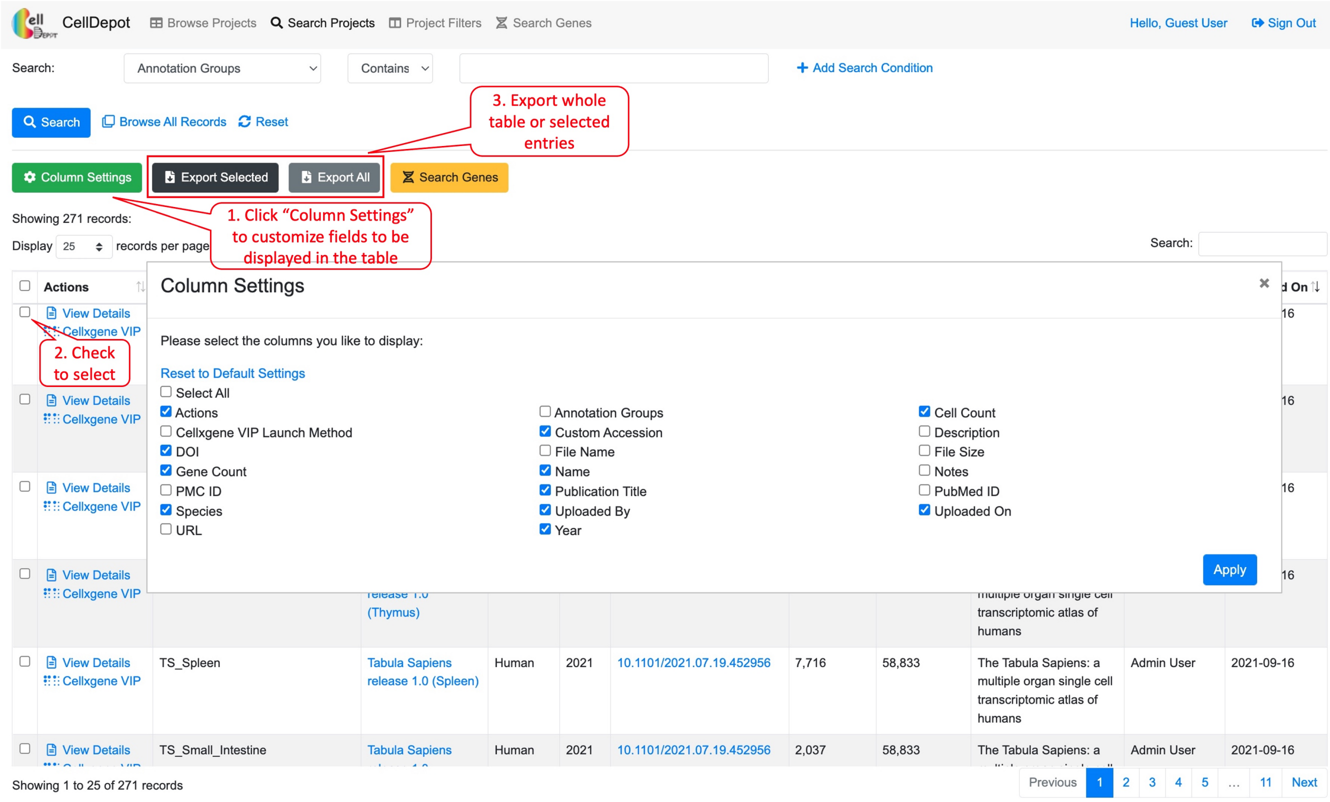
Showing 1 to 25 of 271 records

Figure S1. CellDepot Homepage. Functions designated to different user roles are highlighted in red callout boxes.

The interface contains multiple tabs at the top of the homepage, which correspond to major functionalities of CellDepot. All users can explore the existing datasets loaded in the public CellDepot for visualization and analysis while only admin users can upload datasets to a CellDepot instance, public or private.

4.1 Browse Projects

Users can customize columns to be displayed by clicking on the green 'Column Settings' button. In addition, all or selected entries could be exported in CSV format. A quick search box is provided on the top-right corner of the table while building a complex query is exemplified in Figure S3 (4.2).



1. Click "Column Settings" to customize fields to be displayed in the table

2. Check to select

3. Export whole table or selected entries

Column Settings

Please select the columns you like to display:

Reset to Default Settings

☐ Select All

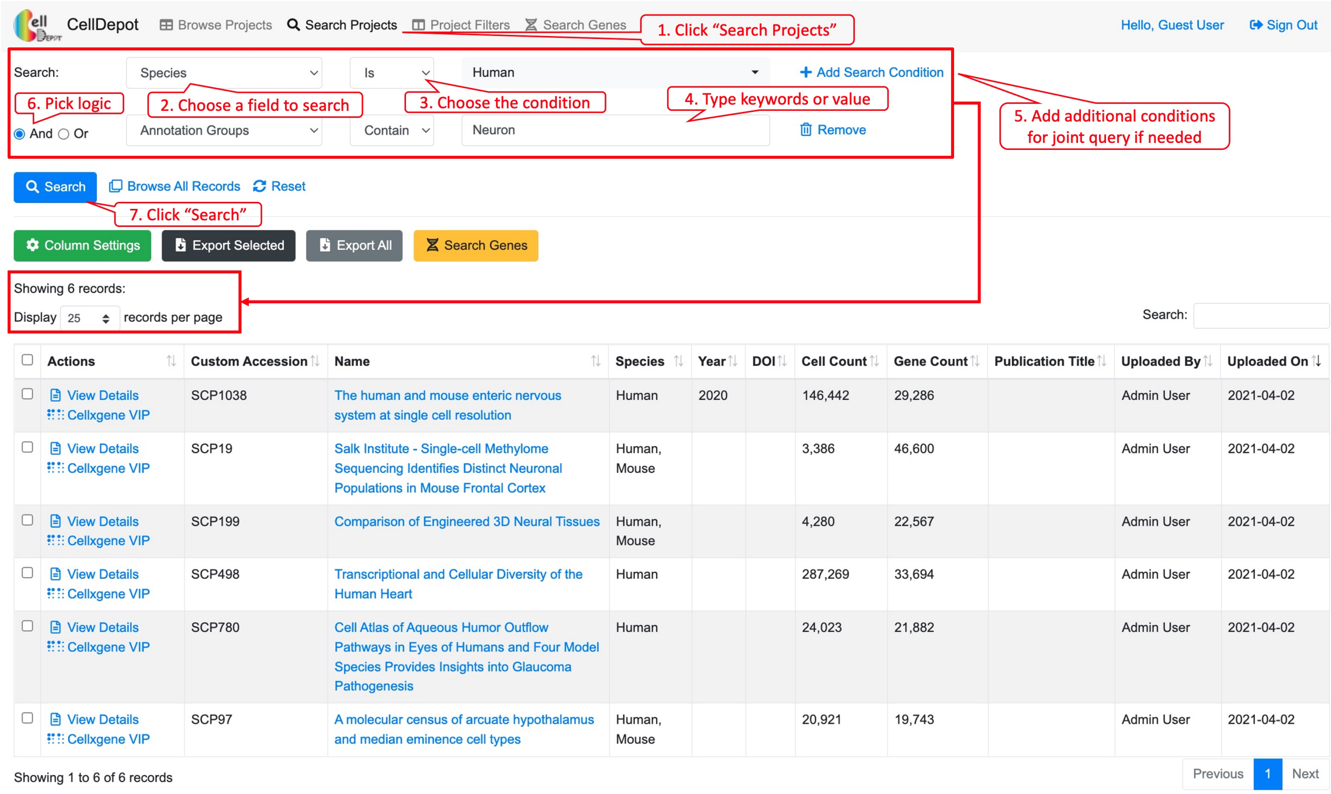
☑ Actions	☐ Annotation Groups	☑ Cell Count
☐ Cellxgene VIP Launch Method	☑ Custom Accession	☐ Description
☑ DOI	☐ File Name	☐ File Size
☑ Gene Count	☑ Name	☐ Notes
☐ PMC ID	☑ Publication Title	☐ PubMed ID
☑ Species	☑ Uploaded By	☑ Uploaded On
☐ URL	☑ Year	

Apply

Figure S2. Browsing projects in a personalized view.

4.2 Search Projects

This function allows users to search projects of interest, which can be accessed through the homepage as well. Users can search projects by 17 attributes in multiple logic conditions: 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, and/or Year.



1. Click "Search Projects"

Search: Species Is Human

● And ○ Or Annotation Groups Contain Neuron

2. Choose a field to search

3. Choose the condition

4. Type keywords or value

5. Add additional conditions for joint query if needed

6. Pick logic

7. Click "Search"

Showing 6 records:

Display 25 records per page

Search:

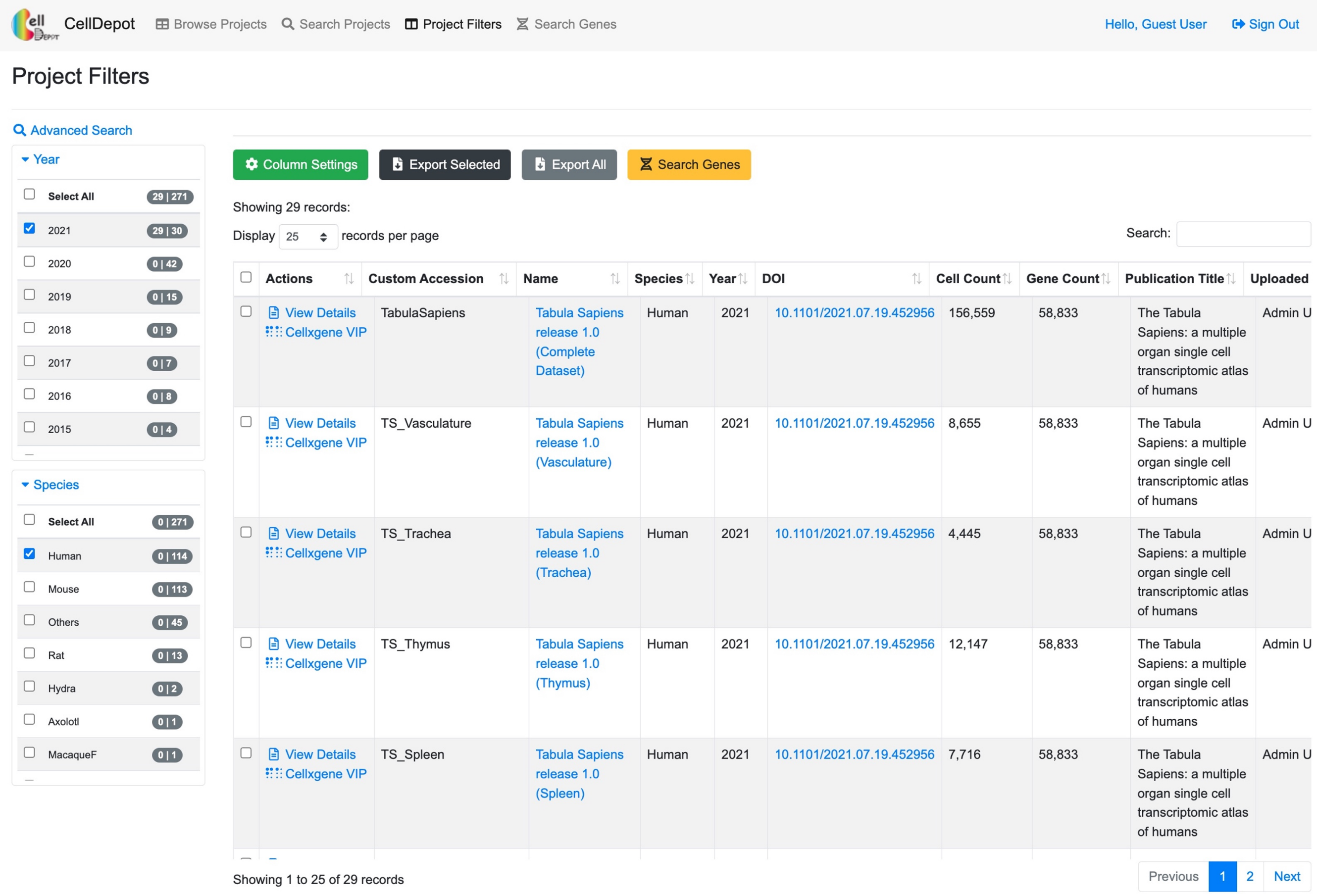
☐	Actions	Custom Accession	Name	Species	Year	DOI	Cell Count	Gene Count	Publication Title	Uploaded By	Uploaded On
☐	View Details Cellxgene VIP	SCP1038	The human and mouse enteric nervous system at single cell resolution	Human	2020		146,442	29,286		Admin User	2021-04-02
☐	View Details Cellxgene VIP	SCP19	Salk Institute - Single-cell Methylome Sequencing Identifies Distinct Neuronal Populations in Mouse Frontal Cortex	Human, Mouse			3,386	46,600		Admin User	2021-04-02
☐	View Details Cellxgene VIP	SCP199	Comparison of Engineered 3D Neural Tissues	Human, Mouse			4,280	22,567		Admin User	2021-04-02
☐	View Details Cellxgene VIP	SCP498	Transcriptional and Cellular Diversity of the Human Heart	Human			287,269	33,694		Admin User	2021-04-02
☐	View Details Cellxgene VIP	SCP780	Cell Atlas of Aqueous Humor Outflow Pathways in Eyes of Humans and Four Model Species Provides Insights into Glaucoma Pathogenesis	Human			24,023	21,882		Admin User	2021-04-02
☐	View Details Cellxgene VIP	SCP97	A molecular census of arcuate hypothalamus and median eminence cell types	Human, Mouse			20,921	19,743		Admin User	2021-04-02

Showing 1 to 6 of 6 records

Figure S3. Workflow of searching projects by using the graphical multi-logic, multi-condition query builder. Six datasets are identified when searching by 'Species is Human' and 'Annotation Group contains Neuron'.

4.3 Project Filters

This function lists the filtered datasets simply based on AND logic operation of checked items under various categories. It is a friendly feature for first-time users as they may not be familiar with the design of the database to construct more complex query.



Project Filters

Advanced Search

Year

☐	Select All	29 | 271
☑	2021	29 | 30
☐	2020	0 | 42
☐	2019	0 | 15
☐	2018	0 | 9
☐	2017	0 | 7
☐	2016	0 | 8
☐	2015	0 | 4

Species

☐	Select All	0 | 271
☑	Human	0 | 114
☐	Mouse	0 | 113
☐	Others	0 | 45
☐	Rat	0 | 13
☐	Hydra	0 | 2
☐	Axolotl	0 | 1
☐	MacaqueF	0 | 1

Showing 29 records:

Display 25 records per page

Search:

☐	Actions	Custom Accession	Name	Species	Year	DOI	Cell Count	Gene Count	Publication Title	Uploaded
☐	View Details Cellxgene VIP	TabulaSapiens	Tabula Sapiens release 1.0 (Complete Dataset)	Human	2021	10.1101/2021.07.19.452956	156,559	58,833	The Tabula Sapiens: a multiple organ single cell transcriptomic atlas of humans	Admin U
☐	View Details Cellxgene VIP	TS_Vasculature	Tabula Sapiens release 1.0 (Vasculature)	Human	2021	10.1101/2021.07.19.452956	8,655	58,833	The Tabula Sapiens: a multiple organ single cell transcriptomic atlas of humans	Admin U
☐	View Details Cellxgene VIP	TS_Trachea	Tabula Sapiens release 1.0 (Trachea)	Human	2021	10.1101/2021.07.19.452956	4,445	58,833	The Tabula Sapiens: a multiple organ single cell transcriptomic atlas of humans	Admin U
☐	View Details Cellxgene VIP	TS_Thymus	Tabula Sapiens release 1.0 (Thymus)	Human	2021	10.1101/2021.07.19.452956	12,147	58,833	The Tabula Sapiens: a multiple organ single cell transcriptomic atlas of humans	Admin U
☐	View Details Cellxgene VIP	TS_Spleen	Tabula Sapiens release 1.0 (Spleen)	Human	2021	10.1101/2021.07.19.452956	7,716	58,833	The Tabula Sapiens: a multiple organ single cell transcriptomic atlas of humans	Admin U

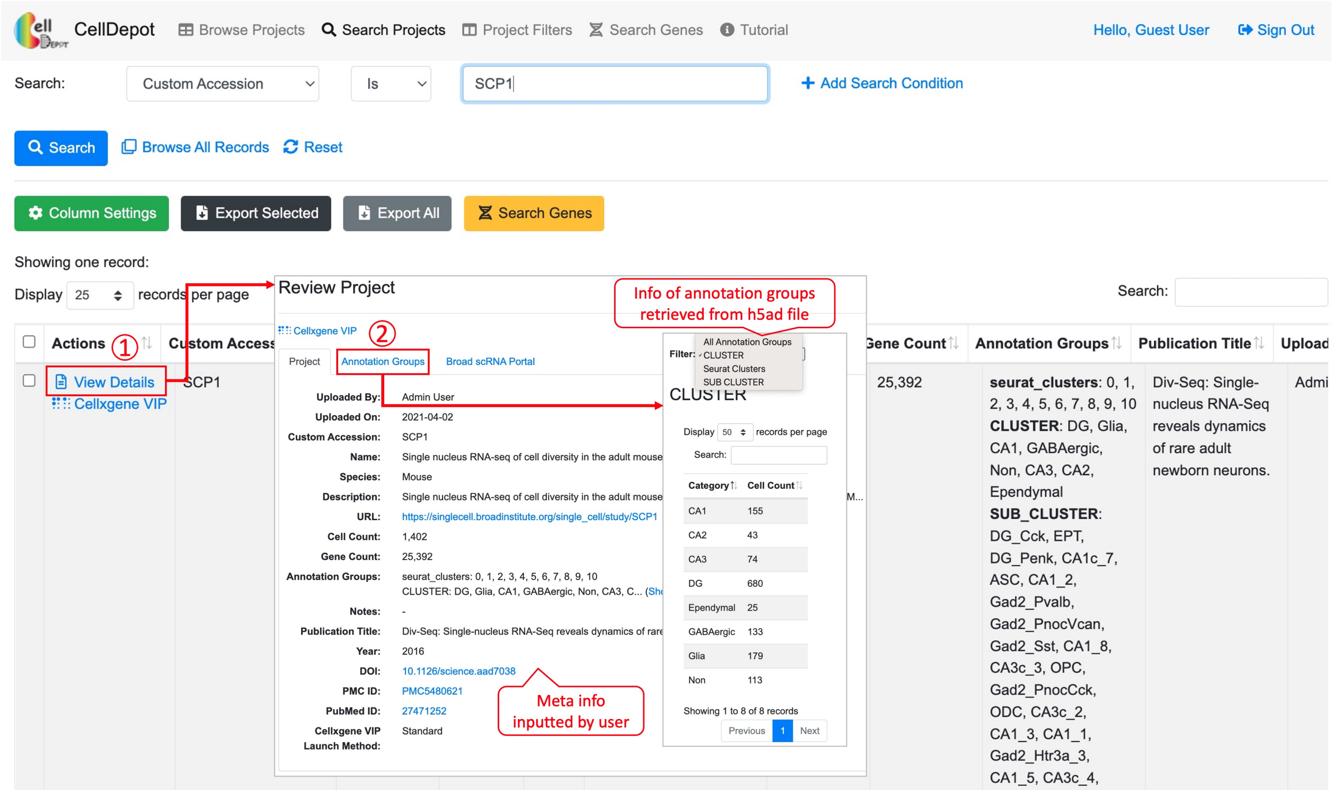
Showing 1 to 25 of 29 records

Figure S4. The ‘Project Filters’ page. 29 records are identified by filtering criteria of ‘Year’ equaling 2021 and ‘Species’ being human.

4.4 Visualize Datasets

4.4.1 View Details

The dataset information consists of project summary and annotation groups. The project summary is provided by admin users when uploading projects while annotation groups are retrieved from uploaded h5ad files.



Showing one record:

SCP1 | 25,392 | **seurat_clusters**: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 **CLUSTER**: DG, Glia, 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, OPC, Gad2_PnocCck, ODC, CA3c_2, CA1_3, CA1_1, Gad2_Htr3a_3, CA1_5, CA3c_4, | Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. | Admi

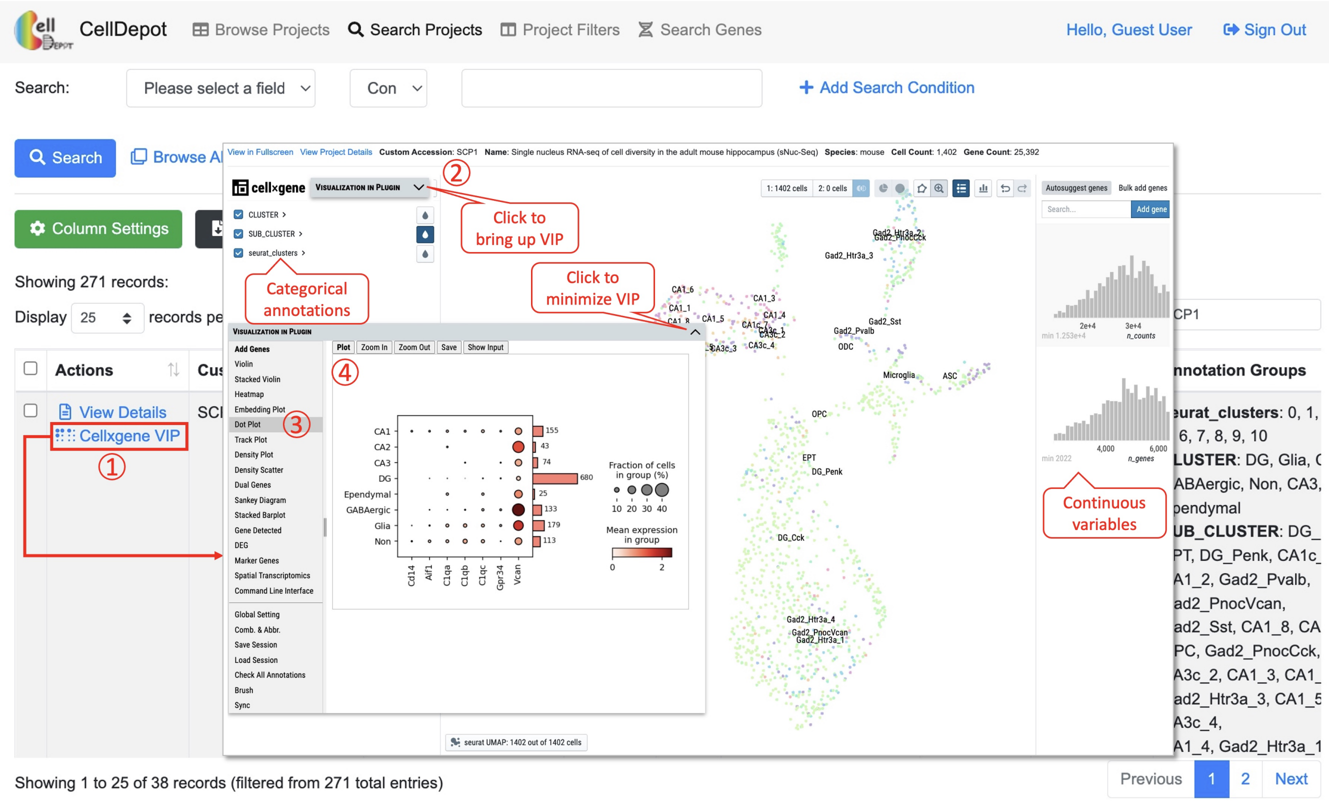
Uploaded By: Admin User
Uploaded On: 2021-04-02
Custom Accession: SCP1
Name: Single nucleus RNA-seq of cell diversity in the adult mouse
Species: Mouse
Description: Single nucleus RNA-seq of cell diversity in the adult mouse
URL: https://singlecell.broadinstitute.org/single_cell/study/SCP1
Cell Count: 1,402
Gene Count: 25,392
Annotation Groups: seurat_clusters: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
CLUSTER: DG, Glia, CA1, GABAergic, Non, CA3, C... (Sh
Notes: -
Publication Title: Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare
Year: 2016
DOI: 10.1126/science.aad7038
PMC ID: PMC5480621
PubMed ID: 27471252
Cellxgene VIP Launch Method: Standard

Category	Cell Count
CA1	155
CA2	43
CA3	74
DG	680
Ependymal	25
GABAergic	133
Glia	179
Non	113

Figure S5. How to view details of a dataset. Steps are outlined by red circled numbers.

4.4.2 Data Visualization and Analysis by cellxgene VIP

CellDepot is not only a database management system, but also a web portal for visualizing and analyzing scRNA-seq datasets through embedded cellxgene VIP tool. By clicking ‘cellxgene VIP’ to access functional modules on the menu, users can perform advanced data visualization and analysis. To learn how to use cellxgene VIP, please go to https://interactivereport.github.io/cellxgene_VIP/tutorial/docs/how-to-use-cellxgene-vip.html.



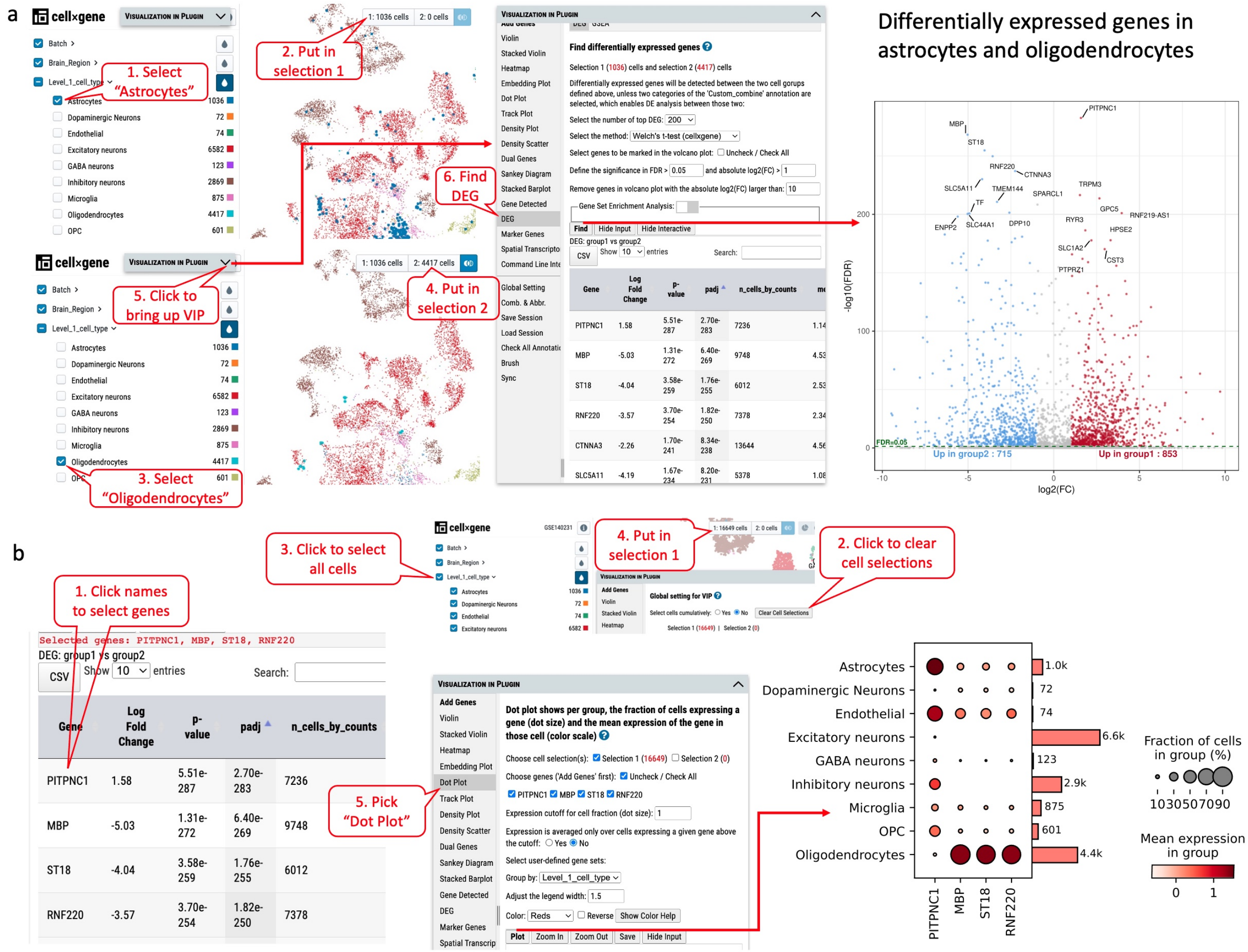
View in Fullscreen View Project Details **Custom Accession**: SCP1 **Name**: Single nucleus RNA-seq of cell diversity in the adult mouse hippocampus (sNuc-Seq) **Species**: mouse **Cell Count**: 1,402 **Gene Count**: 25,392

Figure S6. Visualization and analysis of a scRNA-seq dataset by cellxgene VIP.

4.4.3 Case Study 1

Exploration and visualization of differentially expressed genes (DEGs) between two types of cells.

As shown in Figure S7a, two types of cells, Astrocytes (1036 cells) and Oligodendrocytes (4417 cells) are selected. By running differential gene expression analysis with one of the built-in statistical methods such as Welch's t-test, we detected 1578 (DEGs), including 715 up-regulated and 853 down-regulated genes in astrocytes compared to oligodendrocytes (Figure S7a). The expression of the top four DEGs among the cell types indicates that gene MBP, ST18 and RNF220 are expressed explicitly in oligodendrocytes, while gene PITPNC3 is expressed mainly in astrocytes and endothelial cells (Figure S7b).



(a)

1. Select "Astrocytes"
2. Put in selection 1
3. Select "Oligodendrocytes"
4. Put in selection 2
5. Click to bring up VIP
6. Find DEG

Find differentially expressed genes

Selection 1 (1036) cells and selection 2 (4417) cells

Differentially expressed genes will be detected between the two cell gorups defined above, unless two categories of the 'Custom_combine' annotation are selected, which enables DE analysis between those two:

Select the number of top DEG: 200

Select the method: Welch's t-test (cellxgene)

Select genes to be marked in the volcano plot: ☐ Uncheck / Check All

Define the significance in FDR > 0.05 and absolute log2(FC) > 1

Remove genes in volcano plot with the absolute log2(FC) larger than: 10

Gene Set Enrichment Analysis:

DEG: group1 vs group2

Show 10 entries Search:

Gene	Log Fold Change	p-value	padj	n_cells_by_counts	me
PITPNC1	1.58	5.51e-287	2.70e-283	7236	1.14
MBP	-5.03	1.31e-272	6.40e-269	9748	4.53
ST18	-4.04	3.58e-259	1.76e-255	6012	2.53
RNF220	-3.57	3.70e-254	1.82e-250	7378	2.34
CTNNA3	-2.26	1.70e-241	8.34e-238	13644	4.56
SLC5A11	-4.19	1.67e-234	8.20e-231	5378	1.08

(b)

1. Click names to select genes
2. Click to clear cell selections
3. Click to select all cells
4. Put in selection 1
5. Pick "Dot Plot"

Selected genes: PITPNC1, MBP, ST18, RNF220

DEG: group1 vs group2

Show 10 entries Search:

Gene	Log Fold Change	p-value	padj	n_cells_by_counts
PITPNC1	1.58	5.51e-287	2.70e-283	7236
MBP	-5.03	1.31e-272	6.40e-269	9748
ST18	-4.04	3.58e-259	1.76e-255	6012
RNF220	-3.57	3.70e-254	1.82e-250	7378

Global setting for VIP

Select cells cumulatively: ○ Yes ● No [Clear Cell Selections]

Selection 1 (16649) | Selection 2 (0)

Dot plot shows per group, the fraction of cells expressing a gene (dot size) and the mean expression of the gene in those cell (color scale)

Choose cell selection(s): ☑ Selection 1 (16649) ☐ Selection 2 (0)

Choose genes ('Add Genes' first): ☑ Uncheck / Check All

☑ PITPNC1 ☑ MBP ☑ ST18 ☑ RNF220

Expression cutoff for cell fraction (dot size): 1

Expression is averaged only over cells expressing a given gene above the cutoff: ○ Yes ● No

Select user-defined gene sets:

Group by: Level_1_cell_type

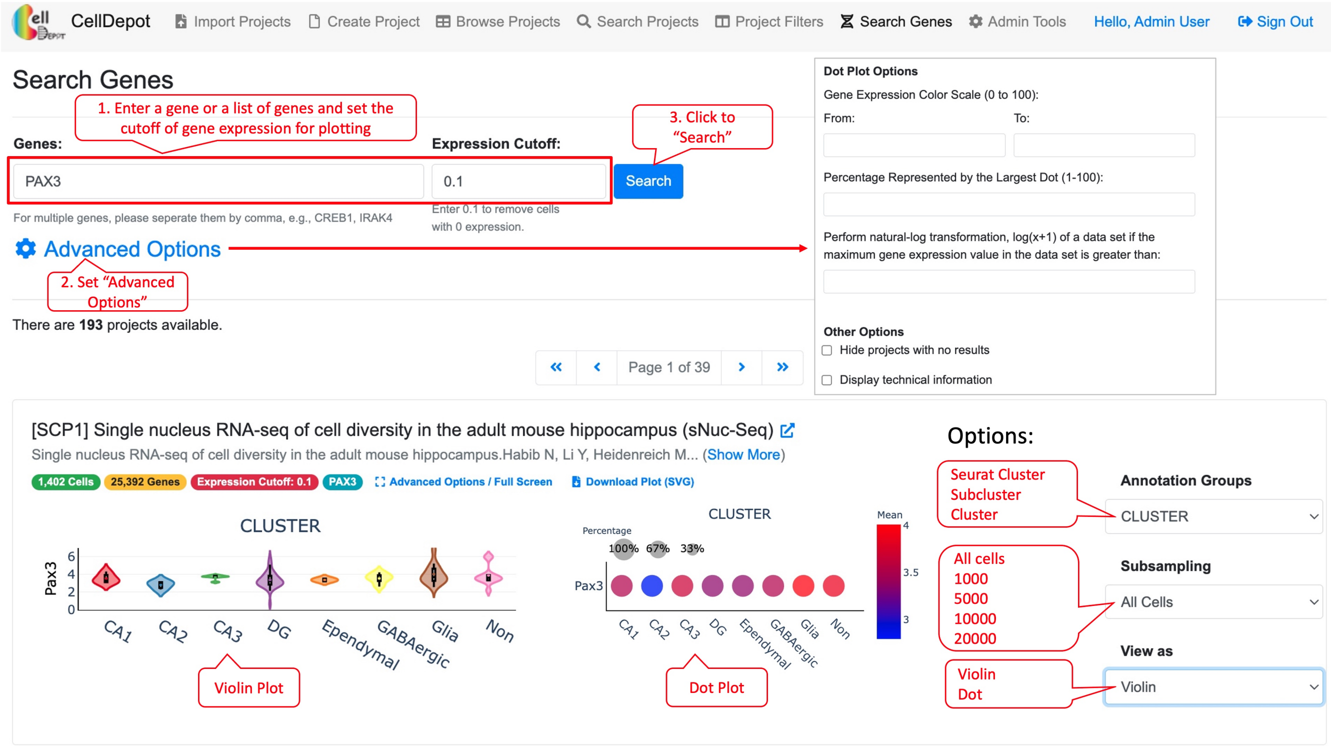
Adjust the legend width: 1.5

Color: Reds ☐ Reverse [Show Color Help]

Figure S7. Exploration of differentially expressed genes in dataset GSE140231 through cellxgene VIP. (a) Identifying differentially expressed genes in astrocytes and oligodendrocytes. (b) The expression of top four genes in various cell types as shown in dot plot.

4.5 Search Genes

This tab allows searching on genes of interest with the expression cutoff. The search outcome provides users a list of projects in which genes of interest are expressed above the cutoff. Each project displays a link to project page and a plot if applicable. This plot can be either a violin plot or dot plot showing the gene expression level in a selected annotation group. Further, under "Advanced Options", users can define the range of expression color scale and the percentage represented by the largest dot to have expression data from various projects plotted in a unified manner.



Search Genes

1. Enter a gene or a list of genes and set the cutoff of gene expression for plotting

3. Click to "Search"

Genes: PAX3

Expression Cutoff: 0.1

For multiple genes, please seperate them by comma, e.g., CREB1, IRAK4

Enter 0.1 to remove cells with 0 expression.

2. Set "Advanced Options"

There are **193** projects available.

Page 1 of 39

Dot Plot Options

Gene Expression Color Scale (0 to 100):

From:

To:

Percentage Represented by the Largest Dot (1-100):

Perform natural-log transformation, log(x+1) of a data set if the maximum gene expression value in the data set is greater than:

Other Options

☐ Hide projects with no results

☐ Display technical information

[SCP1] Single nucleus RNA-seq of cell diversity in the adult mouse hippocampus (sNuc-Seq)

Single nucleus RNA-seq of cell diversity in the adult mouse hippocampus.Habib N, Li Y, Heidenreich M... (Show More)

Violin Plot

Dot Plot

Options:

Seurat Cluster
Subcluster
Cluster

Annotation Groups: CLUSTER

All cells
1000
5000
10000
20000

Subsampling: All Cells

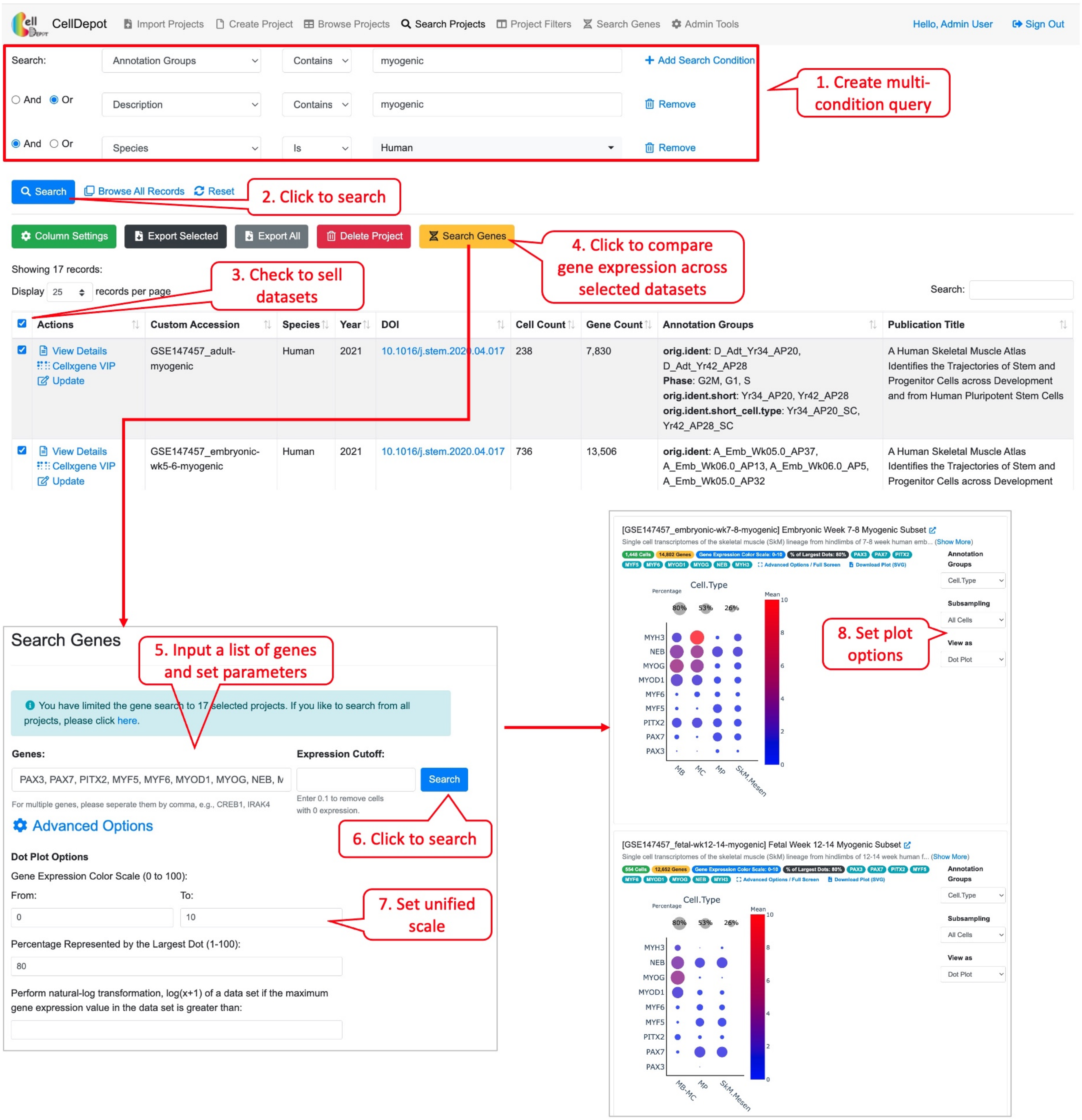
Violin
Dot

View as: Violin

Figure S8. Steps to find gene expression level of a gene in projects under 'Search Genes' tab. The final plot can be customized by available options listed in red callout boxes on the left side.

4.5.1 Case Study 2

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.



1. Create multi-condition query

Search: Annotation Groups | Contains | myogenic

○ And ● Or | Description | Contains | myogenic

● And ○ Or | Species | Is | Human

2. Click to search

3. Check to sell datasets

4. Click to compare gene expression across selected datasets

Showing 17 records:

Display 25 records per page

Search:

☑	Actions	Custom Accession	Species	Year	DOI	Cell Count	Gene Count	Annotation Groups	Publication Title
☑	View Details Cellxgene VIP Update	GSE147457_adult-myogenic	Human	2021	10.1016/j.stem.2020.04.017	238	7,830	**orig.ident**: D_Adt_Yr34_AP20, D_Adt_Yr42_AP28 **Phase**: G2M, G1, S **orig.ident.short**: Yr34_AP20, Yr42_AP28 **orig.ident.short_cell.type**: Yr34_AP20_SC, Yr42_AP28_SC	A Human Skeletal Muscle Atlas Identifies the Trajectories of Stem and Progenitor Cells across Development and from Human Pluripotent Stem Cells
☑	View Details Cellxgene VIP Update	GSE147457_embryonic-wk5-6-myogenic	Human	2021	10.1016/j.stem.2020.04.017	736	13,506	**orig.ident**: A_Emb_Wk05.0_AP37, A_Emb_Wk06.0_AP13, A_Emb_Wk06.0_AP5, A_Emb_Wk05.0_AP32	A Human Skeletal Muscle Atlas Identifies the Trajectories of Stem and Progenitor Cells across Development

Search Genes

5. Input a list of genes and set parameters

You have limited the gene search to 17 selected projects. If you like to search from all projects, please click here.

Genes: PAX3, PAX7, PITX2, MYF5, MYF6, MYOD1, MYOG, NEB, M

For multiple genes, please seperate them by comma, e.g., CREB1, IRAK4

Expression Cutoff:

Enter 0.1 to remove cells with 0 expression.

6. Click to search

Dot Plot Options

Gene Expression Color Scale (0 to 100):

From: 0 To: 10

7. Set unified scale

Percentage Represented by the Largest Dot (1-100): 80

Perform natural-log transformation, log(x+1) of a data set if the maximum gene expression value in the data set is greater than:

[GSE147457_embryonic-wk7-8-myogenic] Embryonic Week 7-8 Myogenic Subset

Single cell transcriptomes of the skeletal muscle (SkM) lineage from hindlimbs of 7-8 week human emb... (Show More)

1,448 Cells | 14,802 Genes | Gene Expression Color Scale: 0-10 | % of Largest Dots: 80% | PAX3 PAX7 PITX2 MYF5 MYF6 MYOD1 MYOG NEB MYH3

Annotation Groups: Cell.Type

Subsampling: All Cells

View as: Dot Plot

8. Set plot options

[GSE147457_fetal-wk12-14-myogenic] Fetal Week 12-14 Myogenic Subset

Single cell transcriptomes of the skeletal muscle (SkM) lineage from hindlimbs of 12-14 week human f... (Show More)

554 Cells | 12,652 Genes | Gene Expression Color Scale: 0-10 | % of Largest Dots: 80% | PAX3 PAX7 PITX2 MYF5 MYF6 MYOD1 MYOG NEB MYH3

Annotation Groups: Cell.Type

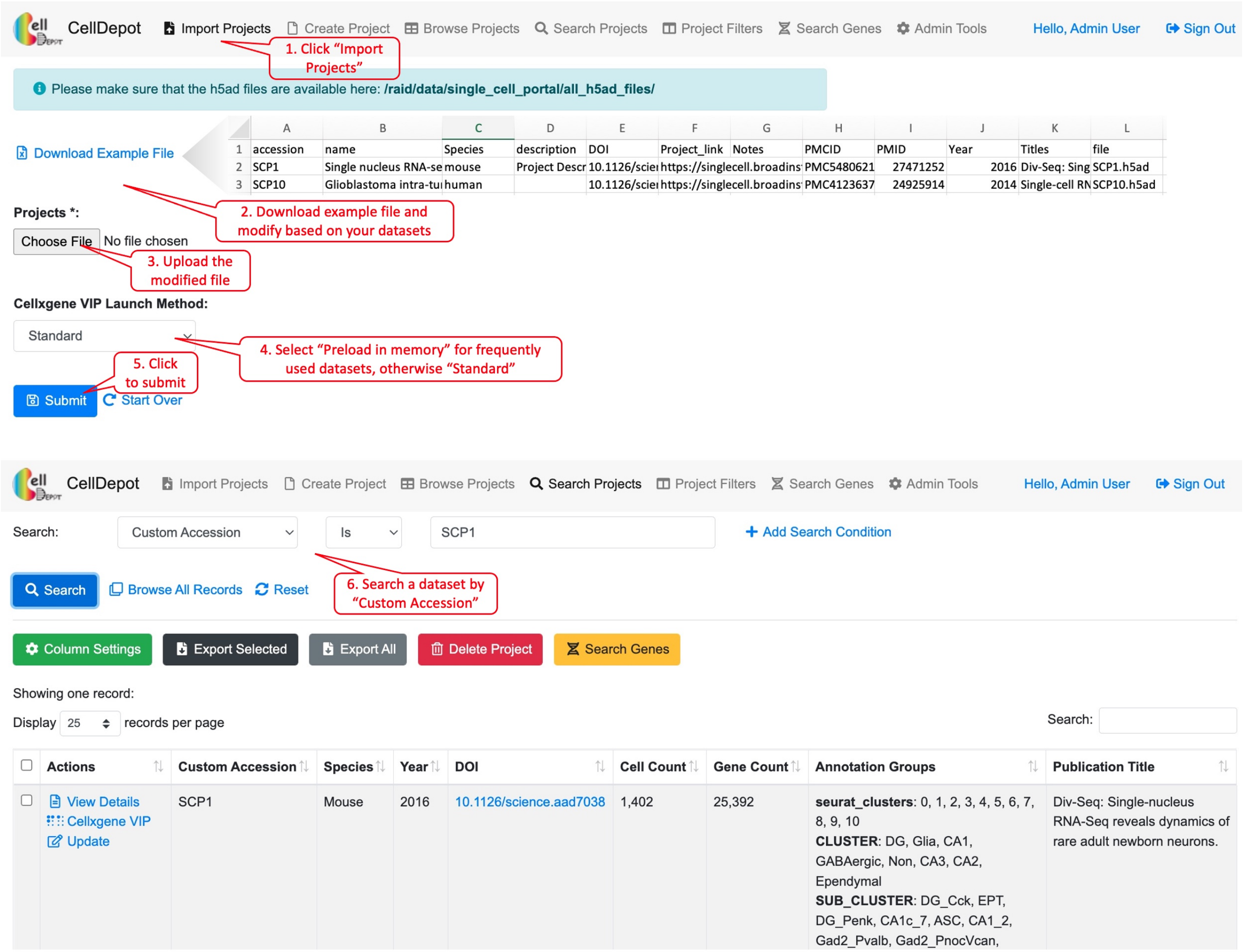
Subsampling: All Cells

View as: Dot Plot

Figure S9. Workflow of conducting the cross-project comparison of a list of genes among the selected datasets.

4.6 Import Projects

The functionality is limited to admin users. To upload new projects to CellDepot database in batch, two types of files are required: 1) .h5ad files and 2) project information file in CSV (Comma Separated Values) format. First, the prepared h5ad files are required to be copied to a folder defined in the configuration file, e.g., /data/celldepot/all_h5ad_files/. Afterwards, admin users navigate to the CellDepot home page, click 'Import Projects' at the top menu, then 'Download Example File' to fill in meta information of datasets into the downloaded template for submission. In addition, there are two cellxgene VIP launch modes to chosen from, 'Standard' and 'Preload in Memory'. 'Standard' mode is for infrequently used datasets while 'Preload in Memory' should be selected to speed up loading and responding time of frequently used large datasets. After the metadata file is uploaded, CellDepot will automatically convert the dataset to CSC format if needed through a cron job (2.4). To explore the detail of imported datasets, users can enter ‘Browse Projects’ page and then search these datasets by user assigned ‘Custom Accession’ identifiers.



1. Click “Import Projects”

Please make sure that the h5ad files are available here: **/raid/data/single_cell_portal/all_h5ad_files/**

2. Download example file and modify based on your datasets

	A	B	C	D	E	F	G	H	I	J	K	L
1	accession	name	Species	description	DOI	Project_link	Notes	PMCID	PMID	Year	Titles	file
2	SCP1	Single nucleus RNA-se	mouse	Project Descr	10.1126/scie	https://singlecell.broadins		PMC5480621	27471252	2016	Div-Seq: Sing	SCP1.h5ad
3	SCP10	Glioblastoma intra-tu	human		10.1126/scie	https://singlecell.broadins		PMC4123637	24925914	2014	Single-cell RN	SCP10.h5ad

Projects *: No file chosen

3. Upload the modified file

Cellxgene VIP Launch Method: Standard

4. Select “Preload in memory” for frequently used datasets, otherwise “Standard”

5. Click to submit

Search: Custom Accession Is SCP1

6. Search a dataset by “Custom Accession”

Showing one record:

Display 25 records per page

Search:

☐	Actions	Custom Accession	Species	Year	DOI	Cell Count	Gene Count	Annotation Groups	Publication Title
☐	View Details Cellxgene VIP Update	SCP1	Mouse	2016	10.1126/science.aad7038	1,402	25,392	**seurat_clusters**: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 **CLUSTER**: DG, Glia, CA1, GABAergic, Non, CA3, CA2, Ependymal **SUB_CLUSTER**: DG_Cck, EPT, DG_Penk, CA1c_7, ASC, CA1_2, Gad2_Pvalb, Gad2_PnocVcan,	Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons.

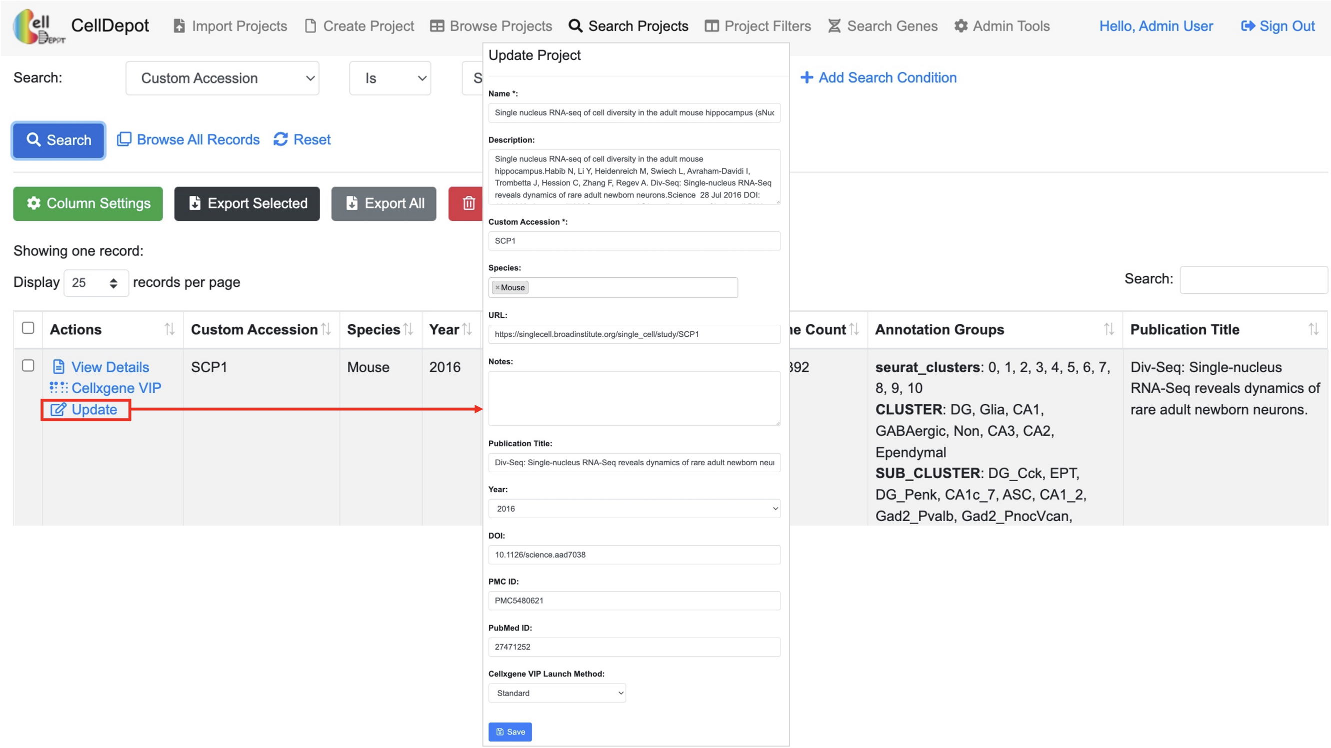
Figure S10. Workflow of how to import new datasets.

4.7 Create Project

Besides batch uploading under “Import Projects” tab, admin user can use the online form under this tab to submit information of a project.

4.8 Update a Project

Project information including launch mode can be modified by admin users by clicking on ‘Update’ link of a project under ‘Actions’ column in the table.



☐	Actions	Custom Accession	Species	Year	e Count	Annotation Groups	Publication Title
☐	View Details Cellxgene VIP Update	SCP1	Mouse	2016	392	**seurat_clusters**: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 **CLUSTER**: DG, Glia, CA1, GABAergic, Non, CA3, CA2, Ependymal **SUB_CLUSTER**: DG_Cck, EPT, DG_Penk, CA1c_7, ASC, CA1_2, Gad2_Pvalb, Gad2_PnocVcan,	Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons.

Update Project

Name *: Single nucleus RNA-seq of cell diversity in the adult mouse hippocampus (sNuc

Description: Single nucleus RNA-seq of cell diversity in the adult mouse hippocampus.Habib N, Li Y, Heidenreich M, Swiech L, Avraham-Davidi 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:

Custom Accession *: SCP1

Species: × Mouse

URL: https://singlecell.broadinstitute.org/single_cell/study/SCP1

Notes:

Publication Title: Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neu

Year: 2016

DOI: 10.1126/science.aad7038

PMC ID: PMC5480621

PubMed ID: 27471252

Cellxgene VIP Launch Method: Standard