

# **Supplementary Materials - CellDepot: A unified repository for scRNASeq data and visual exploration**

Dongdong Lin<sup>11</sup>, Yirui Chen<sup>21</sup>, Soumya Negi<sup>1</sup>, Derrick Cheng<sup>1</sup>, Zhengyu Ouyang<sup>2</sup>, David  
Sexton<sup>1</sup>, Kejie Li<sup>\*31</sup>, and Baohong Zhang<sup>\*41</sup>

<sup>1</sup>Research Department, Biogen, Inc

<sup>2</sup>BioInfoRx, Inc

## Contents

<b>Getting start with CellDepot</b>	<b>2</b>
Sources of annotation and metadata . . . . .	2
Data format, availability, and preparation . . . . .	2
CellDepot platform and installation . . . . .	3
How to set up cron job? . . . . .	3
CellDepot API (Application Programming Interface) . . . . .	4
Code availability . . . . .	4
Online tutorials . . . . .	4
<b>Supplemental Tables</b>	<b>4</b>
Table S1: Comparison matrix of web portal tools . . . . .	4
Table S2 - Criterion for query search . . . . .	5
Table S3 - Criterion for data analysis explorer . . . . .	5
Table S4 - Project metadata captured in CellDepot . . . . .	5

<sup>1</sup>These two authors contributed equally

<sup>2</sup>These two authors contributed equally

<sup>3</sup>Corresponding author, email:kejie.li@biogen.com

<sup>4</sup>Corresponding author, email:baohong.zhang@biogen.com

21	<b>Supplemental Tutorial</b>	<b>5</b>
22	Browse Projects . . . . .	6
23	Search Projects . . . . .	6
24	Project Filters . . . . .	7
25	Visualize Datasets . . . . .	8
26	View Details . . . . .	8
27	Data Visualization and Analysis by cellxgene VIP . . . . .	9
28	Case Study 1 . . . . .	10
29	Search Genes . . . . .	11
30	Case Study 2 . . . . .	12
31	Import Projects . . . . .	13
32	Create Project . . . . .	15
33	Update a Project . . . . .	15
34	This is a supplementary materials written in Markdown, which provides the detailed guide	
35	for CellDepot web portal.	

## 36 **GETTING START WITH CELLDEPOT**

### 37 **Sources of annotation and metadata**

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

### 44 **Data format, availability, and preparation**

45     CellDepot requires scRNA-seq data in h5ad file where the expression matrix is stored in  
46     CSC (compressed sparse column) instead of CSR (compressed sparse row) format to improve

47 the speed of data retrieving. For example, designating genes as columns in the h5ad file cre-  
48 ates the interactive plot five times faster than as rows. Just in case, we provide sample scripts  
49 to help users generate h5ad files. Having gene expression matrix, metadata, and layout files,  
50 users can easily combine and convert their data to h5ad file by following this R script on  
51 <https://github.com/interactivereport/CellDepot/blob/main/toH5ad.R>. In the case  
52 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  
53 python script tailored to their own data. Categorical features extracted from a h5ad file are shown in  
54 the ‘annotation groups’ column of the table on CellDepot home page, while the numerical features  
55 are shown as the histograms in the rightmost panel on cellxgene VIP. (**Figure S6.**)  
56

## 57 CellDepot platform and installation

58 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 im-  
59 plemented with MySQL database, an advanced search engine, and powerful interactive vis-  
60 ualizing tools that allow users to explore attributes of datasets as well as scRNA-seq analy-  
61 sis results. Also, users can intentionally select single-cell RNA-seq datasets on the web in-  
62 terface by simply browsing the online dataset table or applying advanced search to perform  
63 the cross-dataset comparison. Moreover, CellDepot also provides comprehensive data anal-  
64 ysis tools via an embedded interactive visualization plugin. To host private datasets, local  
65 instance of CellDepot on Unix server can be installed by following the guide here, [https://celldepot.bxgenomics.com/celldepot\\_manual/install\\_environment.php](https://celldepot.bxgenomics.com/celldepot_manual/install_environment.php).  
66  
67

## 68 How to set up cron job?

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

```
70 1 @hourly <user-name> cd /var/www/html/celldepot/app/core;  
71 2 php ./api_toCSCh5ad.php
```

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

73       **CellDepot API (Application Programming Interface)**

74       The CellDepot API web service provides a direct way to generate figures for users to share or  
75       embed in web page. For example, the following URL will generate a gene expression violin plot  
76       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](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  
77       URL and explanation of parameters are detailed in the online documentation, [https://celldepot.bxgenomics.com/celldepot\\_manual/api\\_gene\\_plot.php](https://celldepot.bxgenomics.com/celldepot_manual/api_gene_plot.php).

78

79

80

81       **Code availability**

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

85

86       **Online tutorials**

87       To better assist biologists to use CellDepot and integrated cellxgene VIP visual analytical tool,  
88       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 CellDe-  
89       pot and [https://interactivereport.github.io/cellxgene\\_VIP/tutorial/docs/how-to-use-cellx-](https://interactivereport.github.io/cellxgene_VIP/tutorial/docs/how-to-use-cellx-)  
90       [html](https://interactivereport.github.io/cellxgene_VIP/html) for cellxgene VIP, respectively. In addition, a question mark next to the title of each VIP  
91       function module provides a direct way to reach the corresponding section of the HTML document  
92       for help.

93

94       **SUPPLEMENTAL TABLES**

95       **Table S1: Comparison matrix of web portal tools**

96       Table link: <https://github.com/interactivereport/CellDepot/blob/gh-pages/bookdown/>  
97       [S1.csv](#) (Table S1).

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

99 **Table S2 - Criterion for query search**

100 Table link: <https://github.com/interactivereport/CellDepot/blob/gh-pages/bookdown/S2.csv> (Table S2).

102 **Table S3 - Criterion for data analysis explorer**

103 Table link: <https://github.com/interactivereport/CellDepot/blob/gh-pages/bookdown/S3.csv> (Table S3)

105 **Table S4 - Project metadata captured in CellDepot**

106 Table link: <https://github.com/interactivereport/CellDepot/blob/gh-pages/bookdown/S4.csv> (Table S4)

108 **SUPPLEMENTAL TUTORIAL**

109 CellDepot is a scRNA-seq data portal consisting of a relational database management system,  
110 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  
111 tutorial providing detailed instructions. Clicking on a figure will bring up the enlarged view.  
112

Showing 271 records:

Display 25 records per page

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

Showing 1 to 25 of 271 records

Previous [1](#) [2](#) [3](#) [4](#) [5](#) ... [11](#) Next

114

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

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

## 121           Browse Projects

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

The screenshot shows the CellDepot homepage with the 'Browse Projects' tab selected. At the top, there's a search bar with dropdown menus for 'Annotation Groups' and 'Contains'. Below the search bar are buttons for 'Search', 'Browse All Records', and 'Reset'. A red callout box points to the 'Column Settings' button, which is highlighted in green. Another red callout box points to the 'Export Selected' and 'Export All' buttons, both of which are highlighted in yellow. A third red callout box points to the 'Search Genes' button. The main content area displays a table of 271 records. On the left, there's a column with checkboxes and 'View Details' links. A red callout box points to one of these links with the instruction '1. Click "Column Settings" to customize fields to be displayed in the table'. Another red callout box points to a checkbox with the instruction '2. Check to select'. A fourth red callout box points to the 'Export' buttons with the instruction '3. Export whole table or selected entries'. The table itself has columns for Actions, View Details, Release, Name, Description, File Size, Notes, PubMed ID, and Year. An 'Apply' button is located at the bottom right of the table. The bottom of the page shows navigation links for 'Previous', 'Next', and page numbers 1 through 11.

126           Figure S2. Browsing projects in a personalized view.

## 127           Search Projects

128           This function allows users to search projects of interest, which can be accessed through the  
129           homepage as well. Users can search projects by 17 attributes in multiple logic conditions: anno-  
130           tation groups, cell count, cellxgene VIP launch method, Custom accession, description, DOI, file

131 name, file size, gene count, name, notes, PMC ID, Publication Title, PubMed ID, Species, URL,  
132 and/or Year.

The screenshot shows the CellDepot search interface. At the top, there is a search bar with dropdown menus for 'Species' (set to 'Human'), 'Is' (set to 'Human'), and 'Annotation Groups' (set to 'Contain Neuron'). Below the search bar are buttons for 'Add Search Condition' and 'Remove'. Red annotations numbered 1 through 7 point to specific elements: 1. 'Click "Search Projects"', 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' (radio buttons for 'And' or 'Or'), and 7. 'Click "Search"'. Below the search bar are buttons for 'Search', 'Browse All Records', 'Reset', 'Column Settings', 'Export Selected', 'Export All', and 'Search Genes'. A message 'Showing 6 records:' is displayed above a table. The table has columns: Actions, Custom Accession, Name, Species, Year, DOI, Cell Count, Gene Count, Publication Title, Uploaded By, and Uploaded On. The data in the table is as follows:

Actions	Custom Accession	Name	Species	Year	DOI	Cell Count	Gene Count	Publication Title	Uploaded By	Uploaded On
<input type="checkbox"/> <a href="#">View Details</a> 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
<input type="checkbox"/> <a href="#">View Details</a> Cellxgene VIP	SCP19	Salk Institute - Single-cell Methylation Sequencing Identifies Distinct Neuronal Populations in Mouse Frontal Cortex	Human, Mouse			3,386	46,600		Admin User	2021-04-02
<input type="checkbox"/> <a href="#">View Details</a> Cellxgene VIP	SCP199	Comparison of Engineered 3D Neural Tissues	Human, Mouse			4,280	22,567		Admin User	2021-04-02
<input type="checkbox"/> <a href="#">View Details</a> Cellxgene VIP	SCP498	Transcriptional and Cellular Diversity of the Human Heart	Human			287,269	33,694		Admin User	2021-04-02
<input type="checkbox"/> <a href="#">View Details</a> 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
<input type="checkbox"/> <a href="#">View Details</a> 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 Previous 1 Next

133  
134 Figure S3. Workflow of searching projects by using the graphical multi-logic, multi-condition query  
135 builder. Six datasets are identified when searching by ‘Species is Human’ and ‘Annotation Group  
136 contains Neuron’.

## 137 Project Filters

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

CellDepot CellDepot Browse Projects Search Projects Project Filters Search Genes Hello, Guest User Sign Out

## Project Filters

Advanced Search

**Year**

<input type="checkbox"/> Select All	29   271
<input checked="" type="checkbox"/> 2021	29   30
<input type="checkbox"/> 2020	0   42
<input type="checkbox"/> 2019	0   15
<input type="checkbox"/> 2018	0   9
<input type="checkbox"/> 2017	0   7
<input type="checkbox"/> 2016	0   8
<input type="checkbox"/> 2015	0   4
—	

**Species**

<input type="checkbox"/> Select All	0   271
<input checked="" type="checkbox"/> Human	0   114
<input type="checkbox"/> Mouse	0   113
<input type="checkbox"/> Others	0   45
<input type="checkbox"/> Rat	0   13
<input type="checkbox"/> Hydra	0   2
<input type="checkbox"/> Axolotl	0   1
<input type="checkbox"/> MacaqueF	0   1
—	

Column Settings

Showing 29 records:

Display 25 records per page Search:

<input type="checkbox"/> Actions	Custom Accession	Name	Species	Year	DOI	Cell Count	Gene Count	Publication Title	Uploaded	
<input type="checkbox"/>	<input type="button" value="View Details"/> <small>Cellxgene VIP</small>	TabulaSapiens	Tabula Sapiens release 1.0 (Complete Dataset)	Human	2021	<a href="https://doi.org/10.1101/2021.07.19.452956">10.1101/2021.07.19.452956</a>	156,559	58,833	The Tabula Sapiens: a multiple organ single cell transcriptomic atlas of humans	Admin U
<input type="checkbox"/>	<input type="button" value="View Details"/> <small>Cellxgene VIP</small>	TS_Vasculature	Tabula Sapiens release 1.0 (Vasculature)	Human	2021	<a href="https://doi.org/10.1101/2021.07.19.452956">10.1101/2021.07.19.452956</a>	8,655	58,833	The Tabula Sapiens: a multiple organ single cell transcriptomic atlas of humans	Admin U
<input type="checkbox"/>	<input type="button" value="View Details"/> <small>Cellxgene VIP</small>	TS_Trachea	Tabula Sapiens release 1.0 (Trachea)	Human	2021	<a href="https://doi.org/10.1101/2021.07.19.452956">10.1101/2021.07.19.452956</a>	4,445	58,833	The Tabula Sapiens: a multiple organ single cell transcriptomic atlas of humans	Admin U
<input type="checkbox"/>	<input type="button" value="View Details"/> <small>Cellxgene VIP</small>	TS_Thymus	Tabula Sapiens release 1.0 (Thymus)	Human	2021	<a href="https://doi.org/10.1101/2021.07.19.452956">10.1101/2021.07.19.452956</a>	12,147	58,833	The Tabula Sapiens: a multiple organ single cell transcriptomic atlas of humans	Admin U
<input type="checkbox"/>	<input type="button" value="View Details"/> <small>Cellxgene VIP</small>	TS_Spleen	Tabula Sapiens release 1.0 (Spleen)	Human	2021	<a href="https://doi.org/10.1101/2021.07.19.452956">10.1101/2021.07.19.452956</a>	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

141

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

## Visualize Datasets

### *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.

The screenshot shows the CellDepot web interface. At the top, there are navigation links: CellDepot, Browse Projects, Search Projects, Project Filters, Search Genes, Tutorial, Hello, Guest User, and Sign Out. Below the header, there is a search bar with dropdowns for 'Search' (Custom Accession), 'Is' (SCP1), and '+ Add Search Condition'. There are also buttons for 'Search', 'Browse All Records', 'Reset', 'Column Settings', 'Export Selected', 'Export All', and 'Search Genes'.

The main content area displays a 'Review Project' page for 'SCP1'. It includes sections for 'Actions' (with a circled '1'), 'Custom Accession' (with a circled '1'), 'View Details' (with a circled '1'), 'Cellxgene VIP' (with a circled '1'), 'Project' (with a circled '2'), 'Annotation Groups' (with a circled '2'), and 'Broad scRNA Portal'. The 'Project' section shows 'Uploaded By' Admin User and 'Uploaded On' 2021-04-02. The 'Annotation Groups' section shows 'All Annotation Groups' (with a circled '2') and a table of clusters:

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

A red box highlights the 'Info of annotation groups retrieved from h5ad file' section, which contains 'Meta info inputted by user'.

To the right, there is a table titled 'Gene Count' showing cluster details:

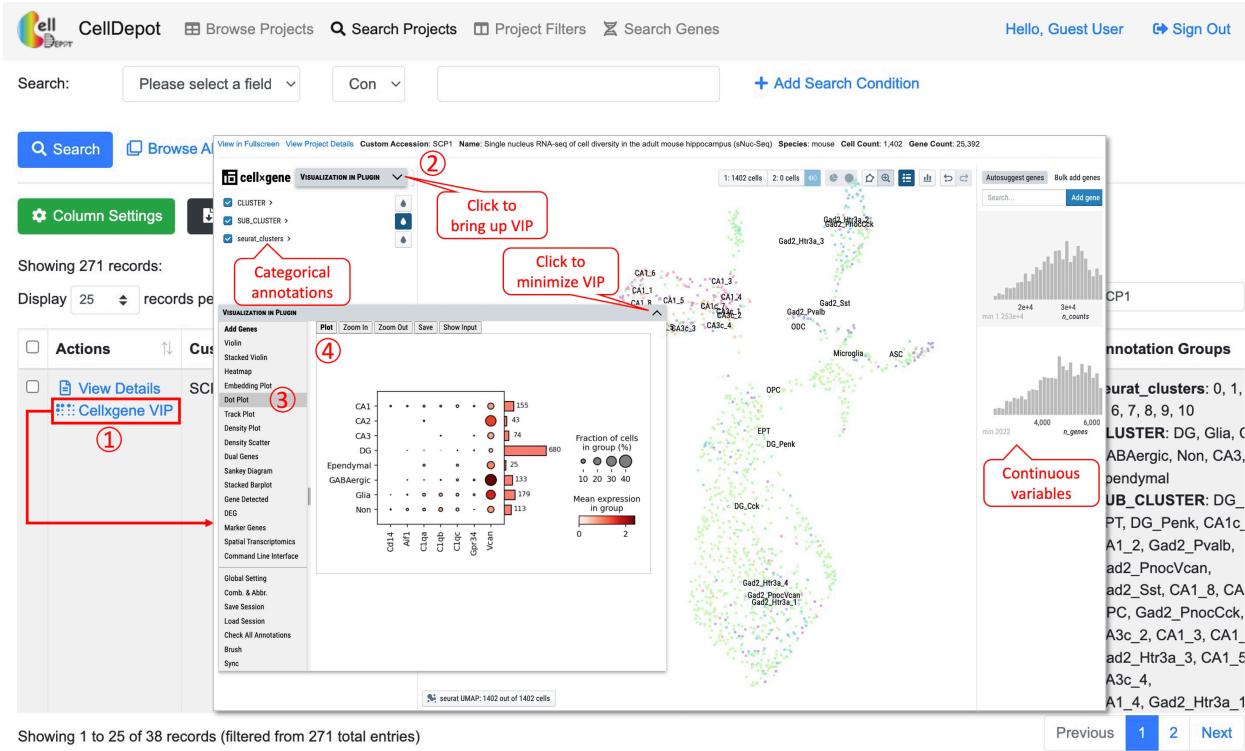
Gene Count	Annotation Groups	Publication Title	Upload
25,392	<b>seurat_clusters</b> : 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 <b>CLUSTER</b> : DG, Glia, CA1, GABAergic, Non, CA3, C...	Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons.	Admin

149

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

151 *Data Visualization and Analysis by cellxgene VIP*

152 CellDepot is not only a database management system, but also a web portal for visualizing  
 153 and analyzing scRNA-seq datasets through embedded cellxgene VIP tool. By clicking ‘cellxgene  
 154 VIP’ to access functional modules on the menu, users can perform advanced data visualization  
 155 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](https://interactivereport.github.io/cellxgene_VIP/tutorial/docs/how-to-use-cellxgene-vip.html).  
 156



157

Showing 1 to 25 of 38 records (filtered from 271 total entries)

Previous 1 2 Next

158

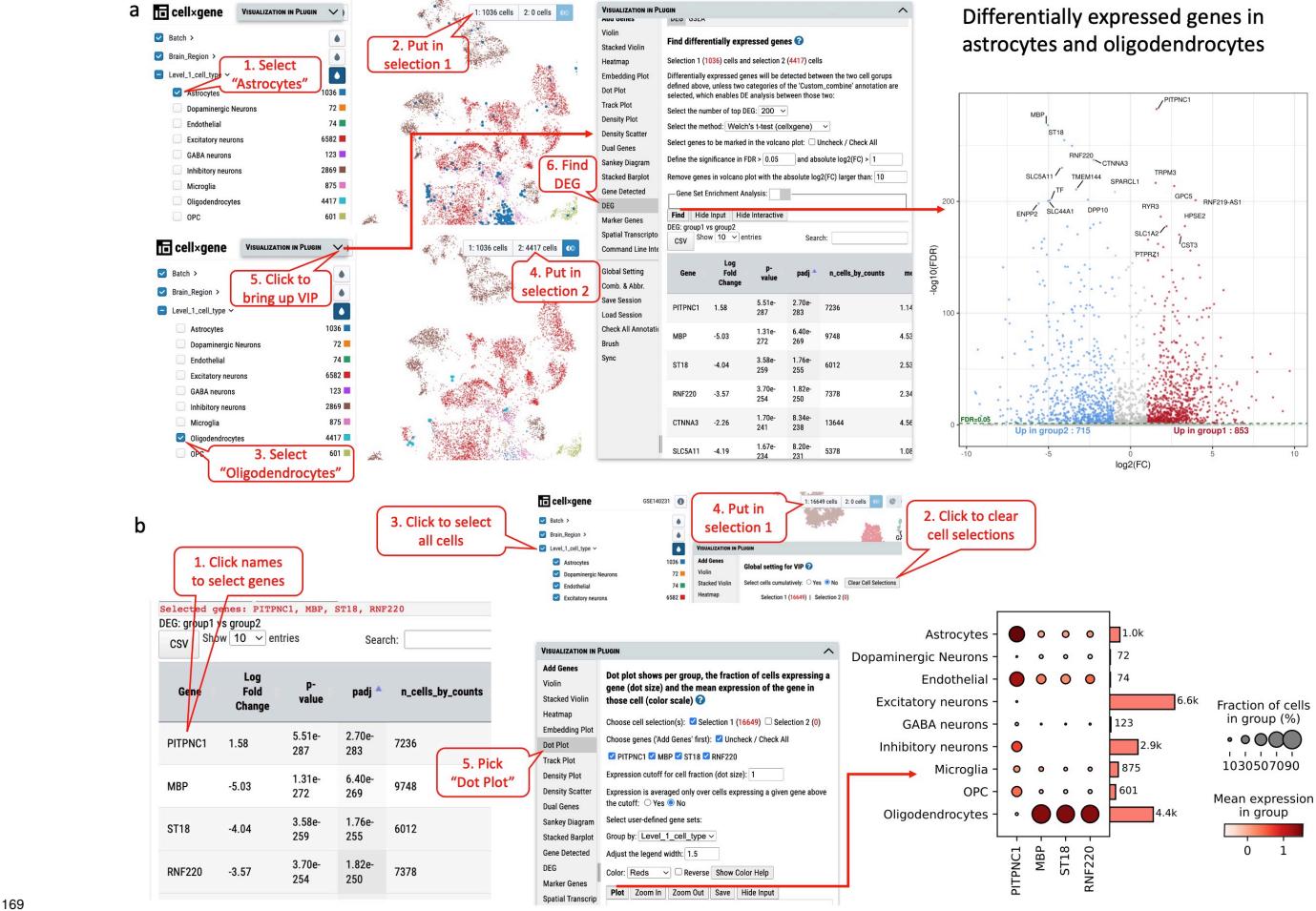
Figure S6. Visualization and analysis of a scRNA-seq dataset by cellxgene VIP.

159

### Case Study 1

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

162 As shown in Figure S7a, two types of cells, Astrocytes (1036 cells) and Oligodendrocytes  
 163 (4417 cells) are selected. By running differential gene expression analysis with one of the built-in  
 164 statistical methods such as Welch's t-test, we detected 1578 (DEGs), including 715 up-regulated  
 165 and 853 down-regulated genes in astrocytes compared to oligodendrocytes (Figure S7a). The  
 166 expression of the top four DEGs among the cell types indicates that gene MBP, ST18 and RNF220  
 167 are expressed explicitly in oligodendrocytes, while gene PTPNC3 is expressed mainly in astrocytes  
 168 and endothelial cells (Figure S7b).

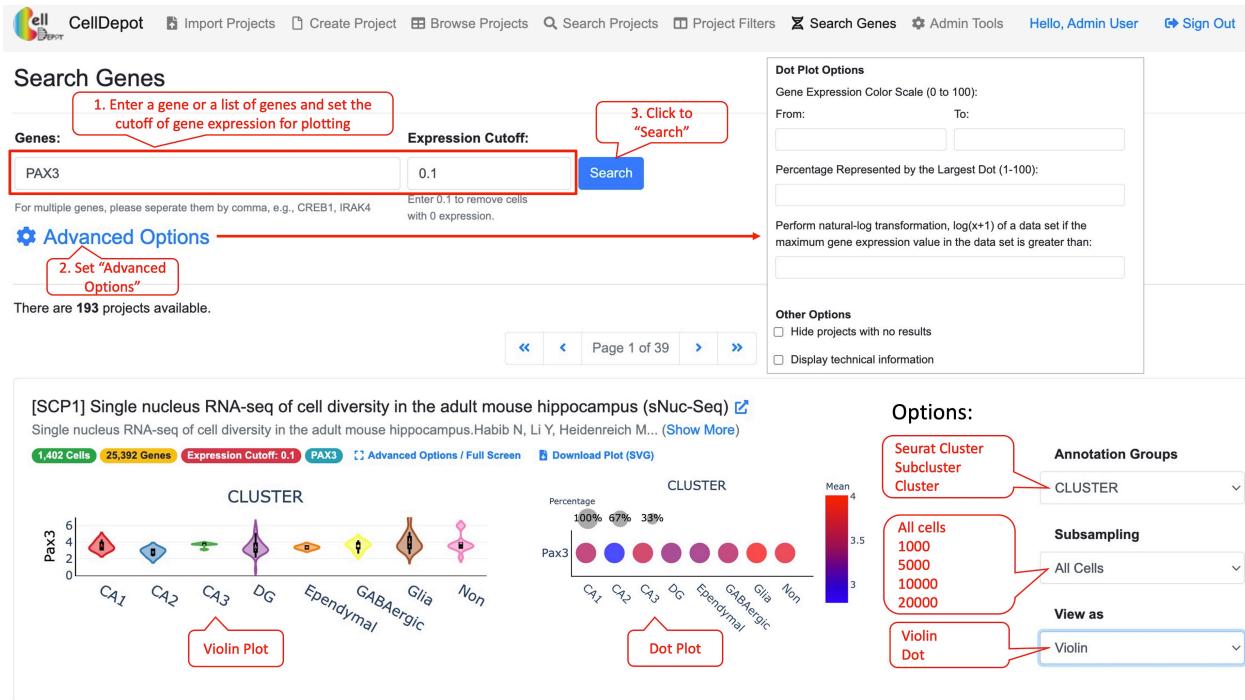


169

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.

### 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.



180  
181 Figure S8. Steps to find gene expression level of a gene in projects under ‘Search Genes’ tab. The  
182 final plot can be customized by available options listed in red callout boxes on the left side.

183 *Case Study 2*

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

1. Create multi-condition query

2. Click to search

3. Check to sell datasets

4. Click to compare gene expression across selected datasets

5. Input a list of genes and set parameters

6. Click to search

7. Set unified scale

8. Set plot options

The screenshot shows the CellDepot web application. At the top, there's a search bar with the query 'Annotation Groups Contains myogenic'. Below it, there are two more search fields: 'Description Contains myogenic' and 'Species Is Human'. A red box highlights this section with the label '1. Create multi-condition query'. Below the search bar are buttons for 'Search', 'Browse All Records', and 'Reset'. A red box highlights the 'Search' button with the label '2. Click to search'. The main area shows a table of 17 records, each with a checkbox in the 'Actions' column. Two specific rows are highlighted with red boxes: 'GSE147457\_adult-myogenic' and 'GSE147457\_embryonic-wk5-6-myogenic'. A red box highlights the 'Actions' column with the label '3. Check to sell datasets'. To the right of the table, a red box highlights the 'Search Genes' button with the label '4. Click to compare gene expression across selected datasets'. Below the table, a 'Search Genes' modal is open. It contains a list of genes: PAX3, PAX7, PITX2, MYF5, MYF6, MYOD1, MYOG, NEB, and CREB1, IRAK4. A red box highlights this list with the label '5. Input a list of genes and set parameters'. Below the genes is an 'Advanced Options' section with 'Dot Plot Options' and a 'Color Scale' from 0 to 100. A red box highlights the 'Color Scale' with the label '7. Set unified scale'. At the bottom of the 'Search Genes' modal is a 'Search' button, which is also highlighted with a red box and labeled '6. Click to search'. To the right of the 'Search Genes' modal are two dot plot graphs. The top graph is titled '[GSE147457\_embryonic-wk7-8-myogenic] Embryonic Week 7-8 Myogenic Subset' and the bottom graph is titled '[GSE147457\_fetal-wk12-14-myogenic] Fetal Week 12-14 Myogenic Subset'. Both plots show gene expression levels across different cell types (e.g., MYH3, NEB, MYOG, MYOD1, MYF6, MYF5, PITX2, PAX7, PAX3) and stages (e.g., MB, NC, MB, SMM, Meso). A red box highlights the 'Color Scale' in the top plot with the label '8. Set plot options'.

187

188 Figure S9. Workflow of conducting the cross-project comparison of a list of genes among the se-  
189 lected datasets.

## 190 Import Projects

191 The functionality is limited to admin users. To upload new projects to CellDepot database in  
192 batch, two types of files are required: 1) .h5ad files and 2) project information file in CSV (Comma  
193 Separated Values) format. First, the prepared h5ad files are required to be copied to a folder defined

194 in the configuration file, e.g., /data/celldepot/all\_h5ad\_files/. Afterwards, admin users navigate to  
 195 the CellDepot home page, click ‘Import Projects’ at the top menu, then ‘Download Example File’  
 196 to fill in meta information of datasets into the downloaded template for submission. In addition,  
 197 there are two cellxgene VIP launch modes to chosen from, ‘Standard’ and ‘Preload in Memory’.  
 198 ‘Standard’ mode is for infrequently used datasets while ‘Preload in Memory’ should be selected to  
 199 speed up loading and responding time of frequently used large datasets. After the metadata file is  
 200 uploaded, CellDepot will automatically convert the dataset to CSC format if needed through a cron  
 201 job (1). To explore the detail of imported datasets, users can enter ‘Browse Projects’ page and then  
 202 search these datasets by user assigned ‘Custom Accession’ identifiers.

The screenshot shows the CellDepot Import Projects interface. At the top, there is a navigation bar with links for Import Projects, Create Project, Browse Projects, Search Projects, Project Filters, Search Genes, Admin Tools, Hello, Admin User, and Sign Out.

**Step 1:** A red box highlights the 'Import Projects' link in the top navigation bar.

**Step 2:** A red box highlights the 'Download Example File' button on the left side of the page.

**Step 3:** A red box highlights the 'Choose File' input field where a file is selected.

**Step 4:** A red box highlights the 'Select "Preload in memory" for frequently used datasets, otherwise "Standard"' dropdown menu.

**Step 5:** A red box highlights the 'Submit' button.

**Step 6:** A red box highlights the 'Search' button on the Browse Projects page.

**Data Tables:**

- Import Projects Table:**

	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/scierhttps://singlecell.broadin	PMC5480621	27471252	2016	Div-Seq: Sing SCP1.h5ad				
3	SCP10	Glioblastoma intra-tumour human			PMC4123637	24925914	2014	Single-cell RN SCP10.h5ad				
- Browse Projects Table:**

	Custom Accession	Is	SCP1									
1	SCP1	Single nucleus RNA-se mouse	Project Descr	10.1126/scierhttps://singlecell.broadin	PMC5480621	27471252	2016	Div-Seq: Sing SCP1.h5ad				

204 Figure S10. Workflow of how to import new datasets.

205 **Create Project**

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

208 **Update a Project**

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

The screenshot shows the CellDepot interface for updating a project. At the top, there's a navigation bar with links for Import Projects, Create Project, Browse Projects, Search Projects, Project Filters, Search Genes, Admin Tools, and a sign-out link. Below the navigation is a search bar and a table titled "Update Project". The table has columns for Name, Description, Custom Accession, Species, URL, Notes, Publication Title, Year, DOI, PMC ID, PubMed ID, and Cellxgene VIP Launch Method. A red arrow points to the "Update" link in the "Actions" column for the first row, which contains SCP1, Mouse, and 2016. To the right of the table, there's a detailed view of the project with sections for Annotation Groups (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) and a Publication Title (Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons).

211



Table S2 - Criterion for query search

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

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**

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	Digital 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	
Species	String	
URL	Link	
Year	String	

Table S1 - Comparison matrix of web portal tools

Web application	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
<b>Main function</b>														
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
<b>Scalability and Capacity</b>														
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
<b>Data type supported</b>														
Datasets	diverse	diverse	hearing/brain	tumor	diverse	diverse	diverse	reproduction	diverse	diverse	diverse	diverse	immune-related	diverse
Datatypes	Single-cell	Single-cell	Epigenetic	Single-cell	Single-cell	Single-cell	Single-cell	Single-cell, Multi-omics	Single-cell	Single-cell, Proteomics	Single-cell	Single-cell	Single-cell, Multi-omics	Single-cell
<b>Links</b>														
Source Code link	<a href="https://github.com/intanzuckerberg/corpora-data-portal">https://github.com/intanzuckerberg/corpora-data-portal</a>	<a href="https://github.com/eractiveresearcher/a-data-portal">https://github.com/eractiveresearcher/a-data-portal</a>	<a href="https://github.com/gearlab/CHARTS">https://github.com/gearlab/CHARTS</a>	<a href="https://github.com/genealogic/SCANNER">https://github.com/genealogic/SCANNER</a>	<a href="https://github.com/Guosalvarez/SingleCellScanner">https://github.com/Guosalvarez/SingleCellScanner</a>	<a href="https://github.com/broadinstitute/singlercellportals">https://github.com/broadinstitute/singlercellportals</a>	<a href="https://github.com/eislab-sfair/SingleCellPortals">https://github.com/eislab-sfair/SingleCellPortals</a>	<a href="https://github.com/fchalmers/SingleCellPortals">https://github.com/fchalmers/SingleCellPortals</a>	<a href="https://github.com/rgfranz/Panorama">https://github.com/rgfranz/Panorama</a>	<a href="https://github.com/broadinstitute/ExpressionAtlas">https://github.com/broadinstitute/ExpressionAtlas</a>	NA	<a href="https://drive.google.com/drive/folders/0BxSFidIDhU1amHmSconeNoSks0sonconquer_comparisoE07resourcekey">https://drive.google.com/drive/folders/0BxSFidIDhU1amHmSconeNoSks0sonconquer_comparisoE07resourcekey</a>	<a href="https://qithub.co/uI1amhttps://github.com/HumanCellSmpMdE07resourcekey">https://qithub.co/uI1amhttps://github.com/HumanCellSmpMdE07resourcekey</a>	<a href="https://drive.google.com/drive/folders/0BxSFidIDhU1amHmSconeNoSks0sonconquer_comparisoE07resourcekey">https://drive.google.com/drive/folders/0BxSFidIDhU1amHmSconeNoSks0sonconquer_comparisoE07resourcekey</a>
Demo link	<a href="http://celldataportal.bx.psu.edu/">http://celldataportal.bx.psu.edu/</a>	<a href="https://xgene.czi.ismge.gearlab.org/">https://xgene.czi.ismge.gearlab.org/</a>	<a href="https://chart.sanger.ac.uk/">https://chart.sanger.ac.uk/</a>	<a href="https://www.morganridge.org/">https://www.morganridge.org/</a>	<a href="https://theacilab.com/scanner/">https://theacilab.com/scanner/</a>	<a href="https://singlecell.broadinstitutelabslabgitlab.com/">https://singlecell.broadinstitutelabslabgitlab.com/</a>	<a href="https://slabgitlab.rgfranz/Panorama">https://slabgitlab.rgfranz/Panorama</a>	<a href="https://rgfranz/Panorama">https://rgfranz/Panorama</a>	<a href="https://rgfranz/Panorama">https://rgfranz/Panorama</a>	<a href="https://pangolin.ebi.ac.uk/">https://pangolin.ebi.ac.uk/</a>	<a href="https://www.ncbi.nlm.nih.gov/geo/">https://www.ncbi.nlm.nih.gov/geo/</a>	<a href="https://bioinfon.uzh.ch/geneexpressiongroup/atlas/">https://bioinfon.uzh.ch/geneexpressiongroup/atlas/</a>	<a href="http://iml.a.humanomics.ac.jp/">http://iml.a.humanomics.ac.jp/</a>	