

Supplementary Materials - 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*

2021-12-09

Contents

1	Preface	3
2	Supplemental Tables	4
	Table S1 - Comparison matrix of web portal tools	4
	Table S2 - Criterion for query search	6
	Table S3 - Criterion for data analysis explorer	6
	Table S4 - Project metadata captured in CellDepot	6
3	Supplemental Figures	7
4	Supplemental Tutorial	10
4.1	Browse Projects	10
4.2	Search Projects	11
4.3	Project Filters	12
4.4	Visualize Datasets	13
4.5	Search Genes	16
4.6	Import Projects	18
4.7	Create Project	19
4.8	Update a Project	19
5	Getting start with CellDepot	21
5.1	Sources of annotation and metadata	21
5.2	Data format, availability, and preparation	21
5.3	CellDepot platform and installation	21
5.4	How to set up cron job?	21
5.5	CellDepot API (Application Programming Interface)	22
5.6	Code availability	22
5.7	Online tutorials	22

Chapter 1

Preface

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

Chapter 2

Supplemental Tables

Table S1 - Comparison matrix of web portal tools

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

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

	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	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	Y
Query Search	Advanced	NA	Basic I	NA	Intermediate	Advanced	Basic II	Intermediate	Advanced	Intermediate	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	NA
Visualization features	Violin, dot, heatmap, bar, QC, scatter, density, embedding plots	Bar, scatter, density, embedding plots	Violin, dot, line, bar, QC, embedding plots	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	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	Single-cell	Single-cell	Single-cell	Single-cell	Multi-omics	Single-cell	Single-cell	Single-cell	Single-cell	Single-cell	Single-cell	Single-cell, Multi-omics
Links															
Source Code link	https://github.com/intan-zuckert/eracterevereg/corpora-data-portal	https://github.com/chuanxiao/xege_czis_umgear.org	https://github.com/stewarts-m/	https://gitlab.com/rt-s/morgridge.org	https://gitlab.com/Guangming_scanner/	https://gitlab.com/rt-huaiCai/scalable_cell_portal/	https://gitlab.com/institute/single_cell_portal/	https://gitlab.com/fair-islab/sfair-a-portal/	https://gitlab.com/rgel/RGV/	https://gitlab.com/oscar-franzen/PanglaoDB/	https://gitlab.com/ebi-efo/resource-expression-group/atlas/	NA	https://drive.google.com/drive/folders/0B_xSFidIDm/csonquer/compariso_n1	https://drive.google.com/drive/folders/0B_xSFidIDm/csonquer/compariso_n1/e0?resource=e7Quw02gR8VhkS7hfdE9nPq	
Demo link	http://celldxepot.m/	https://cellxgene.cziecisms.org/	https://chartlab.m/	https://www.morganridge.com/scanner/	https://singlecell.broadinstabigithub.io/	https://theceilab.com/scanner/	https://singlecell.safrairPortal/	https://panglao.org/sin gle_cell/	https://rgv.ge/nouest.org/	https://panglao.org/aodbs.e/	https://www.ebi.ac.uk/gx/3838/	https://bioinfon.uzh.ch/bioinfon.uzh.ch/	https://o.uth.edu/scellatlas/conquer/	https://spenictic.org/	https://datagenebrowser.org/

Table S2

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

Table S3

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

Table S2 - Criterion for query searchTable link: <https://github.com/interactivereport/CellDepot/blob/gh-pages/bookdown/S2.csv>**Table S3 - Criterion for data analysis explorer**Table link: <https://github.com/interactivereport/CellDepot/blob/gh-pages/bookdown/S3.csv>**Table S4 - Project metadata captured in CellDepot**Table link: <https://github.com/interactivereport/CellDepot/blob/gh-pages/bookdown/S4.csv>

Table S4

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	

Chapter 3

Supplemental Figures

CellDepot Import Projects Create Project Browse Projects Search Projects Project Filters Search Genes Admin Tools Tutorial Hello, Admin User Sign Out

Search: Please select a field to search: Contains + Add Search Condition

Main functionalities available to all users

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

Q Search Browse All Records Reset Column Settings Export Selected Export All Delete Project Search Genes

Showing 271 records: Display 25 records per page Click on a column to sort Search:

Actions	Custom Accession	Species	Year	DOI	Cell Count	Gene Count	Publication Title
<input type="checkbox"/> View Details Cellxgene VIP Update	GSE103892	mouse	2018	10.1101/2018.02.003	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 Update	GSE118614_10X	mouse	2019	10.1101/2019.04.010	107,052	27,998	Single-Cell RNA-Seq Analysis of Retinal Development Identifies NFI Factors as Regulating Mitotic Exit and Late-Born Cell Specification
<input type="checkbox"/> View Details Cellxgene VIP Update	GSE118614_NFI	mouse	2019	10.1101/2019.04.010	26,537	27,998	Single-Cell RNA-Seq Analysis of Retinal Development Identifies NFI Factors as Regulating Mitotic Exit and Late-Born Cell Specification
<input type="checkbox"/> View Details Cellxgene VIP Update	GSE118614_Smart	mouse	2019	10.1101/2019.04.010	747	28,446	Single-Cell RNA-Seq Analysis of Retinal Development Identifies NFI Factors as Regulating Mitotic Exit and Late-Born Cell Specification
<input type="checkbox"/> View Details Cellxgene VIP Update	GSE132042	Mouse	2018	10.1101/2018.08.09.201809	245,389	20,138	Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris
<input type="checkbox"/> View Details Cellxgene VIP Update	GSE132042	Mouse	2018	10.1101/2018.08.09.201809	110,824	22,966	Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris
<input type="checkbox"/> View Details Cellxgene VIP Update	GSE140231	Human	2020	10.1101/2020.02.17.201786	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 Update	GSE143877	Mouse	2020	10.1101/2020.02.17.201786	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 Update	GSE145121	Human	2020	https://doi.org/10.1101/2019.12.31.892166	4,355	15,073	
<input type="checkbox"/> View Details	GSE146373_D85	human	2021	10.1101/2021.02.21.208923	4,370	16,088	GSE146373_D85

Showing 1 to 25 of 271 records Previous 1 2 3 4 5 ... 11 Next

Figure 3.1: Figure S1. Data Filtering. Query search of ‘Species is Human’ and ‘Annotation Groups contains Neuron’ brings about nine datasets of interest.

The screenshot shows the CellDepot web application interface. At the top, there is a navigation bar with links for 'CellDepot', 'Browse Projects', 'Search Projects', 'Project Filters', 'Search Genes', 'Hello, Guest User', and 'Sign Out'. Below the navigation bar is a search bar with dropdowns for 'Search:' (containing 'Annotation Groups' and 'Contains'), a 'Search' button, and a 'Reset' button. To the right of the search bar is a '+ Add Search Condition' link.

The main content area displays a table of search results. The table has columns for 'Actions', 'View Details', 'Cellxgene VIP', 'Actions', 'Cellxgene VIP Launch Method', 'DOI', 'Gene Count', 'PMC ID', 'Species', 'URL', 'Annotation Groups', 'Custom Accession', 'File Name', 'Name', 'Publication Title', 'Uploaded By', 'Cell Count', 'Description', 'File Size', 'Notes', 'PubMed ID', and 'Year'. The table shows 271 records, with 25 displayed per page. A red box highlights the 'Cellxgene VIP' column header, with the annotation '2. Check to select' pointing to it. Another red box highlights the 'Actions' column header, with the annotation '1. Click "Column Settings" to customize fields to be displayed in the table' pointing to it. A third red box highlights the 'Export Selected' and 'Export All' buttons at the bottom of the table, with the annotation '3. Export whole table or selected entries' pointing to them.

Actions	View Details	Cellxgene VIP	Actions	Cellxgene VIP Launch Method	DOI	Gene Count	PMC ID	Species	URL	Annotation Groups	Custom Accession	File Name	Name	Publication Title	Uploaded By	Cell Count	Description	File Size	Notes	PubMed ID	Year
<input type="checkbox"/>	View Details	Cellxgene VIP	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>									
<input type="checkbox"/>	View Details	Cellxgene VIP	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	View Details	Cellxgene VIP	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	View Details	Cellxgene VIP	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Showing 1 to 25 of 271 records

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.

CellDepot Browse Projects Search Projects Project Filters Search Genes 1. Click "Search Projects"

Search: Species Is Human + Add Search Condition
 6. Pick logic 2. Choose a field to search 3. Choose the condition 4. Type keywords or value
 And Or Annotation Groups Contain Neuron Remove
 5. Add additional conditions for joint query if needed

Q Search Browse All Records Reset
 7. Click "Search"

Column Settings Export Selected Export All Search Genes

Showing 6 records:
 Display 25 records per page

Actions	Custom Accession	Name	Species	Year	DOI	Cell Count	Gene Count	Publication Title	Uploaded By	Uploaded On
<input type="checkbox"/>	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
<input type="checkbox"/>	View Details 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"/>	View Details Cellxgene VIP	SCP199	Comparison of Engineered 3D Neural Tissues	Human, Mouse		4,280	22,567		Admin User	2021-04-02
<input type="checkbox"/>	View Details Cellxgene VIP	SCP498	Transcriptional and Cellular Diversity of the Human Heart	Human		287,269	33,694		Admin User	2021-04-02
<input type="checkbox"/>	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
<input type="checkbox"/>	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 Previous 1 Next

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

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.

The screenshot shows the CellDepot homepage with several highlighted features:

- A red callout box points to the "Search Projects" button in the top navigation bar.
- A red callout box points to the "Main functionalities available to all users" button, located below the search bar.
- A red callout box points to the "Import Projects" and "Create Project" buttons in the top navigation bar.
- A red callout box points to the "Admin Tools" and "Tutorial" buttons in the top navigation bar.
- A red callout box points to the "Column Settings" button at the top of the main table.
- A red callout box points to the "Click on a column to sort" button above the table header.
- A red callout box points to the "Import Projects" and "Create Project" buttons in the bottom right corner of the page.

Actions	Custom Accession	Species	Year	DOI	Cell Count	Gene Count	Publication Title
View Details Cellxgene VIP	GSE103892	mouse	2018	10.1101/2018.02.003	17,998	25,024	Massively Parallel Single Nucleus Transcriptional Profiling Defines Spinal Cord Neurons and Their Activity during Behavior
View Details Cellxgene VIP	GSE118614_10X	mouse	2019	10.1101/2019.04.010	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	GSE118614_NFI	mouse	2019	10.1101/2019.04.010	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	GSE118614_Smart	mouse	2019	10.1101/2019.04.010	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	GSE132042	Mouse	2018	10.1101/2018.05.04.180590	245,389	20,138	Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris
View Details Cellxgene VIP	GSE132042	Mouse	2018	10.1101/2018.05.04.180590	110,824	22,966	Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris
View Details Cellxgene VIP	GSE140231	Human	2020	10.1101/2020.02.17.187670	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	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	GSE145121	Human	2020	https://doi.org/10.1101/2019.12.31.892166	4,355	15,073	
View Details	GSE146373_D85	human	2021	10.1101/2021.02.01.208923	4,370	16,088	GSE146373_D85

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).

The screenshot shows the CellDepot search interface for browsing projects. At the top, there are navigation links: CellDepot, Browse Projects, Search Projects, Project Filters, and Search Genes. On the right, there are links for Hello, Guest User and Sign Out.

Below the header, there is a search bar with dropdowns for 'Search:' (Annotation Groups, Contains), 'Annotation Groups' (Contains), and 'Project Filters' (Search Condition). A red callout '3. Export whole table or selected entries' points to the 'Export All' button in the toolbar.

The main area displays a table of 271 records. The first column contains checkboxes and 'View Details' links for each project. A red callout '1. Click "Column Settings" to customize fields to be displayed in the table' points to the 'Column Settings' link in the toolbar. Another red callout '2. Check to select' points to the checkbox in the first row of the table.

The 'Column Settings' dialog box is open, showing a list of columns to select from. It includes checkboxes for Actions, Cellxgene VIP, View Details, DOI, Gene Count, Species, URL, Annotation Groups, Custom Accession, File Name, Name, Publication Title, Uploaded By, Year, Cell Count, Description, File Size, Notes, PubMed ID, and Uploaded On. An 'Apply' button is at the bottom right of the dialog.

The table itself has columns for Actions, Release (e.g., 1.0 (Thymus)), Species (e.g., Human), Year, DOI, File Size, Number of Rows, Number of Genes, Description, Admin User, and Date. Two rows are shown as examples:

<input type="checkbox"/>	View Details	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 User	2021-09-16
<input type="checkbox"/>	View Details	TS_Small_Intestine	Tabula Sapiens	Human	2021	10.1101/2021.07.19.452956	2,037	58,833	The Tabula Sapiens: a multiple organ single cell transcriptomic atlas of humans	Admin User	2021-09-16

At the bottom, a message says 'Showing 1 to 25 of 271 records' and a navigation bar shows pages 1 through 11.

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.

The screenshot shows the CellDepot search interface. At the top, there are navigation links: CellDepot, Browse Projects, Search Projects, Project Filters, and Search Genes. On the right, there are user options: Hello, Guest User, Sign Out.

Search Projects:

- Search: Species Is Human
- Add Search Condition
- 6. Pick logic: And (radio button selected)
- 7. Choose a field to search: Annotation Groups
- 3. Choose the condition: Contain
- 4. Type keywords or value: Neuron
- Remove
- + Add Search Condition
- 5. Add additional conditions for joint query if needed

Action Buttons:

- Q. Search (highlighted in blue)
- Browse All Records
- Reset
- Column Settings
- Export Selected
- Export All
- Search Genes

Results:

Showing 6 records: Display 25 records per page

Actions	Custom Accession	Name	Species	Year	DOI	Cell Count	Gene Count	Publication Title	Uploaded By	Uploaded On
<input type="checkbox"/>	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"/>	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"/>	SCP199	Comparison of Engineered 3D Neural Tissues	Human, Mouse			4,280	22,567		Admin User	2021-04-02
<input type="checkbox"/>	SCP498	Transcriptional and Cellular Diversity of the Human Heart	Human			287,269	33,694		Admin User	2021-04-02
<input type="checkbox"/>	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"/>	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

Search:

Previous 1 Next

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.

The screenshot shows the 'Project Filters' page on the CellDepot website. At the top, there are navigation links: 'CellDepot', 'Browse Projects', 'Search Projects', 'Project Filters', and 'Search Genes'. On the right, there are 'Hello, Guest User' and 'Sign Out' links. Below the header, the title 'Project Filters' is displayed. On the left, there are two filter panels: 'Year' and 'Species'. The 'Year' panel shows a list of years from 2015 to 2021, with '2021' selected. The 'Species' panel shows a list of species including Human, Mouse, Others, Rat, Hydra, Axolotl, and MacaqueF, with 'Human' selected. In the center, a table lists 29 records. The columns are: Actions, Custom Accession, Name, Species, Year, DOI, Cell Count, Gene Count, Publication Title, and Uploaded. Each row represents a dataset, such as 'Tabula Sapiens release 1.0 (Complete Dataset)' or 'Tabula Sapiens release 1.0 (Vasculature)'. All datasets listed are for the Human species and were uploaded in 2021. The table includes buttons for 'View Details' and 'Cellxgene VIP' for each entry.

	Actions	Custom Accession	Name	Species	Year	DOI	Cell Count	Gene Count	Publication Title	Uploaded
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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
<input type="checkbox"/>	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

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:

Display 25 records per page

Actions (1) Custom Accession

View Details SCP1

Cellxgene VIP

Project Annotation Groups Broad scRNA Portal

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... (Show All)

Notes: -

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

Year: 2016

DOI: [10.1126/science.aad7038](https://doi.org/10.1126/science.aad7038)

PMC ID: PMC5480621

PubMed ID: 27471252

Cellxgene VIP: Standard

Launch Method:

All Annotation Groups Filter: CLUSTER Seurat Clusters SUB CLUSTER

CLUSTER

Display 50 records per page

Category Cell Count

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

Showing 1 to 8 of 8 records

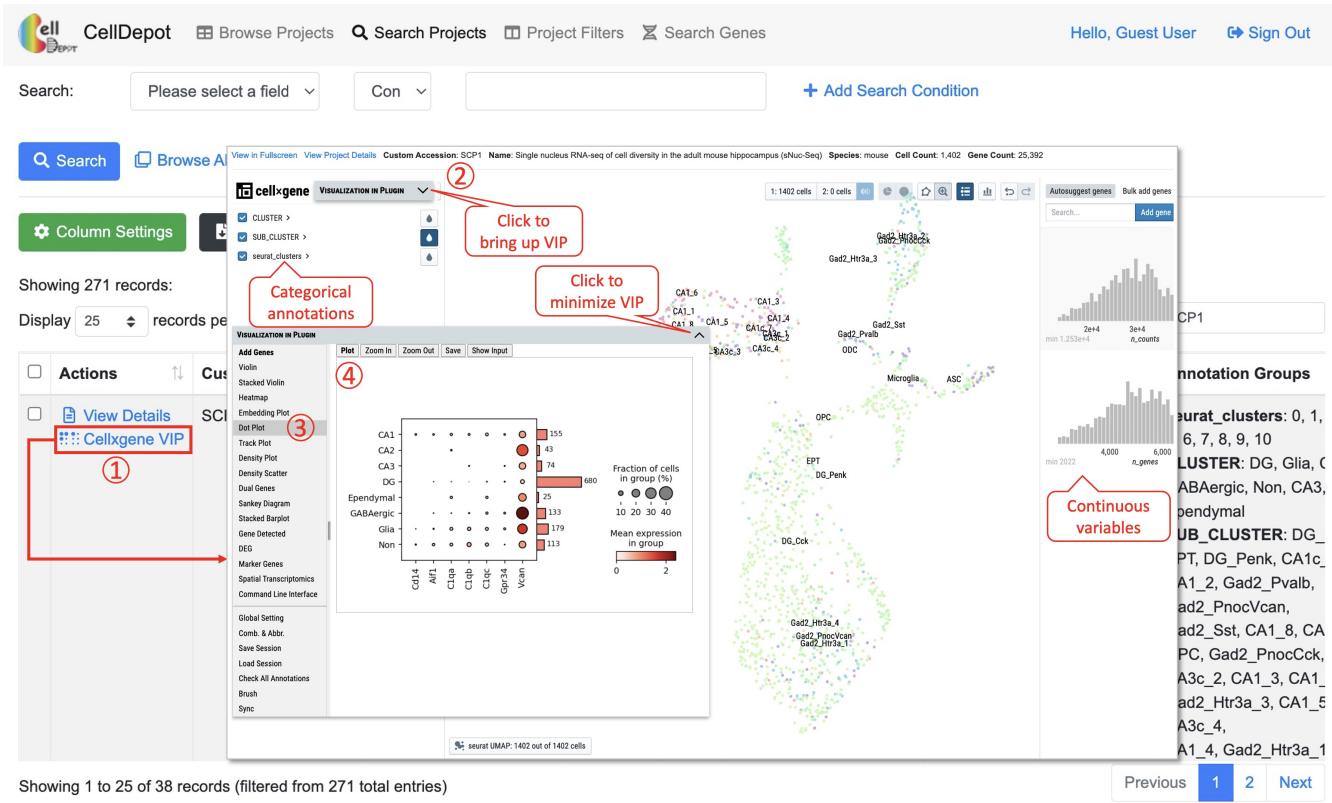
Gene Count Annotation Groups Publication Title Upload

Gene Count	Annotation Groups	Publication Title	Upload
25,392	seurat_clusters: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 CLUSTER: DG, Glia, CA1, GABAergic, Non, CA3, C... 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 neurons	Admin

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.



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

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 PTPNC3 is expressed mainly in astrocytes and endothelial cells (Figure S7b).

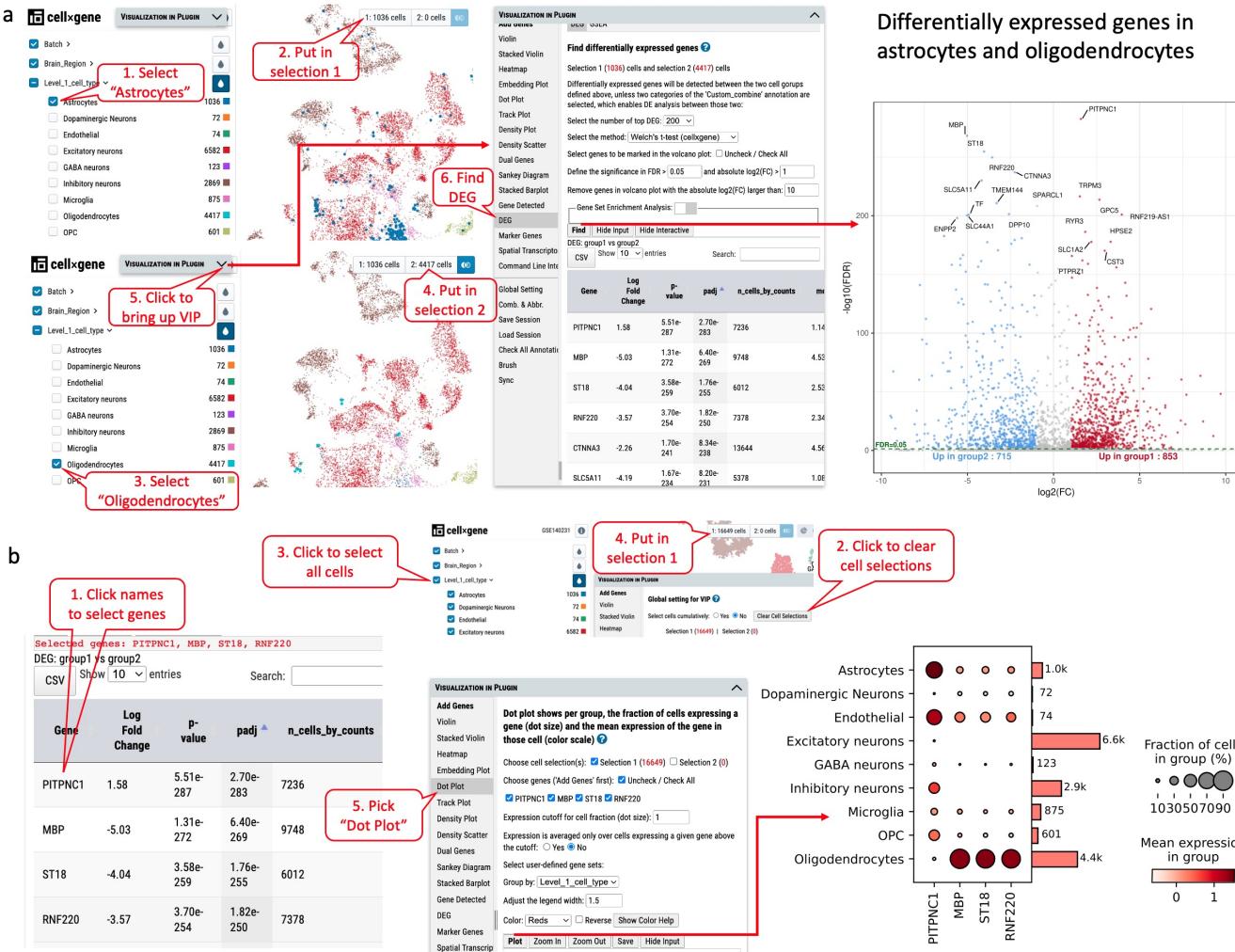


Figure S7. Exploration of differentially expressed genes in dataset GSE140231 through cellxgene VIP. (a) Identifying differentially expressed genes in astrocytes and oligodendrocytes. (b) The expression of top four genes in various cell types as shown in dot plot.

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.

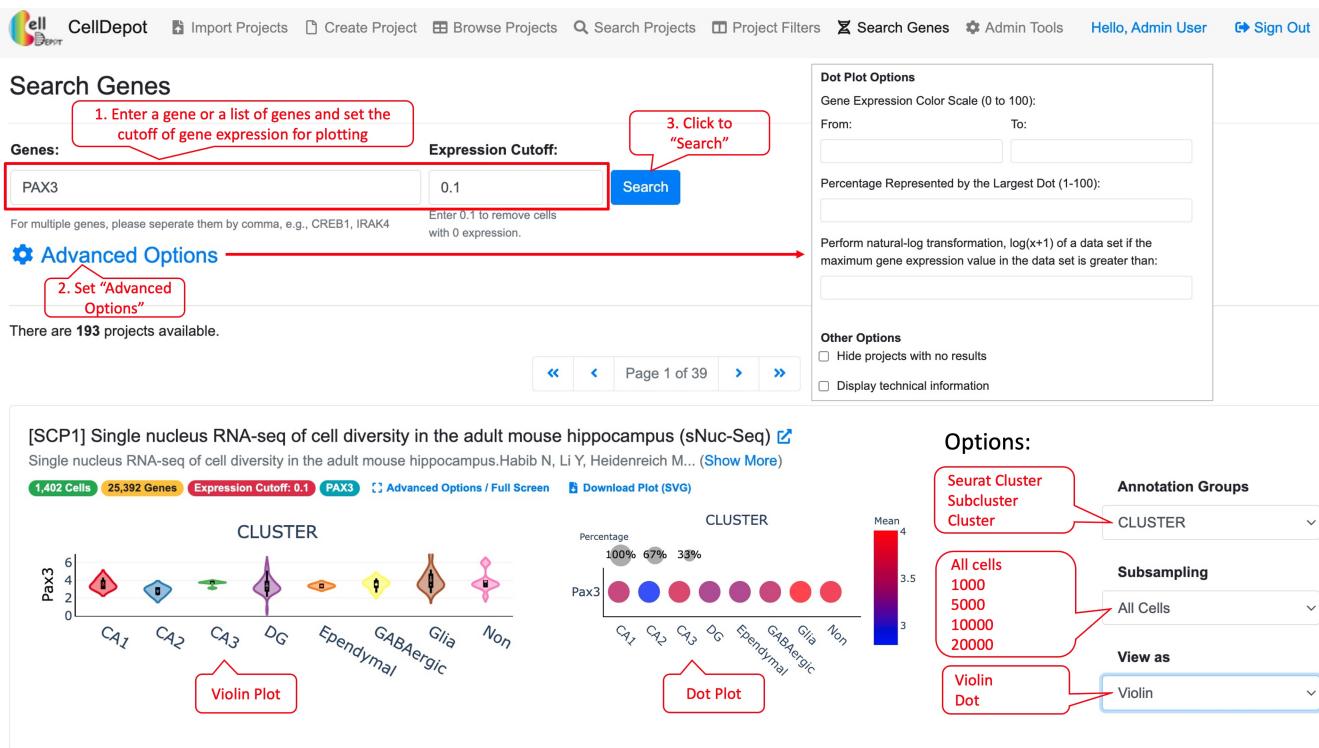


Figure S8. Steps to find gene expression level of a gene in projects under ‘Search Genes’ tab. The final plot can be customized by available options listed in red callout boxes on the left side.

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.

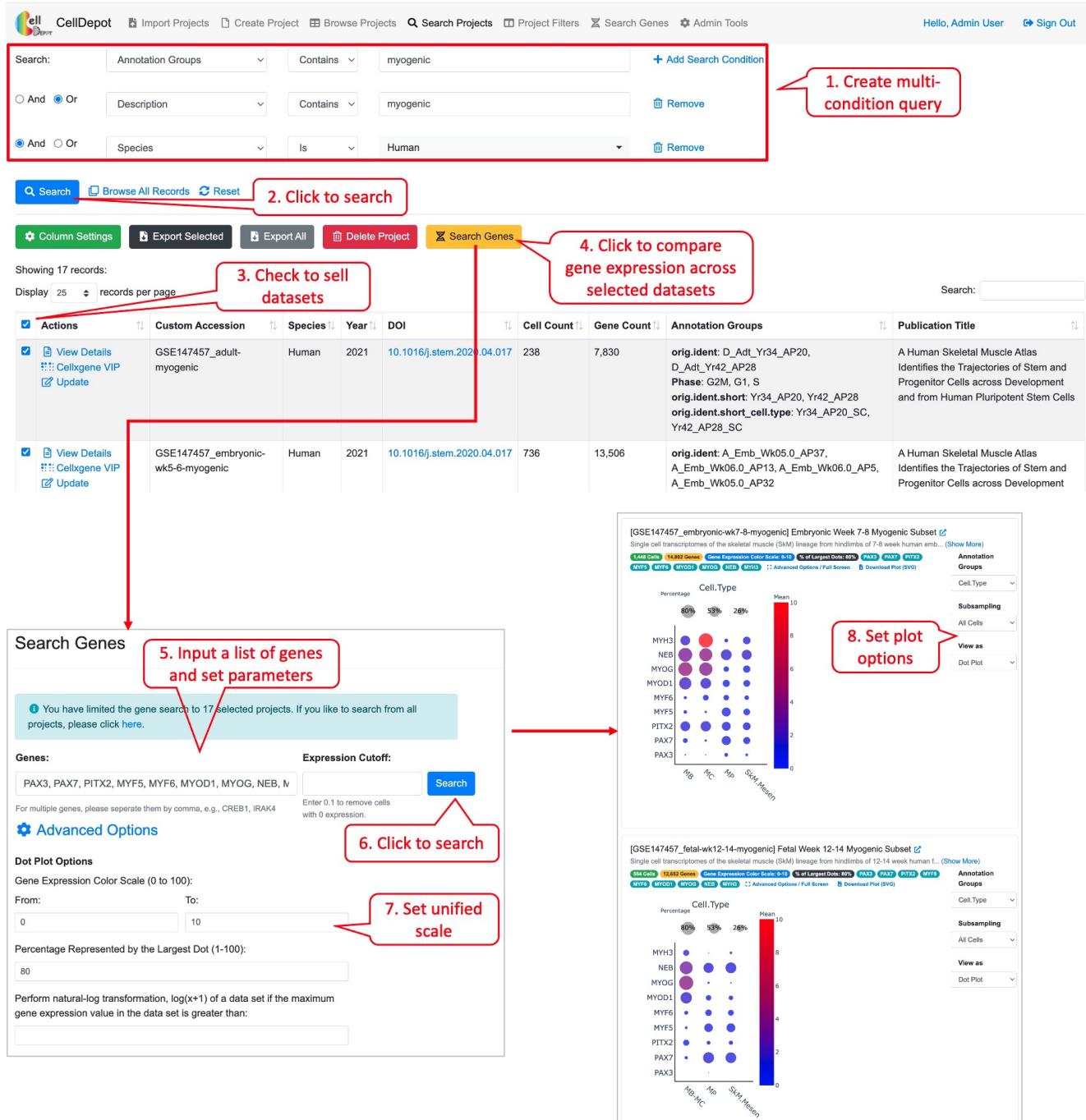


Figure S9. Workflow of conducting the cross-project comparison of a list of genes among the selected datasets.

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/celld depot/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 (5.4). To explore the detail of imported datasets, users can enter ‘Browse Projects’ page and then search these datasets by user assigned ‘Custom Accession’ identifiers.

The screenshot shows the 'Import Projects' page of the CellDepot website. At the top, there is a navigation bar with links like 'CellDepot', 'Import Projects', 'Create Project', 'Browse Projects', 'Search Projects', 'Project Filters', 'Search Genes', 'Admin Tools', 'Hello, Admin User', and 'Sign Out'. A red box highlights the 'Import Projects' link with the instruction '1. Click "Import Projects"'. Below the navigation is a message: 'Please make sure that the h5ad files are available here: /raid/data/single_cell_portal/all_h5ad_files/'. A red box highlights the 'Download Example File' button with the instruction '2. Download example file and modify based on your datasets'. Next, a red box highlights the 'Choose File' input field with the instruction '3. Upload the modified file'. Below these, a dropdown menu for 'Cellxgene VIP Launch Method' is shown with 'Standard' selected; a red box highlights it with the instruction '4. Select "Preload in memory" for frequently used datasets, otherwise "Standard"'. At the bottom left are 'Submit' and 'Start Over' buttons, with a red box highlighting 'Submit' and the instruction '5. Click to submit'. The main area shows a table with columns A through L, containing data for three projects: SCP1, Single nucleus RNA-se mouse; SCP10, Glioblastoma intra-tumoural human; and another entry for SCP1. At the bottom, there is a search interface with fields for 'Search', 'Custom Accession', 'Is', and 'SCP1', and a 'Search' button. A red box highlights the 'Search' button with the instruction '6. Search a dataset by "Custom Accession"'. Below the search interface is a table showing one record for SCP1, with columns for Actions, Custom Accession, Species, Year, DOI, Cell Count, Gene Count, Annotation Groups, and Publication Title. The 'Actions' column contains links for View Details, Cellxgene VIP, and Update, with a red box highlighting 'Update'.

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.

CellDepot Import Projects Create Project Browse Projects Search Projects Project Filters Search Genes Admin Tools Hello, Admin User Sign Out

Search: Custom Accession Is S + Add Search Condition

Search Browse All Records Reset

Column Settings Export Selected Export All

Showing one record:

Display 25 records per page

Actions	Custom Accession	Species	Year	Count	Annotation Groups	Publication Title
<input type="checkbox"/> View Details Cellxgene VIP	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 →

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-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.aad7038

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 neurons

Year: 2016

DOI: 10.1126/science.aad7038

PMC ID: PMC5480621

PubMed ID: 27471252

Cellxgene VIP Launch Method: Standard

Chapter 5

Getting start with CellDepot

5.1 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>.

5.2 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)

5.3 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.

5.4 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.

5.5 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.

5.6 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.

5.7 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.