ScPanStroma：ScRNA-seq of Pan-cancer Stromal Cells

1. Brief introduction

ScPanStroma is a database and interactive website for analyzing tumor microenvironment characteristics based on single-cell RNA sequencing (scRNA).

ScPanStroma has collected single-cell transcriptome data of stromal cells from 258 tumor patients across 16 different cancer types. The data from different samples are aggregated and batch-corrected in ScPanStroma. Subsequently, stromal cells are classified based on classic cellular biomarkers into: Fibroblasts, Endothelial cells, Mural cells, Mesothelial cells, Proliferating cells, and Glial cells. Furthermore, the three main groups (Fibroblasts, Endothelial cells, and Mural cells) are subdivided into subgroups based on distinct gene expressions. Additionally, differential gene analysis is performed

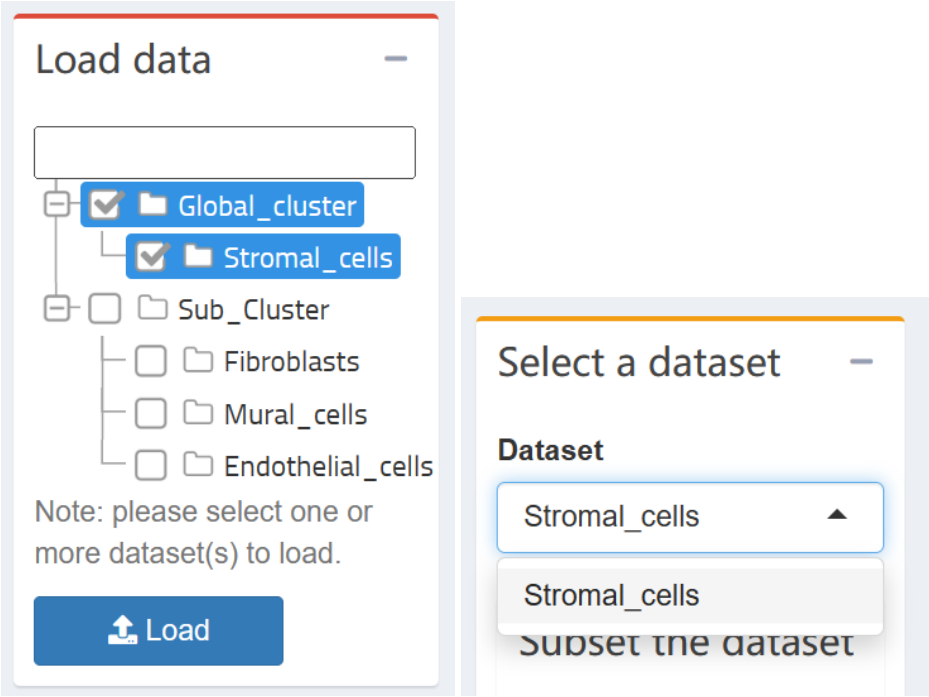
for both the major groups of stromal cells and the three subgroups within each cancer type and across pan-cancer dimensions.

Addressing these analytical outcomes, ScPanStroma offers six interactive analysis modules: Gene Expression Profiling, Distribution Analysis, Differential Gene Analysis, Heatmap Analysis, Data Composition Analysis, and Data Tables.

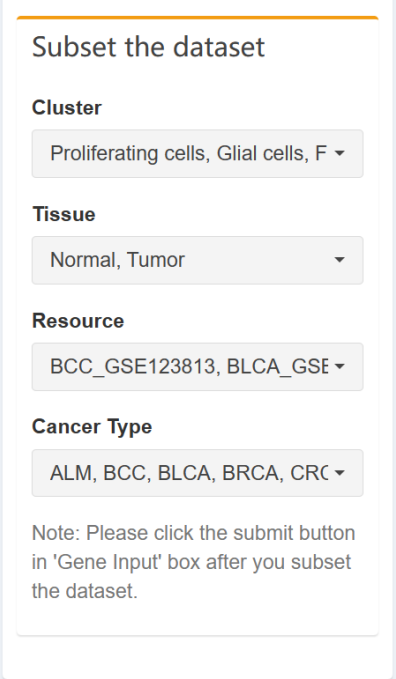
2. Data Loading

Click 'Search' on the page to enter the retrieval page, which includes a combination of option boxes. After selecting the cancer type and dataset of interest in the option box, click the 'Search' button. The page will display all ST sample information corresponding to the selected options. Each sample information has a blue 'Analysis' button in front of it, which, when clicked, redirects to the analysis interface for that sample.

All available datasets are listed in the top right corner of the page (the 'Load data' panel). Users should first select the dataset of interest and then click the 'Load' button. Loading a large expression matrix takes some time. Different f can be loaded multiple times, but deselecting a loaded dataset will not release it from memory. Once loading is complete, the loaded datasets are available in the 'Select a dataset' panel below.

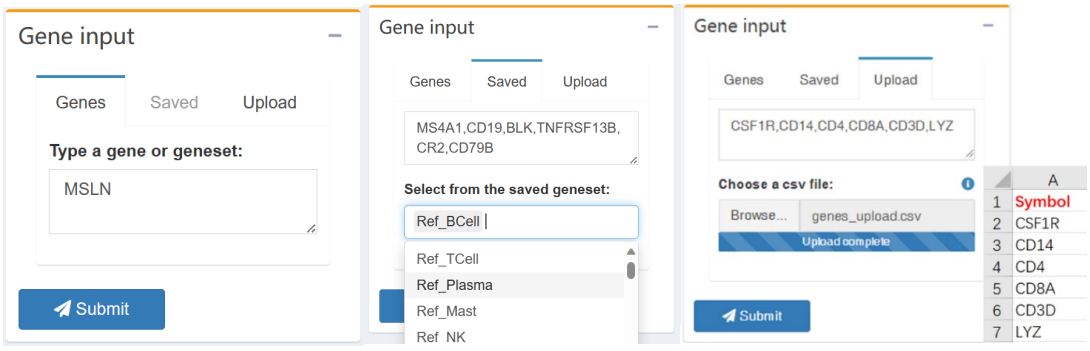


Additionally, in the 'Subset the dataset' panel, the dataset can be filtered to focus more on the data of interest. Users can filter the dataset based on cell clustering, tissue (tumor and adjacent normal tissue), data source, and cancer type. The final subset of the selected dataset is generated by the intersection of all filtering conditions.



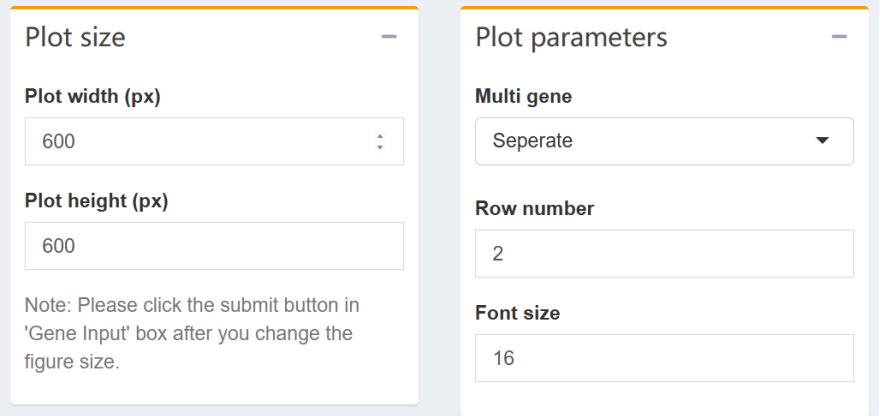
3. Input of Plotting Parameters

Before proceeding with graph analysis, users should enter genes of interest in the 'Gene input' panel to view the data. Here, we offer three different gene input modes: typing in genes of interest, selecting from a pre-stored gene list, or uploading a CSV file. Regardless of the method used, ScPanStroma only accepts case-insensitive gene names (Symbol) or a comma-separated list of gene names as input. If you choose to upload a CSV file or select from a pre-stored gene list, all genes on the list will appear in the text box below and can be manually edited. It's worth noting that the column name containing gene names in the CSV file should be 'Symbol'. After clicking the 'Submit' button in the 'Gene input' panel, all changes to the 'Subset the dataset' panel and 'Gene input' panel will be received.



Users can also adjust plotting parameters at the top of the page. Similarly, changes in 'Plot size' only take effect after clicking the 'Submit' button. When entering multiple genes as input, you can change the "Multi gene" option to "Geometric mean" and present the geometric mean of the expression levels of all input genes as a score. You

can also use the "Row number" option to adjust the number of genes drawn in a column.

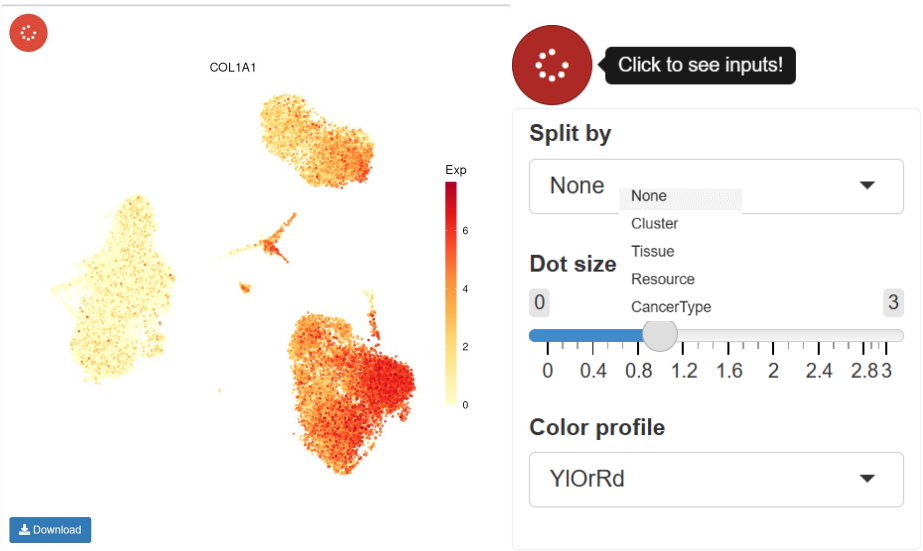


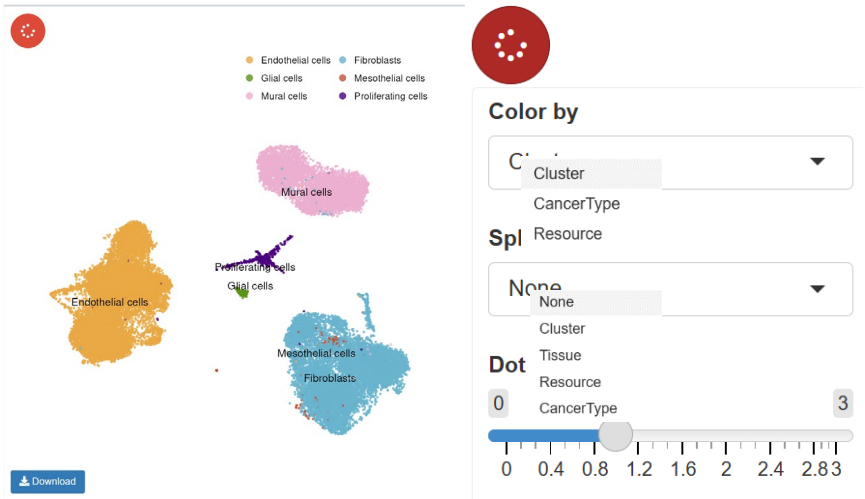
4. Analysis Modules

4.1 Gene Expression Profiling

Users can click 'Embedding' in the side menu to enter the Gene Expression Profiling page. Here, UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction) coordinates are used to explore the gene expression levels or metadata of each cell in two-dimensional space. Gene expression maps and metadata maps are vertically distributed in the plotting area.

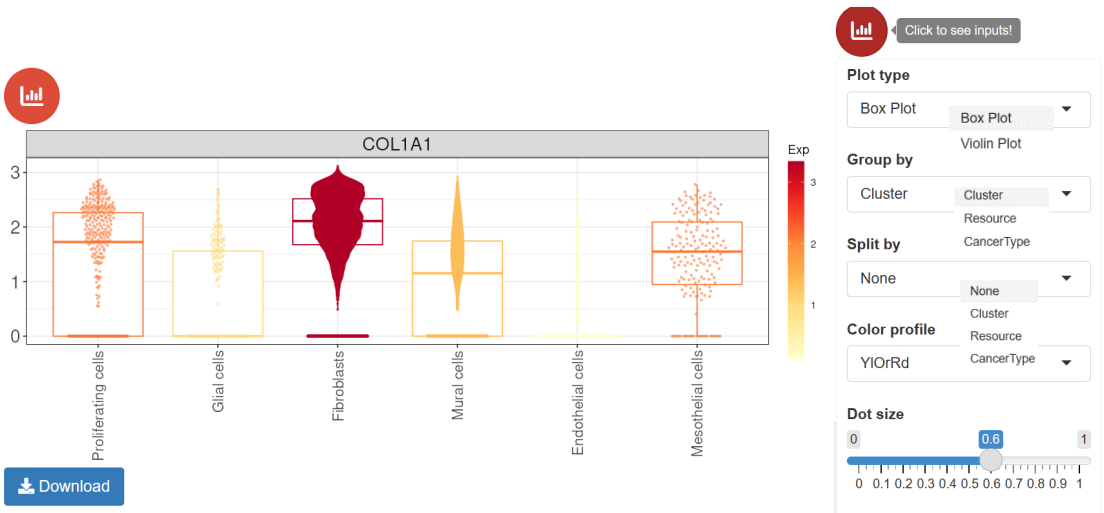
The adjustable data plotting parameters are listed in the red circular button in the top left corner of the plotting area. Users can view data points divided into different groups based on Cluster, Tissue, Resource, CancerType in the metadata. In the metadata map, you can specify Cluster, Resource, CancerType for coloring data points. Additionally, you can configure point size and color profiles.





4.2 Distribution Analysis

Users can click 'Distribution' in the side menu to enter the Distribution Analysis page. Here, you can switch between box plots and violin plots using the 'Plot type' option and group all data points according to Cluster, Resource, CancerType in the metadata using the 'Group by' option. Additionally, selecting the 'Split by' option allows you to split the data graph by different groups.



4.3 Differential Gene Analysis

Users can click 'Signature Gene' in the side menu to enter the Differential Gene Analysis page. Here, we provide a table of differential genes in all cancer types for each cell group (calculated using the findallmarker function from the Seurat package). Additionally, after selecting a differential gene of interest in the 'Gene input' panel, a chart will appear below the table showing the log2FC (Fold change) values of that gene in specified cell groups (specified in the 'Cluster' option) across various cancer types and pan-cancer.

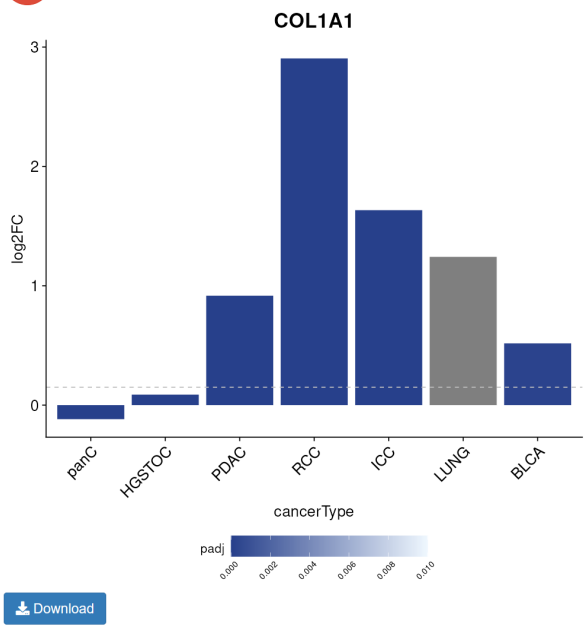
Full talbe of the signature gene detection

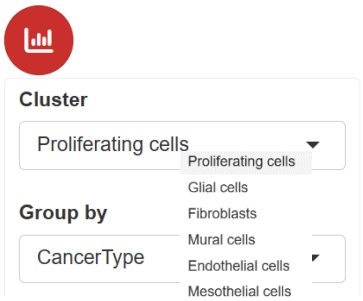
Show 10 entries Search:

	gene	cluster	comb.log2FC	comb.padj	comb.log2FC.rank	sig	cancerType.sig.N	cancerType.s
1	PLVAP	Endothelial cells	4.00411375126914	0	1	true	16	100.00%
2	CD74	Endothelial cells	3.58542589372337	0	2	true	16	100.00%
3	VWF	Endothelial cells	3.34843921837245	0	3	true	15	93.75%
4	RAMP2	Endothelial cells	3.27453137512854	0	4	true	15	93.75%
5	CLDN5	Endothelial cells	3.22743778217614	0	5	true	16	100.00%
6	RNASE1	Endothelial cells	3.19475042029817	0	6	true	16	100.00%
7	HLA-DRB1	Endothelial cells	3.14287384827321	0	7	true	15	93.75%
8	PECAM1	Endothelial cells	3.07874331131076	0	8	true	14	87.50%
9	FABP5	Endothelial cells	3.0584751173984	0	9	true	11	68.75%
10	ACKR1	Endothelial cells	2.89428769636794	0	10	true	14	87.50%

Showing 1 to 10 of 39,105 entries Previous 1 2 3 4 5 … 3,911 Next

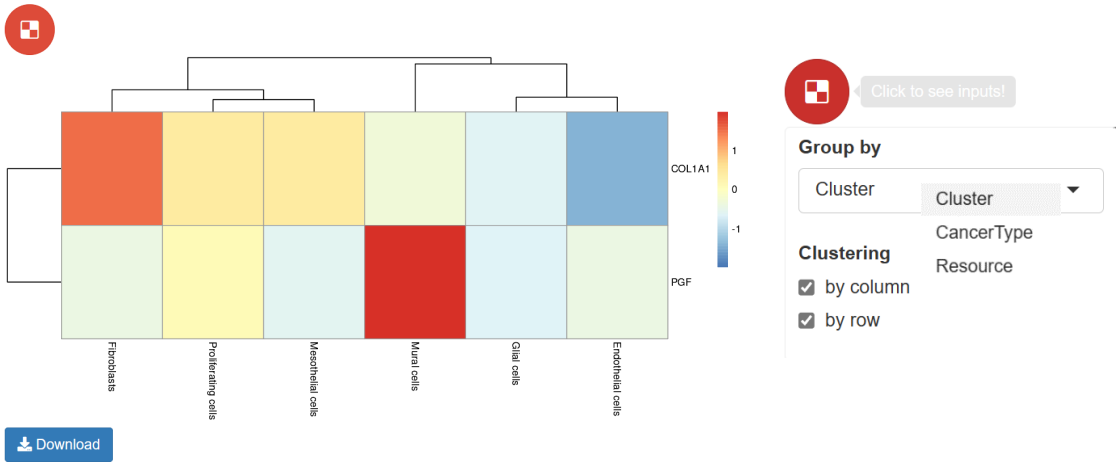
Selected genes and meta.cluster visulized here:





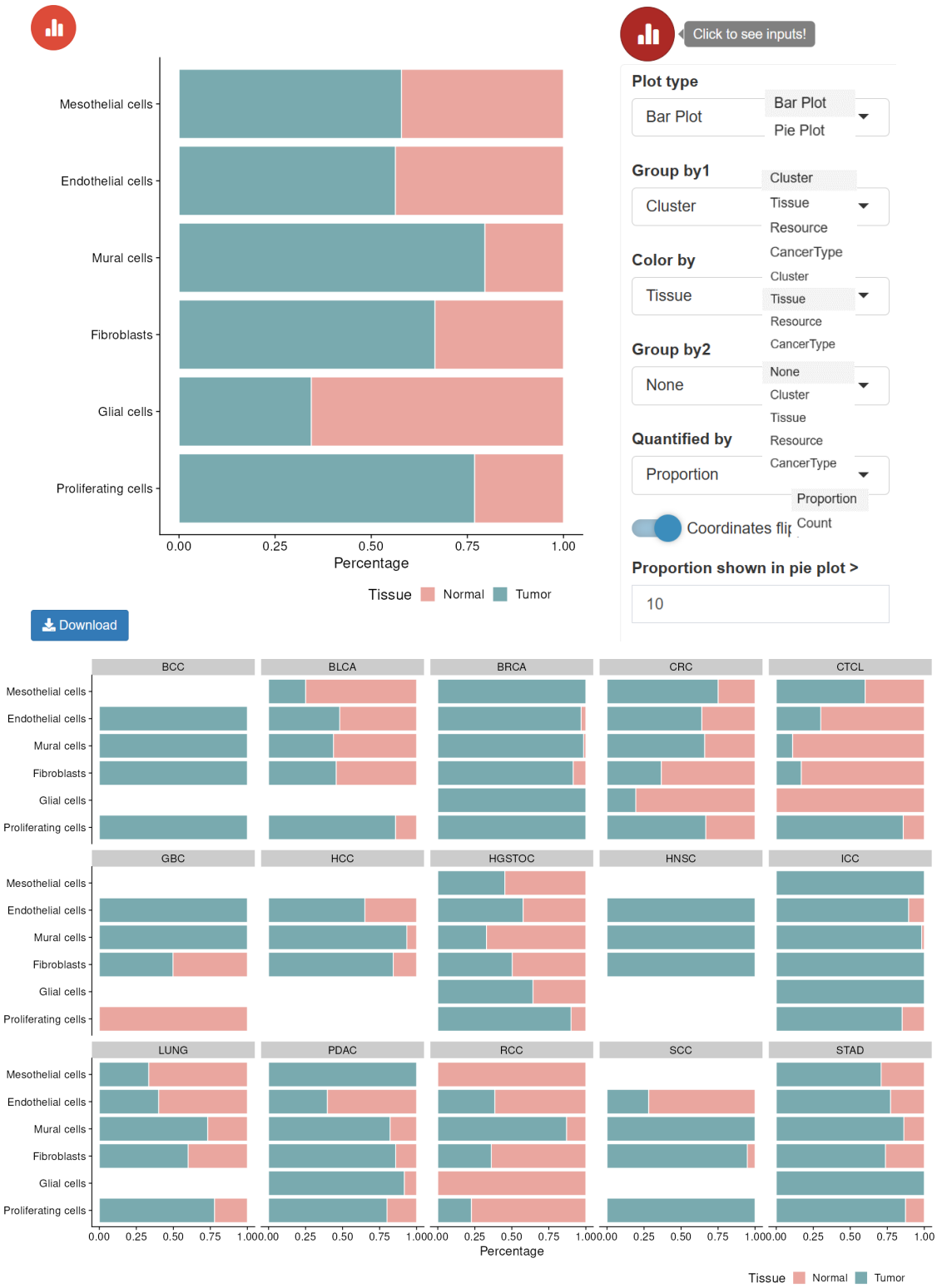
4.4 Heatmap

Users can click 'Heatmap' in the side menu to enter the Heatmap Analysis page. Here, users can select two or more genes of interest in the 'Gene input' panel, and the expression values of the selected genes will be displayed in a heatmap format across different groups (selected in the 'Group by' option). The median clustering of each gene is obtained by group, and the median clustering is z-scored between groups.

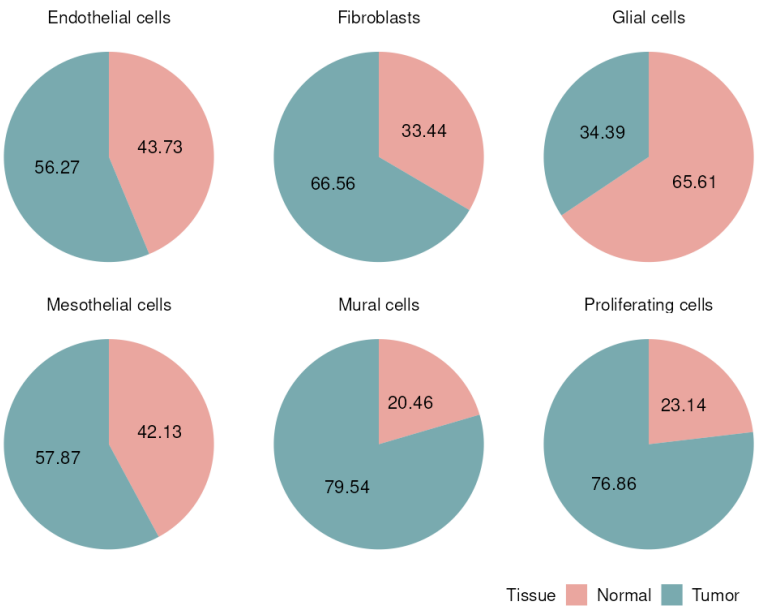


4.5 Metadata

Users can click 'Metadata' in the side menu to enter the Metadata Analysis page. Here, users can explore various metadata distribution combinations. For example, users can group cells by Cluster ('Group by1' option) and calculate tissue distribution in each group ('Color by' option). If 'Group by2' is not set to 'None', the data will be further subdivided, and a grouped chart will be displayed in the plotting area. For instance, you can analyze the tissue distribution of each group of cells in different cancer types as shown in the figure. Sometimes, when the absolute number of cells is very small, this proportion can be confusing or misleading. Therefore, we provide a 'Count' mode in the 'Quantified by' item, which shows the absolute number of data.



Users can select 'Pie plot' in 'Plot type' to explore cell distribution. In pie plot mode, 'Group by2' and 'Quantified by' items are no longer effective. Additionally, we provide the calculation of 'Ro/e', which is the ratio of the observed number of cells in 'Color by' to the expected number of cells for a given 'Group by1'. (The expected number of cells for each combination of 'Group by1' and 'Color by' is obtained from the Chi-squared test.)



4.6 Data Table

Users can click 'Data Table' in the side menu to enter the Data Table page. We integrate input genes and metadata into a data table that can be searched, sorted, and downloaded. The 'Expression' column in the data table indicates the geometric mean expression of all input genes in each cell.

Show 10 entries Search:

	Expression	CellName	nUMI	nGene	CancerType	
edicine_OV_stromal_BT1303_AAACCTGAGTGCAAGC	1.18967433129024	HGSTOC_GMedicine_OV_stromal_BT1303_AAACCTGAGTGCAAGC	2991	1432	HGSTOC	N
edicine_OV_stromal_BT1303_AAACGGGCAGCTCCGA	1.47670195854306	HGSTOC_GMedicine_OV_stromal_BT1303_AAACGGGCAGCTCCGA	31831	5655	HGSTOC	N
edicine_OV_stromal_BT1303_AAAGATGTCACCTCGT	0	HGSTOC_GMedicine_OV_stromal_BT1303_AAAGATGTCACCTCGT	3825	1563	HGSTOC	N
edicine_OV_stromal_BT1303_AAAGCAATCAGCGACC	1.66561231893071	HGSTOC_GMedicine_OV_stromal_BT1303_AAAGCAATCAGCGACC	4650	1751	HGSTOC	N
edicine_OV_stromal_BT1303_AAAGTAGAGCTAGTCT	2.08750884802506	HGSTOC_GMedicine_OV_stromal_BT1303_AAAGTAGAGCTAGTCT	12167	3005	HGSTOC	N
edicine_OV_stromal_BT1303_AAAGTAGGTACCAGTT	1.48443883294975	HGSTOC_GMedicine_OV_stromal_BT1303_AAAGTAGGTACCAGTT	11121	3116	HGSTOC	N
edicine_OV_stromal_BT1303_AAAGTAGGTCCGAGTC	1.46473072421819	HGSTOC_GMedicine_OV_stromal_BT1303_AAAGTAGGTCCGAGTC	7738	2696	HGSTOC	N
edicine_OV_stromal_BT1303_AACTCAGGTACGAAAT	0.452539530466522	HGSTOC_GMedicine_OV_stromal_BT1303_AACTCAGGTACGAAAT	9833	2663	HGSTOC	N
edicine_OV_stromal_BT1303_AACTCAGGTCGCTTTC	0.795543703037281	HGSTOC_GMedicine_OV_stromal_BT1303_AACTCAGGTCGCTTTC	1213	627	HGSTOC	N
edicine_OV_stromal_BT1303_AACTCTTCAAGGCTCC	0	HGSTOC_GMedicine_OV_stromal_BT1303_AACTCTTCAAGGCTCC	2327	1032	HGSTOC	N

Showing 1 to 10 of 29,706 entries Previous 1 2 3 4 5 … 2,971 Next

Download

5. Summary

ScPanStroma provides single-cell transcriptome data of stromal cells for 16 cancer types and, based on clustering results, offers six main analysis modules (Gene Expression Profiling, Distribution Analysis, Differential Gene Analysis, Heatmap Analysis, Metadata

Analysis, and Data Tables), enabling users to easily analyze tumor-related stromal cells' single-cell transcriptome data online. Of course, there may still be some defects in this webpage, but we will continuously improve and develop more practical functions in future studies. If you have any suggestions or questions, please feel free to contact us at shijintong@sjtu.edu.cn.