



FlyBase Features for
scRNAseq Data Survey

Survey Summary

Nov 2022



FlyBase Features for scRNAseq Data Survey

Introduction

Since the end of 2021, FlyBase has started to curate single-cell RNA sequencing (scRNAseq) datasets, in collaboration with the European Bioinformatics Institute's Single-Cell Expression Atlas (SCEA) project. You may have seen, in this year's June release (2022_03), the first visible outcome of this effort, the new Cell Types Ribbon (more on this inside the survey).

As we are planning how to further integrate scRNAseq data with the FlyBase.org website, we need your input to do so in a way that would be the most useful for you.



FlyBase Features for scRNAseq Data Survey

Questions

1) Your use of scRNAseq data

* Q1. What do you currently use to explore scRNAseq data? (Check all that apply)

- [The EBI's Single Cell Expression Atlas](#)
- [The SCoPe platform](#)
- [The SIB's Automated Single-cell Analysis Pipeline](#)
- [The UCSC's Cell Browser](#)
- [The DRSC's scRNA-seq DataBase](#)
- I am not really using scRNAseq data in my work.
- Other (please specify)

2) The Gene Report

Since 2021, the Gene Report has offered linkouts (in the External Crossreferences and Linkouts section) to the EBI's Single Cell Expression Atlas (SCEA) and the DRSC's scRNA-seq DataBase (DRScDB), allowing you to query these resources for any data they may have about the gene currently displayed (Figure 1).

Figure 1. Linkouts to external scRNAseq resources in the Gene Report.

External Crossreferences and Linkouts (186)

Other crossreferences

- DRScDB** - A single-cell RNA-seq resource for data mining and data comparison across species
37044/tissue=All
- EMBL-EBI Single Cell Expression Atlas** - Single cell expression across species
FBgn0283437

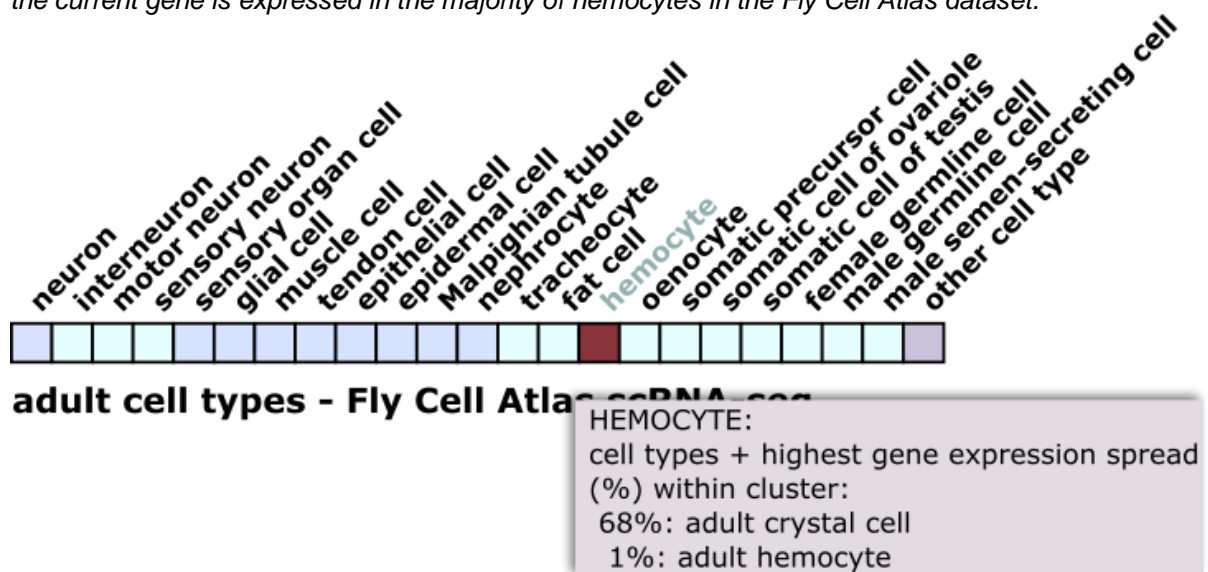
Q2. Did you know you could search for a gene on the SCEA or the DRScDB directly from the Gene Report?

- Yes, and I have already done so.
- Yes, but I have never done so.
- No, I didn't know about this feature.

Since the 2022_03 release (June 2022), we have added a new Expression Summary Ribbon [to complete the existing Anatomy Ribbon and Stages Ribbon](#) (showing the anatomical parts and the developmental stages where a gene is expressed): the Cell Types Ribbon (CTR), showing the cell types where a gene is expressed.

As currently implemented, the Cell Types Ribbon shows expression data from the Fly Cell Atlas project (FCA; [Li et al. 2022](#)). Each tile, representing a major cell type, is coloured based on how many cells of that type were found to express the current gene in the Fly Cell Atlas dataset, with a tooltip giving the exact proportion of cells expressing the gene (Figure 2a).

Figure 2a. The Cell Types Ribbon, as currently implemented. The very bright “hemocyte” tile indicates the current gene is expressed in the majority of hemocytes in the Fly Cell Atlas dataset.

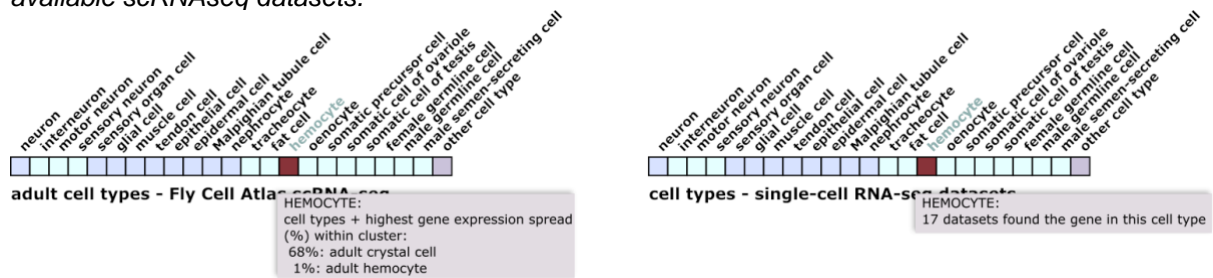


Q3. Had you noticed the Cell Types Ribbon since its introduction earlier this year?

- Yes, and I have already found it useful.
- Yes, but I have not used it a lot.
- No, I am learning about it right now.

As we’ll be accumulating more scRNAseq datasets beyond the Fly Cell Atlas dataset, we’re envisioning switching the Cell Types Ribbon to a new mode where tiles would be coloured based on the number of datasets that found the current gene to be expressed in the corresponding cell type (Figure 2b).

Figure 2b. Proposed new implementation for the Cell Types Ribbon. On the left, the Cell Types Ribbon as currently implemented. On the right, the proposed new behaviour: the very bright “hemocyte” tile would indicate that the current gene was found expressed in hemocytes in most of the currently available scRNAseq datasets.

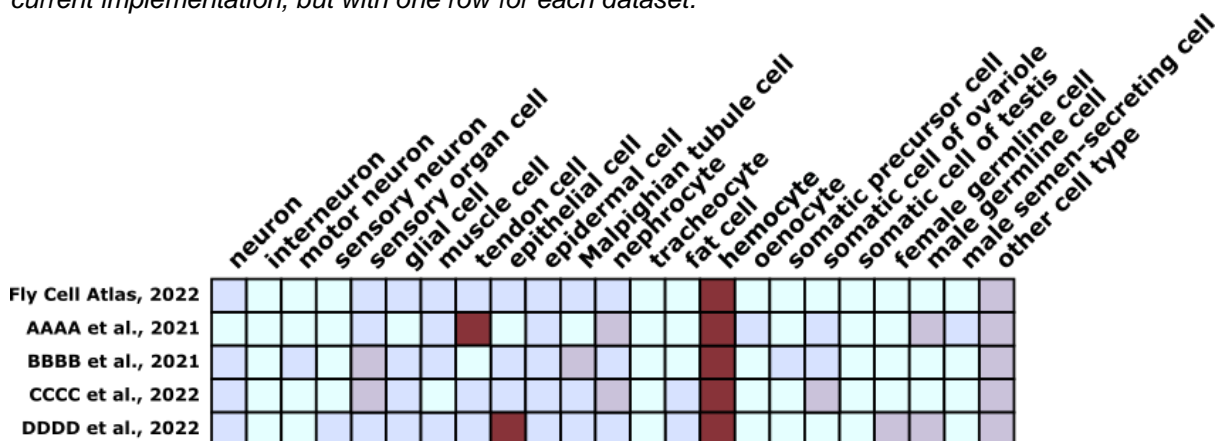


Q4. Would you be satisfied with such a change?

- Yes, knowing at a glance how many datasets found the gene in a cell type would be useful.
- No, I like the way the ribbon currently works, leave it as it is.
- Can't we have both?

Alternatively, we could transform the one-line Cell Types Ribbon into a “stacked ribbon” as in Figure 2c, where each row would present the data from one dataset – with the datasets actually shown being hand picked by FlyBase curators among the most significant dataset. Please note that it is still unclear whether we can technically produce such a stacked ribbon, but we will investigate that possibility if there is enough interest.

Figure 2c. Another proposed implementation for the Cell Types Ribbon. It would basically work as the current implementation, but with one row for each dataset.



Q5. Would you find a stacked Cell Types Ribbon, showing several datasets at once, useful?

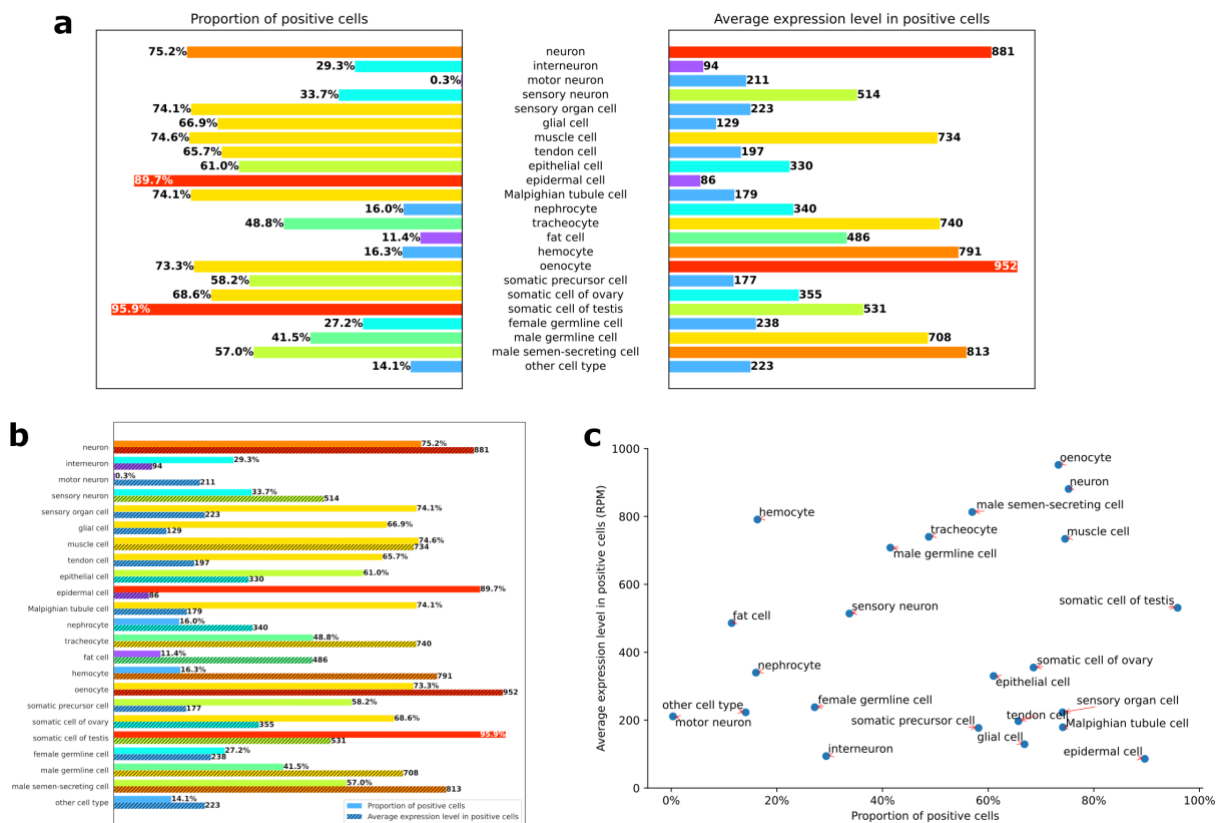
- Yes.
- No.

In addition to the Cell Types Ribbon, we are planning to add a new graph to the High-Throughput Expression Data section of the Gene Report to display the data from the Fly Cell Atlas dataset in a graphical fashion, similarly to what we already do for projects like modENCODE.

The new graph would show the same list of major cell types as in the Cell Types Ribbon, but would display two different values for each cell type: the proportion of cells of that cell type in which the current gene is expressed, and the average level of expression in those cells.

Figure 3 shows the different types of graph we are envisioning for this display.

Figure 3. Proposed types of graph for the FCA data. (a) A dual-sided barplot, with the proportions of positive cells for each cell type on the left, and the average expression levels on the right. (b) A stacked barplot, with solid bars indicating the proportions of positive cells, and hatched bars indicating the average expression levels. (c) A scatterplot. Each dot represents a cell type, with the X coordinate indicating the proportion of positive cells of that cell type and the Y coordinate indicating the average expression level.



Q6. Which of these proposed graphs would you prefer to see?

- I prefer the dual-sided barplot (Figure 3a).
- I prefer the stacked barplot (Figure 3b).
- I prefer the scatterplot (Figure 3c).
- I do not like any of these options.

One possibility we are considering, especially if we opt for a barplot-style graph, is to allow to “expand” the display to also show cell subtypes. For example, if you click on the “hemocytes” label, the graph will be dynamically updated to show more precise data about the different types of hemocytes (lamellocytes, plasmatocytes, crystal cells).

Q7. Would you be interested in such a feature?

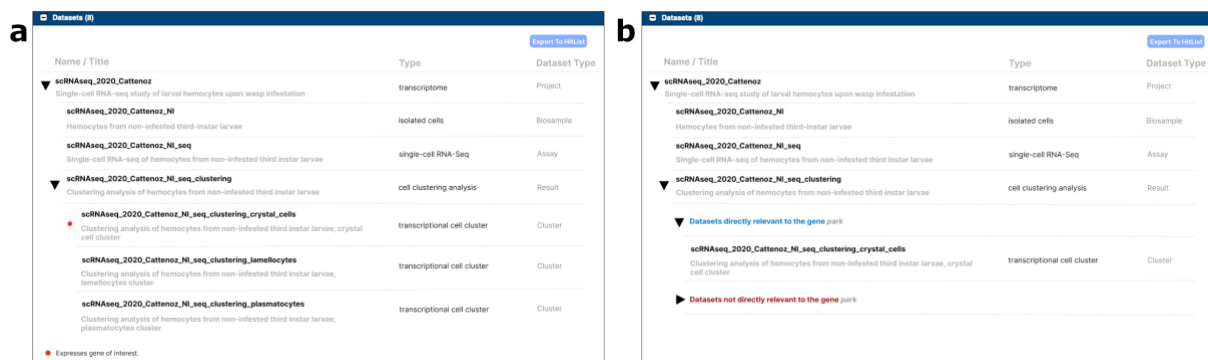
- Yes, that would be great.
- No, showing only the main cell types is enough.

Lastly, a final addition we are considering for the Gene Report is to add a listing of all the scRNAseq datasets that are relevant for this gene – that is, all datasets in which that gene has been found. This could look like what is presented in Figure 4a, which shows that the current gene has been found in the cluster of crystal cells from the non-infested sample of the [Cattenoz et al. 2020](#) dataset.

Alternatively, the list could look like the mockup in Figure 4b, which shows the same thing presented in a slightly different way.

In both cases, items in the listing would be clickable and point towards the corresponding Dataset Report, where more informations about the cluster (such as the number of cells within the cluster and the full list of genes found in that cluster) can be found.

Figure 4. Listing relevant datasets on the Gene Report. (a) A proposed design that lists all samples and clusters within a dataset, with the clusters where the current gene was found being highlighted by a red dot mark. (b) Another design, in which all clusters where the gene was found are grouped together under the label “Datasets directly relevant to the gene”.



Q8. Would such a listing of relevant datasets be useful to you?

- Yes, and I like the mockup in Figure 4a.
- Yes, and I like the mockup in Figure 4b.
- No, I don't think I would need that.

3) The Term Report

As you may know, the Term Report pages describe the terms from our ontologies (controlled vocabularies) that we use to annotate all the data in FlyBase. This includes, in particular, terms from the Drosophila Anatomy Ontology that represent anatomical structures (e.g., “ovary”, “dorsal vessel”, “wing blade”, etc.) and cell types (e.g., “follicle cell”, “cardial cell”, “marginal cell”, etc.).

From the Term Report about a cell type, you can find all the scRNAseq clusters containing cells of that cell type, as shown in Figure 5. Likewise, from the Term Report about an anatomical structure (organ or tissue), you can find all the scRNAseq samples obtained from that anatomical structure.

Figure 5. Getting the list of clusters for a given cell type. (a) From the Term Report for a cell type term (here “adult crystal cell”), expand the “Annotations” section and select the “Libraries/Collections” type of records. (b) This brings a list of the scRNAseq clusters, across all the available scRNAseq datasets, in which cells of that type were identified.

a

| General Information | |
|--|--|
| Term | adult crystal cell |
| Definition | Crystal cell of the adult. FlyBase:FBfr0232226] |
| Comment | |
| Annotations | |
| Records annotated with this term OR any of its CHILD TERMS | |
| Libraries/Collections | 7 |

b

7 selected New Hittist

← Previous 1 Next → Items 1-7 of 7

- scRNAseq_2022_FCA_FEMALE_FULL_seq_clustering_cryst...** Dataset
(FBic0004257) *D. melanogaster*
Class: result Type: transcriptional cell cluster
Title: Clustering analysis of female adult cells from all samples, crystal cells cluster
- scRNAseq_2022_FCA_MIXED_FULL_seq_clustering_crystal...** Dataset
(FBic0005270) *D. melanogaster*
Class: result Type: transcriptional cell cluster
Title: Clustering analysis of adult cells from all samples, crystal cells cluster
- scRNAseq_2022_FCA_MALE_HEART_seq_clustering_crysta...** Dataset
(FBic0004497) *D. melanogaster*
Class: result Type: transcriptional cell cluster
Title: Clustering analysis of cells from 5-days old dorsal vessels, crystal cells cluster

Q9. Did you know you could get a list of scRNAseq clusters and samples from the Term Report?

- Yes, and I have already done it.
- Yes, but I have never done it.
- No, I had never noticed that.

We are considering offering a more direct way to get this list of clusters and samples, on the same page as the Term Report itself, similarly to the list of “relevant datasets” we would like to add on the Gene Report (as described in question 8).

The “relevant datasets” being listed would depend on the term being currently displayed in the report:

- for an anatomical term, this would be either the samples obtained from the organ or tissue that the term represents, or the clusters of the cell type that the term represents;
- for a developmental stage term, this would be all the samples obtained from flies or larvae at that developmental stage;
- for a term representing an experimental method, this would be all the samples, assays, or analyses obtained or performed using that method.

Q10. Would you like a list of relevant datasets directly on the Term Report?

- Yes, that would be useful.
- No, I have no need for that.

Lastly, we are considering enriching the Term Report for cell type terms by adding, when possible, a list of marker genes – genes that can be used to uniquely identify cells of that type. This could be either a list of curated markers (genes that have been experimentally shown to be useful to identify a cell type) or a list of automatically inferred markers (based on scRNAseq data).

Q11. Would a list of marker genes for a cell type be useful?

- Yes, that would be great.
- Yes, but for curated markers only – not for putative markers that have not been verified.
- No, I have no need for that.

4) The Dataset Report

All the curated informations about a scRNAseq dataset ultimately end up on a Dataset Report (such as [this one](#), Figure 6a). This type of report is also used for other kinds of high-throughput experiments, such as bulk RNA-sequencing or mass spectrometry.

Figure 6a. Excerpt of a Dataset Report, showing among other things some of the CV terms used to annotate the dataset and a linkout to the raw data available at the EBI's Single Cell Expression Atlas.

| General Information | | | |
|------------------------|--|---------------|--|
| Name | scRNAseq_2020_Cattenoz | Species | <i>D. melanogaster</i> |
| Project type | transcriptome | FlyBase ID | FBlc0003728 |
| Parent Project | | Data Provider | Giangrande team, IGBMC |
| Title | Single-cell RNA-seq study of larval hemocytes upon wasp infestation | | |
| Accessions | E-MTAB-8698 | | |
| Overview | | | |
| Description | A characterization of the diverse populations of hemocytes in the hemolymph of third instar <i>Drosophila</i> larvae, in control conditions and upon infestation by <i>Leptopilina bouleari</i> . | | (Cattenoz et al., 2020) |
| Project attributes | biotic stimulus study | | (Cattenoz et al., 2020) |
| Biosample type | isolated cells | | (Cattenoz et al., 2020) |
| Assay type | single-cell RNA-Seq | | (Cattenoz et al., 2020) |
| Reagent type | | | |
| Result type | cell clustering analysis | | (Cattenoz et al., 2020) |
| Key genes | | | |
| GO term(s) | defense response to other organism activation of immune response | | (Cattenoz et al., 2020) (Cattenoz et al., 2020) |
| SO term(s) | | | |
| Details | | | |
| Sample preparation | | | |
| Protocol | For each condition, single-cell RNA-seq libraries were prepared from 10,000 cells using the Chromium Single Cell 3' Reagent Kits v2 (10x Genomics). | | (Cattenoz et al., 2020) |
| Mode of Assay | RNA-seq libraries were sequenced on the Illumina HiSeq 4000 platform with paired-end 100 bp reads. | | (Cattenoz et al., 2020) |
| Data analysis | Reads were processed using Cell Ranger v3.0.1 (10x Genomics). They were mapped to the <i>Drosophila melanogaster</i> reference genome BDGP6_ens95. For quality filtering, cells with less than 200 UMI were removed. Further analysis was performed using Seurat v3 (https://satijalab.org/seurat/). Dimensionality reduction and clustering were performed by computing 20 principal components and with a clustering resolution of 0.55. | | (Cattenoz et al., 2020) |
| Comments | | | |
| Files | | | |
| Additional Information | The EMBL-EBI's Single Cell Expression Atlas provides cell-level annotations, clustering data, raw and normalised read counts, and putative marker genes. | | (Cattenoz et al., 2020) |

Q12. Have you already consulted a scRNAseq Dataset Report?

- Yes, and I have found what I wanted.
- Yes, but I couldn't find what I wanted.
- No.

If you couldn't find what you wanted, could you please elaborate?

Currently, there are two ways to land on a Dataset Report. We have already mentioned above how you can get there from the Term Report for a term representing an anatomical structure (where you can get a list of all the dataset samples obtained from that structure) or a cell type (where you can get a list of all the dataset clusters where that cell type was identified).

If you know about the paper where the dataset you're looking for has been described, you can go to the Reference Report for that paper, and look for datasets in the Data From Reference section (Figure 6b).

Figure 6b. Excerpt of a Reference Report, one of the entry points to get to Dataset Reports, by providing a list of all datasets (not limited to scRNAseq) associated with a paper.

| Reference | |
|---------------------|---|
| Citation | Cattenoz, P.B., Sakr, R., Pavlidaki, A., Delaporte, C., Riba, A., Molina, N., Hariharan, N., Mukherjee, T., Giangrande, A. (2020). Temporal specificity and heterogeneity of Drosophila immune cells. EMBO J. 39(12): e104486 . |
| FlyBase ID | FBrf0245988 |
| Publication Type | Research paper |
| Data From Reference | |
| + Alleles (7) | |
| + Genes (4) | |
| - Datasets (1) | |
| Export to HitList | scRNAseq_2020_Cattenoz |

We are considering at least two paths to improve the “discoverability” of the dataset reports.

The first path is the addition of a single page that would list all the available scRNAseq datasets, accessible in a single click. It could be a single flat list, or several lists allowing to browse through the datasets per developmental stages (e.g., adult vs larval datasets), anatomical parts (e.g., gut datasets brain datasets...), or experimental methods (e.g., 10x datasets, Smart-seq2 datasets...).

Q13. Would such a dedicated page for scRNAseq datasets be useful?

- Yes – and a single flat list would be enough.
- Yes – but datasets must be grouped by categories, a flat list would quickly become too long to be of any help.
- Not really.

The second path is to improve our search features to allow finer queries when searching for datasets.

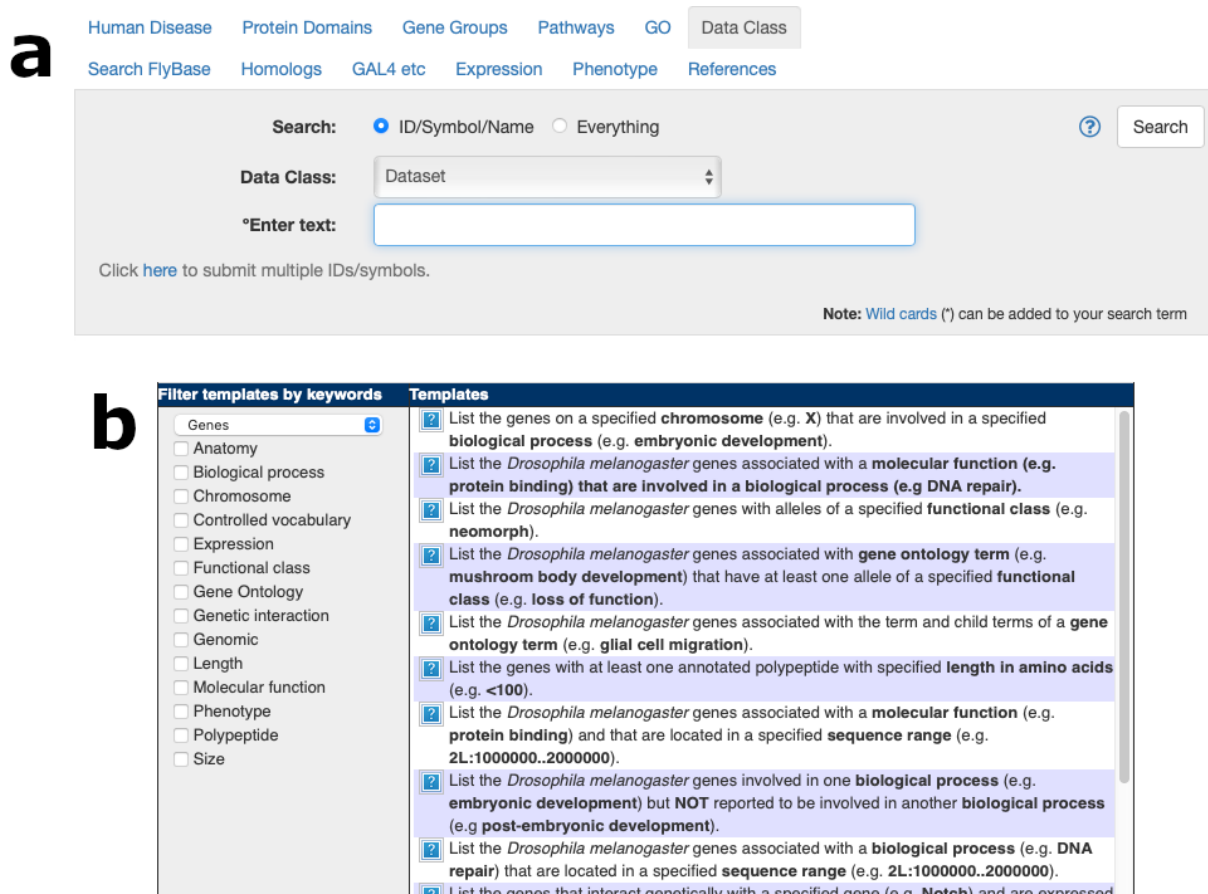
For example, currently datasets can be searched with QuickSearch only by using the generic Data Class tab and setting the data class to Dataset (Figure 7a) – this only allows for either a search through datasets’s ID/Symbol/Name, or a full-text search through datasets’ entire records.

We’re thinking of adding a dedicated “Datasets” tab to the QuickSearch tool, allowing to search through more fields specific to the dataset class – for example, searching only through the projects of a given type (e.g., transcriptomes, proteomes...), or only through the samples of a given nature (e.g., tissues, isolated cells...).

Of note, this would ultimately help finding any kind of datasets we have in FlyBase, not only scRNAseq datasets.

We could also add pre-defined query templates for the QueryBuilder tool (Figure 7b), which would serve the dual purpose of fulfilling some frequent search needs and demonstrating how QueryBuilder can be used to create complex queries.

Figure 7. Better search features for datasets. (a) Searching for datasets using the QuickSearch's generic Data Class search. (b) Pre-defined templates for the QueryBuilder. Currently, there is no such templates for datasets.



Q14. Would a dataset-specific search tab be helpful?

- Yes. I can't believe you don't have one already.
- Not really.

Q15. Would you use pre-defined, dataset-specific query templates?

- Yes, such queries are missing to the QueryBuilder tool.
- Probably not, I have never really used QueryBuilder anyway.
- No, though I do regularly use QueryBuilder.

5) Expression data from a scRNAseq dataset

Raw expression data from scRNAseq datasets are not kept at FlyBase. Instead, we rely on the EBI's Single Cell Expression Atlas to store the raw data. The Dataset Report includes linkouts to the SCEA, listed under the field "Accessions". You can use that linkout to land directly on the SCEA page for a given dataset and from there, either explore the dataset online or download the raw data files for subsequent analysis on your machine.

Q16. Have you already followed a linkout to a SCEA dataset?

- Yes I have.
- No, I never really had a need for it.
- No, I didn't know about such linkouts.

What we do keep at FlyBase, beyond datasets metadata and annotations, are what we call the Summarised Expression Data. Basically, it's a condensed version of the original expression data, summarised at the cell type level in such a way that, for each gene and each cell type in a given dataset, we have two numerical values: the extent of expression, which is the proportion of cells of that cell type in which that gene is expressed (at any level), and the average expression level of that gene across all cells of that type that do express it.

We use the Summarised Expression Data to power the Cell Types Ribbon, and we will also use them to power the new graphical display of the FCA data we have mentioned earlier in this survey.

However, at the moment, the Summarised Expression Data is purely for internal use within FlyBase, and are neither displayed on the website nor available for download.

Q17. Would you like the ability to download our Summarised Expression Data tables?

- Yes, exploring this kind of data would be useful.
- No need, if I ever have a need for it I can get the original data from the SCEA.

6) Priorities and final thoughts

Q18. Of all the features we have outlined in this survey, which one would you like to see the most?

- Revisited Cell Types Ribbon, showing either the number of datasets for each cell type, or several dataset in a "stacked ribbon" mode.

- New type of graph to show the FCA data in the “High-Throughput Expression Data” section.
- Listing of relevant datasets on the Gene Report.
- Listing of relevant datasets on the Term Report.
- Showing putative marker genes on the Term Report (for terms representing cell types).
- A dedicated page to explore the available scRNAseq datasets.
- Better search features for datasets (not only scRNAseq).
- The possibility to download the summarised expression data.

Q19. Is there anything you would like to add? An idea about a feature we didn't think about?

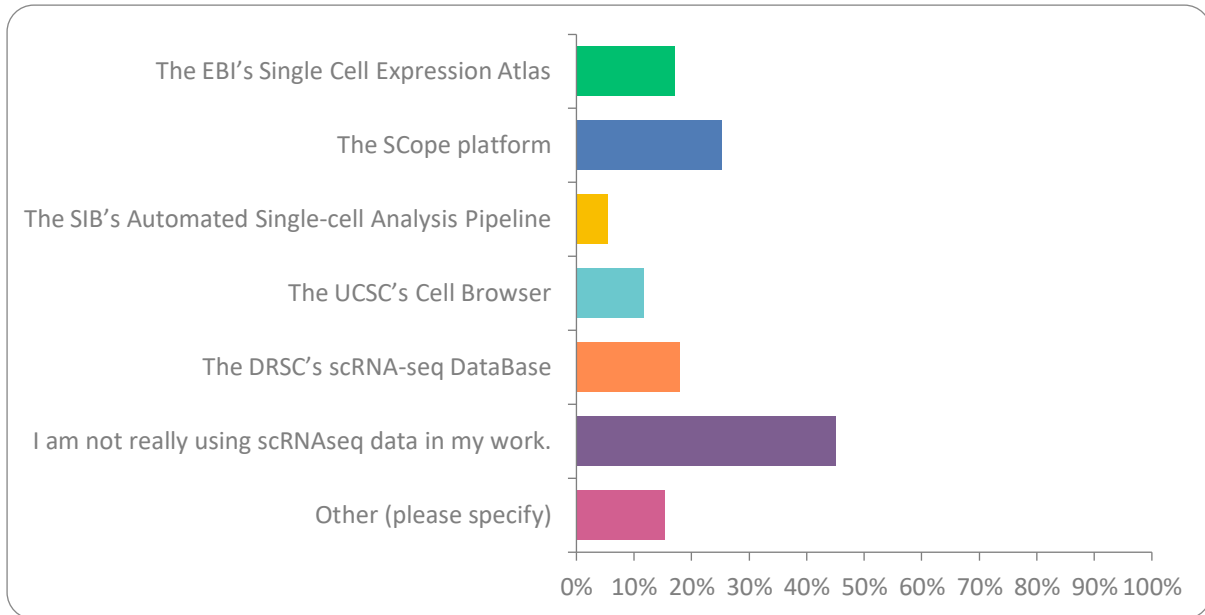


FlyBase Features for scRNAseq Data Survey

Responses

1. What do you currently use to explore scRNAseq data? (Check all that apply)

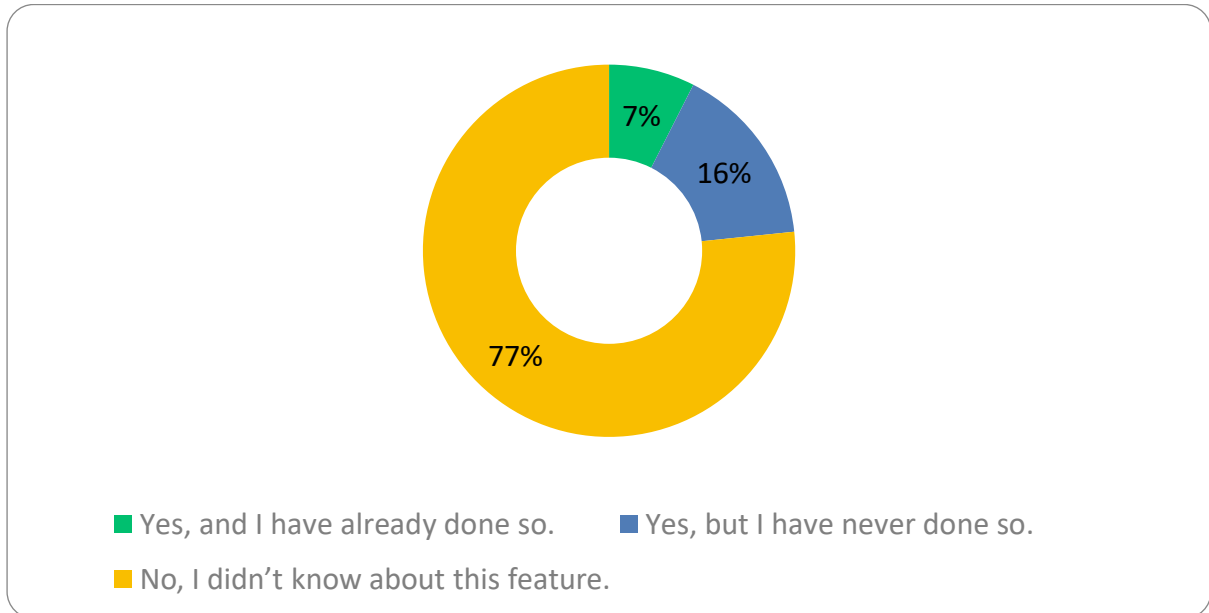
Answered: 111 Skipped: 0



| ANSWER CHOICES | RESPONSES | |
|---|-----------|------------|
| The EBI's Single Cell Expression Atlas | 17.12% | 19 |
| The SCoPe platform | 25.23% | 28 |
| The SIB's Automated Single-cell Analysis Pipeline | 5.41% | 6 |
| The UCSC's Cell Browser | 11.71% | 13 |
| The DRSC's scRNA-seq DataBase | 18.02% | 20 |
| I am not really using scRNAseq data in my work. | 45.05% | 50 |
| Other (please specify) | 15.32% | 17 |
| TOTAL | | 153 |

2. Did you know you could search for a gene on the SCEA or the DRscDB directly from the Gene Report?

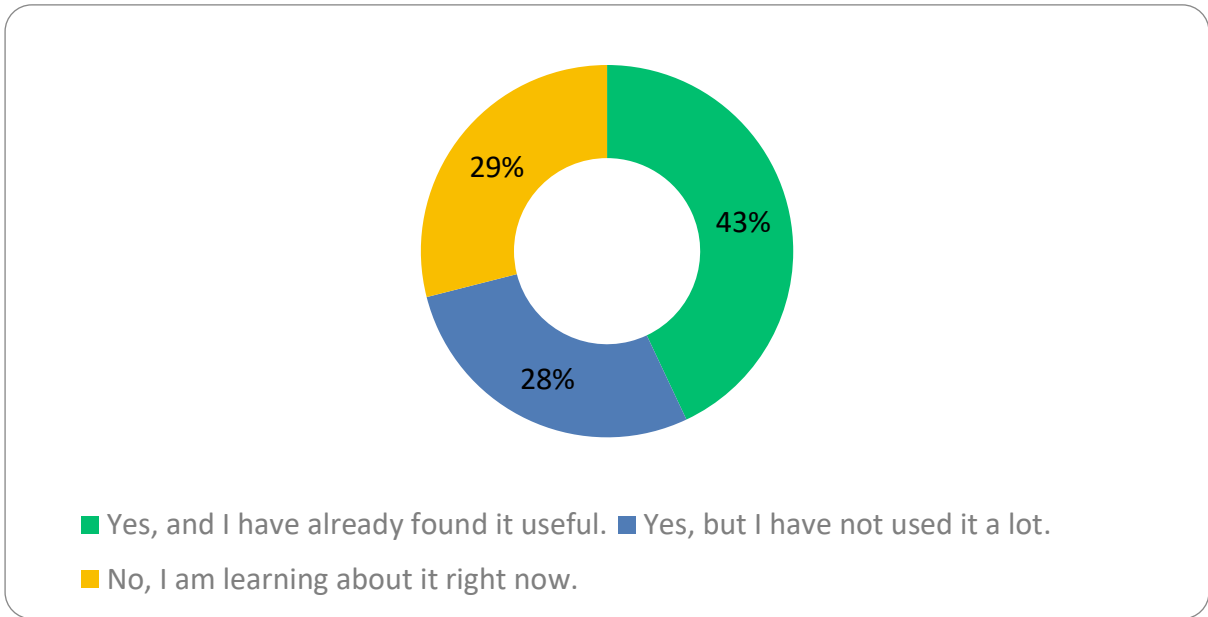
Answered: 107 Skipped: 4



| ANSWER CHOICES | RESPONSES |
|---------------------------------------|-----------|
| Yes, and I have already done so. | 7.48% 8 |
| Yes, but I have never done so. | 15.89% 17 |
| No, I didn't know about this feature. | 76.64% 82 |
| TOTAL | 107 |

3. Had you noticed the Cell Types Ribbon since its introduction earlier this year?

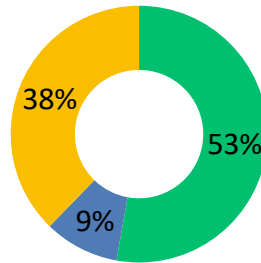
Answered: 107 Skipped: 4



| ANSWER CHOICES | RESPONSES | |
|--|-----------|-----|
| Yes, and I have already found it useful. | 42.99% | 46 |
| Yes, but I have not used it a lot. | 28.04% | 30 |
| No, I am learning about it right now. | 28.97% | 31 |
| TOTAL | | 107 |

4. Would you be satisfied with such a change?

Answered: 106 Skipped: 5

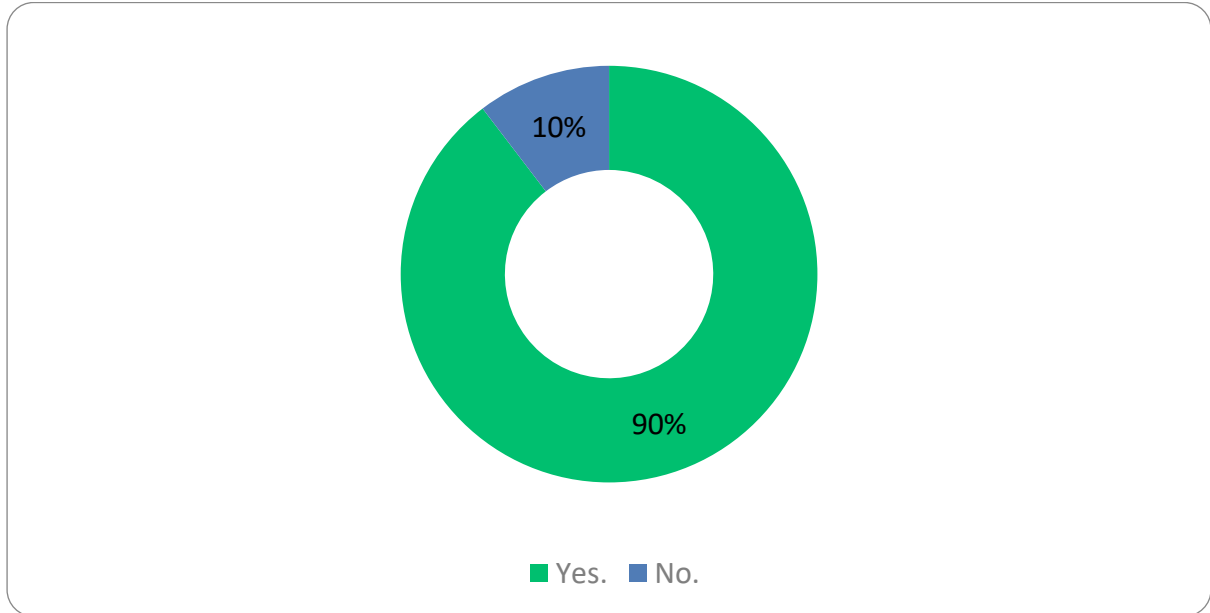


- Yes, knowing at a glance how many datasets found the gene in a cell type would be useful.
- No, I like the way the ribbon currently works, leave it as it is.
- Can't we have both?

| ANSWER CHOICES | RESPONSES | |
|---|-----------|-----|
| Yes, knowing at a glance how many datasets found the gene in a cell type would be useful. | 52.83% | 56 |
| No, I like the way the ribbon currently works, leave it as it is. | 9.43% | 10 |
| Can't we have both? | 37.74% | 40 |
| TOTAL | | 106 |

5. Would you find a stacked Cell Types Ribbon, showing several datasets at once, useful?

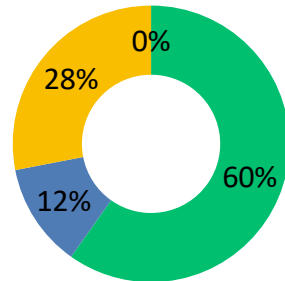
Answered: 106 Skipped: 5



| ANSWER CHOICES | RESPONSES | |
|----------------|-----------|-----|
| Yes. | 89.62% | 95 |
| No. | 10.38% | 11 |
| TOTAL | | 106 |

6. Which of these proposed graphs would you prefer to see?

Answered: 107 Skipped: 4

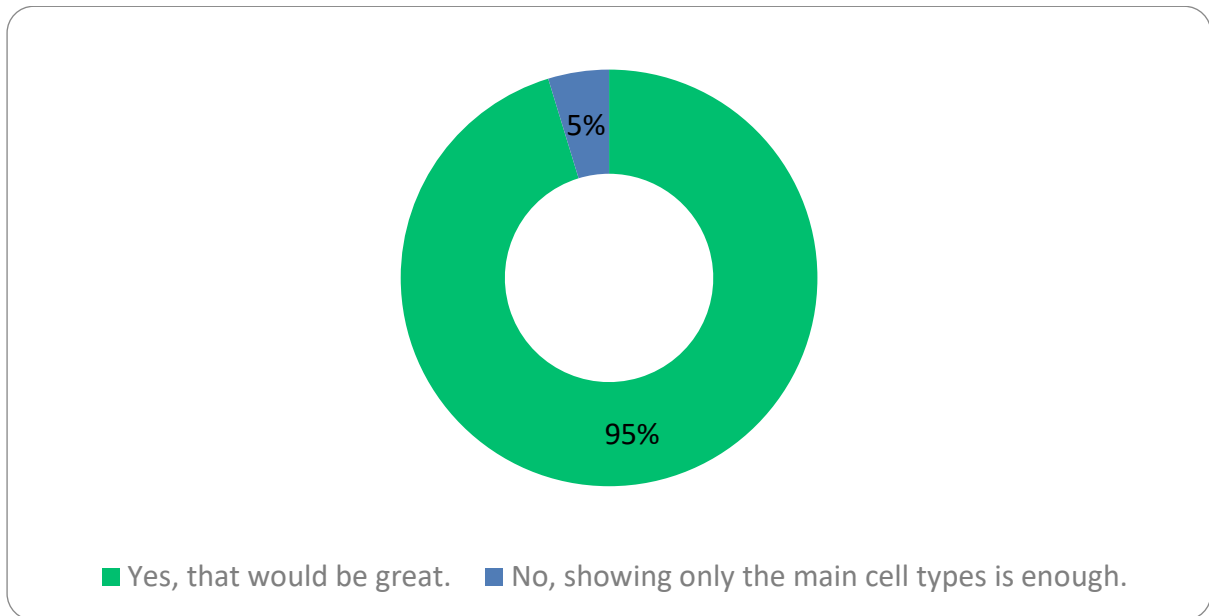


- I prefer the dual-sided barplot (Figure 3a).
- I prefer the stacked barplot (Figure 3b).
- I prefer the scatterplot (Figure 3c).
- I do not like any of these options.

| ANSWER CHOICES | RESPONSES | |
|--|-----------|-----|
| I prefer the dual-sided barplot (Figure 3a). | 59.81% | 64 |
| I prefer the stacked barplot (Figure 3b). | 12.15% | 13 |
| I prefer the scatterplot (Figure 3c). | 28.04% | 30 |
| I do not like any of these options. | 0% | 0 |
| TOTAL | | 107 |

7. Would you be interested in such a feature?

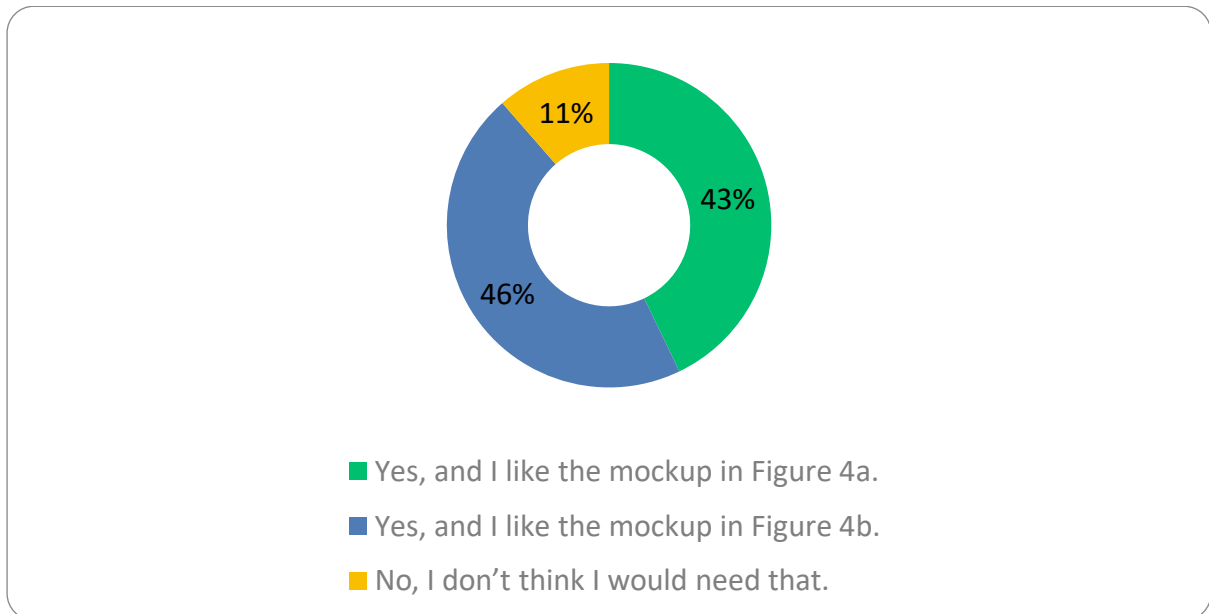
Answered: 106 Skipped: 5



| ANSWER CHOICES | RESPONSES | |
|---|-----------|-----|
| Yes, that would be great. | 95.28% | 101 |
| No, showing only the main cell types is enough. | 4.72% | 5 |
| TOTAL | | 106 |

8. Would such a listing of relevant datasets be useful to you?

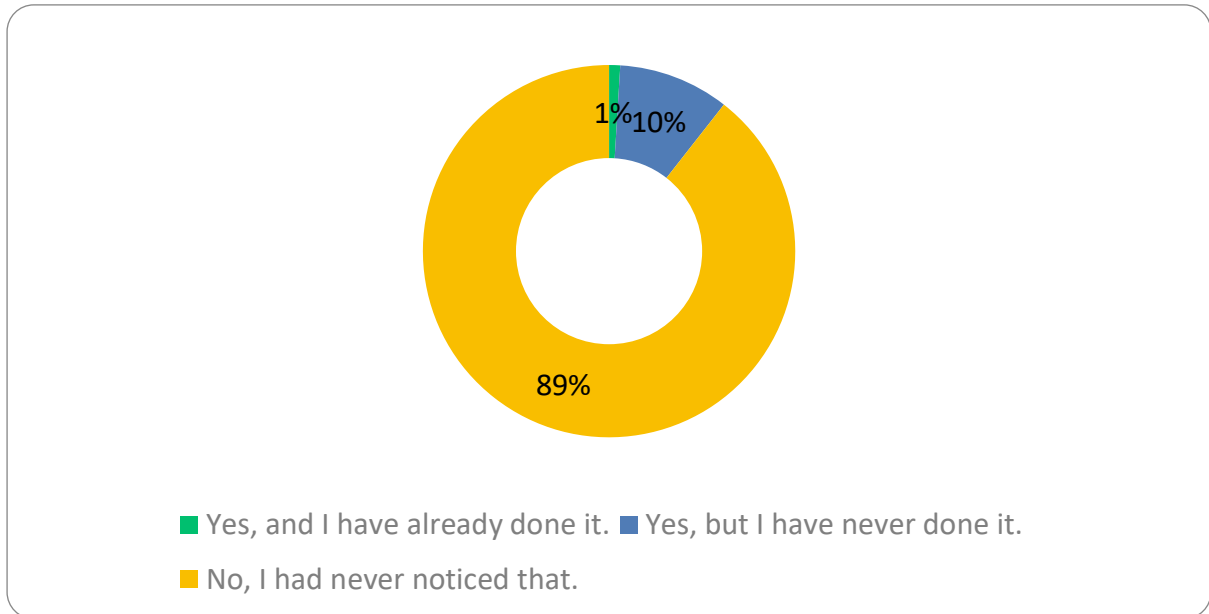
Answered: 105 Skipped: 6



| ANSWER CHOICES | RESPONSES | |
|--|-----------|-----|
| Yes, and I like the mockup in Figure 4a. | 42.86% | 45 |
| Yes, and I like the mockup in Figure 4b. | 45.71% | 48 |
| No, I don't think I would need that. | 11.43% | 12 |
| TOTAL | | 105 |

9. Did you know you could get a list of scRNAseq clusters and samples from the Term Report?

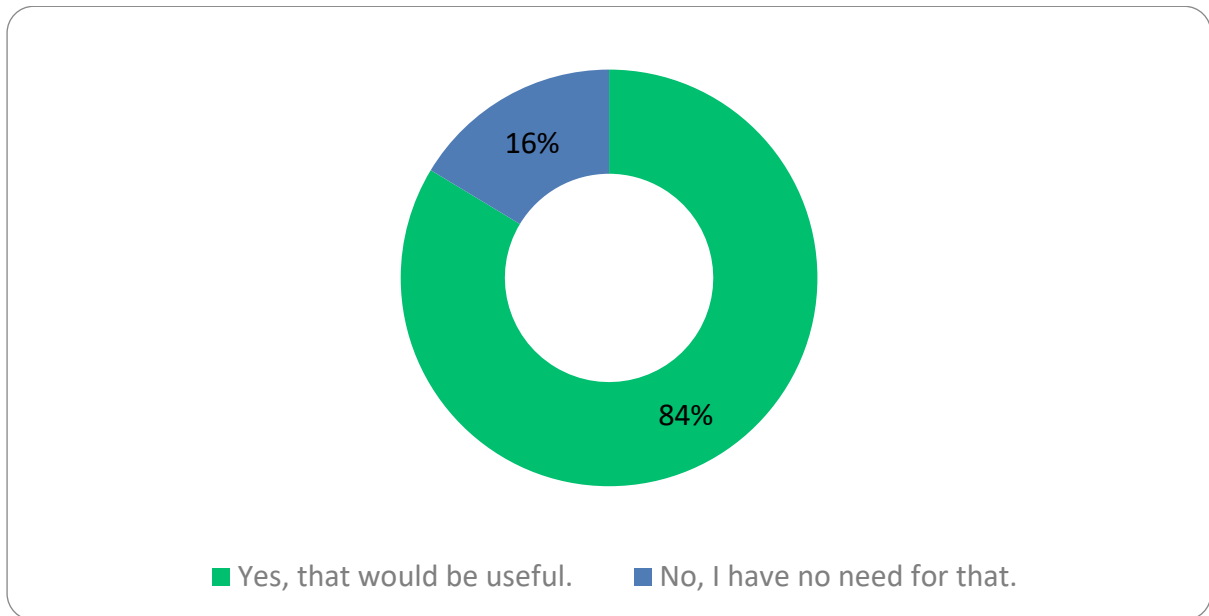
Answered: 104 Skipped: 7



| ANSWER CHOICES | RESPONSES |
|----------------------------------|-----------|
| Yes, and I have already done it. | 0.96% 1 |
| Yes, but I have never done it. | 9.62% 10 |
| No, I had never noticed that. | 89.42% 93 |
| TOTAL | 104 |

10. Would you like a list of relevant datasets directly on the Term Report?

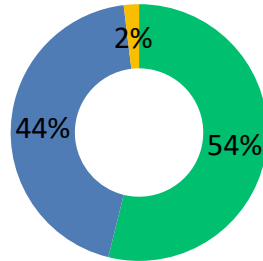
Answered: 104 Skipped: 7



| ANSWER CHOICES | RESPONSES | |
|------------------------------|-----------|-----|
| Yes, that would be useful. | 83.65% | 87 |
| No, I have no need for that. | 16.35% | 17 |
| TOTAL | | 104 |

11. Would a list of marker genes for a cell type be useful?

Answered: 104 Skipped: 7

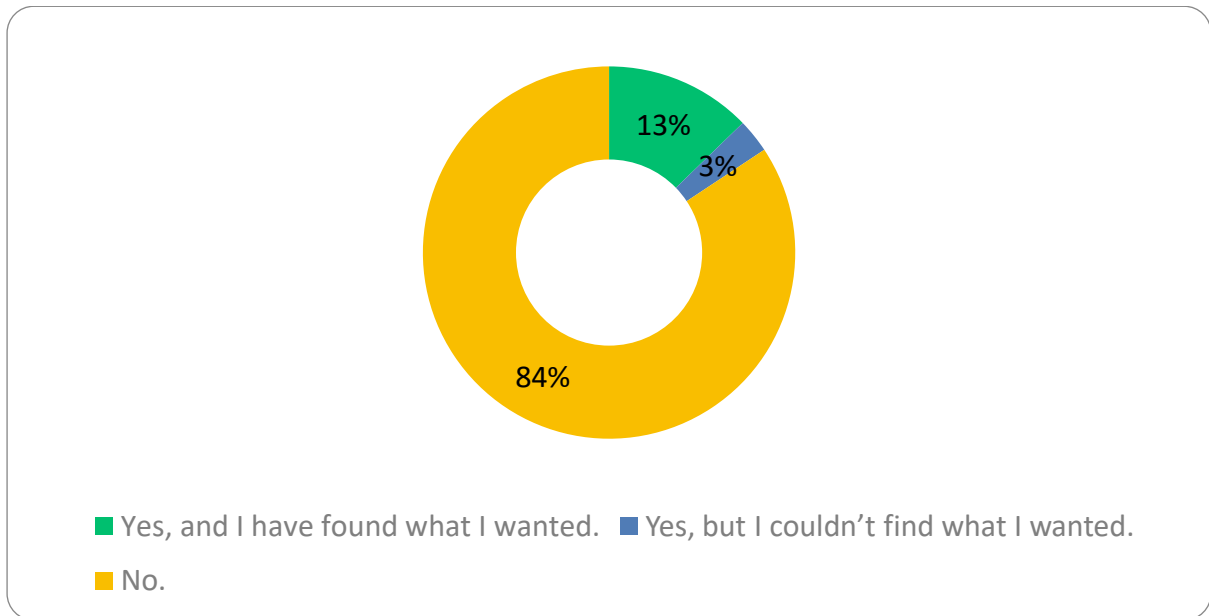


- Yes, that would be great.
- Yes, but for curated markers only – not for putative markers that have not been verified.
- No, I have no need for that.

| ANSWER CHOICES | RESPONSES | |
|---|-----------|------------|
| Yes, that would be great. | 53.85% | 56 |
| Yes, but for curated markers only – not for putative markers that have not been verified. | 44.23% | 46 |
| No, I have no need for that. | 1.92% | 2 |
| TOTAL | | 104 |

12. Have you already consulted a scRNAseq Dataset Report?

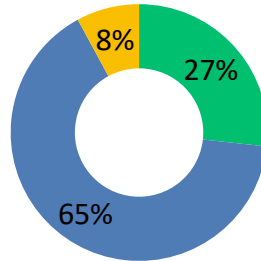
Answered: 102 Skipped: 9



| ANSWER CHOICES | RESPONSES |
|---|-----------|
| Yes, and I have found what I wanted. | 12.75% 13 |
| Yes, but I couldn't find what I wanted. | 2.94% 3 |
| No. | 84.31% 86 |
| TOTAL | 102 |

13. Would such a dedicated page for scRNAseq datasets be useful?

Answered: 101 Skipped: 10

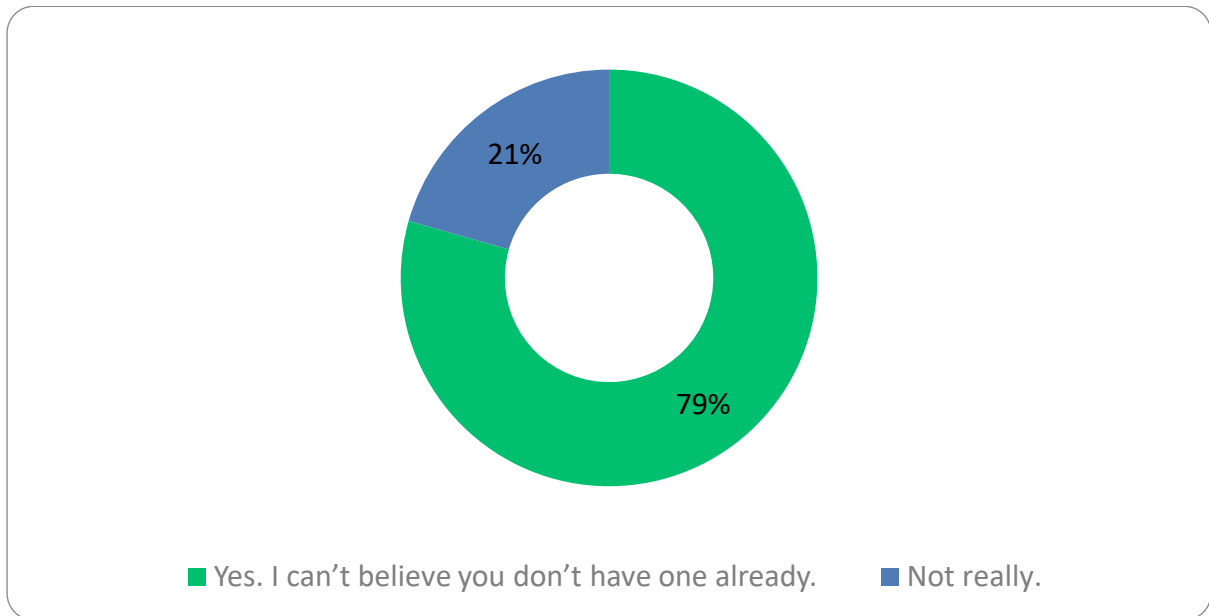


- Yes – and a single flat list would be enough.
- Yes – but datasets must be grouped by categories, a flat list would quickly become too long to be of any help.
- Not really.

| ANSWER CHOICES | RESPONSES | |
|--|-----------|-----|
| Yes – and a single flat list would be enough. | 26.73% | 27 |
| Yes – but datasets must be grouped by categories, a flat list would quickly become too long to be of any help. | 65.35% | 66 |
| Not really. | 7.92% | 8 |
| TOTAL | | 101 |

14. Would a dataset-specific search tab be helpful?

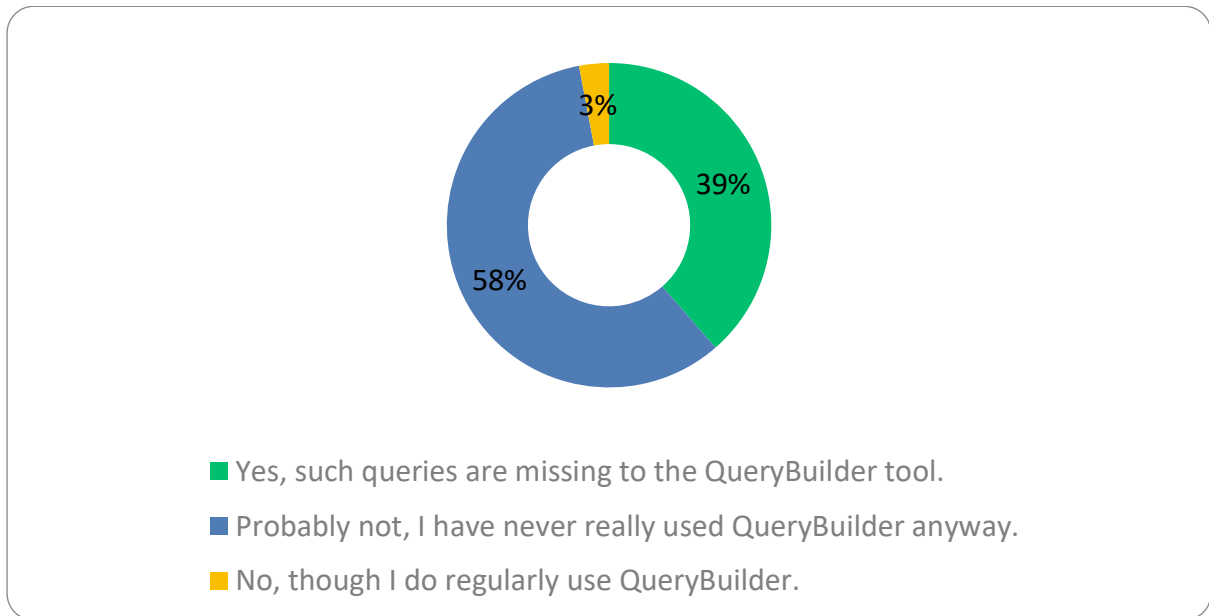
Answered: 102 Skipped: 9



| ANSWER CHOICES | RESPONSES | |
|--|-----------|-----|
| Yes. I can't believe you don't have one already. | 79.41% | 81 |
| Not really. | 20.59% | 21 |
| TOTAL | | 102 |

15. Would you use pre-defined, dataset-specific query templates?

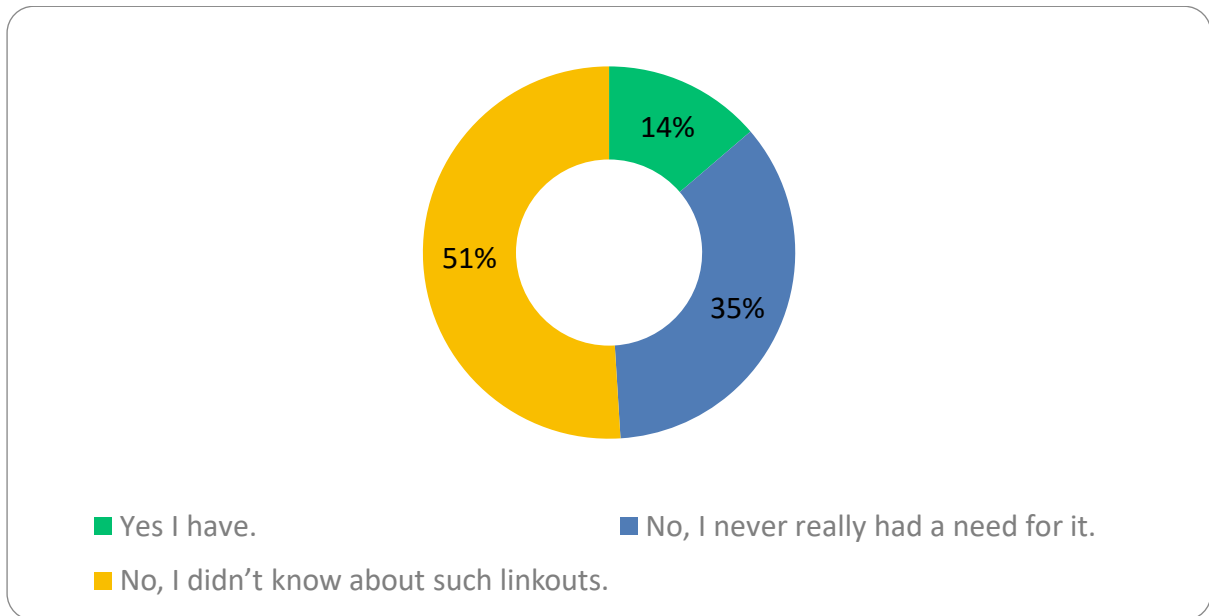
Answered: 101 Skipped: 10



| ANSWER CHOICES | RESPONSES | |
|---|-----------|-----|
| Yes, such queries are missing to the QueryBuilder tool. | 38.61% | 39 |
| Probably not, I have never really used QueryBuilder anyway. | 58.42% | 59 |
| No, though I do regularly use QueryBuilder. | 2.97% | 3 |
| TOTAL | | 101 |

16. Have you already followed a linkout to a SCEA dataset?

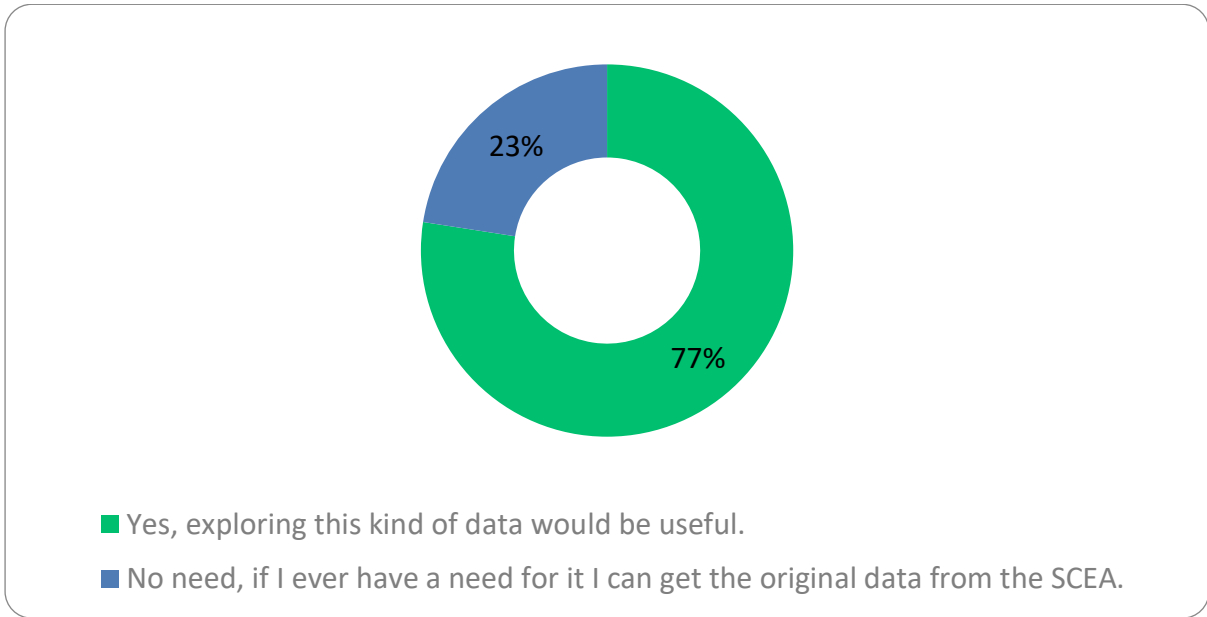
Answered: 102 Skipped: 9



| ANSWER CHOICES | RESPONSES | |
|--|-----------|-----|
| Yes I have. | 13.73% | 14 |
| No, I never really had a need for it. | 35.29% | 36 |
| No, I didn't know about such linkouts. | 50.98% | 52 |
| TOTAL | | 102 |

17. Would you like the ability to download our Summarised Expression Data tables?

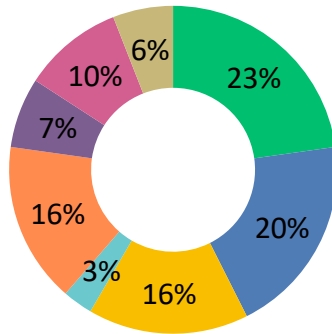
Answered: 102 Skipped: 9



| ANSWER CHOICES | RESPONSES | |
|--|-----------|-----|
| Yes, exploring this kind of data would be useful. | 77.45% | 79 |
| No need, if I ever have a need for it I can get the original data from the SCEA. | 22.55% | 23 |
| TOTAL | | 102 |

18. Of all the features we have outlined in this survey, which one would you like to see the most?

Answered: 101 Skipped: 10



- Revisited Cell Types Ribbon, showing either the number of datasets for each cell type, or several dataset in a “stacked ribbon“ mode.
- New type of graph to show the FCA data in the “High-Throughput Expression Data” section.
- Listing of relevant datasets on the Gene Report.
- Listing of relevant datasets on the Term Report.
- Showing putative marker genes on the Term Report (for terms representing cell types).
- A dedicated page to explore the available scRNAseq datasets.

| ANSWER CHOICES | RESPONSES | |
|---|-----------|----|
| Revisited Cell Types Ribbon, showing either the number of datasets for each cell type, or several dataset in a “stacked ribbon“ mode. | 22.77% | 23 |
| New type of graph to show the FCA data in the “High-Throughput Expression Data” section. | 19.80% | 20 |
| Listing of relevant datasets on the Gene Report. | 15.84% | 16 |
| Listing of relevant datasets on the Term Report. | 2.97% | 3 |
| Showing putative marker genes on the Term Report (for terms representing cell types). | 15.84% | 16 |

| | | |
|--|-------|-----|
| A dedicated page to explore the available scRNAseq datasets. | 6.93% | 7 |
| Better search features for datasets (not only scRNAseq). | 9.90% | 10 |
| The possibility to download the summarised expression data. | 5.94% | 6 |
| TOTAL | | 101 |