Searching vs Browsing/Filtering Data at ZFIN

People search the ZFIN database every day, most frequently by gene symbol. From there, the vast majority of searchers go to the gene page for the gene they searched for. From the gene page, all the information in the ZFIN database about that gene is available. This is a very direct and powerful way to get to the gene record you want to see in ZFIN. However, this is also a narrow view of the potentially insightful related data that may surround that gene record. Often, a data browsing/filtering approach can be more powerful for open-ended or complex queries. By making that cloud of related data more visible, browsing and filtering may also reveal relationships in data that may not have been anticipated. ZFIN supports data browsing/filtering using the “single box search” found on the ZFIN home page and at the top right of all other ZFIN web pages (Figure 1).

Let's look at some examples.

- Which genes are most commonly found in publications discussing **dlx2a**?
- Which genes are associated with a retinal phenotype?
- What other phenotypes are associated with genes producing retinal phenotypes?
- Does the Zebrafish International Resource Center (ZIRC) have any point mutations available for sale which create a premature stop resulting in phenotypes including decreased size of the eye and inner ear?

1. **Which genes are most commonly found in publications discussing **dlx2a**?**
   a. From the ZFIN home page, search for "dlx2a" in the single box search.
   b. Select Publication from the list of categories at the left side of the search results page to get the publication search results.
   c. The left side of the Publication search results page has a collection of filters which can be used to narrow the result set, including "Gene", "Mutation/Tg", etc. Each time one of the filter values is clicked, the publication result set is limited to only publications which have that attribute. For example, if the dlx2a gene in the "Gene" filter is clicked, all the publications will then be required to include the dlx2a gene. Today, 384 publications meeting that criterion (Figure 2).
   d. Here is where the magic is. Not only do the filters on the left side allow you to reduce the publication result set further based on specific attributes, they also contain information of their own. For example, you can immediately see at the top center of the search results that there are 384 publications that include **dlx2a**. The filters at the left show that of the 384 dlx2a publications, 107 also include **shha**, 96 include **foxd3**, 96 include **egr2b**, and 96 include **sox10**. Thus, the answer to our question "Which genes are most commonly found in publications discussing **dlx2a**?" is **shha, foxd3, egr2b, and sox10**.
2. Which genes are associated with a retinal phenotype?

a. Click in the single box search text entry box on the ZFIN home page or top right of all other pages in ZFIN and hit Enter or click the "Go" button. You now have access to all the data in ZFIN in the single box search interface.
b. Click the "Gene/Transcript" category in the left column. Now, only the genes and transcripts in ZFIN are in the search results and the "Gene/Transcript" filters are available in the left side of the page. We want to know which genes are associated with a retina phenotype, so we are looking for genes, not transcripts.
c. In the "Type" filter, select "Gene". This will limit the result set to just records of type "Gene". Click the "Type" filter label to collapse that filter since it is now set up as desired.
d. In the collection of "Phenotype" filters, locate the "Affected Anatomy" filter. This lists any anatomical structure that has been annotated as being affected in a phenotype associated with any gene in the result set. We are looking for genes associated with a retina phenotype.
e. Click "Show All" under "Affected Anatomy" to access the full list of affected anatomical structures. Filter that list to find "retina".
f. Select "retina" from the Affected Anatomy filter. Today, that limits the full set of genes to 520 genes which have a retina phenotype (Figure 3).
g. That list of 520 genes answers our question! These 520 genes are associated with a phenotype affecting the retina.
h. These can then be further filtered if desired using the filters at the left, or they can then be downloaded by clicking on the "Download" button at the top of the search results.

Figure 3. Results of filtering to locate genes having a retinal phenotype.

3. What other phenotypes are associated with genes producing retinal phenotypes?
   i. Starting from the list of 520 genes with a retina phenotype identified in the previous example, what other phenotypes are associated with those genes?
   ii. Clicking the "Phenotype Statement" filter in the "Phenotypes" filter collection in the left column will show the most common phenotypes associated with this set of 520 genes having a retina phenotype, and how many genes in the result set are associated with that additional phenotype.
   iii. Today the filter shows that of the 520 "retina phenotype genes" 262 are associated with decreased eye size, 106 are associated with decreased head size, and 102 are associated with an edematous pericardium (Figure 4).
Figure 4. Genes associated with a retinal phenotype are also associated with reduced eye size, decreased head size, and an edematous pericardium as seen in the "Phenotype Statement" filter.

iv. Clicking on "Show All" at the bottom of the "Phenotype Statement" filter provides access to all the phenotypes associated with this set of genes and the ability to search over those to explore the data further.

v. For example, are any of these 520 genes which have a retina phenotype also associated with the absence of Meckel's cartilage?

1. Click "Show All" under "Phenotype Statement"
2. In the "Filter" box of the popup type "Meckel"
3. The phenotype statement set is filtered to only those involving Meckel's cartilage.
4. Today, five genes are associated with absence of Meckel's cartilage in addition to having the retina phenotype we originally filtered for (Figure 5).
5. Clicking on “Meckel's cartilage absent, abnormal” will filter the gene results down to just those five genes associated with both a retina phenotype and an absence of Meckel's cartilage (aldh1a2, chd7, alx1, med14, and nup107) (Figure 6).

![Figure 5. Filtering the "Phenotype Statement" values for phenotypes involving Meckel's cartilage](image)

Figure 5. Filtering the "Phenotype Statement" values for phenotypes involving Meckel's cartilage.

![Figure 6. Five genes are associated with both a retinal phenotype and the absence of Meckel's cartilage.](image)

Figure 6. Five genes are associated with both a retinal phenotype and the absence of Meckel's cartilage.

4. Does the Zebrafish International Resource Center (ZIRC) have any point mutations available for sale which create a premature stop resulting in phenotypes including decreased size of the eye and inner ear?

   a. Go to the [ZFIN home page](http://www.stanford.edu) and click in the text entry area for single box search and hit enter or click the "Go" button to access all data in the search interface.
   b. We are looking for mutants, so in the list of Categories at the left, select “Mutation/Tg”.
   c. The result set now includes all the mutation and Tg records in ZFIN (currently 68,824). Now, filter that long list down to find just those mutants that meet our criteria.
   d. First, the mutant needs to be available from ZIRC, so click to open the “Source” filter, and select “Zebrafish International Resource Center (ZIRC)”. Currently that filters the set down to 43,865 mutants/transgenics.
   e. Next, the desired mutant must be a point mutation. In the “Type” filter click “Point Mutation”. That limits the set further to only include point mutations (31,566 currently)
   f. Then, the desired mutant should result in a premature stop. Click to open the "Consequence" filter and click "premature stop". That filters the set further to only include those features which have a premature stop as their consequence (20,181 currently).
g. Finally, the desired mutant should have decreased eye and inner ear size. Click to open the "Phenotype Statement" filter and click the "eye decreased size, abnormal". Today, that limits the resulting list of mutants to 25 candidates.

h. Click on "inner ear decreased size, abnormal" in the "Phenotype Statement" filter. Today that filters the set of mutants down to just five mutants (j410, sa964, sa1349, sa1376, and sa3219) (Figure 7).

i. Clicking "Show All" in the "Affected Genomic Region" filter shows that these mutants are associated with the genes car, imx1bb, mdr1, porfa, and tfip11. The "Phenotype Statement" Filter shows that two of these mutants are associated with an abnormal gut phenotype, and three are associated with an edematous pericardium phenotype.

Figure 7. Filtering to locate mutants which are available from ZIRC and are point mutations creating a premature stop and resulting in a phenotype including decreased eye and inner ear size.

Conclusion:
The above examples only scratch the surface of the functionality provided in the single box search filtering interface. Filtering data in the single box search limits the result set to only records which have the attributes you select from the filters. It is also possible to exclude results having one of the filter values by clicking on the small minus sign that appears beside each filter value when your mouse passes over it. The small plus sign that appears has the same effect as clicking on the term itself. This filtering approach shows you where the data are and how many records are available. The only values offered in the filters are values which will produce at least one record in the results. This eliminates blindly searching and getting no search results. Filter values are also marked with the number of search results that will be retained if you select that filter value, providing insight into the composition of your search result set in a way that direct searching cannot. Single box search currently supports searching and filtering for genes, transcripts, expression, phenotype, human disease, Fish, reporter lines, mutation/Tgs, constructs, morpholinos, TALENS, CRISPRs, antibodies, markers/clones, published figures, anatomy and GO terms, publications, and people, labs, and companies...or... you can search across all these categories at once. Come up with your own questions and see if single box search can find the answers! As always, we welcome your feedback on single box searching and filtering, or any other aspect of ZFIN, at zfinadmin [at] zfin.org.