Module 6: Exploring zebrafish research at ZFIN and the ZebrafishMine

Aims

- Introduce the ZebrafishMine data retrieval system
- Work through some ZebrafishMine examples
- Invite workshop attendees to use ZFIN’s Faceted Search prototype and provide feedback

Introduction

ZFIN (http://zfin.org/), the zebrafish model organism database, is a resource of genetic, genomic and developmental research information on the zebrafish. The information available at ZFIN includes curated and submitted data on zebrafish genes, mutants, genotypes, expression, phenotype, orthology, anatomy, publications, and reagents. This module will introduce two new ways of accessing ZFIN’s data: ZebrafishMine and Faceted Search.

ZebrafishMine (http://zebrafishmine.org) is based on the InterMine data warehousing system (http://intermine.github.io/intermine.org/). InterMine-based model organism mines in the InterMOD consortium include FlyMine, MouseMine, RatMine, WormMine, YeastMine and ZebrafishMine. ZebrafishMine allows ZFIN to offer customizable search and download options, and web services. Currently ZebrafishMine includes ZFIN data and data from the Panther homology database (http://pantherdb.org/). ZFIN data in ZebrafishMine is updated weekly. Additional data types, such as Pathways and more sequence data, will continue to be added to ZebrafishMine.
Introduction to the ZebrafishMine home page

Go to http://zebrafishmine.org

Searching options in ZebrafishMine:

- **Keyword Search**
  - Familiar simple search that can be used to find names, symbols, identifiers.

- **Template Searches**
  - Pre-defined searches.

- **Query Builder**
  - Alter pre-defined searches or build searches of your own.
**Templates** are pre-defined searches. A subset of templates can be found in the middle section of the ZebrafishMine home page, divided into categories (see previous page). You can get a complete list of templates by clicking the "Templates" tab.

**Do a template search:**

1. Click on the “Templates” tab at the top of the home page

The template tab:

2. Scroll down the list to find the template “**Gene → OMIM Disease Phenotype**”.
Note: the position of the template will likely differ from the order shown here. Templates are ordered by popularity.

3. Click on the template to see the details.

   This template is a search for the OMIM disease phenotypes of the curated human orthologs of the specified zebrafish gene. The default value in the search box is the “ref” gene. You can run the search for all zebrafish genes by using a wildcard (*). You can also run the search using a list of genes by selecting the “constrain to be IN” checkbox and selecting a list name from the pulldown menu.

4. Change the default value to a wildcard (*). Click the “Show Results” button to run this search.
The results table:

You can use the column header icons to further organize and filter your results:

5. Filter your results:

In the Results table, locate the “Symbol” column and click on the “Column summary” icon. The Column summary popup window lists all genes by order of number of associated OMIM Phenotype Disease annotations. Select the checkboxes for all genes that have 12 or more annotations and click “Filter”. Your filtered results will now be limited to genes with 12 or more associated OMIM phenotypes.

Note: the number of results may differ slightly when you run the template, since data is updated weekly.
Filtering results in the results table:

You can download your results by clicking the “Download” button in the top right hand corner of the results page.

A new popup window will appear where you can select the download format, specify which columns and rows you want to download, specify if you want to compress your download file, and select a destination for your files.

6. Download your results:
Click on the “Download” button. In the popup window, select the “Spreadsheet (comma separated values)” option. Click “Download File”.
Create lists from your results
You can create lists from your results by clicking on the “Lists” button in the top right hand side of the results page. Lists contain information about one type of object; a gene list might include the symbol, full name, and identifier of the genes. You can use this list to run another template search.

7. Create a list
Click on the “Create New List” button on the top right. Click “All 8 Genes”. In the “List Details” popup that comes up, name your list. Click “Create”.

Name your list
Run a template search using a list

In this example, we will use one of the lists of genes already in ZebrafishMine.

1. Go to the full list of templates by clicking on the “Templates” tab on the home page.
2. Find and open the “Gene → Morpholino + MO sequence” template.
3. Select the checkbox next to “constrain to be IN” saved Gene list and select “Hedgehog Signaling Pathway” from the menu of lists.
4. Click on “Show Results” to run the search.

This query returns a list of morpholinos and their sequences for the genes in the “Hedgehog Signaling Pathway” list.

Additional Example: run the same template query with the list you created above.
Lists
You can view, upload, analyze and combine lists. To view lists, go to the “Lists” tab and select “View”. To upload lists, go to the “Lists” tab and select “Upload”.

Note: lists can also be uploaded in the “Analyse” section of the home page.

Lists you have created will be highlighted. To save your lists, create an account and log in (see MyMine below).
To combine lists, check the checkboxes next to the names of the lists, and click “Union”. You will need to name the resulting new list in a popup window. You can also find the intersection of 2 lists, subtract lists from each other, and find the asymmetric difference between 2 lists.

Additional Example: *How many genes are in both of these gene lists?*
- OMIM Cancer Genes
- TGFbeta Signaling Pathway
Templates to Explore in Depth

Go to the “Templates” tab and try answering the questions provided for these 3 templates:

1. **Template: Wild-type expression**

   How many anterior-posterior pattern specification genes are expressed in the heart? (Hint: constrain the gene to be in the “Anterior-Posterior Pattern Specification Genes” list.)

   Additional exploration tips for this template:
   - You can exclude a list of genes by selecting “constrain to be NOT IN”.
   - To find all expression, replace “heart” with “*”.

2. **Template: Find antibodies that detect subcellular structures**

   How many antibodies in ZFIN detect protein expression the ciliary basal body? (Hints: Enter the term “ciliary basal body” into the “Ontology Term > Name (E)” box. Make sure that the antibody symbol is “*”).

   Additional exploration tips for this template:
   - To get a list of all ZFIN antibodies that detect subcellular structures, change the “Antibody > Symbol” field from *cc2d2a* to the wildcard “*”.

3. **Template: Find a gene annotated to a GO term and expressed in structure**

   How many genes involved in kidney development are expressed in the pronephros? (Hints: The default template returns genes annotated with the “fibroblast growth factor receptor binding” GO term and expressed in the “anterior macula”. Alter the defaults to find genes that are involved in “kidney development” and expressed in the “pronephros”.)

   Additional exploration tips for this template:
   - Since this query returns all the “children” of each term, the query will be very slow for higher level ontology terms. The use of a very general term like “protein kinase” is not recommended.
**Modify Templates and Build your own Searches**

You can modify existing templates and build your own templates in InterMine’s custom “Query Builder”. Query Builder is a powerful tool, but can be complicated to use. We will build an extremely basic query here, but you can contact us at zfinadmin@zfin.org if you have questions about putting together more complex queries. Detailed instructions on using Query Builder can be found here: http://flymine.readthedocs.org/en/latest/query-builder/Documentationquerybuilder.html#querybuilder

Click on the Query Builder tab:

On the Query Builder page, go to the “Select a Data Type to Begin a Query” section. Double-click the “Gene” data type to bring up the Gene section in the Model Browser.
The Model browser is on the left, with the “Gene” section of the ZebrafishMine data model expanded:

Click on the “SUMMARY” box next to “Gene”, which adds some of the basic data types associated with genes, such as gene symbol, gene name, and identifier, to the query:
Click on “CONSTRAIN” next to “Symbol” to limit your search. In the popup window, type “fgf8a” in the Gene>Symbol box. Click “Add to query”.

fgf8a has been added to the query:

Click “Show results” to run this query.

**Expand query:** To find where *fgf8a* is expressed, go back to the Model Browser and click on the “+” next to “Expressions”. Under “Expressions”, find “Anatomy” and click on the “+” next to it. Under “Anatomy”, find “Name” and click on “Show” to add the Anatomy term name to your query. Click the green “Show results” box to find out where *fgf8a* is expressed.
You can save your queries, and build templates from your queries, if you are logged in.

**To modify a pre-existing template** in Query Builder, click on the “Edit Query” button on the template page. Here is the “Gene → OMIM Disease Phenotype” template used earlier:

Go to the Model browser to add additional fields. For example, add the Expressions→Anatomy→Name fields we added to the first template:
Click “Show results”. Results will include the expression pattern of *ret*, and the OMIM disease phenotypes associated with the human ortholog of *ret*.

**MyMine**

You can create an account on ZebrafishMine, and save your lists, queries, and templates. To create an account, click on “Log in” at the top right hand corner of the home page. On the “Log in page, click on “Create account now” and follow the instructions. Your saved items can be found by clicking the MyMine tab when you are logged in.
POSTER:
See Poster # 329 (presented Thursday June 26, 9:15-10:30 pm)

URLs:
InterMine project
http://intermine.github.io/intermine.org/
Detailed InterMine user documentation:
http://flymine.readthedocs.org/en/latest/
Detailed InterMine developer documentation:
http://intermine.readthedocs.org/en/latest/

Further reading:
Sullivan, J., Karra, K., Moxon, S.A.T., Vallejos, A., Motenko, H., Wong, J.D.,
Aleksic, J., Balakrishnan, R., Binkley, G., Harris, T., Hitz, B., Jayaraman, P.,
Lyne, R., Neuhauser, S., Pich, C., Smith, R.N., Trinh, Q., Cherry, J.M.,
Richardson, J., Stein, L., Twigger, S., Westerfield, M., Worthey, E., and Micklem,
G. (2013) InterMOD: integrated data and tools for the unification of model

Sullivan, J., Karra, K., Moxon, S.A.T., Vallejos, A., Motenko, H., Wong, J.D.,
Aleksic, J., Balakrishnan, R., Binkley, G., Harris, T., Hitz, B., Jayaraman, P.,
Lyne, R., Neuhauser, S., Pich, C., Smith, R.N., Trinh, Q., Cherry, J.M.,
Richardson, J., Stein, L., Twigger, S., Westerfield, M., Worthey, E., and Micklem,
G. (2013) InterMOD: integrated data and tools for the unification of model
Answers to the questions in module:

*How many genes are in both of these gene lists?*
  OMIM Cancer Genes
  TGFbeta Signaling Pathway
  7 genes

*How many anterior-posterior pattern specification genes are expressed in the heart?* 6 genes

*How many antibodies in ZFIN detect protein expression the ciliary basal body?*
  5 antibodies

*How many genes involved in kidney development are expressed in the pronephros?*
  43 genes
Faceted Search

During the Madison 2014 meeting and for a limited time after the meeting, you will be able to access ZFIN's faceted search at:

http://zfinlabs.zfin.org/

A password will be provided.

Please use the prototype and provide feedback!