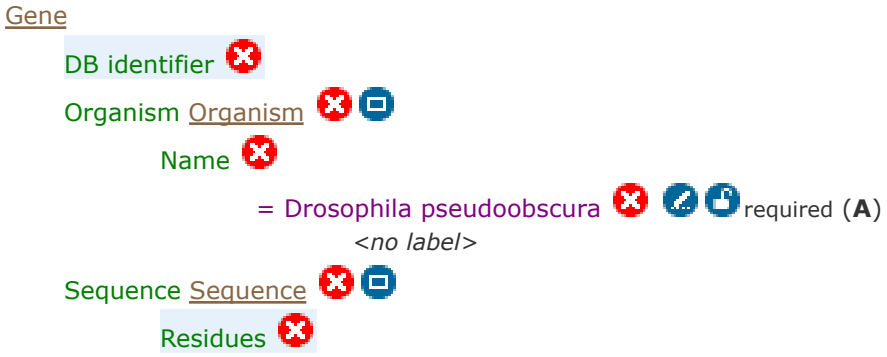
InterMine use-cases based on FlyMine and HumanMine

Hints and Answers

Use case 1: Analysis of transcripts from a fly transcriptomics study

A. The lists are transcripts - convert each list to a gene list. This can be done from the list analysis page, using the link in the top right pane.
B. To find genes in common, do a list intersect - you should get 239 genes in common
C. The genes in common are enriched for serine proteases (protein domain enrichment)
D. Use the template Gene→ Pathways to identify pathways for the two sets of genes.
E. The set of pathway names can be saved as a list.
F. Use the asymmetric difference function to identify pathways unique to each set
G. You could also do this analysis by identifying pathways enriched in one set but not the other. To do this save the set of pathways for each gene set from the pathways enrichment widget - select all pathways, save the pathways from the results table. Use the asymmetric difference to identify enriched pathways unique to each set.
H. You should find 1 enriched pathway shared by the two sets (metabolism) if you used the default enrichment settings.

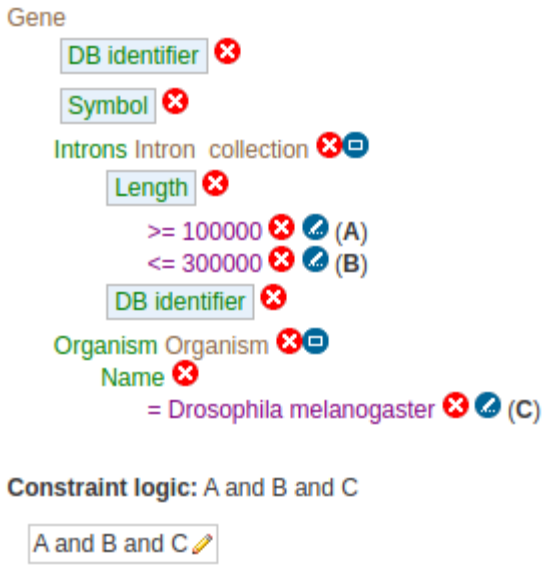
Use case 2: Creation of a blast database:



Once you have run the query you can export all the sequences in FASTA format: select EXPORT and change the export format to fasta.

Use case 3: Intron size and expression in Drosophila

1. You can create a query to search for introns of a certain size. Either create the query directly using the query builder or edit the template Gene → Introns:



HINT: you could also use the column summary to set the intron size range, rather than defining it in the query builder.

2. From the results of this query, save the set of introns.
3. Use the template **Introns → Overlapping regulatory regions** to identify regulatory regions within these introns (or try and build this yourself using the query builder].
4. Use the column summary on the Intron identifier column to find the intron with the most regulatory regions - intron intron_FBgn0052816:7_FBgn0052816:1 has 186 regulatory regions within it.
5. The gene symbol can be added by either using the Manage Columns function on the results table or by going to the query builder and adding the Gene.symbol field to the output. The Gene class is referenced from the Intron class.

Use case 4: analysis of transcription factor binding sites

The set of Dichaete coordinates need to be uploaded through the region search. For this exercise just select Gene from the Feature type menu and then upload the set of coordinates by choosing the appropriate file.

From the uploaded results, save the set of genes for all regions - you should get a list of 790 genes from the chr3 coordinates set.

A. Expression: use the widgets to look at the expression. From the expression graphs on the list analysis page you can see that many of these genes are expressed in the eye, brain, larval CNS, ovary and thoracioabdominal ganglion. From the FlyFish and BDGP graphs it is possible to identify genes from the list expressed in early development and save these as sub-lists.

B. Look at the Gene Ontology enrichment widget to identify genes annotated with developmental process terms.

C. One way to identify whether any of the overlapping genes are also transcription factors is to make use of the public list: PL flyTF_trusted_TFs - a set of genes experimentally verified to be transcription factors. A list intersection with this set of genes gives 30 genes (for the Chr 3 coordinates). You could also make use of template searches to look at the REDfly transcription factor data.

Use case 5: Analysis of genes from a Drosophila RNAi screen

A. Link to HumanMine from list analysis page -> 1170 Human genes
B. GWAS template gives a list of around 120-140 genes, depending on terms selected
C. Back to FlyMine gives around 300 genes but this needs to be intersected with the original set of 510 screen hits to give around 100- 120 genes.
D. Gene ontology enrichment of this set - set the UseCase1_backgroundGeneList as the background population.
E. Gene Ontology enrichment: Depends on the gene returned, you may not get any enrichment. Since we are dealing with smaller numbers of genes, you could turn off the multiple test correction. You should get terms for ion transport enriched.
F. Protein domain enrichment: Neurotransmitter-gated ion-channel or Epithelial sodium channel
G. Large number expressed in brain and larval CNS
H. Use gene ontology template

Use Case 6: Analysis of genes involved in Parkinson's disease

1. PL_GenomicsEngland_GenePanel:Parkinson_Disease_and_Complex_Parkinsonism

2. Template: Gene → Disease (OMIM)
3. Gene → Alleles and Disease (clinVar data). Filter on clinical significance and disease name.
4. Could again use the Gene → Disease (OMIM) template and the Gene → Alleles and Disease (clinVar data) template and filter for each type. Create lists and intersect.
5. Template: Gene A --> Interaction <-- Gene B
6. Many gene expression templates