

# Exercises

---

Biological Data Analysis Using InterMine  
workshop exercises with answers

# Exercise1: Faceted Search

## Use HumanMine for this exercise

1. Search for one or more of the following using the keyword search (result for PPARG only shown below):

- **PPARG**
- **rs10509540**
- **diabetes\***

Human *PPARG* is the first gene result returned when HumanMine is searched for *PPARG*. Click on this gene to be taken to the report page.

### Search results 1 to 100 out of 273 for *pparg*

<< First < Previous | Next > Last >>  
0.029s

Categories

Hits by Category

- Publication: 250
- Gene: 8
- Protein: 6
- Protein Structure: 4
- UniProt Feature: 4
- Protein Domain: 1

Hits by Organism

- H. sapiens: 7
- M. musculus: 4
- R. norvegicus: 3

Type	Details	Score
Gene	<b>PPARG</b>   peroxisome proliferator activated receptor gamma   5468   ENSG00000132170 Length: 183164 FASTA... Organism . Name: Homo sapiens	.....
Gene	<b>Pparg</b>   peroxisome proliferator activated receptor gamma   MGI:97747   - Length: Organism . Name: Mus musculus	.....
Gene	<b>Pparg</b>   peroxisome proliferator-activated receptor gamma   RGD:3371   - Length: Organism . Name: Rattus norvegicus	.....
Protein	<b>Q4FJR2_MOUSE</b>   <b>Q4FJR2</b> Organism . Name: Mus musculus Length: 505 FASTA...	.....
Gene	<b>PPARGC1B</b>   PPARG coactivator 1 beta   133522   ENSG00000155846 Length: 127623 FASTA... Organism . Name: Homo sapiens	.....

# Exercise1: Faceted Search:

## 2. Filter and create a list:

- Search for diabetes\*
- Filter for genes
- Filter for *H. sapiens*
- Make a list of these genes

Use the categories to filter for “Genes” from “Homo Sapiens”

Search results 1 to 100 out of 25034 for *diabetes*\*

<< First < Previous | Next > Last >>  
0.026s

**Categories**

Hits by Category

- Gene: 548
- Author: 105
- Disease: 99
- Protein Domain: 95
- GWAS: 44
- UniProt Feature: 31
- Ontology Term Synonym: 23
- Interaction Experiment: 17
- ... and 11 more values »

Hits by Organism

- H. sapiens: 53
- M. musculus: 23
- ... and 11 more values »

Type	Details	Score
Publication	<b>18547236</b> First Author: Diabetes Research in Children Network (DirecNet) Study Group Title: Adiponectin and catecholamine concentrations during acute exercise in children with type 1 diabetes. Year: 2008 Journal: Pediatr Diabetes Volume: 9 Pages: 221-7	● ● ● ● ●
Publication	<b>20628086</b> First Author: Bailey Swneke D Title: Variation at the NFATC2 locus increases the risk of thiazolidinedione-induced edema in the Diabetes Reduction Assessment with ramipril and rosiglitazone Medication (DREAM) study. Year: 2010 Journal: Diabetes Care Volume: 33 Pages: 2250-3	● ● ● ● ●
Publication	<b>18003760</b> First Author: Villarreal-Molina M Teresa Title: Association of the ATP-binding cassette transporter A1 R230C variant with early-onset type 2 diabetes in a Mexican population. Year: 2008 Journal: Diabetes Volume: 57	● ● ● ● ●

Use the checkbox in the header to select all the genes and make a list

Search results 1 to 33 out of 33 for *diabetes*\*

Category restricted to Gene

Organism restricted to H. sapiens

0.101s

**Categories**

Category: Gene  
« show all

Organism: H. sapiens  
« show all

<input checked="" type="checkbox"/>	Type	Details	Score
<input checked="" type="checkbox"/>	Gene	<b>TNDM</b>   diabetes mellitus, transient neonatal   7952   - Length: - Organism . Name: Homo sapiens	● ● ● ● ●
<input checked="" type="checkbox"/>	Gene	<b>IDDM23</b>   Diabetes mellitus, insulin-dependent, 23   100271697   - Length: - Organism . Name: Homo sapiens	● ● ● ● ●
<input checked="" type="checkbox"/>	Gene	<b>IDDMX</b>   Diabetes mellitus, insulin-dependent, X-linked, susceptibility to   8245   - Length: - Organism . Name: Homo sapiens	● ● ● ● ●
<input checked="" type="checkbox"/>	Gene	<b>RRAD</b>   RRAD, Ras related glycolysis inhibitor and calcium channel regulator   6236   ENSG00000166592 Length: 3858 FASTA... Organism . Name: Homo sapiens	● ● ● ● ●

## Exercise2: Exploring a Gene:

You are interested in the Human *PPARG* gene and want to know the following things about it:

### Use HumanMine for this exercise

1. On which chromosome is *PPARG* located?
2. Can I access the sequence for the *PPARG* gene?
3. With which diseases is *PPARG* associated?
4. In which tissues is *PPARG* most highly expressed?
5. Does the *PPARG* protein have any know isoforms?
6. Is there a *PPARG* orthologue in *D. melanogaster*?
7. Does this orthologue interact with any other Genes/proteins? Identify the interaction type (genetic/physical)
8. For the interaction with Kr, what was the original experiment and publication that determined this interaction

## Exercise2: Exploring a Gene:

You are interested in the Human *PPARG* gene and want to know the following things about it:

### Use HumanMine for this exercise

1. On which chromosome is *PPARG* located?
2. Can I access the sequence for the *PPARG* gene?
3. With which diseases is *PPARG* associated?
4. In which tissues is *PPARG* most highly expressed?
5. Does the *PPARG* protein have any know isoforms?
6. Is there a *PPARG* orthologue in *D. melanogaster*?
7. Does this orthologue interact with any other Genes/proteins? Identify the interaction type (genetic/physical)
8. For the interaction with Kr, what was the original experiment and publication that determined this interaction

## Exercise2: Exploring a Gene:

1. On which chromosome is *PPARG* located?
2. Can I access the sequence for the *PPARG* gene?

The first section on the gene report page provides information about the chromosome location of the gene along with identifiers and synonyms and a link to the FASTA DNA sequence

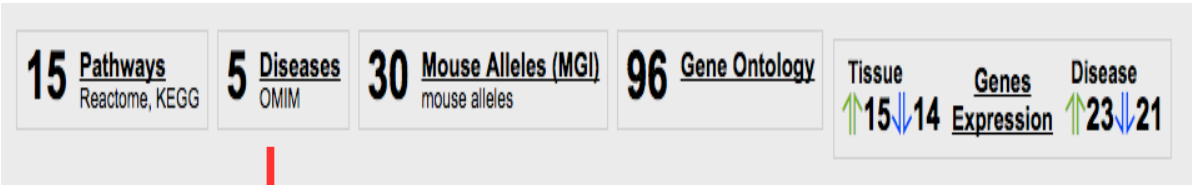
Gene : **PPARG** *Homo sapiens*

Name	peroxisome proliferator activated receptor gamma	Brief Description	peroxisome proliferator-activated receptor gamma
synonyms:	PPARG1, PPARG1, PPARGgamma, CIMT1, PPARG2, NM_005037, HGNC:9236, PPARG2, <a href="#">Show more</a>		
identifiers:	5468, ENSG00000132170, PPARG		
Region:	gene	Length:	183164 <a href="#">FASTA...</a>
Location:	3:12287850-12471013		

# Exercise2: Exploring a Gene:

## 3. With which diseases is PPARG associated?

In HumanMine a summary of data is provided at the top of the report page (note this feature is not available in all InterMine's).



“Diseases” links to a table for OMIM data

**Disease**

5 Diseases

Trait: Gene

☐ Manage Columns ☐ Manage Filters

☐ Manage Relationships

Showing 1 to 5 of 5 rows

Diseases Identifier	Diseases Name
OMIM:125853	DIABETES MELLITUS, NONINSULIN-DEPENDENT
OMIM:601665	OBESITY
OMIM:604367	LIPODYSTROPHY, FAMILIAL PARTIAL, TYPE 3
OMIM:609338	CAROTID INTIMAL MEDIAL THICKNESS 1
ORPHANET:528	NO VALUE

Further disease information is sometimes available from the “Curated comments from Uniprot”:

Curated comments from UniProt		Show proteins
Type	Comment	
disease	Defects in PPARG can lead to type 2 insulin-resistant diabetes and hypertension. PPARG mutations may be associated with colon cancer.	
disease	MIM:137800; Glioma 1; GLM1; Gliomas are benign or malignant central nervous system neoplasms derived from glial cells. They comprise astrocytomas and glioblastoma multiforme that are derived from astrocytes, oligodendrogliomas derived from oligodendrocytes and ependymomas derived from ependymocytes. Disease susceptibility may be associated with variations affecting the gene represented in this entry. Polymorphic PPARG alleles have been found to be significantly over-represented among a cohort of American patients with sporadic glioblastoma multiforme suggesting a possible contribution to disease susceptibility.	
disease	MIM:601665; Obesity; OBESITY; A condition characterized by an increase of body weight beyond the limitation of skeletal and physical requirements, as the result of excessive accumulation of body fat. Disease susceptibility may be associated with variations affecting the gene represented in this entry.	
disease	MIM:604367; Lipodystrophy, familial partial, 3; FPLD3; A form of lipodystrophy characterized by marked loss of subcutaneous fat from the extremities. Facial adipose tissue may be increased, decreased or normal. Affected individuals show an increased preponderance of insulin resistance, diabetes mellitus and dyslipidemia. The disease is caused by mutations affecting the gene represented in this entry.	

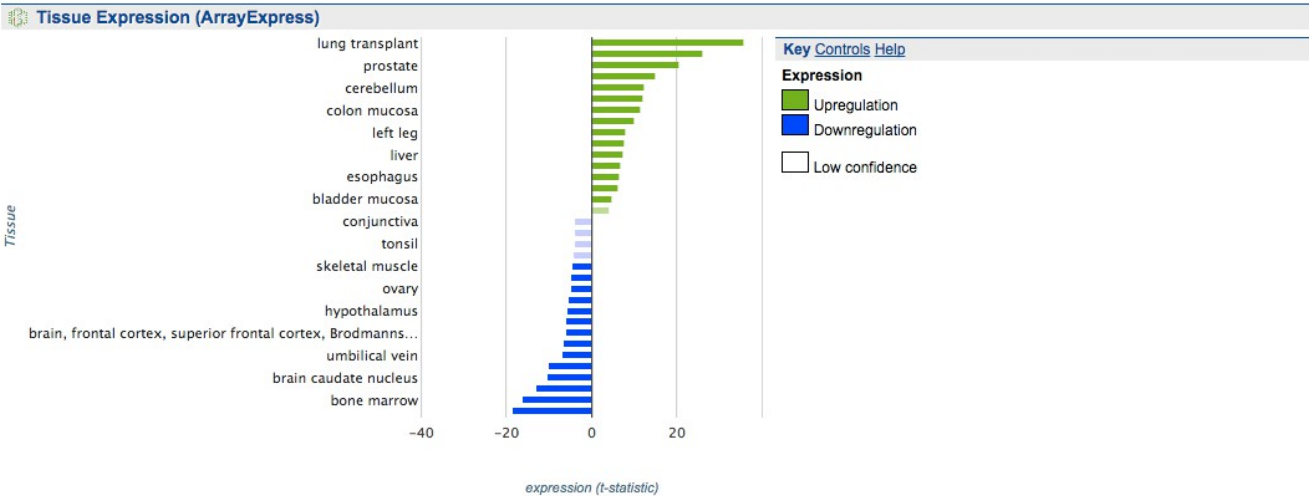


# Exercise2: Exploring a Gene:

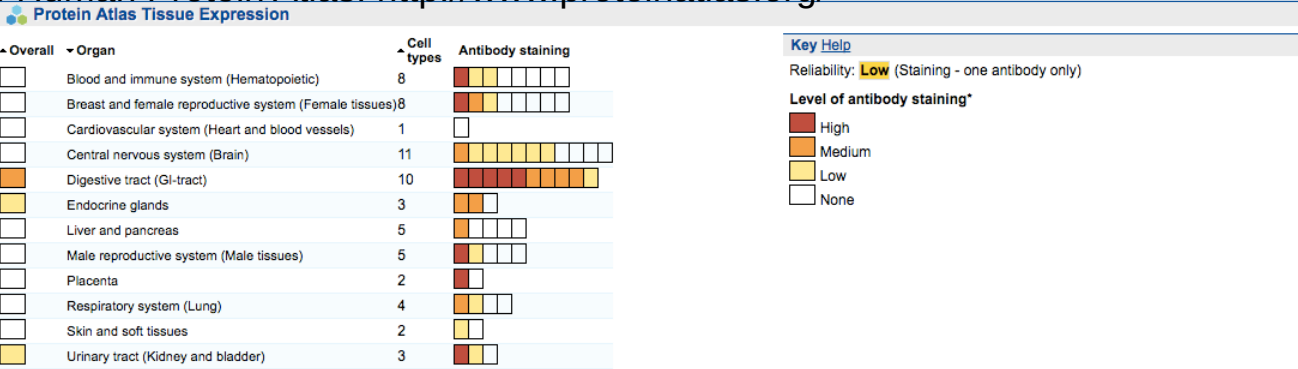
## 4. In which tissues is PPARG most highly expressed?

Data on tissue expression can be found from three sources:

1. Human gene expression atlas of 5372 samples representing 369 different cell and tissue types, disease states and cell lines: from <http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-62/>



2. Human Protein Atlas: <http://www.proteinatlas.org/>



3. Curated comments from Uniprot:

Curated comments from UniProt		Show proteins
Type	Comment	
disease	Defects in PPARG can lead to type 2 insulin-resistant diabetes and hypertension. PPARG mutations may be associated with colon cancer.	
disease	MIM:137800; Glioma 1; GLM1; Gliomas are benign or malignant central nervous system neoplasms derived from glial cells. They comprise astrocytomas and glioblastoma multiforme that are derived from astrocytes, oligodendrogliomas derived from oligodendrocytes and ependymomas derived from ependymocytes. Disease susceptibility may be associated with variations affecting the gene represented in this entry. Polymorphic PPARG alleles have been found to be significantly over-represented among a cohort of American patients with sporadic glioblastoma multiforme suggesting a possible contribution to disease susceptibility.	
disease	MIM:601665; Obesity; OBESITY: A condition characterized by an increase of body weight beyond the limitation of skeletal and physical requirements, as the result of excessive accumulation of body fat. Disease susceptibility may be associated with variations affecting the gene represented in this entry.	
disease	MIM:604367; Lipodystrophy, familial partial, 3; FPLD3; A form of lipodystrophy characterized by marked loss of subcutaneous fat from the extremities. Facial adipose tissue may be increased, decreased or normal. Affected individuals show an increased preponderance of insulin resistance, diabetes mellitus and dyslipidemia. The disease is caused by mutations affecting the gene represented in this entry.	
function	Nuclear receptor that binds peroxisome proliferators such as hypolipidemic drugs and fatty acids. Once activated by a ligand, the nuclear receptor binds to DNA specific PPAR response elements (PPRE) and modulates the transcription of its target genes, such as acyl-CoA oxidase. It therefore controls the peroxisomal beta-oxidation pathway of fatty acids. Key regulator of adipocyte differentiation and glucose homeostasis. ARF6 acts as a key regulator of the tissue-specific adipocyte P2 (aP2) enhancer. Acts as a critical regulator of gut homeostasis by suppressing NF-kappa-B-mediated proinflammatory responses. Plays a role in the regulation of cardiovascular circadian rhythms by regulating the transcription of ARNTL/BMAL1 in the blood vessels (By similarity).	
similarity	Belongs to the nuclear hormone receptor family.	
similarity	Belongs to the nuclear hormone receptor family. NR1 subfamily.	
similarity	Contains 1 nuclear receptor DNA-binding domain.	
similarity	Contains nuclear receptor DNA-binding domain.	
subcellular location	Redistributed from the nucleus to the cytosol through a MAP2K1/MEK1-dependent manner. NUCY1 enhances its nuclear translocation.	
tissue specificity	Highest expression in adipose tissue. Lower in skeletal muscle, spleen, heart and liver. Also detectable in placenta, lung and ovary.	



# Exercise2: Exploring a Gene:

## 5. Does the *PPARG* protein have any know isoforms?

a. Navigate to the proteins table using the quick links

Quick Links: [Summary](#) [Function](#) [Genomics](#) [Proteins](#) [SNPs](#) [Disease](#) [Homology](#) [Interactions](#) [Gene Ontology](#) [Other](#)

**Proteins**

Gene --> Proteins. (5 rows)

[Manage Columns](#) [Manage Filters](#) [Manage Relationships](#) [Save as List](#) [Generate Python code](#) [Export](#)

Showing 1 to 5 of 5 rows

Gene Primary Identifier	Gene Symbol	Proteins Primary Identifier	Proteins Primary Accession	Proteins Name
5468	PPARG	D2KUA6_HUMAN	D2KUA6	Peroxisome proliferator-activated receptor gamma isoform 1
5468	PPARG	E9PFX5_HUMAN	E9PFX5	Peroxisome proliferator-activated receptor gamma
5468	PPARG	PPARG_HUMAN	P37231	Peroxisome proliferator-activated receptor gamma
5468	PPARG	PPARG_HUMAN-2	P37231-2	Peroxisome proliferator-activated receptor gamma
5468	PPARG	PPARG_HUMAN-3	P37231-3	Peroxisome proliferator-activated receptor gamma

b. Select the PPARG\_HUMAN protein (the canonical uniprot annotated protein) to be taken to the protein report page. Note: we can also see from this table that PPARG has two isoforms: PPARG\_HUMAN2 and PPARG\_HUMAN3

c. Navigate to the Isoforms table on the protein report page. Note that this table links to a report page for each of the isoforms.

**2 Isoforms**

[Manage Columns](#) [Manage Filters](#) [Manage Relationships](#) [Save as List](#) [Generate Python code](#) [Export](#)

Showing 1 to 2 of 2 rows

Isoforms Primary Identifier	Isoforms Primary Accession	Isoforms Organism . Name	Isoforms Length
PPARG_HUMAN-2	P37231-2	Homo sapiens	477
PPARG_HUMAN-3	P37231-3	Homo sapiens	186

# Exercise2: Exploring a Gene:

## 6. Is there a *PPARG* orthologue in *D. melanogaster*?

Links to other Mines

RatMine

R. norvegicus

Pparg

YeastMine

No results

FlyMine

D. melanogaster

Eip75B, Hr96, Eip75C

MouseMine

M. musculus

Pparg

ZebrafishMine

D. rerio

ZDB-GENE-990415-213

Use the “Links to other Mines” to navigate to the *D. melanogaster* orthologue in FlyMine

Note that there are three othologous fly genes. For this exercise select the first.

## 7. Does this orthologue interact with any other proteins? Identify the interaction type (genetic/physical).

Use the “Interactions” quick link to navigate to protein and genetic Interaction data. Eip75B has a genetic interaction with genes ph-p and Kr

Quick Links: Summary Function Genes Homology Expression Regulation **Interactions** Other

Interactions

Interaction Network

Eip75B

ph-p

Kr

Show the following interaction types:

☒ Genetic

☐ All

☐ Physical

Reset view

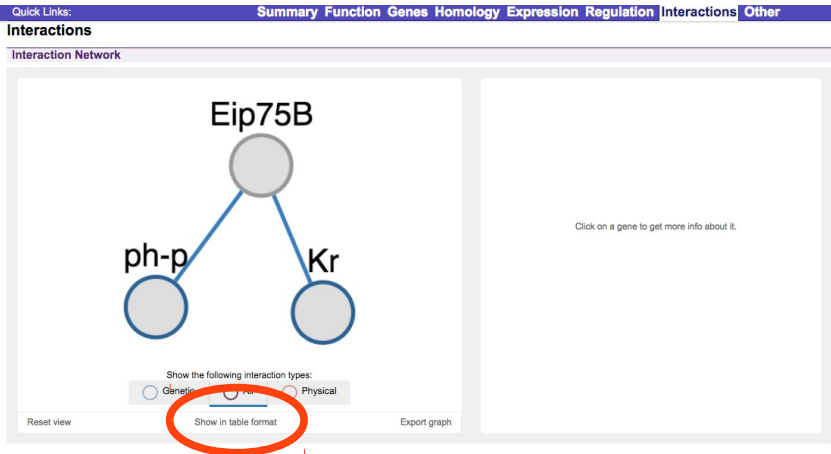
Show in table format

Export graph

Click on a gene to get more info about it.

# Exercise2: Exploring a Gene:

8. For the interaction with Kr, what was the original experiment and publication that determined this interaction



a. Select "Show in table"

b. Select the interaction name

Gene DB identifier	Gene Symbol	Participant 2 Symbol	Interactions Participant 2 . Primary Identifier	Details Name	Details Role 1	Details Role 2	Details Type	Data Sets Name
FBgn0000568	Eip75B	Kr	FBgn0001325	Eip75B-Kr	bait	prey	genetic	BioGRID interaction data set
FBgn0000568	Eip75B	Kr	FBgn0001325	Eip75B-Kr	bait	prey	genetic	BioGRID interaction data set
FBgn0000568	Eip75B	Kr	FBgn0001325	FlyBase:FBgn0000568_FBgn0001325	unspecified	unspecified	genetic	FlyBase data set for Drosophila melanogaster

## 1 Experiment

Add tags

Manage Columns

Manage Filters

Manage Relationships

Save as List

Generate Python code

Export

Showing 1 to 1 of 1 row

c. Select the experiment name

Experiment Name	Experiment Host Organism	Experiment Description
1 Carrera P (1998)	NO VALUE	A modifier screen in the eye reveals control genes for Krueppel activity in the Drosophila embryo.

Add tags

Manage Columns

Manage Filters

Manage Relationships

Save as List

Generate Python code

Export

Showing 1 to 1 of 1 row

d. Information about the publication is on the experiment page

Publication First Author	Publication Title	Publication Year	Publication Journal	Publication Volume	Publication Pages	Publication PubMed ID
Carrera P	A modifier screen in the eye reveals control genes for Krüppel activity in the Drosophila embryo.	1998	Proc. Natl. Acad. Sci. U.S.A.	95	10779-8	9724781

# Exercise3: List Upload:

## Use FlyMine for this exercise

1. Navigate to the lists tab and upload sub-tab
2. Select the example list (leave type and organism as the default values).
3. Click “Create list”.
4. Examine and understand the list page, name and save your list.

E2f has matched two genes (**duplicates**) - in this case you need to decide which of the two genes you want in your list (or both). The action column allows you to do this.

Duplicates found - which one(s) do you want? [?](#)

Add all

Remove all

Identifier you provided	Matches <a href="#">?</a>					Action <a href="#">?</a>
	organism name	primary identifier	symbol	class	secondary identifier	
E2f	Drosophila melanogaster	FBgn0011766	E2f1	Gene	CG6376	<a href="#">Add</a>
	Drosophila melanogaster	FBgn0024371	E2f2	Gene	CG1071	<a href="#">Add</a>

Two of the identifiers in the list matched the same gene: FBgn0010433 and ato. This is indicated in the **direct hits**.

Direct Hits [?](#) Non-Genes Identifiers [?](#) Synonyms [?](#)

Page 1 of 5

1 2 3 4 5

5 rows per page

Identifier you provided	Match <a href="#">?</a>				
	organism name	primary identifier	symbol	class	secondary identifier
FBgn0000000	Drosophila melanogaster	FBgn0000099		Gene	
FBgn0010433, ato <a href="#">?</a>	Drosophila melanogaster	FBgn0010433	ato	Gene	CG7508
CG2328	Drosophila melanogaster	FBgn0000606	eve	Gene	CG2328
CG9786	Drosophila melanogaster	FBgn0001180	hb	Gene	CG9786
so	Drosophila melanogaster	FBgn0003460	so	Gene	CG11121

## Exercise3: List Upload:

One of the identifiers is a protein identifier (TWIST\_DROME). As the associated gene could be identified, this has been added to the list. This is shown under **non-gene identifiers**.

Direct Hits

Non-Gene Identifiers

Synonyms

Identifier you provided	Match				
	organism name	primary identifier	symbol	class	secondary identifier
TWIST_DROME	Drosophila melanogaster	FBgn0003900	twi	Gene	CG2956

One of the identifiers only matched a **synonym** (rather than a current identifier). As this matched only one gene, this has been added to the list.

Direct Hits

Non-Gene Identifiers

Synonyms

Identifier you provided	Match				
	organism name	primary identifier	symbol	class	secondary identifier
FBgn0001251	Drosophila melanogaster	FBgn0001325	Kr	Gene	CG3340

# *Exercise4: List Analysis Pages:*

---

## **Use FlyMine for this exercise**

### **Examine the FlyMine public list: PL\_FlyTF\_putativeTFs**

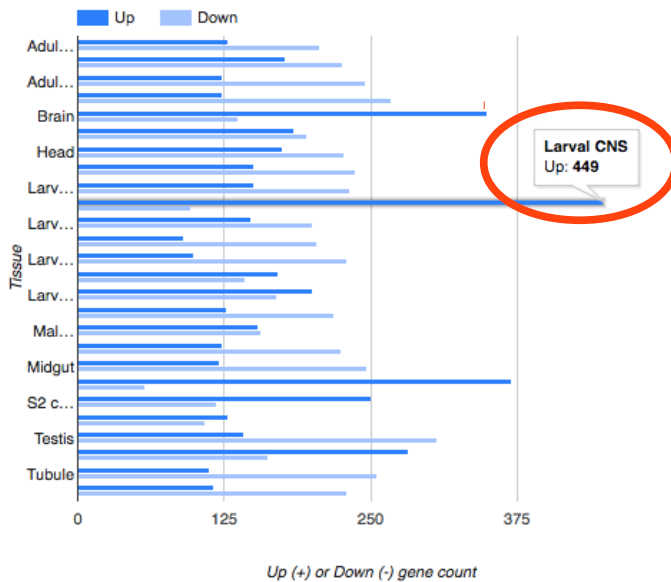
1. In which tissues are these genes most highly expressed?
2. What is the most enriched GO term for this list?
3. How many genes in the list are annotated with this GO term?

Note: you could make a sub-list containing only genes from this list annotated with this term by clicking on the matches number

4. Navigate to the MouseMine database to examine the mouse orthologues for this list.
5. How many mouse orthologues are there for this list?
6. Are these mouse genes enriched for any phenotypes (Mammalian Phenotype Ontology)?

# Exercise4: List Analysis Pages:

## 1. In which tissues are these genes most highly expressed?



449 genes are up-regulated in the larval CNS

## 2. What is the most enriched GO term for this list?

## 3. How many genes in the list are annotated with this GO term?

### Gene Ontology Enrichment

GO terms enriched for items in this list.

Number of Genes in this list not analysed in this widget: 212

Test Correction

Max p-value

Ontology

Holm-Bonferroni

0.05

biological\_process

Background population

Default Change

View Download

<input type="checkbox"/> GO Term	p-Value	Matches
<input type="checkbox"/> transcription, DNA-templated [GO:0006351]	0.000000e+0	439
<input type="checkbox"/> regulation of transcription, DNA-templated [GO:0006355]	0.000000e+0	430
<input type="checkbox"/> regulation of biosynthetic process [GO:0009889]	0.000000e+0	431
<input type="checkbox"/> regulation of gene expression [GO:0010468]	0.000000e+0	438

439 genes in the list are annotated with the GO term "transcription, DNA templated". This is the most enriched GO term

You can click on this number to create A sub-list of just these 458 genes



## Exercise4: List Analysis Pages:

4. Navigate to the MouseMine database to examine the mouse orthologues for this list.

5. How many mouse orthologues are there for this list?

**View homologues in other Mines:**

- RatMine  
R. norvegicus
- YeastMine  
S. cerevisiae
- MouseMine**  
M. musculus
- HumanMine  
H. sapiens
- ZebrafishMine  
D. rerio

**List Analysis for link\_2 (1123 Genes)**

Columns Manage Filters Generate Python code Export Save as List

relationships

25 of 1,123 rows Rows per page: 25

Gene Primary Identifier	Gene Symbol	Gene Name	Gene Type	Gene Chromosome
MGI:101786	Mtf1	metal response element binding transcription factor 1	protein_coding_gene	4
MGI:101791	Isl1	ISL1 transcription		
MGI:101838	Tbp	TATA box binding		

**Mammalian Phenotype Ontology Enrichment**

MP terms enriched for items in this list.

Number of Genes in this list not analysed in this widget: 388

Test Correction: Holm-Bonferroni Max p-value: 0.05 Background population: Default

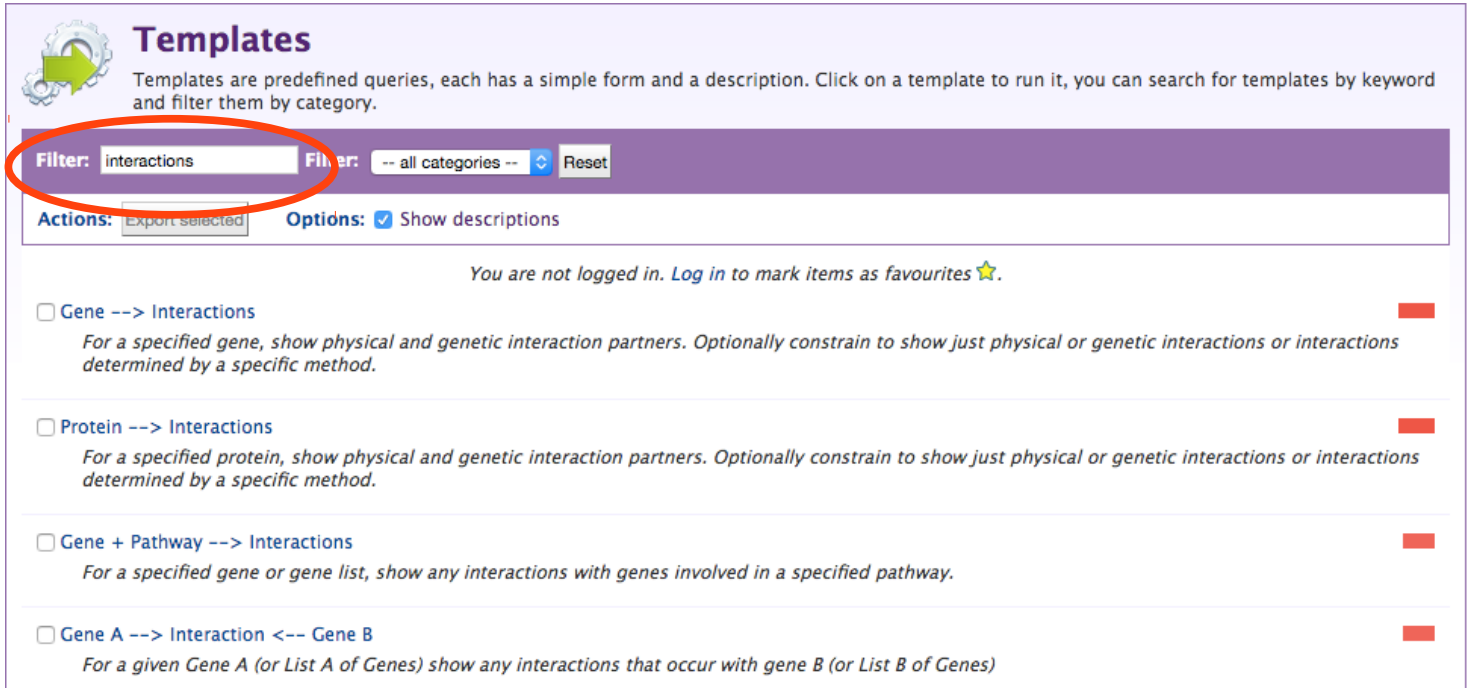
View Download

MP Term	p-Value	Matches
lethality during fetal growth through weaning [MP:0010832]	7.682887e-104	335
preweaning lethality [MP:0010770]	1.091598e-93	445
abnormal survival [MP:0010769]	8.950768e-88	484
mortality/aging [MP:0010768]	2.838162e-85	498
perinatal lethality [MP:0002081]	1.181186e-77	222
embryo phenotype [MP:0005380]	4.521712e-67	244
abnormal embryonic tissue morphology [MP:0002085]	3.571684e-65	179
neonatal lethality [MP:0002058]	5.066740e-65	172
abnormal embryo morphology	1.329409e-63	207

6. Are these mouse genes enriched for any phenotypes?

# Exercise 5: Template searches:

1. Browse the template searches in FlyMine and HumanMine - try running a few or changing the filters.
2. Use the search box to find template searches for interactions



**Templates**

Templates are predefined queries, each has a simple form and a description. Click on a template to run it, you can search for templates by keyword and filter them by category.

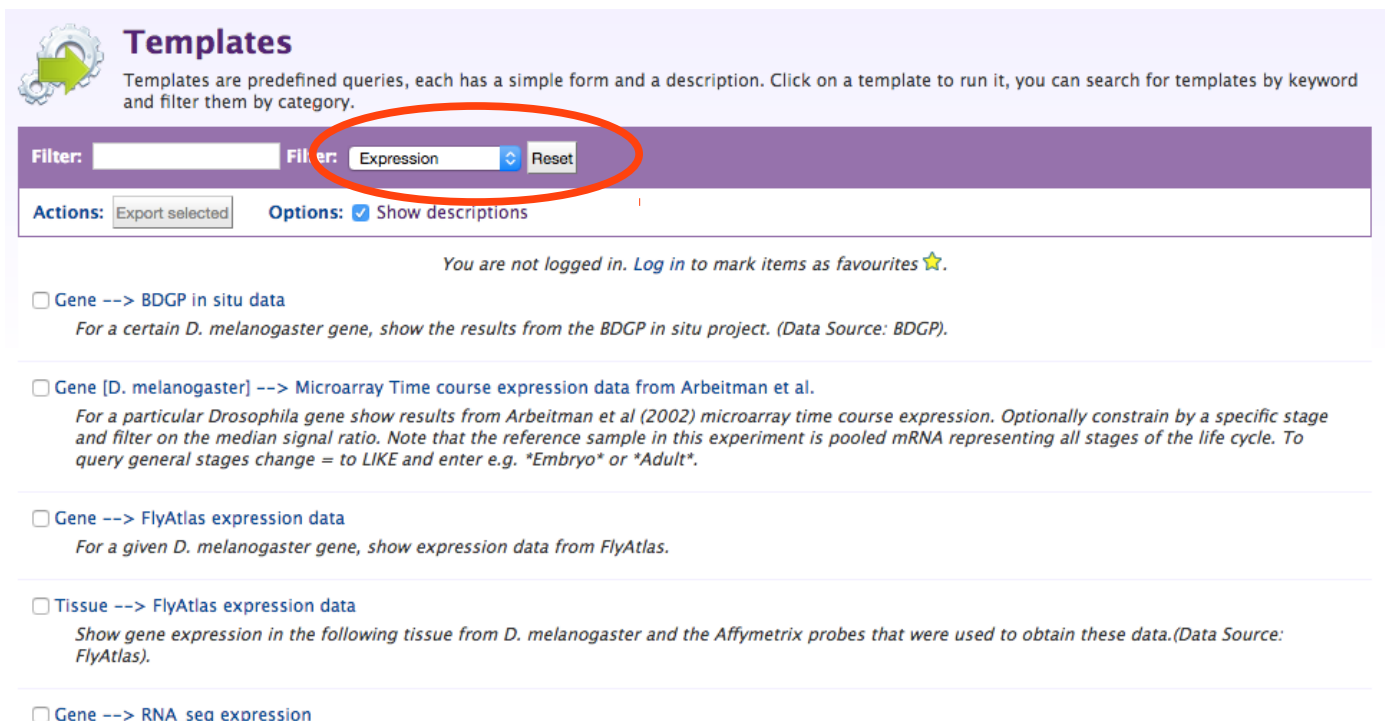
Filter:  Filter:

Actions:  Options: ☒ Show descriptions

You are not logged in. [Log in](#) to mark items as favourites ☆.

- ☐ **Gene --> Interactions**  
For a specified gene, show physical and genetic interaction partners. Optionally constrain to show just physical or genetic interactions or interactions determined by a specific method.
- ☐ **Protein --> Interactions**  
For a specified protein, show physical and genetic interaction partners. Optionally constrain to show just physical or genetic interactions or interactions determined by a specific method.
- ☐ **Gene + Pathway --> Interactions**  
For a specified gene or gene list, show any interactions with genes involved in a specified pathway.
- ☐ **Gene A --> Interaction <-- Gene B**  
For a given Gene A (or List A of Genes) show any interactions that occur with gene B (or List B of Genes)

3. Filter the FlyMine template searches to show only “expression” templates.



**Templates**

Templates are predefined queries, each has a simple form and a description. Click on a template to run it, you can search for templates by keyword and filter them by category.

Filter:  Filter:

Actions:  Options: ☒ Show descriptions

You are not logged in. [Log in](#) to mark items as favourites ☆.

- ☐ **Gene --> BDGP in situ data**  
For a certain *D. melanogaster* gene, show the results from the BDGP in situ project. (Data Source: BDGP).
- ☐ **Gene [D. melanogaster] --> Microarray Time course expression data from Arbeitman et al.**  
For a particular *Drosophila* gene show results from Arbeitman et al (2002) microarray time course expression. Optionally constrain by a specific stage and filter on the median signal ratio. Note that the reference sample in this experiment is pooled mRNA representing all stages of the life cycle. To query general stages change = to LIKE and enter e.g. "Embryo" or "Adult".
- ☐ **Gene --> FlyAtlas expression data**  
For a given *D. melanogaster* gene, show expression data from FlyAtlas.
- ☐ **Tissue --> FlyAtlas expression data**  
Show gene expression in the following tissue from *D. melanogaster* and the Affymetrix probes that were used to obtain these data. (Data Source: FlyAtlas).
- ☐ **Gene --> RNA\_seq expression**

## *Exercise 6: Using template searches:*

---

In our exploration of the PPARG gene we could see that it is Up-regulated in adipose tissue. Use the template searches in HumanMine to answer the following questions:

1. What other genes are up-regulated in adipose tissue according to the array express dataset? Don't forget to add a filter for "UP" expression. Save a list of these genes.
2. Do any of the genes identified in 1. interact with PPARG? Save a list of these genes.

# Exercise 6: Using template searches:

## Use HumanMine for this exercise

1. What other genes are up-regulated in adipose tissue according to the Array express dataset? Don't forget to add a filter for "UP" expression. Save a list of these genes.

Find the following template and set "Condition = Adipose tissue"


 **Tissue → Gene Expression (Array Express)** ☆

Show expression for a gene or list of genes associated with a specific tissue(s). Optionally constrain the P-value. Data source: The Gene Expression Atlas (<http://www.ebi.ac.uk/gxa/>). Keywords: expression, expressed, Genes, Disease state, P value, T statistic.

**Atlas Expression > Condition**

**Atlas Expression > P Value**

**Show Results**

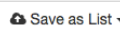
 Edit Query  Edit Template

Filter the column "Atlas Expression.Expression" for "UP"

"Save as List" : Atlas Expression > Gene. You should have a list of 1, 654 genes


### Tissue → Gene Expression (Array Express)

Show expression for a gene or list of genes associated with a specific tissue(s). Optionally constrain the P-value. Data source: The Gene Expression Atlas (<http://www.ebi.ac.uk/gxa/>). Keywords: expression, expressed, Genes, Disease state, P value, T statistic.

☐ Manage Columns ☐ Manage Filters ☐ Manage Relationships  Save as List ☐ Generate Python code ☐ Export

Showing 1 to 25 of 5,659 rows

Rows per page: 25

Atlas Expression Condition	Gene Primary Identifier	Gene Symbol	Gene Name	Atlas Expression	2 Atlas Expression Expressions	Data Sets Name
adipose tissue	1000	CDH2	cadherin 2	DOWN	 2,204 Items Selected Filter values Atlas Expression Expression Count <input type="checkbox"/> DOWN 3,455 <input checked="" type="checkbox"/> UP 2,204 Filter Download data	ArrayExpress accession: E-MTAB-62
adipose tissue	1000	CDH2	cadherin 2	DOWN		ArrayExpress accession: E-MTAB-62
adipose tissue	10000	AKT3	AKT serine/threonine kinase 3	UP		ArrayExpress accession: E-MTAB-62
adipose tissue	10005	ACOT8	acyl-CoA thioesterase 8	DOWN		ArrayExpress accession: E-MTAB-62
adipose tissue	10007	GNPDA1	glucosamine-6-phosphate deaminase 1	DOWN		ArrayExpress accession: E-MTAB-62
adipose tissue	10010	TANK	TRAF family member associated NFKB activator	UP		ArrayExpress accession: E-MTAB-62
adipose tissue	10013	HDAC6	histone deacetylase 6	DOWN		ArrayExpress accession: E-MTAB-62
adipose tissue	100130348	TMEM262	transmembrane protein 262	DOWN		ArrayExpress accession: E-MTAB-62
adipose tissue	10016	PDCD6	programmed cell death 6	DOWN		ArrayExpress accession: E-MTAB-62
adipose tissue	10019	SH2B3	SH2B adaptor protein 3	UP		ArrayExpress accession: E-MTAB-62
adipose tissue	1002	CDH4	cadherin 4	DOWN		ArrayExpress accession: E-MTAB-62

# Exercise 6: Using template searches:

## 2. Do any of the genes identified in 1. interact with PPARG? Save a list of these genes.

Use the following template: Set the first condition to “PPARG” and the second condition to your saved list.

**Gene A → Interaction ← Gene B** ☆  
For a given Gene A (or List A of Genes) show any interactions that occur with gene B (or List B of Genes)

**Gene - Show interactions between gene(s):**  
LOOKUP:  for Organism:   
☐ constrain to be  saved Gene list

**Gene > Participant 2 - and gene(s):**  
LOOKUP:  for Organism:   
☒ constrain to be  saved Gene list

**Show Results** [Edit Query](#) [Edit Template](#)

[web service URL](#) [Perl](#) [Python](#) [Ruby](#) [Java \[help\]](#) [export XML](#)

Save the genes from Gene > Interaction > Participant 2 (24 genes)

Trail: Query

**Gene A → Interaction ← Gene B** ☆  
For a given Gene A (or List A of Genes) show any interactions that occur with gene B (or List B of Genes)

[Manage Columns](#) [Manage Filters](#) [Manage Relationships](#) [Save as List](#) [Generate Python code](#) [Export](#)

Showing 1 to 25 of 60 rows

Gene Primary Identifier	Gene Symbol	Interactions Participant 2 . Primary Identifier	Publication PubMed ID	Data Sets Name
5468	PPARG	10133	20195357	IntAct interactions data set
5468	PPARG	1051	17785586	BioGRID interaction data set
5468	PPARG	1051	18310298	BioGRID interaction data set
5468	PPARG	1191	20195357	IntAct interactions data set
5468	PPARG	1191	20195357	IntAct interactions data set
5468	PPARG	1191	20195357	IntAct interactions data set
5468	PPARG	1956	26718225	BioGRID interaction data set

Gene (1 Gene)  
Gene > Interactions > Participant 2 (24 Genes)  
Gene > Interactions > Details (60 Interaction Details)  
Gene > Interactions > Details > Experiment > Interaction Detection Methods (10 Interaction Terms)  
Gene > Interactions > Details > Experiment > Publication (30 Publications)  
Gene > Interactions > Details > Data Sets (2 Data Sets)

Pick items from the table

CLU	physical	bait	prey	display technology
CLU	physical	bait	prey	display technology
CLU	physical	bait	prey	tandem affinity purification
EGFR	physical	bait	prey	affinity chromatography technology

[Create List](#) [Add to List](#)

## Exercise 7: Query Builder:

---

Using HumanMine: we will build a query to show Human genes and OMIM diseases, and then add a further constraint to show genes associated with all types of Diabetes.

1. Start your query from Gene
2. Constrain “Organism” to Homo Sapiens
3. Add the columns of data we want in our results:

Gene: Primary identifier and Symbol

Disease: name

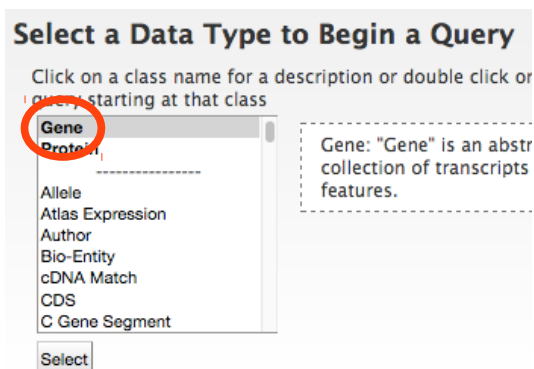
4. Run this search - ‘Show results’.
5. Return to the query (Use the “Trail” in the top left) and add a constraint to Disease name for “CONTAINS *Diabetes*”
6. Run the search and save the set of genes



# Exercise 7: Query Builder:

Use HumanMine for this exercise

## 1. Start your query from Gene

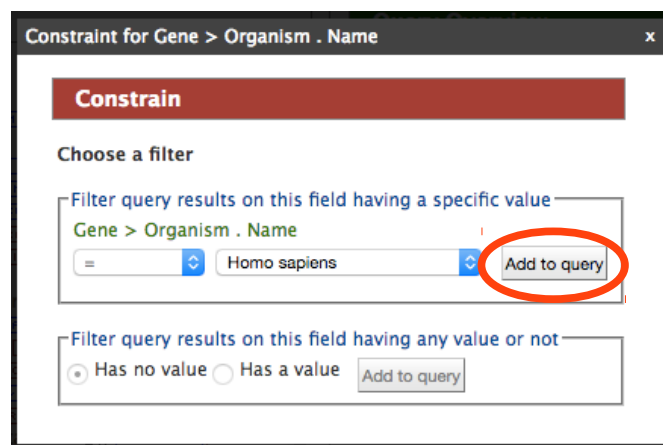


## 2. Constrain "Organism.name" to Homo Sapiens



Gene  
Organism Organism (X) (A)  
Name (X)  
= Homo sapiens (X) (A)

Constraint logic:  
one constraint





# Exercise 7: Query Builder:

## 3. Add the columns of data we want in our results:

Gene: Primary identifier and Symbol

Disease: name

Gene

- Brief Description
- Description
- Id Integer
- Length Integer
- Name
- Primary Identifier
- Secondary Identifier
- Symbol
- Alleles Allele
- Atlas Expression Atlas Expression
- Child Features Sequence Feature
- Chromosome Chromosome
- Chromosome Location Location
- Cross References Cross Reference
- Data Sets Data Set
- Diseases Disease
- Genes Gene
- Id Integer
- Identifier
- Name
- Alleles Allele
- Hpo Annotations HPO Annotation

Gene

- Primary Identifier
- Symbol
- Diseases Disease collection
- Name
- Organism Organism
- Name
- = Homo sapiens (A)

Constraint logic:

one constraint

## 4. Run this search - 'Show results'.

### Columns to Display

Use the [SHOW](#) or [SUMMARY](#) links to add fields to the results table. Click and drag the blue output boxes to choose the output column order. Click [A-Z](#) to choose a column to sort results by, click again to select ascending [A-Z](#) or descending [Z-A](#). Use the [REMOVE ALL](#) link to remove all fields from the results table.

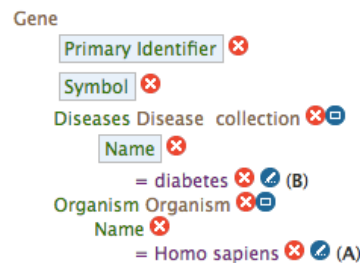
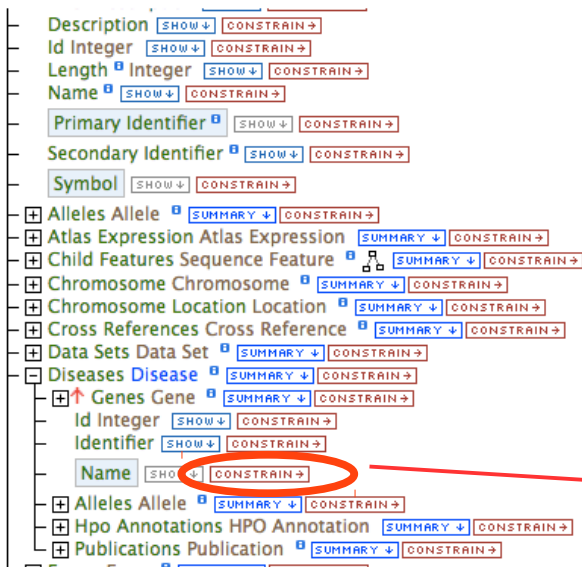
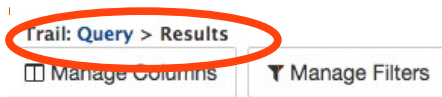
[REMOVE ALL](#)

Gene > Primary Identifier (no description)	Gene > Symbol (no description)	Gene > Diseases > Name (no description)
---	-----------------------------------	--

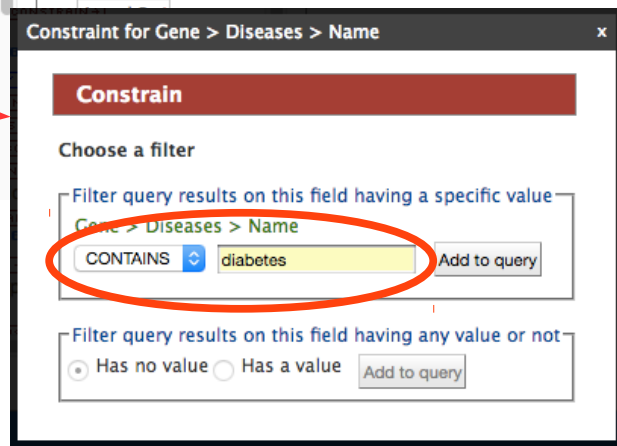
Show results

# Exercise 7: Query Builder:

5. Return to the query (Use the “Trail” in the top left) and add a constraint to Disease name for “CONTAINS *Diabetes*”



Constraint logic: A and B



NOTE: Make sure you change the Constraint from “=” to “CONTAINS”

6. Run the search and save the set of genes (68 genes)

Trail: Query > Results

Manage Columns Manage Filters Manage Relationships

Save as List Generate Python code Export

Showing 1 to 25 of 88 rows

Rows per page: 25

Gene Primary Identifier	Gene Symbol	Diseases Name
100188782	NIDDM4	DIABETES MELLITUS, NONINSULIN-DEPENDENT
1056	CEL	MATURITY-ONSET DIABETES OF THE YOUNG, TYPE 8, WITH EXOCRINE DYSFUNCTION
10644	IGF2BP2	DIABETES MELLITUS, NONINSULIN-DEPENDENT
11132	CAPN10	DIABETES MELLITUS, NONINSULIN-DEPENDENT, 1
1234	CCR5	DIABETES MELLITUS, INSULIN-DEPENDENT, 22

# Exercise 8: Analysis Workflows:

## Use HumanMine for this exercise

1. Identify the sets of genes you have created under the lists “view” tab.
2. Use the list set operations available on this page to intersect the list of diabetes genes you created with the query builder with your previous set of genes (genes expressed in adipose tissue that interact with pparg) created in exercise 6.


The screenshot shows the 'Lists' page in HumanMine. At the top, there's a 'Filter' bar and a 'Reset' button. Below that, the 'Actions' bar shows 'Union', 'Intersect', 'Subtract', and 'Asymmetric Difference' with radio buttons. 'Intersect' is selected. To the right, there are 'Copy', 'Delete', and 'Options' (with 'Show descriptions' checked). A message says 'You are not logged in. Log in to save lists permanently and to mark items as favourites'. The list of items includes 'Genes\_Diabetes 68 Genes', 'GenesExpressed\_AdiposeTissue\_InteractPPARG 24 Genes', 'GenesExpressed\_AdiposeTissue 1654 Genes', and 'PL\_DiabetesGenes 68 Genes'. A modal window titled 'Intersect' is open, showing 'Enter a new List name:' with 'ExpressedAdip' entered and a 'Save' button.

✓ Created list "ExpressedAdipose\_InteractPPARG\_Diabetes" as INTERSECT of Genes\_Diabetes and GenesExpressed\_AdiposeTissue\_InteractPPARG. [Hide](#)

The screenshot shows the 'Lists' page after the intersection operation. The 'Intersect' action is still selected. The list of items now includes 'ExpressedAdipose\_InteractPPARG\_Diabetes 1 Gene' at the top, followed by 'Genes\_Diabetes 68 Genes', 'GenesExpressed\_AdiposeTissue\_InteractPPARG 24 Genes', and 'GenesExpressed\_AdiposeTissue 1654 Genes'. The new list is circled in red.

# Exercise 8: Analysis Workflows:

- 3. We now want to know if any of these genes have been identified in GWAS studies. Run a template on your intersected list to find this out.
- 4. Use the column summary to find if any of the GWAS phenotypes are related to diabetes.

**Gene** → **GWAS hit**

For a given Gene (or List of Genes) returns a list of SNPs associated with the Gene in GWAS experiments. Results can be constrained by P-value [Format Examples: 0.05 or 2.0E-6; default = 0.01] or Phenotype [Example: \*diabetes\* or \*obesity\* or try typing your keyword to bring up the autocomplete options] [Keywords: genome wide association, study, studies, disease]

Gene > Associated Genes

LOOKUP:  For Organism:

☒ constrain to be IN

GWAS Result > P Value

optional

ON | OFF

GWAS Result > Phenotype

optional

ON | OFF

Show Results

Edit Query

web service URL

Perl | Python | Ruby | Java [help]

export XML

Trail: Query

**Gene** → **GWAS hit**

For a given Gene (or List of Genes) returns a list of SNPs associated with the Gene in GWAS experiments. Results can be constrained by P-value [Format Examples: 0.05 or 2.0E-6; default = 0.01] or Phenotype [Example: \*diabetes\* or \*obesity\* or try typing your keyword to bring up the autocomplete options] [Keywords: genome wide association, study, studies, disease]

Manage Columns

Manage Filters

Manage Relationships

Save as List

Generate Python code

Export

Showing 1 to 7 of 7 rows

Associated Genes Primary Identifier	Associated Genes Symbol	Associated Genes Name	Results SNP . Primary Identifier	Results P Value	Results Phenotype	GWAS Name	Publication PubMed ID	Organism Short Name
6925	TCF4	transcription factor 4	rs613872	1e-18	Fuch's corneal dystrophy	Fuch's corneal dystrophy	20825314	H. sapiens
6925	TCF4	transcription factor 4	rs4765905	1e-7	Schizophrenia	Bipolar disorder	22688191	H. sapiens
6925	TCF4	transcription factor 4	rs1970671	0.000003	Diabetic retinopathy	Diabetic retinopathy	21441570	H. sapiens
6925	TCF4	transcription factor 4	rs17594526	1e-7	Schizophrenia	Schizophrenia	19571811	H. sapiens
6925	TCF4	transcription factor 4	rs9960767	4e-9	Schizophrenia	Schizophrenia	19571808	H. sapiens
6925	TCF4	transcription factor 4	rs17512836	0.000001	Schizophrenia	Schizophrenia	21926974	H. sapiens
6925	TCF4	transcription factor 4	rs1631486	0.000004	Obesity (early onset extreme)	Obesity (early onset extreme)	23563609	H. sapiens

4 GWAS Result Phenotypes

Filter values

GWAS Result Phenotype	Count
Schizophrenia	4
Diabetic retinopathy	1
Fuch's corneal dystrophy	1
Obesity (early onset extreme)	1

Filter

Download data

# Exercise 9: Region Search:

Use FlyMine for this exercise

1. Select the example set of regions
2. De-select the features and re-select Genes and Regulatory regions
3. Extend the search by 5kb
4. Run the search

## Search for features within Genomic Regions

Search for features that overlap a list of genome coordinates you enter or upload, e.g. 2L:11334..12296

Genome coordinates help

1. Select Organism: D. melanogaster genome build: not available

2. **Select Feature Types:**

<input type="checkbox"/> cDNA Clone	<input type="checkbox"/> Exon	<input type="checkbox"/> Probe Set
<input type="checkbox"/> CDS	<input type="checkbox"/> Five Prime UTR	<input type="checkbox"/> rRNA
<input type="checkbox"/> CRM	<input checked="" type="checkbox"/> Gene	<input checked="" type="checkbox"/> Regulatory Region
<input type="checkbox"/> Chromosomal Deletion	<input type="checkbox"/> Intergenic Region	<input type="checkbox"/> snRNA
<input type="checkbox"/> Chromosomal Duplication	<input type="checkbox"/> Intron	<input type="checkbox"/> snoRNA
<input type="checkbox"/> Chromosomal Inversion	<input type="checkbox"/> mRNA	<input type="checkbox"/> TF Binding Site
<input type="checkbox"/> Chromosomal Translocation	<input type="checkbox"/> miRNA	<input type="checkbox"/> tRNA
<input type="checkbox"/> Chromosomal Transposition	<input type="checkbox"/> miRNA Target	<input type="checkbox"/> Three Prime UTR
<input type="checkbox"/> Chromosome Band	<input type="checkbox"/> Microarray Oligo	<input type="checkbox"/> Transcript
<input type="checkbox"/> Chromosome Structure Variation	<input type="checkbox"/> Natural Transposable Element	<input type="checkbox"/> Transposable Element
<input type="checkbox"/> EST	<input type="checkbox"/> ncRNA	<input type="checkbox"/> Transposable Element Insertion Site
<input type="checkbox"/> Enhancer	<input type="checkbox"/> Point Mutation	

3. Type/Paste in genomic regions in ☒ base coordinate ☐ interbase coordinate

(click to see an example)

2L:14615455..14619002  
2R:5866646..5868384  
3R:2578486..2580016

or Upload genomic regions from a .txt file...

Choose File No file chosen

4. Extend your regions at both sides:

5.00k 1k 10k 100k 1M 10M

Result Search

# Exercise 9: Region Search:

Examine the results and:

- 5. Create a list of all genes found.
- 6. Create a list of the regulatory regions found in the first genomic span.

Selected organism: *D. melanogaster*

Selected feature types: Gene, Regulatory Region

Extend Regions: 5 kbp

Export for all regions: TAB CSV GFF3 FASTA BED FASTA or Create List by feature type: Gene Go

2L:14610455..14624002

Original input: 2L:14615455..14619002

TAB CSV GFF3 FASTA BED FASTA GALAXY

Create List by Regulatory Region Go

GENOME REGION	FEATURE	FEATURE TYPE	LOCATION
	osp FBgn0003016	Gene	2L:14599196..14689340
	FBsf0000436668	Regulatory Region	2L:14610557..14614897
	Adh_ALE_enhancer	CRM	2L:14610557..14614897
	CR43411 FBgn0263330	Gene	2L:14613258..14613621
	Adh_pD-1320CAT	CRM	2L:14614238..14615611
	Adh_pD-663CAT	CRM	2L:14614895..14615611