

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 Pax6 only shown below):

- Pax6
- rs10509540
- *insulin*

Human Pax6 is the first gene result returned when HumanMine is searched for Pax6. Click on this gene to be taken to the report page.

Categories

Hits by Category

- Publication: 869
- Gene: 16
- Interaction Experiment: 8
- Protein: 8
- Exon: 4
- UniProt Feature: 2
- Mesh Term: 1
- Protein Domain: 1

Hits by Organism

- H. sapiens: 21
- M. musculus: 6
- R. norvegicus: 1

Search results 1 to 100 out of 909 for *pax6*

<< First < Previous | Next > Last >>

Type	Details	Score
Gene	PAX6 paired box 6 5080 ENSG00000007372 Length: 33170 FASTA... Organism . Name: Homo sapiens
Gene	Pax6 paired box 6 MGI:37490 - Length: Organism . Name: Mus musculus
Gene	Pax6 paired box 6 RGD:3258 - Length: Organism . Name: Rattus norvegicus
Gene	- - 105941491 - Length: Organism . Name: Homo sapiens
Protein	Q8VBZ1_MOUSE Q8VBZ1 Organism . Name: Mus musculus Length: 64 FASTA...
Gene	- - 106014249 - Length: Organism . Name: Homo sapiens
Protein	Q8VHH9_MOUSE Q8VHH9 Organism . Name: Mus musculus Length: 146 FASTA...
Protein	Q8VBX9_MOUSE Q8VBX9 Organism . Name: Mus musculus Length: 221 FASTA...

Exercise 1: Faceted Search:

2. Filter and create a list:

- Search for ***diabetes***
- Filter for publications
- Make a list of the publications

Search results 1 to 100 out of 34890 for ***diabetes***

<< First < Previous | Next > Last >>
3.225s

Categories

Hit by Category

- Publication: 30764
- GWAS Result: 272
- Gene: 674
- GWAS: 152
- Protein Domain: 145
- Author: 134
- Disease: 113
- UniProt Feature: 38
- Ontology Term Synonym: 29
- Interaction Experiment: 26
- ... and 8 more values »

Hit by Organism

- R. norvegicus: 537
- H. sapiens: 143
- M. musculus: 16

Type	Details	Score
Gene	Ad adult obesity and diabetes MGI:87915 - Length: Organism . Name: Mus musculus
Gene	Ms2 resistance to diabetes induction by MS2 MGI:97170 - Length: Organism . Name: Mus musculus
Gene	Rrad Ras-related associated with diabetes MGI:1930943 - Length: Organism . Name: Mus musculus
Gene	Asip agouti signaling protein RGD:2003 - Length: Organism . Name: Rattus norvegicus
Gene	Aanat aralkylamine N-acetyltransferase RGD:2006 - Length: Organism . Name: Rattus norvegicus
Gene	Acp5 acid phosphatase 5, tartrate resistant RGD:2022 - Length: Organism . Name: Rattus norvegicus

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

Search results 1 to 100 out of 30764 for ***diabetes***

Category restricted to Publication

<< First < Previous | Next > Last >>
2.239s

Categories

Category: Publication
« show all



Type	Details	Score
<input type="checkbox"/> Publication	7926745 First Author: Michaud E J Title: Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. Year: 1994 Journal: Genes Dev. Volume: 8 Pages: 1463-72
<input type="checkbox"/> Publication	1473152 First Author: Bultman S J Title: Molecular characterization of the mouse agouti locus. Year: 1992 Journal: Cell Volume: 71 Pages: 1195-204
	23903354 First Author: Shimomura Kenji	

Exercise2: Exploring a Gene:

You are interested in the Human *Pax6* gene and want to know the following things about it:

Use HumanMine for this exercise

1. On which chromosome is *Pax6* located?
2. Can I access the sequence for the *Pax6* gene?
3. With which diseases is *Pax6* associated?
4. In which tissues is *Pax6* most highly expressed?
5. Does the *Pax6* protein have any known isoforms?
6. Does the *Pax6* protein have known domains?
7. Is there a *Pax6* orthologue in *D. melanogaster*?
8. Does this orthologue interact with any other genes/proteins? Identify the interaction type (genetic/physical)
9. For the interaction with 4E-T, what was the original experiment and publication that determined this interaction

Exercise2: Exploring a Gene:

1. On which chromosome is *Pax6* located?
2. Can I access the sequence for the *Pax6* 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 : PAX6 *Homo sapiens*

Name: **paired box 6** Cytological Location: **11p13**

Brief Description: **paired box 6**

description: This gene encodes paired box protein Pax-6, one of many human homologs of the *Drosophila melanogaster* gene *prd*. In addition to a conserved paired box domain, a hallmark feature of this gene family, the encoded protein also contains a homeobox domain. Both domains are known to bind DNA and function as regulators of gene transcription. Activity of this protein is key in the development of neural tissues, particularly the eye. This gene is regulated by multiple enhancers located up to hundreds of kilobases distant from this locus. Mutations in this gene or in the enhancer regions can cause ocular disorders such as aniridia and Peter's anomaly. Use of alternate promoters and alternative splicing results in multiple transcript variants encoding different isoforms. Interestingly, inclusion of a particular alternate coding exon has been shown to increase the length of the paired box domain and alter its DNA binding specificity. Consequently, isoforms that carry the shorter paired box domain regulate a different set of genes compared to the isoforms carrying the longer paired box domain. [provided by RefSeq, Mar 2019]

synonyms: HGNC:8620, uc058abq.1, MGDA, D11S812E, NM_001604, FVH1, Show more

identifiers: 5080, ENSG00000007372, PAX6

Region: **gene** Length: 33170 FASTA...
Location: 11:31784792-31817961 reverse strand Cytol location: 11p13

11 Pathways Reactome, KEGG **8** Diseases OMIM **45** Mouse Alleles (MGI) mouse alleles **76** Gene Ontology

Tissue **14** Genes **20** Expression Disease **15** **8**

>5080
accctcttttcttatcattgacatttaaacctctggggcagggtcctcgcgtagaacgcggc
tgtcagatctgccacttccctgccgagcggcggtgagaagtgtgggaacggcgctgcc
aggctcacctgctccccgcctccgctcccaggtaaccgcccgggctccggccccggcc
cggtcggggcccgccgggctcctccgctgccagcgactgctgtcccaaatcaaaagccc
gcccgaagtggccccgggcttgatttttgcttttaaaaggagggcatacaaagatggaaag
cgagttactgagggagggataggaaagggggtggaggagggaacttgtcttgcgagtg
gctcttctgcataagtagcaaaatgttccactcctaagagtggaacttccagtcggccct
gagctgggagtagggggcgggagtgctgctgctgtctgctaaagccactcgcgacccg
gaaaaatgcaggaggtggggacgcactttgcatccagacctcctctgcactgcagttcac
gacatccacgcttgggaaagtccgtaccgcgcctggagcgcttaaaagacacctgccc
gggtcgggcgaggtgcagcagaagtttcccgcggttgcaagtgacagatggctggaccgc
aacaaagtctagagatggggttcgtttctcagaaagacgcggagtagcaaaagatgcggc
cgacagagctgggcagcgctaaagctcccagcgtgtgatttgagcttcaacttcggaaga
cctaataatagcgattctcactgagctagaacgcgggctccgggtactgcgggcgctgc
gctggctcctcggcggaagcgcgcgggcgccatgggagccgggcccgaagcccgga
aagagaagcggccaccacctgcggccttgaggggccagagccggggccgggaagtgtc
cccttagtgagtcagggaagggacgacccaaacttagactaacttgtggttcagccgggc
gaaagtcctgggagcccgccctaaaagcagctggagacgcccagccggcttaggcggg
gcacccaagaaccccgccgggagtttggggcgggcctctgggcccaggcccggttagccc
ccaaccccaactcccaactcgcgctcctgcgccccctttctaggtctcttgaggacactt
cggagctcagtcacctgtgacaggtgttgggacctccgcccagctccgagcgggtggcgcc
ccctgctcacctcacctggaaacggggacgagcgagtgagcagtgagcagtcgacgctt
acctgggcccgggatcctctcctccggggttccctcctcctggagctgggcccagcagc
tcgctgcttcgagtggaagtggaccttctcctccagtcataaatcaaaacccagccat
cctcgggctcctcctcattagagatgtttattggagatcgtgtttattcgggtgtcac
ggcgagaaaacgggtgacataattacctctgaccagagtcctcgcctccgcgccagggc
gagcgagagactcctctatagcttctcatcccatgacctgctgagcagagagcagc

Exercise2: Exploring a Gene:

3. With which diseases is Pax6 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).

11 Pathways
Reactome, KEGG

8 Diseases
OMIM

45 Mouse Alleles (MGI)
mouse alleles

76 Gene Ontology

Tissue
↑14↓20

Genes
Expression

Disease
↑15↓8



“Diseases” links to a table for OMIM data

8 Diseases

Trail: Gene

Manage Columns

Manage Filters

Manage Relationships

Save as List

Generate Python code

Export

Showing 1 to 8 of 8 rows

Diseases Primary Identifier	Diseases Name
OMIM:106210	ANIRIDIA 1
OMIM:120200	COLOBOMA, OCULAR, AUTOSOMAL DOMINANT
OMIM:120430	COLOBOMA OF OPTIC NERVE
OMIM:136520	FOVEAL HYPOPLASIA 1
OMIM:148190	KERATITIS, HEREDITARY
OMIM:165550	OPTIC NERVE HYPOPLASIA, BILATERAL
OMIM:194072	WILMS TUMOR, ANIRIDIA, GENITOURINARY ANOMALIES, AND MENTAL RETARDATION SYNDROME
OMIM:604229	ANTERIOR SEGMENT DYSGENESIS 5

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

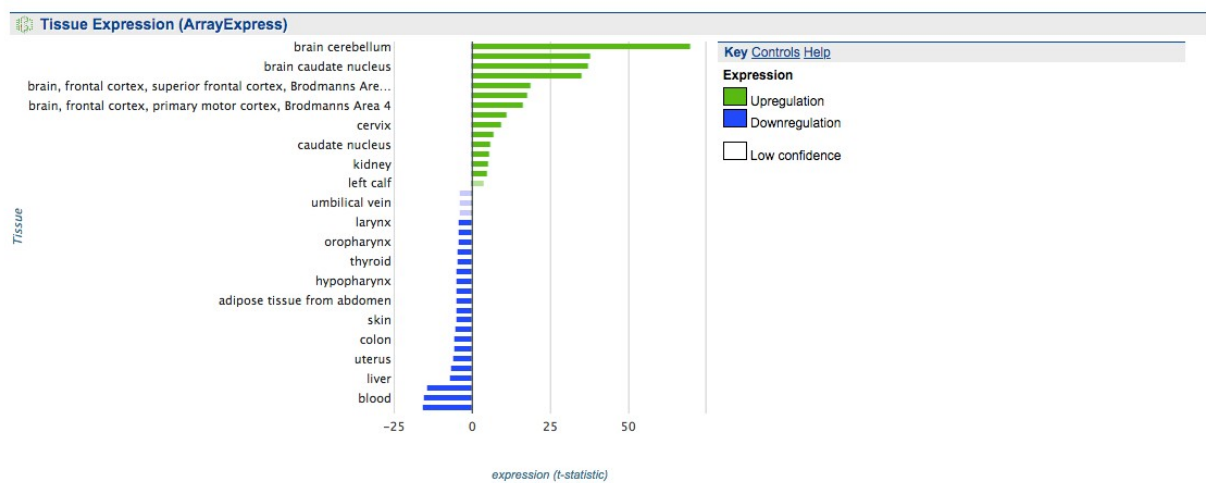
Curated comments from UniProt		Show proteins
Type	Comment	
developmental stage	Expressed in the developing eye and brain. Expression in the retina peaks at fetal days 51-60. At 6-week old, in the retina, is predominantly detected in the neural layer (at protein level). At 8- and 10-week old, in the retina, the expression is strongest in the inner and middle layer of the neural part (at protein level).	
disease	MIM:106210; Aniridia 1; AN1; A congenital, bilateral, panocular disorder characterized by complete absence of the iris or extreme iris hypoplasia. Aniridia is not just an isolated defect in iris development but it is associated with macular and optic nerve hypoplasia, cataract, corneal changes, nystagmus. Visual acuity is generally low but is unrelated to the degree of iris hypoplasia. Glaucoma is a secondary problem causing additional visual loss over time. The disease is caused by mutations affecting the gene represented in this entry.	
disease	MIM:120200; Coloboma, ocular, autosomal dominant; COAD; A set of malformations resulting from abnormal morphogenesis of the optic cup and stalk, and the fusion of the fetal fissure (optic fissure). The clinical presentation is variable. Some individuals may present with minimal defects in the anterior iris leaf without other ocular defects. More complex malformations create a combination of iris, uveoretinal and/or optic nerve defects without or with microphthalmia or even anophthalmia. The disease is caused by mutations affecting the gene represented in this entry.	
disease	MIM:120430; Coloboma of optic nerve; COLON; An ocular defect that is due to malclosure of the fetal intraocular fissure affecting the optic nerve head. In some affected individuals, it appears as enlargement of the physiologic cup with severely affected eyes showing huge cavities at the site of the disk. The disease is caused by mutations affecting the gene represented in this entry.	
disease	MIM:136520; Foveal hypoplasia 1; FVH1; An isolated form of foveal hypoplasia, a developmental defect of the eye defined as the lack of foveal depression with continuity of all neurosensory retinal layers in the presumed foveal area. Clinical features include absence of foveal pit on optical coherence tomography, absence of foveal hyperpigmentation, absence of foveal avascularity, absence of foveal and macular reflexes, decreased visual acuity, and nystagmus. Anterior segment anomalies and cataract are observed in some FVH1 patients. The disease is caused by mutations affecting the gene represented in this entry.	

Exercise2: Exploring a Gene:

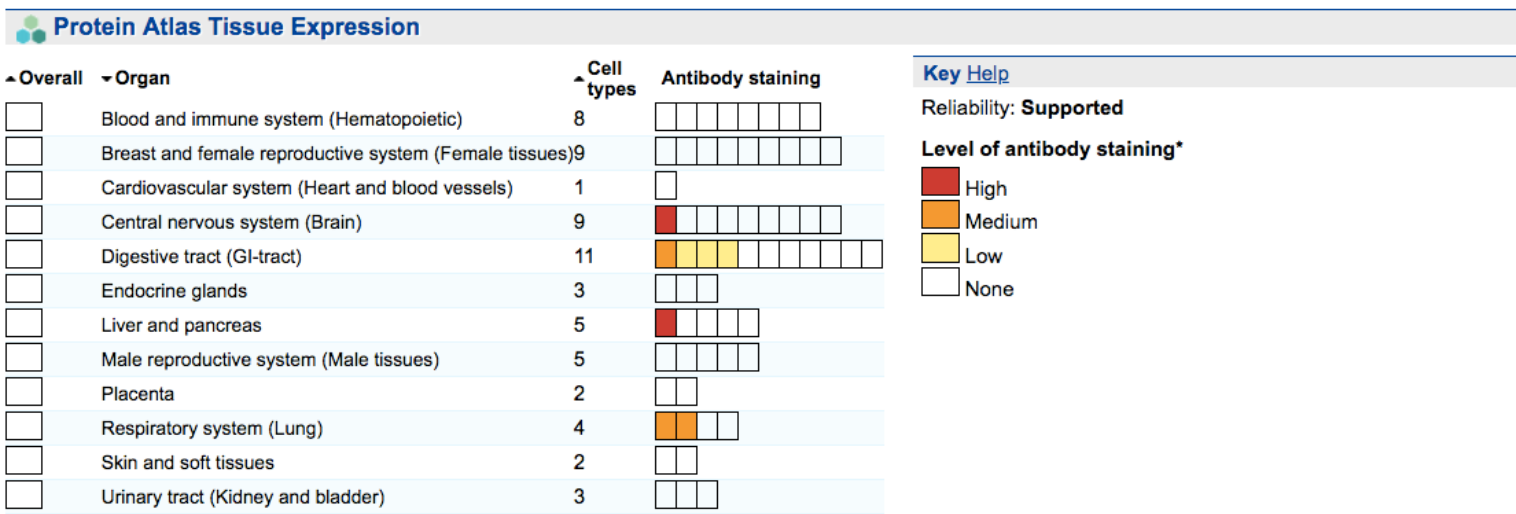
4. In which tissues is Pax6 most highly expressed?

Data on tissue expression can be found from three sources:

A. 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/>



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



C. Curated comments from Uniprot:

disease	MIM:617141; Aniridia 2; AN2; A form of aniridia, a congenital, bilateral, panocular disorder characterized by complete absence of the iris or extreme iris hypoplasia. Aniridia is not just an isolated defect in iris development but it is associated with macular and optic nerve hypoplasia, cataract, corneal changes, nystagmus. Visual acuity is generally low but is unrelated to the degree of iris hypoplasia. Glaucoma is a secondary problem causing additional visual loss over time. The gene represented in this entry is involved in disease pathogenesis. A mutation in a PAX6 long-range cis-regulatory element, known as SIMO, affects PAX6 expression in the developing eye and has pathological consequences. The mutation is located in ELP4 intron 9, 150 kb downstream of PAX6.
function	Transcription factor with important functions in the development of the eye, nose, central nervous system and pancreas. Required for the differentiation of pancreatic islet alpha cells (By similarity). Competes with PAX4 in binding to a common element in the glucagon, insulin and somatostatin promoters. Regulates specification of the ventral neuron subtypes by establishing the correct progenitor domains (By similarity). Isoform 5a appears to function as a molecular switch that specifies target genes.
similarity	Belongs to the paired homeobox family.
tissue specificity	Fetal eye, brain, spinal cord and olfactory epithelium. Isoform 5a is less abundant than the PAX6 shorter form.

Exercise2: Exploring a Gene:

6. Is there a *Pax6* orthologue in *D. melanogaster*?

Links to other Mines

FlyMine

D. melanogaster
ey¹, sy¹, toy¹

MouseMine

M. musculus
Pax6

YeastMine **No results**

RatMine

R. norvegicus
Pax6

ZebrafishMine

D. rerio
ZDB-GENE-041210-244,
ZDB-GENE-001031-1,
ZDB-GENE-990415-200

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

Note that there are three orthologous fly genes. For this exercise select the first (ey).

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

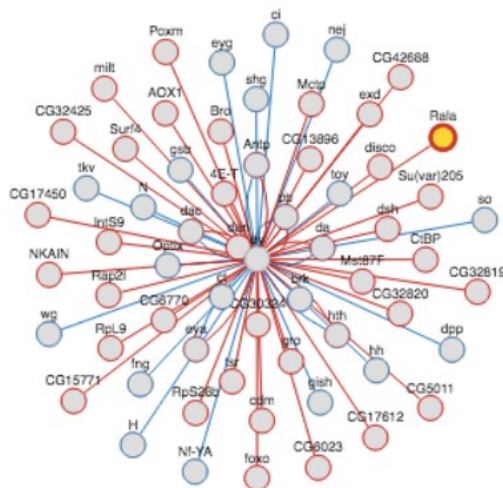
Use the “Interactions” quick link to navigate to protein and genetic Interaction data. Ey has both genetic and physical interactions with a number of genes.

Quick Links:

Summary Function Genes Homology Expression Regulation **Interactions** Other

Interactions

Interaction Network



Show the following interaction types:

Genetic

Physical

Reset view

Show in table format

Export graph

Gene Rala

Id: FBqn0015286

Symbol: Rala

Interaction Details:

Rala-Ey:

Type: Physical

Roles:

Bait: Rala

Prey: Ey

Data Set:

IntAct Molecular Interactions

Rala-Ey:

Type: Physical

Roles:

Bait: Rala

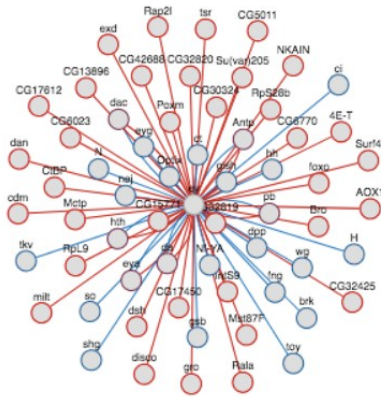
Prey: Ey

Data Set:

BioGRID Interaction Data Set

Exercise2: Exploring a Gene:

8. For the interaction with 4E-T, what was the original experiment and publication that determined this interaction



a. Select “Show in table format”

Show the following interaction types:

☐ Genetic ☒ All ☐ Physical

Reset view Export graph

b. Select the interaction name for the 4E-T interaction:

Rows per page: 25

Showing 1 to 25 of 153 rows

Gene DB identifier	Gene Symbol	Participant 2 Symbol	Interactions Participant 2 . Primary Identifier	ey-4E-T	Details Role 1	Details Role 2	Details Type	Data Sets Name
FBgn0005558	ey	4E-T	FBgn0052016	ey-4E-T	bait	prey	physical	BioGRID interaction data set
FBgn0005558	ey	AOX1	FBgn0267408	aox1-ey-1	prey	bait	physical	IntAct molecular interactions

c. Select the experiment name

1 Experiment

Showing rows 1 to 1 of 1

Experiment Description	Experiment Host Organism	Experiment Name
A protein interaction map of Drosophila melanogaster.	NO VALUE	Giot L (2003)

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
Giot L	A protein interaction map of Drosophila melanogaster.	2003	Science	302	1727-36	14605208

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.

You entered: 27 identifiers
We found: 25 Genes

Why are the numbers different? See below.

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					Action
	symbol	organism name	secondary identifier	primary identifier	class	
E2f	E2f1	Drosophila melanogaster	CG6376	FBgn0011766	Gene	Add
	E2f2	Drosophila melanogaster	CG1071	FBgn0024371	Gene	Add

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

Direct Hits					
Non-Gene Identifiers					
Synonyms					
Page 1 of 5					
5 rows per page					
Identifier you provided	Match				
	organism name	primary identifier	symbol	class	secondary identifier
FBgn0000000	Drosophila melanogaster	FBgn0000099		Gene	
FBgn0010433, ato	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 automatically 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

Two of the identifiers matched a **synonym** (rather than a current identifier). As the synonyms matched only one gene, these are automatically 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
FBgn0000099	Drosophila melanogaster	FBgn0267978	ap	Gene	CG8376

Exercise4: List Analysis Pages:

Use HumanMine for this exercise

Examine the HumanMine public list: PL_Pax6_Targets (319 genes)

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

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

Exercise4: List Analysis Pages:

1. What is the most enriched GO term for this list?
2. 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: 22

Test Correction: Holm-Bonferroni | Max p-value: 0.05 | Ontology: biological_process

Background population: Default | Change

View | Download

<input type="checkbox"/> GO Term	p-Value	Matches
<input type="checkbox"/> regulation of transcription by RNA polymerase II [GO:0006357]	1.109890e-50	150
<input type="checkbox"/> transcription by RNA polymerase II [GO:0006366]	1.205653e-50	154
<input type="checkbox"/> pattern specification process [GO:0007389]	1.799504e-43	71
<input type="checkbox"/> regionalization [GO:0003002]	2.577953e-41	63
<input type="checkbox"/> transcription, DNA-templated [GO:0006351]	7.562950e-40	172
<input type="checkbox"/> regulation of transcription, DNA-templated [GO:0006355]	2.243194e-39	168
<input type="checkbox"/> nucleic acid-templated transcription [GO:0097659]	4.338361e-39	172

150 genes in the list are annotated with the GO term “regulation of transcription by RNA”.
This is the most enriched GO term

You can click on this number to create
A sub-list of just these 150 genes

Exercise4: List Analysis Pages:

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

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

View homologues in other Mines:

- FlyMine
- D. melanogaster
- MouseMine**
- M. musculus
- RatMine
- R. norvegicus
- ZebrafishMine
- D. rerio

Could not retrieve results

List Analysis for link_1 (328 Genes)

Manage Columns Manage Filters Generate Python code Export Save as List

Manage Relationships

Showing 1 to 25 of 328 rows Rows per page: 25

Gene Primary Identifier	Gene Symbol	Gene Name	Gene Type	Gene Chromosome
MGI:101762	Elk3	ELK3, member of ETS oncogene family	protein_coding_gene	10
MGI:101876	Tead1	TEA domain family member 1	protein_coding_gene	7
MGI:101877	Tcf12	transcription factor 12	protein_coding_gene	9
MGI:101895	Pou3f2	POU domain, class 3, transcription factor 2	protein_coding_gene	4
MGI:101897	Pou2f2	POU domain, class 2, transcription factor 2	protein_coding_gene	7
MGI:101926	Dlx5	distal-less homeobox 5	protein_coding_gene	6
MGI:101927	Dlx6	distal-less homeobox 6	protein_coding_gene	6
MGI:102524	Pou4f2	POU domain, class 4, transcription factor 2	protein_coding_gene	8
MGI:102764	Six3	sine oculis-related homeobox 3	protein_coding_gene	17

5. Are these mouse genes enriched for any phenotypes?

Mammalian Phenotype Ontology Enrichment

MP terms enriched for items in this list

Number of Genes in this list not analysed in this widget: 60

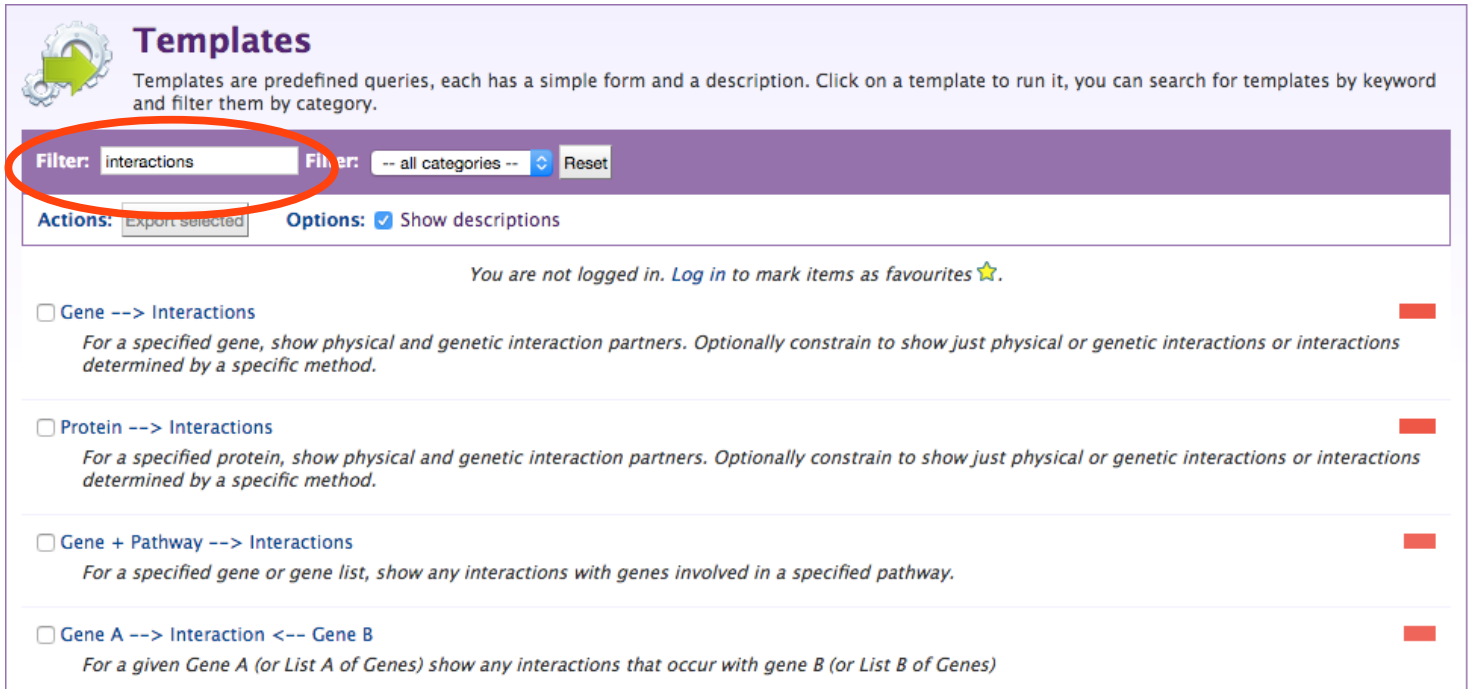
Test Correction: Holm-Bonferroni Max p-value: 0.05 Background population: Default

View Download

MP Term	p-Value	Matches
lethality during fetal growth through weaning [MP:0010832]	8.766171e-35	119
preweaning lethality [MP:0010770]	1.483056e-34	156
abnormal survival [MP:0010769]	1.718415e-30	168
mortality/aging [MP:0010768]	5.172224e-27	170
abnormal nervous system development [MP:0003861]	1.104814e-23	80
abnormal nervous system morphology [MP:0003632]	1.981481e-22	129
neonatal lethality, complete penetrance [MP:0011087]	2.561135e-22	52
perinatal lethality [MP:0002081]	1.758625e-21	73

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.

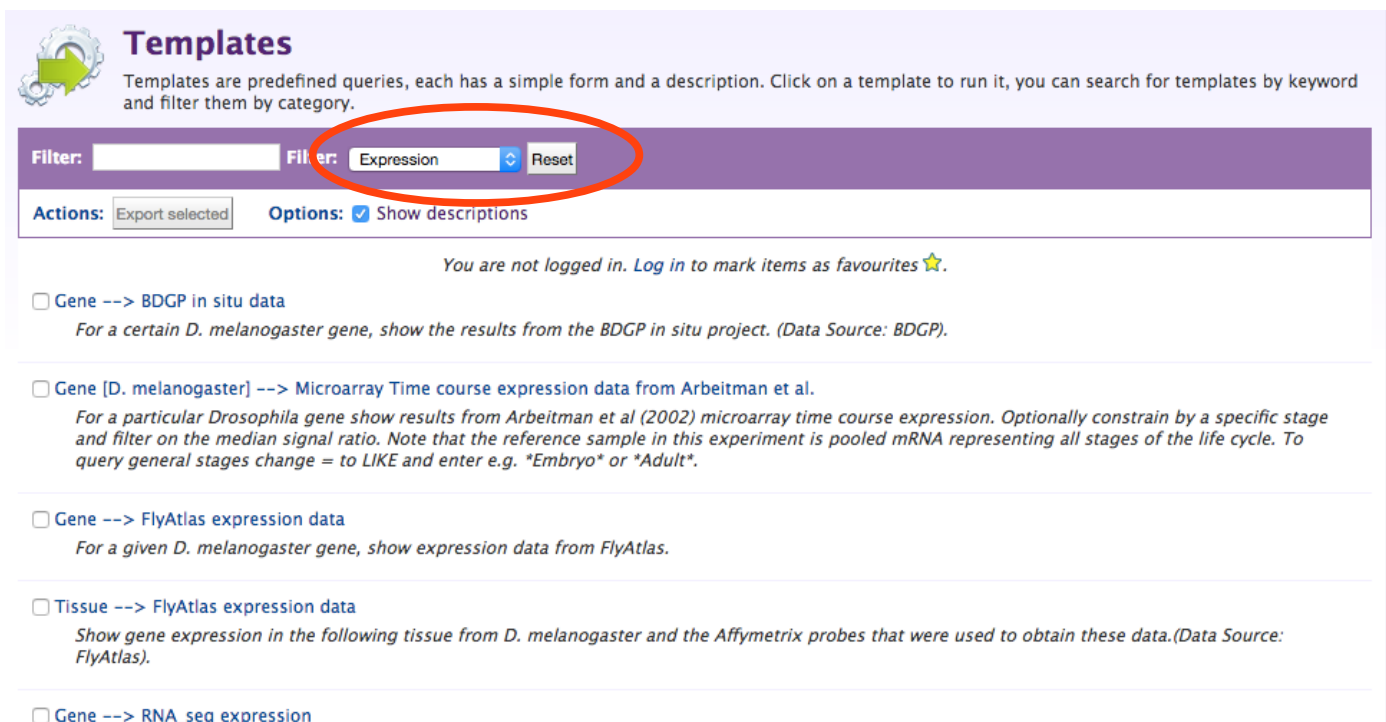
Filter: Filter:

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:

We will continue our exploration of the Pax6 gene in pancreatic tissue. Use a template search in HumanMine to answer the following question:

Are any of the known targets of Pax6 expressed in the pancreas. Use the public list of known Pax6 target genes in HumanMine (PL_Pax6_Targets).

Use the Protein Atlas dataset for the expression measurement


Save the list of target genes expressed in the pancreas as a list.

Exercise 6: Using template searches:

Use HumanMine for this exercise

Are any of the known targets of Pax6 expressed in the pancreas.

1. Find the following template and for Gene, select the PL_Pax6_Targets list


**Gene → Protein tissue Localisation** ☆
For a given human gene returns tissues in which the corresponding protein has been identified. Data Source: Protein Atlas

Gene
LOOKUP: for Organism:
☒ constrain to be saved Gene list

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

2. Filter the column “Tissue.name” for “Pancreas”

Gene Primary Identifier	Gene Symbol	Protein Atlas Expression Cell Type	Protein Atlas Expression Level	Tissue Name	Count
10097	ACTR2	Langerhans	Not detected		
10097	ACTR2	Leydig cells	Medium		
10097	ACTR2	Purkinje cells	Medium		
10097	ACTR2	adipocytes	Low		
10097	ACTR2	adipocytes	Not detected		
10097	ACTR2	bile duct cells	Medium		
10097	ACTR2	cells in endometrial stroma	Not detected		
10097	ACTR2	cells in endometrial stroma	Not detected		
10097	ACTR2	cells in glomeruli	Medium	Approved	Kidney
10097	ACTR2	cells in granular layer	Not detected	Approved	Cerebellum
10097	ACTR2	cells in molecular layer	Medium	Approved	Cerebellum
10097	ACTR2	cells in red pulp	Medium	Approved	Spleen
10097	ACTR2	cells in seminiferous ducts	Medium	Approved	Testis

54 Tissue Names

431 Items Selected

Tissue Name	Count
Pancreas	431

☒ Filter

Exercise 6: Using template searches:

3. Filter the Expression.level column for “Medium” and “High”.

Gene Primary Identifier	Gene Symbol	Protein Atlas Expression Cell Type	Protein Atlas Level	Protein Atlas Expression Tissue . Name
10097	ACTR2	exocrine glandular cells	Medium	Pancreas
10097	ACTR2	islets of Langerhans	Medium	Pancreas
10196	PRMT3	exocrine glandular cells	Medium	Pancreas
10196	PRMT3	islets of Langerhans	Medium	Pancreas
1021	CDK6	exocrine glandular cells	Low	Pancreas
1021	CDK6	islets of Langerhans	Not detected	Pancreas
10257	ABCC4	exocrine glandular cells	Not detected	Pancreas
10257	ABCC4	islets of Langerhans	Not detected	Pancreas
10320	IKZF1	exocrine glandular cells	Not detected	Pancreas
10320	IKZF1	islets of Langerhans	Not detected	Pancreas
1112	FOXN3	exocrine glandular cells	Not detected	Pancreas
1112	FOXN3	islets of Langerhans	Not detected	Pancreas

4 Protein Atlas Expression Levels

196 Items Selected

Filter values

Protein Atlas Expression Level	Count
<input type="checkbox"/> Not detected	156
<input checked="" type="checkbox"/> Medium	133
<input type="checkbox"/> Low	79
<input checked="" type="checkbox"/> High	63

Filter Download data

4. Save the genes from the resulting table as a list (129 genes)

Gene → Protein tissue Localisation

For a given human gene returns tissues in which the corresponding protein has been identified. Data Source: Protein Atlas

Manage Columns Manage Filters

Manage Relationships

Save as List Generate Python code Export

Showing rows 1 to 25 of 196

Gene Primary Identifier	Gene Symbol	Protein Atlas Cell Type	Protein Atlas Level	Protein Atlas Expression Tissue . Name
10097	ACTR2	exocrine glandular cells	Medium	Pancreas
10097	ACTR2	islets of Langerhans	Medium	Pancreas
10196	PRMT3	exocrine glandular cells	Medium	Pancreas
10196	PRMT3	islets of Langerhans	Medium	Pancreas
1121	CHM	exocrine glandular cells	Medium	Pancreas
1121	CHM	islets of Langerhans	Medium	Pancreas
11340	EXOSC8	exocrine glandular cells	High	Pancreas

Gene (129 Genes)

Gene > Protein Atlas Expression (196 Protein Atlas Expressions)

Gene > Protein Atlas Expression > Tissue (1 Tissue)

Pick items from the table

Create List Add to List

Exercise 7: Query Builder:

Using HumanMine: we will build a query to show Human genes and associated 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:

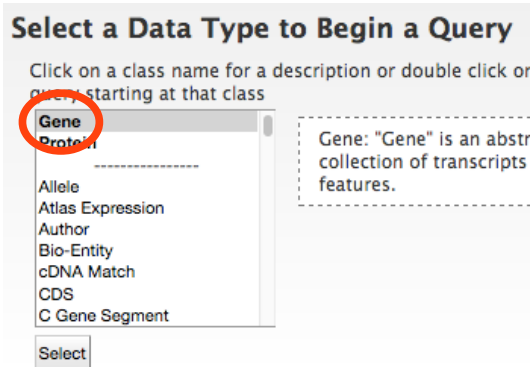
Gene: Primary identifier and Symbol
Disease: name

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

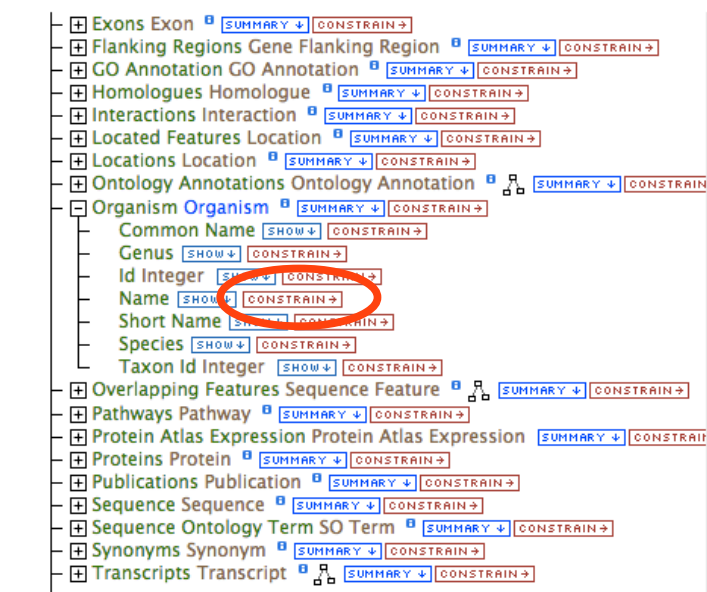
Exercise 7: Query Builder:

Use HumanMine for this exercise

1. Start your query from Gene

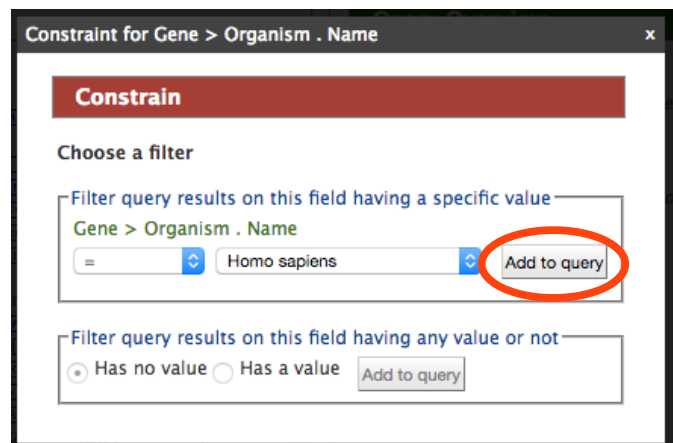


2. Constrain "Organism.name" to Homo sapiens



Gene
Organism Organism
Name
= Homo sapiens (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

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 [\[2\]](#) to choose a column to sort results by, click again to select ascending [\[2\]](#) or descending [\[2\]](#). Use the [REMOVE ALL](#) link to remove all fields from the results table.

[REMOVE ALL](#)

Gene > Primary Identifier (no description) [\[2\]](#)

Gene > Symbol (no description) [\[2\]](#)

Gene > Diseases > Name (no description) [\[2\]](#)

4. Run this search - 'Show results'.

Show results

Trail: [Query](#) > Results

[Manage Columns](#) [Manage Filters](#)

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

Showing rows 1 to 25 of 6,169 Rows per page: 25

Gene Primary Identifier	Gene Symbol	Diseases Name
10	NAT2	ACETYLATION, SLOW
100	ADA	SEVERE COMBINED IMMUNODEFICIENCY, AUTOSOMAL RECESSIVE, T CELL-NEGATIVE, B CELL-NEGATIVE, NK CELL-NEGATIVE, DUE TO ADENOSINE DEAMINASE DEFICIENCY
10000	AKT3	MEGALENCEPHALY-POLYMICROGYRIA-POLYDACTYL-HYDROCEPHALUS SYNDROME 2
10002	NR2E3	ENHANCED S-CONE SYNDROME
10002	NR2E3	RETINITIS PIGMENTOSA 37
100033413	SNORD116-1	PRADER-WILLI SYNDROME

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*”

Trail: **Query** > Results

☐ Manage Columns ☒ Manage Filters

Gene

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

Constraint logic: A and B

Constraint for Gene > Diseases > Name

Constrain

Choose a filter

Filter query results on this field having a specific value

Gene > Diseases > Name

CONTAINS diabetes Add to query

Filter query results on this field having any value or not

☒ Has no value ☐ Has a value Add to query

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

6. Run the search and save the set of genes (60 genes)

Trail: **Query** > Results

☐ Manage Columns ☒ Manage Filters ☒ Manage Relationships

Showing 1 to 25 of 60 rows

Rows per page: 25

Gene (60 Genes)

Gene > Organism (1 Organism)

Gene > Diseases (48 Diseases)

Pick items from the table

Create List Add to List

Gene Symbol	Gene Name	Primary Identifier	Organism . Name	
ABCC8	ATP binding cassette subfamily C member 8	0006071	Homo sapiens	
ACE	angiotensin I converting enzyme	1636	ENSG00000159640	Homo sapiens
AKT2	AKT serine/threonine kinase 2	208	ENSG00000105221	Homo sapiens
APPL1	adaptor protein, phosphotyrosine interacting with PH domain and leucine zipper 1	26060	ENSG00000157500	Homo sapiens
AQP2	aquaporin 2	359	ENSG00000167580	Homo sapiens
AVP	arginine vasopressin	551	ENSG00000101200	Homo sapiens
AVPR2	arginine vasopressin receptor 2	554	ENSG00000126895	Homo sapiens
BLK	BLK proto-oncogene, Src family tyrosine kinase	640	ENSG00000136573	Homo sapiens

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 (Pax6 target genes expressed in the Pancreas) created in exercise 6.

The screenshot shows the 'Lists' page in HumanMine. At the top, there's a 'Filter' section with a search bar and a 'Filter by a tag' dropdown. Below this, there's an 'Actions' bar with buttons for Union, Intersect, Subtract, and Asymmetric Difference. The 'Intersect' button is circled in red. To the right of the 'Intersect' button is a 'Delete' button. Below the 'Actions' bar, there's a list of gene sets. The first set is 'Diabetes_OMIM' with 60 Genes, and the second set is 'Pax6targets_ExpressedPancreas' with 129 Genes. Both sets have a checkbox next to them, and the 'Diabetes_OMIM' checkbox is circled in red. A modal dialog box is open over the 'Intersect' button, asking for a 'new List name' and showing 'Intersect_Pax' as the input. The 'Save' button is also circled in red.

Created list "Intersect_Pax6Diabetes" as INTERSECT of Diabetes_OMIM and Pax6targets_ExpressedPancreas.


Hide

The screenshot shows the 'Lists' page in HumanMine after the intersection operation. The 'Intersect' button is still circled in red. Below the 'Actions' bar, the list of gene sets now includes a new entry: 'Intersect_Pax6Diabetes' with 3 Genes. This entry has a checkbox next to it, which is circled in red. Below it are the original two sets: 'Diabetes_OMIM' with 60 Genes and 'Pax6targets_ExpressedPancreas' with 129 Genes. The 'Filter' section at the top now includes a 'Reset' button.

Three genes are found that are Pax6 targets expressed in the pancreas, and that are associated with diabetes. You can find out more about these genes by examining the list analysis page and their report pages.

Exercise 8: Analysis Workflows:

- 3. We now want to know if these genes have been identified in GWAS studies. Run a template on your intersected list.
- 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 saved Gene list

GWAS Result > P Value

optional ☐ ON | OFF

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

Showing 1 to 25 of 252 rows

Rows per page: page 1

Associated Genes Primary Identifier	Associated Genes Symbol	Associated Genes Name	Results SNP . Primary Identifier	Results P Value	Results Phenotype	GWAS First Author	GWAS Name	Publication PubMed ID	Organism Short Name
6934	TCF7L2	transcription factor 7 like 2	rs7903146	1e-18	Body mass index	Adeyemo AA	ZRANB3 is an African-specific type 2 diabetes locus associated with beta-cell mass and insulin response.	31324766	H. sapiens
6934	TCF7L2	transcription factor 7 like 2	rs7903146	9e-12	Body mass index	Bonas-Guarch S	Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes.	29358691	H. sapiens
3172	HNF4A	hepatocyte nuclear factor 4 alpha	rs1800961	6e-12	Neutrophil count	Chen J	Genome-wide association study of type 2 diabetes in Africa.	31049640	H. sapiens
3172	HNF4A	hepatocyte nuclear factor 4 alpha	rs1800961	9e-12	Sum neut counts	Cook JP	Multi-ethnic genome-wide association study identifies novel locus for type 2 diabetes susceptibility.	27189021	H. sapiens
3172	HNF4A	hepatocyte nuclear factor 4 alpha	rs1800961	1e-11	Sum baso counts	Ghassibe-Sabbagh M	T2DM GWAS in the Lebanese population confirms the role of TCF7L2 and CDKAL1 in disease susceptibility.	25483131	H. sapiens
3172	HNF4A	hepatocyte nuclear factor 4 alpha	rs1800961	2e-11	Granulocyte count				

98 GWAS Result Phenotypes

64 Items Selected

Filter values

GWAS Result Phenotype	Count
Type 2 diabetes	64
Body mass index	18
Breast cancer	7
Colorectal cancer	6

2 Gene Symbols

Filter values

Gene Symbol	Count
TCF7L2	50
HNF4A	9

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.
- 7. Notice that you can also export the sequences for your regions in fasta format.

Selected organism: *D. melanogaster*

Selected feature types: Gene, Regulatory Region

Extend Regions: 5 kbp

Hide

Export data for all features within all regions:

TAB CSV GFF3 BED FASTA

Export entire sequences for all regions:

FASTA

Create list by feature type:

Gene

Go

2L:14610455..14624002

Export sequence for entire region:

FASTA

Original input: 2L:14615455..14619002

TAB CSV GFF3 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_AdhF_3.2wt	CRM	2L:14614895..14618065
	FBsf0000435297	Regulatory Region	2L:14614942..14615084
	Adh_AAE_enhancer	CRM	2L:14614942..14615084
	TF000004	TF Binding Site	2L:14614981..14615006
	TF000003	TF Binding Site	2L:14615019..14615042
	TF000002	TF Binding Site	2L:14615019..14615042
	TF000001	TF Binding Site	2L:14615066..14615084
	TF000219	TF Binding Site	2L:14615221..14615228
	Adh_distal_promoter	CRM	2L:14615316..14615706