

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: 845
- 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 885 for *pax6*

<< First < Previous | Next > Last >>

0.386s

Type	Details	Score
Gene	PAX6 paired box 6 5080 ENSG00000007372 Length: 33170 FASTA... Organism . Name: Homo sapiens
Gene	Pax6 paired box 6 RGD:3258 - Length: - Organism . Name: Rattus norvegicus
Gene	Pax6 paired box 6 MGI:97490 - Length: - Organism . Name: Mus musculus
Gene	PAX6DRR PAX6 downstream regulatory region 105941491 - Length: - Organism . Name: Homo sapiens
Protein	Q8VBZ1_MOUSE Q8VBZ1 Organism . Name: Mus musculus Length: 64 FASTA...
Gene	LOC106014249 PAX6 upstream regulatory region 106014249 - Length: - Organism . Name: Homo sapiens
Protein	Q8VBX9_MOUSE Q8VBX9 Organism . Name: Mus musculus Length: 221 FASTA...
Protein	L8E7A6_HUMAN L8E7A6 Organism . Name: Homo sapiens Length: 65 FASTA...
Protein	Q8VHH9_MOUSE Q8VHH9 Organism . Name: Mus musculus Length: 146 FASTA...


Exercise1: 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 32255 for ***diabetes***

<< First < Previous | Next > Last >>
19.151s

 **Categories**

Hits by Category

- Publication: 29176
- GWAS Result: 1601
- Gene: 694
- Protein Domain: 144
- Author: 131
- GWAS: 131
- Disease: 108
- UniProt Feature: 36
- Ontology Term Synonym: 29
- Interaction Experiment: 23
- ... and 9 more values »

Hits by Organism

- R. norvegicus: 536
- H. sapiens: 164
- M. musculus: 26

Type	Details	Score
Gene	TNDM1 Diabetes mellitus, transient neonatal, 1 114253699 - Length: Organism . Name: Homo sapiens
Gene	IDDM3 insulin dependent diabetes mellitus 3 3402 - Length: Organism . Name: Homo sapiens
Gene	IDDMX Diabetes mellitus, insulin-dependent, X-linked, susceptibility to 8245 - Length: Organism . Name: Homo sapiens
Gene	IDDM4 insulin dependent diabetes mellitus 4 3403 - Length: Organism . Name: Homo sapiens
Gene	IDDM13 insulin dependent diabetes mellitus 13 3412 - Length: Organism . Name: Homo sapiens
Gene	IDDM8 insulin dependent diabetes mellitus 8 3407 - Length: Organism . Name: Homo sapiens
	IDDM23 Diabetes mellitus, insulin-dependent, 23 100271697 -	

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

Search results 1 to 100 out of 29176 for ***diabetes***

Category restricted to Publication ✕

<< First < Previous | Next > Last >>
7.399s

 **Categories**

Category: Publication
« show all

<input checked="" type="checkbox"/>	Type	Details	Score
<input checked="" type="checkbox"/>	Publication	19199708 First Author: Gonzalez-Begne Mireya Title: Proteomic analysis of human parotid gland exosomes by multidimensional protein identification technology (MudPIT). Year: 2009 Journal: J. Proteome Res. Volume: 8 Pages: 1304-14
<input checked="" type="checkbox"/>	Publication	20833797 First Author: Zhao Xiaolu Title: Phosphoproteome analysis of functional mitochondria isolated from resting human muscle reveals extensive phosphorylation of inner membrane protein complexes and enzymes. Year: 2011 Journal: Mol. Cell Proteomics Volume: 10 Pages: M110.000299
<input checked="" type="checkbox"/>	Publication	16099819 First Author: Volinic Jamie L Title: Overexpression of the coactivator bridge-1 results in insulin deficiency and diabetes. Year: 2006 Journal: Mol. Endocrinol. Volume: 20 Pages: 167-82

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 CDX2, 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	<p>This gene encodes paired box protein Pax-6, one of many human homologs of the <i>Drosophila melanogaster</i> gene <i>prd</i>. 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]</p>		
synonyms:	HGNC:8620, uc058abq.1, MGDA, D11S812E, NM_001604, FVH1, Show more		
identifiers:	5080, ENSG00000007372, PAX6		
Region:	gene	length:	33170
Location:	11:31784792-31817961 reverse strand	Chromosome location:	11p13

11 Pathways
Reactome, KEGG

8 Diseases
OMIM

45 Mouse Alleles (MGI)
mouse alleles

76 Gene Ontology

Tissue
↑14↓20

Genes
Expression

Disease
↑15↓8

>5080
accctcttttcttatcattgacatttaaactctggggcagggtcctcgcgtagaacgcggc
tgtcagatctgccacttccctgcgcgagcggcggtgagaagtgtgggaaccggcgctgcc
aggctcacctgcctccccgcctcgcctcccaggtaaccgccccgggtcccgccccggcc
cggtcggggcccgccggggcctcgcctgccagcgactgctgtccccaatcaaagccc
gccccaaagtggccccggggctgatttttgcttttaaaggaggcatacaaagtggaaag
cgagttactgagggaggataggaaggggggtggaggagggaacttgtcttgcgagtggt
gctcttctgcaaaagtagcaaaatgttccactcctaagagtggacttccagtcggccct
gagctgggagtagggggcgggagtgctgctgctgctgtctgctaaagccactcgcgaccgc
gaaaaatgcaggaggtggggacgcactttgcatccagacctcctctgcactcgcagttcac
gacatccacgcttgggaaagtccgtaccgcgcctggagcgcgttaaagacacctgcccgc
gggtcgggcgaggtgcagcagaagtttcccgcggttgcaaaagtgcatggctggaccgc
aacaaagtctagagatgggggttcgtttctcagaaaagacgcggagtagcaaaagatgcggc
cgacagagctgggcagcgcgtaaagctcccagcgtgtgatttagagcttccacttcggaaga
cctaataattagcgattctcactgagctagaacgcgggctccgggtactcggggcgctgc
gctggctgcctcggcggggaagcgcgcggcgccatgggagccgggcccgaagcccgga
aagagaagcggccaccacctgcggcccttgaggggccagagccggggcccggaagtgtc
cccttagtgagtcagggaagggacgacccaaacttagactaacttgtgttcagccgggc
gaaagtcctgggagcccgccctaaaagcagctggagacgcccagccggcttaggcggg
gcacccaagaaccccgccgggagtttggggcgccctctgggcccaggccccggctagccc
ccaacccccactcccactcgcgtcctcgcgccccctttctaggtctcttgggagacactt
cggagctcagtcacctgtgacaggtgttgggacctccgcccagctccgagcgggtggcgcc
ccctgctcacctcacctggaacggggacgagcgcagtggaagtcggcagtcggccgcacgctt
acctgggcgggatcctctcctccggggttcctctccctggagctgggcggcgagcagg
tcgctgcttcgcagtggaagtggaccttctcctccagtcataaatcaaaacccagccat
cctcgggcctcctcctcattagagatgtttattggagatcgtgtttattcggtgtcac
ggcgagaaaacgggtgacataattacctctgaccagagtcctcgtccgcgccagggc
gagcgaggagactcctctatcgacttctcctcgcctgacctgacctgctgagagagcgc

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	
<input type="checkbox"/> Manage Columns	<input type="checkbox"/> Manage Filters
<input type="checkbox"/> Manage Relationships	<input type="button" value="Save as List"/> <input type="button" value="Generate Python code"/> <input type="button" value="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”:

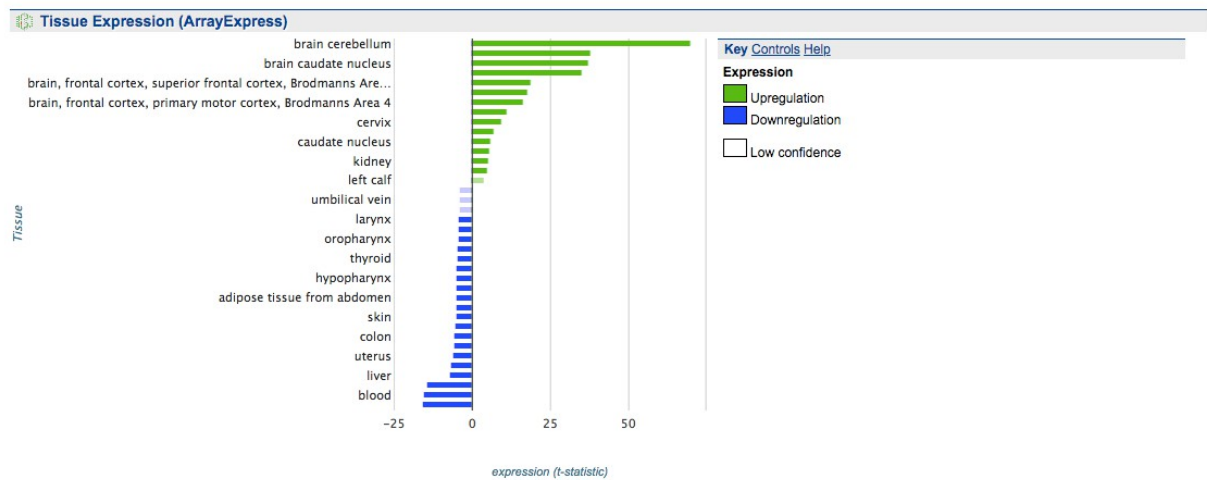
Curated comments from UniProt		Show proteins
Type	Comment	
developmental stage	Expressed in the developing eye and brain. Expression in the retina peaks at fetal days 51-60. At 6-week old, in the retina, is predominantly detected in the neural layer (at protein level). At 8- and 10-week old, in the retina, 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:

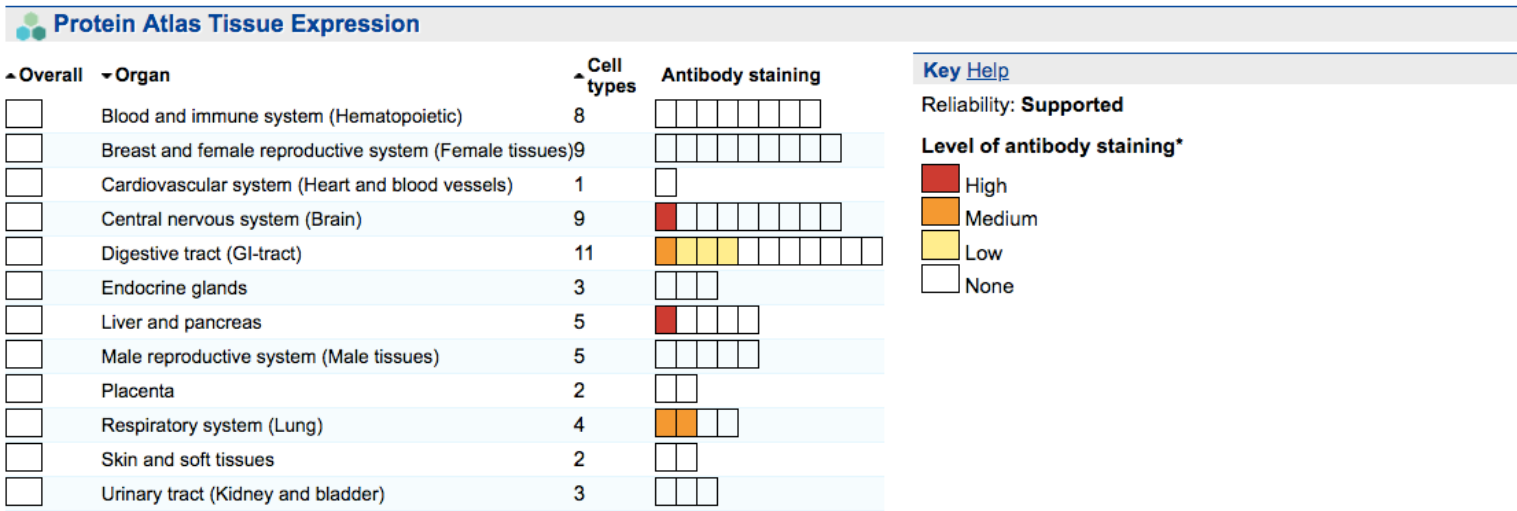
4. In which tissues is Pax6 most highly expressed?

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
ey, s, toy

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 othologous 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

Gene Rala

Id: FBgn0015286

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

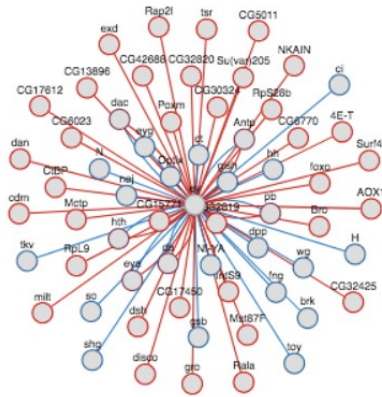
Show the following interaction types:

☒ Genetic ☐ All ☐ Physical

Reset view Show in table format Export graph

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

12345

5 rows per page

Identifier you provided	Match				
	organism name	primary identifier	symbol	class	secondary identifier
FBgn0000099	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 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: 19

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"/> transcription by RNA polymerase II [GO:0006366]	1.208028e-44	160
<input type="checkbox"/> regulation of transcription by RNA polymerase II [GO:0006357]	1.568312e-44	156
<input type="checkbox"/> regionalization [GO:0003002]	2.821877e-42	63
<input type="checkbox"/> pattern specification process [GO:0007389]	5.471655e-42	69
<input type="checkbox"/> transcription, DNA-templated [GO:0006351]	3.460283e-40	175
<input type="checkbox"/> regulation of transcription, DNA-templated [GO:0006355]	4.031986e-40	171
<input type="checkbox"/> positive regulation of transcription by RNA polymerase II [GO:0045944]	4.941925e-40	103

160 genes in the list are annotated with the GO term “transcription by RNA polymerase II”. This is the most enriched GO term

You can click on this number to create A sub-list of just these 160 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 Manage Relationships Generate Python code Export Save as List

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: 61

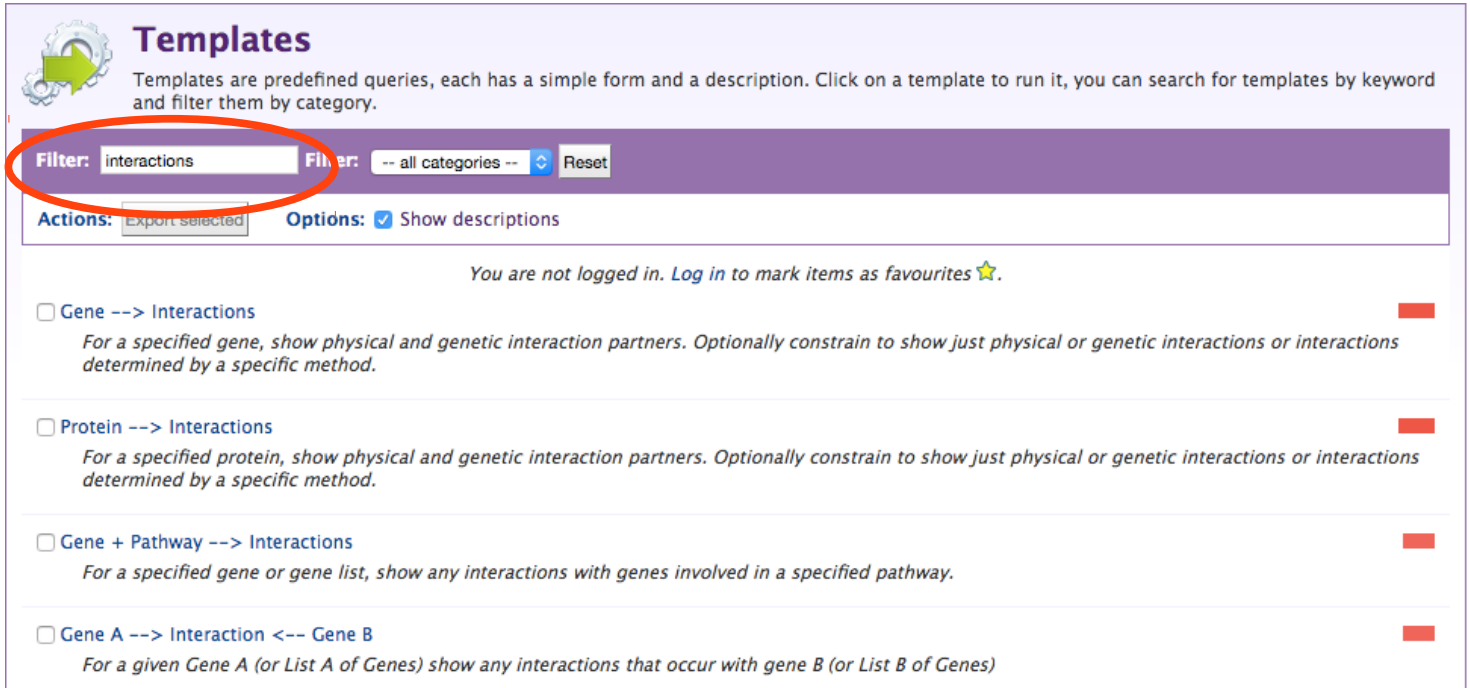
Test Correction: Holm-Bonferroni Max p-value: 0.05 Background population: Default Change

View Download

MP Term	p-Value	Matches
preweaning lethality [MP:0010770]	5.271869e-36	173
lethality during fetal growth through weaning [MP:0010832]	3.758746e-34	118
abnormal survival [MP:0010769]	2.649423e-32	184
mortality/aging [MP:0010768]	2.202941e-29	186
neonatal lethality, complete penetrance [MP:0011087]	4.339789e-22	52
perinatal lethality [MP:0002081]	3.659297e-21	73
neonatal lethality [MP:0002058]	6.531424e-21	61
abnormal nervous system morphology [MP:0003632]	1.572482e-20	123

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.

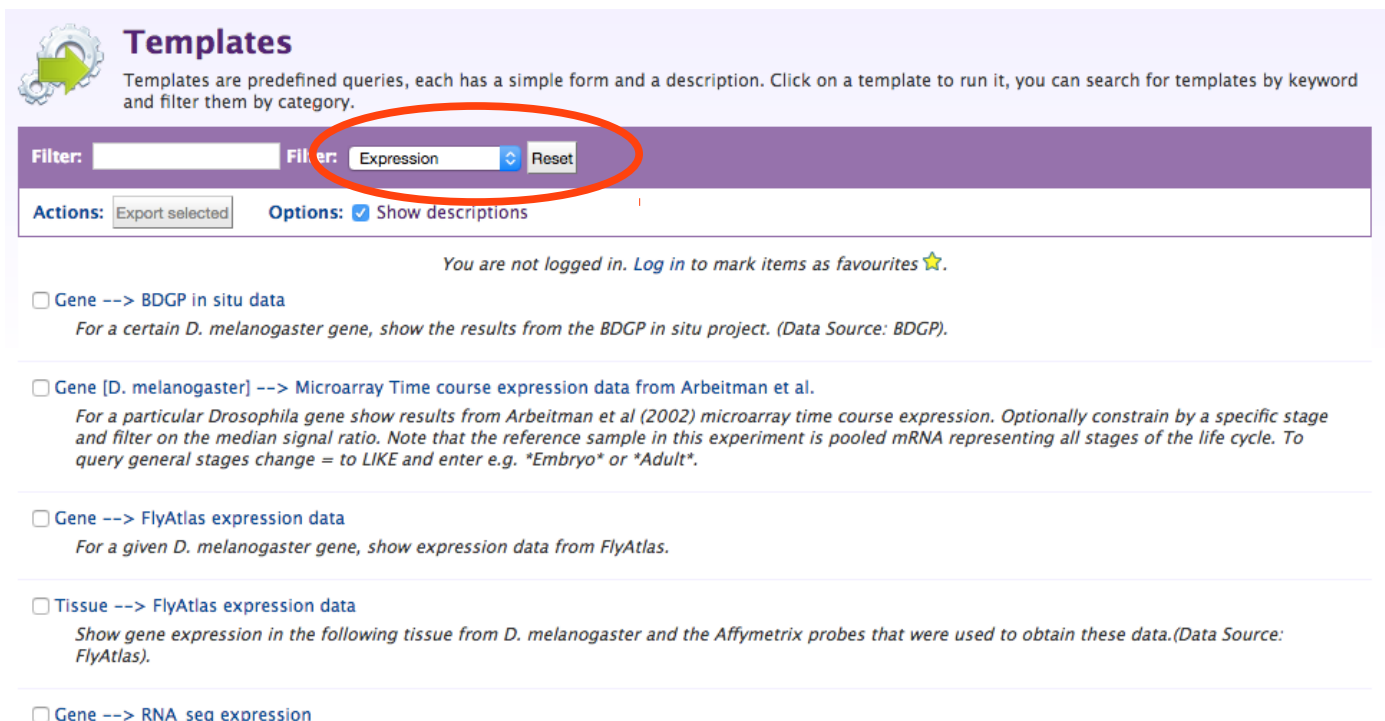
Filter: Filter:

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	Protein Atlas Expression Time
10097	ACTR2	Langerhans	Not detected	
10097	ACTR2	Leydig cells	Medium	
10097	ACTR2	Purkinje cells	Medium	
10097	ACTR2	adipocytes	Low	
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	

52 Tissue Names

422 Items Selected

pancreas

Tissue Name	Count
Pancreas	422

Filter

Download data

Exercise 6: Using template searches:

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

Showing 1 to 25 of 422 rows

Rows per page: 25

page 1

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

4 Protein Atlas Expression Levels

194 Items Selected

Filter values

Protein Atlas Expression Level	Count
<input type="checkbox"/> Not detected	148
<input checked="" type="checkbox"/> Medium	132
<input type="checkbox"/> Low	80
<input checked="" type="checkbox"/> High	62

Filter

Download data

4. Save the genes from the resulting table as a list (127 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

Undo

Save as List

Generate Python code

Export

Showing 1 to 25 of 194 rows

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
1121	CHM	exocrine glandular cells	Medium	Pancreas
1121	CHM	islets of Langerhans	Medium	Pancreas

Gene (127 Genes)

Gene > Protein Atlas Expression (194 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

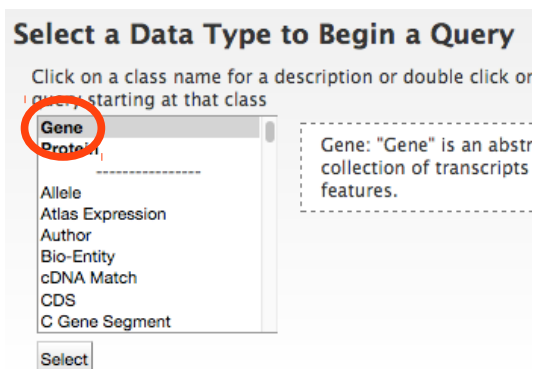
Disease: name

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

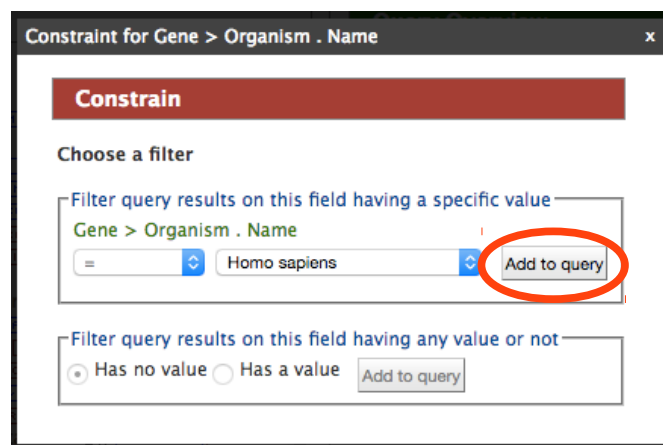


2. Constrain "Organism.name" to Homo sapiens



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

The screenshot shows the Query Builder interface with a tree view on the left. The 'Gene' section is expanded, showing fields like 'Primary Identifier' and 'Symbol'. The 'Diseases' section is also expanded, showing the 'Name' field. Red circles highlight the 'Primary Identifier', 'Symbol', and 'Name' fields.

The screenshot shows the Query Builder interface with a tree view on the left. The 'Gene' section is expanded, showing fields like 'Primary Identifier' and 'Symbol'. The 'Diseases' section is also expanded, showing the 'Name' field. Red circles highlight the 'Primary Identifier', 'Symbol', and 'Name' fields.

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](#)

The screenshot shows the 'Columns to Display' section with three blue output boxes. Each box contains the text 'Gene > Primary Identifier (no description)', 'Gene > Symbol (no description)', and 'Gene > Diseases > Name (no description)'. Each box has a blue output box and a red 'X' icon.

4. Run this search - 'Show results'.

Show results

The screenshot shows the 'Results' section of the Query Builder. The 'Trail: Query > Results' is displayed. The 'Manage Columns' button is active. The 'Rows per page' is set to 25. The results table shows 6,178 rows. The table has columns: Gene Primary Identifier, Gene Symbol, and Diseases Name. The first few rows are displayed.

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-POLYDACTYLY-HYDROCEPHALUS SYNDROME 2
10002	NR2E3	ENHANCED S-CONE SYNDROME
10002	NR2E3	RETINITIS PIGMENTOSA 37
100033413	SNORD116-1	PRADER-WILLI SYNDROME
10008	KCNE3	BRUGADA SYNDROME 6

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*”

The screenshot shows the InterMine Query Builder interface. At the top, the trail is "Query > Results". Below it are buttons for "Manage Columns" and "Manage Filters". The main area shows a tree of fields. The "Diseases" field is expanded, and the "Name" field is selected. A red arrow points from the "Name" field to a "Constraint for Gene > Diseases > Name" dialog box. The dialog box has a "Constrain" header and two sections: "Choose a filter" and "Filter query results on this field having a specific value". In the "Choose a filter" section, "CONTAINS" is selected, and "diabetes" is entered in the text box. The "Filter query results on this field having a specific value" section has "CONTAINS" selected, and "diabetes" is entered in the text box. The "Add to query" button is visible.

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

6. Run the search and save the set of genes (64 genes)

The screenshot shows the InterMine Query Builder interface after running the search. The trail is "Query > Results". Below it are buttons for "Manage Columns", "Manage Filters", and "Manage Relationships". The main area shows a table of results. The table has three columns: "Gene", "Gene Symbol", and "Diseases Name". The table shows 85 rows. The "Rows per page" dropdown is set to 25. A "Save as List" dialog box is open, showing the "Gene (64 Genes)" option selected. The dialog box also shows "Gene > Diseases (46 Diseases)" and "Gene > Organism (1 Organism)". The "Create List" button is visible.

Gene	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
1493	CTLA4	DIABETES MELLITUS, INSULIN-DEPENDENT, 12
1636	ACE	MICROVASCULAR COMPLICATIONS OF DIABETES, SUSCEPTIBILITY TO, 3

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 description: 'View your own and public lists, search by keyword and compare or combine the contents of lists. Click on a list to view graphs and summaries in an analysis page, select lists using checkboxes to perform set operations. Click 'Upload' above to import a new list.' Below this is a filter bar with a search box, a filter icon, a tag filter dropdown, and a 'Reset' button. The 'Actions' bar shows 'Union', 'Intersect' (circled in red), 'Subtract', and 'Asymmetric Difference'. There are also 'Copy' and 'Delete' icons, and 'Options' for 'Show descriptions' and 'Show Tags'. The list table shows two selected lists: 'Pax6TargetsExpressedPancreas' (127 Genes) and 'Diabetes_Omim' (64 Genes). A modal window titled 'Intersect' is open, showing 'Enter a new List name:' with 'IntersectDiabe' entered and a 'Save' button.

✓ Created list "IntersectDiabetes" as INTERSECT of Pax6TargetsExpressedPancreas and Diabetes_Omim.

Hide

The screenshot shows the 'Lists' page after the intersection operation. The 'IntersectDiabetes' list is now at the top of the list table, with 3 genes. Below it are the original lists: 'Pax6TargetsExpressedPancreas' (127 Genes) and 'Diabetes_Omim' (64 Genes). The 'IntersectDiabetes' list is circled in red.

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.

4. Use the column summary to find if any of the GWAS phenotypes are related to diabetes.

Associated Genes Primary Identifier	Associated Genes Symbol	Results	GWAS First Author	GWAS Name	Publication PubMed ID	Organism Short Name
6934	TCF7L2	2 Gene Symbols	Bonas-Guarch S	Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes.	29358691	H. sapiens
6934	TCF7L2	Gene Symbol Count	Cook JP	Multi-ethnic genome-wide association study identifies novel locus for type 2 diabetes susceptibility.	27189021	H. sapiens
6934	TCF7L2	TCF7L2 43	Ghassibe-Sabbagh M	T2DM GWAS in the Lebanese population confirms the role of TCF7L2 and CDKAL1 in disease susceptibility.	25483131	H. sapiens
6934	TCF7L2	HNF4A 3	Hara K	Genome-wide association study identifies three novel loci for type 2 diabetes.	23945395	H. sapiens
6934	TCF7L2	transcription factor 7 like 2	Imamura M	Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes.	26818947	H. sapiens

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: 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...

No file chosen

4. Extend your regions at both sides:

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 sequence 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