

# InterMine & MODs

Richard Smith  
richard@flymine.org



[www.intermine.org](http://www.intermine.org)

# InterMine

- data integration
- query-optimised data warehouse
- works with any data model
- flexible web interface
- web services

# Data Integration

*Existing data sources*



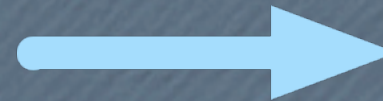
+

**Custom Data Sources**



*Java and Perl APIs*

*Configurable data integration*



**InterMine  
data warehouse**

# Web Interface

- **Template queries** - pre-defined queries
- **QueryBuilder** - custom queries, advanced
- **Report pages** - configurable, templates
- **Lists** - upload, use in queries, analysis widgets
- **Export**
- **Search**

**Gene → Proteins.**  
Show the proteins for a particular gene.

Gene:  ? for Organism:

constrain to be  list

CREATE LIST ADD TO LIST ADD COLUMN EXPORT PAGE SIZE 25 << FIRST < PREVIOUS | NEXT > LAST >>

Gene > primaryIdentifier ?	Gene > symbol	Protein > primaryIdentifier	Protein > primaryAccession
<input type="checkbox"/> FBgn0003380	Sh	<input type="checkbox"/> B6IDJ2_DROME	B6IDJ2
<input type="checkbox"/> FBgn0003380	Sh	<input type="checkbox"/> FBpp0088600	[no value]
<input type="checkbox"/> FBgn0003380	Sh	<input type="checkbox"/> FBpp0088601	[no value]
<input type="checkbox"/> FBgn0003380	Sh	<input type="checkbox"/> FBpp0088603	[no value]
<input type="checkbox"/> FBgn0003380	Sh	<input type="checkbox"/> FBpp0088605	[no value]
<input type="checkbox"/> FBgn0003380	Sh	<input type="checkbox"/> KCNAS_DROME	P08510
<input type="checkbox"/> FBgn0003380	Sh	<input type="checkbox"/> KCNAS_DROME-2	P08510-2

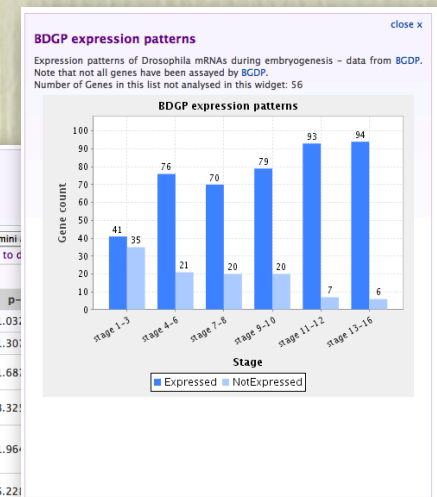
**Gene Ontology Enrichment**

GO terms enriched for items in this list.  
Number of Genes in this list not analysed in this widget: 2

Options  
Multiple Hypothesis Test Correction   
Ontology:  Maximum value to display:

[View in results table](#) [Download](#)

GO Term	P-value
<input type="checkbox"/> regulation of transcription [GO:0045449] <a href="#">ⓘ</a>	1.03E-163
<input type="checkbox"/> transcription [GO:0006350] <a href="#">ⓘ</a>	1.30E-163
<input type="checkbox"/> regulation of transcription, DNA-dependent [GO:0006355] <a href="#">ⓘ</a>	1.68E-163
<input type="checkbox"/> regulation of macromolecule biosynthetic process [GO:0010556] <a href="#">ⓘ</a>	3.32E-163
<input type="checkbox"/> regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process [GO:0019219] <a href="#">ⓘ</a>	1.96E-163
<input type="checkbox"/> regulation of nitrogen compound metabolic process [GO:0051171] <a href="#">ⓘ</a>	5.22E-163
<input type="checkbox"/> regulation of biosynthetic process [GO:0009889] <a href="#">ⓘ</a>	2.8382E-159 163
<input type="checkbox"/> regulation of cellular biosynthetic process [GO:0031326] <a href="#">ⓘ</a>	3.004E-159 163
<input type="checkbox"/> transcription, DNA-dependent	1.6476E-155 164



# New Features

- report page update
- region search
- better galaxy integration
- expanded web services
- new results tables (soon)

# InterMOD



Yeast



Worm



Mouse



Zebrafish



Rat



InterMine



FlyMine

**3 years funding, thanks NHGRI!**

# InterMOD



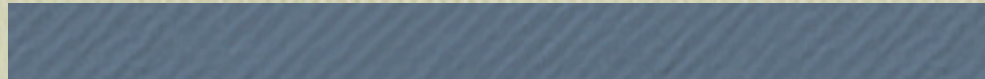
Julie Sullivan



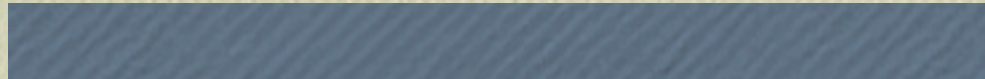
Alex Kalderimis

# Progress

RGD



SGD



ZFIN



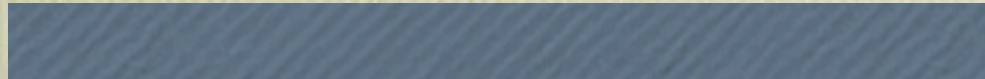
MGI



WormBase



FlyMine



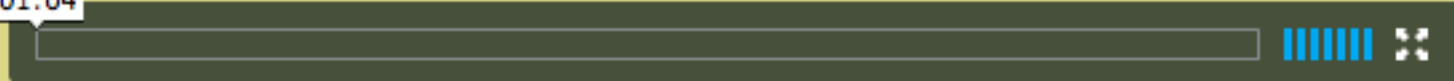


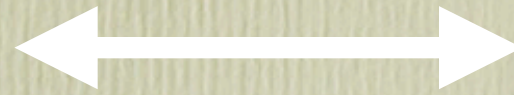
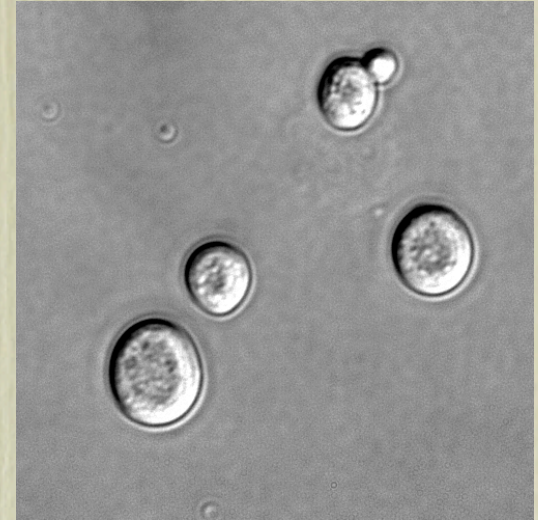


YeastMine is *Awesome!*

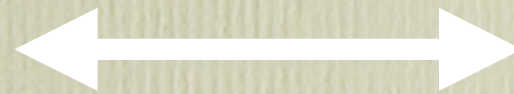


01:04





**Orthologues**  
Protein domains  
Pathways  
Gene Ontology



Demo

General Array IDs

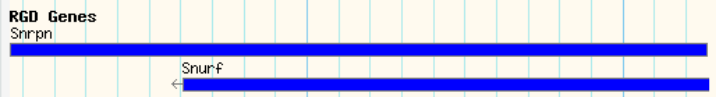
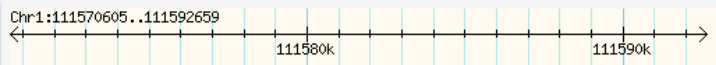
Gene: **Snrpn** (small nuclear ribonucleoprotein polypeptide N) *Rattus norvegicus*

**Symbol:** Snrpn  
**Name:** small nuclear ribonucleoprotein polypeptide N  
**Description:** EXHIBITS protein binding (orthology); INVOLVED IN response to hormone stimulus; ASSOCIATED WITH Prader-Willi Syndrome (ortholog); FOUND IN U1 snRNP; U2 snRNP  
**Type:** protein-coding  
**Status:** PROVISIONAL  
**Also known as:** FE 294; MGC105406; FE 294 psi; FE294; FE294 psi; small nuclear ribonucleoprotein polypeptide N; sm-N; sm-D; sm-N; snRNP-N; sm protein N; FE294 psi pseudogene; FE294 gene for snRNP-associated polypeptide N; small nuclear ribonucleoprotein-associated protein N; small nuclear ribonucleoprotein-associated protein (snRNP) mRNA clone Sm51; sm protein D; small nuclear ribonucleoprotein-associated protein (snRNP) mRNA, clone Sm51; small nuclear ribonucleoprotein N;  
**Orthologs:** Homo sapiens : [SNRPN \(small nuclear ribonucleoprotein polypeptide N\)](#)  
Mus musculus : [Snrpn \(small nuclear ribonucleoprotein N\)](#)

**Latest Assembly:** RGSC Genome Assembly v3.4

**Position:**

Map	Chr	Position	Strand	Source
Rat Cytogenetic Map	1	q22		RGD
<b>RGSC Genome Assembly v3.4</b>	1	111570605 - 111592659	-	RGD
Rat Celera Assembly	1	104690957 - 104713013	-	RGD
genome assembly 3.1	1	111648947 - 111671002	-	RGD



EntrezGene Gene Features



Contact Us **Lists**

**Search**

Search RatMine by entering names of genes, proteins, pathways, papers, etc.

e.g Snrpn, Blood Pressure, Cancer

**SEARCH**

**Analyse**

Enter a list of identifiers.

ENSRNOG00000010349, Snrpn, 25723, 1310526

**ANALYSE**

GENES PROTEINS QTLs DISEASE GENE ONTOLOGY PATHWAYS **enrichment widgets**

Read more

- Query for disease:
- Gene → Disease Ontology
  - Disease Term [Identifier and children of this term] → Genes
  - Disease Term [Chromosome and children of this term] → Genes
  - Disease Term and Chromosome → Genes
  - Chromosome and Disease [and child terms] → Genes

» More queries

**more data**



**General** | Array IDs

**Gene:** Snrpn (small nuclear ribonucleoprotein polypeptide N) Rattus norvegicus

**Symbol:** Snrpn

**Name:** E

**Description:** A

**Type:** protein coding

**Status:** PROVISIONAL

**Also known as:** FE 294; MGC105406; FE 294 psi; FE294; FE294 psi; small nuclear ribonucleoprotein polypeptide N; sm-N; sm-D; sm-N; snRNP-N; sm protein N; FE294 psi pseudogene; FE294 gene for snRNP-associated polypeptide N; small nuclear ribonucleoprotein-associated pr

**Orthologs:** H

**Latest Assembly:** RGSC Genome Assembly v3.4

**Position:**

Map	Chr	Position	Strand	Source
Rat Cytogenetic Map	1	q22		RGD

Chr 1  
111580k 111590k

**RGD Genes**  
Snrpn

**Model**  
← Snurf

EntrezGene Gene Features

**RatMine** v 2.2011-09-26 Powered by InterMine 0.96

Home | Templates | Lists | QueryBuilder | Data | API | MyMine

Contact Us

**Search**

Search RatMine by entering names of genes, proteins, pathways, papers, etc.

e.g Snrpn, Blood Pressure, Cancer

**SEARCH**

**Analyse**

Enter a list of identifiers.

ENSRNOG00000010349, Snrpn, 25723, 1310526

**ANALYSE**

GENES | PROTEINS | QTLs | DISEASE | GENE ONTOLOGY | PATHWAYS

Read more

Query for disease:

- Gene -> Disease Ontology
- Disease Term Identifier and children of this term -> Genes
- Disease Term [and children of this term] -> Genes
- Disease Term and Chromosome -> Genes
- Chromosome and Disease [and child terms] -> Genes

» More queries



*Template query embed links....*

# The plan

- Delete existing webapp
  - Java/JSP - old and slow to develop
- All functionality as web services
- Key components as JavaScript
- Lightweight new webapp

# Web Services

```
PathQuery q = new PathQuery(model);  
  
q.setView("Protein.primaryIdentifier,  
         Protein.genes.primaryIdentifier");  
  
q.addConstraint("Protein.proteinDomains.name",  
               Constraints.eq("Homeobox"));  
q.addConstraint("Protein.organism.shortName",  
               Constraints.eq("D. melanogaster"));
```

Perl, Python, Java, Ruby APIs

XML & JSON



# Result formats

JSON  
(objects)

JSON  
(rows)

JSONP

FASTA

TSV

GFF<sub>3</sub>

BED

CSV

- MODs maintain data
- regular releases (weekly)
- similar data models
- powerful web services



# mash ups



# Crossmodel.organism *...query model organism data*



# Crossmodel.organism *...query model organism data*

1. users can query across model organisms
2. InterMine API demos for developers

Demo

# Caching

1. Query results are cached in the InterMine server

2. Each Mine has a release number:

`yeastmine.org/service/release`

if the number is the same no data has changed.

Easy to add local caches and check version numbers.



# Why is this useful

- Data integration is hard
  - resolution of ids, old ids, update schedules
- Problems with stale data
- Small databases can present more data without local storage

# Conclusion

- MODs maintain data
- InterMine provides query software
  - focus on performance
- Bioinformaticians build webapps!
  - easy to attach to multiple MODs

# InterMine Team

<b>Biologists</b>	Rachel Lyne, Adrian Carr, Mike Lyne
<b>Developers</b>	Sergio Contrino, Julie Sullivan, Alex Kalderimis, Fengyuan Hu, Daniela Butano, Radek Stepan, Richard Smith
<b>Sys Admin</b>	Dan Tomlinson
<b>PI</b>	<b>Gos Micklem</b>

## MODs

<b>RGD</b>	<b>Andrew Vallejos</b> , Elizabeth Worthey, (Simon Twigger)
<b>SGD</b>	<b>Kalpana Karra</b> , Ben Hitz, Gail Binkley, Mike Cherry
<b>ZFIN</b>	<b>Sierra Taylor</b> , Judy Blake, Nathan Dunn, Monte Westerfield
<b>MGI</b>	(Michael O'Keefe), Joel Richardson
<b>WormBase</b>	Todd Harris, Lincoln Stein

**[www.intermine.org](http://www.intermine.org)**  
**[richard@flymine.org](mailto:richard@flymine.org)**