

InterMine

Collaborative Data Mining



InterMine

Collaborative Data Mining

Who/what is InterMine?

What's new?

- Better client library
- Region Search
- Collaboration tools

What's next?

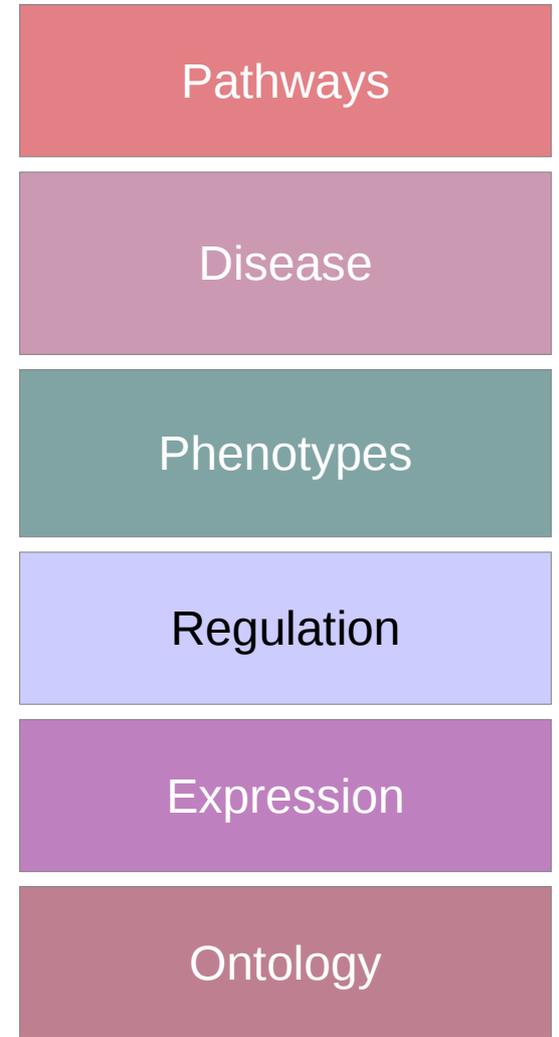
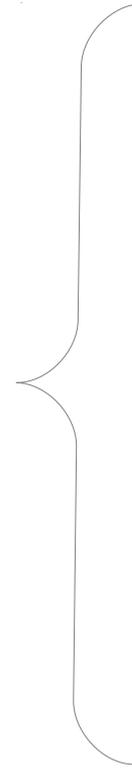
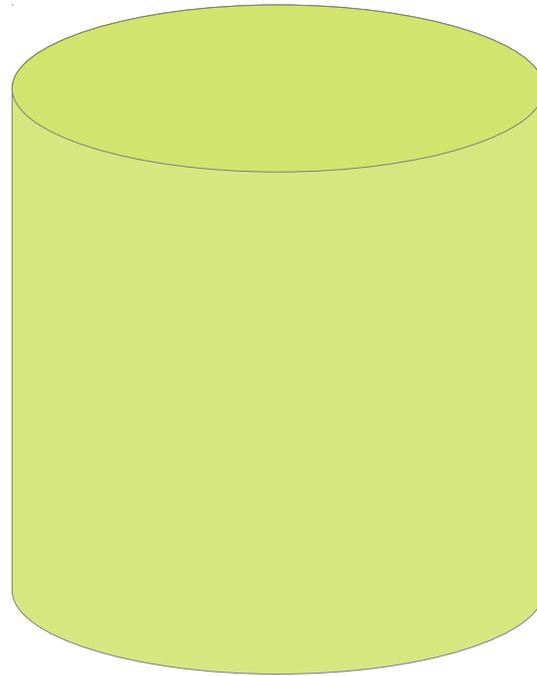
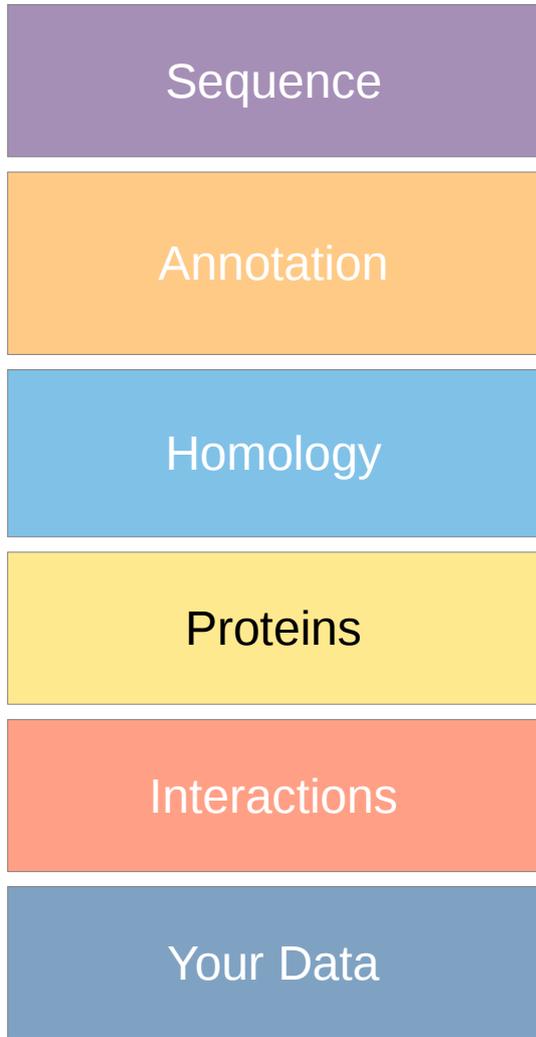


What is InterMine?

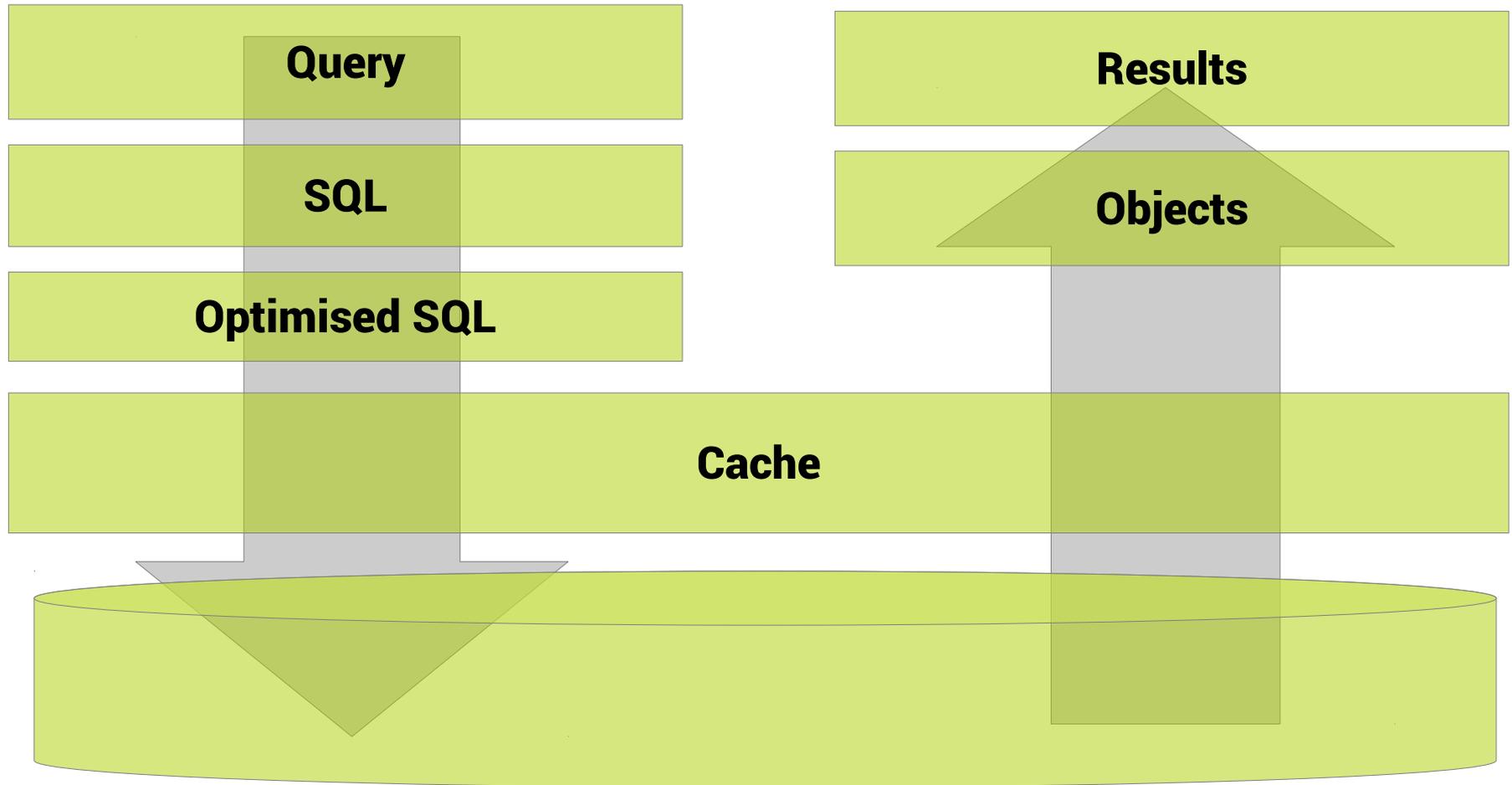
- A data-integration system
- An optimising query engine
- A set of data analysis tools
- A restful API
- Graphical web-app components



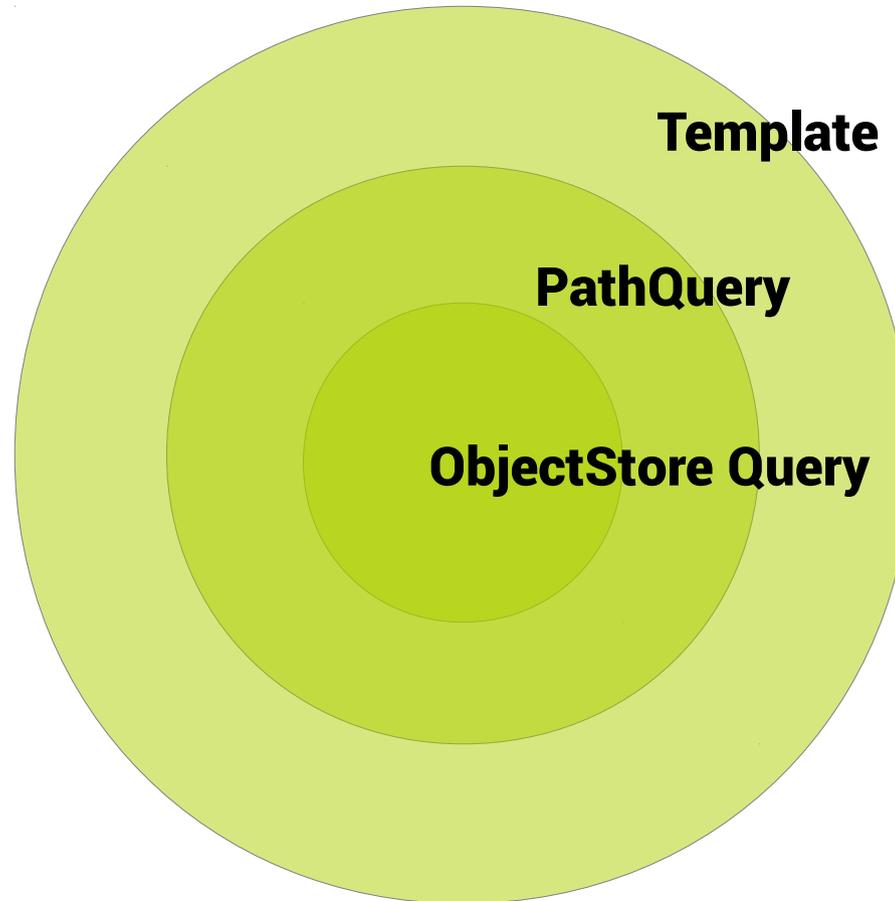
Data Integration



Query Engine



Query Abstraction



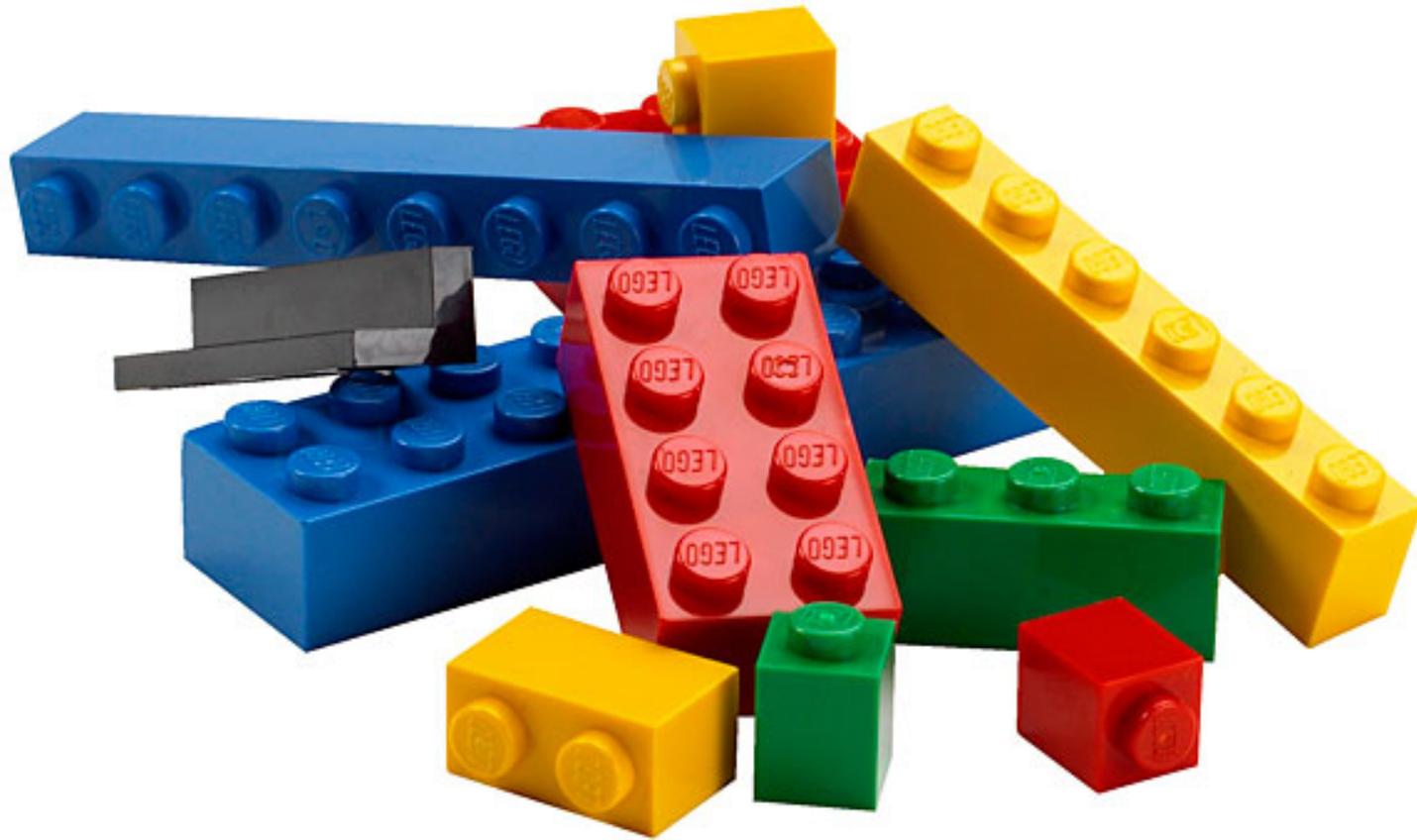
Why use InterMine?



*** may contain LGPL**

 **InterMine**

Why use InterMine?



 **InterMine**

Who is/uses InterMine?

- Developed at University of Cambridge
- Used by groups that curate/integrate data
 - SGD
 - RGD
 - MGI
 - Wormbase
 - FlyMine
 - modEncode
 - ZFIN
 - GO
 - NBIO
 - MRC
 - OICR
 - ...



{InterMOD

SGD *Saccharomyces*
GENOME DATABASE



 **InterMine**

InterMOD

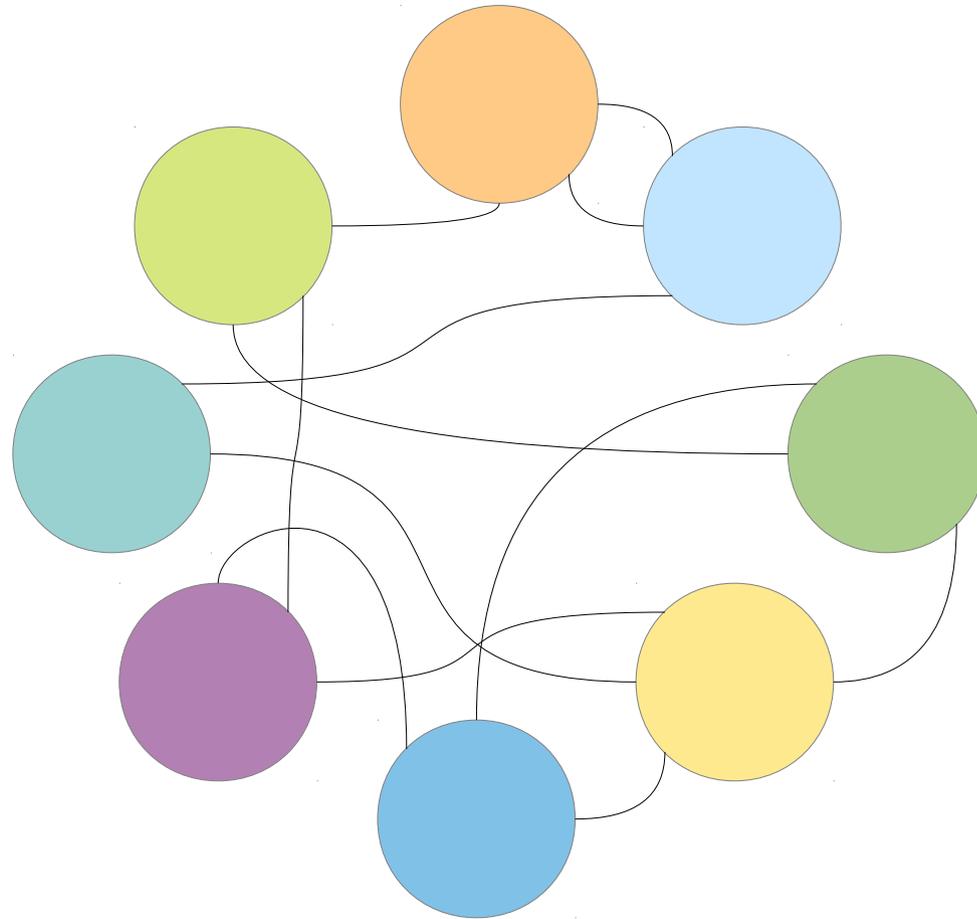
Homology

Pathways

Synteny

Phenotypes

...



What's New? - imjs

- Dog-fooding our javascript.
- Works in **node.js** and the **browser**
- Promise based
- Comprehensive Test suite
- Documentation

```
> npm install imjs
```



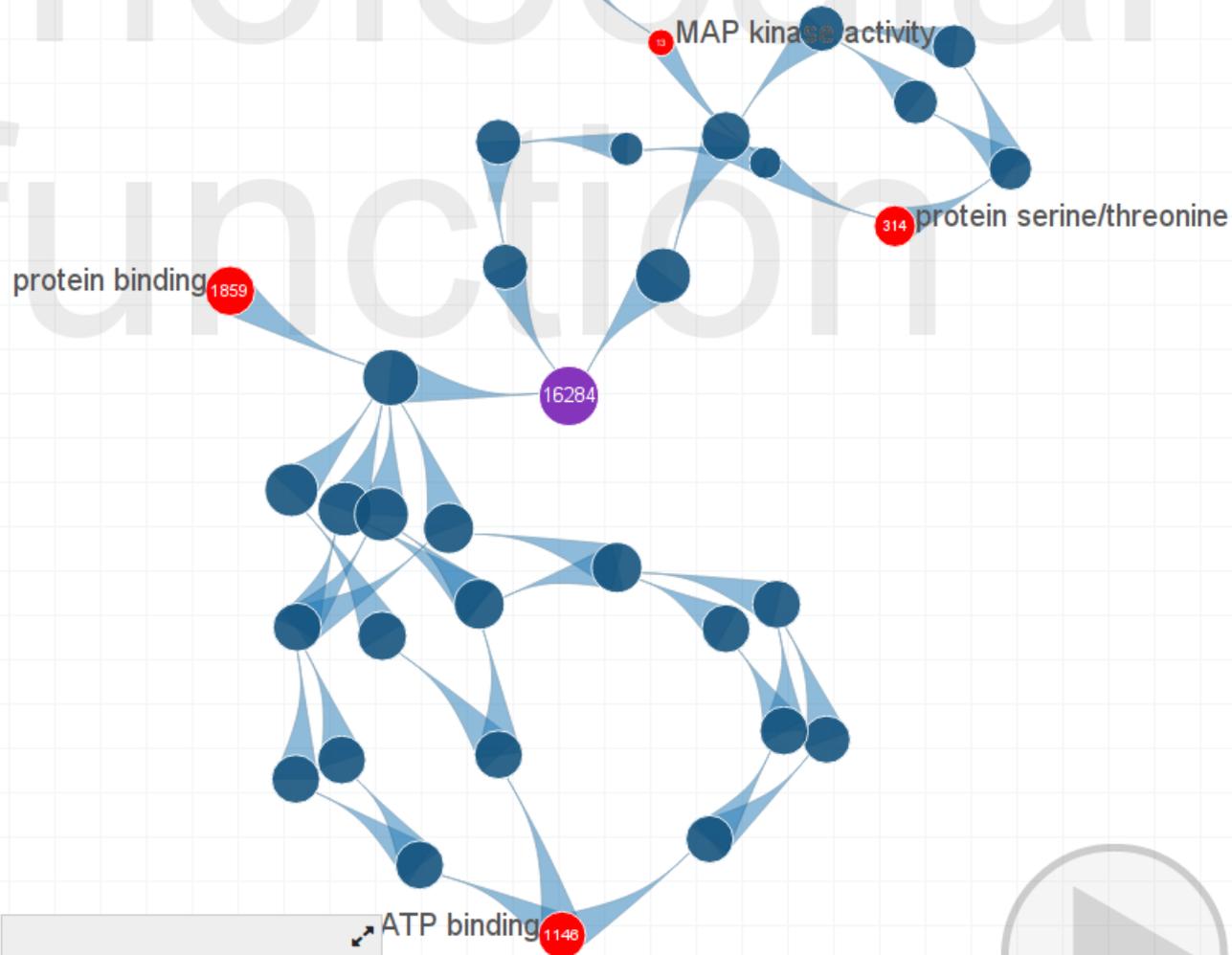
imjs → better tools

- Results tables
- Enrichment tools
- Ontology Browser
- List management tools



JUN kinase activity

Ontology Browser



InterMine

What's new - Region Search

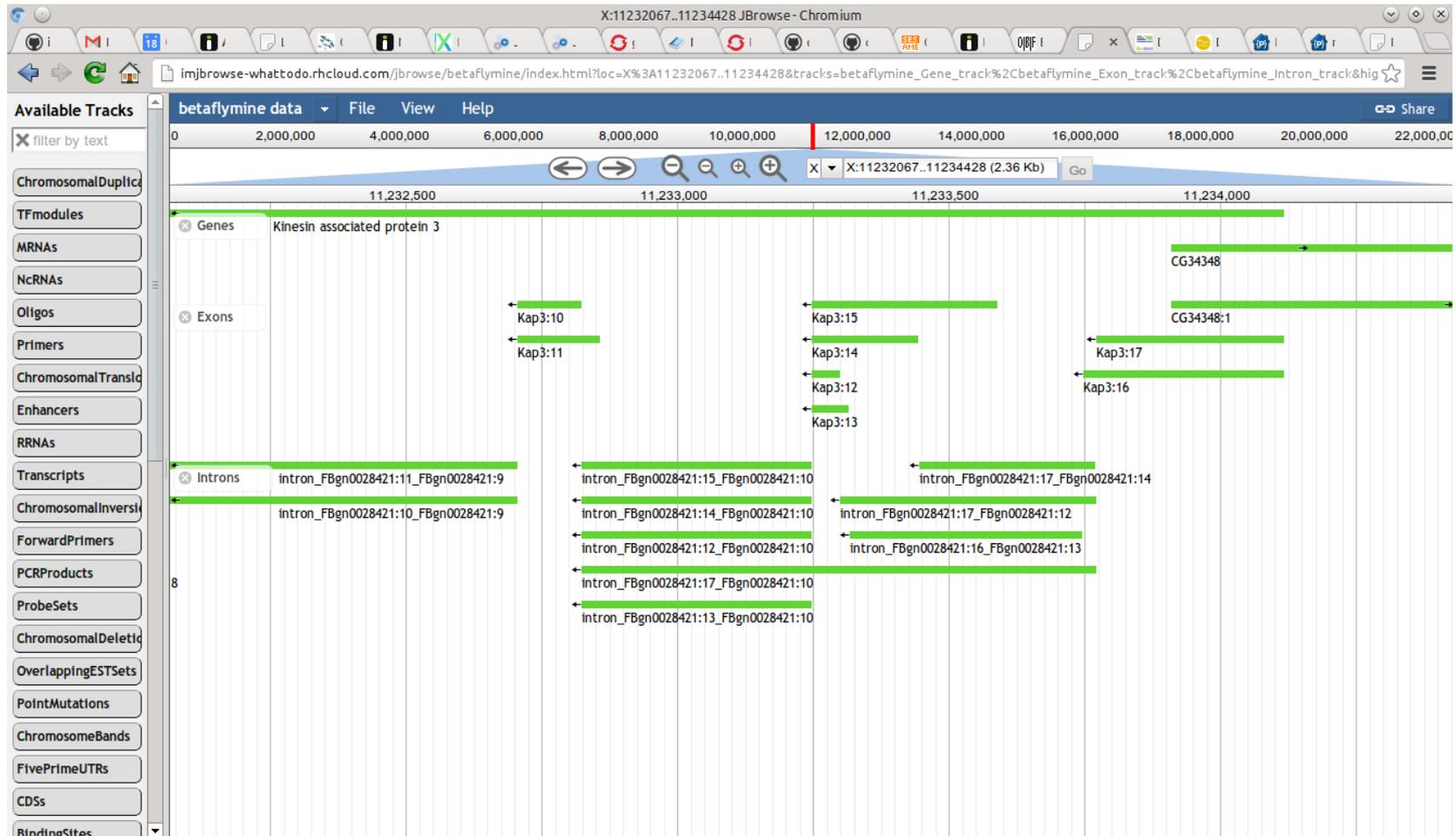
```
<query model="genomic"  
  view="Gene.primaryIdentifier Gene.symbol">  
  <constraint  
    path="Gene.organism.taxonId"  
    op="="  
    value="7227" />  
  <constraint  
    path="Gene.chromosomeLocation"  
    op="OVERLAPS">  
    <value>X:12345..67890</value>  
  </constraint>  
</query>
```



What's new - Region Search

```
{  
  "from": "Gene",  
  "select": ["primaryIdentifier", "symbol"],  
  "where": [  
    ["organism.taxonId", "=", 7227],  
    ["chromosomeLocation", "OVERLAPS", ["X:1..2"]]  
  ]  
}
```

Region Search → Better Tools

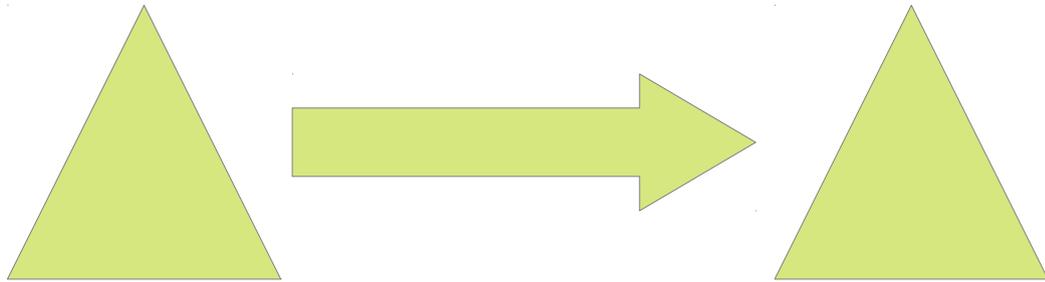


 **InterMine**

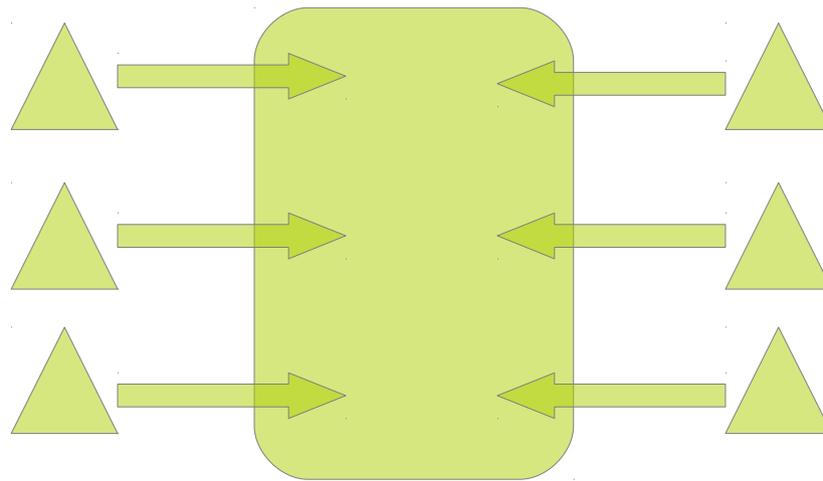
What's new - Collaboration

- Sharing lists

- One to one



- Group based



Collaboration - API

```
var flymine = new Service({root: "www.flymine.org/query"});  
  
flymine.get("groups");  
  
flymine.post("groups", {name: "a new group"});  
  
flymine.post("groups/xyz/lists", {name: "a list to share"})
```



What's next?

- A new web interface built on composable embedded tools
 - Using what we already have (libraries, widgets)
 - Building new tools (query builders, visualisation)
 - Working with third party tools (Blast, data stores)
 - Working within other sites and frameworks



› UPLOAD A LIST › LIST WIDGET › UPLOAD A LIST

Upload a List a few seconds ago

#1: Input Identifiers

#2: Resolve Identifiers

#3: Convert to a List

#4: See Results

Showing 1 to 7 of 7 rows

Rows per page: 10

« ‹ ‹‹ p. 1 ›› › »

Secondary Identifier	Symbol	DB identifier	Organism
CG1046	zen	FBgn0004053	<i>D. melanogaster</i>
CG12399	Mad	FBgn0011648	<i>D. melanogaster</i>
CG2047	ftz	FBgn0001077	<i>D. melanogaster</i>
CG3481	Adh	FBgn0000055	<i>D. melanogaster</i>
NO VALUE	PPARG	5468	<i>H. sapiens</i>
NO VALUE	Pparg	MGI:97747	<i>M. musculus</i>
NO VALUE	Pparg	RGD:3371	<i>R. norvegicus</i>

NEXT STEPS

RESEARCH NOTES

List Widgets

- Gene Expression in the Fly (FlyAtlas)
- mRNA subcellular localisation (fly-FISH)
- BDGP expression patterns
- MiRNA Enrichment
- Gene Ontology Enrichment
- Protein Domain Enrichment
- BDGP Enrichment
- Publication Enrichment
- Pathway Enrichment
- Orthologues
- Chromosome Distribution

[LIST WIDGET](#) > [UPLOAD A LIST](#) > [LIST WIDGET](#)**List Widget** *4 minutes ago*[NEXT STEPS](#)[RESEARCH NOTES](#)

#1: Choose a list

#2: See a Widget

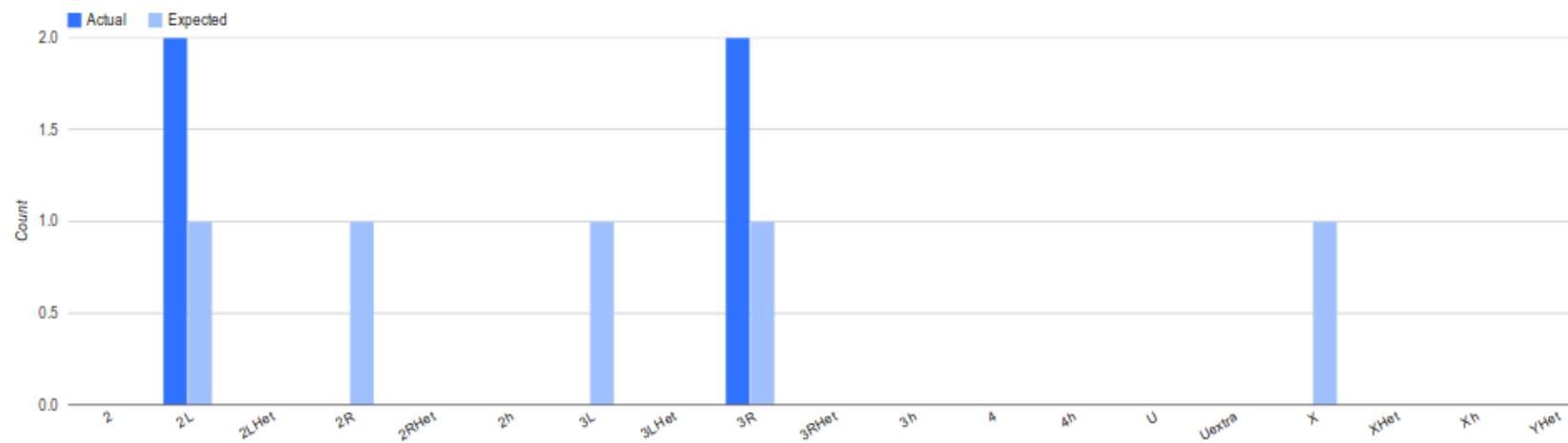
Chromosome Distribution

Actual: number of items in this list found on each chromosome. Expected: given the total number of items on the chromosome and the number of items in this list, the number of items expected to be found on each chromosome.

Number of Genes in this list not analysed in this widget: **3**

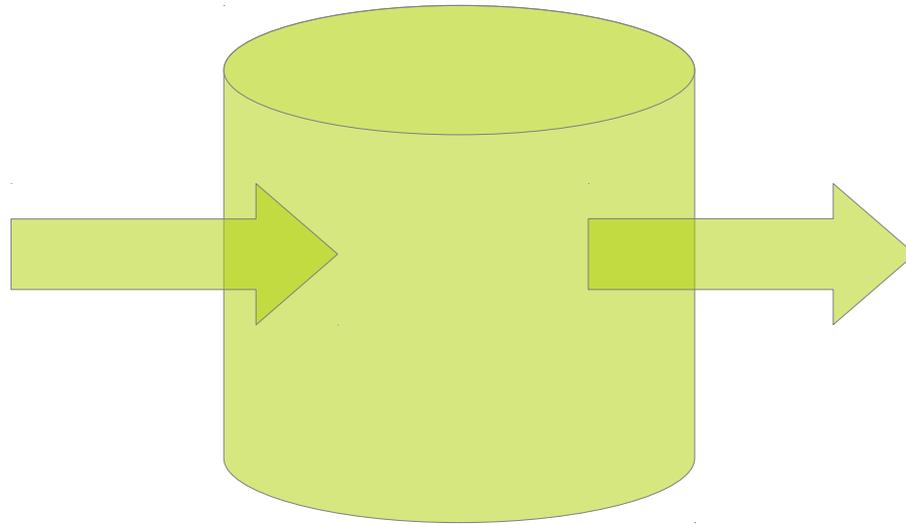
Organism

Drosophila melanx ▾

[View in table](#)

More support for standards

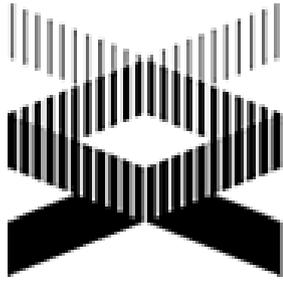
- **GFF3**
- **FASTA**
- **OWL**
- ...
- **SPARQL**



- **GFF3**
- **FASTA**
- **RDF**
- **DAS**



UNIVERSITY OF
CAMBRIDGE



National
Human Genome
Research Institute

Supported by

welcometrust

dev@intermine.org
alex@intermine.org

<http://www.intermine.org>
<https://github.com/intermine>

Gos Micklem (PI)
Sergio Contrino
Hu Fengyuan
Mike Lyne
Rachel Lyne
Alex Kalderimis
Radek Štěpán
Julie Sullivan

 **InterMine**