InterMine

Collaborative Data Mining



InterMine

Collaborative Data Mining

Who/what is InterMine? What's new?

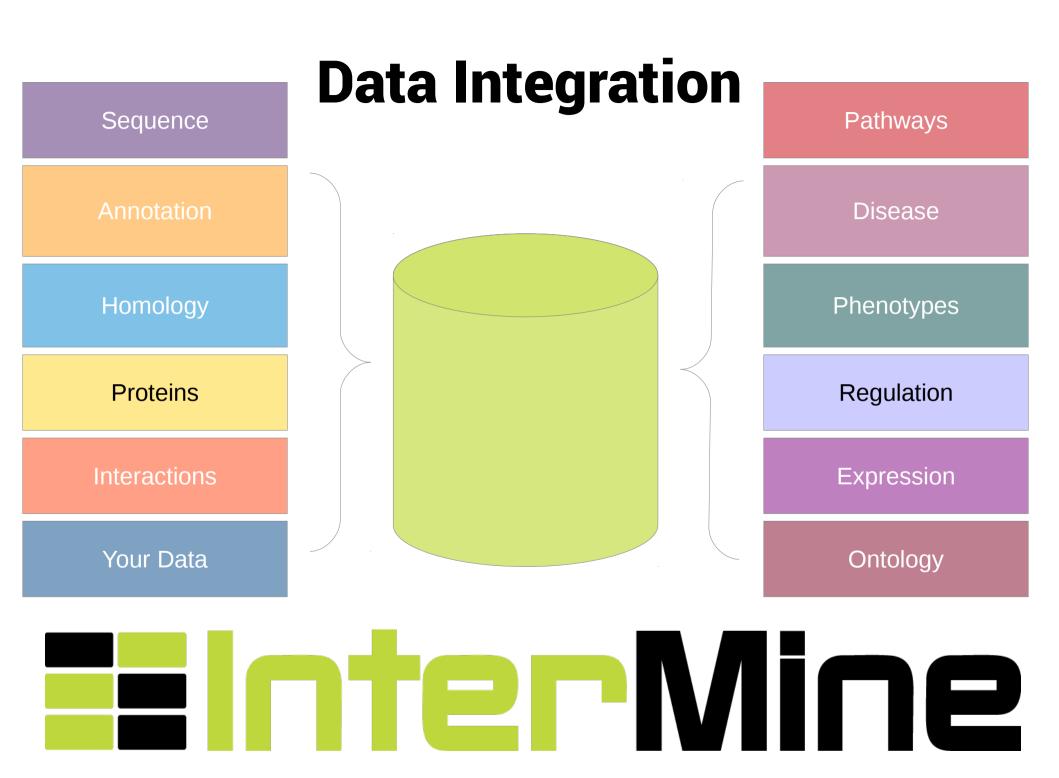
- Better client library
- Region Search
- Collaboration tools
 What's next?



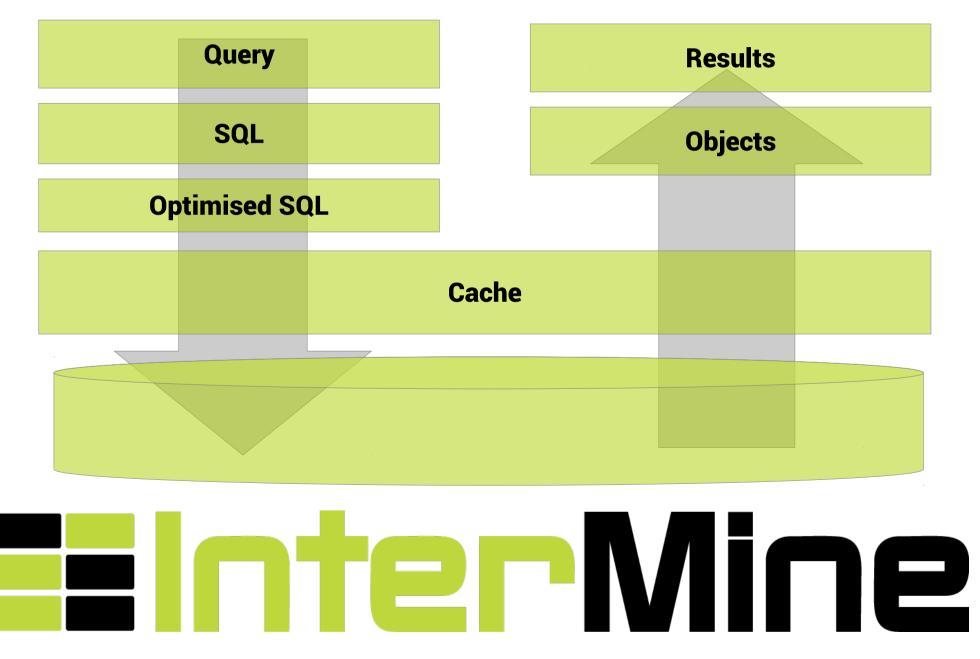
What is InterMine?

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

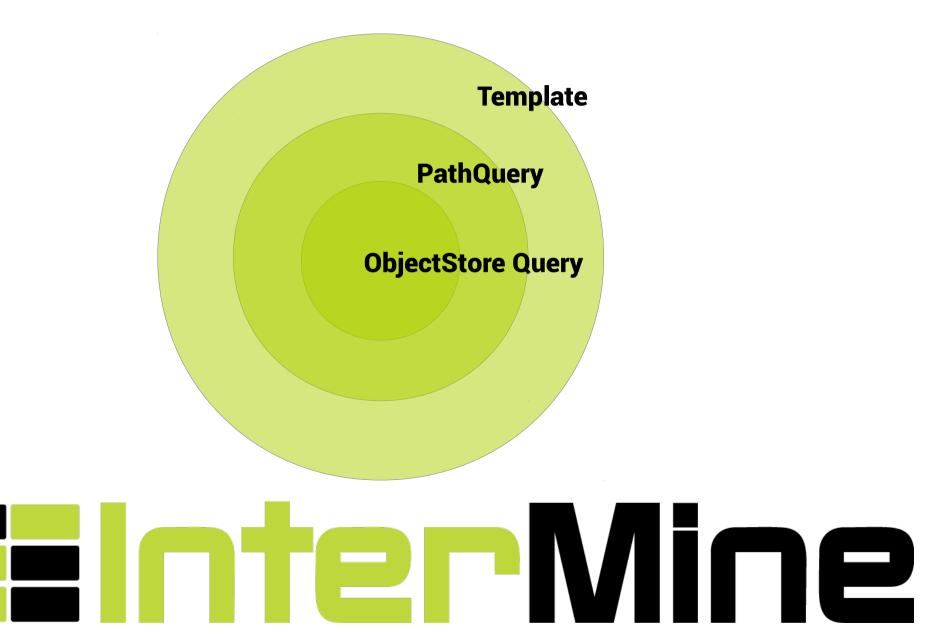




Query Engine



Query Abstraction



Why use InterMine?





Why use InterMine?



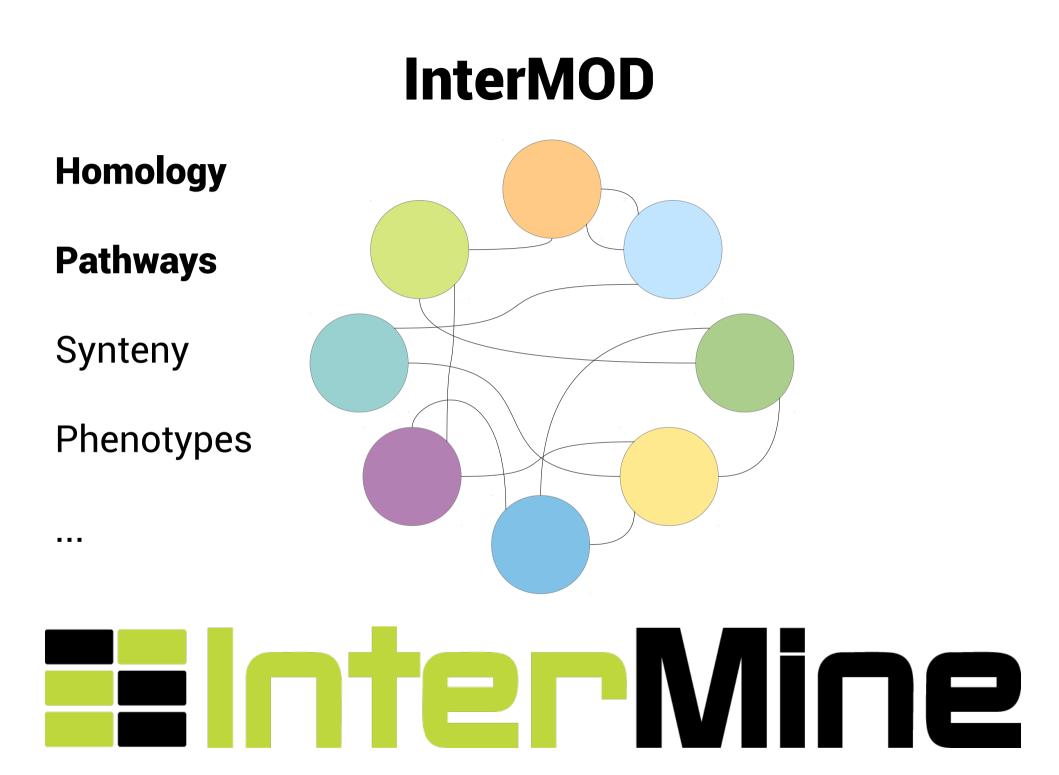
Who is/uses InterMine?

- Developed at University of Cambridge
- Used by groups that curate/integrate data
 - SGD
 - RGD
 - MGI
 - Wormbase
 - FlyMine
 - modEncode

- ZFIN
- GO
- NBIO
- MRC
- OICR







What's New? - imjs

- Dog-fooding our javascript.
- Works in node.js and the browser
- Promise based
- Comprehensive Test suite
- Documentation
 - > npm install imjs

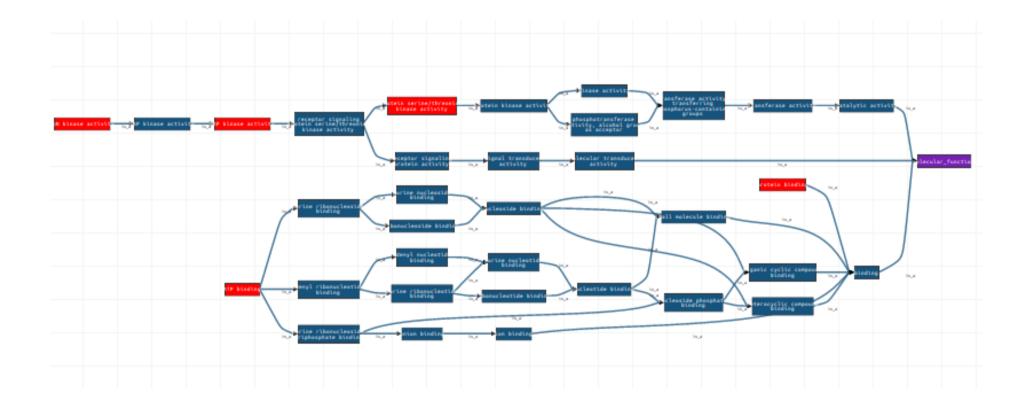


imjs \rightarrow better tools

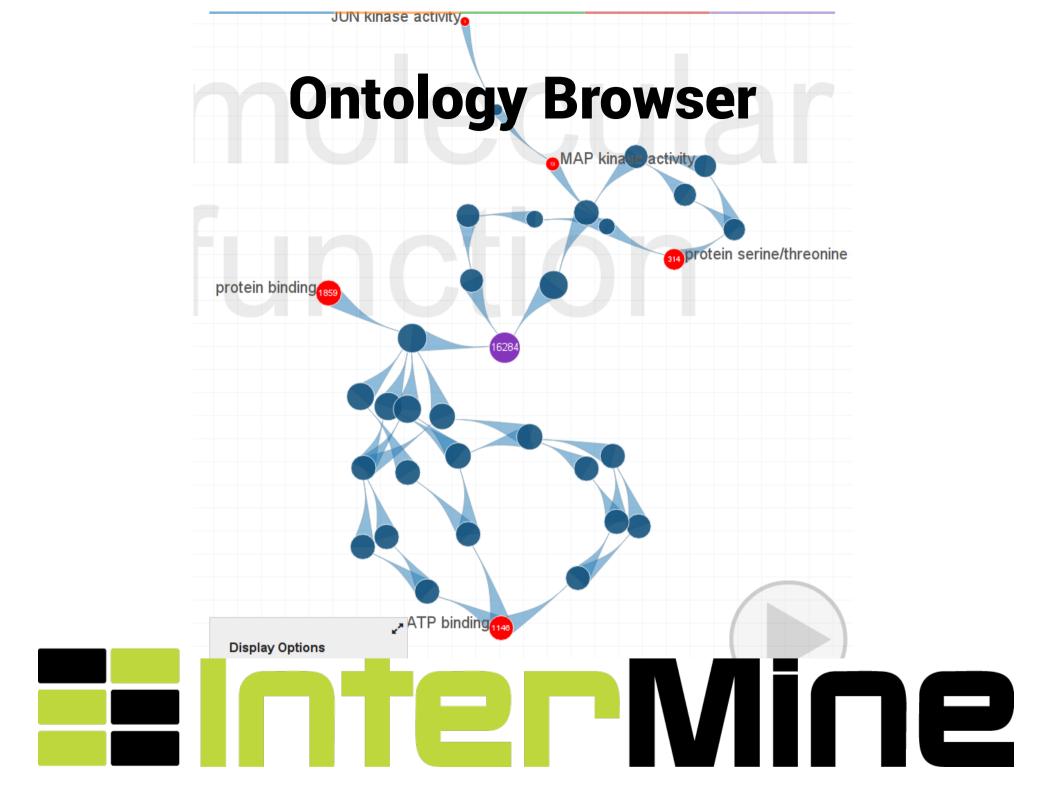
- Results tables
- Enrichment tools
- Ontology Browser
- List management tools



Ontology Browser







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"]]
]
```



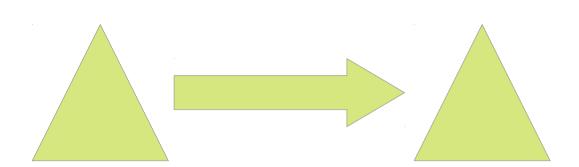
Region Search → **Better Tools**

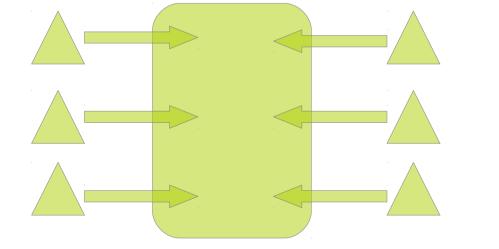
filter by text	0 2	2,000,000 4,000,000 6,00	00,000	8,000,000 10,000,000	12,000,000	14,000,000	16,000,000	18,000,000	20,000,000	22,000,00
romosomalDuplica			<	$\supset \supset Q \land \land \Theta$	X 🔻 X:112320	6711234428 (2.36 K	(b) Go			
modules		11,232,500		11,233,000		11,233,500		11,234,000		
	Genes	Kinesin associated protein 3								
NAS								CG34348	-	
RNAs E			•		+					
gos	Exons		Kap3:10		Kap3:15			CG34348:1		
mers			Kap3:11		Kap3:14		Kap3:17			
romosomalTransld					← Kap3:12		← Kap3:16			
hancers					← Kap3:13					
NAS										
anscripts	Introns	intron_FBgn0028421:11_FBgn0028421:9		<pre> intron_FBgn0028421:15_FBgn0028421 </pre>	:10	intron_FBgn002842	1:17_FBgn0028421:14			
romosomalinversi	•	intron_FBgn0028421:10_FBgn0028421:9		← intron_FBgn0028421:14_FBgn0028421	:10 intron_FBgr	n0028421:17_FBgn0028	8421:12			
rwardPrimers				← intron_FBgn0028421:12_FBgn0028421	:10 intron_FB	gn0028421:16_FBgn00	28421:13			
RProducts	8			← intron_FBgn0028421:17_FBgn0028421						
obeSets				← intron_FBgn0028421:13_FBgn0028421						
romosomalDeletio					.10					
verlappingESTSets										
intMutations										
romosomeBands										
vePrimeUTRs										
Ss										
35										

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



Inter	Mine S	tepsα	e.g. list	upload,	PPARG
-------	--------	-------	-----------	---------	-------

> UPLOAD A LIST > LIST WIDGET > UPLOAD A LIST				NEXT STEPS RESEARCH NOTES
Upload a List a few seconds ago				List Widgets
\$1: Input Identifiers	 Gene Expression in the Fly (FlyAtlas) 			
#2:Resolve Identifiers				 mRNA subcellular localisation (fly-FISH) BDGP expression patterns
#3: Convert to a List				MiRNAEnrichment Gene Ontology Enrichment
#4: See Results				Protein Domain Enrichment BDGPEnrichment Publication Enrichment
Showing 1 to 7 of 7 rows		Rows per page: 10	$\blacksquare \qquad \blacksquare \qquad$	 Pathway Enrichment Orthologues
¢0@¶.	÷ C @ T III	÷ C) @ T III	÷ 0 @ T III	Chromosome Distribution
Secondary Identifier	Symbol	DB identifier	Organism	
@ CG1046	🥥 zen	FBgn0004053	D. melanogaster	
© CG12399 @ Mad		FBgn0011648	D. melanogaster	
Q CG2047 Q ftz		FBgn0001077	D. melanogaster	
@ CG3481	Adh	FBgn0000055	D. melanogaster	
NO VALUE	PPARG	5468	H. sapiens	
NO VALUE	Pparg	MGI:97747	@ M. musculus	
NO VALUE	Pparg	RGD:3371	R. norvegicus	

nterMine Steps α	e.g. list upload, PPAF	RG		UPLOAD	ALIST							Monsieur To	ut-le-Monde • Logout
+ LIST WIDGET + UPLOAD A	LIST → LIST WIDGET											NEXT STEPS	RESEARCH NOTES
List Widget	4 minutes ago												
#1: Choose a list													
#2: See a Widget													
Chromosome Dist Actual: number of items in the chromosome.		nosome. Expected: giv	ven the total number	f items on the chromo	some and the number	of items in this lis	t, the number of iter	ns expecte	d to be found	on each			
Number of Genes in this list r	not analysed in this widget	: 3											
Organism													
Drosophila melant													
View in table													
2.0 Actual Exped	ted												
1.5													
**													
1.0													
0.5													
0.0 -2 22	21 ⁴⁰¹ 2R	2RHe1 2h	3L 3LHer	3R 3RH ^{al}	4 1'C	A ^{TA}	U Uortra	*	XHe1	*"	VH81		



More support for standards

- GFF3
- FASTA
- OWL
- •
- SPARQL

- GFF3
- FASTA
 - RDF
 - DAS







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