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

- Sequence
- Annotation
- Homology
- Proteins
- Interactions
- Your Data

- Pathways
- Disease
- Phenotypes
- Regulation
- Expression
- Ontology

InterMine
Query Engine

- Query
- SQL
- Optimised SQL
- Cache
- Results
- Objects

InterMine
Query Abstraction

- Template
- PathQuery
- ObjectStore Query

InterMine
Why use InterMine?

* may contain LGPL
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
  - ...

[InterMine logo]
InterMOD

Homology
Pathways
Synteny
Phenotypes
...

InterMine
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
Ontology Browser

[Diagram showing a network of nodes and edges representing biological processes and activities, with labels such as "kinase activity," "receptor signaling," and "nucleotide binding."

InterMine logo at the bottom right]
Ontology Browser

molecular function

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
What's new - Collaboration

• Sharing lists
  – One to one
  – Group based
Collaboration - API

```javascript
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
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 melanog.
More support for standards

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

- GFF3
- FASTA
- RDF
- DAS

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

dev@intermine.org
test@intermine.org
http://www.intermine.org
https://github.com/intermine

UNIVERSITY OF CAMBRIDGE
National Human Genome Research Institute
Supported by
wellcome trust

InterMine