Check for updates

WEB TOOL

BioJS InterMine List Analysis: A BioJS component for displaying graphical or statistical analysis of collections of items from InterMine endpoints [version 1; peer review: 1 approved]

Alexis Kalderimis, Radek Stepan, Julie Sullivan, Rachel Lyne, Michael Lyne, Gos Micklem

Department of Genetics and Cambridge Systems Biology Centre, Cambridge University, Cambridge, CB2 3EH, UK

V1 First published: 13 Feb 2014, 3:45
<https://doi.org/10.12688/f1000research.3-45.v1>
Latest published: 13 Feb 2014, 3:45
<https://doi.org/10.12688/f1000research.3-45.v1>

Abstract

Summary: The InterMineTable component is a reusable JavaScript component as part of the BioJS project. It enables users to embed powerful table-based query facilities in their websites with access to genomic data-warehouses such as <http://www.flymine.org>, which allow users to perform flexible queries over a wide range of integrated data types.

Availability: <http://github.com/alexkalderimis/im-tables-biojs>;
<http://github.com/biojs/biojs>; <http://dx.doi.org/10.5281/zenodo.8301>.

Open Peer Review

Approval Status

1

version 1

13 Feb 2014

[view](#)

1. **Clemens Wrzodek**, Roche Diagnostics GmbH, Penzberg, Germany

Any reports and responses or comments on the article can be found at the end of the article.

BIOJS

This article is included in the **BioJS** collection.

Corresponding author: Gos Micklem (g.micklem@gen.cam.ac.uk)

Competing interests: No competing interests were disclosed.

Grant information: InterMine has been developed with the support of the following grants, awarded to Dr. G. Micklem: the Wellcome Trust (Grant number: 090297), and the National Human Genome Research Institute (Grant number: R01HG004834). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding bodies.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2014 Kalderimis A *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

How to cite this article: Kalderimis A, Stepan R, Sullivan J *et al.* **BioJS InterMine List Analysis: A BioJS component for displaying graphical or statistical analysis of collections of items from InterMine endpoints [version 1; peer review: 1 approved]** F1000Research 2014, 3:45 <https://doi.org/10.12688/f1000research.3-45.v1>

First published: 13 Feb 2014, 3:45 <https://doi.org/10.12688/f1000research.3-45.v1>

Introduction

InterMine¹ is a platform for building data warehouses which includes specialisations for the life-sciences. As part of the InterMOD² project, a number of InterMine data-warehouses have been developed and released to the public containing high-quality integrated data curated by the major model organism database (MOD) organisations. In addition, the InterMine platform is widely used by other projects, such as the modENCODE project³, as well as a range of other resources including metabolicMine⁴, TargetMine⁵, FlyTFMine⁶, and MitoMiner⁷. This means that reliable integrated data sets exist for use by researchers working in a wide range of fields in the life-sciences, which can be accessed by a common interface.

One of the features of the InterMine system is the ability to store named sets of entities, called *lists*, and refer to them in queries and other analysis. This allows a user, for example, to save a list of genes and reuse this saved collection easily. The InterMine system also allows specialised analysis to be performed taking advantage of the integrated nature of the data warehouse system. For example the system can run queries that aggregate information about relationships between data types, and provide indications of levels of statistical significance for the results (*enrichment queries*).

Until recently, the output of these list analysis tools was only accessible through the web-application built into the InterMine system. Recent work on the InterMine web services has enabled this functionality to be externalised into the list-widgets⁸ project: separate JavaScript-based components that can be used in third party websites. These developments have already been incorporated into the standard InterMine web-application configuration, meaning that users of the tools described here have access to the same query and display mechanisms in their own sites that are available through the standard InterMine web-application.

InterMine supports the aims of the BioJS⁹ initiative to provide well-designed, robust website components to application developers in order to foster code reuse and minimise duplicated effort. This leads us to contribute to the BioJS project this set of components for running list analysis tools and displaying their output, so that they may be widely distributed, and interoperate with tools from other developers.

Installation

As a JavaScript web component, these tools are designed to be run within the JavaScript virtual machines provided by modern browsers, and render to HTML pages. Installation means indicating to the remote client (the user), which resources to load as dependencies, as well as where these are located. Typically this is done by adding references to these resources in the `head` section of a page through the use of `script` element (see [code sample 1](#)). Recent practice suggests loading these resources in at the end of the `body` improves page load time. The dependencies that must be loaded to use these tools are listed in [Supplementary materials A](#).

The BioJS InterMine list analysis library needs to be downloaded from the BioJS registry¹⁰ and hosted in an accessible location.

Listing 1. Loading the list analysis tools library.

```
<script
  src= "Biojs.InterMine.ListAnalysis.js" ></script>
```

Usage

Once the BioJS component and its dependencies are loaded, the component itself may be instantiated, which creates a new list analysis display, inserts it into the document, and populates it with the appropriate data by calling to the InterMine web-services. This requires that an element exists within the document (see [code listing 2](#)) into which the component can be inserted.

Listing 2. The target document element

```
<div id= "list-analysis-example"></div>
```

The JavaScript code to instantiate the component refers to this element as the `target`, and provides the other arguments required to specify which `list` we wish to analyse, the `url` of the service where that list is to be found, and which specific analysis `tool` we wish to run. The example below uses a list of genes encoding putative *Drosophila melanogaster* transcription factors made available as a public list at FlyMine¹¹ and runs the pathway enrichment statistical analysis tool. The full list of available lists (which each user can extend by creating personal lists) and analysis tools can be accessed from the InterMine service being used.

Relationship enrichment

One category of tools is the *enrichment tools*, which run queries that attempt to find relationships that are statistically significant for the set of entities as a whole. For example, FlyMine¹¹ contains both genes, loaded from sources such as FlyBase¹², and biochemical pathways, loaded from sources such as KEGG¹³ and Reactome¹⁴. The pathways enrichment tool lists pathways of which genes in the list are members, ordered by the degree of significance for the list of genes as a whole.

For example, if one gene in a list is in a particular pathway, but none of the others are, it would be considered less significant than a pathway that all or most genes in a list belonged to. Similarly, the background probability that a particular relationship exists for an item is taken into account, meaning for example that finding a publication that lists many or even all genes for an organism, such as Clark 2007¹⁵, would not be considered as significant as a publication that mentions fewer genes, but with most of them being in the list of interest.

The p-values used as measures of statistical significance are calculated by modelling the relationships as a hypergeometric distribution (as Rivals 2007¹⁶ and Beissbarth 2004¹⁷), which determines the probability that a relationship between two entities would be selected at random given the set of items to choose from. Let n be the number of items in the list, and N be the size of the reference population, and k be the number of items in the list which are involved in the given relationship (are mentioned in the publication, for example, or belong to a particular biochemical pathway), and M be the number of items in the reference population which share that same relationship. Then for each relationship

$$P = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$$

The options made available for multiple test correction include the Bonferroni, Holm-Bonferroni, and Benjamini Hochberg¹⁸ algorithms.

The tools in this category are all prefixed with `enrichment:`, and can be loaded as follows:

Listing 3. Loading an enrichment list analysis tool.

```
var ListAnalysis =
  Biojs.InterMine.ListAnalysis;
var analysis = new ListAnalysis({
  target: "list-analysis-example",
  url: "http://www.flymine.org/query",
  list: "PL FlyTF_putativeTFs",
  tool: "enrichment:pathway_enrichment"
});
```

Once run, the component should be inserted into the document (see Figure 1). The component allows the user to adjust the parameters of the analysis, including the multiple test correction method used, the p-value threshold and the background population.

The component also allows the user to interact with the results in a number of ways, specifically: by clicking on an individual item that was matched; by clicking on a button to show a set of matches; and by clicking on a button to request that the selected items be saved to some location. All these actions cause the component to emit events, which can be listened for and handled by the host JavaScript application. For example, to alert a string such as Gene – FGBN0123 when a user clicks on the corresponding element, one might attach an event listener to capture the `onClickMatch` event, see code listing 4.

Listing 4. Listening for a click event.

```
analysis.onClickMatch(function (ident, type) {
  alert(type + " - " + ident);
});
```

This enables the behaviour of the component to be integrated into the hosting application. The full listing of events and their arguments is included in the BioJS API documentation¹⁹.

The canonical example for the use of statistical enrichment in bioinformatics is enrichment of Gene Ontology (GO) terms for sequence annotations (Rivals 2007¹⁶). This functionality is supported as one of the statistical analysis tools (see Figure 2), within this more generic enrichment analysis framework. The GO enrichment tool merits some further notes, however, as it supports some of the more advanced parameters.

Pathway Enrichment

Pathways enriched for genes in this list - data from KEGG and Reactome

Number of Genes in this list not analysed in this widget: 633

Test Correction	Max p-value	DataSet
Holm-Bonferroni	0.05	All

Background population

Default [Change](#)

[View](#) [Download](#)

☐ Pathways	p-Value	Matches
☐ Generic Transcription Pathway	8.600689e-23	26
☐ Regulation of gene expression in early pancreatic precursor cells	5.967443e-14	11
☐ Regulation of gene expression in late stage (branching morphogenesis) pancreatic bud precursor cells	1.987925e-11	11
☐ Nuclear Receptor transcription pathway	1.987925e-11	11
☐ Developmental Biology	1.455013e-8	38
☐ Regulation of gene expression in endocrine-committed (NEUROG3+) progenitor cells	1.703169e-6	6
☐ PPARA Activates Gene Expression	1.220927e-5	15
☐ Regulation of Lipid Metabolism by Peroxisome proliferator-activated receptor alpha (PPARalpha)	1.220927e-5	15
☐ Transcriptional activity of SMAD2/SMAD3/SMAD4 heterotrimer	4.486373e-5	10

Figure 1. A list analysis tool displaying the results of a statistical analysis query.

Gene Ontology Enrichment

GO terms enriched for items in this list.

Number of Genes in this list not analysed in this widget: 230

Test Correction Max p-value Ontology

Holm-Bonferroni 0.05 biological_process

Background population Normalise by length

Default Change Normalise

View Download

☐ GO Term	p-Value	Matches
☐ transcription, DNA-dependent [GO:0006351]	0.000000e+0	435
☐ regulation of transcription, DNA-dependent [GO:0006355]	0.000000e+0	428
☐ regulation of biosynthetic process [GO:0009889]	0.000000e+0	429
☐ regulation of gene expression [GO:0010468]	0.000000e+0	435
☐ regulation of macromolecule biosynthetic process [GO:0010556]	0.000000e+0	428

Figure 2. A list analysis tool displaying the results of the Gene Ontology (GO) statistical analysis query.

The GO enrichment tool demonstrates the use of optional filter parameters to limit the results in some way. In the GO tool, it allows the user to select the sub-ontology they are interested in. The user can also choose to normalise the results of this tool, in this case by transcript length.

Charts

The other main category of analysis tools is the *chart tools*. These run aggregate queries over the items in a list, and present the information graphically in interactive charts. The InterMine system supports both numerical and categorical charting, reflected in the supported chart formats: bar charts, line charts, pie charts and scatterplots.

Loading a chart analysis tool is identical to loading a statistical enrichment tool - only the name of the tool need differ (see [code listing 5](#)).

Listing 5. Loading a chart list analysis tool.

```
var chart = new Biojs.InterMine.ListAnalysis({
  target: "list-analysis-example",
  url: "http://www.flymine.org/query",
  list: "PL FlyTF_putativeTFs",
  tool: "chart:flyfish"
});
```

This code will request data for the particular tool (flyfish), as run against the given input list (PL FlyTF_putativeTFs), and then display the results in the appropriate chart format ([Figure 3](#)). The chart tools have fewer parameters; they may take a single parameter, as detailed in the tool description available from the relevant service (e.g. <http://www.flymine.org/query/service/widgets>).

In most cases they do not provide mechanisms for the user to change the results displayed. They do however provide several mechanisms

mRNA subcellular localisation (fly-FISH)

Expression patterns of *Drosophila* mRNAs at the subcellular level during early embryogenesis - data from [Fly-FISH](#). Note that not all genes have been assayed by [Fly-FISH](#).

Number of Genes in this list not analysed in this widget: [576](#)

[View in table](#)

Figure 3. A list analysis component displaying the results of a the `chart:flyfish` tool (loaded in [Code Listing 5](#)), which queries against [Fly-FISH²⁰](#) data.

for the user to interact with the results displayed. The user can click on the groupings or data-points represented on the chart (see [Figure 4](#)), which allows the user to trigger the same events available to enrichment tools, which can be captured the same way (see [code listing 4](#)).

Discussion

This tool addresses an important set of needs for bioinformatics developers: the ability to perform enrichment analysis, and the the visualisation of typed relationships between entities. The InterMine platform, and this BioJS component make performing these analyses and displaying the output straightforward. It allows the developers to focus on integrating this functionality where it is needed, and users to focus on interpreting rather than retrieving the data. It is expected that wide availability of these tools will provide significant savings in time for typically stretched developers and researchers. By providing this functionality as a BioJS component, it is hoped that integration between different tools will result in the creation

of applications that are able to integrate analysis and visualisation from different platforms.

Conclusions

It is hoped that this component will prove useful to those developing tools for researchers in the life-sciences. Significant work has gone into creating, curating and combining high quality data sets. The recent work in exposing these resources through web-services and producing reusable web-based components allows this investment to benefit not just visitors to sites based on InterMine applications, but any developer or user who aims to include this kind of statistical analysis and visualisation in their platform. By providing bioinformatics web-developers, and their users, with access to a broad range of data sources meeting the needs of many diverse research communities, we expect to help reduce the development burden on projects with limited resources, and help minimise redundancy of effort.

Figure 4. The result of a user clicking on the “stage 6–7, expressed” bar of the chart.

Software availability

Zenodo: BioJS InterMine List Analysis Widgets, doi: [10.5281/zenodo.830221](https://doi.org/10.5281/zenodo.830221).

GitHub: BioJS, [http://github.com/biojs/biojs](https://github.com/biojs/biojs).

Author contributions

Alex Kalderimis wrote the manuscript and implemented the BioJS wrapper, under the supervision of Gos Micklem, to a set of user specifications supplied by Julie Sullivan. Radek Štěpán implemented the list analysis component, based on designs and specification from Julie Sullivan, Rachel Lyne, Mike Lyne and Alex Kalderimis. Rachel Lyne and Mike Lyne contributed to the component design and revised the manuscript. All authors have approved the manuscript.

Competing interests

No competing interests were disclosed.

Grant information

InterMine has been developed with the support of the following grants, awarded to Dr. G. Micklem: the Wellcome Trust (Grant number: 090297), and the National Human Genome Research Institute (Grant number: R01HG004834). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding bodies.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

The authors thank Manuel Corpas for useful feedback.

Supplementary materials

A Dependencies

```
<link rel="stylesheet" type="text/css"
  href="http://cdn.intermine.org/
js/intermine/apps-c/list-widgets/2.0.
4/app.bundle.prefix.min.css">
<script src="http://cdn.intermine.org
```

```
/js/intermine/apps-c/list-widgets/2.0
.4/app.bundle.min.js"></script>
```

Here we are referring to resources which are made publicly available as part of a Content Delivery Network (CDN). These resources could just as well be hosted locally.

References

1. Smith RN, Aleksic J, Butano D, *et al.*: **InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data.** *Bioinformatics*. 2012; **28**(23): 3163–3165.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
2. Sullivan J, Karra K, Moxon SA, *et al.*: **InterMOD: integrated data and tools for the unification of model organism research.** *Sci Rep*. 2013; **3**: 1802.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
3. Contrino S, Smith RN, Butano D, *et al.*: **modMine: flexible access to modENCODE data.** *Nucleic Acids Res*. 2012; **40**(Database issue): D1082–D1088.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Lyne M, Smith RN, Lyne R, *et al.*: **metabolicMine: an integrated genomics, genetics and proteomics data warehouse for common metabolic disease research.** *Database (Oxford)*. 2013; **2013**: bat060.
[PubMed Abstract](#) | [Publisher Full Text](#)
5. Chen YA, Tripathi LP, Mizuguchi K: **TargetMine, an integrated data warehouse for candidate gene prioritisation and target discovery.** *PLoS ONE*. 2011; **6**(3): e17844.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. Adryan B, Teichmann SA: **FlyTF: a systematic review of site-specific transcription factors in the fruit fly *Drosophila melanogaster*.** *Bioinformatics*. 2006; **22**(12): 1532–1533.
[PubMed Abstract](#) | [Publisher Full Text](#)
7. Smith AC, Robinson AJ: **MitoMiner, an integrated database for the storage and analysis of mitochondrial proteomics data.** *Mol Cell Proteomics*. 2009; **8**(6): 1324–1337.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. **List widgets project.**
[Reference Source](#)
9. Gómez J, García LJ, Salazar GA, *et al.*: **BioJS: an open source JavaScript framework for biological data visualization.** *Bioinformatics*. 2013; **29**(8): 1103–1104.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. **Biojs project registry.**
[Reference Source](#)
11. Lyne R, Smith R, Rutherford K, *et al.*: **FlyMine: an integrated database for *Drosophila* and *Anopheles* genomics.** *Genome Biol*. 2007; **8**(7): R129.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Marygold SJ, Leyland PC, Seal RL, *et al.*: **Fly-Base: improvements to the bibliography.** *Nucleic Acids Res*. 2013; **41**(Database issue): D751–757.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. Ogata H, Goto S, Sato K, *et al.*: **KEGG: Kyoto Encyclopedia of Genes and Genomes.** *Nucleic Acids Res*. 1999; **27**(1): 29–34.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
14. Joshi-Tope G, Gillespie M, Vastrik I, *et al.*: **Reactome: a knowledgebase of biological pathways.** *Nucleic Acids Res*. 2005; **33**(Database issue): D428–432.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
15. Clark AG, Eisen MB, Smith DR, *et al.*: **Evolution of genes and genomes on the *Drosophila* phylogeny.** *Nature*. 2007; **450**(7167): 203–218.
[PubMed Abstract](#) | [Publisher Full Text](#)
16. Rivals I, Personnaz L, Taing L, *et al.*: **Enrichment or depletion of a GO category within a class of genes: which test?** *Bioinformatics*. 2007; **23**(4): 401–407.
[PubMed Abstract](#) | [Publisher Full Text](#)
17. Beissbarth T, Speed TP: **GStat: find statistically overrepresented Gene Ontologies within a group of genes.** *Bioinformatics*. 2004; **20**(9): 1464–1465.
[PubMed Abstract](#) | [Publisher Full Text](#)
18. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: A practical and powerful approach to multiple testing.** *J R Stat Soc Series B Methodol*. 1995; **57**(1): 289–300.
[Publisher Full Text](#)
19. **Biojs project api documentation.**
[Reference Source](#)
20. Lécuyer E, Yoshida H, Parthasarathy N, *et al.*: **Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function.** *Cell*. 2007; **131**(1): 174–187.
[PubMed Abstract](#) | [Publisher Full Text](#)
21. Kalderimis A, Micklem G, Stepan R: **BioJS InterMine List Analysis Widgets.** *Zenodo*. 2014.
[Data Source](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 13 October 2014

<https://doi.org/10.5256/f1000research.3699.r5874>

© 2014 Wrzodek C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Clemens Wrzodek

Roche Innovation Center Penzberg, Roche Diagnostics GmbH, Penzberg, Germany

The Manuscript:

The Article is very clearly written and formatted. It strongly focuses on the end-users that want to use the published library and describes how to include it and what possibilities it offers. The examples shown in the manuscript are nice and well-picked. The manuscript is interesting and easy-to-read.

I did not find it useful to include the formula for the hypergeometric test in the manuscript. Nearly every manuscript that even mentions enrichments of lists of genes depicts this formula. It is already known very well by researchers in this area and for beginners, the information provided is rarely sufficient. Personally, I would remove it.

I would recommend the authors to rethink the title. Maybe something shorter like "BioJS InterMine List Analysis: A BioJS component for displaying InterMine analysis results" might be less-confusing (even though it misses the information about the various possible endpoints).

What-is-published-here:

However, what is published here is only the approximate 400 lines-of-code long BioJS-wrapper for InterMine (available as a single JS file on GitHub). It's not the implementation of the described analysis methods, nor is it the InterMine library itself. It's just the plain BioJS wrapper for the analysis methods offered by the InterMine endpoints.

Actual source code:

I tested the provided Demo on GitHub:

- Works very well in Chrome 37.
- It does not work in IE10 ("Unable to construct query: 8070000c").
- Works in Firefox (although the grid on-mouse over-popup behaves different than in Chrome).

It may be nice to mention some information about Browser compatibility in the manuscript and

on GitHub. I further tested the automated code generation, available from a button in the upper-right corner (specifically, the Java-Code). That worked well. I extended the Demo JavaScript file and played a bit with the information, provided in the manuscript. Everything seemed to work well.

Generally, the code listings in the manuscript are very helpful when working with the library. Also, the Demo file on Github helps getting started. The JavaScript code itself is well documented with comments.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research