

Systems biology

Advance Access publication September 26, 2013

## INVEX—a web-based tool for integrative visualization of expression data

Jianguo Xia<sup>1,2</sup>, Ngan H. Lyle<sup>1,2</sup>, Matthew L. Mayer<sup>1,2</sup>, Olga M. Pena<sup>1,2</sup> and Robert E. W. Hancock<sup>1,2,3,\*</sup>

<sup>1</sup>Department of Microbiology and Immunology, University of British Columbia, Vancouver, Canada V6T 1Z3, <sup>2</sup>Centre for Microbial Diseases and Immunity Research, University of British Columbia, Vancouver, Canada V6T 1Z4 and <sup>3</sup>Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK

Associate Editor: Michael Brudno

### ABSTRACT

**Summary:** Gene expression or metabolomics data generated from clinical settings are often associated with multiple metadata (i.e. diagnosis, genotype, gender, etc.). It is of great interest to analyze and to visualize the data in these contexts. Here, we introduce INVEX—a novel web-based tool that integrates the server-side capabilities for data analysis with the browser-based technology for data visualization. INVEX has two key features: (i) flexible differential expression analysis for a wide variety of experimental designs; and (ii) interactive visualization within the context of metadata and biological annotations. INVEX has built-in support for gene/metabolite annotation and a fully functional heatmap builder.

**Availability and implementation:** Freely available at <http://www.invex.ca>.

**Contact:** bob@hancocklab.ubc.ca

Received on July 22, 2013; revised on September 13, 2013; accepted on September 20, 2013

### 1 INTRODUCTION

‘Omics’ technologies such as microarrays, next-generation sequencing and metabolomics are increasingly used in clinical studies. In many cases, a single dataset will be associated with multiple clinical parameters (metadata) such as diagnosis, genotype, gender and so forth. It is of great interest to analyze and to visualize the data within these contexts as well as biological annotations to enable one to capture dynamic changes that correlate with clinical factors. Linear models have proven to be a powerful and flexible approach for analysis of gene expression experiments (Smyth, 2005). However, users need to have a deep understanding of statistical concepts and R language to use this approach properly. Heatmaps have proven to be useful in visualizing expression data. Heatmaps coupled with clustering have become hallmarks for the presentation of gene expression data. However, most web-based tools provide only static images with limited support for user interactions. Interactive visualization is mostly limited to stand-alone tools or Java applet plugin (Pavlidis and Noble, 2003; Perez-Llamas and Lopez-Bigas, 2011; Reich *et al.*, 2006; Saeed *et al.*, 2003; Saldanha, 2004). The rapid development of information technology, especially

HTML5 and JavaScript, has presented new opportunities to overcome this limitation (Miller *et al.*, 2013; Tan *et al.*, 2013). Here, we present INVEX (Integrative Visualization of Expression data)—an intuitive web-based tool that seamlessly integrates server-side capabilities for data analysis and annotation with the browser-based technology for data visualization. INVEX allows researchers to perform flexible data analysis and to visually explore the results as interactive heatmaps within the context of associated metadata and biological annotations.

### 2 IMPLEMENTATION

INVEX is composed of two modules. The server-side module was implemented using the latest JavaServer Faces 2.0 technology. The data analysis was based on R and several packages from Bioconductor (Gentleman *et al.*, 2004). The client-side module was developed based on the HTML5 canvas and JavaScript using the jQuery library ([www.jquery.com](http://www.jquery.com)). INVEX has been tested using Google Chrome (5.0+), Firefox (3.0+) and Internet Explorer (9.0+) browsers. The performance of visualization depends on the users’ computer. We recommend accessing INVEX from a computer with at least a 15-inch screen and 2 GB memory.

### 3 APPLICATION EXAMPLE

INVEX provides four example datasets each associated with multiple metadata—three gene expression datasets (*Estrogen*, *Sepsis* and *TimeSeries*) and one metabolomics data (*Cachexia*). Here, we illustrate the main features of INVEX using the *Sepsis* dataset.

#### 3.1 Data upload, annotation and analysis

INVEX accepts an expression data table annotated with various metadata. The data can be uploaded as a tab-delimited text (.txt) or in its compressed format (.zip) (see INVEX ‘Help’ page for detailed instructions). Click the ‘Start’ menu on the home page to enter the analysis page. To access our test datasets, click the ‘Try Examples’ button on the bottom left. The four datasets are listed with detailed descriptions. The *Sepsis* data were generated from a study comparing gene expression changes from Lipopolysaccharide (LPS)-induced inflammation to endotoxin tolerance in human peripheral blood mononuclear cells

\*To whom correspondence should be addressed.



Fig. 1. Visual data exploration using INVEV. (A) Overview, (B) Focus view, (C) Annotation view and (D) Heatmap builder

(PBMC) from four donors (Pena *et al.*, 2011). There are three experimental conditions: control, LPS (pro-inflammatory) and LPS\_LPS indicating two doses of LPS treatments within a day (leading to endotoxin tolerance). Thus, there are two types of metadata: Treatment and Donor. Select the *Sepsis* data and click ‘Yes’ to upload the file. For the convenience of testing, INVEV sets the default parameters for the remaining steps—annotation, normalization, differential expression and enrichment analysis. Click ‘Submit’ to proceed in the first two steps. For differential analysis, INVEV can deal with two- or multiple-group, paired or block design, time-series, common-control and nested-comparisons. Here, it is particularly interesting to compare genes that respond differently between the two treatments. To perform such analyses, we choose ‘nested comparisons’ between ‘control vs. LPS’ and ‘control vs. LPS\_LPS’, and select ‘Interactions only’. The analysis returns 1791 significant genes, which is reduced to 251 after setting  $\log_2$  fold change cutoff to 1.0 (2-fold change in expression). Use ‘Kyoto Encyclopedia of Genes and Genomes (KEGG)’ for enrichment analysis. Finally, click ‘Proceed to visualization’ to enter the visualization page.

### 3.2 Visual data exploration

A screenshot of the visualization page is shown in Figure 1. There are four views—*Overview*, *Focus view*, *Annotation view* and *Heatmap builder*, with a top toolbar containing menus for adjusting resolution, colors, clustering and so forth. The *Overview* on the left provides a bird’s-eye view of the expression profile of all significant genes (Fig. 1A). By default, genes are ordered by their adjusted *P*-values. Select ‘Euclidean distance’ to cluster genes for pattern discovery. Users can drag to select any region of interest for detailed inspection in the *Focus view*. The *Focus view* in the center shows the gene expression profiles of current interest with three adjustable resolutions (Fig. 1B). The metadata, sample IDs and color keys are displayed on the top and bottom panels, respectively. Double click the ‘Treatment’ metadata row to sort all samples accordingly. On the right side, the *Overall Enriched Themes* pane shows enriched pathways, *P*-values and matched gene numbers (Fig. 1C). Double click the name of the top hit ‘cytokine–cytokine receptor

interaction’ to visualize the expression profiles of its 22 genes. The *Enriched Themes in Current Focus* bar allows users to identify enriched functional modules for genes in *Focus view*.

### 3.3 Building custom heatmaps

Click the ‘Heatmap builder’ from the top toolbar to activate the *Heatmap builder* (Fig. 1D). This pane is a ‘playground’ that allows users to easily create custom heatmaps to reveal specific features. Users can now double click to select a single gene or drag to select multiple genes from *Focus view* to *Heatmap builder*. Within the builder, users can double click to delete or drag to reorganize a row. Separators (blank rows) can be added and then dragged to specific positions to create visual cluster boundaries. When all genes of interest are included and organized, users can edit samples using the ‘Edit samples’ option. Finally, all heatmaps can be exported as portable network graphics (PNG) images labeled with metadata, sample IDs and color keys by using the ‘Download’ function from the top toolbar.

## 4 CONCLUSIONS

In the context of integrative analysis of expression data from clinical studies, there are two general scenarios: multiple expression datasets collected for the same disease or multiple metadata collected for a single dataset. We have recently developed INMEX, a tool to support data analysis in the former scenario (Xia *et al.*, 2013). In this article, we introduce INVEV that has been designed for the latter scenario. By coupling the conventional server-side functions for data analysis and annotation with the client-side visualization technologies, INVEV provides a promising approach for developing efficient bioinformatics tool in the ‘omics’ era.

**Funding:** Canadian Institutes for Health Research (CIHR). J. Xia was supported by a CIHR Postdoctoral Fellowship and Killam Postdoctoral Research Fellowship. R.E.W.H. holds a Canada Research Chair. The charge for open access fee was covered by UBC Killam Advanced Studies Fund.

**Conflict of Interest:** none declared.

## REFERENCES

- Gentleman,R.C. et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.
- Miller,C.A. et al. (2013) Scribl: an HTML5 Canvas-based graphics library for visualizing genomic data over the web. *Bioinformatics*, **29**, 381–383.
- Pavlidis,P. and Noble,W.S. (2003) Matrix2png: a utility for visualizing matrix data. *Bioinformatics*, **19**, 295–296.
- Peña,O.M. et al. (2011) Endotoxin tolerance represents a distinctive state of alternative polarization (M2) in human mononuclear cells. *J. Immunol.*, **186**, 7243–7254.
- Perez-Llamas,C. and Lopez-Bigas,N. (2011) Gitoools: analysis and visualisation of genomic data using interactive heat-maps. *PLoS One*, **6**, e19541.
- Reich,M. et al. (2006) GenePattern 2.0. *Nat. Genet.*, **38**, 500–501.
- Saeed,A.I. et al. (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*, **34**, 374–378.
- Saldanha,A.J. (2004) Java Treeview—extensible visualization of microarray data. *Bioinformatics*, **20**, 3246–3248.
- Smyth,G.K. (2005) Limma: linear models for microarray data. In: *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, New York, Springer, pp. 397–420.
- Tan,C.M. et al. (2013) Network2Canvas: network visualization on a canvas with enrichment analysis. *Bioinformatics*, **29**, 1872–1878.
- Xia,J. et al. (2013) INMEX—a web-based tool for integrative meta-analysis of expression data. *Nucleic Acids Res.*, **41**, W63–W70.