**ExpressionPlot: A web-based framework for analysis of RNA-Seq and microarray gene expression data**

Brad A. Friedman\(^1\)*, Hemali Phatnani\(^2\), Tom Maniatis\(^1,2\)

\(^1\)Department of Molecular and Cell Biology, Harvard University, Cambridge, MA
\(^2\)Department of Biochemistry and Molecular Biophysics, Columbia University College of Physicians and Surgeons, New York, NY

*brad.aaron.friedman@gmail.com*

**Abstract**

Exon array and RNA-Seq platforms have emerged as important tools for detecting changes in gene expression and RNA processing in biological samples. We have developed a software package called *ExpressionPlot* (EP) that can process both types of data and provides an easy-to-use exploratory web interface. The package consists of a back end, which processes and prepares the raw high-throughput data, and the web-based front end, which allows non-specialists to browse, visualize, and compare different data sets.

**EP Backend: Data Pre-processing**

**Platforms for Expression Profiling**

<table>
<thead>
<tr>
<th>Platform</th>
<th>Measurement</th>
<th>Queried Regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affy 3' UTR Arrays</td>
<td>Probe Intensities</td>
<td>Designed: 3' UTRs</td>
</tr>
<tr>
<td>Affy Exon Arrays</td>
<td>Probe Intensities</td>
<td>Designed: All Exons</td>
</tr>
<tr>
<td>RNA-Seq</td>
<td>Normalized Read Counts</td>
<td>Unrestricted: genomic, spliced, exogenous</td>
</tr>
</tbody>
</table>

**Microarrays**

Input: CEL files

- Background subtraction, normalization
- Generate probe intensities for gene models, altsplicing events
- Calculate fold changes and P-values for gene level and altsplicing changes
- Register data set with website, set access privelages

**RNA-Seq**

Input: FASTQ files

- Generate genomic, splice junction alignments
- Count reads for gene models, altsplicing events
- Calculate fold changes and P-values for gene level and altsplicing changes
- Table Browser showing brain-enriched genes can be filtered and sorted. Each row is linked to the readmap browser.

**ExpressionPlot User Interface**

- Current tool
- Navigation bar
- User login
- Plotting options

**EP RNA-Seq Overviews**

- Percent of aligning reads by region (Human Tissue Panel)
- Pairwise correlations of gene level profiles

**2 way: Examining Gene Level Changes**

- Two-way plot showing brain-enriched and brain-depleted genes

**4 way: Comparing different datasets**

- Four-way plots comparing brain-enriched expression in different experiments. Left, human exon array tissue panel. Right, mouse RNA-Seq tissue panel. Colors indicate genes significantly changed in one (red or green) or both (blue) experiments.

**Readmaps: Visualizing RNA-Seq data**

- Readmap showing brain-restricted expression of myelin-associated glycoprotein. Black bars indicate reads mapping to genomic positions, blue brackets indicate reads mapping to splice junctions.

**Other tools**

- genelev: Barplots of the level of a gene across different samples
- ecdf: Empirical Cumulative Distribution Function of selected sets of gene levels over different samples or fold changes over comparisons
- event_heatmap: generate heatmap showing gene levels or changes across multiple events and data sets
- pairplot: Visualize paired-end alignments

**Case Study: Tissue-specific gene expression and Splicing**

- ExpressionPlot was used to investigate tissue-specific gene expression using exon array data from Affymetrix and RNA-Seq data from the Burge and Wold labs.

**Want to try ExpressionPlot?**

Go to friedman-lab.com/ep

- Try ExpressionPlot on our public datasets.
- Join our google group google.com/group/expressplot
- Download and install ExpressionPlot on your own server (Available Sept 1, 2010).