Functional enrichment analysis of high-throughput omics data
Contents

1 Gene expression-based enrichment analysis
2 Genomic region enrichment analysis
3 Multi-omics enrichment analysis
Contents

1 Gene expression-based enrichment analysis
2 Genomic region enrichment analysis
3 Multi-omics enrichment analysis
Gene expression data analysis

Experiment

Preprocessing

Gene analysis

Gene set/network analysis

Biological interpretation
**EnrichmentBrowser**

Seamless navigation through combined results of set-based and network-based enrichment analysis

**Bioconductor version: Release (3.2)**

The EnrichmentBrowser package implements essential functionality for the enrichment analysis of expression data. The analysis combines the advantages of set-based and network-based enrichment analysis in order to derive high-confidence gene sets and biological pathways that are differentially regulated in the expression data under investigation. Besides, the package facilitates the visualisation and exploration of such sets and pathways.

Author: Ludwig Geistlinger, Gergely Csaba, Ralf Zimmer

Maintainer: Ludwig Geistlinger <Ludwig.Geistlinger at bioifi.lmu.de>

Citation (from within R, enter citation("EnrichmentBrowser")):


<table>
<thead>
<tr>
<th>Month</th>
<th>Nb of distinct IPs</th>
<th>Nb of downloads</th>
</tr>
</thead>
<tbody>
<tr>
<td>May/2015</td>
<td>132</td>
<td>190</td>
</tr>
<tr>
<td>Jun/2015</td>
<td>140</td>
<td>220</td>
</tr>
<tr>
<td>Jul/2015</td>
<td>147</td>
<td>226</td>
</tr>
<tr>
<td>Aug/2015</td>
<td>125</td>
<td>172</td>
</tr>
<tr>
<td>Sep/2015</td>
<td>124</td>
<td>200</td>
</tr>
<tr>
<td>Oct/2015</td>
<td>152</td>
<td>287</td>
</tr>
<tr>
<td>Nov/2015</td>
<td>125</td>
<td>205</td>
</tr>
<tr>
<td>Dec/2015</td>
<td>148</td>
<td>276</td>
</tr>
<tr>
<td>Jan/2016</td>
<td>196</td>
<td>339</td>
</tr>
<tr>
<td>Feb/2016</td>
<td>186</td>
<td>362</td>
</tr>
<tr>
<td>Mar/2016</td>
<td>210</td>
<td>345</td>
</tr>
<tr>
<td>Apr/2016</td>
<td>280</td>
<td>399</td>
</tr>
<tr>
<td>All months</td>
<td><strong>1535</strong></td>
<td><strong>3221</strong></td>
</tr>
</tbody>
</table>

**The following locations:**

- **Support site** - for questions about Bioconductor packages
- **BioC-devel** mailing list - for package developers
Intro: EnrichmentBrowser

- Implements an end-to-end gene expression analysis pipeline

- Preparing the data
- Preprocessing the data
- Differential expression (DE) analysis
- Defining gene sets and networks of interest
- Executing individual enrichment methods
- Combining results of different methods
- Visualization and exploration of results
A primer on gene set enrichment analysis

• GO/KEGG overrepresentation
• Dozens of methods published, partly available (in BioC)
• Development of additional methods ongoing
• Lack of standards for method evaluation
  – No „best“ method
  – Existing methods have never been compared consistently

→ No easy solution
→ Pragmatic approach: enable comparative exec of methods
→ Detection of gene sets with evidence across methods
Why combining results!? 

- gene sets reported by multiple methods 
- downgrades spurious hits of individual methods
Contents

1  Gene expression-based enrichment analysis
2  Genomic region enrichment analysis
3  Multi-omics enrichment analysis
ENCODE (Encyclopedia of DNA Elements)
Setup: genomic region enrichment

- Experimentally-derived set of genomic regions
  - CNVs, ChIP-seq peaks, open chromatin, ...

- Functional feature set of genomic regions
  - Genes, QTLs, ...

- Enrichment of functional features in experimentally-derived regions?

Packages: regioneR, LOLA, rGREAT, ...
Contents

1 Gene expression-based enrichment analysis
2 Genomic region enrichment analysis
3 Multi-omics enrichment analysis
TCGA (The Cancer Genome Atlas)

12 tumor types
- Leukemia (LAML)
- Lung adenocarcinoma (LUAD)
- Lung squamous (LUSC)
- Kidney (KIRK)
- Bladder (BLCA)
- Endometrial (UCEC)

- Glioblastoma (GBM)
- Head and neck (HNSC)
- Breast (BRCA)
- Ovarian (OV)
- Colon (COAD)
- Rectum (READ)

Oomics characterizations
- Mutation
- Copy number
- Gene expression
- DNA methylation
- MicroRNA
- RPPA
- Clinical data

Thematic pathways

Platforms

Samples

Genes/loci
Setup: multi-omics enrichment

- Multi-omics (enrichment) methods just begin to emerge

- Basic strategies
  - integrate & enrich vs. enrich & integrate

  ➔ Relationships between data layers are not well understood
    - Cooperative (mRNA / protein), inhibitory (miRNA / mRNA), ...

Packages: mogsa, RTopper, (...?)

Outside R/Bioc: PARADIGM, MONA