Experimental Data Analysis using Pathway Studio® Web for Mammal and Plant – An Introduction

Gene Expression • Proteomics • Metabolomics

Introduction to Experimental Analysis in Pathway Studio Web

The Gene Set Enrichment Analysis (GSEA) algorithm is a powerful analytical tool for interpreting experimental data to gain insights into biological mechanisms. This is the tool to use to answer the following questions:

- What know pathways are affected in my experiment?
- What cell processes or ontologies are enriched in my experiment?
- What regulators are implicated in my experiment?

The algorithm ranks experiment results by the absolute value of the fold change and identifies known gene sets (pathways and ontologies) that are statistically enriched based on this ranking.

Experimental Data

<table>
<thead>
<tr>
<th>Name</th>
<th>OA vs. Ctrl</th>
<th>OA vs. Ctrl p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEG1</td>
<td>0.5961</td>
<td>5.000e-01</td>
</tr>
<tr>
<td>AK1</td>
<td>0.5962</td>
<td>5.000e-02</td>
</tr>
<tr>
<td>CACNB4</td>
<td>0.6981</td>
<td>5.000e-01</td>
</tr>
<tr>
<td>GSK3B</td>
<td>1.6137</td>
<td>5.000e-01</td>
</tr>
<tr>
<td>PRKAP2</td>
<td>0.3284</td>
<td>5.000e-01</td>
</tr>
<tr>
<td>APOB2</td>
<td>0.3123</td>
<td>5.000e-01</td>
</tr>
<tr>
<td>SLCA1A1</td>
<td>0.3648</td>
<td>5.000e-01</td>
</tr>
<tr>
<td>TRIM31</td>
<td>-0.8307</td>
<td>5.000e-01</td>
</tr>
<tr>
<td>STK3</td>
<td>-0.4781</td>
<td>5.000e-01</td>
</tr>
<tr>
<td>DNAJC9</td>
<td>0.4823</td>
<td>5.000e-01</td>
</tr>
<tr>
<td>SHB2</td>
<td>1.1538</td>
<td>5.000e-01</td>
</tr>
<tr>
<td>DDR2</td>
<td>-0.3512</td>
<td>5.000e-01</td>
</tr>
<tr>
<td>KCNA62</td>
<td>0.1896</td>
<td>5.000e-01</td>
</tr>
<tr>
<td>STBD1</td>
<td>-0.6748</td>
<td>5.000e-01</td>
</tr>
<tr>
<td>FOSON</td>
<td>0.3473</td>
<td>5.000e-01</td>
</tr>
<tr>
<td>KDAC064</td>
<td>-0.1856</td>
<td>5.000e-01</td>
</tr>
<tr>
<td>BSG</td>
<td>-0.2755</td>
<td>5.000e-02</td>
</tr>
<tr>
<td>GPRML</td>
<td>0.8047</td>
<td>5.000e-01</td>
</tr>
<tr>
<td>ANK2</td>
<td>0.7261</td>
<td>5.000e-01</td>
</tr>
<tr>
<td>SEC238B</td>
<td>0.6773</td>
<td>5.000e-01</td>
</tr>
</tbody>
</table>

Algorithm

Enriched gene sets results

Mammal:
- Cell Process Pathways
- Disease Collections
- Expression Targets Pathways
- Immunological Pathways
- Metabolic Pathways
- Nociception Pathways
- Signaling Pathways
- Gene Ontology
- Pathway Studio Ontology
- Chromosomal localization

Plant:
- AraCyc
- Arabidopsis Signaling Pathways
- MaizeCyc
- RiceCyc
- Plant Ontology
- Pathway Studio Ontology
- Gene Ontology

De novo user-defined sub-networks (identifies the regulators)

The results of the analysis are a list of most significantly enriched pathway and ontologies or sub-networks ranked by p-value. You can examine this list to better understand the underlying mechanisms and cellular processes affected in your experimental conditions.

The gene set used in the Gene Set Enrichment Analysis in Pathway Studio consists of publically curated and Elsevier curated pathway collections and ontologies as well as sub-networks (calculated de novo) when using the Sub-Network Enrichment Analysis (SNEA) option. Elsevier will continue to add more pathways to these collections as the databases grow. Let’s take a look at each of these collections more closely.
Pathways collections provided in the Mammal database

- **Cell Process Pathways – 71 total** *(Example: Transcytosis)*
The Cell Process Pathways collection represents the major processes of a cell, including such areas as apoptosis, cell division, cellular contacts, chromatin remodeling, complement activation, cytoskeleton assembly, DNA repair, histone modification, mitochondrial processes, transcription and vesicular transport.

- **Disease Collections – 99 total** *(Includes both pathways and groups)* *(Example Acute Myeloid Leukemia)*
The Disease Collection includes both curated pathways and genes sets for commonly studied diseases.
• **Immunological Pathways** – 41 total *(Example: MHC1-Mediated Antigen Presentation)*
  Illustrate molecular events occurring in immune cells related to immune response.

• **Metabolic Pathways** – 78 total *(Example: Heme Biosynthesis)*
  The Metabolic Pathway collection represents common metabolic processes. The enzymes in the metabolic pathways are indicated by functional classes.
- **Nociception (pain perception) Pathways – 98 total** *(Example: GRM2-4/6-8 (presynaptic -> glutamate release attenuation))*

Illustrate molecular mechanisms of noxious stimuli perception, consisting of: nociceptors (pain receptors) signaling, neurotransmitter release/reuptake cycles in synapses, and summarized pathways for general events associated with pain perception.

- **Signaling Pathways – 258 total**

The Signaling Pathways collection includes two sections: Cell Signaling and Receptor Signaling. The Cell Signaling is represented mostly by functional classes and the pathways end with cell processes. The Receptor Signaling Pathways generally start with an external signal and usually end at the level of transcription factors.

  - **Cell Signaling** *(Example: Adipocytokine Signaling)*
- **Receptor Signaling** *(Example: Notch -> RBPJ/HES/HEY Signaling)*

- **Expression Targets Pathways – 387 total** *(available in the ChemEffect+DiseaseFx database only)* *(Example: Notch Expression Targets)*

  The Expression Targets Pathways are similar to the Receptor Signaling Pathways, but add additional information about the targets of the transcription factors, thus expanding the information in these pathways downstream an additional level. As some transcription factors can have many targets, these pathways may be large. The targets of the transcription factors in these pathways have thus been limited to the targets that are known to be regulated (indirectly) by the initial receptors.
• **Toxicity Pathways – 22 total** available in the ChemEffect+DiseaseFx database only
  (Example: Ethanol-Induced Hepatotoxicity)

The Toxicity pathways include both well characterized drug toxicity pathways as well as general mechanism toxicities.

Pathways collections provided in the Plant database
- **Metabolic Pathways:** AraCyc - 406 total, MaizeCyc – 434 total and RiceCyc – 383 total
  (Example: Proline Biosynthesis II from arginine)
- **Arabidopsis Signaling Pathways** – 28 total AraCyc, MaizeCyc and RiceCyc *(Example: Systemin Signaling)*

**Ontology Collections**

**Mammal Ontologies**
- **Pathway Studio Ontology** gene groups and **Gene Ontology** gene groups

  Pathway Studio Ontology sub-groups:
  - cellular process
  - molecular function
  - disease regulation
  - biomarkers

  Gene Ontology sub-groups:
  - cellular component
  - molecular function
  - biological process

The Pathway Studio curated ontology is an ontology that has been optimized for best results in enrichment analysis. The ontology is relatively flat and designed to minimize redundancy in classification. The ontology consists of slightly more than 500 ontology groups representing almost 9000 genes.

Gene Ontology, assembled by the Gene Ontology Consortium, consists of hierarchical controlled vocabularies (ontologies) that describe gene products in terms of their associated biological processes, cellular components and molecular functions.
Plant Ontologies

- **Plant Ontology, Pathway Studio Ontology** and **Gene Ontology** gene groups

Plant Ontology sub-groups:
  - plant structure development stage
  - plant anatomical entity

Pathway Studio Ontology sub-groups:
  - plant traits
  - protein classes

Gene Ontology sub-groups:
  - cellular component
  - molecular function
  - biological process


Gene Ontology, assembled by the Gene Ontology Consortium, consists of hierarchical controlled vocabularies (ontologies) that describe gene products in terms of their associated biological processes, cellular components and molecular functions.

Both enriched pathways and ontologies are identified when analyzing array data using Gene Set Enrichment Analysis.

*Example of Gene Set Enrichment Analysis Results*
Sub-Networks

The Sub-Network Enrichment Analysis (SNEA) in Pathway Studio is an extension of the Gene Set Enrichment Analysis where the “gene sets” used in the enrichment analysis are small regulatory networks calculated de novo from all the data in the ResNet database by the SNEA algorithm. The enriched sub-network results of this analysis include the identification of the regulators of the experiment.

A sub-networks consist of a regulator and its targets.

![Sub-network diagram]

This is an example of an expression sub-network, where the regulator has expression or promoter binding relationships with its downstream targets.

When analyzing gene expression analysis, the most commonly utilized sub-network is the “expression targets” sub-network which identifies expression regulators whose targets are enriched in your microarray results. This analysis can identify transcriptional regulators that are responsible for the gene expression profile in your experiment.

Running Gene Set Enrichment Analysis and Sub-Network Enrichment Analysis in Pathway Studio

Explore

After importing an experiment, use the “Analyze Experiment” option in the Tools menu to access the enrichment algorithms.

![Analyze Experiment window]

The Analyze Experiment window allows you to select:
- Gene Set Enrichment Analysis
- Sub-Network Enrichment Analysis
- Positional Gene Set Enrichment Analysis
If you choose:

**Gene Set Enrichment Analysis** - then chose the pathway and ontology gene set categories of interest to include in the analysis.

**Sub-Network Enrichment Analysis** – then chose the type of sub-network appropriate for your data set
Positional Gene Set Enrichment Analysis – (this identifies enriched gene clusters based on chromosomal localization) then select the appropriate species for your experimental data.

![Gene Set Categories: ○ Human chromosome position ○ Mouse chromosome position ○ Rat chromosome position](image)

The results of each analysis will be displayed in the list pane at the bottom of your screen, ranked by the most significant result (smallest $p$-value) at the top.

The pathways, ontologies and sub-networks in your results table provide you valuable information as to what mechanisms are being impacted in your experimental conditions. Results tables can be exported to MS Excel® to be saved for future reference.

More Information about using Pathway Studio Web

If you have any questions about Pathway Studio please contact Customer Care:

<table>
<thead>
<tr>
<th>Region</th>
<th>Contact Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA, Canada and Latin America:</td>
<td>(8am-8pm CET - St.Louis) Tel: US toll-free: +1 (888) 615 4500 Tel: Non toll-free: +1 (314) 523 4900 Email: <a href="mailto:usinfo@elsevier.com">usinfo@elsevier.com</a> Email Brazil: <a href="mailto:brinfo@elsevier.com">brinfo@elsevier.com</a></td>
</tr>
<tr>
<td>Europe, Middle East and Africa:</td>
<td>(9am-6pm GMT+1, Amsterdam office) Tel: +31 20 485 3767 Email: <a href="mailto:nlinfo@elsevier.com">nlinfo@elsevier.com</a></td>
</tr>
<tr>
<td>Japan:</td>
<td>(9,30am-5,30pm JST, Tokyo office) Tel: +81 (3) 5561 5035 Email: <a href="mailto:jpinfo@elsevier.com">jpinfo@elsevier.com</a> Website: japan.elsevier.com</td>
</tr>
<tr>
<td>Asia and Australasia:</td>
<td>(9am-6pm SST, Singapore office) Tel: +65 6349 0222 Email: <a href="mailto:sginfo@elsevier.com">sginfo@elsevier.com</a></td>
</tr>
</tbody>
</table>