INGENUITY

Understand complex 'omics data at multiple levels

Ester Feldmesser, Bioinformatics Unit Biological Services
IPA provides many kinds of biological insights
Quickly analyze your data to identify key insights with IPA

Transcriptomics
IPA can help with almost any transcriptomics-related question or application

Biomarker Discovery
Identifies the most promising and relevant biomarker candidates within experimental datasets

microRNA Research
Combines filtering tools and microRNA-mRNA content to provide insight into the biological effects of microRNAs

Toxicogenomics
Delivers a focused toxicity and safety assessment of candidate compounds, and provides a more complete understanding of pharmacological response, drug mechanism of action, and mechanism of toxicity

Metabolomics
Overcomes the metabolomics data analysis challenge by providing the critical context necessary to gain biological insight into cell physiology and metabolism from metabolite data

Drug Repositioning
Expression profiling of approved drugs and comparison to profiles of diseased tissue can lead to discovery of new uses for these already approved entities

Proteomics
Perform a comprehensive analysis of your proteomics for a deep understanding of proteins and related biological processes

Target Discovery
Genes that are shown to be activated in a pathological condition may serve as promising targets for therapeutic development efforts
HIGHLIGHTS

✓ Quickly analyze your experimental data to identify key insights

✓ Easily search the scientific literature and find insights most relevant to your experimental model or question

✓ Build dynamic pathway models to extend your understanding of your research systems
THE KEY COMPONENTS OF THE IPA CORE ANALYSIS ARE:

✓ Signaling and Metabolic Pathways Analysis
✓ Cellular and Disease Process Analysis
✓ Molecular Network Analysis
✓ Upstream (Transcription Factor) Analysis
✓ Regulator Analysis
✓ Evaluation of Downstream Effects
Supported Species:

IPA supports the upload and analysis of human, mouse, rat, and canine identifiers, plus chemical identifiers. Additionally, IPA supports analysis of molecular data for the following species through ortholog mapping of Entrez Gene IDs:

**Arabidopsis thaliana** (plant)
**Bos taurus** (bovine)
**Caenorhabditis elegans** (c. elegans)
**Canis lupus familiaris** (canine)
**Danio rerio** (zebrafish)
**Drosophila melanogaster** (fruit fly)
**Gallus gallus** (chicken)
**Macaca mulatta** (Rhesus Monkey)
**Pan troglodytes** (chimpanzee)
**Saccharomyces cerevisiae** (yeast)
**Schizosaccharo-myces pombe** (yeast)
What’s in IPA

The Ingenuity Knowledge Base includes modeled relationships between:

• Chemicals
• Proteins and molecular complexes
• Genes and mutations
• Cells, cellular components, cellular processes and tissues
• Drugs, diseases and clinical phenotypes

IPA includes relationships that have been manually curated from the literature (Ingenuity® Expert Findings) as well as information that has been manually reviewed and automatically extracted from the literature (Ingenuity® ExpertAssist Findings).

IPA also contains information that has been manually curated by Ingenuity scientists (Ingenuity® Expert Knowledge) and that describes:
• FDA approved drugs and clinical candidates
• Cell Signaling, Metabolic, and Disease Pathways
• Toxicity Lists and Pathways
• Predicted and experimentally demonstrated microRNA targets
IPA includes manually reviewed content from selected third party sources (Ingenuity® Supported Third Party Information), including:

Entrez Gene
RefSeq
OMIM
GWAS Database
Gene Ontology
Human Metabolome Database (HMDB)
GNF Tissue Expression Body Atlas
NCI-60 Cell Line Expression Atlas
KEGG metabolic pathway information
LIGAND enzyme/substrate reactions
BIND, DIP, MINT, MIPS, BIOGRID, INTACT, COGNIA protein-protein interactions
TarBase
TargetScan
miRecords
Clinicaltrials.gov
Drugs@FDA
Mosby’s Drug Consult
Goodman & Gilman’s Pharmacological Basis of Therapeutics
DrugBank
Hazardous Substance Database (HSDB)
Chemical Carcinogenesis Research Information System database (CCRIS)
Content Updates
(December 2015)

• ~94,000 new findings (total of ~5.4 million findings), including:
• ~37,000 new Expert findings
• ~700 new ExpertAssit findings
• ~39,000 new cancer mutation disease association findings from COSMIC
• ~31,000 new protein-protein interactions from the BioGRID database
• ~1000 new disease target findings from ClinicalTrials.gov
• ~700 new mouse knockout-to-phenotype findings in Mouse Genome Database (JAX Labs)
• ~300 new gene to disease findings from OMIM
• ~200 new protein-protein interactions from the intAct database
Using Enrichment Analysis in Practice

- statistical measure
  - how likely your differentially regulated genes fall into that category by chance

![Bar chart showing enrichment analysis results]

- microarray
  - 1000 genes

- experiment

- 100 genes differentially regulated

- mitosis – 80/100
  - apoptosis – 40/100
  - p. ctrl. cell prol. – 30/100
  - glucose transp. – 20/100
Using Enrichment Analysis in Practice

- Looking at the distribution of all genes on the microarray:

<table>
<thead>
<tr>
<th>Process</th>
<th>Genes on array</th>
<th># genes expected in 100 random genes</th>
<th>occurred</th>
</tr>
</thead>
<tbody>
<tr>
<td>mitosis</td>
<td>800/1000</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>apoptosis</td>
<td>400/1000</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>p. ctrl. cell prol.</td>
<td>100/1000</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>glucose transp.</td>
<td>50/1000</td>
<td>5</td>
<td>20</td>
</tr>
</tbody>
</table>
# Analysis Summary

## Top Canonical Pathways

<table>
<thead>
<tr>
<th>Name</th>
<th>p-value</th>
<th>Overlap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antigen Presentation Pathway</td>
<td>9.40E-14</td>
<td>32.4 % 12/87</td>
</tr>
<tr>
<td>Communication between Innate and Adaptive Immune Cells</td>
<td>5.33E-11</td>
<td>15.4 % 14/91</td>
</tr>
<tr>
<td>Graft-versus-Host Disease Signaling</td>
<td>1.44E-09</td>
<td>20.8 % 10/48</td>
</tr>
<tr>
<td>Interferon Signaling</td>
<td>2.35E-08</td>
<td>23.5 % 8/34</td>
</tr>
<tr>
<td>Activation of IRF by Cytosolic Pattern Recognition Receptors</td>
<td>2.71E-08</td>
<td>15.6 % 10/64</td>
</tr>
</tbody>
</table>

## Top Upstream Regulators

<table>
<thead>
<tr>
<th>Upstream Regulator</th>
<th>p-value of overlap</th>
<th>Predicted Activation</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFNA2</td>
<td>1.04E-05</td>
<td>Activated</td>
</tr>
<tr>
<td>IFNL1</td>
<td>1.04E-05</td>
<td>Activated</td>
</tr>
<tr>
<td>IFNG</td>
<td>4.37E-05</td>
<td>Activated</td>
</tr>
<tr>
<td>IRF</td>
<td>5.09E-05</td>
<td>Activated</td>
</tr>
<tr>
<td>MAPK1</td>
<td>1.21E-02</td>
<td>Inhibited</td>
</tr>
</tbody>
</table>

## Top Diseases and Bio Functions

<table>
<thead>
<tr>
<th>Disease</th>
<th>p-value range</th>
<th># Molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immunological Disease</td>
<td>1.76E-04 - 3.70E-29</td>
<td>113</td>
</tr>
<tr>
<td>Endocrine System Disorders</td>
<td>5.06E-04 - 6.42E-29</td>
<td>84</td>
</tr>
<tr>
<td>Gastrointestinal Disease</td>
<td>5.07E-04 - 6.42E-29</td>
<td>95</td>
</tr>
<tr>
<td>Metabolic Disease</td>
<td>1.43E-20 - 6.42E-29</td>
<td>78</td>
</tr>
<tr>
<td>Dermatological Diseases and Conditions</td>
<td>9.12E-05 - 9.87E-23</td>
<td>69</td>
</tr>
</tbody>
</table>

## Molecular and Cellular Functions

<table>
<thead>
<tr>
<th>Function</th>
<th>p-value range</th>
<th># Molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cellular Function and Maintenance</td>
<td>3.29E-04 - 3.34E-13</td>
<td>73</td>
</tr>
<tr>
<td>Cellular Development</td>
<td>5.67E-04 - 5.70E-13</td>
<td>113</td>
</tr>
<tr>
<td>Cellular Growth and Proliferation</td>
<td>4.58E-04 - 5.70E-12</td>
<td>120</td>
</tr>
<tr>
<td>Cell Signaling</td>
<td>4.34E-04 - 8.22E-12</td>
<td>47</td>
</tr>
<tr>
<td>Cell-To-Cell Signaling and Interaction</td>
<td>5.87E-04 - 6.71E-12</td>
<td>81</td>
</tr>
</tbody>
</table>

## Physiological System Development and Function

<table>
<thead>
<tr>
<th>Function</th>
<th>p-value range</th>
<th># Molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organismic Survival</td>
<td>9.48E-07 - 8.14E-14</td>
<td>44</td>
</tr>
<tr>
<td>Hematological System Development and Function</td>
<td>5.87E-04 - 4.39E-13</td>
<td>102</td>
</tr>
<tr>
<td>Tissue Morphology</td>
<td>4.39E-04 - 3.99E-13</td>
<td>71</td>
</tr>
<tr>
<td>Immune Cell Trafficking</td>
<td>5.87E-04 - 1.16E-10</td>
<td>61</td>
</tr>
<tr>
<td>Hematopoiesis</td>
<td>5.87E-04 - 4.17E-08</td>
<td>53</td>
</tr>
</tbody>
</table>
Networks generation

• Chosen genes are combined into networks that maximize their specific connectivity, according to pre-defined relationships in the Ingenuity Knowledge Base.

• Additional molecules from the database are used to specifically connect two or more smaller networks by merging them into a larger one.
Example of the calculation of specific connectivity

Specific connectivity = \frac{\text{Number of genes in intersection of the neighborhood and network}}{\text{Number of genes in union of neighborhood and network}}
View of a network
Cannonical pathways

• More than 600 predefined pathways
  – HumanCyc Metabolic Pathways
  – Signaling Pathways
Downstream Effects Analysis
The purpose of the regulation z-score is to identify increased or decreased biological functions that are implicated by the observed gene expression changes.
More Downstream Effects Analysis
Upstream Regulator Analysis

- The upstream regulator analysis is based on prior knowledge of expected effects between transcriptional regulators and their target genes stored in the Ingenuity® Knowledge Base.

- The analysis examines how many known targets of each transcription regulator are present in the user’s dataset, and also compares their direction of change to what is expected from the literature in order to predict likely relevant transcriptional regulators.

- IPA’s definition of upstream transcriptional regulator is quite broad – any molecule that can affect the expression of other molecules, which means that upstream regulators can be almost any type of molecule, from transcription factor, to microRNA, kinase, compound or drug.
The primary purpose of the activation z-score is to infer the activation states of predicted transcriptional regulators.
Regulator Effectors

- Generate a hypothesis for how a phenotype, function or disease is regulated in your dataset by activated or inhibited upstream regulators
- Explain the biological impact of upstream molecules and the underlying mechanism for a phenotype
- The algorithm goes through one or more iterations to merge upstream and downstream results from the Upstream Regulator Analysis tab and Downstream Effects Analysis tabs
How does the Regulator Effects algorithm work?

1. **Upstream Regulator Analysis**
   - Hypotheses for how activated or inhibited upstream regulators cause downstream effects on biology.

2. **Downstream Effects Analysis**
   - First iteration: Targets in the dataset.
   - Second iteration: Disease or Function.
   - Displays a relationship between the regulator and disease/function if it exists.

3. **Upstream Regulator Analysis**
   - Algorithm: Next iteration.

4. **Downstream Effects Analysis**
   - Disease or Function "A" and Disease or Function "B".
How does the Regulator Effects algorithm work?

Consistency Score = \( \frac{P_c \cdot W_c + P_i \cdot W_i + P_n \cdot W_n}{(S)^{W_s}} \)

Definitions of terms in the formula:

- \( P_c \): the total number of consistent paths from regulator to function (through dataset targets)
- \( W_c \): the weight that rewards consistent paths and that weight is equal to +1
- \( P_i \): the total number of inconsistent paths
- \( W_i \): the weight that penalizes inconsistent paths and that penalty is equal to -15
- \( P_n \): the total number of non-causal paths
- \( W_n \): the weight for non-causal paths. Set to 0 currently (e.g., non-causal paths don’t affect the score).
- \( S \): the size (the total number of dataset targets)
- \( W_s \): a penalty weight for the size and weight is equal to 0.5
Custom filters help you identify relevant information
Gain more insight by customizing pathway and network overlays

Now you can choose which measurement to overlay on networks and pathways from your dataset or analysis, and then quickly customize the range of color to emphasize the genes of interest.
Comparison Analysis

Quickly visualize canonical pathway scores across dose, time, or other condition using the new Comparison Analysis heat map. Prioritize by score, hierarchical cluster, or trend.
Which microRNA is predicted to target a given mRNA, and how good is the prediction?

Based on my expression data, which microRNAs have regulation that supports the target prediction?

Which mRNAs participate in a relevant disease, subcellular location, or pathway?

How do certain mRNAs and microRNAs interact, and what’s downstream?

What is the predicted impact of changes in microRNA expression on cellular processes, pathways, diseases, and phenotypes?
Pro-angiogenic Genes and microRNA deregulated in Ovarian Cancer

- MIRNLET7F2
- MIRN21
- VEGF
- LGALS3
- KDR
- pazopanib
- BCL2
- TNFRSF1A
- MYC
- HDAC1
Path Designer 2
Batch Upload

IPA will read, parse and upload all dataset files in this directory/folder C:\Users\jaykumar\Documents\IPA\IPA Training Datasets\Prostate Disease.
Once a dataset file is successfully uploaded, it will be moved to subdirectory Successfully Uploaded 2013-03-21 08:34:23.
Sometimes moving the dataset file to this directory/folder might not succeed. Please move such datasets manually.
In general, batch upload works better when the datasets are stored on your local hard drive rather than remotely.
A step by step log of batch upload will also be written to a log file in this directory/folder.
Note that there may be limits to how many datasets you are allowed to upload on a daily and monthly basis. When your allocation for that period is used up, the dataset upload will stop.

Parsing file 'batch upload dataset 3.xla'
Parsing complete. File has 976 rows and 4 columns
Guessing identifier types
Guessed identifier types=GenBank
Validating data in file.
685 rows mapped.
Number of observations=3
Expression types=Fold Change
Validation complete. Uploading file.
Upload complete.

Parsing file 'batch upload dataset 2.xla'
Parsing complete. File has 976 rows and 4 columns
Guessing identifier types
Guessed identifier types=GenBank
Validating data in file.
685 rows mapped.
Number of observations=3
Expression types=Fold Change
Validation complete. Uploading file.
Upload complete.

Parsing file 'batch upload dataset 1.xla'
Parsing complete. File has 976 rows and 4 columns
Guessing identifier types
Guessed identifier types=GenBank
Validating data in file.
685 rows mapped.
Number of observations=3
Expression types=Fold Change
Validation complete. Uploading file.