Canadian Bioinformatics Workshops

www.bioinformatics.ca

In collaboration with
Cold Spring Harbor Laboratory
&
New York Genome Center
Module 3 Lab
Cytoscape & Enrichment Map

Jüri Reimand
High-throughput Biology: From Sequence to Networks
April 27-May 3, 2015
Network Analysis Workflow

- Collect genomics data (e.g. mRNA expression)
- Normalize and score (e.g. compute differential expression)
- Generate gene list
- Learn about underlying cellular mechanism using pathway and network analysis
- Visualize and identify interesting pathways and networks
- Drill down to understand molecular mechanism
- Publish model explaining data

A specific example of this workflow:

Pathways vs Networks

- Detailed, high-confidence consensus
- Biochemical reactions
- Small-scale, fewer genes
- Concentrated from decades of literature

- Simplified cellular logic, noisy
- Abstractions: directed, undirected
- Large-scale, genome-wide
- Constructed from omics data integration

Module 3 Lab: Cytoscape and Enrichment Map
bioinformatics.ca
Networks

- Represent relationships of biological molecules
  - Physical, regulatory, genetic, functional interactions
- Useful for discovering relationships in large data sets
  - Better than tables in Excel
- Visualize multiple data types together
  - Discover interesting patterns
- Network analysis
  - Finding sub-networks with certain properties (densely connected, co-expressed, frequently mutated, clinical characteristics)
  - Finding paths between nodes (or other network “motifs”)
  - Finding central nodes in network topology (“hub” genes)

Network basics (i): nodes and edges

A simple mapping
- one molecule - node
- one interaction - edge
A more realistic mapping
- Cell localization
- cell cycle
- cell type
- taxonomy
- Physiologically relevant
- Edges - diverse relationships

Critical: what do nodes and edges mean?
Network basics (i): nodes and edges

A simple mapping
- one molecule - node
- one interaction - edge

A more realistic mapping
- Cell localization
- cell cycle
- cell type
- taxonomy
- Physiologically relevant
- Edges - diverse relationships

Critical: what do nodes and edges mean?

Network Representations

Module 3 Lab: Cytoscape and Enrichment Map

bioinformatics.ca
Network basics (ii) - layout and visual attributes

Use layout to interpret network
Local relationships
Guilt-by-association
Dense clusters
Global relationships

Visualise multiple types of data

Automatic network layout

Module 3 Lab: Cytoscape and Enrichment Map
bioinformatics.ca
Force-directed network layout

- Force-directed: nodes *repel* and edges *pull*
- Good for up to 500 nodes
  - Bigger networks give hairballs - reduce number of edges
- Advice: try force directed first, or hierarchical for tree-like networks
- Tips for better looking networks
  - Manually adjust layout
  - Load network into a drawing program (e.g. Illustrator) and adjust labels

force directed layout
http://sydney.edu.au/engineering/it/~aquigley/avi/spring.avi
Dealing with ‘hairballs’: zoom or filter

Focus

PKC
Cell
Wall
Integrity

Zoom

Synthetic Lethal
Transcription Factor Regulation
Protein-Protein Interaction

- Up Regulated Gene Expression
- Down Regulated Gene Expression

Visual Features

- Node and edge attributes
  - Represent properties of genes, interactions
  - Text (string), integer, float, Boolean, list
- Visual attributes
  - Node, edge visual properties
  - Colour, shape, size, borders, opacity...

Module 3 Lab: Cytoscape and Enrichment Map

bioinformatics.ca
What Have We Learned?

- Networks help discover relationships in large data sets
- Networks help integrate several datasets or types
- Important to understand meaning of nodes and edges
- Avoid hairballs by focusing analysis
- Automatic layout is required to visualize networks
- Visual attributes enable multiple types of data to be shown at once – useful to see their relationships

Cytoscape - Network Visualization

Cytoscape is

- an open source software platform
- for visualizing complex networks and integrating these with any type of attribute data.
- a lot of apps are available for various kinds of problem domains, including bioinformatics, social network analysis, and semantic web.
Manipulate Networks
Filter/Query
Automatic Layout
Interaction Database Search

The Cytoscape App Store
http://apps.cytoscape.org

Pathway analysis
Gene expression analysis
Complex detection
Literature mining
Network motif search
Pathway comparison
Introduction to Cytoscape (3.1.0)

- Control panel
- Results panel
- Table panel

save your session
save image

Navigate through the network (3.1.0)

move the blue square to navigate through the network
Try different Cytoscape layouts (3.1.0)

Circular Layout

yFiles Circular
Module 3 Lab: Cytoscape and Enrichment Map

Change visual Features (3.1.0)

- Colors
- Shapes
- Sizes
- Brushes
- Transparency
Load “Your Favorite Network”
File > Import > Network > File …

Load “Your Favorite Expression”
Dataset
File > Import > Table > File …
Map expression values to node colours using a continuous mapper

Expression data mapped to node colours
Fine-tuning network layout

- Move, zoom/pan, rotate, scale, align

Create Subnetwork
Create Subnetwork

Module 3 Lab: Cytoscape and Enrichment Map

Network Filtering

Select nodes with at least 20 interactions

Module 3 Lab: Cytoscape and Enrichment Map
Query the BioGrid database for interactions of the cancer gene KRAS

**What have we learned**

- Cytoscape is a useful, free software tool for network visualisation and analysis
- Provides basic network features
  - Automated layouts
  - Mapping of genomic attributes to visual attributes
- Apps are available to extend functionality to diverse analyses
Enrichment Map

- A Cytoscape app to visualize and interpret results of pathway enrichment analysis
- Enriched pathways are visualized as networks
- Edges connect pathways with many shared genes
Creating an Enrichment Map with data from g:Profiler

2. Select and copy all genes in the tutorial file
   - `MCF7_24hr_topgenes.txt` in the Query box
3. In Options, check Significant only, No electronic GO annotations
4. Set the Output type to Generic Enrichment Map (TAB)
5. Show advanced options
6. Set Max size of functional category to 1000 and Min size to 5.
7. Set Min size of gene list and functional category overlap Q&T to 2.
8. Set Significance threshold to Benjamini-Hochberg FDR.
9. Choose GO biological process, molecular function, KEGG and Reactome from the color legend.
10. Download g:Profiler data as gmt: name
11. Click on g:Profiler! to run the analysis
12. Download the result file: Download data in Generic Enrichment Map (GEM) format
**g:Profiler files for Enrichment Map**

- Two files are required:
  
  A. Tabular text file with enriched processes & pathways
  
  B. GMT file with all processes & pathways and associated genes

**TAB file – Enrichment Map**

- Process ID
- Process name
- P-value, FDR
- Phenotype (positive 1; negative -1)

**Genes common to input list and process/pathway**
 GMT file – Enrichment Map
http://biit.cs.ut.ee/gprofiler/gmt/gprofiler_hsapiens.NAME.gmt.zip

Getting the Enrichment Map app

A. Cytoscape App Store
http://apps.cytoscape.org/apps/enrichmentmap

B. In Cytoscape software
Apps > App manager > Search > Install
### Constructing an Enrichment Map

A. Set input file format to “Generic” (format of g:Profiler results)

B. Select GMT file from file system (this file contains gene lists of pathways)

C. Select enrichments file from file system

D. Set equal P-value and Q-value cutoffs (because g:Profiler only provides corrected P-values=Q-values)

Click Build!

### Enrichment Map – a network of pathways

- **Nodes** – pathways, processes, functions
- **Edges** – between pathways with many shared genes

- Node size – genes in pathway
- Node color – enrichment strength
- Edge weight – genes shared
Enrichment Map – a network of pathways

What are major functional themes?

Module 3 Lab: Cytoscape and Enrichment Map

Edge definition determines granularity

- Edges between pathways defined by % of common genes
- Increasing edge stringency reveals finer granularity of functional themes
Mapping attributes to network
Node labels mapped to pathway IDs by default

Pathway IDs

Pathway names

Compare two pathway enrichment analyses

g:Profiler analyses of MCF7_12hr_topgenes.txt and MCF7_24hr_topgenes.txt

1st dataset maps to node fillings
2nd dataset maps to node edges
Mapping further gene sets to Enrichment Map

E.g. which processes associate to known cancer genes?

B. Select GMT file from file system
(this file contains gene lists used in original Enrichment Map)

B. Select another GMT file from file system
(this file contains gene lists mapped on top of original Enrichment Map)

Post-Analysis: New edges link further gene lists to existing pathways

Purple edges – pathway genes in this analysis are significantly enriched in known cancer genes
What have we learned?

- Enrichment map uses network visualization to summarize results of pathway enrichment analysis
- Edges connect pathways with shared genes, and edge stringency determines granularity of maps
- Two sets of pathways can be compared
- Post-analysis links further gene sets to existing map

Tips for publishable Enrichment Maps

- Manually curate clusters of connected pathways as “functional themes”
  - e.g. by picking most general process/pathway in the group
  - Check genes involved
- Review small groups and singleton nodes
  - Many are probably part of larger groups – safe to remove
  - Some interesting singletons may be the only representatives!
- Assign further visual attributes using your omics data
- Export as PDF and use graphics software (AI) for finalizing
- Sometimes less is more
Cytoscape Tips & Tricks

• Network views
  – When you open a large network, no view is shown by default
  – To improve interactive performance, Cytoscape has the concept of “Levels of Detail”
    • Some visual attributes will only be apparent when you zoom in
    • The levels for various attributes can be changed in the preferences
    • To see what things will look like at full detail:
      – View ➔ Show Graphics Details

Cytoscape Tips & Tricks

• Sessions
  – Sessions save pretty much everything:
    • Networks
    • Properties
    • Visual styles
    • Screen sizes
  – Saving a session on a large screen may require some resizing when opened on your laptop
Cytoscape Tips & Tricks

• Memory
  – Cytoscape uses lots of it
  – Doesn’t like to let go of it
  – An occasional restart when working with large networks helps
  – Destroy views when you don’t need them
  – Java doesn’t have a good way to get the memory right at start
    • Since version 2.7, Cytoscape does a much better job at “guessing” good
      default memory sizes than previous versions

Cytoscape Tips & Tricks

• CytoscapeConfiguration directory
  – In your user/home directory
  – Your defaults and any apps downloaded from the app
    store will go here
  – Sometimes, if things get really messed up, deleting (or
    renaming) this directory can give you a “clean slate”
We are on a Coffee Break & Networking Session