

Integrative Network Analysis using R and RCytoscape

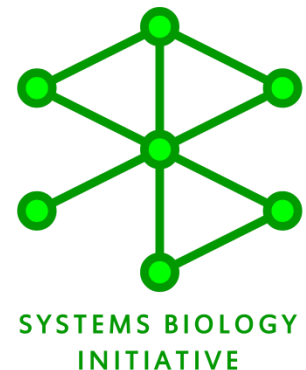
Ignatius Pang, Apurv Goel, Marc
Wilkins

10th May 2017

Sydney Users of R (SURF) Meetup
SMSA Townhall



i.pang (at) unsw.edu.au

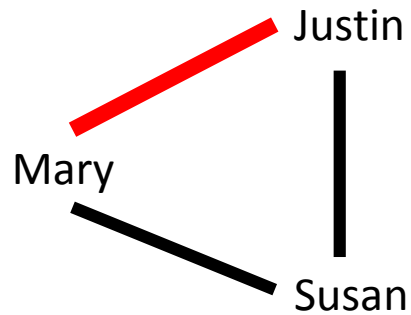


Outline

- Why is network being used?
- R and RCytoscape
- What are negative genetic interactions?
- Applications of negative genetic interactions
- Investigating the network basis of genetic interactions
- Results & Code Examples
- Summary

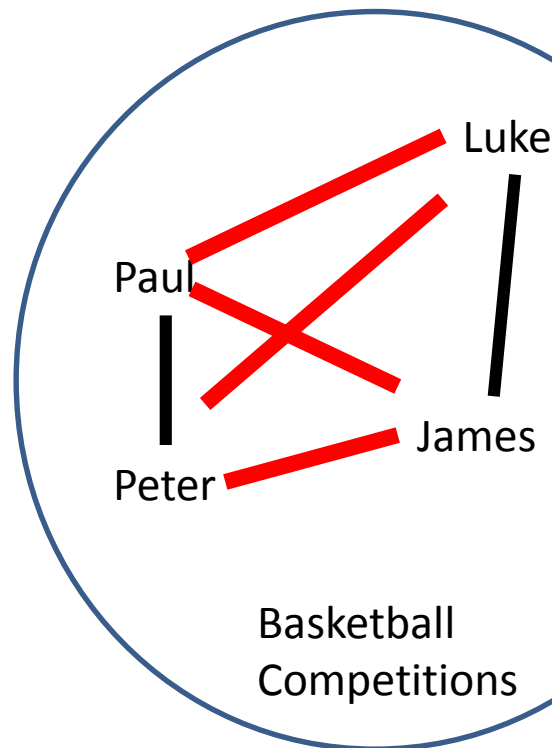
Where is network being used?

- Friends recommendation on facebook



Friends of Friends

Red = relationships
predicted/suggested by Facebook



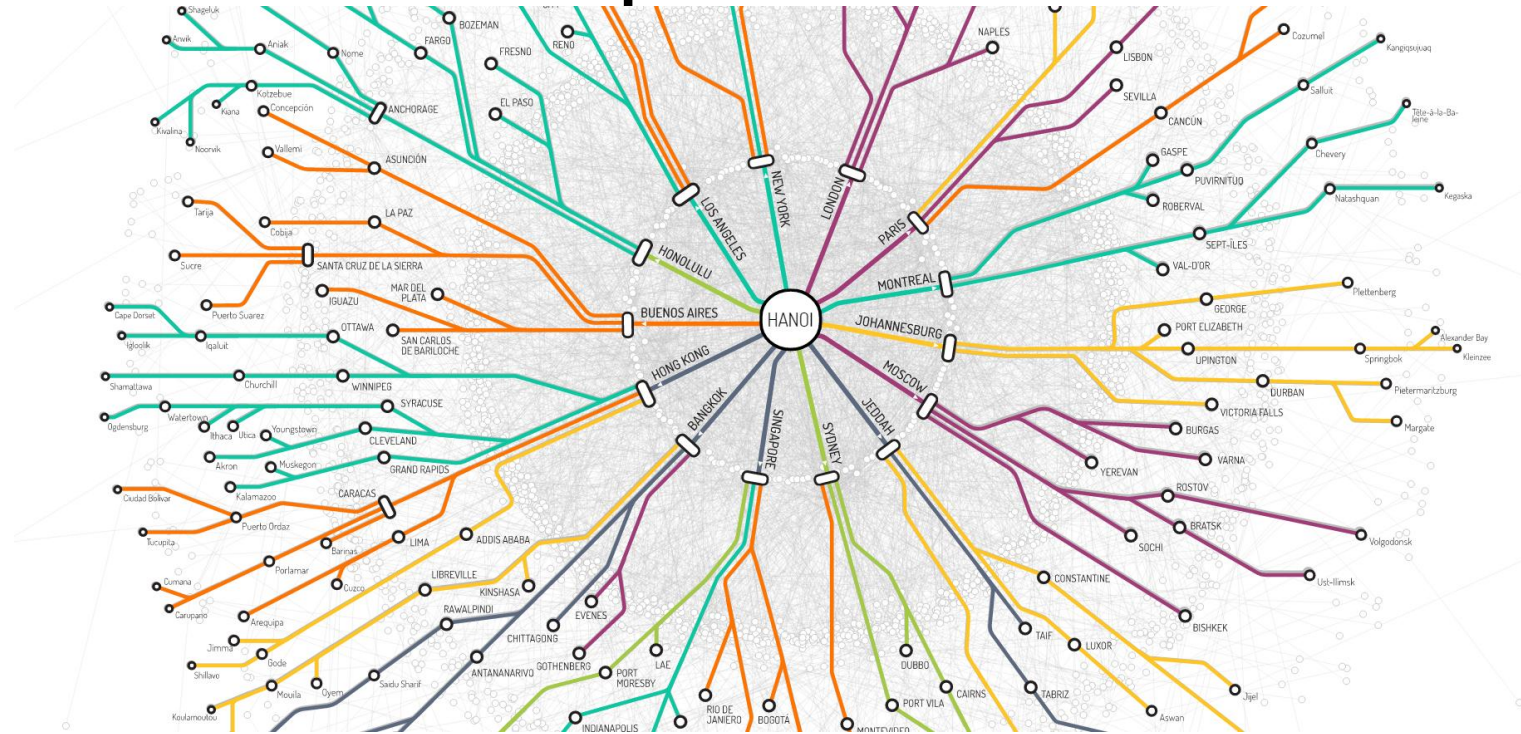
Basketball
Competitions

Same Social Group



Also suggest
advertisement
for Basketball
match tickets /
memorabilia

Modelling the Spread of Viruses in an Epidemic

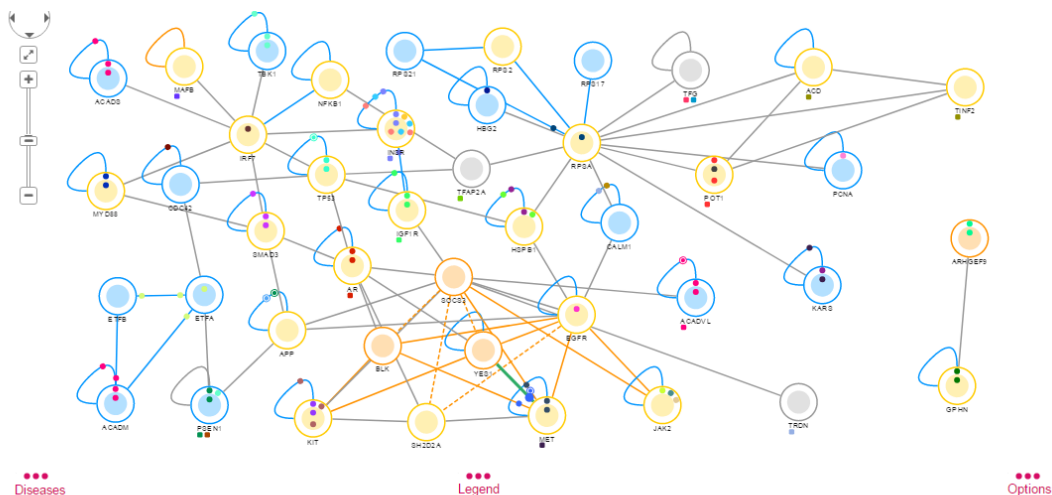


EPIDEMIC RAPID TRANSIT MAP

<http://www.mobs-lab.org/epi-rail.html>

Network of disease related mutations for proteins **ACADM, ACADS, ACADVL, ETF, ETFB...** [43 proteins and 98 interactions]

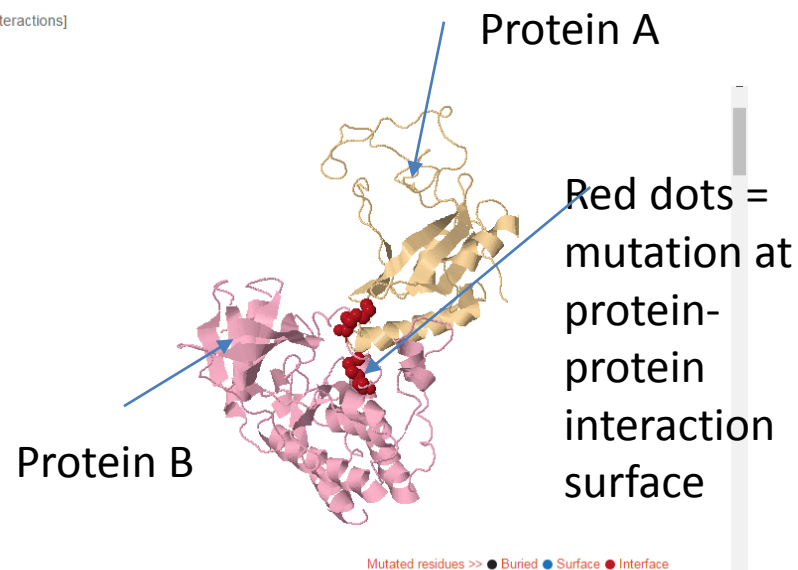
Reload the network including all neighbors ?



● Acyl-CoA dehydrogenase deficiency

An autosomal recessive metabolic disorder associated with mitochondrial complex I deficiency, resulting in multisystemic and variable manifestations. Clinical features include infantile onset of acute metabolic acidosis, Reye-like episodes (brain edema and vomiting that may rapidly progress to seizures, coma and death), exercise intolerance, hypertrophic cardiomyopathy, liver failure, muscle weakness, and neurologic dysfunction.

Phenotype	Protein	Mutations (● Buried ● Surf. ● Int. ○ Not Classified)
Acyl-CoA dehydrogenase medium-chain deficiency	ACADM Medium-chain specific acyl-CoA dehydrogenase... P11310	● I78T ● C116Y ● M149I ● G195R ● C244R ● M326T ● Y67H ● T121I ● T193A ● R206L ● S245L ● G267R ● R281T ● R53C ● G310R ● K328E ● S336R ● Y352C ● I375T
Acyl-CoA dehydrogenase short-chain deficiency	ACADS Short-chain specific acyl-CoA dehydrogenase, m... P16219	● S353L ● R46W ● G90S ● G92C ● R107C ● W177R ● A192V ● R325W ● R380W ● R383C
Acyl-CoA dehydrogenase very long-chain deficiency	ACADVL Very long-chain specific acyl-CoA dehydrogenase... P49748	● Q159R ● V174M ● G185S ● A213P ● L243R ● K247E ● K247T ● T260M ● A281D ● V283A ● Y317A ● M352V ● G441D ● F458L ● G290D ● G284E ● D405H ● R450H



SOCS3 :: O14543

MET :: P08581

Source

Intact

Detection method

Fluorescence polarization spe...

Publication

24728074

Ext. Ref.

EBI-9473568

Structure

Protein

Interface mutations

View

Model from 4gl9_H-B

MET

K1244R

M1250I

M1250T

Protein

Templ.

Seq. id.

Cov

SOCS3

4gl9_H

76.7%

72.4% (22-184)

MET

4gl9_B

35%

18.9% (1078-1340)

Download

Interactome3D

Disease

Protein

Interface mutations

Hepatocellular carcinoma

MET

K1244R

M1250I

Renal cell carcinoma papillary

MET

M1250T

<http://dsysmap.irbbarcelona.org/>

Why would you want to drive Cytoscape using R scripts?

- Animate your network, make the network change dynamically, visualize time series data
- Support reproducibility, sharing of methods via code
- Another way of developing applications, without the difficulty of developing Cytoscape apps in Java
- No need to click buttons

Installation Guide – How to get R and Cytoscape to talk to each other?

Step 1: Install Java 8

The screenshot shows the Oracle Technology Network website. The main navigation bar includes links for Sign In/Register, Help, Country, Communities, I am a..., I want to..., and a Search box. Below this is a secondary navigation bar with links for Products, Solutions, Downloads, Store, Support, Training, Partners, About, and OTN. The breadcrumb trail indicates the current location: Oracle Technology Network > Java > Java SE > Downloads.

On the left side, there is a sidebar menu with links to Java SE, Java EE, Java ME, Java SE Support, Java SE Advanced & Suite, Java Embedded, Java DB, Web Tier, Java Card, Java TV, New to Java, Community, and Java Magazine.

The main content area is titled "Java SE Runtime Environment 8 Downloads". It contains the following text:

Java SE Runtime Environment 8 Downloads
Do you want to run Java™ programs, or do you want to develop Java programs? If you want to run Java programs, but not develop them, download the Java Runtime Environment, or JRE™.

If you want to develop applications for Java, download the Java Development Kit, or JDK™. The JDK includes the JRE, so you do not have to download both separately.

JRE 8u131 Checksum

Below the text is a section titled "Java SE Runtime Environment 8u131" with a sub-header "You must accept the Oracle Binary Code License Agreement for Java SE to download this software." It includes two radio buttons: "Accept License Agreement" (selected) and "Decline License Agreement".

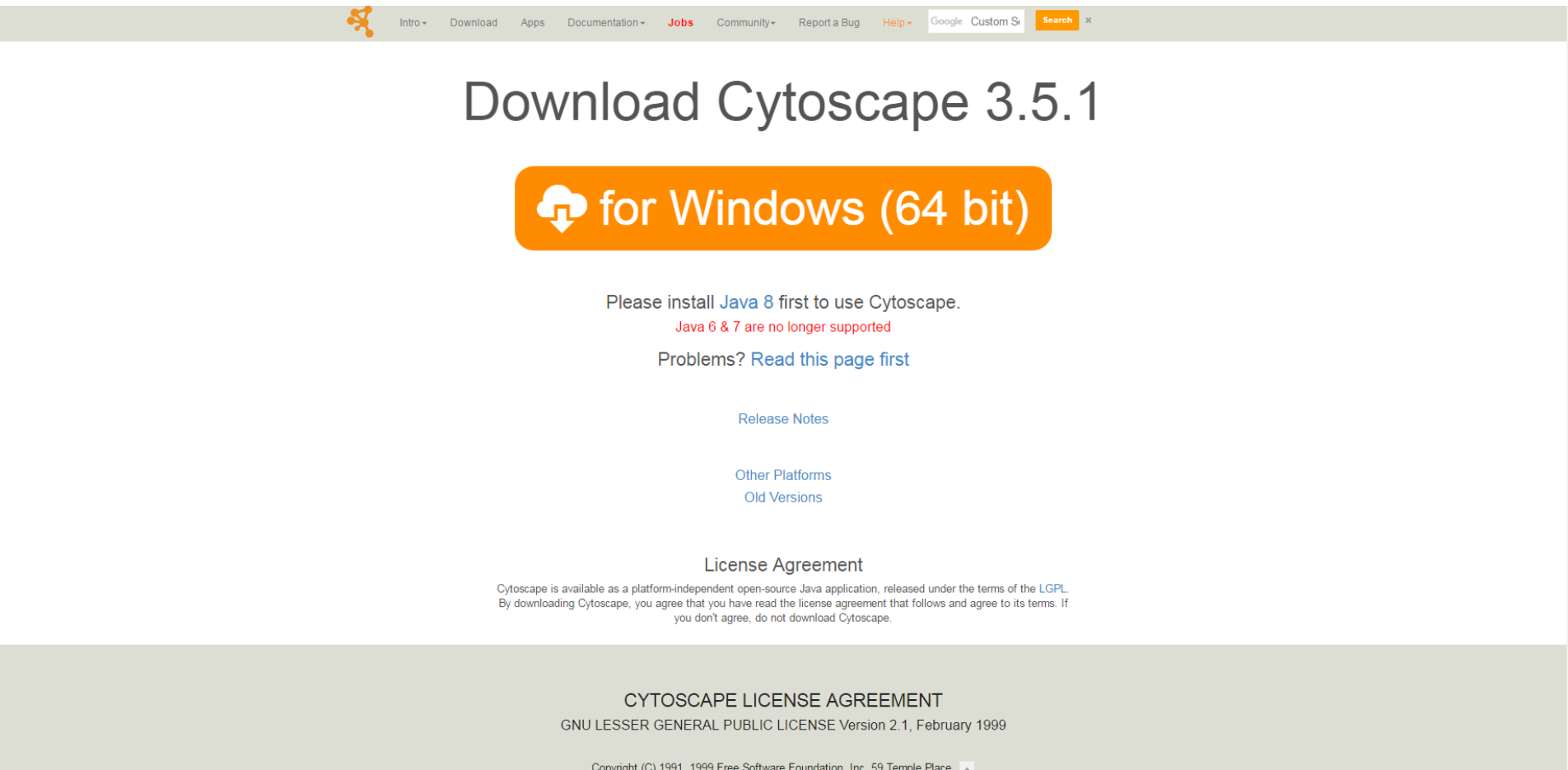
Below the license agreement is a table with three columns: Product / File Description, File Size, and Download.

Product / File Description	File Size	Download
Linux x86	59.13 MB	jre-8u131-linux-i586.rpm
Linux x86	74.98 MB	jre-8u131-linux-i586.tar.gz
Linux x64	56.47 MB	jre-8u131-linux-x64.rpm
Linux x64	72.4 MB	jre-8u131-linux-x64.tar.gz
Mac OS X	63.92 MB	jre-8u131-macosx-x64.dmg
Mac OS X	55.54 MB	jre-8u131-macosx-x64.tar.gz
Solaris SPARC 64-bit	52.05 MB	jre-8u131-solaris-sparcv9.tar.gz
Solaris x64	49.92 MB	jre-8u131-solaris-x64.tar.gz
Windows x86 Online	0.7 MB	jre-8u131-windows-i586-iftw.exe
Windows x86 Offline	54.83 MB	jre-8u131-windows-i586.exe
Windows x86	60.18 MB	jre-8u131-windows-i586.tar.gz
Windows x64 Offline	62.62 MB	jre-8u131-windows-x64.exe
Windows x64	63.97 MB	jre-8u131-windows-x64.tar.gz

On the right side, there are two sections: "Java SDKs and Tools" with links to Java SE, Java EE and Glassfish, Java ME, Java Card, NetBeans IDE, and Java Mission Control; and "Java Resources" with links to Java APIs, Technical Articles, Demos and Videos, Forums, Java Magazine, Java.net, Developer Training, Tutorials, and Java.com.

- <http://www.oracle.com/technetwork/java/javase/downloads/jre8-downloads-2133155.html>


Step 2: Install Cytoscape



The screenshot shows the Cytoscape website's download page. At the top is a navigation bar with links: Intro, Download, Apps, Documentation, Jobs, Community, Report a Bug, Help, and a search bar. The main heading is "Download Cytoscape 3.5.1". Below it is a large orange button with a cloud icon and the text "for Windows (64 bit)". Under the button, a message states: "Please install Java 8 first to use Cytoscape. Java 6 & 7 are no longer supported". Below this is a link "Problems? Read this page first". Further down are links for "Release Notes", "Other Platforms", and "Old Versions". A "License Agreement" section follows, stating that Cytoscape is available as a platform-independent open-source Java application under the terms of the LGPL, and that users agree to the license by downloading. At the bottom, the text "CYTOSCAPE LICENSE AGREEMENT" and "GNU LESSER GENERAL PUBLIC LICENSE Version 2.1, February 1999" is displayed, along with a small copyright notice for 1991-1999 Free Software Foundation, Inc.

Intro Download Apps Documentation Jobs Community Report a Bug Help Google Custom Search

Download Cytoscape 3.5.1

 for Windows (64 bit)

Please install [Java 8](#) first to use Cytoscape.
Java 6 & 7 are no longer supported

Problems? [Read this page first](#)

[Release Notes](#)

[Other Platforms](#)
[Old Versions](#)

License Agreement

Cytoscape is available as a platform-independent open-source Java application, released under the terms of the [LGPL](#). By downloading Cytoscape, you agree that you have read the license agreement that follows and agree to its terms. If you don't agree, do not download Cytoscape.

CYTOSCAPE LICENSE AGREEMENT
GNU LESSER GENERAL PUBLIC LICENSE Version 2.1, February 1999

Copyright (C) 1991, 1999 Free Software Foundation, Inc. 59 Temple Place

- <http://www.cytoscape.org/download.php>

Step 3: Understand what the Cytoscape app CyREST is for?

- A Cytoscape Core app for driving Cytoscape from R/Python/Julia/Node.js/etc.
- **IMPORTANT:** Now this app is a part of the Cytoscape Core!
- Ono, Keiichiro, et al. "[CyREST: Turbocharging Cytoscape Access for External Tools via a RESTful API](#)." F1000Research 4 (2015).

Step 4: Rstudio and libraries

- Install R and/or Rstudio
- Install igraph
- `install.packages("igraph")`
- Install Bioconductor download manager (lite version)
- `source("https://bioconductor.org/biocLite.R")`
- Install RCy3 in R (which connects R to Cytoscape, R side application).
- `biocLite("RCy3")` # previously called RCytoscape

Step 5: Running R and Cytoscape together

- In the same desktop environment
- Open Rstudio and load the following packages:
 - `library(igraph)`
 - `library(RCy3)`
- Open Cytoscape and have it side by side. CyREST is automatically loaded once you start Cytoscape. Now you're ready to draw some networks!

All Apps

Categories

collections
data visualization
network generation
graph analysis
online data import
network analysis
clustering
integrated analysis
utility
enrichment analysis
systems biology
data integration
visualization
layout
ontology analysis
pathway database
network comparison
local data import
core app
annotation

more »

Newest Releases

Get Started with the App Store »



PSFC

3.0+

Pathway Signal Flow Calculator:
Calculates pathway activity based



cyREST

3.0+

Core App: Language-agnostic
RESTful API



Rene

3.0+

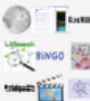
Regulatory Network Enhancing
Plugin - ReNE allows users to



MetScape

3.0+

Maps human, mouse and rat
metabolomics and gene



Omics Analysis Collection

3.0+

Collection of apps to facilitate
network analysis of omic data. By



**Functional Enrichment
Collection**

3.0+

Collection of apps to facilitate
functional enrichment analysis. By

[more newest releases »](#)

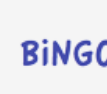
Top Downloaded Apps



ClueGO

3.0+

Creates and visualizes a
functionally grouped network of



BiNGO

3.0+

Calculates overrepresented GO
terms in the network and display

GeneMANIA

GeneMANIA

3.0+

Imports interaction networks from
public databases from a list of



CluePedia

3.0+

CluePedia: A ClueGO plugin for
pathway insights using integrated

<http://apps.cytoscape.org/>

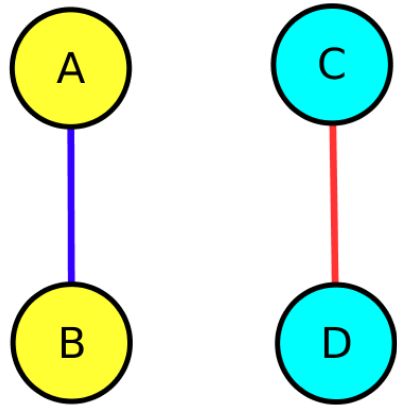
Code example 1 – Drawing a triplet in Cytoscape

Notes:

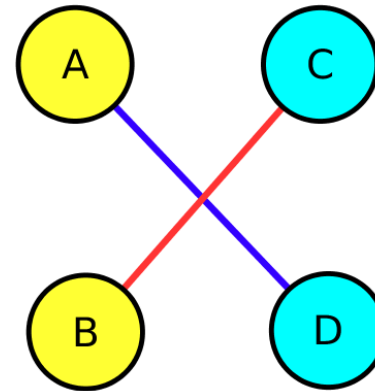
- Please refer to the script 'draw_a_single_triplet.Rmd' in the Github page.
- I've used the Bioconductor RCy3 vignette extensively for this example. Here is the reference :
<https://bioconductor.org/packages/release/bioc/vignettes/RCy3/inst/doc/RCy3.R>

Generating Randomized Networks but Keeping the Degree Distribution

Before Edge Swap



After Edge Swap



- The node 'degree' is the number of interaction partners each node has
- Establishing the background level of triplets
- The number of times the edges were swapped was equal to the size of the network.
- Each type of biological network (e.g. protein-protein interaction network) was randomized independently of other types of networks.
- These were combined to form the randomized integrated network.

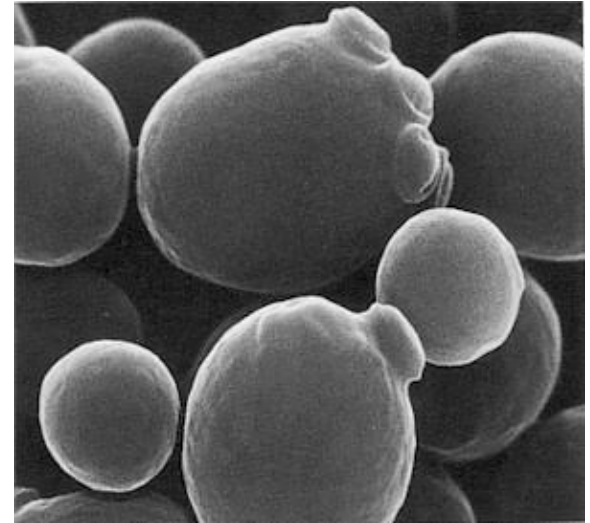
Code example 2 – Performing Network Randomization

Notes:

- Please refer to the script 'rewire_network.Rmd' in the Github page.
- This script show you how to perform network randomization, by edge swapping, while keeping the degree distribution the same. The number of interaction partners for every node should remain the same after this type of network randomization. The script also show you what happens if we randomize the network without conserving the degree distribution.

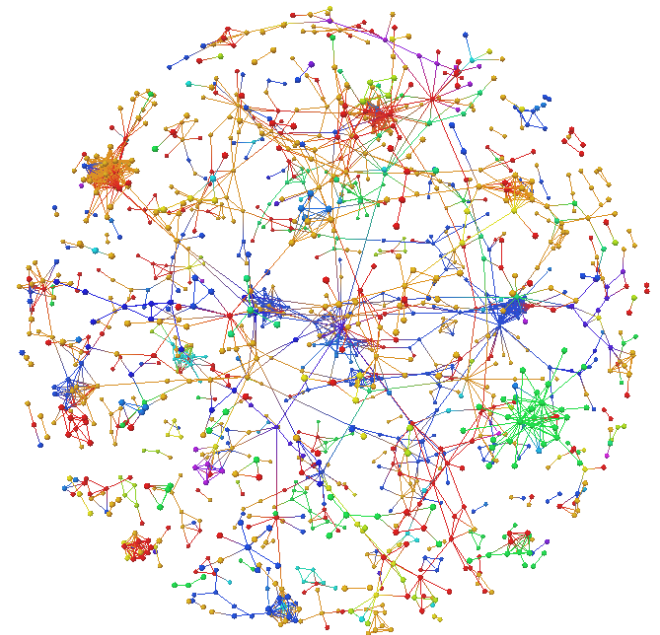
Saccharomyces cerevisiae

- **The world's most characterized eukaryote**
 - Vast genome, transcriptome and proteome resources
 - Knock out and over-expression mutants available for many genes
 - Databases for protein-protein, protein-DNA, protein-RNA interactions

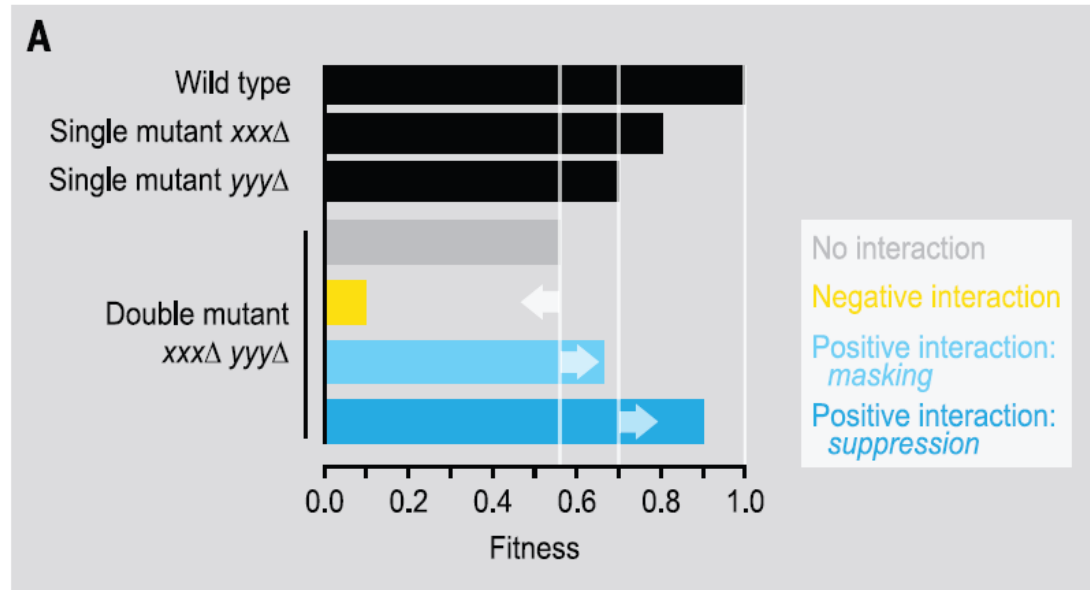


<http://tinyurl.com/q36n6dh>

→ Start with *Saccharomyces*
→ Testing grounds for algorithms
→ Map results to genes that are evolutionarily conserved in humans (e.g. essential genes, drug targets)



Genetic Interactions



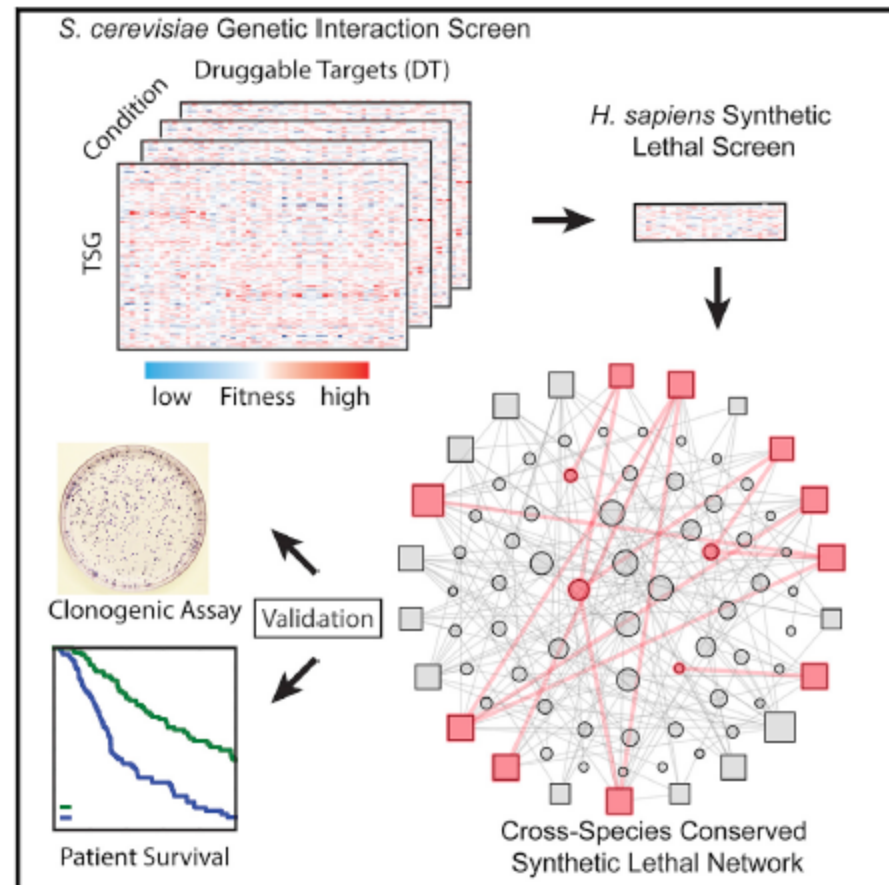
Expected Score If there is No Interaction
= $F(\text{wild type}) - F(xxx\Delta) * F(yyy\Delta)$

van Leeuwen et al. (2016) *Science* 354(6312):aag0839

Molecular Cell

A Network of Conserved Synthetic Lethal Interactions for Exploration of Precision Cancer Therapy

Graphical Abstract



Authors

Rohith Srivas, John Paul Shen, Chih Cheng Yang, ..., Pedro Aza-Blanc, Robert W. Sobol, Trey Ideker

Correspondence

tideker@ucsd.edu

In Brief

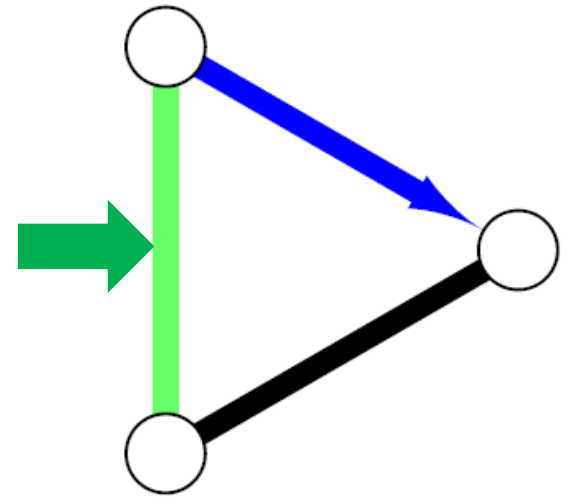
An emerging strategy for precision cancer therapy is to induce synthetic lethality in a tumor based on its pattern of genetic mutations. Here we map a large network of conserved synthetic lethal interactions between selective drugs and tumor suppressor genes somatically mutated in human cancer. All of these interactions are mirrored in budding yeast, an accessible model for researching drug mode of action.

Srivas et al., 2016, Molecular Cell 63, 514–525
August 4, 2016 © 2016 Elsevier Inc.
<http://dx.doi.org/10.1016/j.molcel.2016.06.022>

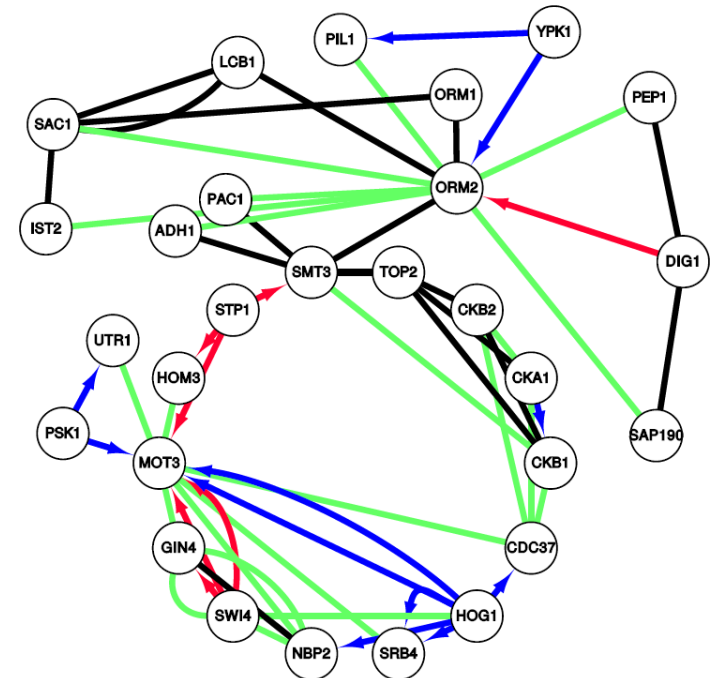
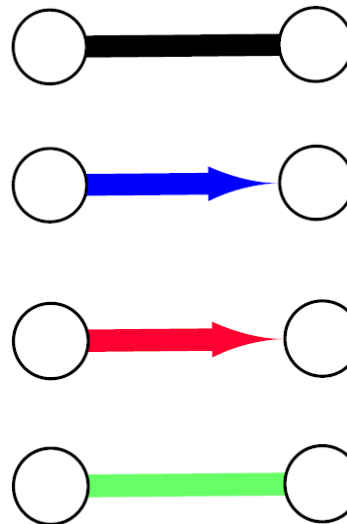
Applications

- Searching for Synergistic drug targets
- One drug target Protein A
- Another drug target Protein B
- Their combined effect is greater than expected (i.e. kills cancer cells or microbes quicker)
- Limit the chance of drug resistance developing
- New drugs are hard to develop, use pairs of existing drugs

A triplet consists of one
negative genetic interaction
and two other interactions.



- Protein-protein interactions
- Kinase-substrate interactions
- Transcription factor-target gene interactions
- Negative genetic interactions



Identification of Triplet Motifs

- There are 15 possible types of motifs that could occur, from the different combinations of the edges.
- We have identified all 15 possible types of motifs, with 30,850 triplets, 3,293 proteins and 23,225 interactions in total.

- **Protein-protein interactions**



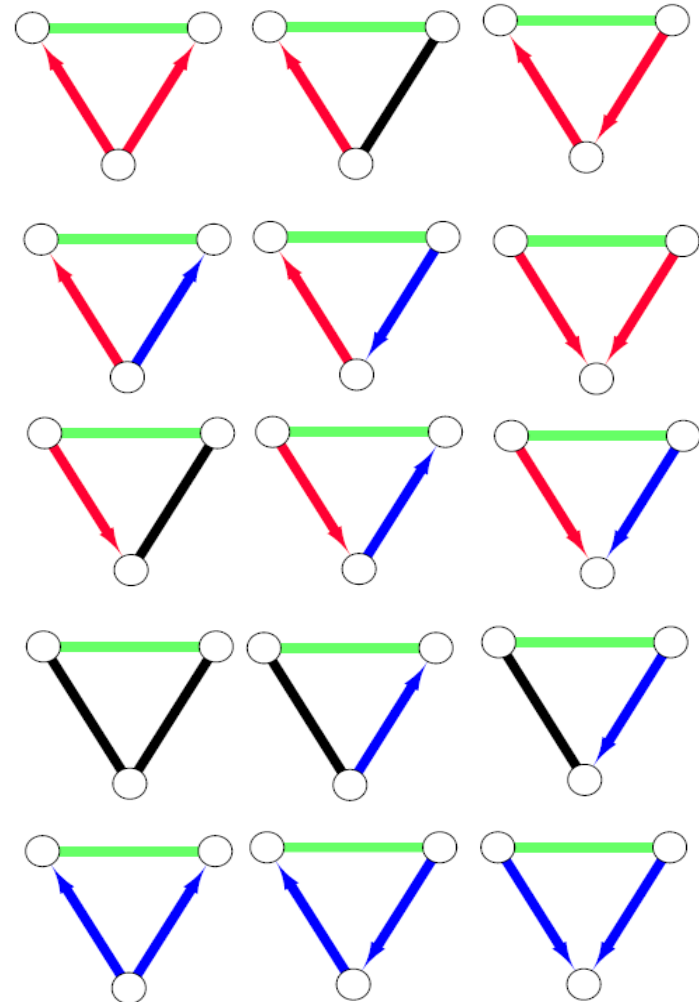
- **Kinase-substrate interactions**



- **Transcription factor-target gene interactions**



- **Negative genetic interactions**



Edge list table - caveat

Node A	Node B
Adam	Xavier
Xavier	Adam

* The edges can be entered into the edge list table in two different directions. Sometimes you need to filter / clean the table so that you only have one of these entry (e.g. Node A string > Node B string, “Adam” is alphabetically before “Xavier” etc...). Other times you need to keep both entries. I’ve done these filtering several times in the script ‘form_triplet_motifs.Rmd’.

Code example 3 – Setting random number seeds and distributing tasks to multiple cores

Notes:

- Please refer to the script 'random_number_generator.Rmd' in the Github page.
- This script show you how to set the seed so that the results are replicable each time you run it. It also show you how to do this in a setting where you send the execution to multicores (in a Linux / Unix environment).
- I did not show this example in the actual talk, mainly due to time constraint, but it is included here as a useful resource.

Code example 4 – counting the number of each type of triplet motif

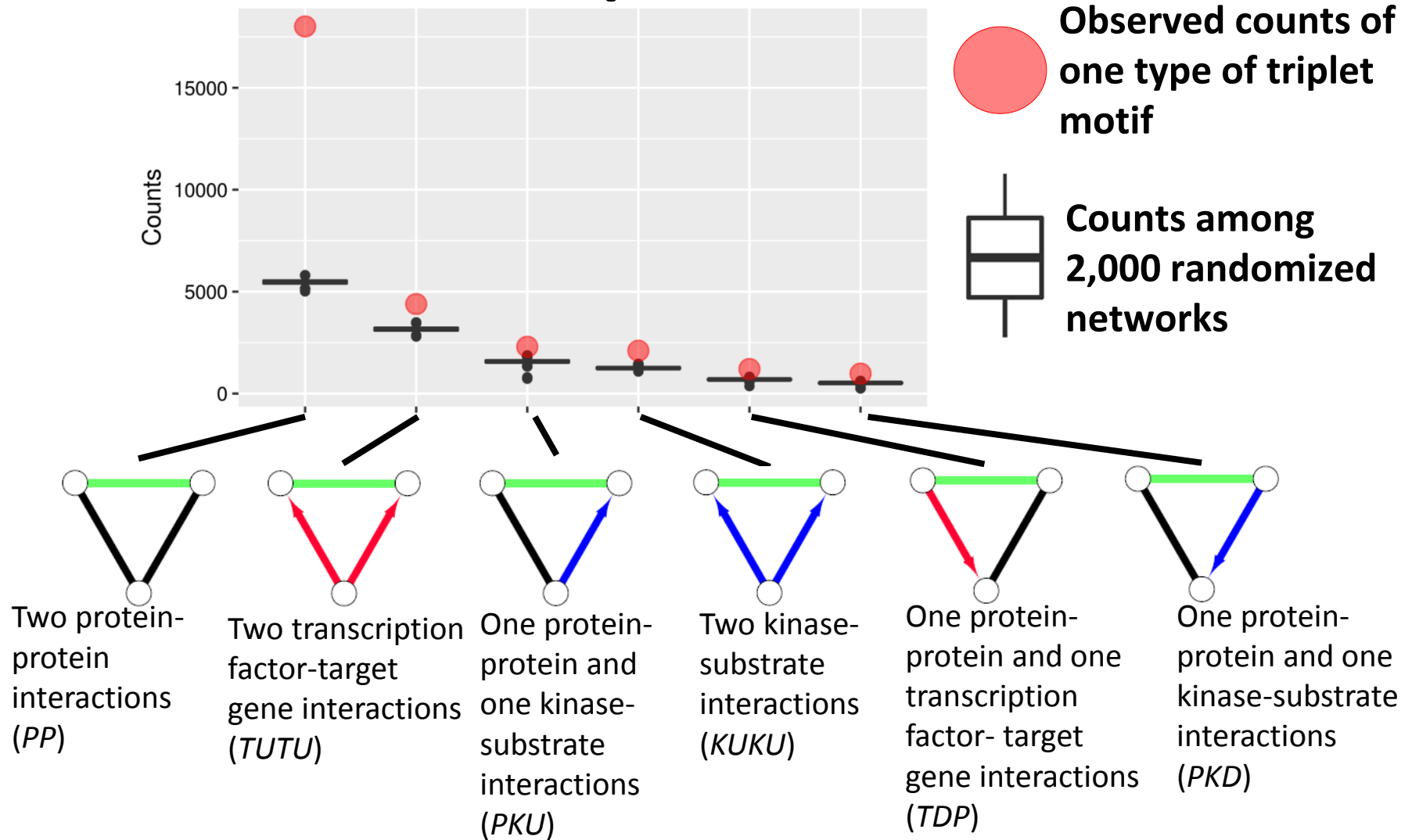
Notes:

- Please refer to the script ‘form_triplet_motifs.Rmd’ in the Github page.
- It counts the number of times each type of triplet motif has been identified in the network.

Calculating the Significance of Triplet Motifs

- p-value calculated by the number of times 2000 randomized networks has triplet motif counts greater than counts of the same motif in the actual network
- The Bonferroni method was used for multiple testing correction.
- A triplet motif is significant if:
 - The adjusted p-value is < 0.05 , and
 - Triplet motifs with observed counts less 2% of the total number of triplets is deemed un-reliable under a 2% false discovery rate.

Overrepresentation of 6 out of 15 Types of Triplet Motifs

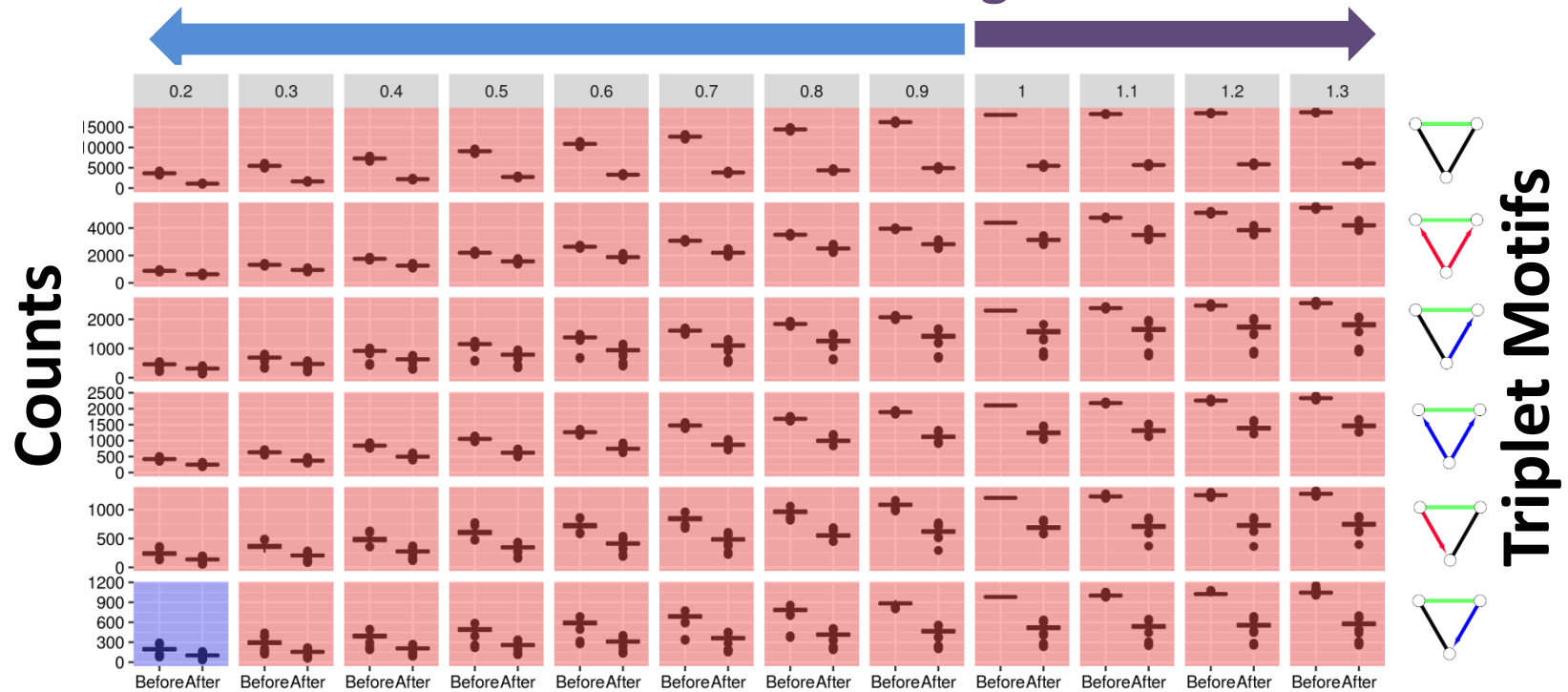


P= protein-protein, *T*= transcription factor-target gene, *K*=kinase-substrate, *U*=up, *D*=down

Triplet Motifs are Robust

Removing 10-80% of genetic interactions

Adding 10-30% of genetic interactions



Before or After Network Randomization

Significant

Non-significant

Known Limitations

- Network is only a reference map, static not dynamic
- Lacking information on whether interaction is present or absent in a specific condition
- Condition-specific interactions need to be tested for under various stresses, drug treatments, genetic backgrounds etc...
- False positive or false negative interactions

Summary

- Cytoscape can be used to visualize networks
- RCytoscape can be used to drive network analyses from R, providing a programmable workflow
- Network randomization can be used to generate many randomized networks. For these background set, we can calculate the 'null' or expected distribution. We can then compared our observed data with the expected distribution and infer the significance of the differences.

Resources and Databases

- Github page for today's talk
 - https://github.com/IgnatiusPang/Network_SURF
- Biological Networks Databases
 - <https://thebiogrid.org/>
 - <http://string-db.org/>
 - <http://www.ebi.ac.uk/intact/>
- Cytoscape, Rcytoscape, igraph
 - <http://www.cytoscape.org>
 - <https://bioconductor.org/packages/release/bioc/html/RCy3.html>
 - <http://igraph.org/r/>
 - Igraph tutorial: <http://kateto.net/networks-r-igraph>

References

- Alon, U. (2007) *Nature Reviews Genetics* **8**, 450-461.
- Costanzo, M. *et al.* (2010) *Science* **327**, 425–431.
- Costanzo, M. *et al.* (2016) **353**, p.aaf1420-aaf1420
- Pang, C. N. I. *et al.* (2012) *J. Proteome Res.* **11**, 5204–5220.
- Bertin, N. *et al.* (2007) *PLoS Biol* 5(6): e153.
- Sharifpoor, S. *et al.* (2011) *Genome Biol* **12**, R39.
- Maclsaac, K. D. *et al.* (2006) *BMC Bioinformatics* **7**, 113.
- Reimand, J. *et al.* (2012) *Genome Biol* **13**, R55.
- Balakrishnan, R. *et al.* (2012) *Database (Oxford)*. **2012**, bar062.
- Shannon, P. *et al.* (2003) *Genome Res* **13**, 2498–2504 (2003).
- Yu *et al.* (2008) *Nucleic Acids Res.* 36, 6494-6503.

Acknowledgements

- We would like to acknowledge funding from the Bioplatforms Australia and the Australian Research Council (ARC).
- We would also like to thank people from the Wilkins lab and Edwards lab for feedback and suggestions.



Australian Government

Australian Research Council