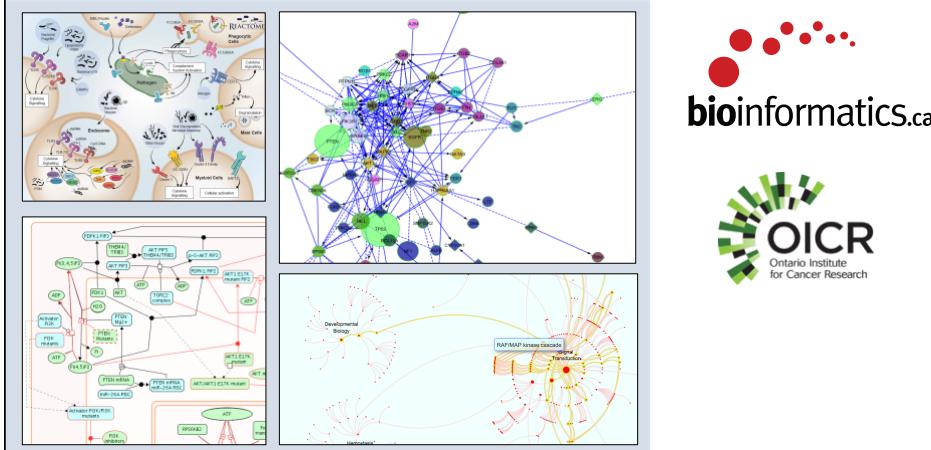


Module 4

More Depth on Pathway and Network Analysis

Lincoln Stein
Pathways and Network Analysis of -omics Data
June 14th, 2016



Why Pathway/Network Analysis?

- Dramatic data size reduction: 1000's of genes => dozens of pathways.
- Increase statistical power by reducing multiple hypotheses.
- Find meaning in the “long tail” of rare cancer mutations.
- Tell biological stories:
 - Identifying hidden patterns in gene lists.
 - Creating mechanistic models to explain experimental observations.
 - Predicting the function of unannotated genes.
 - Establishing the framework for quantitative modeling.
 - Assisting in the development of molecular signatures.

What is Pathway/Network Analysis?

- Any analytic technique that makes use of biological pathway or molecular network information to gain insights into a biological system.
- A rapidly evolving field.
- Many approaches.

Module 4

bioinformatics.ca

Why Pathway Analysis? Mutational landscape and significance across 12 major cancer types

Cyriac Kandoth^{1*}, Michael D. McLellan^{1*}, Fabio Vandin², Kai Ye^{1,3}, Beifang Niu¹, Charles Lu¹, Mingchao Xie¹, Qunyuan Zhang^{1,3}, Joshua F. McMichael¹, Matthew A. Wyczalkowski¹, Mark D. M. Leiserson², Christopher A. Miller¹, John S. Welch^{4,5}, Matthew J. Walter^{2,5}, Michael C. Wendl^{1,3,6}, Timothy J. Ley^{1,3,4,5}, Richard K. Wilson^{1,3,5}, Benjamin J. Raphael² & Li Ding^{1,3,4,5}

127 Cancer Driver Genes

ACVR1B, ACVR2A, AJUBA, AKT1, APC, AR, ARHGAP35, ARID1A, ARID5B, ASXL1, ATM, ATR, ATRX, AXIN2, B4GALT3, BAP1, BRAF, BRCA1, BRCA2, CBFB, CCND1, CDH1, CDK12, CDKN1A, CDKN1B, CDKN2A, CDKN2C, CEBPA, CHEK2, CRIPAK, CTCF, CTNNB1, DNMT3A, EGFR, EGR3, EIF4A2, ELF3, EP300, EPHA3, EPHB6, EPPK1, ERBB4, ERCC2, EZH2, FBXW7, FGFR2, FGFR3, FLT3, FOXA1, FOXA2, GATA3, H3F3C, HGF, HIST1H1C, HIST1H2BD, IDH1, IDH2, KDM5C, KDM6A, KEAP1, KIT, KRAS, LIFR, LRRK2, MALAT1, MAP2K4, MAP3K1, MAPK8IP1, MECOM, MIR142, MLL2, MLL3, MLL4, MTOR, NAV3, NCOR1, NF1, NFE2L2, NFE2L3, NOTCH1, NPM1, NRAS, NSD1, PBRM1, PCBP1, PDGFRA, PHF6, PIK3CA, PIK3CG, PIK3R1, POLQ, PPP2R1A, PRX, PTEN, PTPN11, RAD21, RB1, RPL22, RPL5, RUNX1, SETBP1, SETD2, SF3B1, SIN3A, SMAD2, SMAD4, SMC1A, SMC3, SOX17, SOX9, SPOP, STAG2, STK11, TAF1, TBL1XR1, TBX3, TET2, TGFBR2, TLR4, TP53, TSHZ2, TSHZ3, U2AF1, USP9X, VEZF1, VHL, WT1

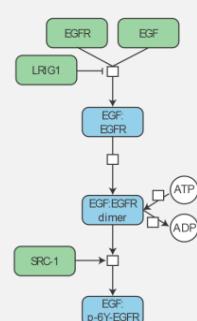
Nature 502 (2013): 333-339

Module 4

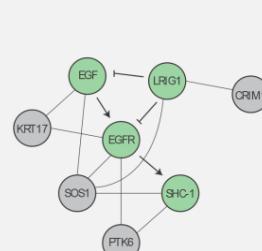
bioinformatics.ca

Pathways vs Networks

EGFR-centered Pathway



EGFR-centered Network

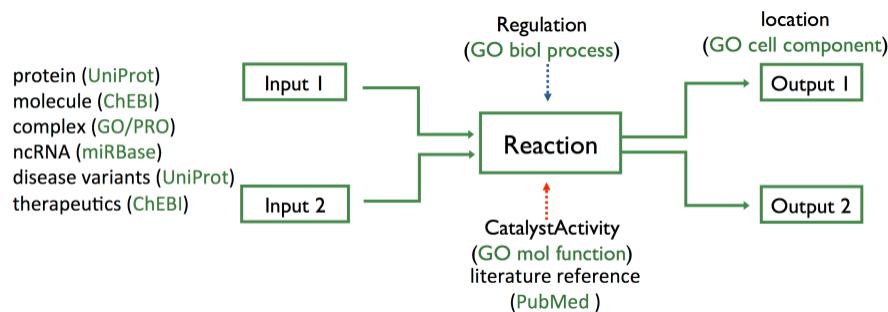


Module 4

bioinformatics.ca

Reaction-Network Databases

- Reactome & KEGG
 - explicitly describe biological processes as a series of biochemical reactions.
 - represents many events and states found in biology.



Module 4

bioinformatics.ca

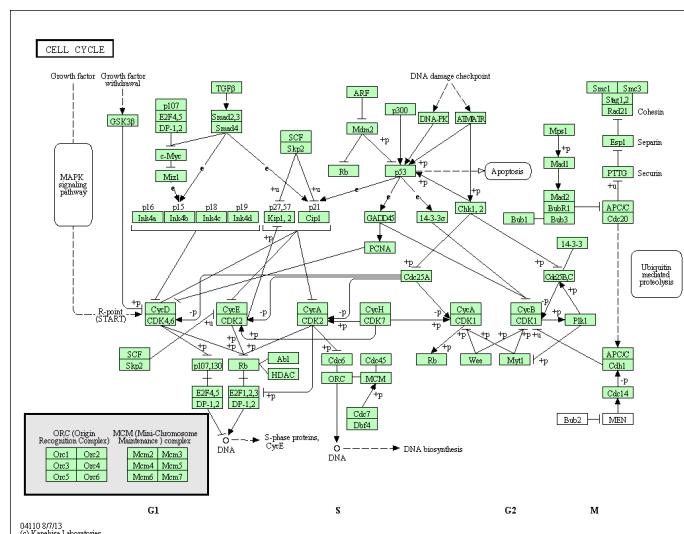
KEGG

- KEGG is a collection of biological information compiled from published material → **curated database**.
- Includes information on genes, proteins, metabolic pathways, molecular interactions, and biochemical reactions associated with specific organisms
- Provides a relationship (map) for how these components are organized in a cellular structure or reaction pathway.

Module 4

bioinformatics.ca

KEGG Pathway Diagram



Module 4

bioinformatics.ca

Reactome

- Open source and open access pathway database
- Curated human pathways encompassing metabolism, signaling, and other biological processes.
- Every pathway is traceable to primary literature.
- Cross-reference to many other bioinformatics databases.
- Provides data analysis and visualization tools

Module 4

bioinformatics.ca

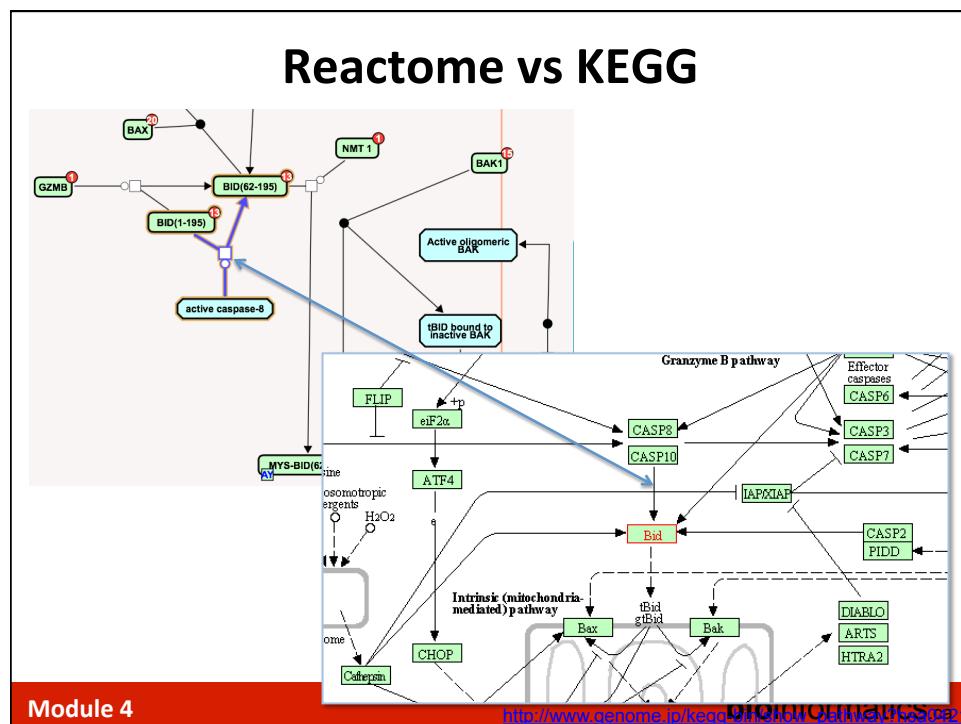
Pathway Browser: Pathway Enrichment Analysis

| Pathway name | Entities found | Entities Total | Entities ratio | pValues | Entities FDR | Reactions found | Reactions Total | Reactions ratio | Species name |
|--|----------------|----------------|----------------|---------|--------------|-----------------|-----------------|-----------------|--------------|
| GAB1 signaling | 21 | 125 | 0.11 | 1.1E-15 | 4.8E-14 | 34 | 38 | 0.004 | Homo sapiens |
| PI3KAKT Signaling in Cancer | 22 | 116 | 0.11 | 3.2E-15 | 1.3E-13 | 21 | 21 | 0.003 | Homo sapiens |
| Developmental Biology | 47 | 904 | 0.082 | 7.4E-14 | 2.9E-12 | 153 | 415 | 0.047 | Homo sapiens |
| Downstream signaling events of B Cell Receptor (BCR) | 23 | 196 | 0.018 | 7.4E-14 | 2.9E-12 | 24 | 37 | 0.004 | Homo sapiens |
| Signaling by ERBB4 | 28 | 349 | 0.031 | 8.7E-13 | 3.3E-11 | 82 | 109 | 0.012 | Homo sapiens |
| Signaling by ERBB2 | 28 | 360 | 0.033 | 2.3E-12 | 8.6E-11 | 77 | 118 | 0.013 | Homo sapiens |
| Signaling by SCF-KIT | 27 | 339 | 0.031 | 3.6E-12 | 1.2E-10 | 82 | 100 | 0.011 | Homo sapiens |
| Downstream signaling of activated FGFR2 | 27 | 349 | 0.032 | 6.9E-12 | 2.1E-10 | 71 | 103 | 0.012 | Homo sapiens |
| Downstream signaling of activated FGFR3 | 27 | 349 | 0.032 | 6.9E-12 | 2.1E-10 | 71 | 103 | 0.012 | Homo sapiens |
| Downstream signaling of activated FGFR4 | 27 | 349 | 0.032 | 6.9E-12 | 2.1E-10 | 60 | 103 | 0.012 | Homo sapiens |
| Downstream signaling of activated FGFR1 | 27 | 349 | 0.032 | 6.9E-12 | 2.1E-10 | 60 | 103 | 0.012 | Homo sapiens |

127 Cancer Driver Genes

Module 4

bioinformatics.ca



Pathway and Network Compendia

[Search Gene](#)
[Downloads](#)
[Apps](#)
[FAQ](#)
[Publications](#)
[Contact](#)

Version 8: Over 42 000 Pathways and 1 350 000 Interactions from 22 Data Sources X

Pathway Commons

Pathway information. Single point of access.

Pathway Commons aims to store and disseminate knowledge about biological pathways. Information is sourced from [public pathway databases](#) and is readily searched, visualized, and downloaded. The data is freely available under the license terms of each contributing database.

Pathway Commons, a web resource for biological pathway data. Cerami E et al. Nucleic Acids Research (2011).

Module 4

bioinformatics.ca

What is a Interaction Network?

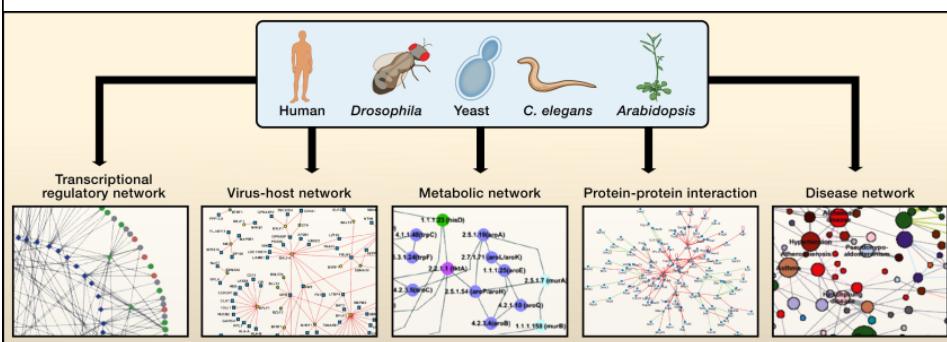
- An Interaction Network is a collection of:
 - Nodes (or vertices).
 - Edges connecting nodes (directed or undirected, weighted, multiple edges, self-edges).
- Nodes can represent proteins, genes, metabolites, or groups of these (e.g. complexes) - any sort of object.
- Edges can be either physical or functional interactions, activators, regulators, reactions - any sort of relations.



Module 4

bioinformatics.ca

Types of Interactions Networks



Vidal, Cusick and Barabasi, Cell 144, 2011.

Module 4

bioinformatics.ca

Network Databases

- Can be built automatically or via curation.
- More extensive coverage of biological systems.
- Relationships and underlying evidence more tentative.
- Popular sources of curated human networks:
 - BioGRID – Curated interactions from literature; 21K interactors & 363K interactions.
 - IntAct – Curated interactions from literature; 91K interactors & 591K interactions.
 - MINT – Curated interactions from literature; 32K interactors, 241K interactions.

Module 4

bioinformatics.ca

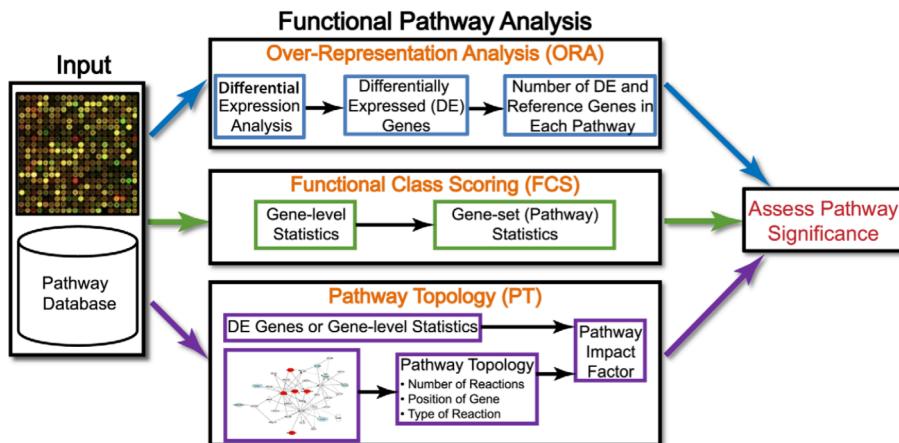
IntAct

| ID | Molecule 'A' | Links 'A' | Molecule 'B' | Links 'B' | Interaction Detection Method | Source Database |
|----|--------------|----------------------|--------------|----------------------|---------------------------------|---|
| 1 | TP53 | P04637 EBI-356083 | MDM2 | Q00987 EBI-359668 | anti bait coimmunoprecipitation | EBI-5735247 MINT-683475 IM-11641-7 |
| 2 | | | | | anti bait coimmunoprecipitation | EBI-2573567 IntAct |
| 3 | | | | | anti bait coimmunoprecipitation | EBI-2573607 IntAct |
| 4 | | | | | molecular sieving | EBI-8045759 MINT-790532 IM-15586-6 |
| 5 | | | | | molecular sieving | EBI-8045899 MINT-790532 IM-15586-14 |

Module 4

bioinformatics.ca

Pathway/Network Analysis Workflow

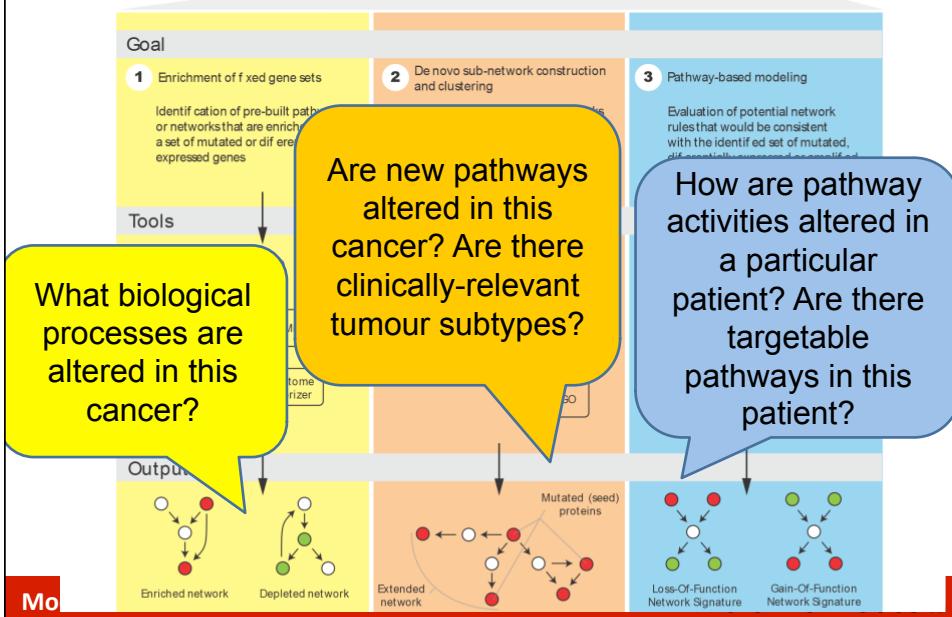


Khatri P et al. PLoS Comput Biol. 2012; 8(2): e1002375

Module 4

bioinformatics.ca

Types of Pathway/Network Analysis



Issues in Pathway-based Analysis

- How to handle the pathway hierarchical organization?
 - Flatten pathways organized in a hierarchy into a systems-wide network
- How to handle pathway cross-talks?
 - Shared genes/proteins will be listed once in a network
 - Interactions causing cross-talks are displayed in the same network

Module 4

bioinformatics.ca

More Challenged Issues in Pathway Data Analysis

- Many omics data types available for one single patient
 - CNV, gene expression, methylation, somatic mutations, etc.
- Pathway and network-based Simulation
 - How to use topological structures?
 - Predict drug effects: one drug or multiple drugs together?

Module 4

bioinformatics.ca

Network-based Data Analysis

- Based on systems-wide biological networks
 - Covering the majority of human genes
 - Usually protein-protein interaction networks
- Modules (or clusters)-based network patterns
 - Pathway annotations via enrichment analysis
 - Gene signature or biomarker discovery
- Disease genes discovery
 - Cancer drivers
 - Disease modules

Module 4

bioinformatics.ca

2) De Novo Subnetwork Construction & Clustering

- Apply list of altered {genes,proteins,RNAs} to a biological network.
- Identify “topologically unlikely” configurations.
 - E.g. a subset of the altered genes are closer to each other on the network than you would expect by chance.
- Extract clusters of these unlikely configurations.
- Annotate the clusters.

Module 4

bioinformatics.ca

Network clustering

- Clustering can be defined as the process of grouping objects into sets called clusters (communities or modules), so that each cluster consists of elements that are similar in some ways.
- Network clustering algorithm is looking for sets of nodes [proteins] that are joined together in tightly knit groups.
- Cluster detection in large networks is very useful as highly connected proteins are often sharing similar functionality.

Module 4

bioinformatics.ca

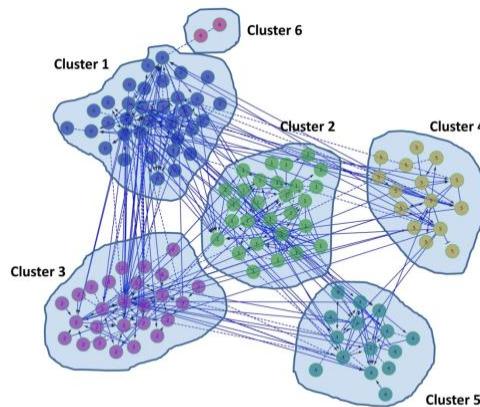
Popular Network Clustering Algorithms

- Girvin-Newman
 - a hierarchical method used to detect communities by progressively removing edges from the original network
- Markov Cluster Algorithm
 - a fast and scalable unsupervised cluster algorithm for graphs based on simulation of (stochastic) flow in graphs
- HotNet
 - Finds “hot” clusters based on propagation of heat across metabolic lattice.
 - Avoids ascertainment bias on unusually well-annotated genes.
- HyperModules Cytoscape App
 - Find network clusters that correlate with clinical characteristics.
- Reactome FI Network Cytoscape App
 - Offers multiple clustering and correlation algorithms (including HotNet, PARADIGM and survival correlation analysis)

Module 4

bioinformatics.ca

Typical output of a network clustering algorithm



This hypothetical subnetwork was decomposed onto 6 clusters.

Different clusters are marked with different colors.
Cluster 6 contains only 2 elements and could be ignored in the further investigations.

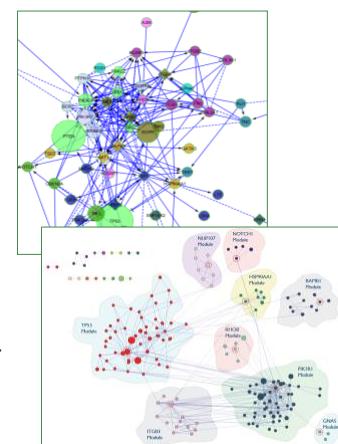
Clusters are mutually exclusive meaning that nodes are not shared between the clusters.

Module 4

bioinformatics.ca

Reactome Functional Interaction (FI) Network and ReactomeFIViz app

- No single mutated gene is necessary and sufficient to cause cancer.
 - Typically one or two common mutations (e.g. TP53) plus rare mutations.
- Analyzing mutated genes in a network context:
 - reveals relationships among these genes.
 - can elucidate mechanism of action of drivers.
 - facilitates hypothesis generation on roles of these genes in disease phenotype.
- Network analysis reduces hundreds of mutated genes to < dozen mutated pathways.

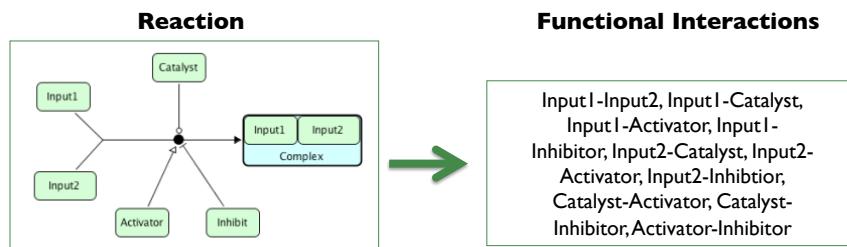


Module 4

bioinformatics.ca

What is a Functional Interaction?

- Convert reactions in pathways into pair-wise relationships
 - Functional Interaction:** an interaction in which two proteins are involved in the same reaction as input, catalyst, activator and/or inhibitor, or as components in a complex

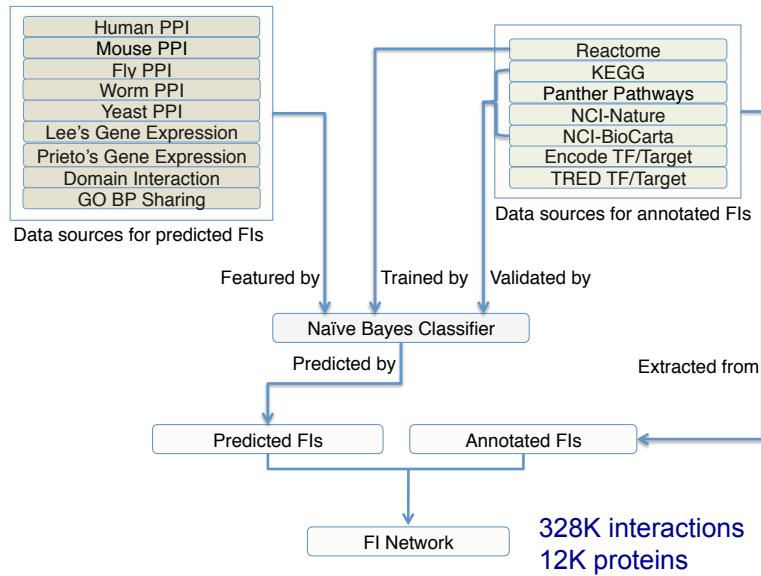


Method and practical application: A human functional protein interaction network and its application to cancer data analysis, [Vu et al. 2010 Genome Biology](#).

Module 4

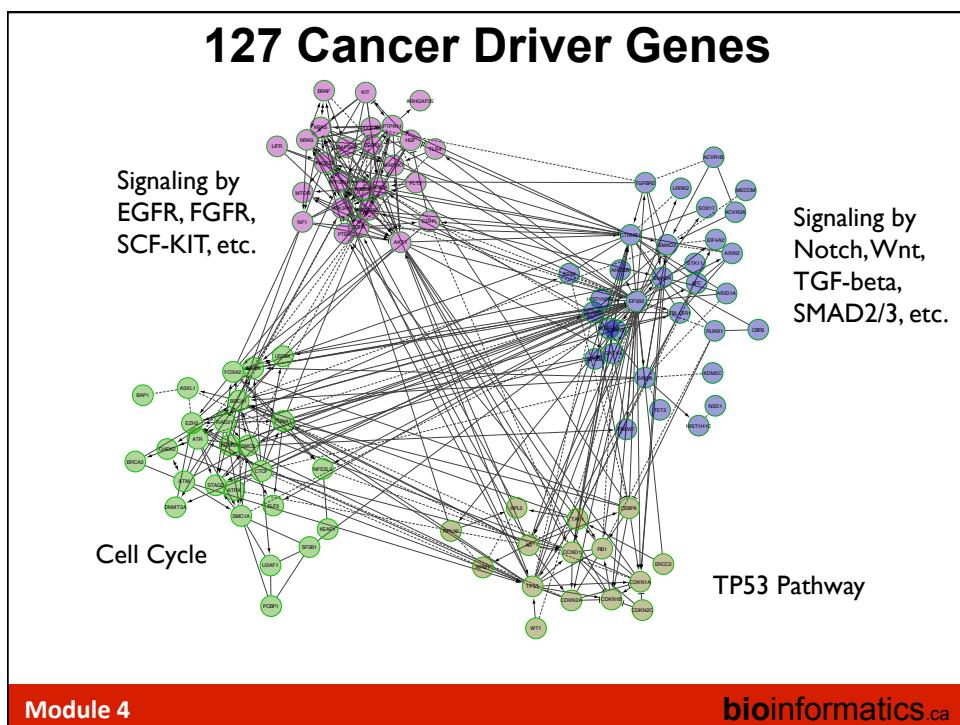
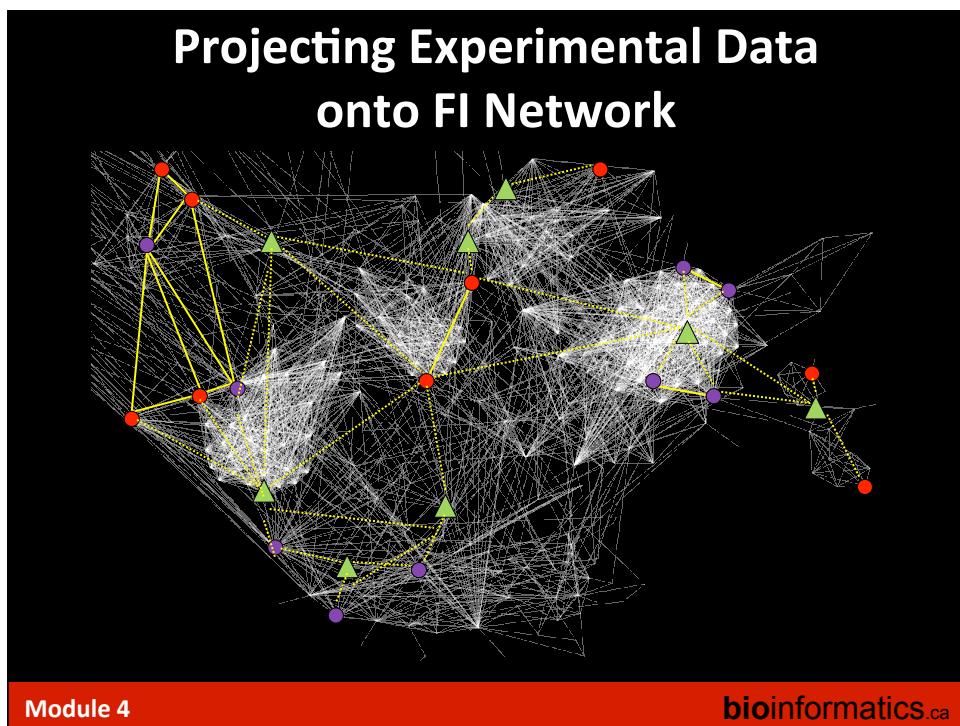
bioinformatics.ca

Construction of the FI Network

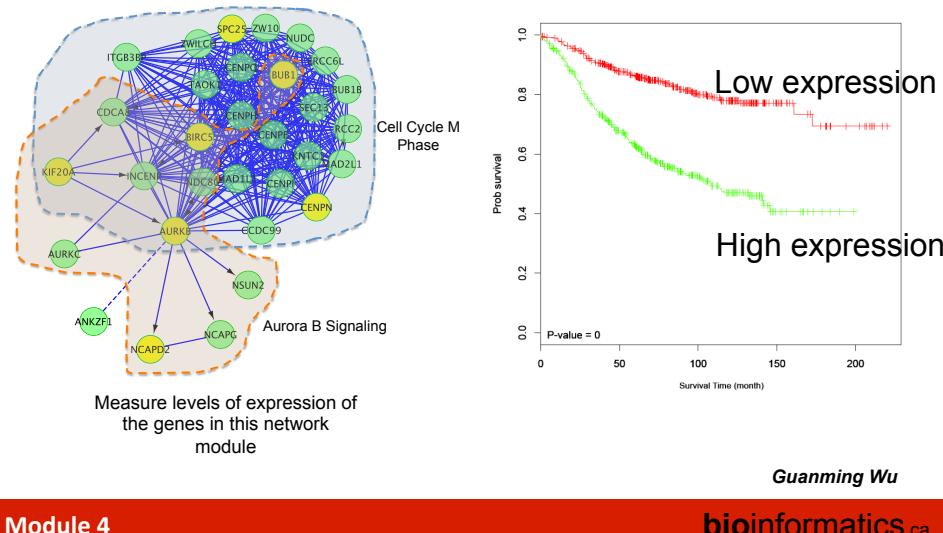


Module 4

bioinformatics.ca



Module-Based Prognostic Biomarker in ER+ Breast Cancer



3) Pathway-Based Modeling

- Apply list of altered {genes, proteins, RNAs} to biological pathways.
- Preserve detailed biological relationships.
- Attempt to integrate multiple molecular alterations together to yield lists of altered pathway activities.
- Pathway modeling shades into Systems Biology

Types of Pathway-Based Modeling

- Partial differential equations/boolean models, e.g. CellNetAnalyzer
 - Mostly suited for biochemical systems (metabolomics)
- Network flow models, e.g. NetPhorest, NetworKIN
 - Mostly suited for kinase cascades (phosphorylation info)
- Transcriptional regulatory network-based reconstruction methods, e.g. ARACNe (expression arrays)
- Probabilistic graph models (PGMs), e.g. PARADIGM
 - Most general form of pathway modeling for cancer analysis at this time.

Module 4

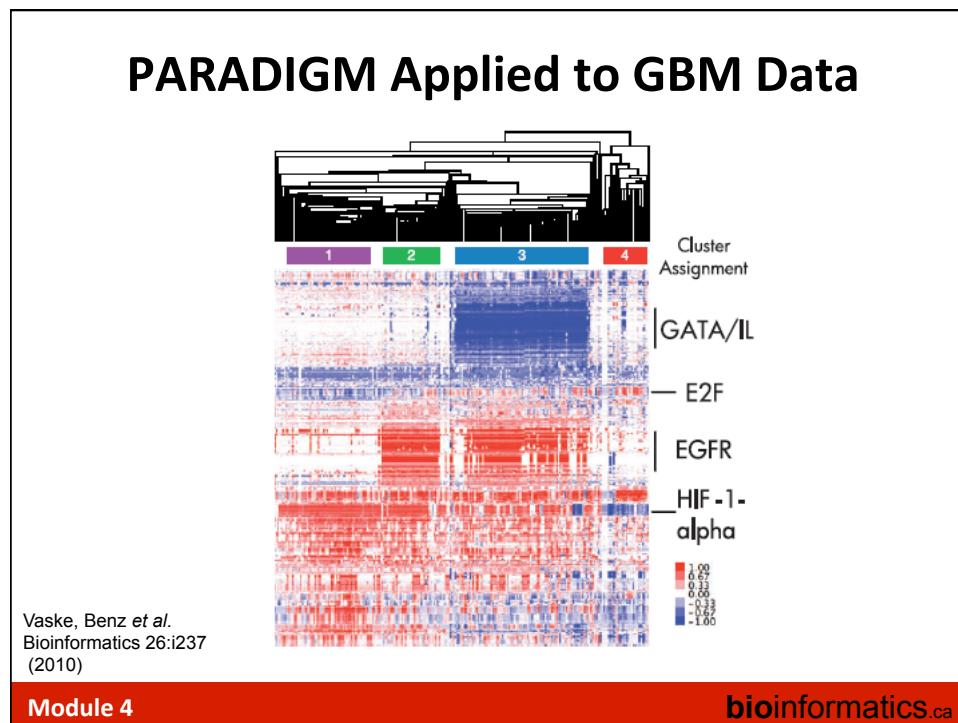
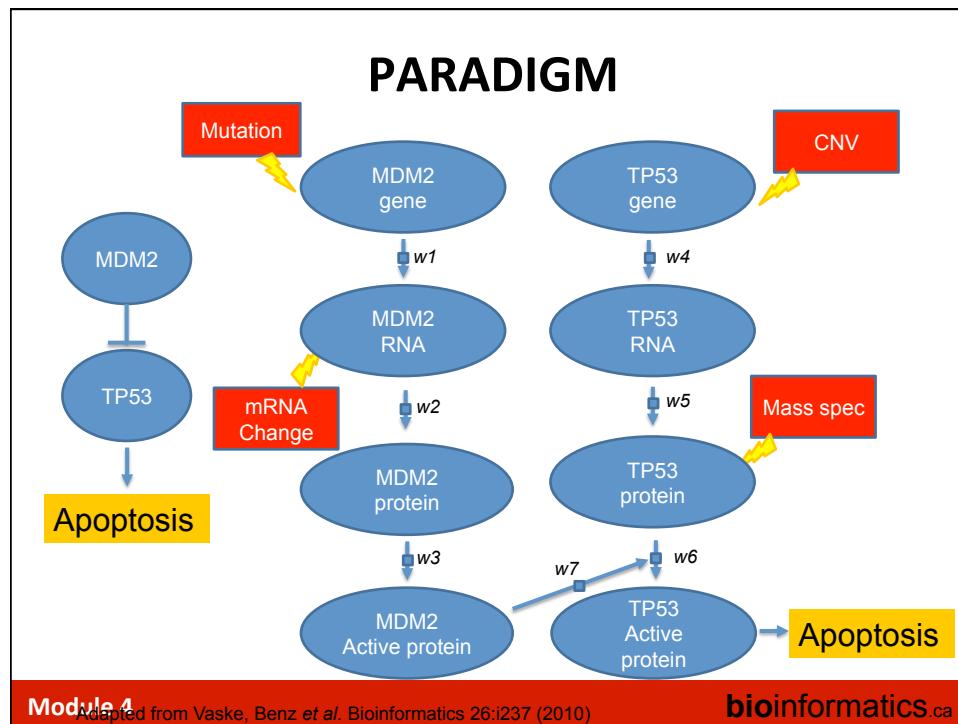
bioinformatics.ca

Probabilistic Graphical Model (PGM) based Pathway Analysis

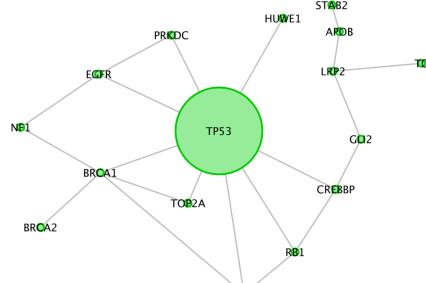
- Integrate multiple ‘omics’ data types into pathway context using PGM models
 - CNV
 - mRNA
 - *Mutation, Protein, etc*
- Find significantly impacted pathways for diseases
- Link pathway activities to patient phenotypes

Module 4

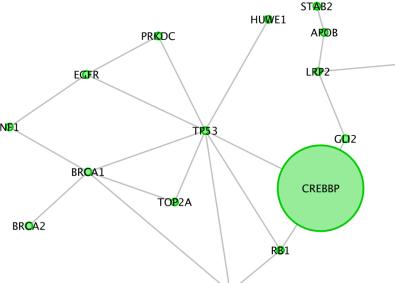
bioinformatics.ca



PGM-based Single OvCa Patient Network View



TCGA-04-1517 (TCGA OV)



TCGA-04-1525 (TCGA OV)

Module 4

bioinformatics.ca

PARADIGM: Good & Bad News

- Bad News
 - Distributed in source code form & hard to compile.
 - No pre-formatted pathway models available.
 - Scant documentation.
 - Takes a long time to run.
- Good News
 - Reactome Cytoscape app supports PARADIGM (beta testing).
 - Includes Reactome-based pathway models.
 - We have improved performance; working on further improvements.

Module 4

bioinformatics.ca

Pathway/Network Database URLs

- BioGRID
 - <http://www.thebiogrid.org>
- IntAct
 - <http://www.ebi.ac.uk/intact/>
- KEGG
 - <http://www.genome.jp/kegg>
- MINT
 - <http://mint.bio.uniroma2.it>
- Reactome
 - <http://www.reactome.org>
- Pathway Commons
 - <http://www.pathwaycommons.org>
- Wiki Pathways
 - <http://wikipathways.org>

Module 4

bioinformatics.ca

De novo network construction & clustering

- GeneMANIA
 - <http://www.genemania.org>
- HotNet
 - <http://compbio.cs.brown.edu/projects/hotnet/>
- HyperModules
 - <http://apps.cytoscape.org/apps/hypermodules>
- Reactome Cytoscape FI App
 - <http://apps.cytoscape.org/apps/reactomefi>

Module 4

bioinformatics.ca

Pathway Modeling

- CellNetAnalyzer
 - <http://www.ebi.ac.uk/research/saez-rodriguez/software>
- NetPhorest/NetworKIN
 - <http://netphorest.info>, <http://networkin.info>
- ARACNe
 - <http://wiki.c2b2.columbia.edu/califanolab/index.php/Software/ARACNE>
- PARADIGM
 - <http://paradigm.five3genomics.com/>