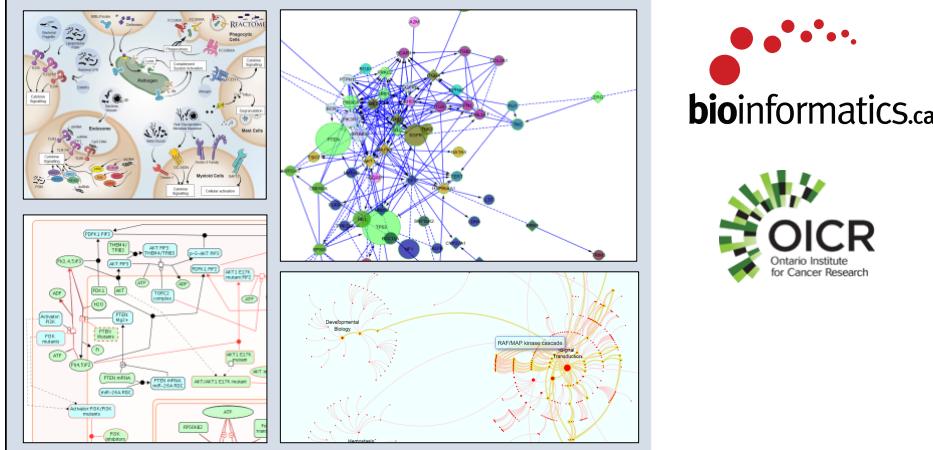


Module 4

Networks and Pathways

Lincoln Stein
 Pathway and Network Analysis of -omics Data
 June 1-3, 2015



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?

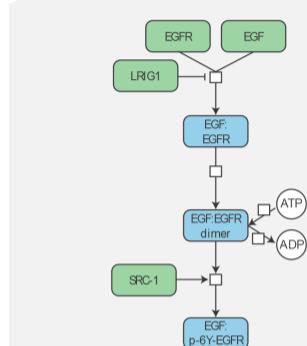
- Any analytic technique that makes use of biological pathway or molecular network information to gain insights into a biological system.
- A rapidly evolving field.
- Many approaches.

Module 4

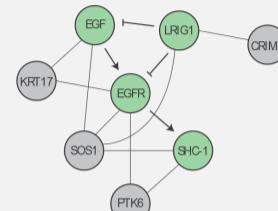
bioinformatics.ca

Pathways vs Networks

EGFR-centered
Pathway



EGFR-centered
Network



Ingredients you will Need

1. A list of altered genes, proteins, RNAs, etc.
2. A source of pathways or networks.

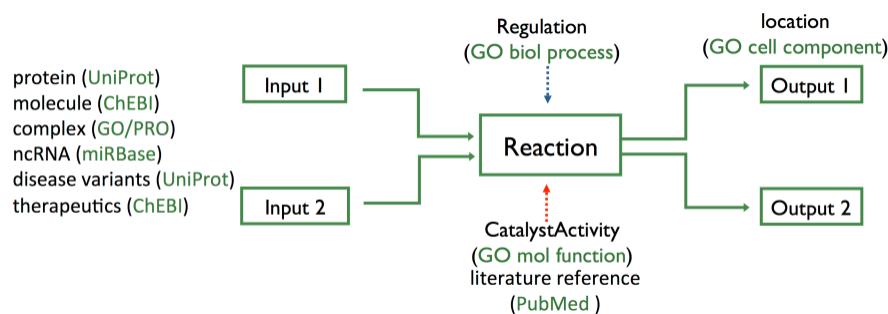


Module 4

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

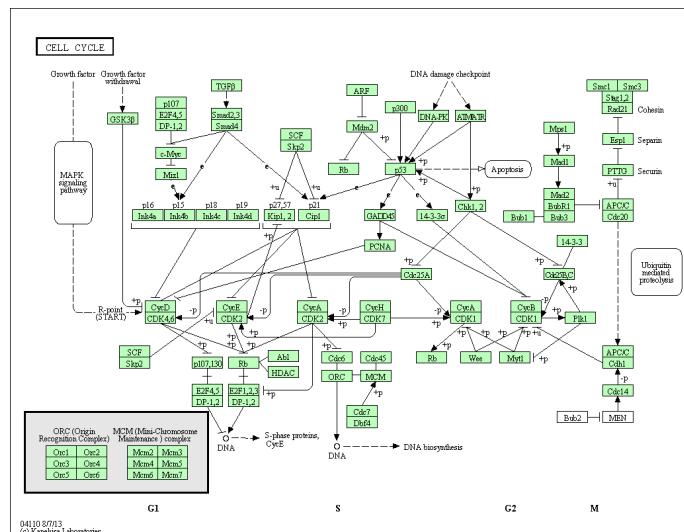
KEGG

- 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

Reactome

The screenshot shows the Reactome Pathway Browser interface. The main panel displays a complex biological pathway involving DNA repair, specifically Double-Strand Break Repair. Key components shown include XRCC6, XRCC5, LIG4, XRCC4, BRCA1, PRKDC, and various enzymes like nucleases and phosphatases. A color scale on the right indicates pathway scores from 8.15 to 13.50. Below the main diagram, a table provides detailed results for the pathway:

Pathway name	Entities found	Entities Total	Entities ratio	pValue	For	Reactions found	Reactions total	Reactions ratio	10h control
truncated APC mutants destabilize the destruction complex phosphorylation site mutants of CTNNB1 are not targeted to the proteasome by the destruction complex	38	62	0.01	1.47E-13	1.18E-12	7	24	0.003	9.426
Extension of Telomeres	22	29	0.004	3.09E-13	2.26E-12	14	21	0.003	8.134
Double-Strand Break Repair	20	23	0.003	3.23E-13	2.26E-12	31	32	0.004	7.454
E2F mediated regulation of DNA replication	23	33	0.004	4.97E-13	3.48E-12	6	6	0.001	7.822
RNA Polymerase III Transcription	25	40	0.005	5.12E-13	3.59E-12	25	25	0.003	7.613

Module 4

bioinformatics.ca

Pathway and Network Compendia

Pathway Commons Download F.A.Q. Publications Contact

March 4, 2015 - We have released a new version of Pathway Commons 2 (v7), which serves 31,698 pathways and 1,151,476 interactions from 18 data sources. [X](#)

Pathway Commons

Search and visualize public biological pathway information. Single point of access.

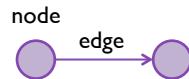
BRCA1, BRCA2, MDM2 [Start exploring »](#)

Pathway Commons is a network biology resource and acts as a convenient point of access to biological pathway information collected from public pathway databases, which you can search, visualize and download. All data is freely available, under the license terms of each contributing database.

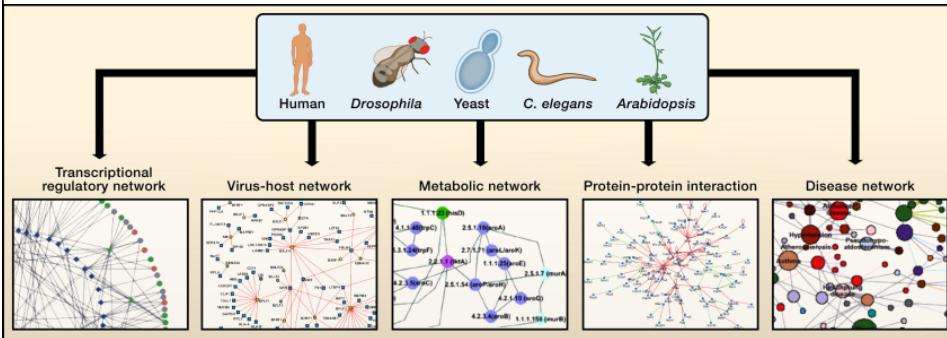
Module 4 bioinformatics.ca

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.



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 networks:
 - BioGRID – Curated interactions from literature; 529,000 genes, 167,000 interactions.
 - InTact – Curated interactions from literature; 60,000 genes, 203,000 interactions.
 - MINT – Curated interactions from literature; 31,000 genes, 83,000 interactions.

Module 4

bioinformatics.ca

IntAct

IntAct > IntAct Search Results Show more data from EMBL-EBI

7,708 binary interactions found for search term *tp53*

[Interactions \(7708\)](#) [Browse](#) [Lists](#) [Interaction Details](#) [Molecule View](#) [Graph](#)

- Filter out the spoke expanded co-complexes
- (2112)
- Include 8 negative interactions matching the query (pink row, not exported in MITAB 2.5, RDF, Biopax, XGMML)

Your query also matches **8,423** interaction evidences [What is this view?](#)
from 7 other databases.

Select format to Download [Download](#) [Customize view](#)

Dts	Molecule 'A'	Links 'A'	Molecule 'B'	Links 'B'	Interaction Detection Method	Interaction AC	Source Database
	TP53 P04637 EBI-366083		MDM2 Q00987 EBI-389668		anti bait coimmunoprecipitation	EBI-8552247 MINT-6823775 imex : IM-11641-7	MINT
					anti bait coimmunoprecipitation	EBI-626521	UniProt
					anti bait coimmunoprecipitation	EBI-1381306 imex : IM-14945-14	IntAct
					anti bait coimmunoprecipitation	EBI-7316419 MINT-4545014 imex : IM-11374-7	MINT
					enzyme linked immunosorbent assay	EBI-1540050	IntAct

Types of Pathway/Network Analysis

Goal

- 1 Enrichment of fixed gene sets
Identification of pre-built pathways or networks that are enriched in a set of mutated or differentially expressed genes
- 2 De novo sub-network construction and clustering
- 3 Pathway-based modeling
Evaluation of potential network rules that would be consistent with the identified set of mutated, differentially expressed genes

Tools

What biological processes are altered in this cancer?

Are new pathways altered in this cancer? Are there clinically-relevant tumour subtypes?

How are pathway activities altered in a particular patient? Are there targetable pathways in this patient?

Output

Mo

Enriched network Depleted network Extended network Mutated (seed) proteins Loss-Of-Function Network Signature Gain-Of-Function Network Signature

1) Enrichment of Fixed Gene Sets

- Covered in previous Modules.
- Most popular form of pathway/network analysis.
- Advantages:
 - Easy to perform.
 - Many good end-user tools.
 - Statistical model well worked out.
- Disadvantages:
 - Many possible gene sets
 - Gene sets are heavily overlapping; need to sort through lists of enriched gene sets!
 - “Bags of genes” obscure regulatory relationships among them.

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

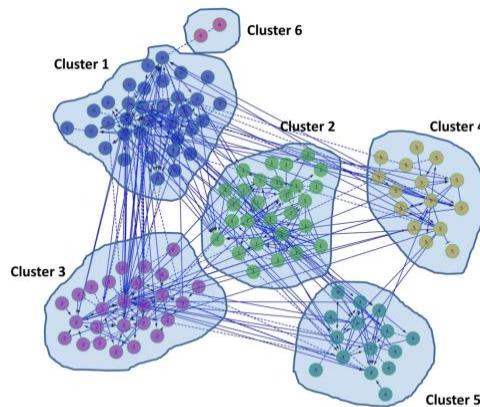
Popular Network Clustering Algorithms

- 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.

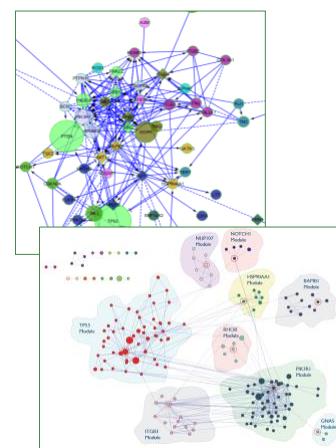
Clusters are mutually exclusive meaning that nodes are not shared between the clusters.

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.

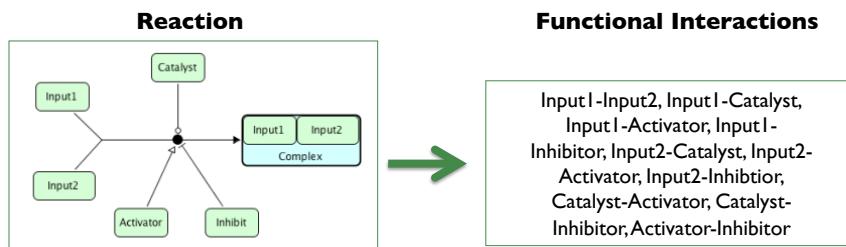


Module 4

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

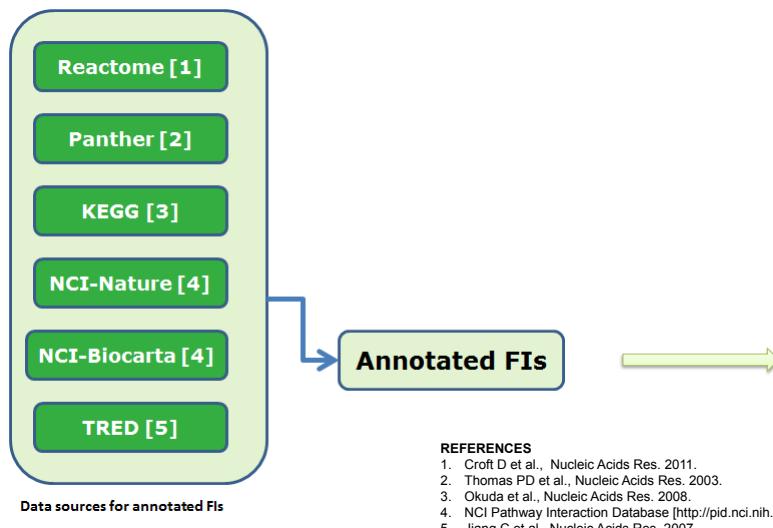


Method and practical application: A human functional protein interaction network and its application to cancer data analysis, [Wu et al. 2010 Genome Biology](#).

Module 4

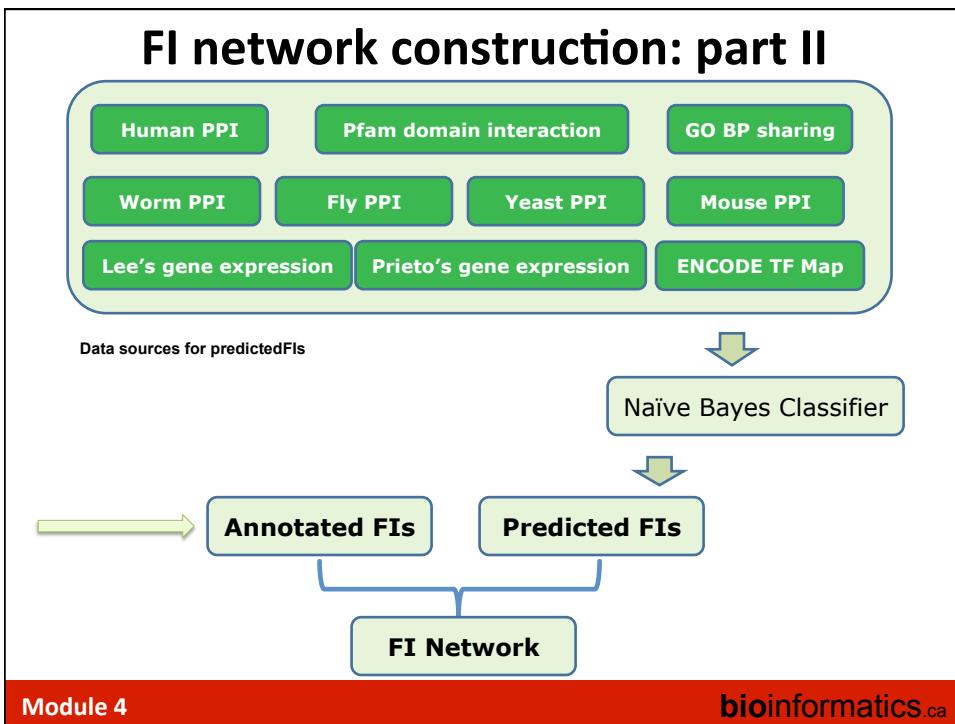
bioinformatics.ca

FI network construction: part I



Module 4

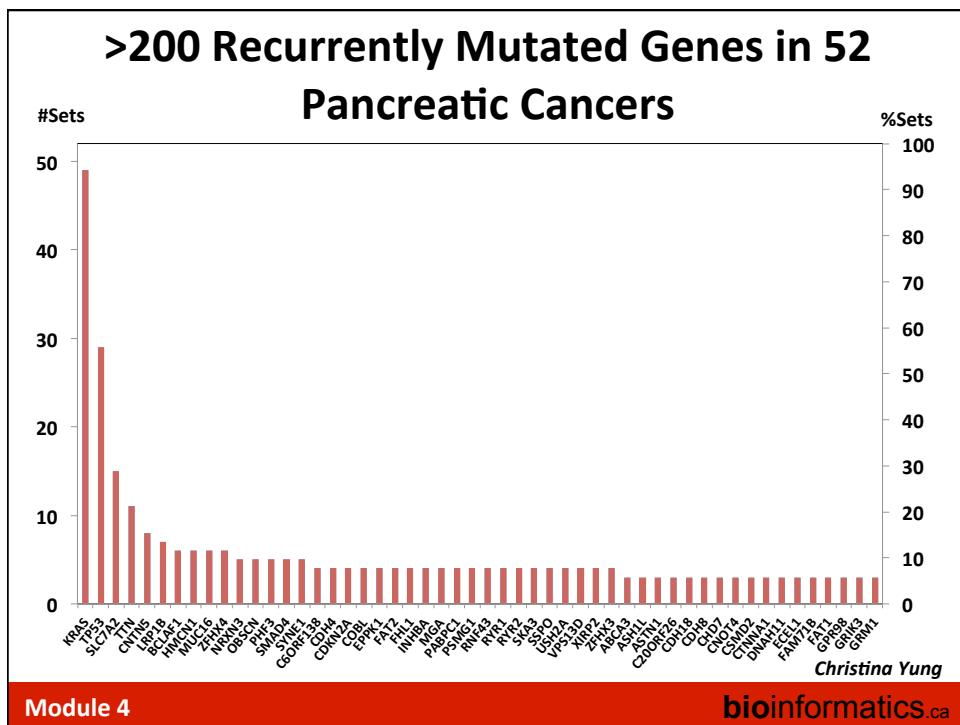
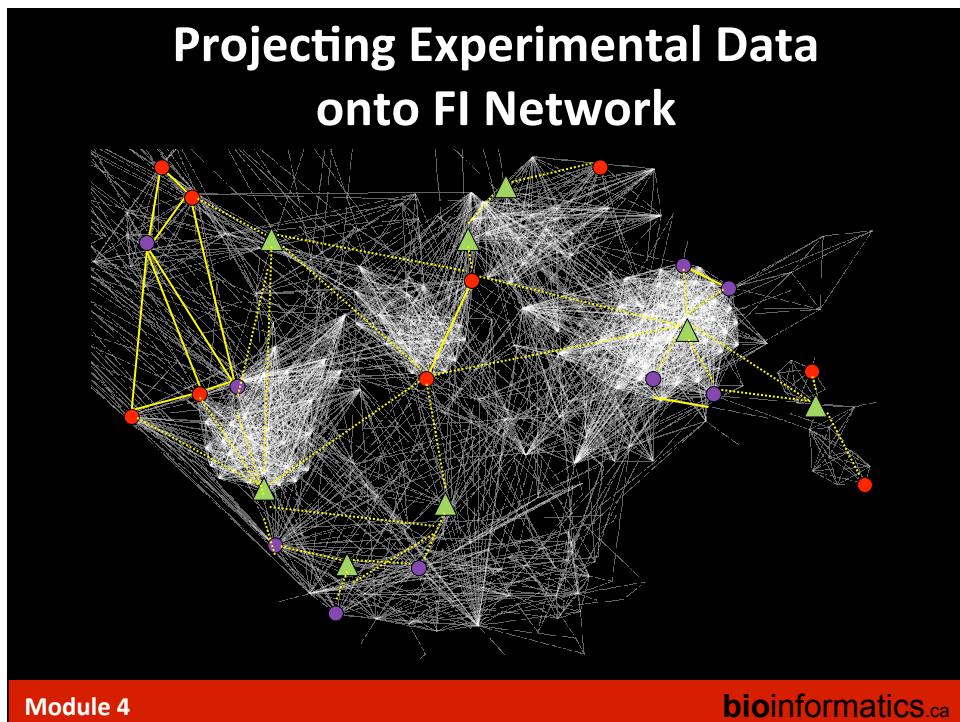
bioinformatics.ca

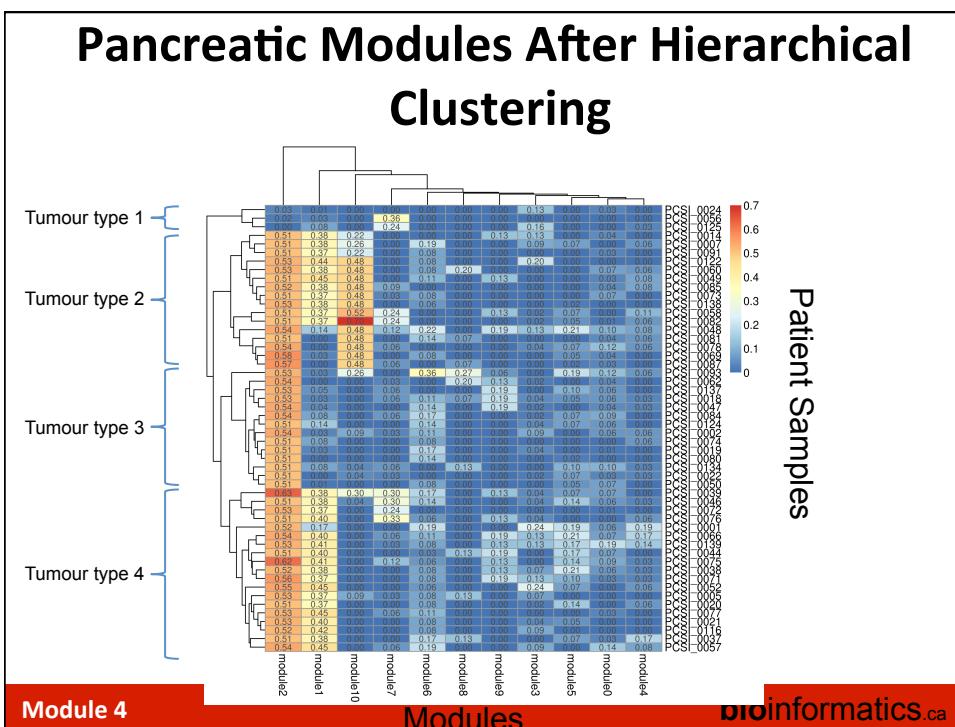
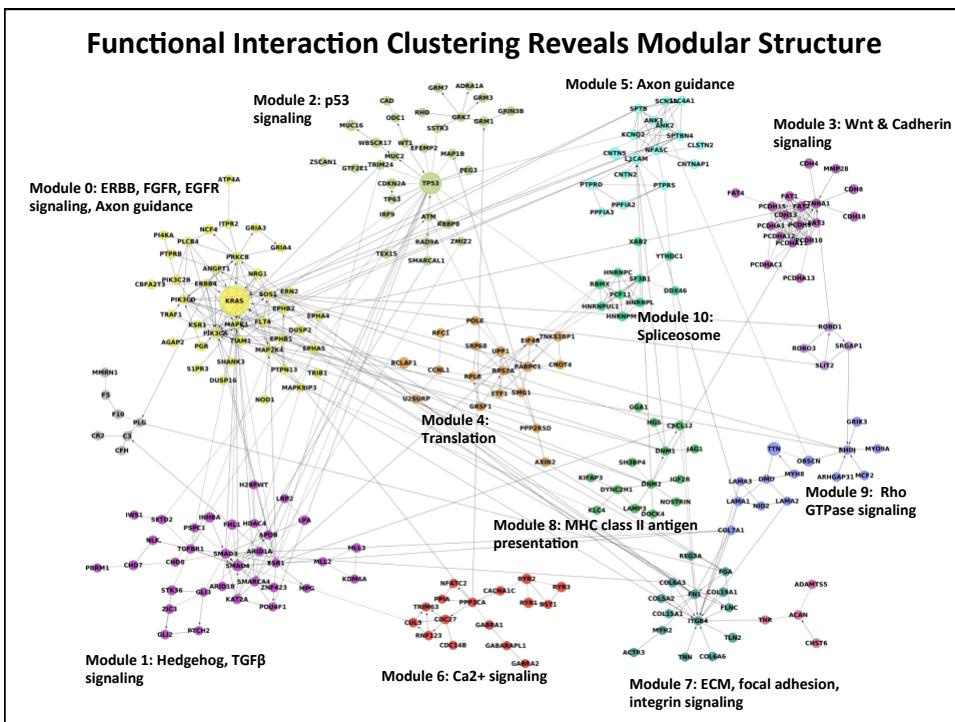


Composition of the FI network (2014)

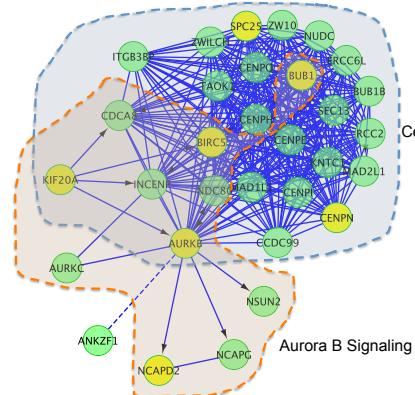
Data Source	Total FIs	SwissProt Ids	Percentage
Pathway	205,158	8,804	44%
Predicted	131,061	9,081	45%
Total	336,219	11,780	58%

Module 4 **bioinformatics.ca**

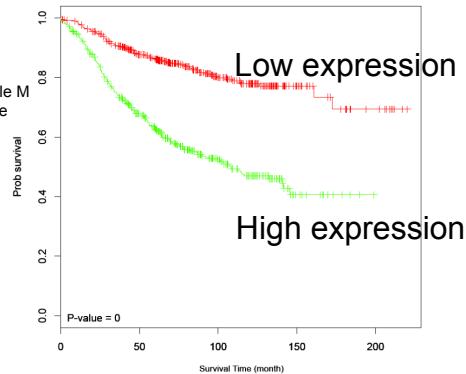




Module-Based Prognostic Biomarker in ER+ Breast Cancer



Measure levels of expression of the genes in this network module



Guanming Wu

Module 4

bioinformatics.ca

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

Module 4

bioinformatics.ca

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

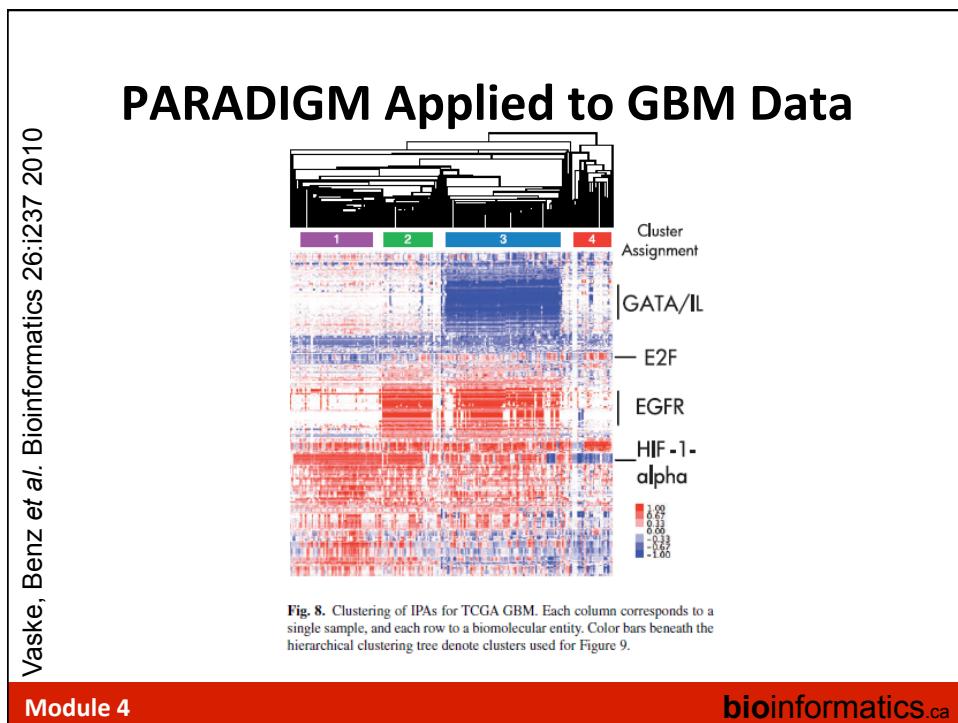
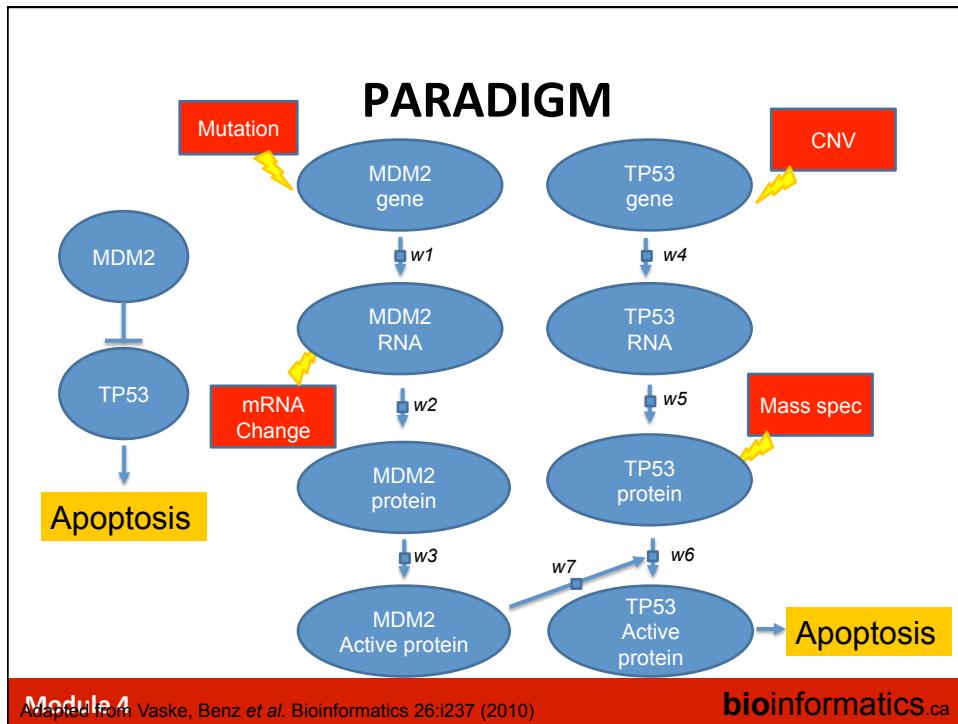
bioinformatics.ca

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



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/reactomefis>

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/>

Module 4

bioinformatics.ca

We are on a Coffee Break &
Networking Session



Module 4

bioinformatics.ca