bioinformatics.ca

bioinformaticsdotca.github.io

CC BY-SA 4.0 DEED

Attribution-ShareAlike 4.0 International

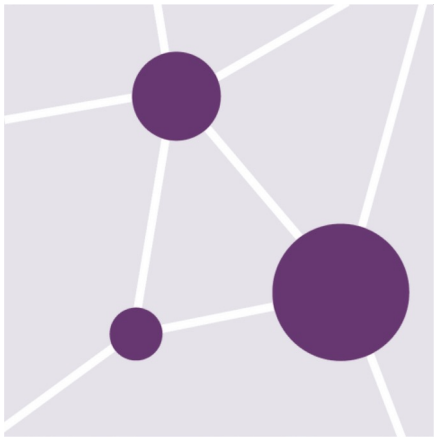
Canadian Bioinformatics Workshops

Module 4: Integrative Assignment

Constance Li

Pathway and Network Analysis

May 13, 2026



UNIVERSITY OF CALGARY

Arnie Charbonneau Cancer Institute

Learning Objectives

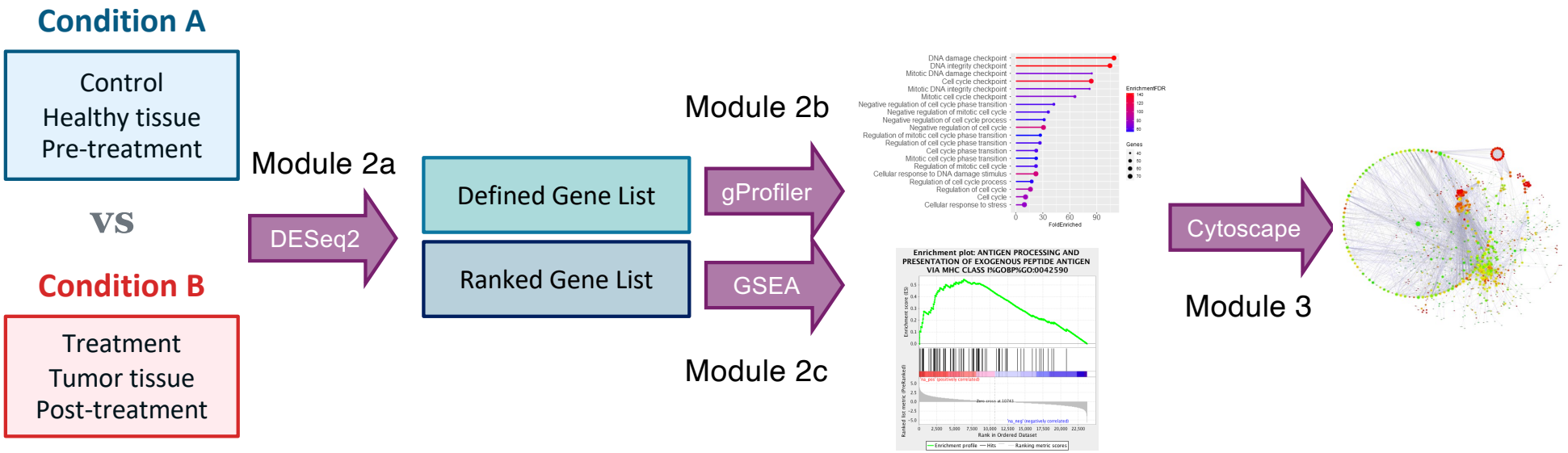
1. Describe similarities and differences between pathways and networks

2. Distinguish between types of pathway/network analysis and their goals

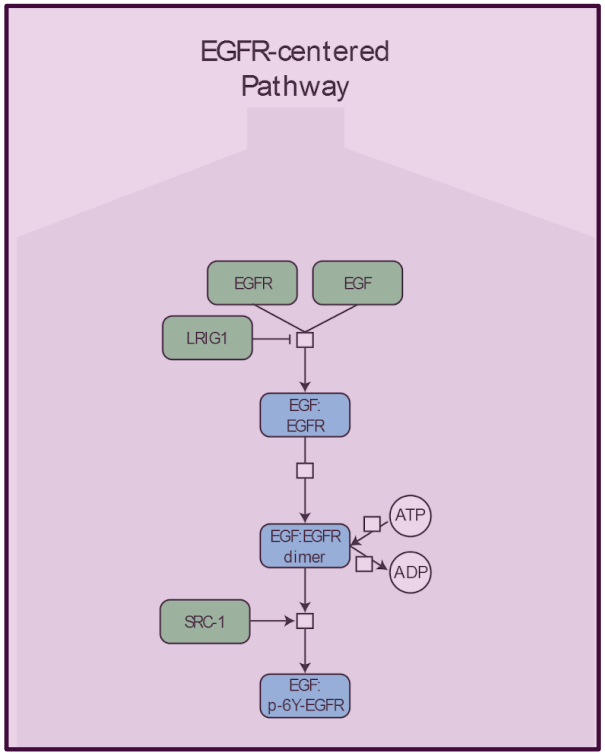
3. Identify and apply an appropriate method for mutations-based pathway analysis

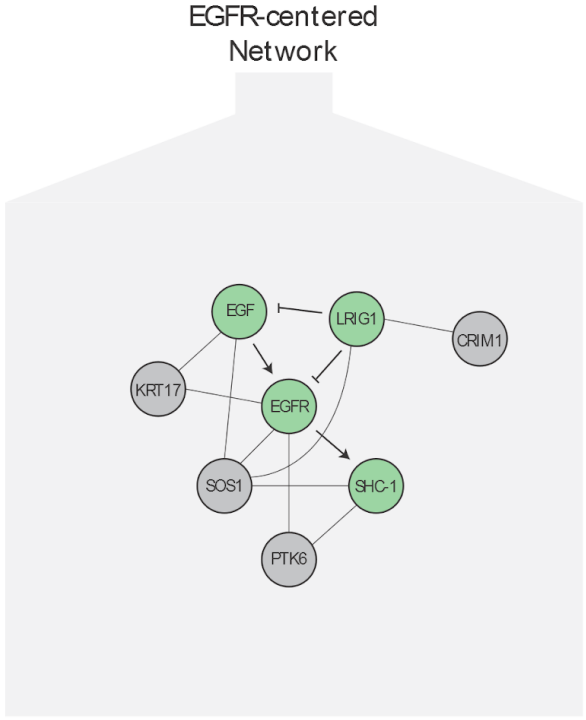
4. List examples of pathway/network databases and analysis tools

What we've learned so far



Pathways and networks





Pathway databases

Advantages

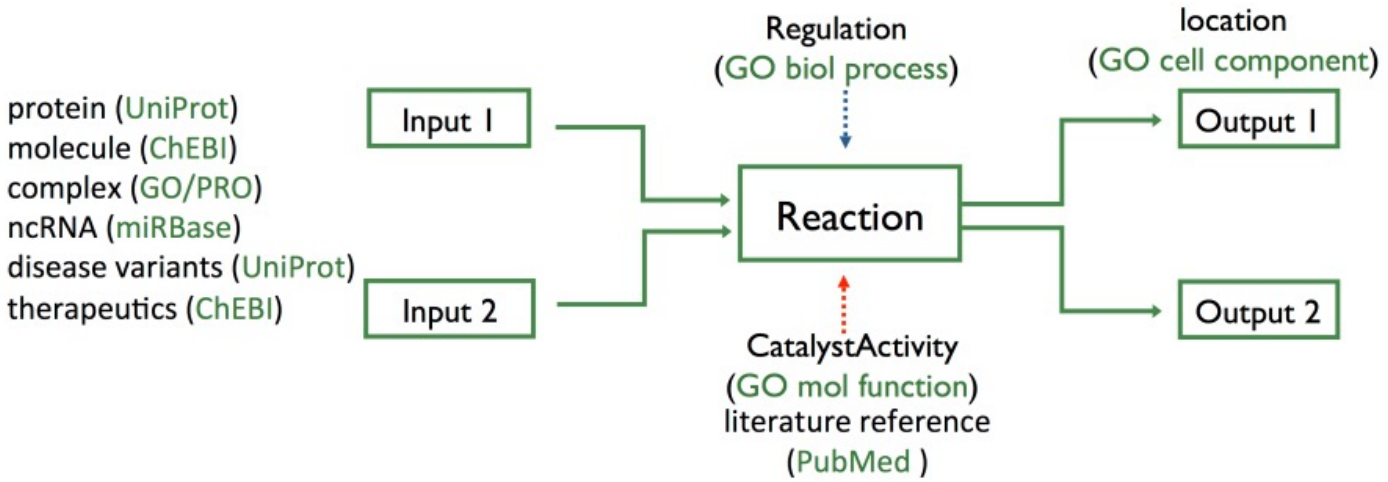
- Usually curated
- Biochemical view of biological processes
- Cause and effect captured
- Human-interpretable visualizations

Disadvantages

- Sparse coverage of genome
- Different databases disagree on boundaries of pathways

Reaction-Network databases

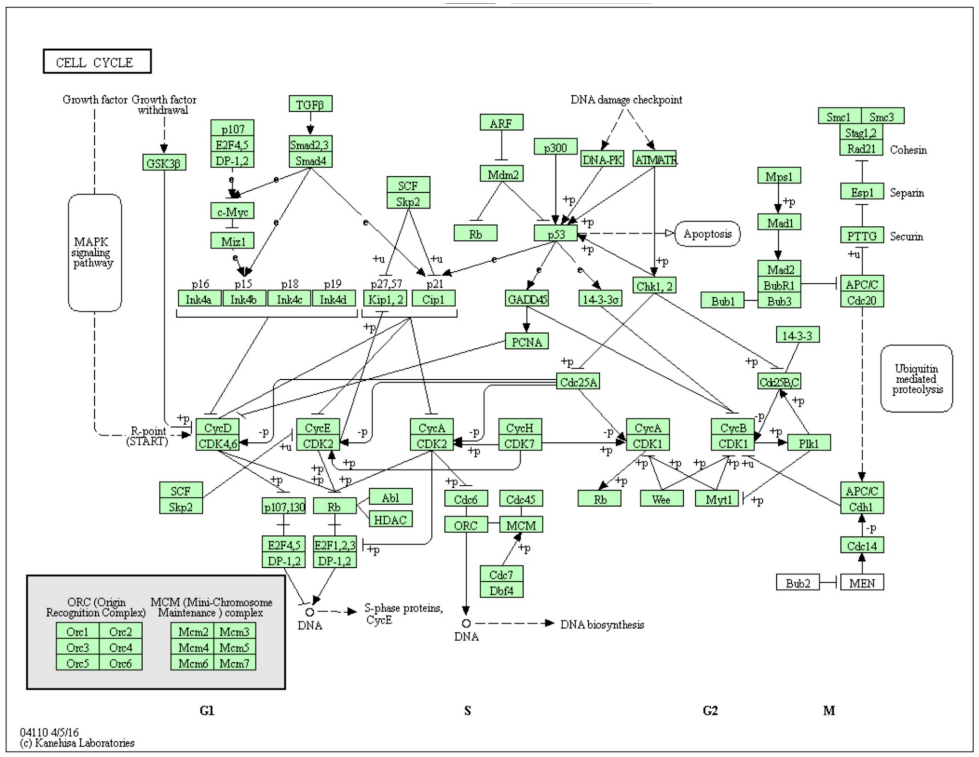
- Reactome & KEGG
  - explicitly describe biological processes as a series of biochemical reactions.
  - represents many events and states found in biology.

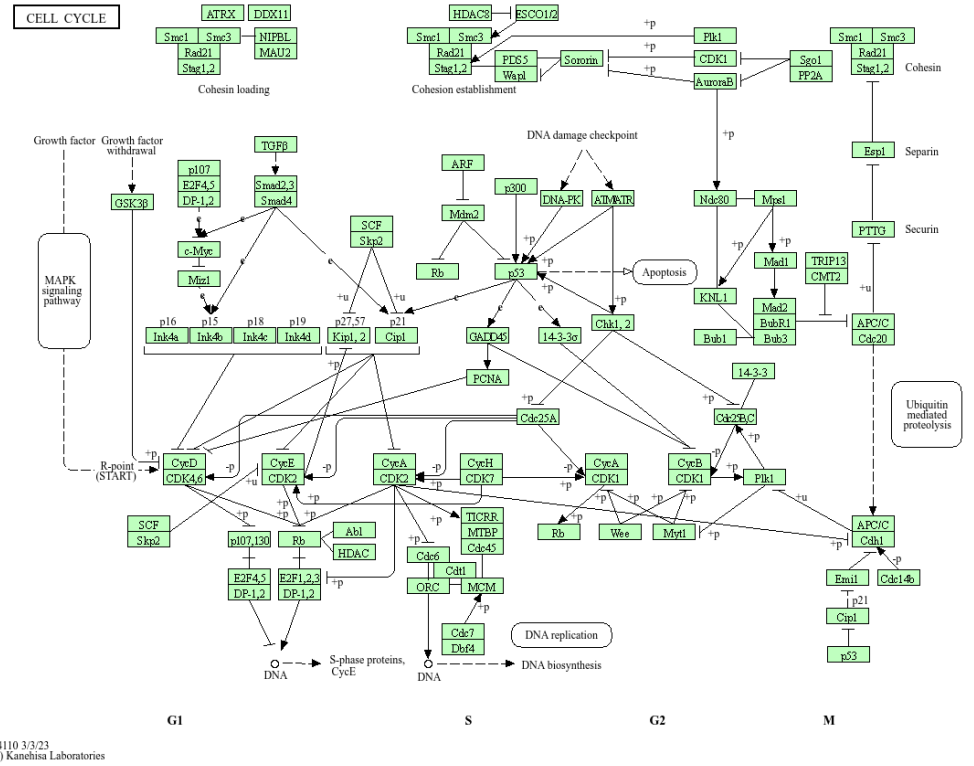


KEGG

- Kyoto Encyclopedia of Genes and Genomes(KEGG):
  - A vast library of information: fully sequenced genomes, genes, proteins, pathways, chemical compounds pertaining to over a hundred different species of both prokaryotes and eukaryotes.
  - KEGG PATHWAY is a collection of manually drawn pathway maps representing knowledge on the molecular interaction and reaction networks for Metabolism, Cellular Processes, Organismal Systems, Human Diseases and Drug Development
- Subscription required for access to underlying data for analysis use

KEGG Cell Cycle

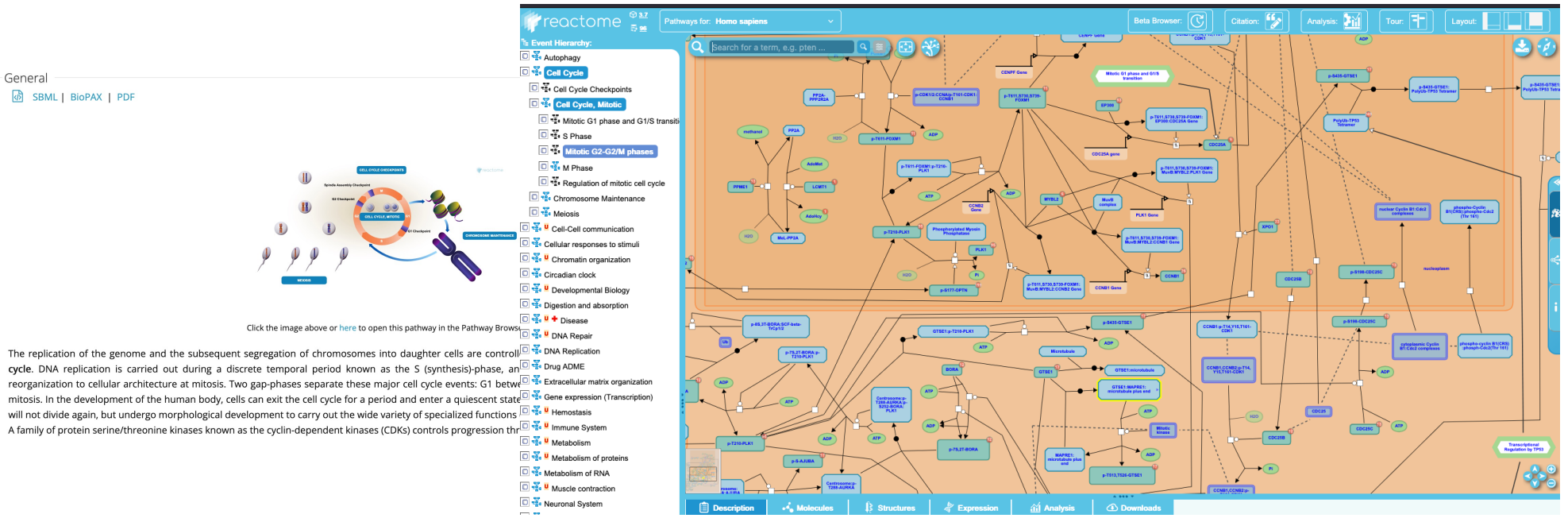




Reactome

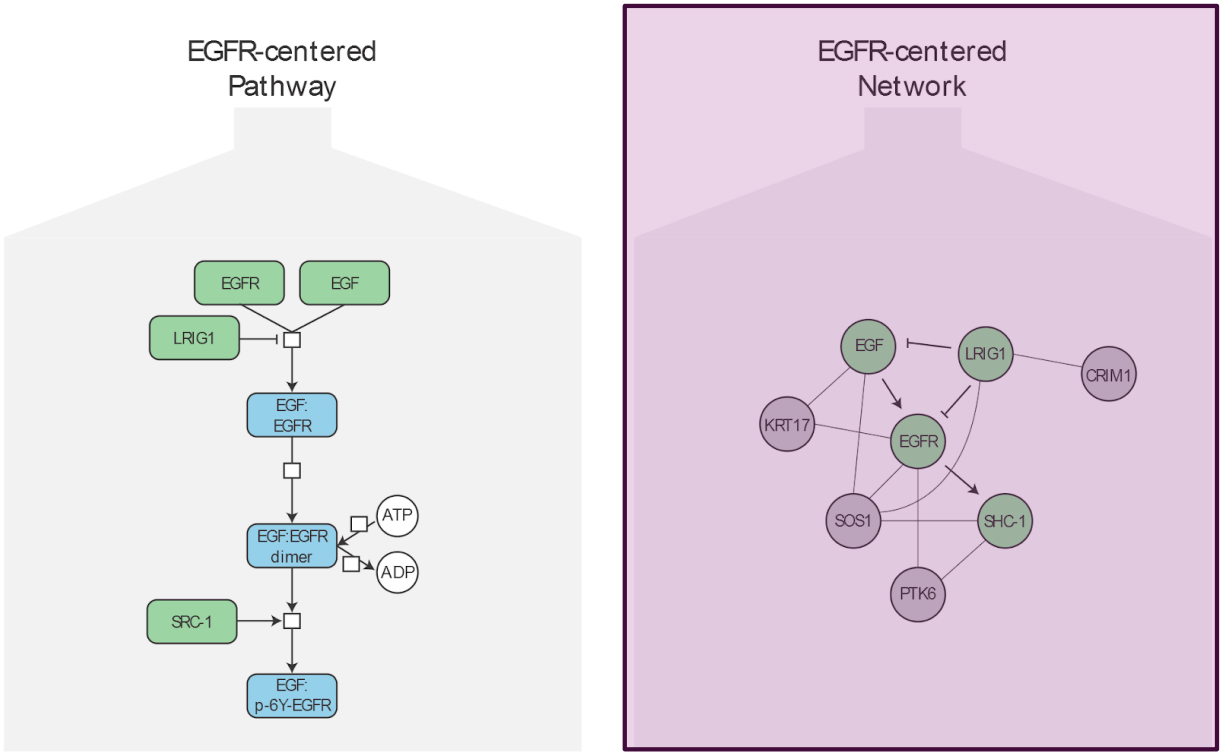
- Open source and open access pathway database
- Curated human pathways encompassing metabolism, signaling, and other biological processes
- Rigorous curation standards – every pathway is traceable to primary literature
- Cross-reference to many other bioinformatics databases
- Provides data visualization and analysis tools
  - Google-map style reaction diagrams and textbook-style illustrations with overlays
  - Find pathways containing your gene list
  - Calculate gene overrepresentation in pathways
  - Find corresponding pathways in other species

Reactome Cell Cycle



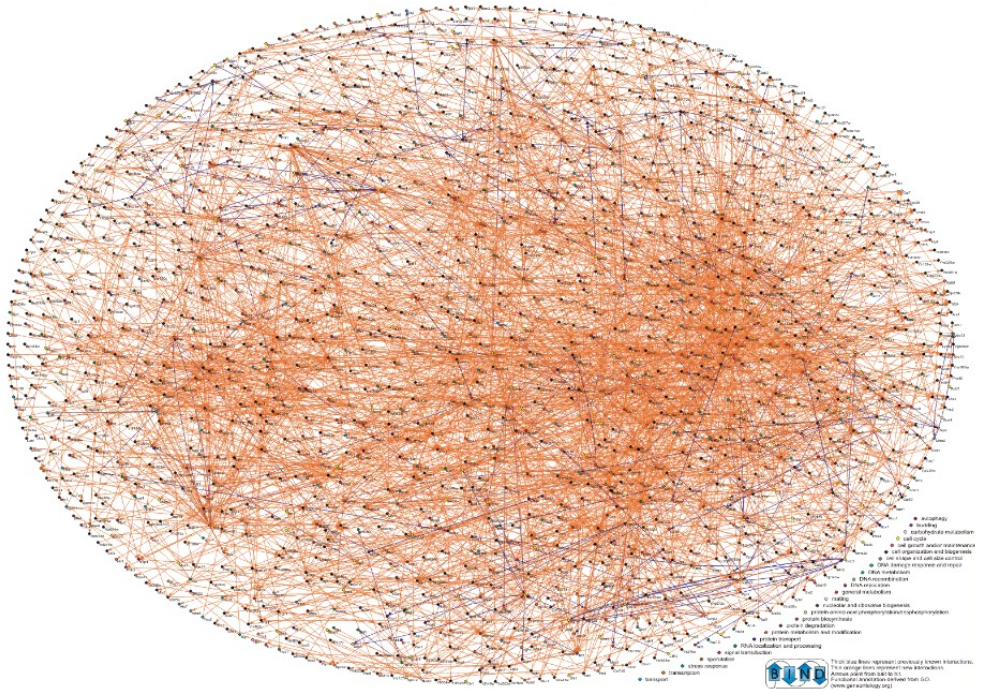
https://reactome.org/PathwayBrowser/#/R-HSA-453274&PATH=R-HSA-1640170,R-HSA-69278,R-HSA-453274

Pathways and networks



Networks

- Pathways capture only the “well understood” portion of biology
- Networks cover less well understood relationships:
  - Genetic interactions
  - Physical interaction
  - Coexpression
  - GO term sharing
  - Adjacency in pathways

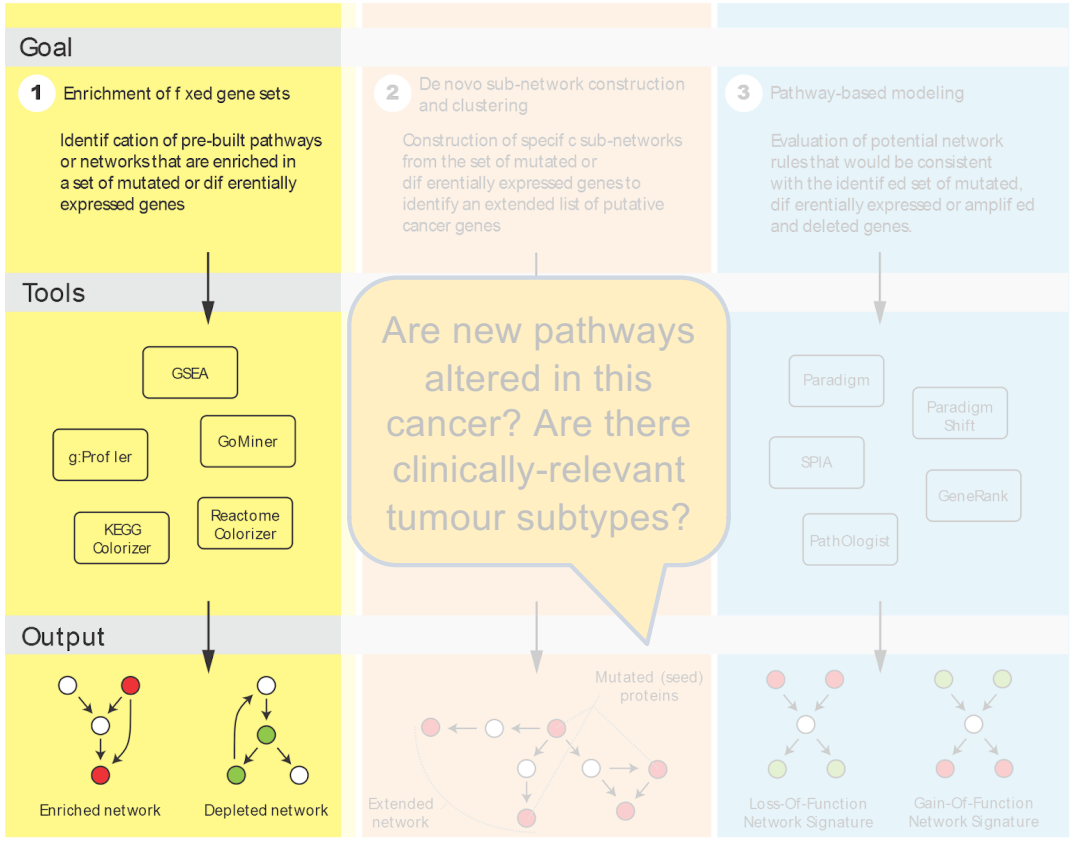


Network Databases

- 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 physical and genetic interactions from literature; 89K genes & 2.1M interactions from 80 species (https://thebiogrid.org/)
  - IntAct – Curated interactions from literature; 143K interactors & 1.5M interactions from 9000 species. (https://www.ebi.ac.uk/intact/home)
  - GeneMANIA - Compendium of 2.8K gene association networks representing 167K genes and 660M interactions from 9 species

Pathway/Network Analysis

What biological processes are altered in this cancer?



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

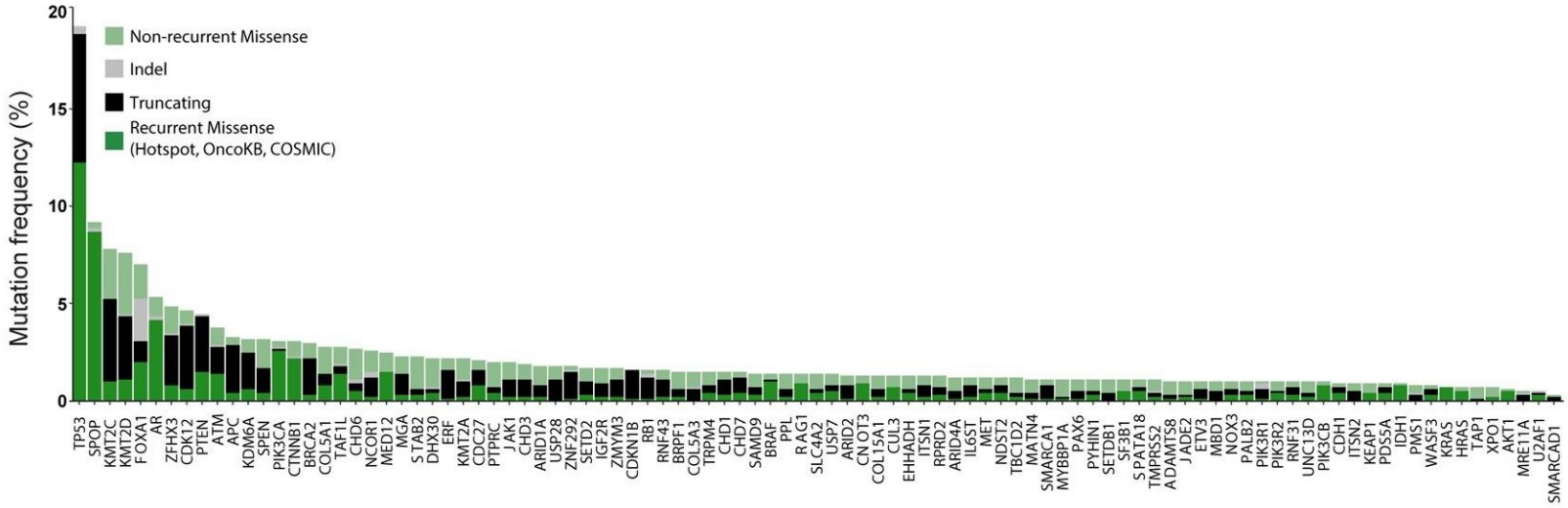
Given the pathway structure, what functional states explain my observations? Are there targetable pathways in this patient?

Enrichment of Fixed Gene Sets

- Covered in Module 2
- Most popular form of pathway/network analysis
- Overrepresentation analysis vs functional class scoring
- Advantages:
  - Easy to perform
  - Many good end-user tools
  - Statistical model well worked out
- Disadvantages:
  - Many possible gene sets
  - Gene sets are heavily overlapping
  - “Bags of genes” obscure regulatory relationships among them
  - Not well-suited for mutations data

Working with mutation data

- Standard pathway analysis doesn't work well for somatic mutations
  - Sparseness and heterogeneity result in a 'long tail' of rare cancer mutations



Nat Genet. 2018 Apr 2;50(5):645–651. doi: 10.1038/s41588-018-0078-z

Working with mutation data

- Standard pathway analysis doesn't work well for somatic mutations
  - Sparseness and heterogeneity result in a 'long tail' of rare cancer mutations
  - We often have small gene lists of ~20-50 significant genes
  - 'Driver' and 'passenger' mutations have different biological implications
- What does this mean for ORA and GSEA?
  - ORA needs many genes hitting same pathway: mutations too sparse
  - GSEA needs ranked list: mutations aren't naturally ranked
  - Both miss connections between related pathways
  - Neither identifies convergent mutations in the same process

Nat Genet. 2018 Apr 2;50(5):645–651. doi: 10.1038/s41588-018-0078-z

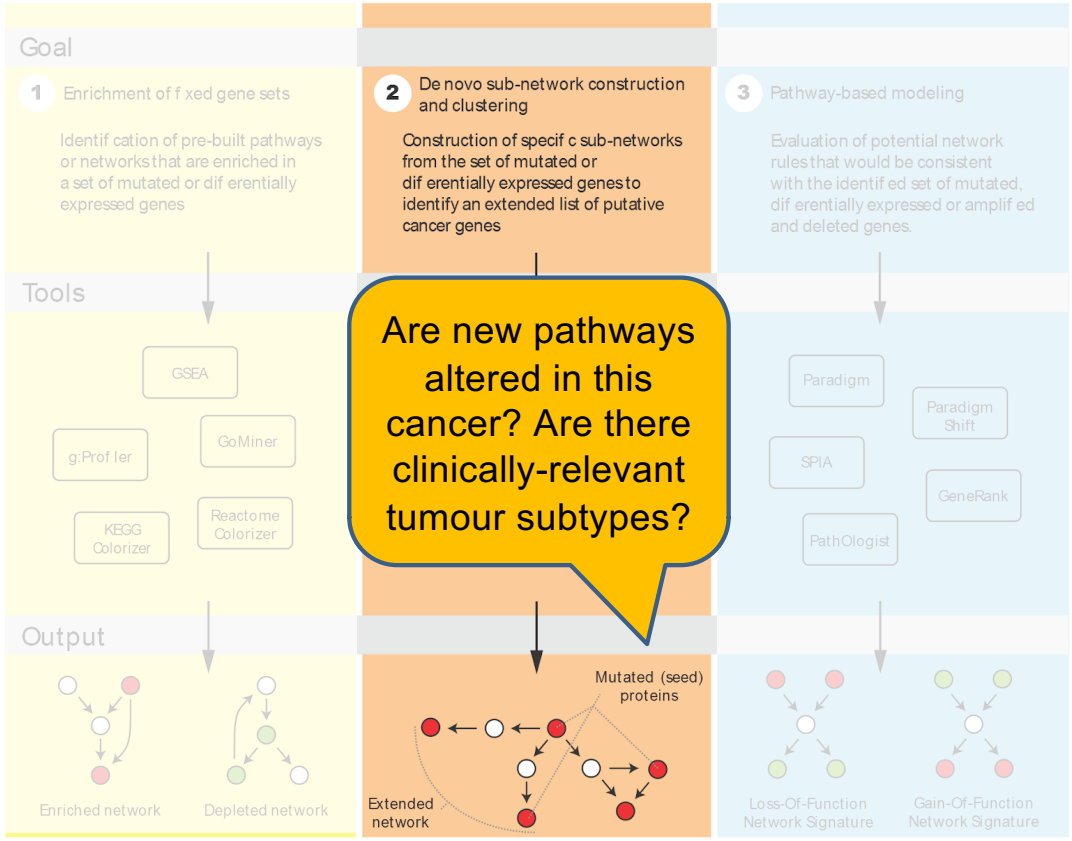
Leverage networks to solve the sparse data problem

Key Insight

Mutated genes that are functionally connected are more likely to be drivers affecting the same biological process.

Pathway/Network Analysis

What biological processes are altered in this cancer?



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

Given the pathway structure, what functional states explain my observations? Are there targetable pathways in this patient?

https://www.nature.com/articles/nmeth.3440

De Novo Subnetwork Construction & Clustering

- Are new pathways altered in this cancer? Are there clinically-relevant tumour subtypes?
- Apply list of altered molecular entities (genes, proteins, RNAs) to a biological network
- Identify “topologically unlikely” configurations
  - *eg*. 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

Finding convergence in mutations

- Look for shared functional consequence in different genes

Example: RTK/RAS/PI3K Signaling Module

Patient 1	Patient 2	Patient 3
EGFR mutation	**ERBB2 mutation**	**KRAS mutation**

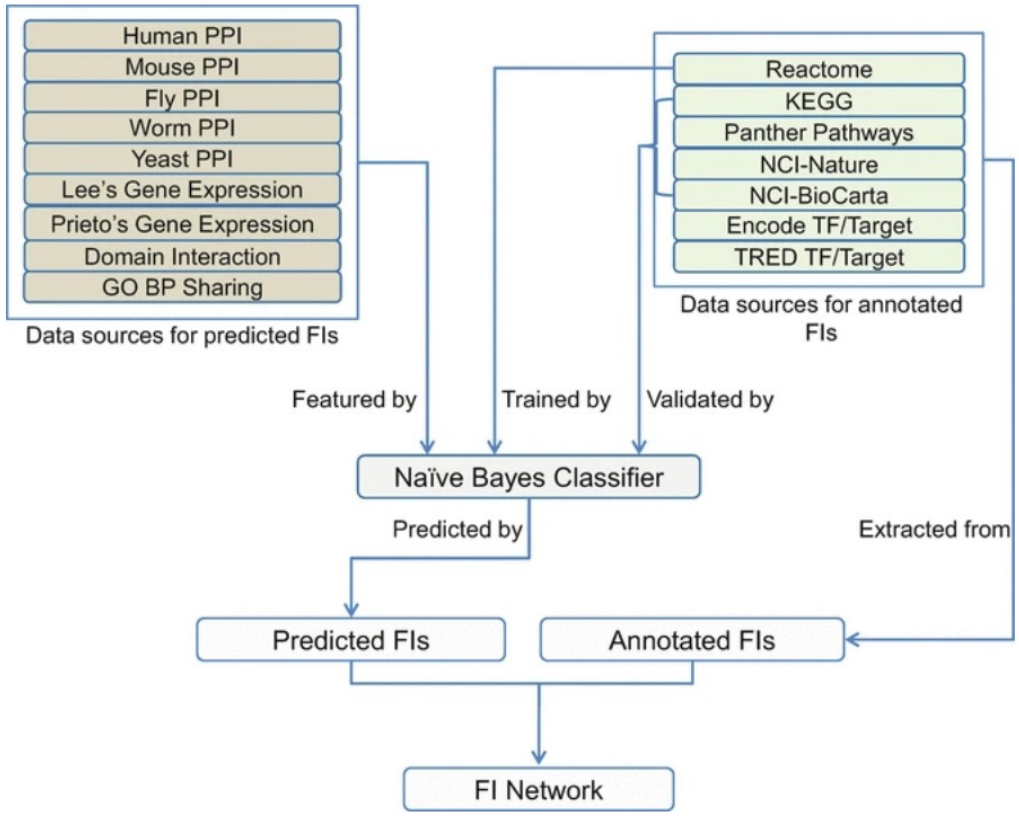
Same Module: RTK Signaling

Statistical Power Through Aggregation

Gene X mutated in 5% + Gene Y mutated in 3% = Not significant individually.
But if X and Y form a connected module, then 8% of tumors affect this pathway →
Now statistically significant!

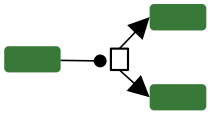
ReactomeFI

- Reactome Functional Interaction allows us to analyze genes in a network context
  - To reveal relationships between genes
  - To inform our biological understanding of genes and gene products
- A functional interaction is an interaction where two proteins are involved in the same reaction



Protein Bioinformatics. Methods in Molecular Biology, vol 1558. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-6783-4_11

The Reactome FI Network

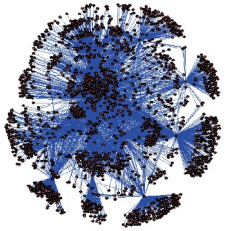


Curated Pathway Dbs

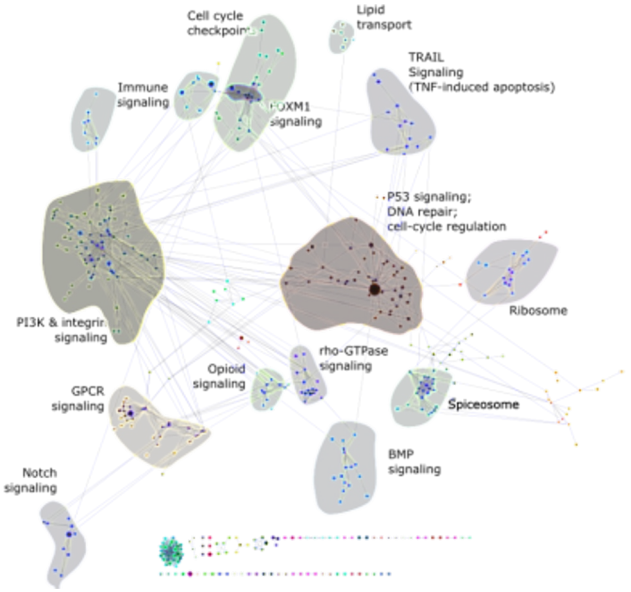
+

Uncurated Interaction Evidence

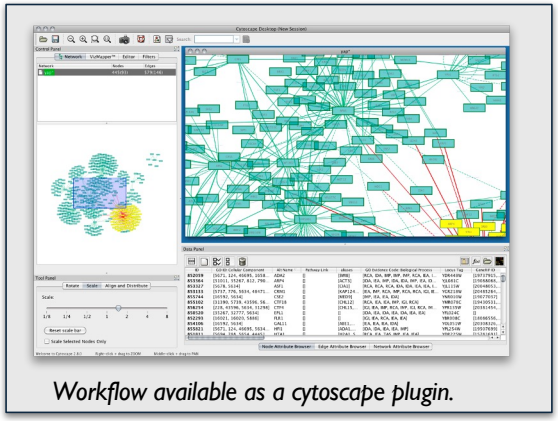
Machine Learning →



Extract and Cluster "Disease Genes" →



Reactome Functional Interaction Network
(~11,000 proteins; 270,000 interactions)



Workflow available as a cytoscape plugin.

A human functional protein interaction network and its application to cancer data analysis, Wu et al. 2010 Genome Biology

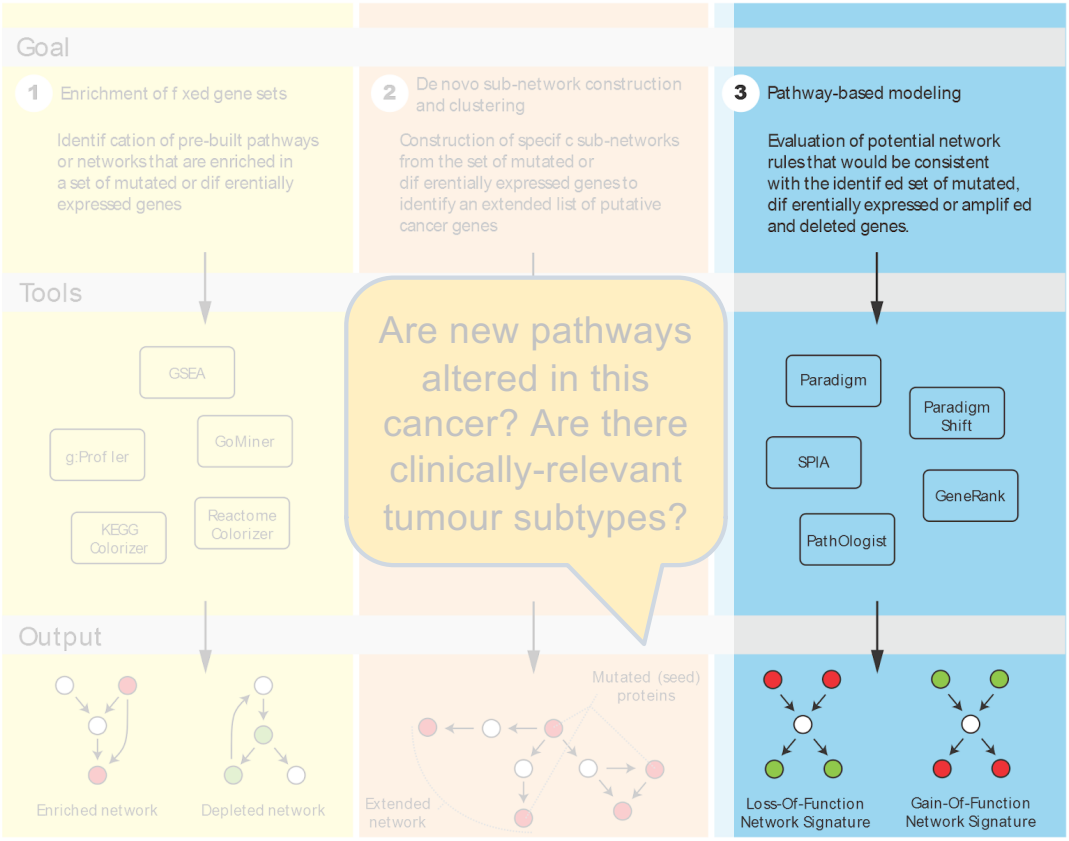
Other network clustering algorithms

- GeneMANIA
  - “Birds of a feather” principle.
  - Very useful for finding genes that are related to an experimentally defined set.
- HotNet
  - Finds “hot” clusters based on propagation of heat across metallic 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, and survival correlation analysis)

Pathway/Network Analysis

What biological processes are altered in this cancer?

Covered in Module 2
Most popular form of pathway/network analysis.



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

Given the pathway structure, what functional states explain my observations? Are there targetable pathways in this patient?

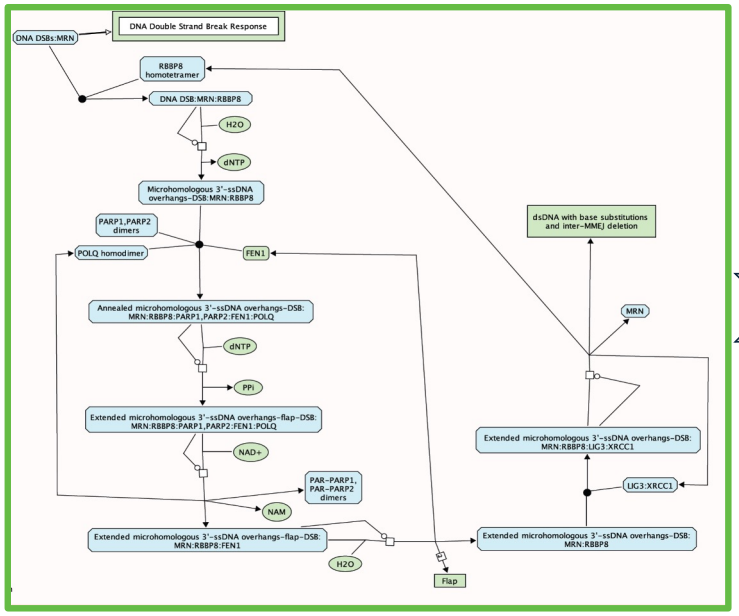
Pathway-Based Modeling

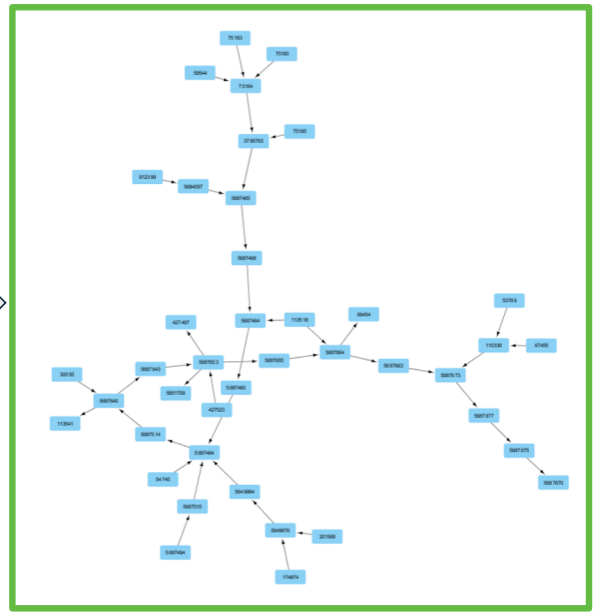
- Given the pathway structure, what functional states explain my observations?
- Apply list of altered molecular entities (genes, proteins, RNAs) to biological pathways
- Preserve detailed biological relationships
- Attempt to integrate multiple molecular alterations together to yield lists of altered pathway activities
- Pathway modeling shades into Systems Biology

Types of Pathway-Based Modeling

- Partial differential equations, e.g. CellNetAnalyzer
  - Mostly suited for biochemical systems (metabolomics)
- Network flow models, e.g. NetPhorest
  - Mostly suited for kinase cascades (phosphorylation info)
- Transcriptional regulatory network-based reconstruction methods, e.g. ARACNe
- Logic Graphs and probabilistic graph models (PGMs)
  - Capture the logic of a pathway without needing rate/binding constants.
- Generative AI Models

Boolean Network Inference (PARADIGM)





Pathway View

Logic Graph View

Pathway Database (Reactome/NCI-PID/KEGG) → Convert to Graph → Overlay Patient Data (CNA + Expression) → Belief Propagation (inference) → Inferred Pathway Levels per node per patient

Generative AI Network Models

- Large Language Models (GPT-3) are trained to predict masked text:

  `The quick brown ? jumped over the lazy dog.`

- Generative pathway models are trained to predict gene network perturbations:

How to choose a tool?

- What is your biological question?
- Will the tool work for your model?
- Does the tool include reasonable gene sets (pathway databases)?
- Are the pathway databases up to date?
- Is the statistical output sufficient?
- Do you like the output style? Will it work for downstream analysis?
- Can you connect it with network visualization tools like Cytoscape?

Some tips for good PNA and beyond

Name files clearly

Include analysis type and data source

Record parameters

Document FDR, gene set sizes, databases

Start stringent

Begin at FDR 0.01, relax if needed

Look for convergence

Same pathways across methods = confidence

Think biologically

Do results fit disease mechanism?

Identify hubs

Genes connecting data types may be drivers

Summary

- Pathway/network analysis is an active and changing field
- There are three major goals for pathway analysis
  - Enrichment of fixed gene sets (Module 2)
  - De novo sub-network construction and clustering (Module 4 lab)
  - Pathway-based modeling (Module 4 lab Extra content)
- The appropriate method depends on your goal and data

Resources

- BioGRID
  - http:// www.thebiogrid.org
- IntAct
  - http://www.ebi.ac.uk/intact/
- KEGG
  - http:// www.genome.jp/kegg
- Reactome
  - http:// www.reactome.org
- GeneMANIA
  - http://www.genemania.org
- HotNet
  - http://compbio.cs.brown.edu/projects/hotnet/
- HyperModules
  - http://apps.cytoscape.org/apps/hypermodules
- Reactome Cytoscape FIViz App
  - http://apps.cytoscape.org/apps/reactomefis
- 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
- Pathway Prediction Evaluation Paper
  - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9216552/

Let's Take a Break

Workshop Sponsors