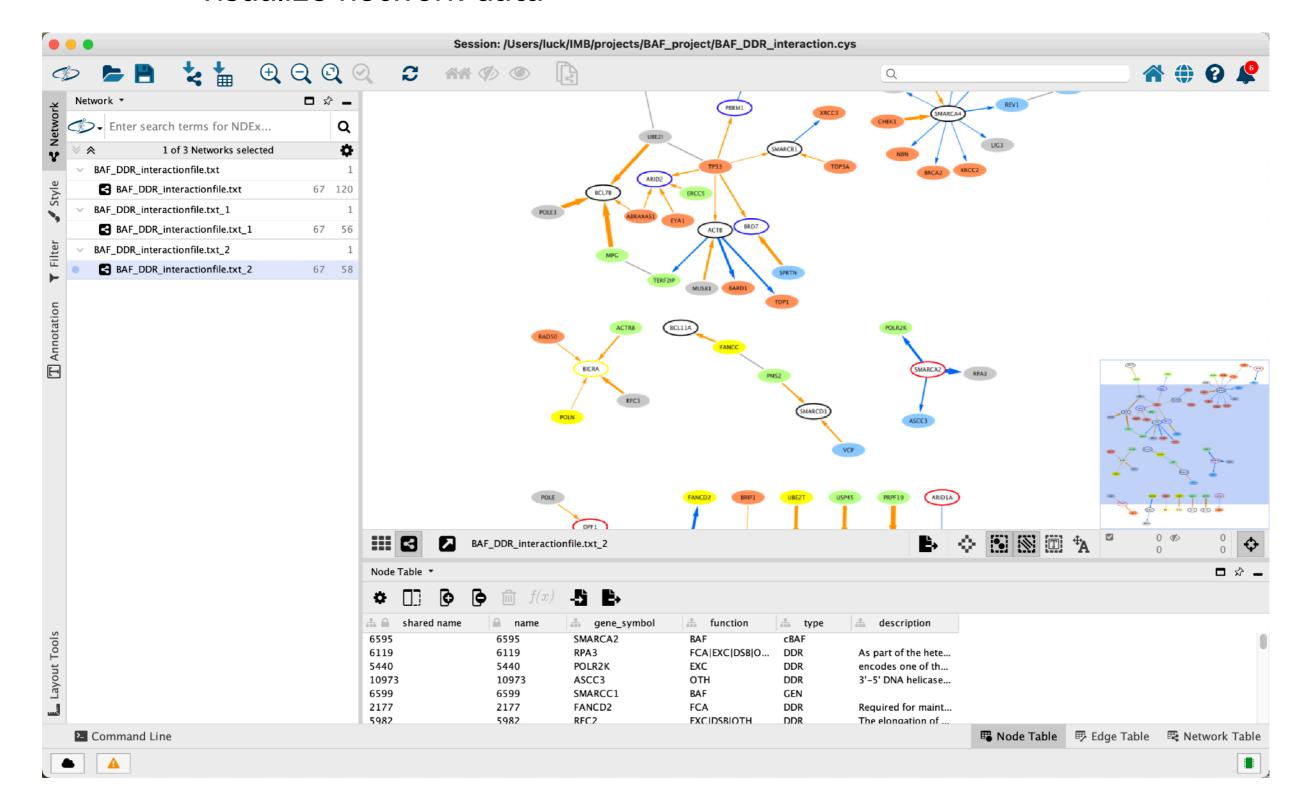Master Module
Proteinbiochemistry and Bioinformatics
March 2022

Session: Protein interaction networks

5. Visualizing and analyzing networks using Cytoscape

Introduction into Cytoscape

Introduction into Cytoscape

Software to:
- explore a set of genes with protein interaction data
- visualize network data

Introduction into Cytoscape

Example data:

Introduction into Cytoscape

Example data:

Evidence for 28 genetic disorders discovered by combining healthcare and research data

De novo mutations in protein-coding genes are a well-established cause of developmental disorders[1]. However, genes known to be associated with developmental disorders account for only a minority of the observed excess of such de novo mutations[1,2]. Here, to identify previously undescribed genes associated with developmental disorders, we integrate healthcare and research exome-sequence data from 31,058 parent–offspring trios of individuals with developmental disorders, and develop a simulation-based statistical test to identify gene-specific enrichment of de novo mutations. We identified 285 genes that were significantly associated with developmental disorders, including 28 that had not previously been robustly associated with developmental disorders. Although we detected more genes associated with developmental disorders, much of the excess of de novo mutations in protein-coding genes remains unaccounted for. Modelling suggests that more than 1,000 genes associated with developmental disorders have not yet been described, many of which are likely to be less penetrant than the currently known genes. Research access to clinical diagnostic datasets will be critical for completing the map of genes associated with developmental disorders.

Kaplanis et al *Nature* 2020

Introduction into Cytoscape

Example data:

Evidence for 28 genetic disorders discovered by combining healthcare and research data

De novo mutations in protein-coding genes are a well-established cause of developmental disorders[1]. However, genes known to be associated with developmental disorders account for only a minority of the observed excess of such de novo mutations[1,2]. Here, to identify previously undescribed genes associated with developmental disorders, we integrate healthcare and research exome-sequence data from 31,058 parent–offspring trios of individuals with developmental disorders, and develop a simulation-based statistical test to identify gene-specific enrichment of de novo mutations. We identified 285 genes that were significantly associated with developmental disorders, including 28 that had not previously been robustly associated with developmental disorders. Although we detected more genes associated with developmental disorders, much of the excess of de novo mutations in protein-coding genes remains unaccounted for. Modelling suggests that more than 1,000 genes associated with developmental disorders have not yet been described, many of which are likely to be less penetrant than the currently known genes. Research access to clinical diagnostic datasets will be critical for completing the map of genes associated with developmental disorders.

Kaplanis et al *Nature* 2020

What is the relationship between these genes?

Introduction into Cytoscape

Example data:

Evidence for 28 genetic disorders discovered by combining healthcare and research data

De novo mutations in protein-coding genes are a well-established cause of developmental disorders[1]. However, genes known to be associated with developmental disorders account for only a minority of the observed excess of such de novo mutations[1,2]. Here, to identify previously undescribed genes associated with developmental disorders, we integrate healthcare and research exome-sequence data from 31,058 parent–offspring trios of individuals with developmental disorders, and develop a simulation-based statistical test to identify gene-specific enrichment of de novo mutations. We identified 285 genes that were significantly associated with developmental disorders, including 28 that had not previously been robustly associated with developmental disorders. Although we detected more genes associated with developmental disorders, much of the excess of de novo mutations in protein-coding genes remains unaccounted for. Modelling suggests that more than 1,000 genes associated with developmental disorders have not yet been described, many of which are likely to be less penetrant than the currently known genes. Research access to clinical diagnostic datasets will be critical for completing the map of genes associated with developmental disorders.

Kaplanis et al *Nature* 2020

What is the relationship between these genes?

In which processes are they involved?

Introduction into Cytoscape

Example data:

Evidence for 28 genetic disorders discovered by combining healthcare and research data

De novo mutations in protein-coding genes are a well-established cause of developmental disorders[1]. However, genes known to be associated with developmental disorders account for only a minority of the observed excess of such de novo mutations[1,2]. Here, to identify previously undescribed genes associated with developmental disorders, we integrate healthcare and research exome-sequence data from 31,058 parent–offspring trios of individuals with developmental disorders, and develop a simulation-based statistical test to identify gene-specific enrichment of de novo mutations. We identified 285 genes that were significantly associated with developmental disorders, including 28 that had not previously been robustly associated with developmental disorders. Although we detected more genes associated with developmental disorders, much of the excess of de novo mutations in protein-coding genes remains unaccounted for. Modelling suggests that more than 1,000 genes associated with developmental disorders have not yet been described, many of which are likely to be less penetrant than the currently known genes. Research access to clinical diagnostic datasets will be critical for completing the map of genes associated with developmental disorders.

Kaplanis et al *Nature* 2020

What is the relationship between these genes?

In which processes are they involved?

Which genes to focus on for downstream studies?

Introduction into Cytoscape

What you need on your computer:

- Cytoscape
- Kaplanis_28_genes_annotated.txt
- Kaplanis_285_genes.txt

- Kaplanis_285_genes_annotated.txt
- Kaplanis_285_IntAct.cys
- Cytoscape_tasks.txt

What's the plan?

- Quick Cytoscape tour by lecturer (~10 minutes)
- Network visualization and exploration on your own following some guidance provided in file Cytoscape_tasks.txt (45 minutes)
- Presentation of networks and discussion of questions (30 minutes)