Introduction to Bioinformatics

Pathway Analysis and Pathway Visualizations

The p-values????

detection p-values

- Is the signal there or not there, is it just some background noise or real signal?
- The p-value is calculated by comparing the detection p-values to the predefined value for more info you can refer to this link
- Used to filter out the probes/samples that contain too much background noise
- The smaller the better (0.01)

differential p-values

- Is the probe/gene is different in both of the samples
- Mostly using T-test if you have 2 different sample groups
- Used to filter out in which genes the different of methylation level is significant
- The smaller the better (0.01)

GSEA p-values

- How significant is the match between our gene set of interest to the one in the database? The likelihood of our gene set of interest belongs to the gene set in the database?
- Mostly using hypergeometric distribution
- Used to find out to which gene does our gene set of interest belong to
- The smaller the better (0.01)

The p-values????

detection p-values

differential p-values

GSEA p-values

Pathway Enrichment Analysis

- Analysis and interpretation of long list of data from array experiment represents a major challenge for many researchers.
- Analyses often result in long lists of genes that require an impractically large amount of manual literature searching to interpret.

Reimand J. 2019. Pathway enrichment analysis and visualization of omics data

Pathway Enrichment Analysis

- Analysis and interpretation of long list of data from array experiment represents a major challenge for many researchers.
- Analyses often result in long lists of genes that require an impractically large amount of manual literature searching to interpret.

Pathway enrichment analysis, which summarizes the large gene list as a smaller list of more easily interpretable pathways is a standard approach to address this problem

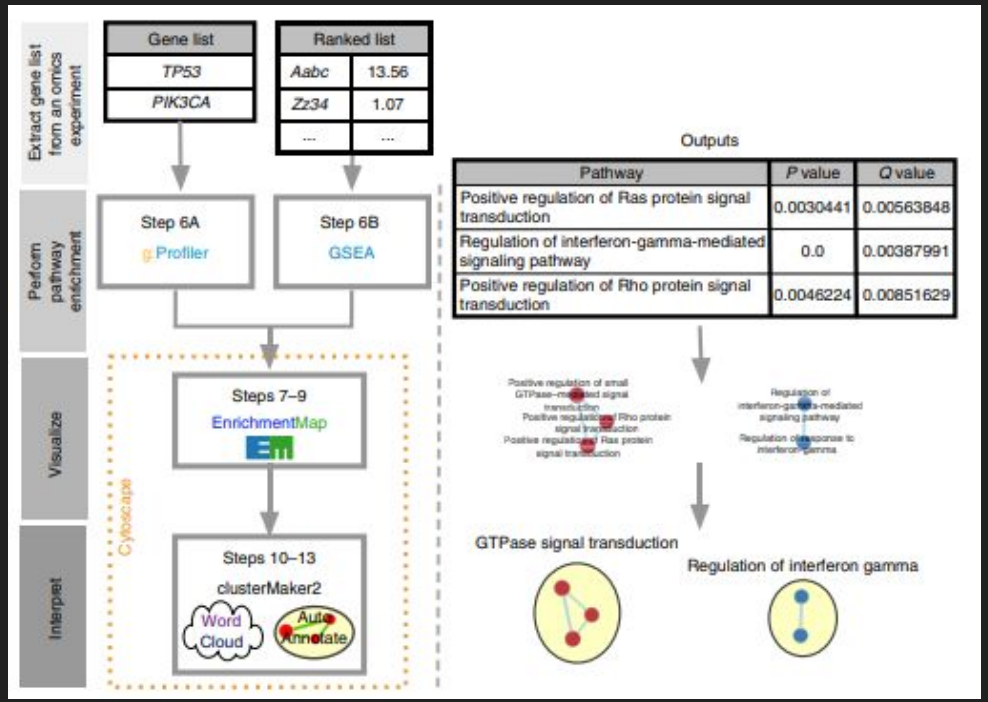
Reimand J. 2019. Pathway enrichment analysis and visualization of omics data

What do we get out of Pathway Analysis?

1. In-depth and contextualized findings to help understand the mechanisms of disease in question
2. Identification of genes and proteins associated with the etiology of a specific disease
3. Prediction of drug targets
4. Understand how to intervene therapeutically in disease processes
5. Conduct targeted literature searches

https://bioinformatics.mdanderson.org/MicroarrayCourse/Lectures09/Pathway%20Analysis.pdf

Pathway Enrichment Analysis



The image on the left is the usual process that we have to go through when we're performing PEA.

Reimand J. 2019. Pathway enrichment analysis and visualization of omics data

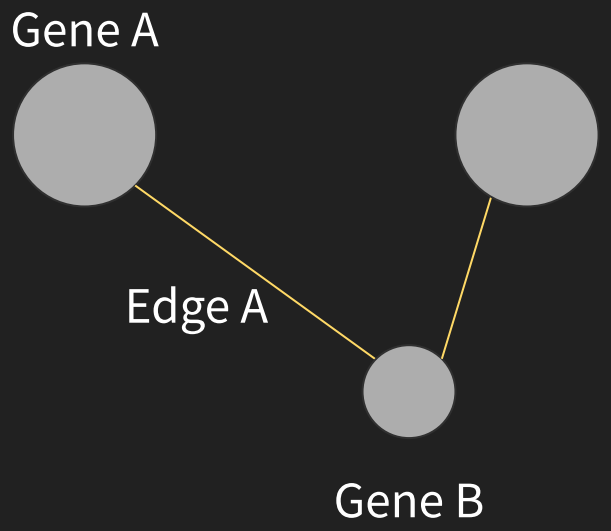
Important Keywords

1. **Pathway**. Genes that work together to carry out a biological process.
2. **Gene set**. A set of related genes. Gene sets can be based on various relationships between genes
3. **Gene list of interest**. The list of genes derived from an omics experiment that is input to pathway enrichment analysis.
4. **Ranked gene list**. List of genes that are ranked based on some value to provide more information for pathway enrichment analysis.

Reimand J. 2019. Pathway enrichment analysis and visualization of omics data

When working with Pathway Analysis and Network Viz we will also hear these keywords

1. **Nodes**
    a. Symbolizes a list of, for example, genes.
    b. This is essentially a one-dimensional representation of the data
2. **Edge**
    a. The thing that links nodes together

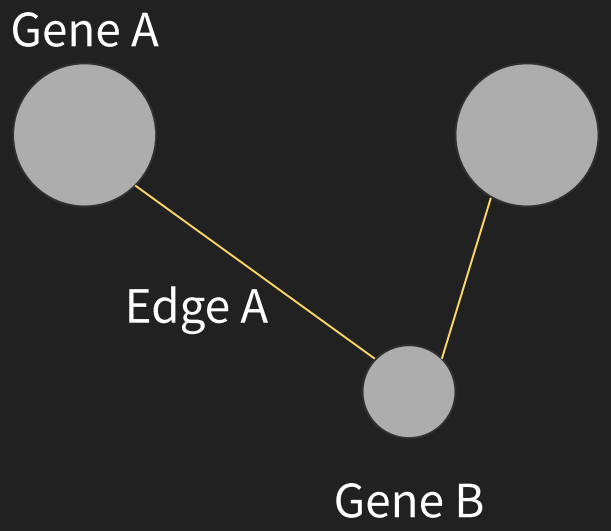


Reimand J. 2019. Pathway enrichment analysis and visualization of omics data

Networks are defined by their edges

Undirected

No difference between Gene A - Gene B and Gene B - Gene A

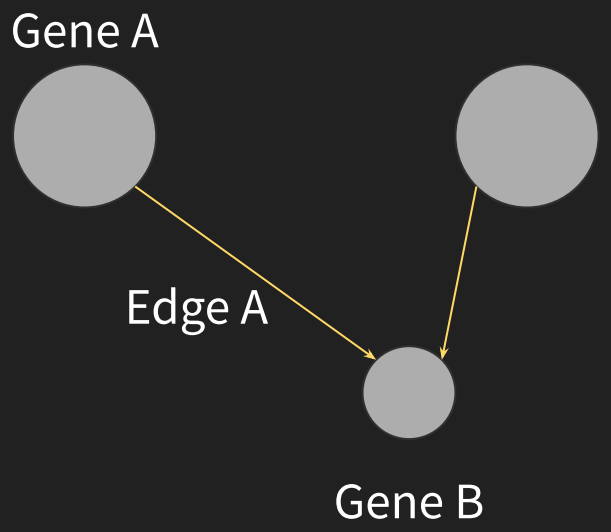


Lars Juhl Jensen. Network biology: A short introduction to the core concepts.

Networks are defined by their edges

Directed

Gene A - Gene B and Gene B - Gene A are not the same thing, how so? (eg. gene regulation)

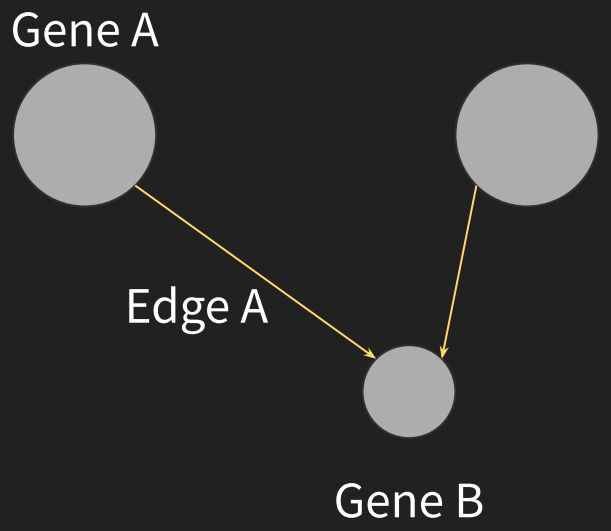


Lars Juhl Jensen. Network biology: A short introduction to the core concepts.

Edges can also be different too

Unweighted

All edges are equal

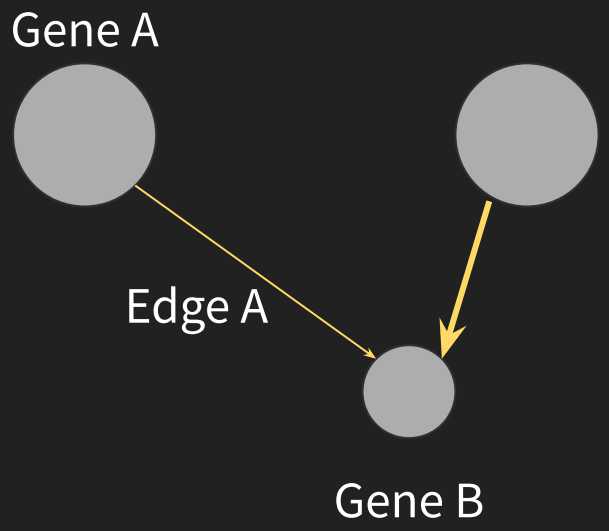


Lars Juhl Jensen. Network biology: A short introduction to the core concepts.

Edges can also be different too

Weighted

Some edges are stronger than the other edges



Lars Juhl Jensen. Network biology: A short introduction to the core concepts.

Pathway Enrichment Analysis

In general the procedure can be summarized into this 3 big steps:

1. Definition of a gene list of interest using omics data
2. Pathway enrichment analysis
3. Visualization and interpretation of pathway enrichment analysis results

Reimand J. 2019. Pathway enrichment analysis and visualization of omics data

Stage 1: definition of a gene list of interest using omics data

We mostly have done this stage

During this stage we perform the differential methylation analysis using R.

- Doing this step we then can discover our gene of interest based on their differentially methylated significance of adjusted p-value.

Reimand J. 2019. Pathway enrichment analysis and visualization of omics data

Stage 2: Pathway Enrichment Analysis

We also have done this stage, we're using one of the way, in this case GSEA

- GSEA is a threshold-free method that analyzes all genes on the basis of their differential expression rank, or other score
- GSEA is particularly suitable and is recommended when ranks are available for all or most of the genes in the genome
- GSEA searches for pathways whose genes are enriched at the top or bottom of the ranked gene list. For instance, if the topmost differentially expressed genes are involved in the cell cycle, this suggests that the cell cycle pathway is regulated in the experiment.

Reimand J. 2019. Pathway enrichment analysis and visualization of omics data

Stage 3: visualization and interpretation of pathway enrichment analysis results

This one we haven’t done it yet

- **Pathway information is inherently redundant**, as genes often participate in multiple pathways, and databases may organize pathways hierarchically by including general and specific pathways with many shared genes
- Consequently, pathway enrichment analysis often **highlights several versions of the same pathway**
- To address the redundancy problem we usually use the following tools EnrichmentMap and ClueGO

Reimand J. 2019. Pathway enrichment analysis and visualization of omics data

Stage 3: visualization and interpretation of pathway enrichment analysis results

An enrichment map helps identify interesting pathways and themes.

1. Expected themes **should be identified to help validate the pathway enrichment analysis results** . For instance, growth-related pathways and other hallmarks of cancer are expected to be identified in analyses of cancer genomics datasets
2. Pathways **not previously associated with the experimental context are evaluated more carefully** as potential discoveries. Pathways and themes with the strongest ESs should be studied first, followed by progressively weaker signals
3. **Interesting pathways are examined in more detail**, examining genes within the pathways (e.g., expression heat maps and the GSEA leading edge genes)

Reimand J. 2019. Pathway enrichment analysis and visualization of omics data