

Programming for Biomedical Informatics

Lecture 14 “Network Analysis in Practice”

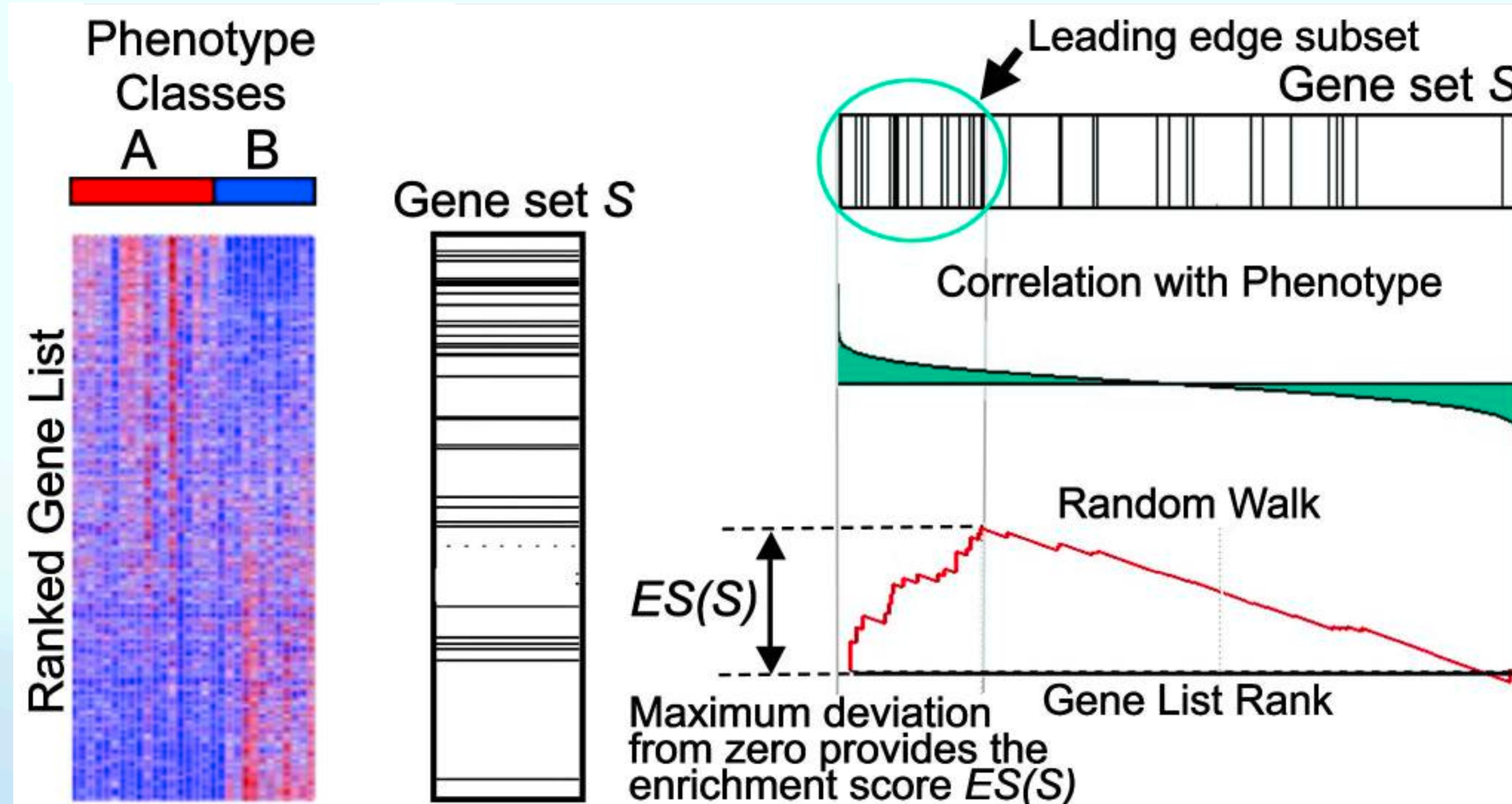
<https://github.com/tisimpson/pbi>

Ian Simpson

ian.simpson@ed.ac.uk

Coding

Gene Set Enrichment Analysis (GSEA)



Gene Set Enrichment Analysis (GSEA)

Input data

- Expression data set D with N genes and k samples
- Ranking procedure to produce Gene List L against a phenotype or profile of interest C.
- Independently derived Gene Set S of NH genes (e.g., a pathway, a cytogenetic band, or a GO category).

Enrichment Score (ES)

- Rank order the N genes in D to form $L = \{g_1, \dots, g_N\}$ according to the correlation of their expression profiles with C.
- Evaluate the fraction of genes in S ("hits") weighted by their correlation and the fraction of genes not in S ("misses") present up to a given position i in L.
- The ES is the maximum deviation from zero of $P_{hit} - P_{miss}$.

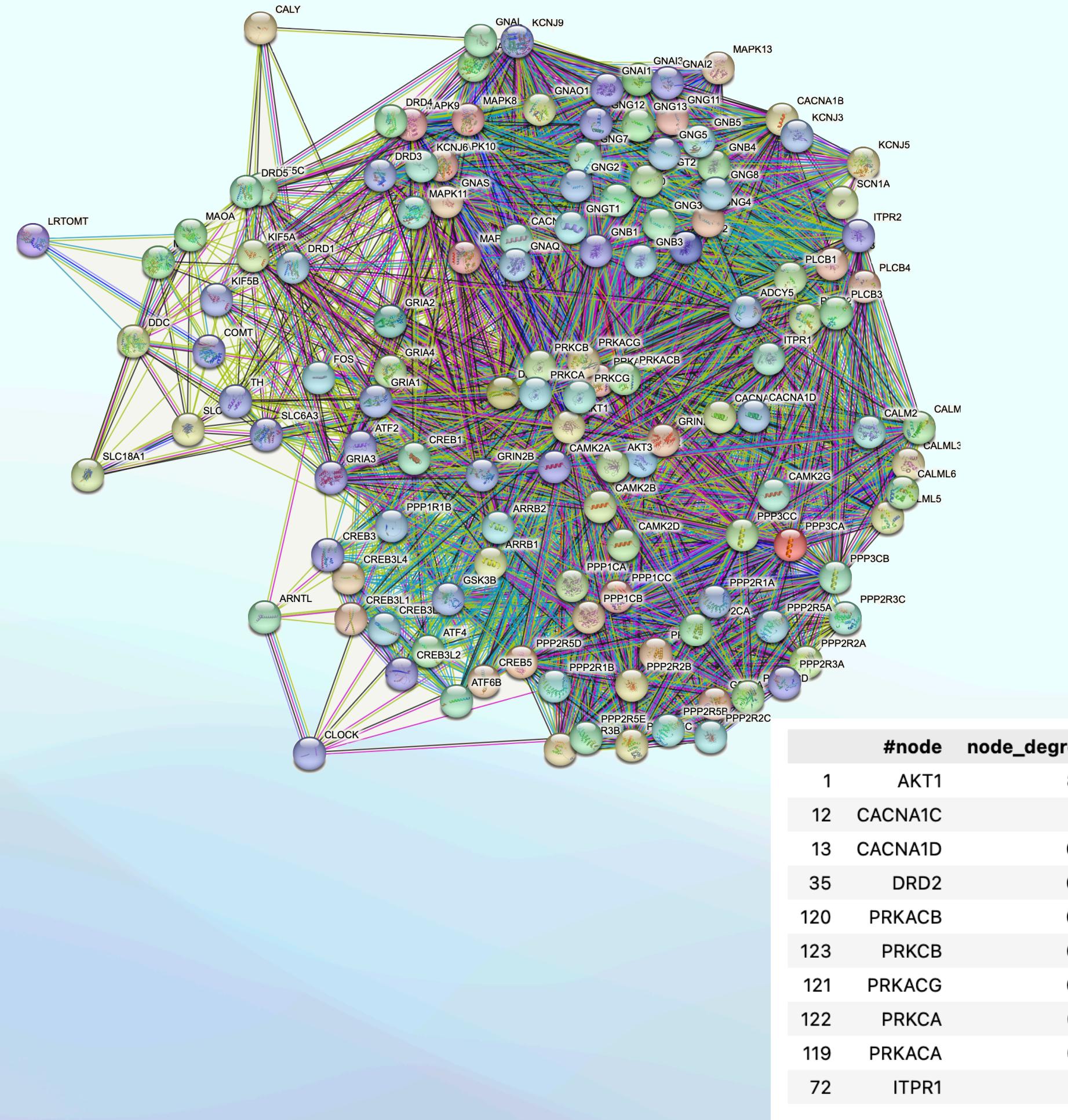
Estimating significance

- Randomly assign the original phenotype labels to samples, reorder genes, and re-compute $ES(S)$.
- Repeat for 1,000 permutations, and create a histogram of the corresponding enrichment scores ES_{Null} .
- Estimate nominal P value for S from ES_{Null} by using the positive or negative portion of the distribution corresponding to the sign of the observed $ES(S)$.

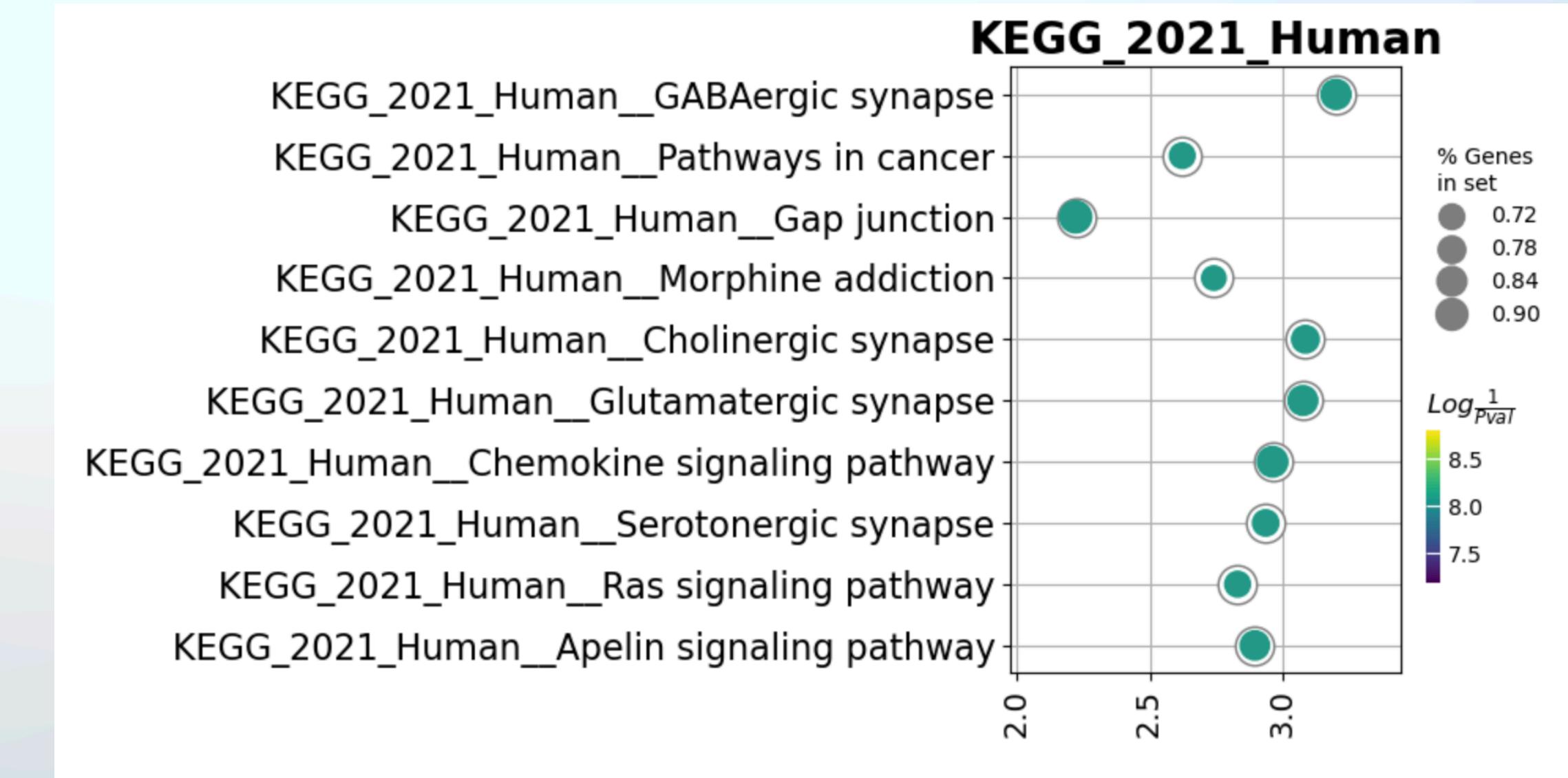
Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545-50

Gene Set Enrichment Analysis (GSEA Example)

Created a protein-protein interaction network from the Dopaminergic Synapse protein list and used node-degree to rank the genes - then we do a pre-rank GSEA analysis



Name	Term	ES	NES	NOM p-val	FDR q-val	FWER p-val	Tag %	Gene %	Lead_genes
0	prerank KEGG_2021_Human_GABAergic synapse	0.843099	3.197717	0.0	0.0	0.0	28/33	24.62%	CACNA1C;CACNA1D;PRKACB;PRKCB;PRKACG;PRKCA;PRKA...
1	prerank KEGG_2021_Human_Cholinergic synapse	0.77439	3.083102	0.0	0.0	0.0	45/58	40.00%	AKT1;CACNA1C;CACNA1D;PRKACB;PRKCB;PRKACG;PRKCA...
2	prerank KEGG_2021_Human_Glutamatergic synapse	0.77133	3.073702	0.0	0.0	0.0	42/50	40.00%	CACNA1C;CACNA1D;PRKACB;PRKCB;PRKACG;PRKCA;PRKA...
3	prerank KEGG_2021_Human_Chemokine signaling pathway	0.769659	2.96008	0.0	0.0	0.0	32/37	35.38%	AKT1;PRKACB;PRKCB;PRKACG;PRKACA;GNB1;GNB3;GNG3...
4	prerank KEGG_2021_Human_Serotonergic synapse	0.741352	2.9343	0.0	0.0	0.0	37/50	35.38%	CACNA1C;CACNA1D;PRKACB;PRKCB;PRKACG;PRKCA;PRKA...



GSEAPY - Python Implementation of GSEA

The screenshot shows the GSEAPY documentation page on ReadTheDocs. The top navigation bar includes a logo for 'GSEApY' and a 'latest' tag. A search bar is present. The main content area features a header 'Welcome to GSEAPY's documentation!' and a sub-header 'GSEAPY: Gene Set Enrichment Analysis in Python.' Below this are several badges: 'pypi package 1.1.3', 'GSEApY passing', 'docs passing', 'license MIT', and 'python 3'. A link to 'Release notes : https://github.com/zqfang/GSEApY/releases' is provided. A 'Citation' section contains a box with the text: 'Zhuoqing Fang, Xinyuan Liu, Gary Peltz, GSEApY: a comprehensive package for performing gene set enrichment analysis in Python. Bioinformatics, 2022;, btac757, https://doi.org/10.1093/bioinformatics/btac757'. At the bottom, there is an 'Installation' section with the text 'Install gseapy package from bioconda or pypi.' and a download button labeled 'latest'.

Home GSEApY latest Search docs

TABLE OF CONTENTS

- 1. Welcome to GSEAPY's documentation!
- 2. GSEAPY Example
- 3. GSVA example
- 4. scRNA-seq Example
- 5. A Protocol to Prepare files for GSEApY
- 6. Developmental Guide
- 7. Frequently Asked Questions

>Welcome to GSEAPY's documentation!

GSEAPY: Gene Set Enrichment Analysis in Python.

pypi package 1.1.3 GSEApY passing docs passing license MIT python 3

Release notes : <https://github.com/zqfang/GSEApY/releases>

Citation

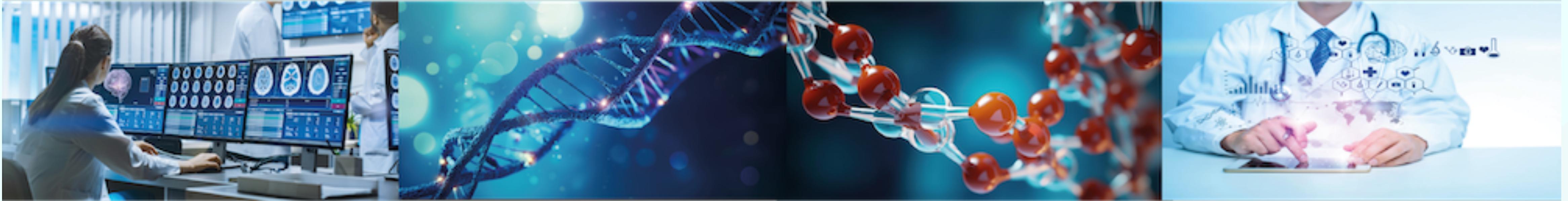
Zhuoqing Fang, Xinyuan Liu, Gary Peltz, GSEApY: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics*, 2022;, btac757, <https://doi.org/10.1093/bioinformatics/btac757>

Installation

Install gseapy package from bioconda or pypi.

latest

<https://gseapy.readthedocs.io/en/latest/>



Programming for Biomedical Informatics

Next Lecture - Tuesday 12th November
“Structuring Biomedical Data with Ontologies”

Ask Questions on the EdStem Discussion Board

Background

<https://github.com/tisimpson/pbi>