



Canadian Bioinformatics Workshops

www.bioinformatics.ca

bioinformaticsdotca.github.io

Supported by



Creative Commons

This page is available in the following languages:

Afrikaans Български Català Dansk Deutsch Ελληνικά English English (CA) English (GB) English (US) Esperanto
Castellano Castellano (AR) Español (CL) Español (Ecusdor) Castellano (MX) Castellano (PE)
Euskara Suomeksi français français (CA) Galego മലയാളം hrvatski Magyar Italiano 日本語 한국어 Macedonian Melayu
Nederlands Norsk Sesotho sa Leboa polski Português română slovenščiški jezik српски (latinica) Sotho svenska
中文 草語 (台灣) isiZulu



Attribution-Share Alike 2.5 Canada

You are free:



to Share — to copy, distribute and transmit the work



to Remix — to adapt the work



Under the following conditions:



Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).



Share Alike. If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar licence to this one.

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- The author's moral rights are retained in this licence.

Disclaimer

Your fair dealing and other rights are in no way affected by the above.

This is a human-readable summary of the Legal Code (the full licence) available in the following languages:

English French

Learn how to distribute your work using this licence

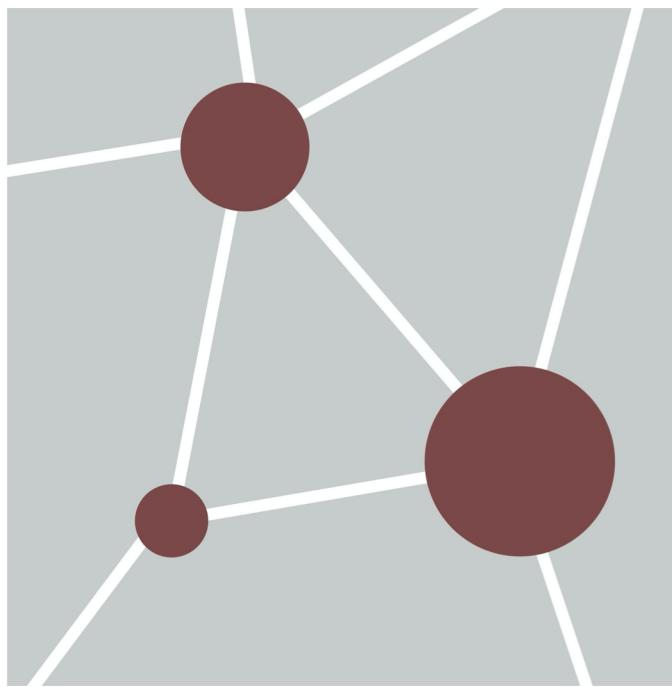
Final Slides



Veronique Voisin

Pathway and Network Analysis of -omics Data

June 5-7, 2023



Summary

Cytoscape apps that we used during the workshop...more apps in the app store

 ClueGO Creates and visualizes a functionally grouped network of	 stringApp Import and augment Cytoscape networks from STRING
 BiNGO Calculates overrepresented GO terms in the network and display	 CluePedia CluePedia: A ClueGO plugin for pathway insights using integrated
 MCODE Clusters a given network based on topology to find densely	 yFiles Layout Algorithms Highly sophisticated algorithms for arranging networks.
 enhancedGraphics A passthrough mapper for charts and gradients	 clusterMaker2 Multi-algorithm clustering app for Cytoscape
 cyREST Core App: Language-agnostic RESTful API	 EnrichmentMap Visualizes enrichments of pathways as an enrichment map,
 GeneMANIA Imports interaction networks from public databases from a list of	 cytoHubba Predicts and explores important nodes and subnetworks in a given
 ReactomeFIPPlugin Explore Reactome pathways and search for diseases related	 KEGGscape File reader and pathway visualizer for KEGG XML (KGML) files

The Cytoscape App Store

cytoscape app store

Submit an App ▾

Search the App Store

Sign In

Wall of Apps 184 total

network
generation



online data
import



graph
analysis



<http://apps.cytoscape.org>

Creating Networks

gene list

large
(100- 2000 genes)

medium
(100 genes)

small
(1-50 genes)

summarize by pathways

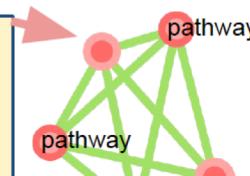
- represent as a network of pathways

- represent as a network of genes (gene products)

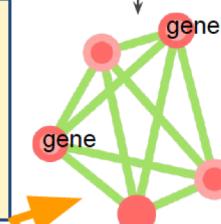
expand the list;
use function prediction

- represent as a network of gene (gene products) and add gene linkers

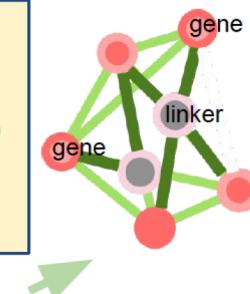
network



Enrichment Map

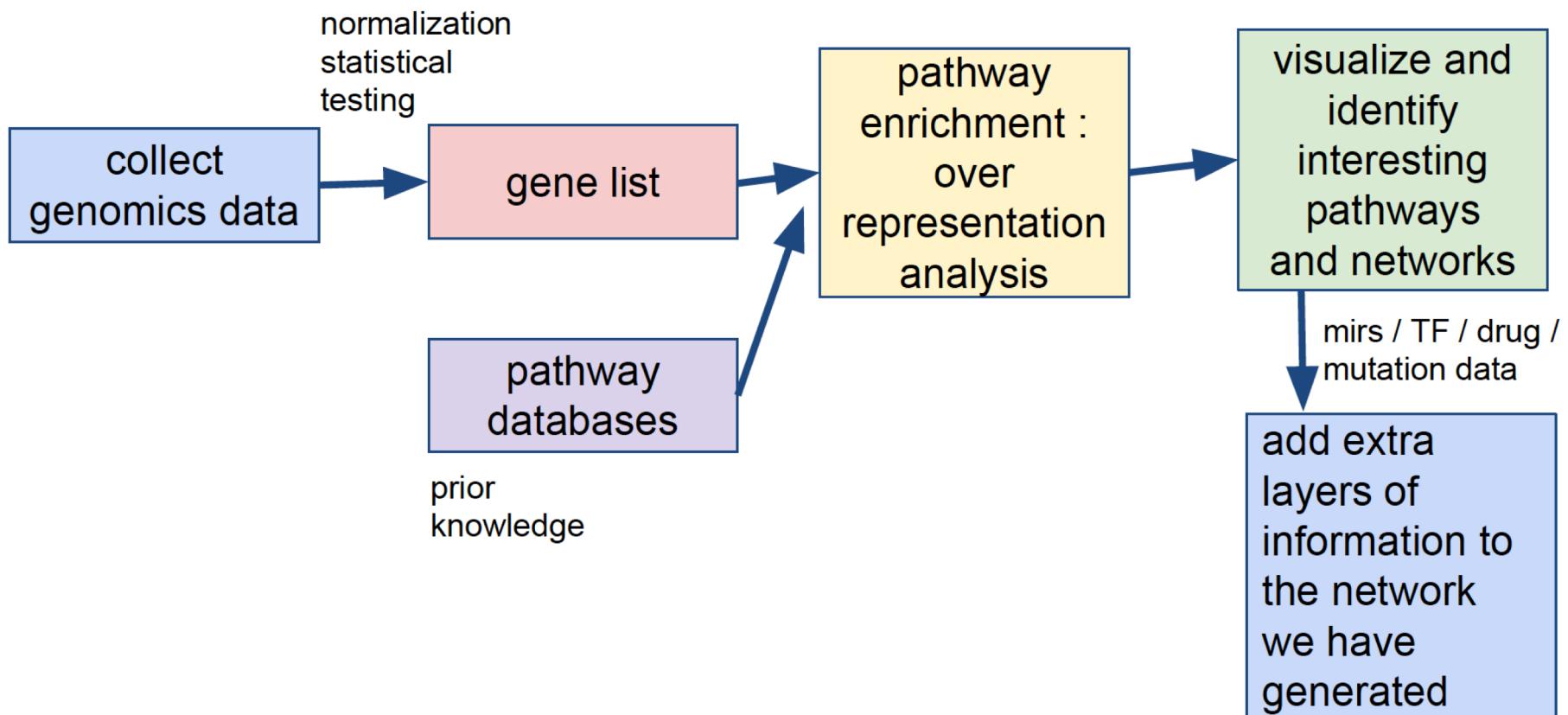


ReactomeFI

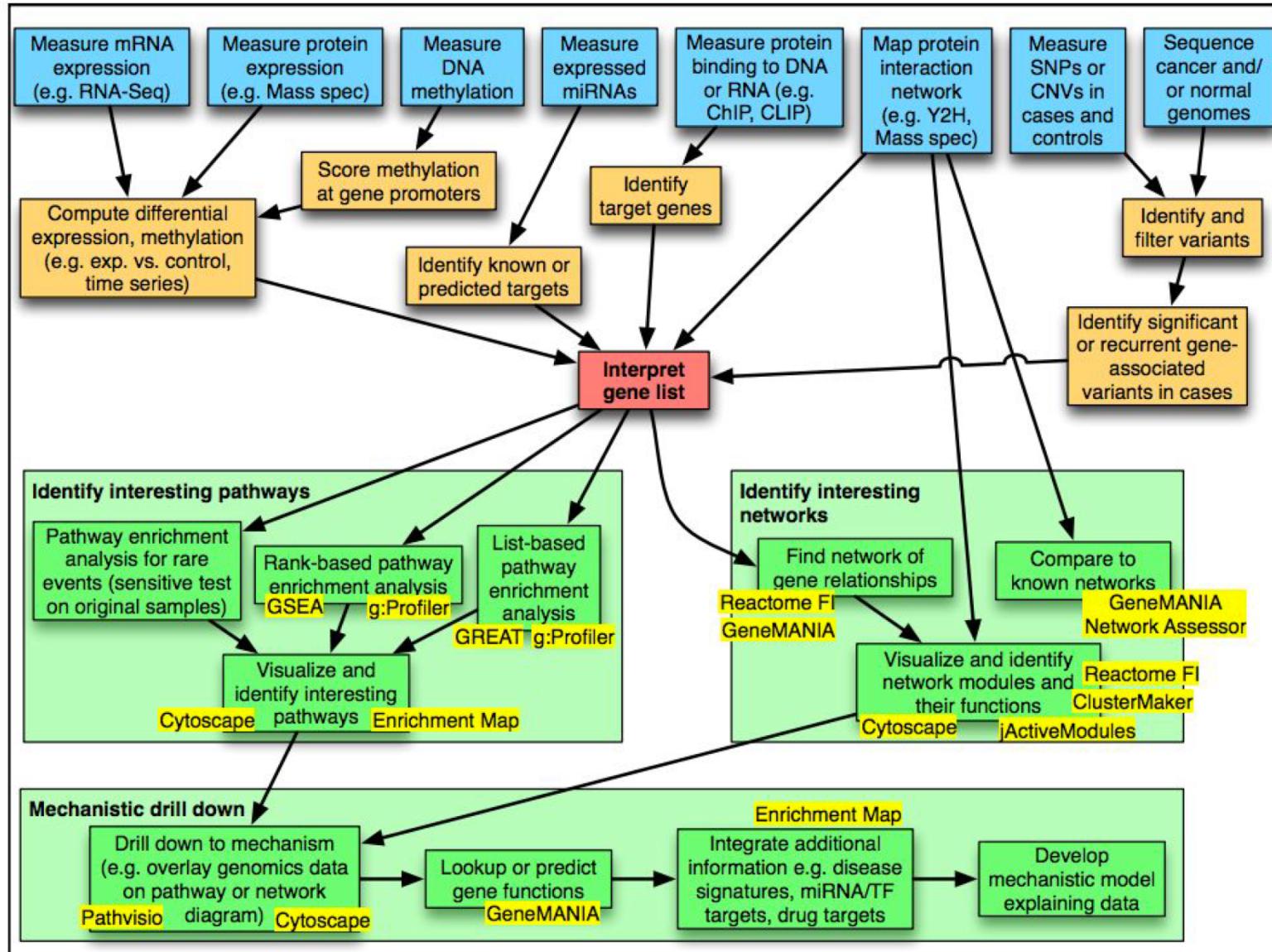


ReactomeFI with linkers
or
geneMANIA
stringApp

WORKFLOW SUMMARY



ROADMAP



Mirs, pathways and targets



miRPathDB v2.0

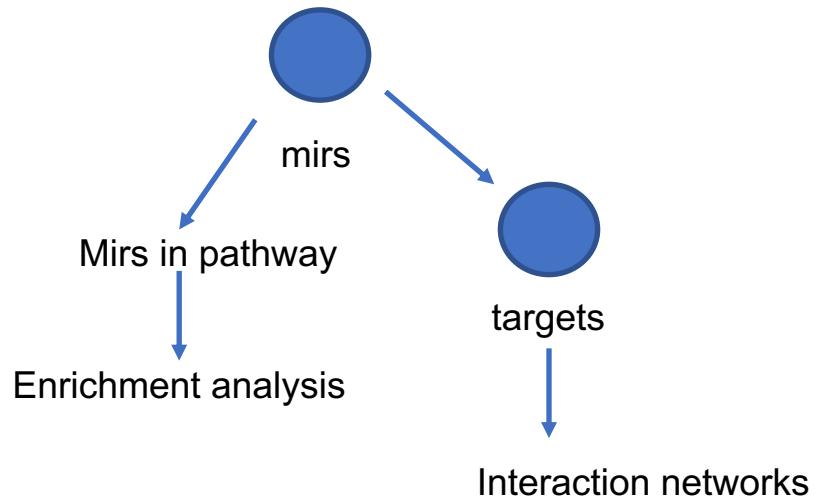
Home About Documentation Download



miRPathDB 2.0

Search

Enter a miRNA or pathway name:



miEAA: microRNA enrichment analysis and annotation

http://wwwccb.uni-saarland.de/mieaa_tool/

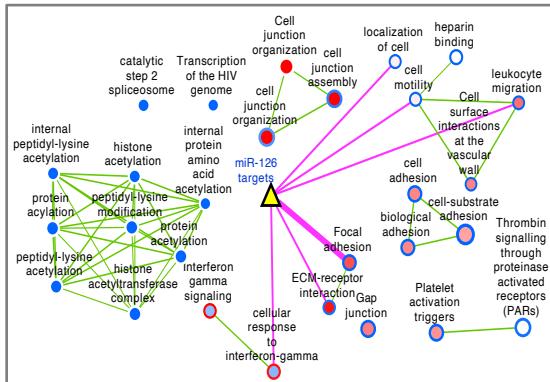
<http://www.lirmed.com/tam2/>

EnrichmentMap

Post analysis

Mir targets

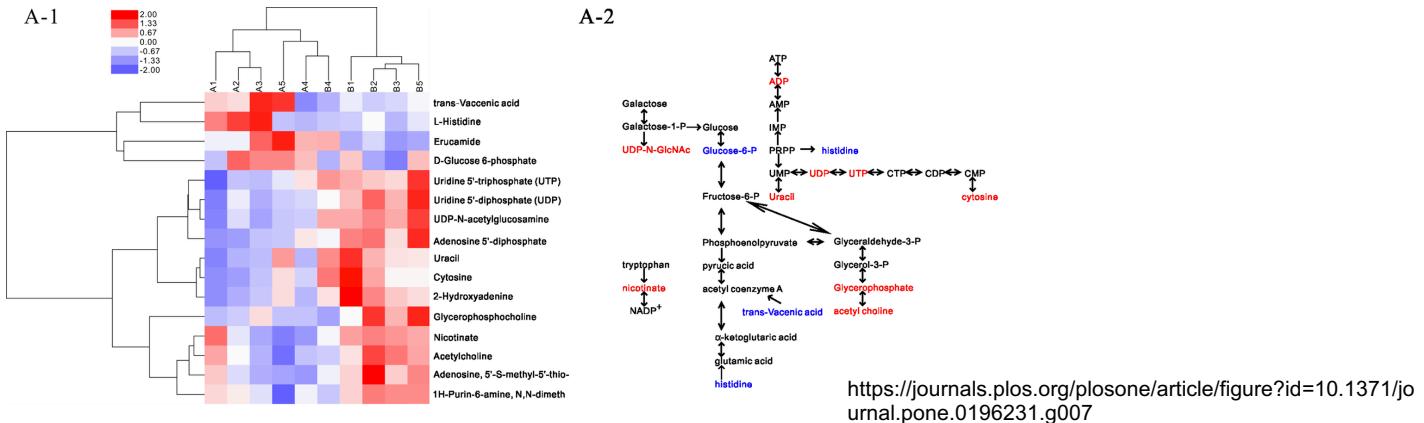
Post analysis:
Good for drug target
and transcription
factors



Result						
Enrichment analysis results						
Text file of results		Results Visualization				
Term	Count	Percent	Fold	P-value	Bonferroni	FDR
Category: Cluster (4 Items)						
hsa-mir-106b cluster [details]	1	0.33333	33.63889	0.0295	1	0.3755
hsa-mir-17 cluster [details]	2	0.33333	33.63889	1.32e-3	0.3569	0.08
hsa-mir-423 cluster [details]	1	0.5	50.45833	0.0197	1	0.3365
hsa-mir-6081 cluster [details]	1	0.2	20.18333	0.0487	1	0.479
Category: Disease (194 Items)						
Acute Cerebral Infarction [details]	1	0.16687	16.81944	0.0581	1	0.5292
Acute Ischemic Stroke [details]	2	0.14286	14.41667	7.67e-3	1	0.1858
Acute Myocardial Infarction [details]	2	0.04348	4.38768	0.0731	1	0.5944
Acute Pancreatitis [details]	1	0.14286	14.41667	0.0675	1	0.5676
Adenocarcinoma, Colon [details]	2	0.08696	8.77536	0.0203	1	0.2926
Adenocarcinoma, Esophageal [details]	1	0.04545	4.58712	0.1983	1	1
Adenocarcinoma, Gastric [details]	1	0.02632	2.6557	0.3191	1	1
Adenocarcinoma, Lung [details]	2	0.0198	1.99835	0.2642	1	1
Adrenal Cortex Neoplasms [details]	1	0.08333	8.40972	0.1131	1	0.7828

Metabolomics Data Analysis

A) list of metabolites from your experiment



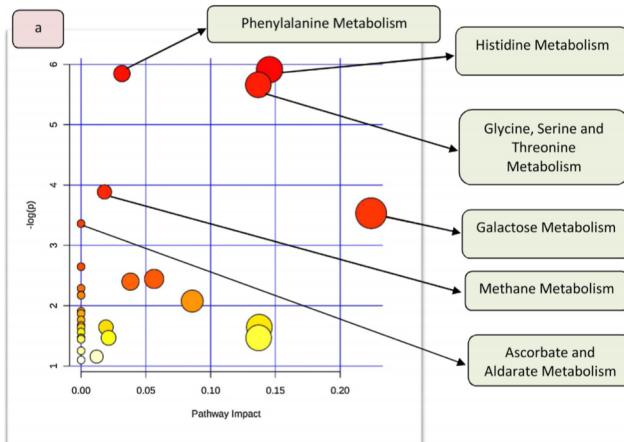
B) pathway: a set of metabolites known to be involved in metabolic pathways

The colored compounds indicate potential matches from the user's input, with red colors indicating significant hits and blue colors indicating non-significant hits.

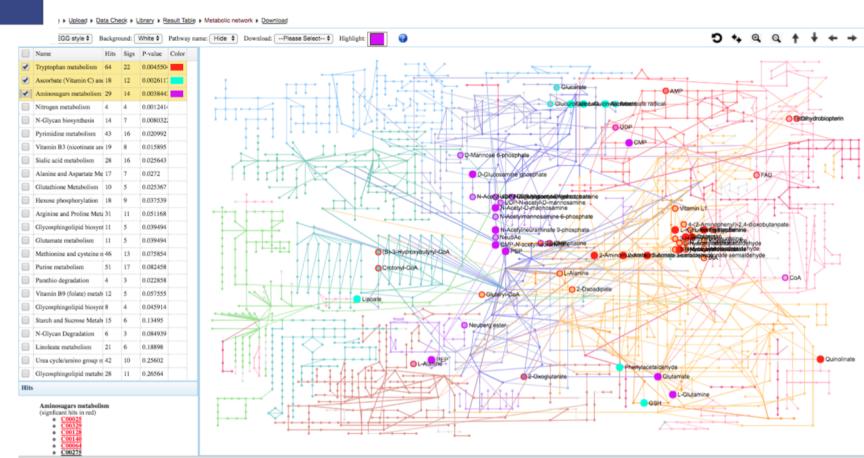
Pathway	Metabolites
Tryptophan metabolism	C00025; C00024; C00027; C00026; C00021; C00020; CE2119; C00028; C03722; C05647; C05645; CE1395; C05643; C05640; C00780; C05648; C00704; C01342; C00010; C00014; C00016; C00019; C15605; C00067; thbpt4acam; C05651; C03512; C05653; C02693; C05660; C00398; CE5982; C00643; C02220; C00078; C00978; C00877; C05642; CE5860; C01252; C05637; C02406; C00108; C00272; CE1918; C01652; C02470; C01144; C06212; C06213; CE2949; CE2948; CE3140; CE2122; C00479; CE2947; C01717; CE3092; C06205; CE5899; C03161; C00268; C00632; CE2152; CE2153; C02700; CE2095; C00328; CE3087; CE3086; C00322; C04409; C01352; C00051; C00058; C00030; C00331; C00332; C00936; C03824; C05636; C05635; C05634; C05639; C05638; C10164; C00637; C03227; C01598; C00525; C00527; C00041; C03024; C00954



Metabolomics Data Analysis

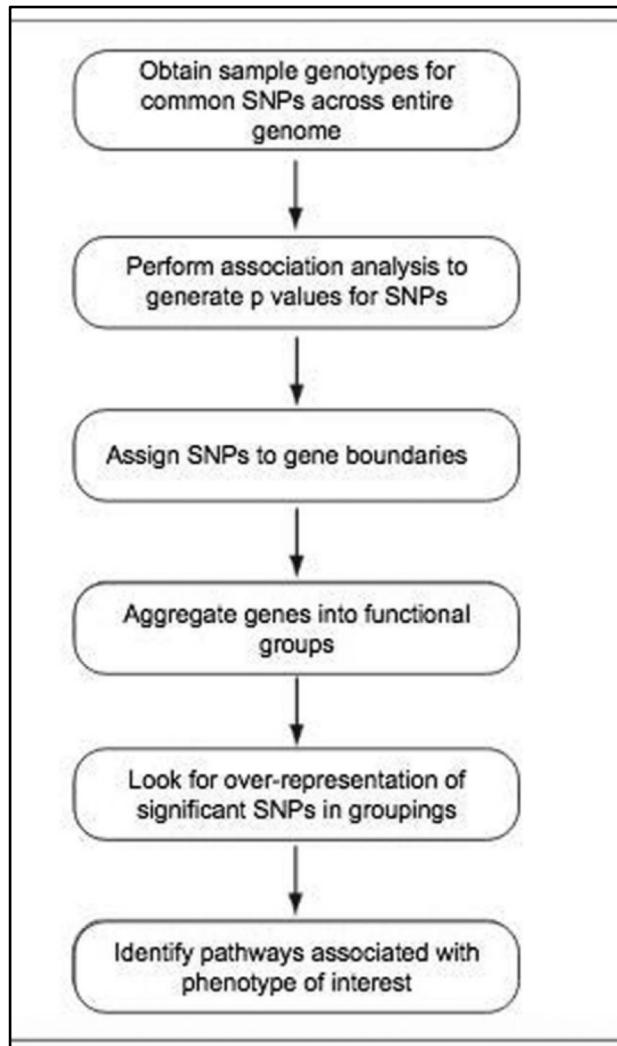


<https://sciez.com/applications/life-science-research/metabolomics/metabolomics-pathway-analysis>



List of metabolites: Fishers' exact test If it is possible to rank all the metabolites: GSEA

Overall protocol for pathway-based GWAS / WGS analysis



GWAS --> MAGENTA
(Meta-Analysis Geneset Enrichment of variaNT Associations)
<https://software.broadinstitute.org/mpg/magenta/>

The only **input** required is a table with variant association p-values and their chromosome positions taken from a genome-wide association study or meta-analysis. **Optional:** pathway/s or gene set/s of interest. Otherwise, we provide a set of pathways from public databases (see below).

The main **output** of MAGENTA is a nominal **gene set enrichment analysis (GSEA) p-value** and a **false discovery rate** for each gene set or pathway tested. There are several options, including running MAGENTA in the absence of a subset of genes, such as a predefined set of disease or trait genes. Additional information is provided, such as the expected and observed number of genes above the enrichment cutoff, and the number and name of genes in each tested gene set that lie near validated disease or trait SNPs if inputed by the user.

Ref: <https://pubmed.ncbi.nlm.nih.gov/30387919/>

ATACseq / CHIPseq

- GREAT (Stanford) is a recommended tool (from chromosomal position to gene + enrichment analysis)
- All tools compatible with EnrichmentMap

GREAT Overview News Use GREAT Demo Video How to Cite Help Forum

GREAT version 4.0.4 current (08/19/2019 to now) ▾

GREAT predicts functions of cis-regulatory regions.

Many coding genes are well annotated with their biological functions. Non-coding regions typically lack such annotation. GREAT assigns biological meaning to a set of non-coding genomic regions by analyzing the annotations of the nearby genes. Thus, it is particularly useful in studying cis functions of sets of non-coding genomic regions. Cis-regulatory regions can be identified via both experimental methods (e.g. ChIP-seq) and by computational methods (e.g. comparative genomics). For more see our [Nature Biotech Paper](#).

News

- Aug. 19, 2019: GREAT version 4 adds support for human hg38 assembly and updates ontology datasets for all supported assemblies.
- Sep. 8, 2018: GREAT has served over 1 million job submissions.
- Oct. 23, 2017: GREAT is moved to a VM to eliminate proxy errors.
- June 22, 2017: GREAT hardware upgrade to meet increasing submission volume.
- Nov. 16, 2015: The [GREAT user help forums](#) are frozen.
- Feb. 15, 2015: GREAT version 3 switches to Ensembl genes, adds support for zebrafish danRer7 and mouse mm10 assemblies, and adds new ontologies.
- Apr. 3, 2012: GREAT version 2 adds new annotations to human and mouse ontologies and visualization tools for data exploration.
- Feb. 18, 2012: The [GREAT user help forums](#) are opened.
- May 2, 2010: GREAT version 1 is launched, concurrent to [Nature Biotechnology publication \(reprint\)](#), Faculty of 1000 "Must Read"). How to Cite GREAT?

More news items...

Tips:

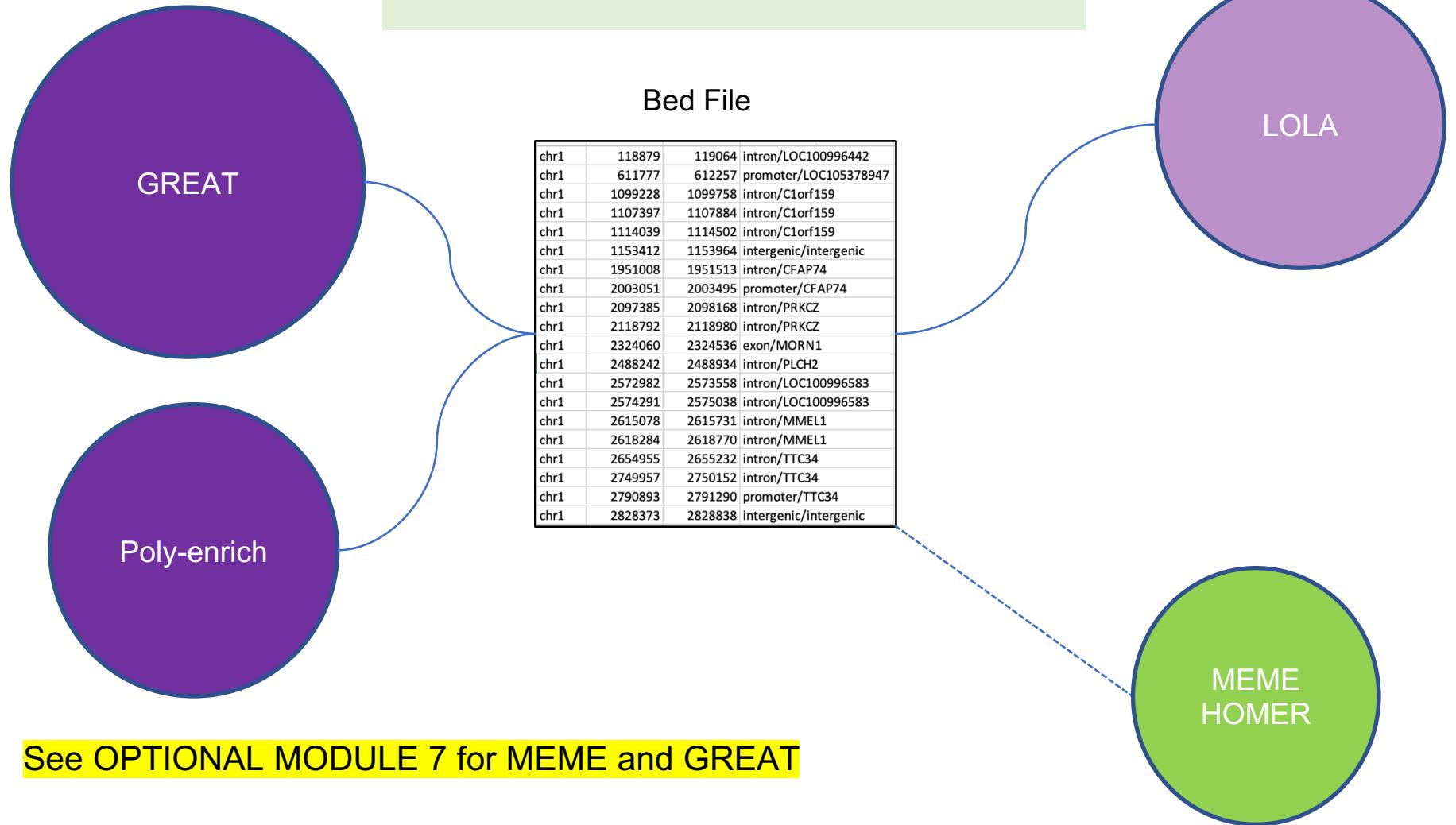
Do

- 1) Proximal analysis (+-2kb around TSS of genes)

And

- 2) Distal analysis (+-50kb around genes, filter genomic regions using tools like Segway or BEHST)

ATACseq / CHIPseq



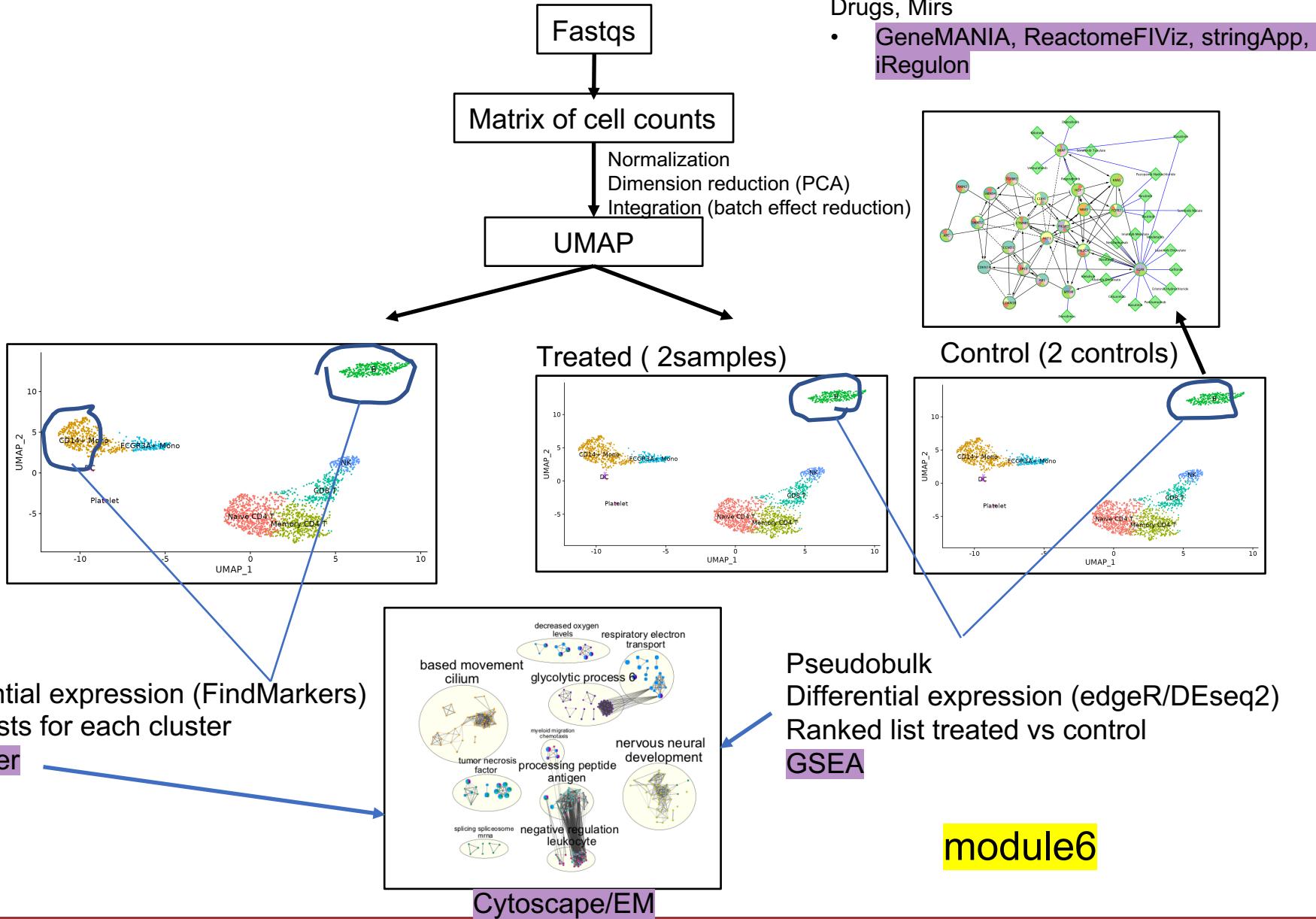
See OPTIONAL MODULE 7 for MEME and GREAT

Included in module 7: iRegulon

2 class design

- GSEA
- Enrichment Map

scRNA-seq



Non model organisms

1. Find a pathway database/ gmt file which is the closest to your organism
2. Convert your gene identifier to the gene identifiers used in the gmt file that you found using g:Convert and g:Orth
3. Both GSEA and g:Profiler accept custom gmt file.
4. GeneMANIA offers several organisms and the option to build your interaction networks.

The screenshot shows the g:GOST Functional profiling interface. At the top, there are four tabs: g:GOST (Functional profiling), g:Convert (Gene ID conversion), g:Orth (Orthology search), and g:SNPense (SNP id to gene name). Below the tabs, there are three input fields: 'Query' (selected), 'Upload query', and 'Upload bed file'. A note says 'Input is whitespace-separated list of genes'. To the right, there is an 'Options' section with a dropdown menu for 'Organism'. The dropdown shows 'Homo sapiens (Human)' selected. Below it is a 'Favourites' list with 14 items: Arabidopsis thaliana, Aspergillus fumigatus A1163, Bos taurus (Cow), Caenorhabditis elegans, Danio rerio (Zebrafish), Drosophila melanogaster, Gallus gallus (Chicken), Homo sapiens (Human) (selected), Mus musculus (Mouse), Oryza sativa Japonica Group, Pyrenopora teres f. teres 0-1 (Pyrenopora teres), Rattus norvegicus (Rat), and Saccharomyces cerevisiae.

Script to create a gmt file from the GO ontology:
<https://www.dropbox.com/s/wm3kq4lsdlfwcoq/creategmt.R?dl=0>

Module 6 (Review of the tools)

PART I: scRNA labs:

- Classic (gene lists with FindMarkers and GSEA from pseudobulk)
- scNetViz

Wednesday, options:

- Work with your own data
- Integrated Assignment
- Q and A with instructors
- Automation (R coding for advanced users)

Module 7 (optional)

- Chipseq and ATAC data: MEME and GREAT, EnrichmentMap
- iregulon: TF network

Workshop Sponsors:



Canadian Centre for
Computational
Genomics



GenomeCanada