

# Canadian Bioinformatics Workshops

[www.bioinformatics.ca](http://www.bioinformatics.ca)

[bioinformaticsdotca.github.io](https://bioinformaticsdotca.github.io)



## CC BY-SA 4.0 DEED

Attribution-ShareAlike 4.0 International

Canonical URL : <https://creativecommons.org/licenses/by-sa/4.0/>

[See the legal code](#)


### You are free to:


**Share** — copy and redistribute the material in any medium or format for any purpose, even commercially.

**Adapt** — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

### Under the following terms:

 **Attribution** — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

 **ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the [same license](#) as the original.

**No additional restrictions** — You may not apply legal terms or [technological measures](#) that legally restrict others from doing anything the license permits.

### Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable [exception or limitation](#).

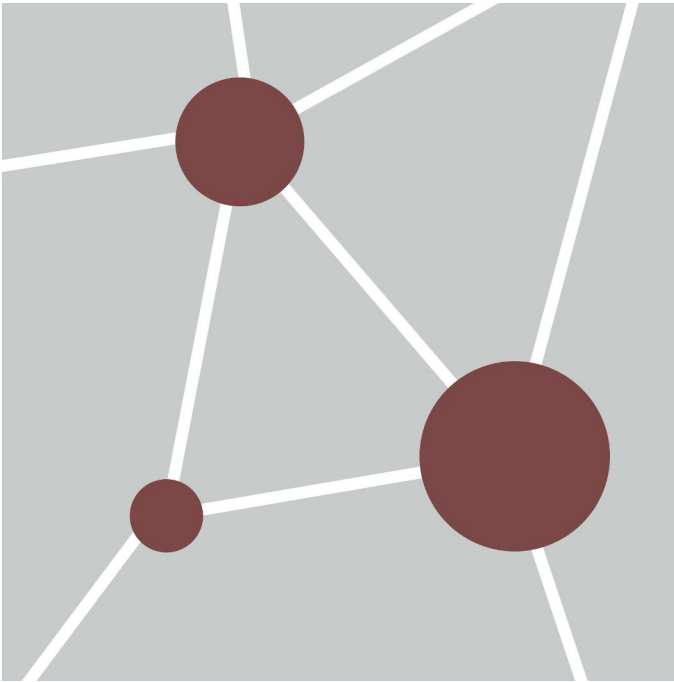
No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as [publicity, privacy, or moral rights](#) may limit how you use the material.

# Final Slides

Veronique Voisin

Pathway and Network Analysis of -omics Data

June 26-28, 2024



# Summary

# Cytoscape apps are organized by categories

The screenshot shows the Cytoscape App Store interface. At the top, there is a navigation bar with the Cytoscape logo, the text "cytoscape app store", and links for "Submit an App", "Search the App Store", and "Sign In". Below the navigation bar, the main heading reads "Wall of Apps 184 total" with the URL "http://apps.cytoscape.org" displayed in a large font. The interface is organized into three main categories, each with a grid of app icons:

- network generation**: This category includes icons for DiGeNET, GENEMANIA, LitSearch, BioGRID, and others.
- online data import**: This category includes icons for BioGRID, Drolic, LitSearch, and others.
- graph analysis**: This category includes icons for various network analysis tools.

All Apps

## Categories

data visualization

network generation

graph analysis

network analysis

online data import

automation

integrated analysis

clustering

utility

systems biology

enrichment analysis

visualization

data integration

core app

layout

annotation

pathway database

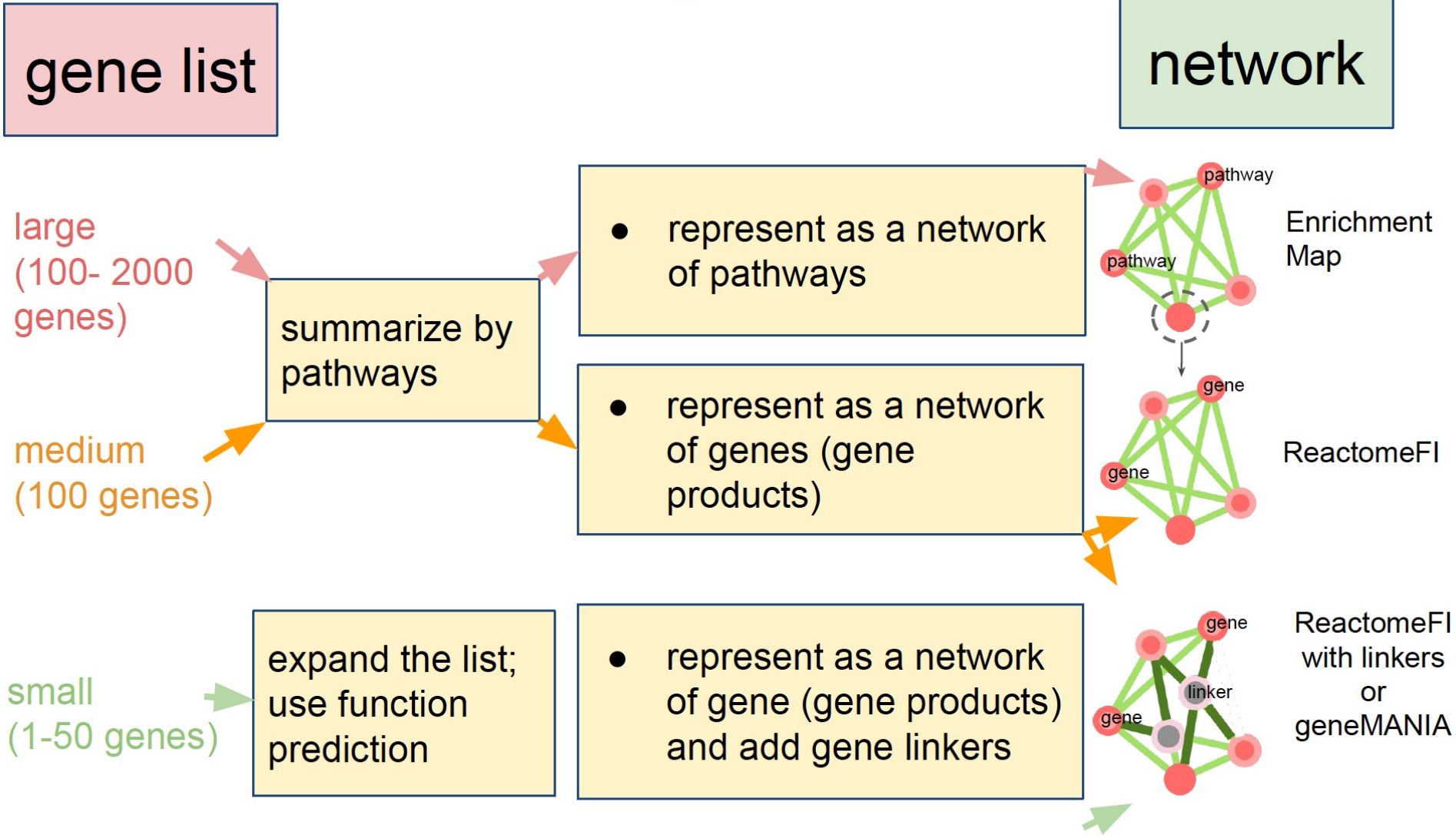
ontology analysis

import

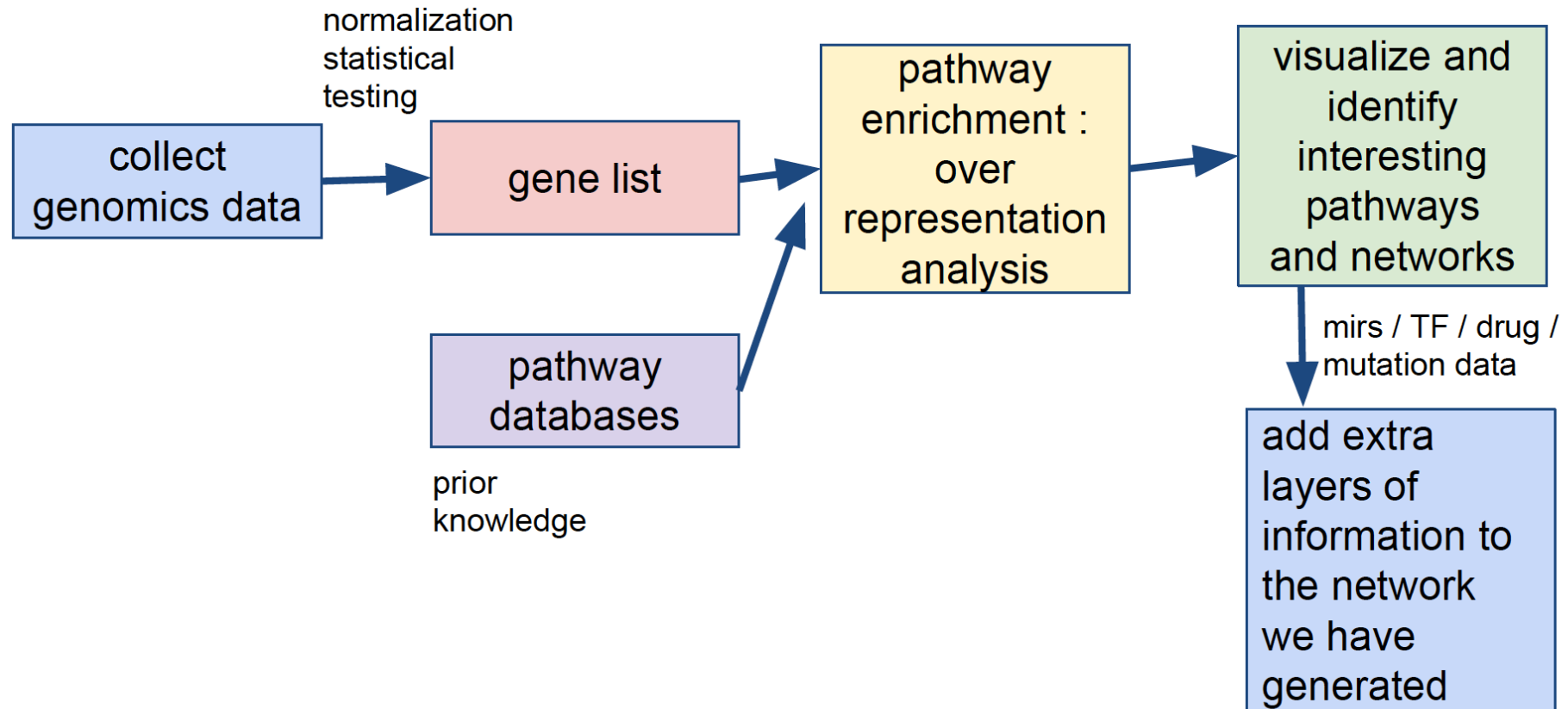
gene expression

more »

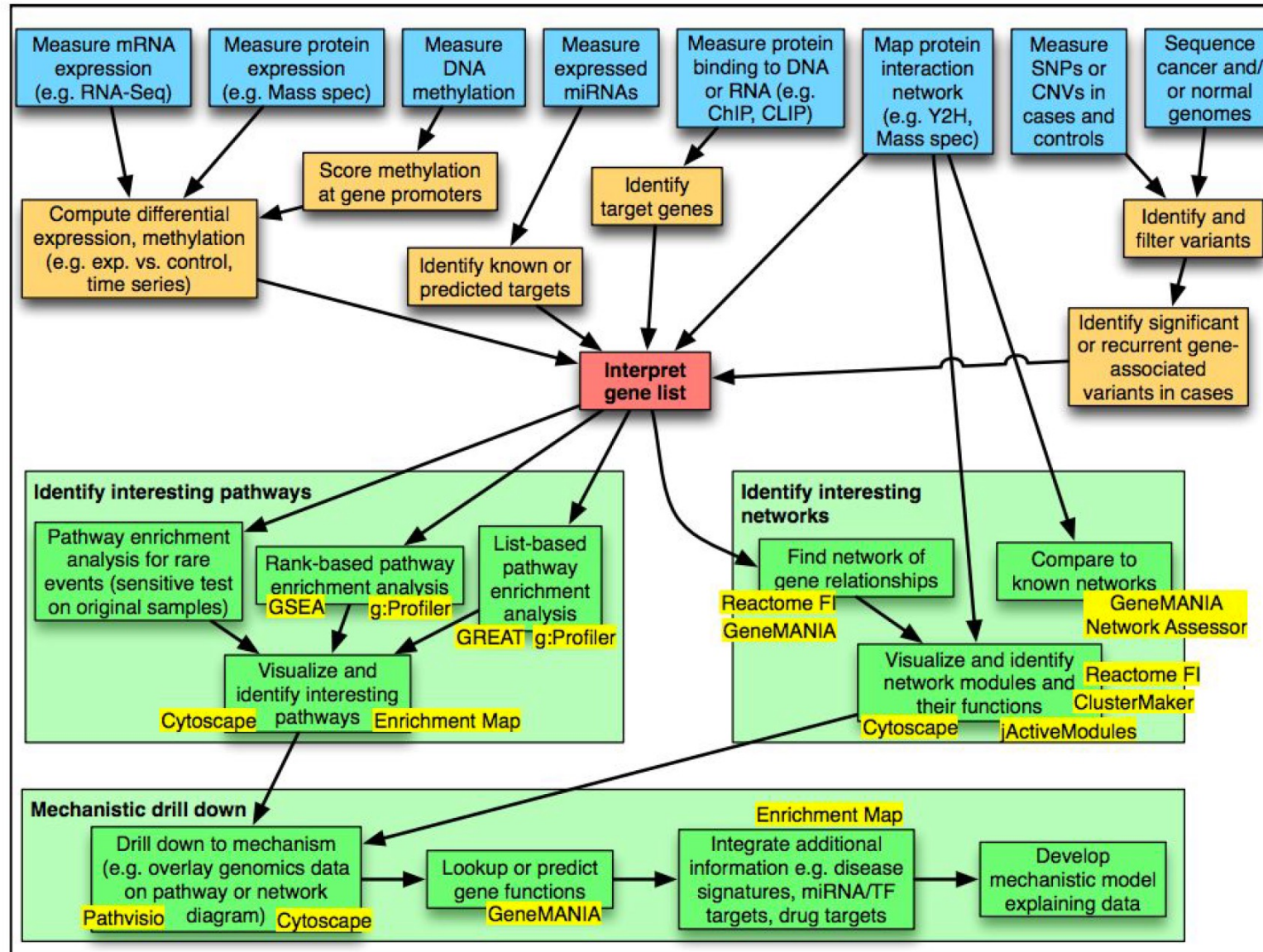
# Creating Networks



# WORKFLOW SUMMARY



# ROADMAP





# Mirs, pathways and targets



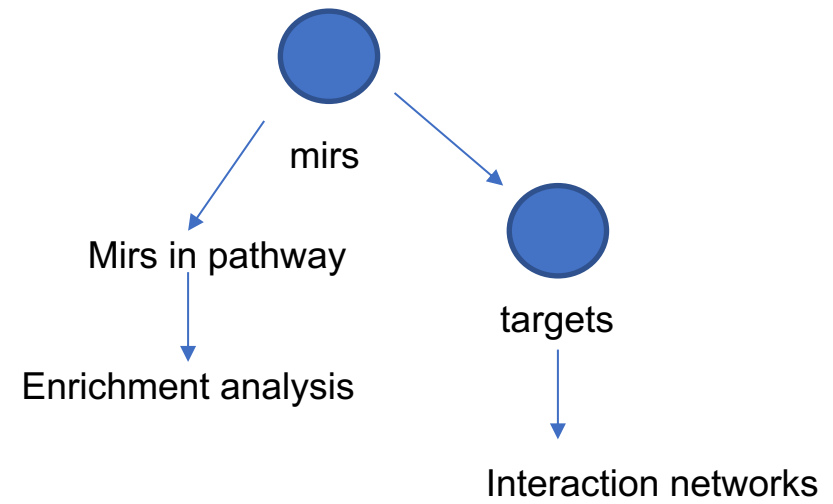
**Fc gamma R-mediated phagocytosis**

miRNAs that are significantly enriched for this pathway

In this table miRNAs are depicted that have significantly more targets in this pathway than expected by chance.

Show 10 entries

Database	miRNA	Evidence	Hits	Expected hits	P-value	Targets
miRBase	hsa-miR-126-3p	experimental (any)	7	0.281231	3.34e-8	AKT1,AKT2,CRK,CRKL,PKGCG,PKC
miRBase	hsa-miR-184	experimental (any)	5	0.297456	1.61e-4	AKT1,AKT2,INPPL1,PLPP3,PRKCB
miRCarta	m-5765	predicted (union)	56	36.0553	3.01e-4	AKT2,AMPH,ARPC2,ARPC3,ARPC4
miRCarta	m-17342	predicted (intersection)	12	2.26763	3.86e-4	ARPC2,CRKL,AT,MAPK1,NCF1,PPP
miRCarta	m-152	predicted (union)	57	38.8709	4.14e-4	AKT2,AMPH,ARPC1B,ARPC2,ARPC
miRCarta	m-12614	predicted (union)	55	36.7072	4.47e-4	AKT2,ARPC1B,ARPC2,ARPC4,ARPC
miRBase	hsa-miR-184	experimental (strong)	5	0.336283	4.48e-4	AKT1,AKT2,INPPL1,PLPP3,PRKCB
miRBase	hsa-miR-550a-3p	experimental (strong)	2	0.0353982	6.66e-4	MAPK1,MAPK3

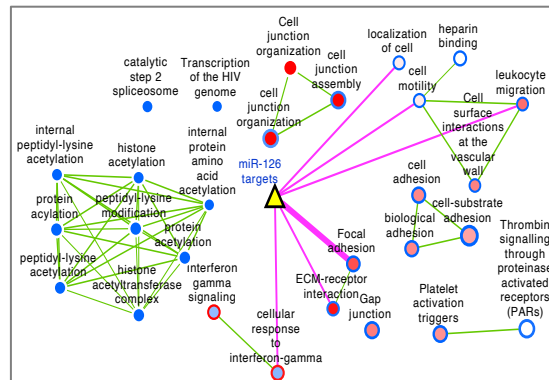


[http://www.ccb.uni-saarland.de/mieaa\\_tool/](http://www.ccb.uni-saarland.de/mieaa_tool/)

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

EnrichmentMap  
Post analysis  
Mir targets

Post analysis:  
Good for drug target  
and transcription  
factors



**Result**

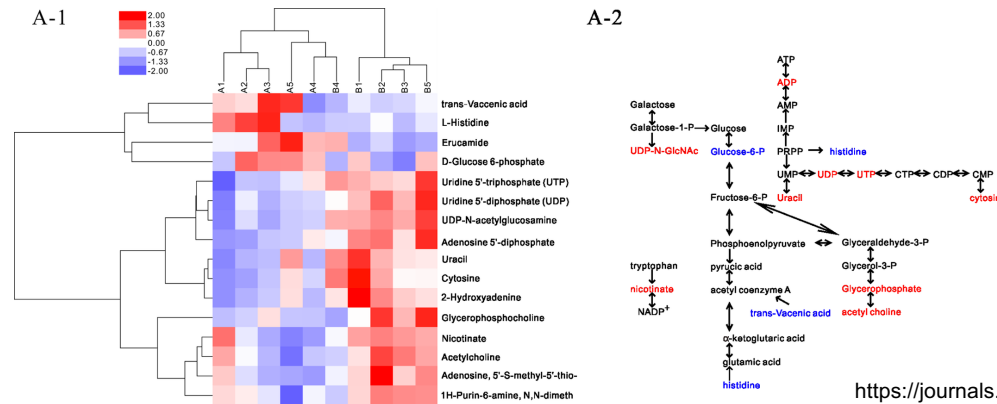
Enrichment analysis results

Text file of results Results Visualization

Term	Count	Percent	Fold	P-value	Bonferroni	FDR
<b>Category: Cluster (4 Items)</b>						
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
<b>Category: Disease (194 Items)</b>						
Acute Cerebral Infarction [details]	1	0.16667	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



<https://journals.plos.org/plosone/article/figure?id=10.1371/journal.pone.0196231.g007>

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

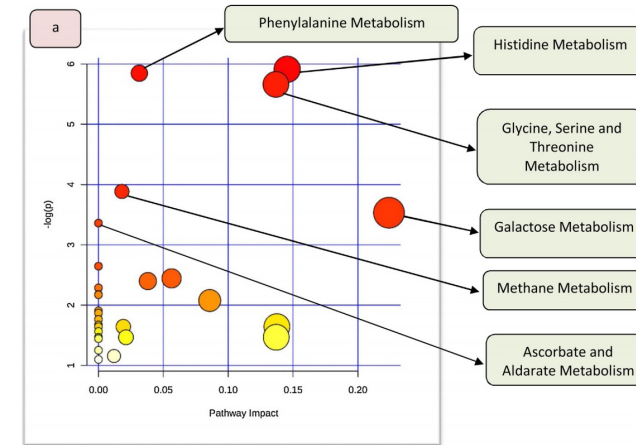
The colored compounds indicate potential matches from the user's input, with red colors indicating significant hits and blue colors indicating non-significant hits.

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; CE6205; 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

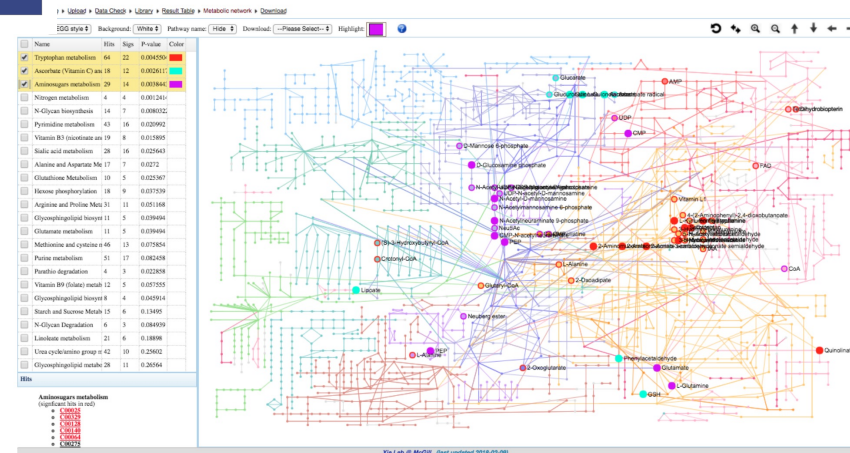
# Metabolomics Data Analysis



Input Data Type	Available Modules (click on a module to proceed, or scroll down for more details)					
Raw Spectra (mzML, mzXML or mzData)			LC-MS Spectra Processing			
MS Peaks (peak list or intensity table)			Functional Analysis	Functional Meta-analysis		
Annotated Features (compound list or table)		Enrichment Analysis	Pathway Analysis	Joint-Pathway Analysis	Network Analysis	
Generic Format (.csv or .txt table files)	Statistical Analysis [one factor]	Statistical Analysis [metadata table]	Biomarker Analysis	Statistical Meta-analysis	Power Analysis	Other Utilities

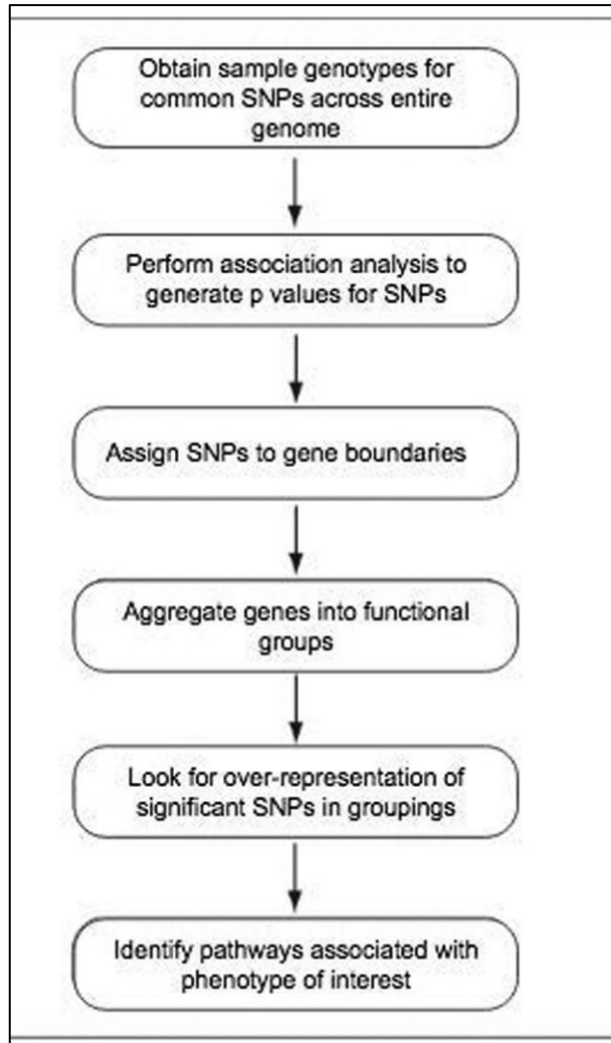


Pathway Name	Total	Hits (all)	Hits (sig.)	Fisher's Pvalue	EASE Score	Gamma Pvalue	Match Details
Tryptophan metabolism	94	64	21	0.0045504	0.0098086	0.0046682	<a href="#">View</a>
Ascorbate (Vitamin C) and Aldarate Metabolism	29	18	9	0.0026117	0.010691	0.0046835	<a href="#">View</a>
Aminosugars metabolism	69	29	12	0.0038443	0.011655	0.0047003	<a href="#">View</a>
Nitrogen metabolism	6	4	4	0.0012414	0.022604	0.0048951	<a href="#">View</a>
N-Glycan biosynthesis	48	14	7	0.0080322	0.032406	0.0050767	<a href="#">View</a>
Pyrimidine metabolism	70	43	14	0.020992	0.045203	0.0053243	<a href="#">View</a>
Vitamin B3 (nicotinate and nicotinamide) metabolism	28	19	8	0.015895	0.042225	0.0054047	<a href="#">View</a>
Sialic acid metabolism	107	28	10	0.025643	0.062907	0.0056878	<a href="#">View</a>
Alanine and Aspartate Metabolism	30	17	7	0.0272	0.08133	0.0060936	<a href="#">View</a>
Glutathione Metabolism	19	10	5	0.025367	0.099714	0.0065288	<a href="#">View</a>
Hexose phosphorylation	20	18	7	0.037539	0.10359	0.0066246	<a href="#">View</a>
Arginine and Proline Metabolism	45	31	10	0.051168	0.10984	0.0067823	<a href="#">View</a>
Glycosphingolipid biosynthesis - ganglioseries	62	11	5	0.039494	0.13447	0.0074426	<a href="#">View</a>
Glutamate metabolism	15	11	5	0.039494	0.13447	0.0074426	<a href="#">View</a>
Methionine and cysteine metabolism	94	46	13	0.075854	0.13715	0.0075184	<a href="#">View</a>
Purine metabolism	80	51	14	0.082458	0.1436	0.0077704	<a href="#">View</a>
Parathio degradation	6	4	3	0.022856	0.16222	0.0082678	<a href="#">View</a>
Vitamin B9 (folate) metabolism	33	12	5	0.067555	0.17336	0.0086255	<a href="#">View</a>
Glycosphingolipid biosynthesis - globoseries	16	8	4	0.045914	0.17647	0.0087285	<a href="#">View</a>
Starch and Sucrose Metabolism	33	15	5	0.13495	0.30626	0.014409	<a href="#">View</a>



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 inputted by the user.

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

# GWAS and pathway analysis

- **BridGE: a pathway-based analysis tool for detecting genetic interactions from GWAS.**  
Published: 21 March 2024. Nature Protocol

# ATACseq / CHIPseq

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

**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...](#)

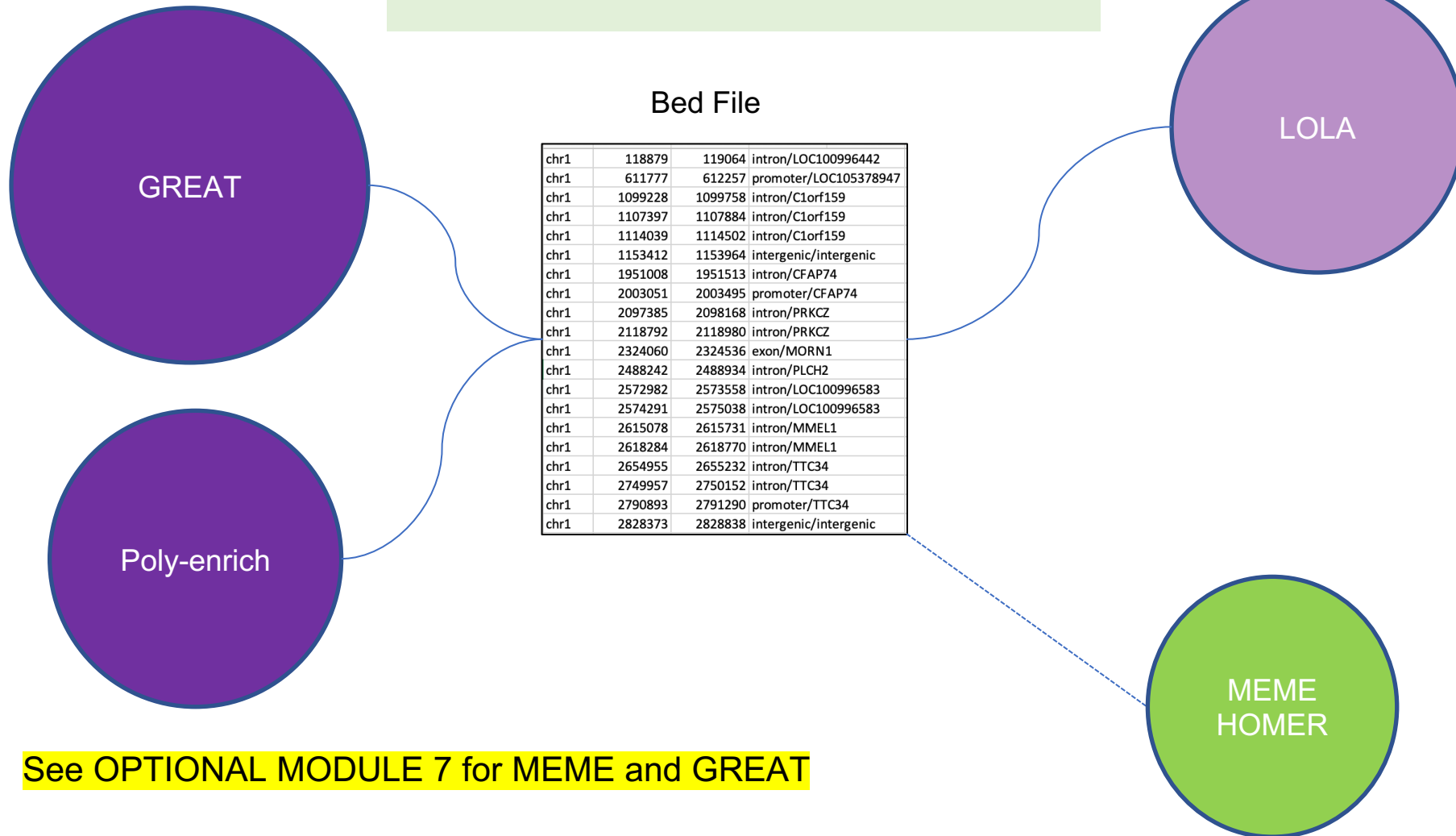
Tips:  
Do

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

# Transcriptomics: bulk RNA-seq 2 class design

- Module 2 and 3:
- GSEA
- Enrichment Map



# scRNA-seq

## Module 6

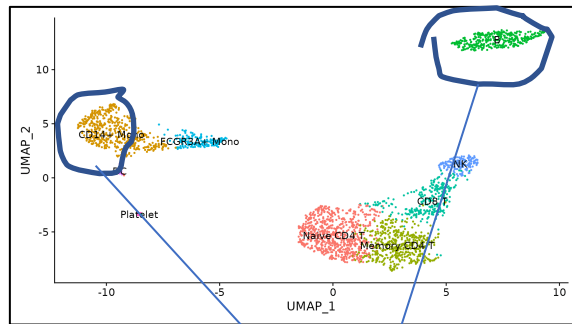
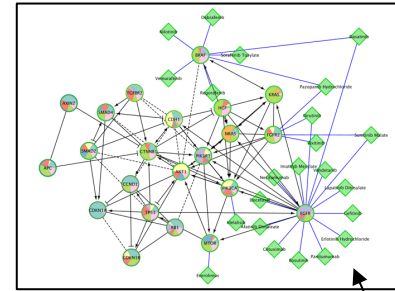
- Functional interaction network
- Possibility to add TF regulators, Drugs, Mirs
- GeneMANIA, ReactomeFIViz, stringApp, iRegulon

Fastqs

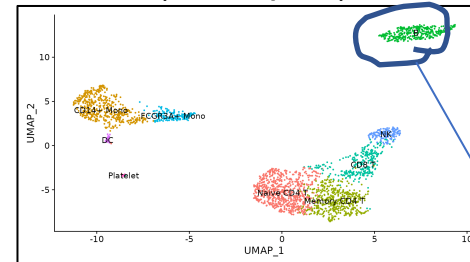
Matrix of cell counts

Normalization  
Dimension reduction (PCA)  
Integration (batch effect reduction)

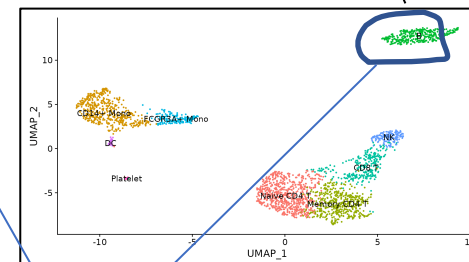
UMAP



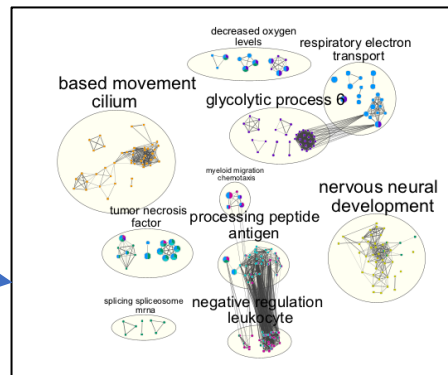
Treated ( 2samples)



Control ( 2 controls)



Pseudobulk  
Differential expression (edgeR/DEseq2)  
Ranked list treated vs control  
GSEA



Cytoscape/EM

Differential expression (FindMarkers)  
Gene lists for each cluster

g:Profiler

# 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 displays the g:GOST web interface. At the top, there are four navigation tabs: 'g:GOST Functional profiling' (highlighted in orange), 'g:Convert Gene ID conversion', 'g:Orth Orthology search', and 'g:SNPense SNP id to gene name'. Below the tabs, there are buttons for 'Query', 'Upload query', and 'Upload bed file'. A text input field is present with the placeholder text 'Input is whitespace-separated list of genes'. Below the input field, there are buttons for 'Run query', 'random example', and 'mixed query example'. On the right side, the 'Options' section is visible, showing a dropdown menu for 'Organism:'. The dropdown is open, displaying a list of organisms. 'Homo sapiens (Human)' is currently selected and highlighted in orange. Other organisms listed include Arabidopsis thaliana, Aspergillus fumigatus A1163, Bos taurus (Cow), Caenorhabditis elegans, Danio rerio (Zebrafish), Drosophila melanogaster, Gallus gallus (Chicken), Mus musculus (Mouse), Oryza sativa Japonica Group, Pyrenophora teres f. teres 0-1 (Pyrenophora teres), Rattus norvegicus (Rat), and Saccharomyces cerevisiae.

g:GOST performs functional enrichment analysis, also known as over-representation

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

## Module 7 (Review of the tools)

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

## Module 7 (optional)

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

# We are on a Coffee Break & Networking Session

Workshop Sponsors:



Canadian Centre for  
Computational  
Genomics



HPC4Health



GenomeCanada