bioinformatics.ca

Canadian Bioinformatics Workshops

www.bioinformatics.ca

bioinformaticsdotca.github.io

Supported by

McGill UNIVERSITY


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) Castellano (CO) Español (Ecuador) Castellano (MX) Castellano (PE)
Euskara Suomeksi français français (CA) Galego עברית hrvatski Magyar Italiano 日本語 한국어 Macedonian Melayu
Nederlands Norsk Sesotho sa Leboa polski Português română slovenski jezik српски srpski (latinica) Sotho svenska
中文 華語 (台灣) isiZulu
creative commons
Attribution-Share Alike 2.5 Canada
You are free:
to Share — to copy, distribute and transmit the work
to Remix — to adapt the work
Free Cultural
APPROVED FOR
Works
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

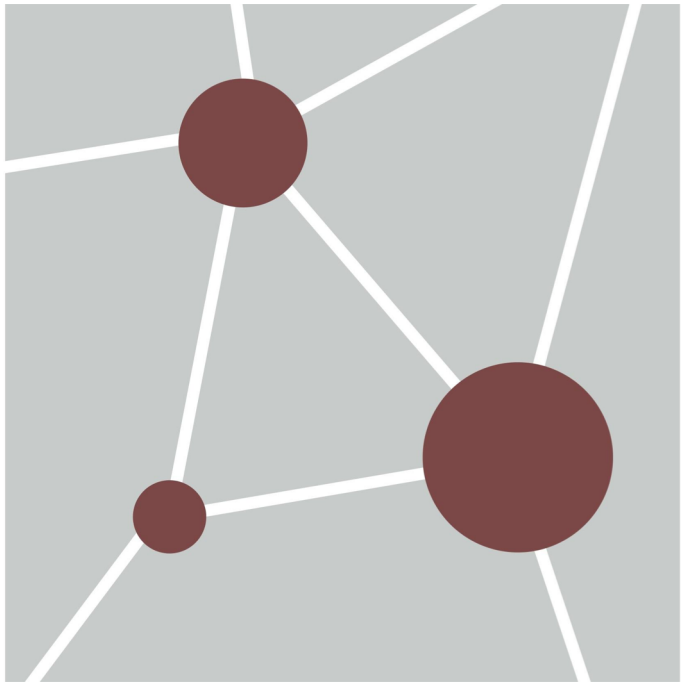
Final Slides

Veronique Voisin

Pathway and Network Analysis of –omics Data

July 27-29, 2020

bioinformatics.ca



Instructor affiliation logos

Creating Networks

gene list

network

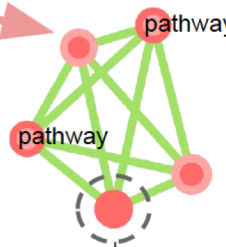
large (100- 2000 genes)

medium (100 genes)

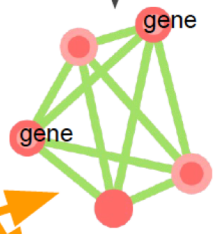
summarize by pathways

- represent as a network of pathways

- represent as a network of genes (gene products)



Enrichment Map

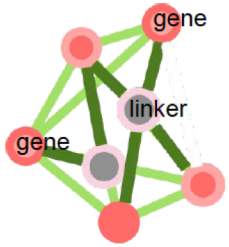


ReactomeFI

small (1-50 genes)

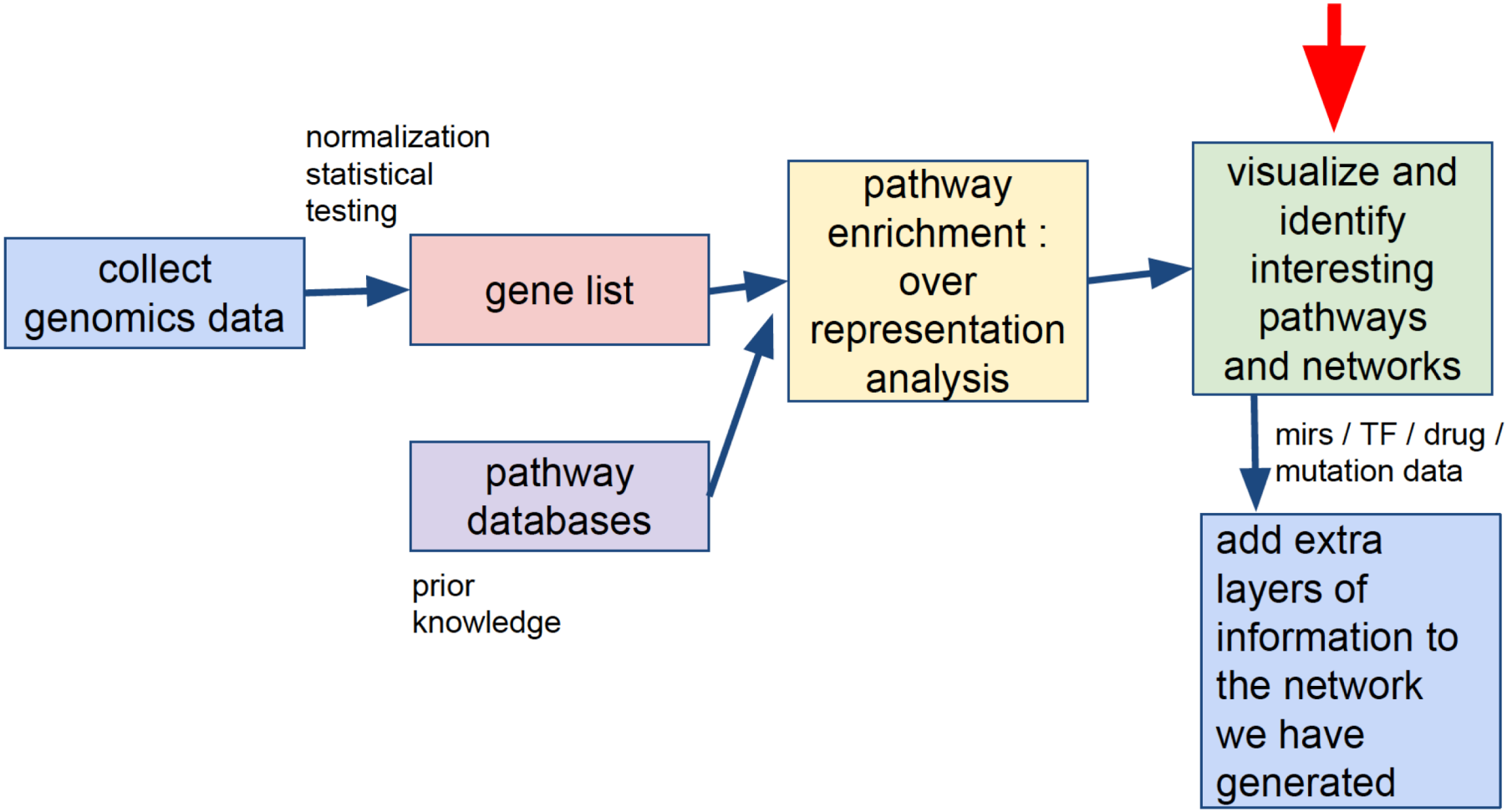
expand the list; use function prediction

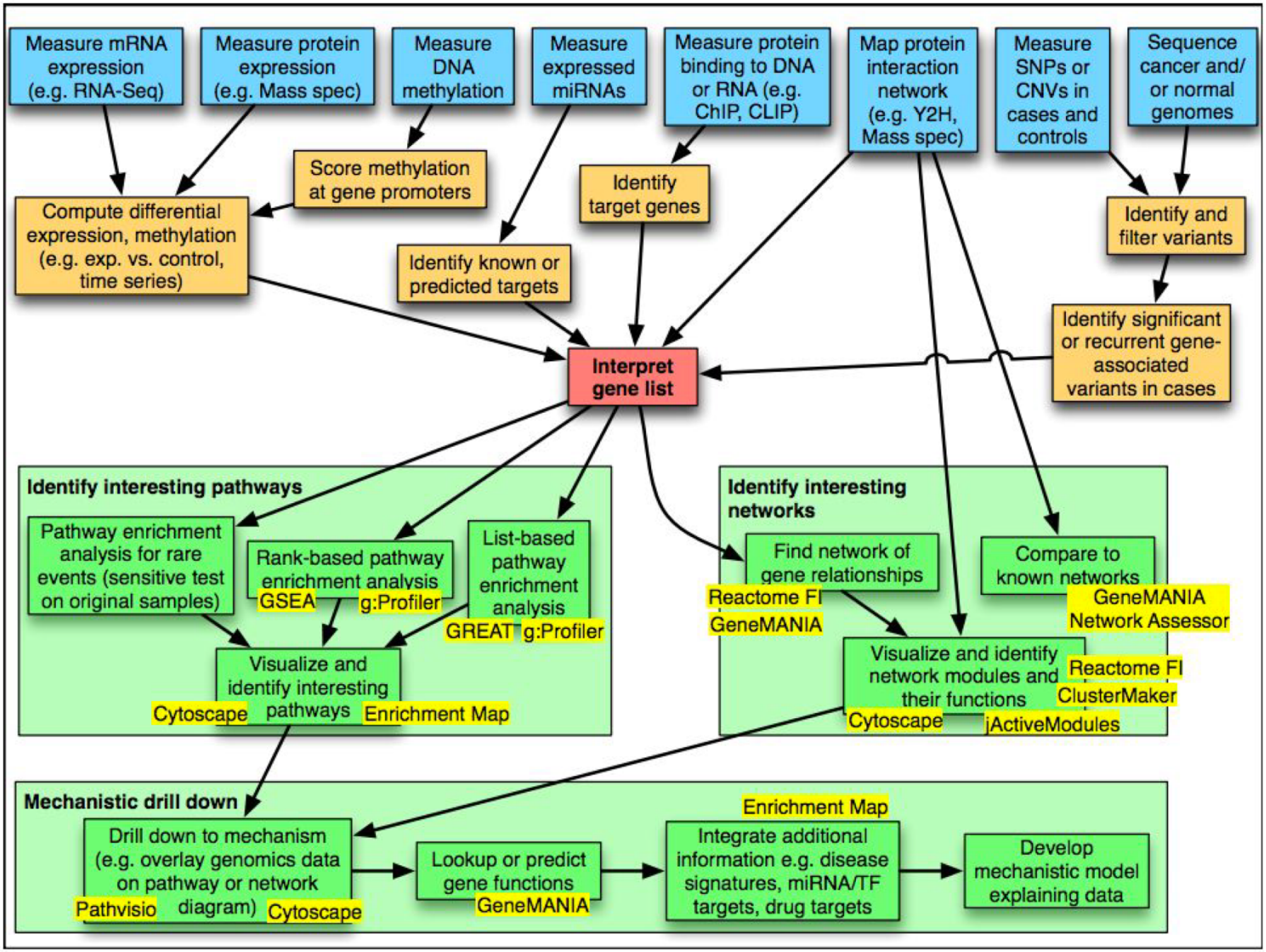
- represent as a network of gene (gene products) and add gene linkers



ReactomeFI with linkers or geneMANIA

Where are we in the workflow?




Measure mRNA expression (e.g. RNA-Seq)
Measure protein expression (e.g. Mass spec)
Measure DNA methylation
Measure expressed miRNAs
Measure protein binding to DNA or RNA (e.g. ChIP, CLIP)
Map protein interaction network (e.g. Y2H, Mass spec)
Measure SNPs or CNVs in cases and controls
Sequence cancer and/ or normal genomes
Score methylation at gene promoters
Identify target genes
Compute differential expression, methylation (e.g. exp. vs. control, time series)
Identify known or predicted targets
Identify and filter variants
Identify significant or recurrent gene-associated variants in cases
Interpret gene list
Identify interesting pathways
Pathway enrichment analysis for rare events (sensitive test on original samples)
Rank-based pathway enrichment analysis
GSEA
g:Profiler
List-based pathway enrichment analysis
GREAT
g:Profiler
Visualize and identify interesting pathways
Cytoscape
Enrichment Map
Identify interesting networks
Find network of gene relationships
Reactome FI
GeneMANIA
Compare to known networks
GeneMANIA
Network Assessor
Visualize and identify network modules and their functions
Reactome FI
ClusterMaker
Cytoscape
jActiveModules
Mechanistic drill down
Drill down to mechanism (e.g. overlay genomics data on pathway or network diagram)
Pathvisio
Cytoscape
Lookup or predict gene functions
GeneMANIA
Enrichment Map
Integrate additional information e.g. disease signatures, miRNA/TF targets, drug targets
Develop mechanistic model explaining data

Create custom gmt file from GO (R script)

```
################## hsapiens
library(biomaRt)


### get annotations
mart=useMart(biomart="ensembl",dataset="hsapiens_gene_ensembl")


go_annotation <-
getBM(attributes=c("hgnc_symbol","ensembl_gene_id","ensembl_transcript_id","go_id","name_1006","namespace_1003","go_linkage_type"),filters=list(biotype='protein_coding'),mart=mart);


go_annotation_bp <- go_annotation[which(go_annotation$namespace_1003=="biological_process"),]
head(go_annotation_bp)


##create gmt
go_pathway_sets <- aggregate(go_annotation_bp[1],by=list(go_annotation_bp$go_id),FUN=function(x){list(unique(x))})


m = match( go_pathway_sets[,1], go_annotation_bp$go_id)
go_pathway_names <- go_annotation_bp$name_1006[m]

### write the gmt

fname = "gobp.gmt"
object = go_pathway_sets[,2]
for ( e in 1: length(object) ){


write.table( t(c(go_pathway_sets[e,1], go_pathway_names[e],object[[e]])),sep="\t",quote=FALSE,file=fname,append=TRUE,col.names=FALSE,row.names=FALSE)
}
############### horse
library(biomaRt)

### get annotations
mart=useMart(biomart="ensembl",dataset="ecaballus_gene_ensembl")


go_annotation <-
getBM(attributes=c("uniprot_gn","ensembl_gene_id","ensembl_transcript_id","go_id","name_1006","namespace_1003","go_linkage_type"),filters=list(biotype='protein_coding'),mart=mart);


go_annotation_bp <- go_annotation[which(go_annotation$namespace_1003=="biological_process"),]
head(go_annotation_bp)


##create gmt
go_pathway_sets <- aggregate(go_annotation_bp[1],by=list(go_annotation_bp$go_id),FUN=function(x){list(unique(x))})


m = match( go_pathway_sets[,1], go_annotation_bp$go_id)
go_pathway_names <- go_annotation_bp$name_1006[m]


### write the gmt
fname = "gobp_horse.gmt"
object = go_pathway_sets[,2]
for ( e in 1: length(object) ){
write.table( t(c(go_pathway_sets[e,1], go_pathway_names[e],object[[e]])),sep="\t",quote=FALSE,file=fname,append=TRUE,col.names=FALSE,row.names=FALSE)
}
```

https://www.dropbox.com/s/wm3kq4lsdlfwcoq/creategmt.R?dl=0

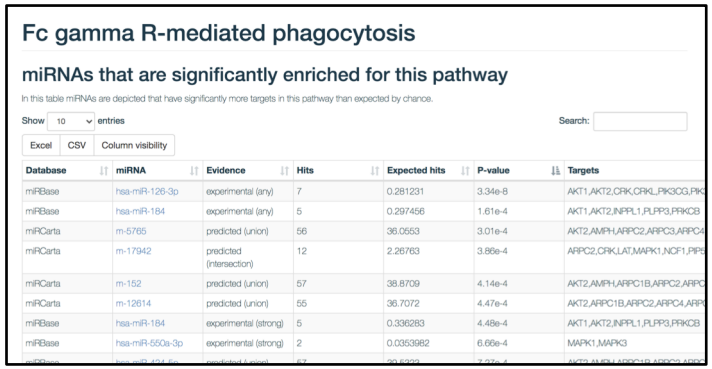
GWAS -- > MAGENTA

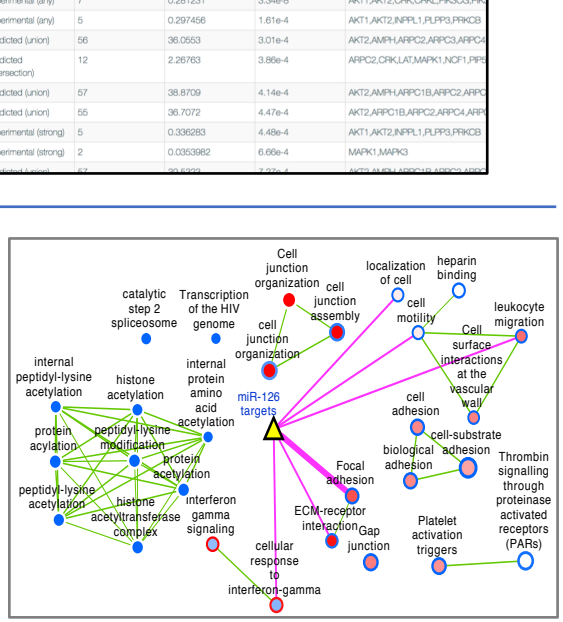
https://software.broadinstitute.org/mpg/magenta/

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.

Mirs, pathways and targets





miEAA: microRNA enrichment analysis and annotation

http://www.ccb.uni-saarland.de/mieaa_tool/

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

EnrichmentMap
Post analysis
Mir targets

Result

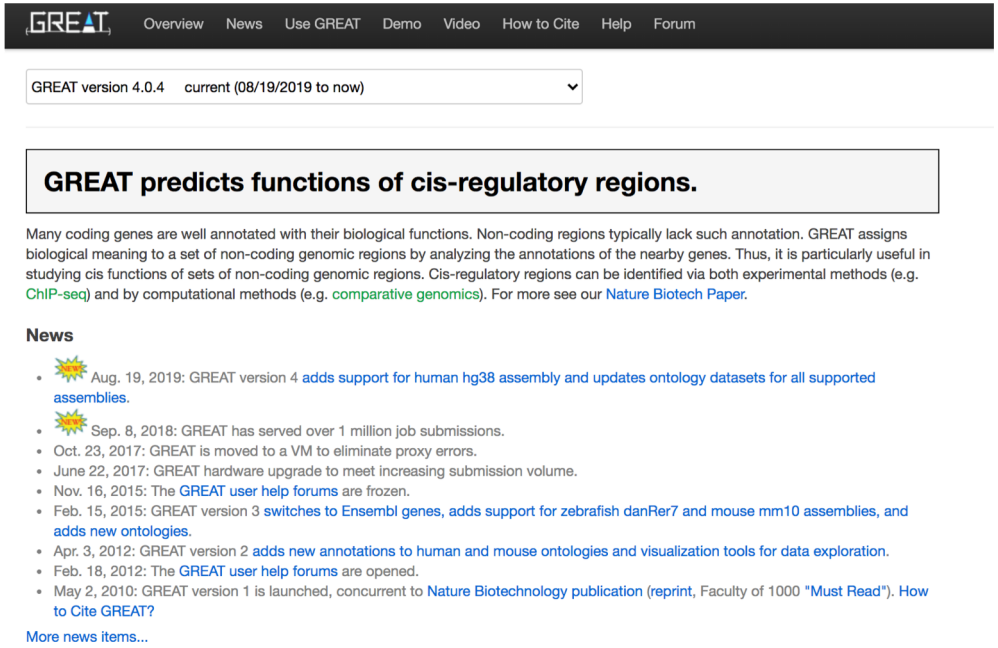
Enrichment analysis results

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	0.16667	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

ATACseq / CHIPseq

- EnrichR and g:Profiler accept bed files as input
- GREAT (Standford) is also a recommended tool
- HOMER: to look for enrichment factors in transcription factors

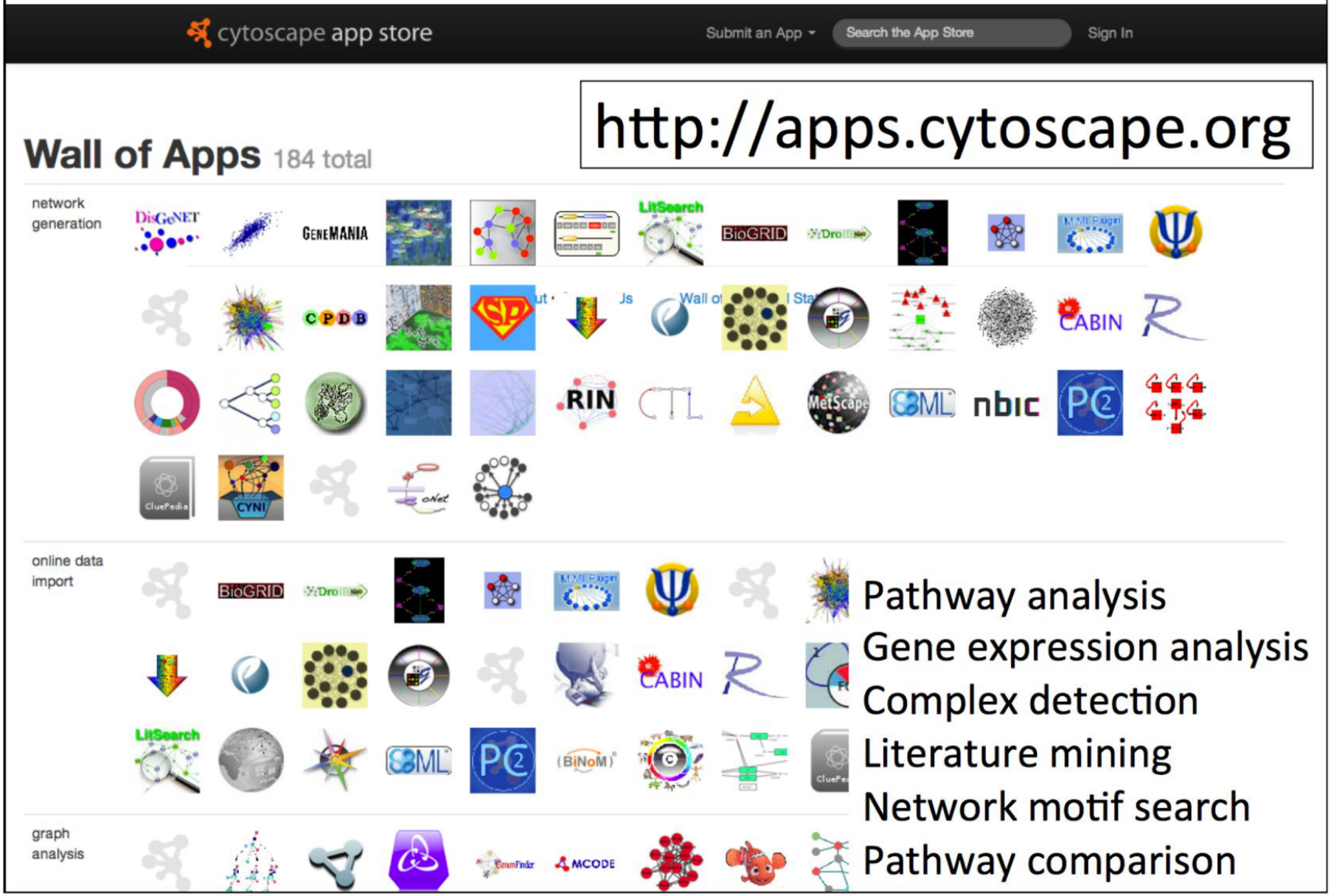


RNAseq :2 class design

- GSEA
- Enrichment Map

- Single cell Data
  - GSVA() in R or Wilcoxon Rank sum test (R, Panther)

The Cytoscape App Store



We are on a Coffee Break & Networking Session

Workshop Sponsors: