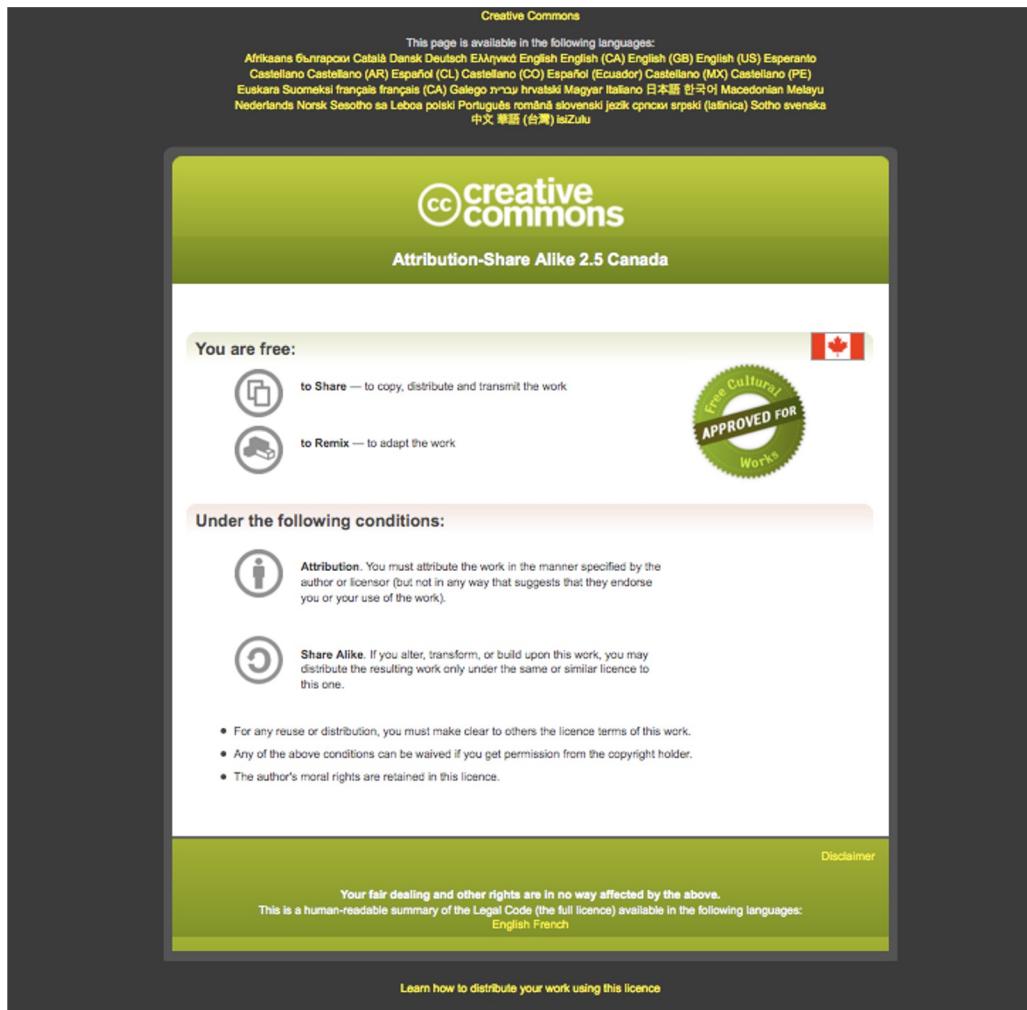




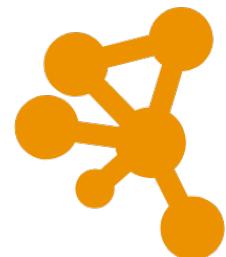
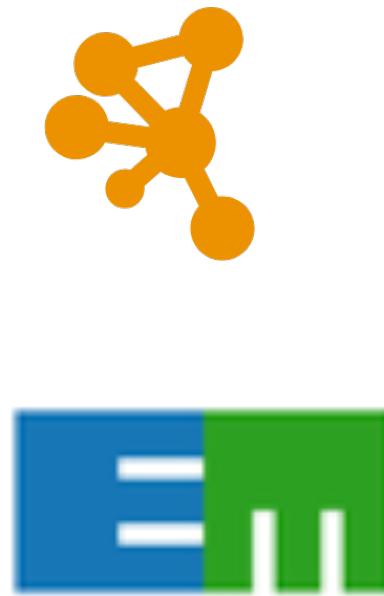
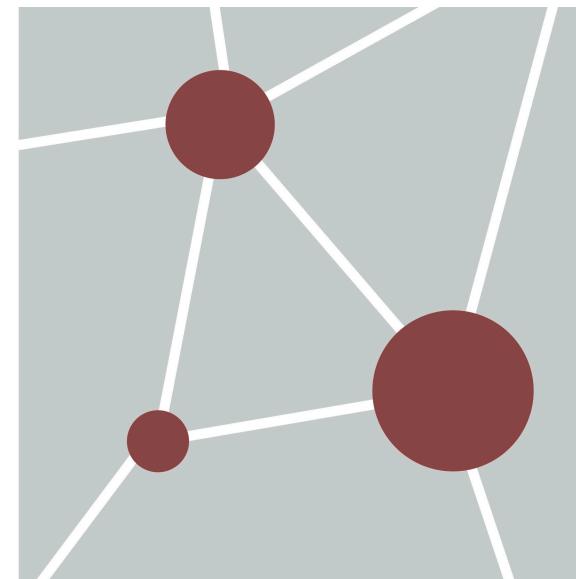
Canadian Bioinformatics Workshops

www.bioinformatics.ca

bioinformaticsdotca.github.io



Module 3 Enrichment Maps



**BADER
LAB**



UNIVERSITY OF
TORONTO



Learning Objectives

- By the end of this lecture, you will:
 - Understand...
 - Be able to define...
 - Know...

Learning Objectives

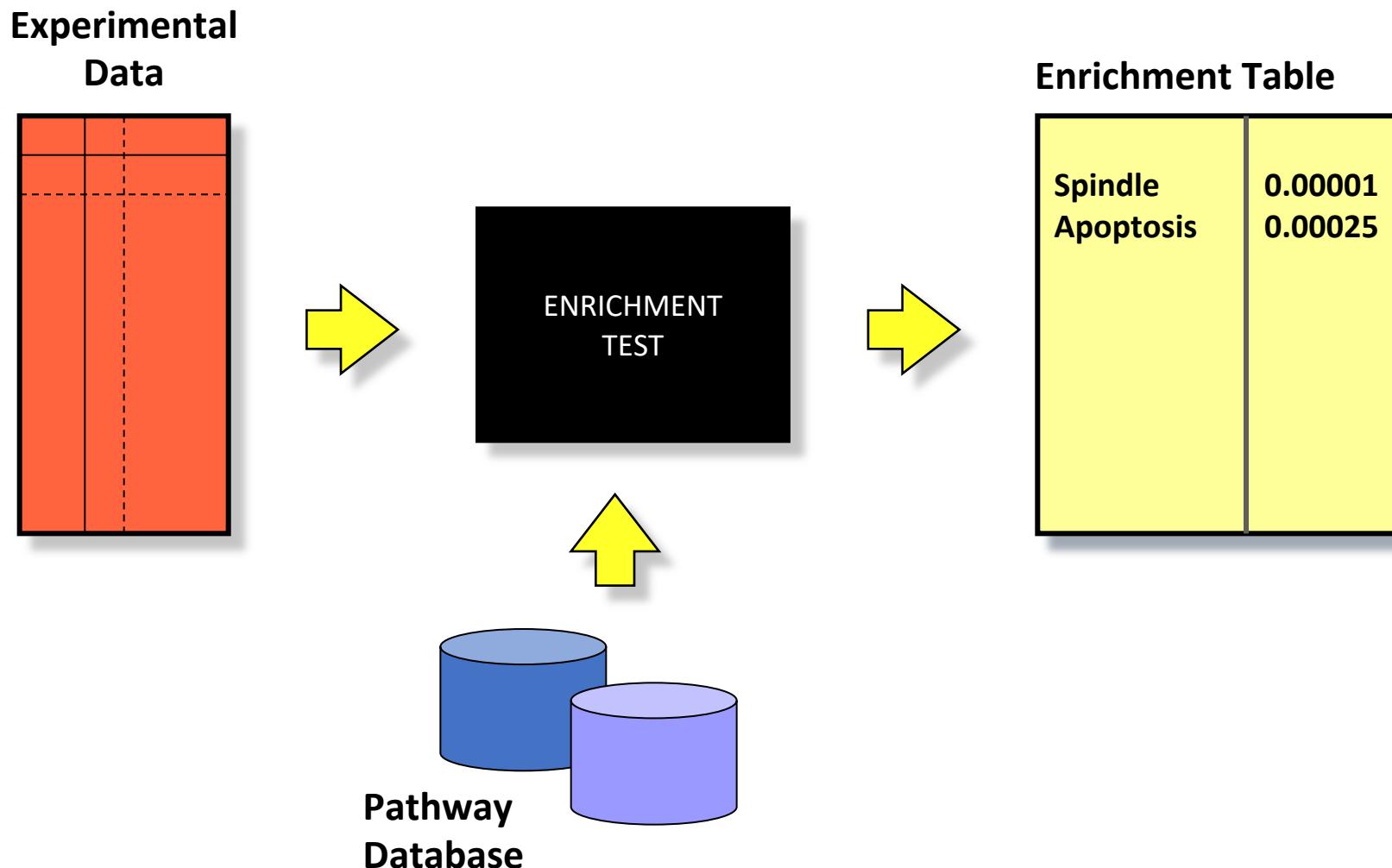
- By the end of this lecture, you will:
 - Understand how to transform your enrichment results from g:profiler, gsea or other enrichment algorithms to a Network
 - Understand the differences between a network and an Enrichment Map
 - Be able to summarize your enrichment results with annotation using the Autoannotate App.

Results of Lab 2 - Enrichment Results

GO.ID	Description	p.Val	FDR	Phenotype	Genes	GS follow GS DETAILS	SIZE	ES	NES	NOM p-val	FDR q-val
GO:0060255	regulation of macromolecule metabolic process	2.34E-30	2.34E-30		1 TP53,PTEN,APC,VHL,KRAS,ARID1A,EGFR,NF1,PIK3R1,CDKN2A,GATA3,RB1,NOTCH	HALLMARK_INTERFERON_GAMMA_RESPONSE%MSIGDB_C2%HALLMARK_INTERFERON_GAMMA_RESPONSE	195	-0.8124175	-3.1241286	0	0
GO:0051171	regulation of nitrogen compound metabolic process	6.44E-30	6.44E-30		1 TP53,PTEN,APC,VHL,KRAS,ARID1A,EGFR,NF1,PIK3R1,CDKN2A,GATA3,RB1,NOTCH	HALLMARK_INTERFERON_ALPHA_RESPONSE%MSIGDB_C2%HALLMARK_INTERFERON_ALPHA_RESPONSE	94	-0.871453	-2.969929	0	0
GO:0031323	regulation of cellular metabolic process	2.80E-29	2.80E-29		1 TP53,PIK3CA,PTEN,APC,VHL,KRAS,ARID1A,EGFR,NF1,PIK3R1,CDKN2A,GATA3,RB1,NOTCH	RESPONSE TO INTERFERON-GAMMA%GOBP%GO:0034341	119	-0.7410603	-2.8620038	0	0
GO:0019222	regulation of metabolic process	3.50E-29	3.50E-29		1 TP53,PIK3CA,PTEN,APC,VHL,KRAS,ARID1A,EGFR,NF1,PIK3R1,CDKN2A,GATA3,RB1,NOTCH	DEFENSE RESPONSE TO VIRUS%GOBP%GO:0051607	102	-0.7547711	-2.8492434	0	0
GO:0080090	regulation of primary metabolic process	4.34E-29	4.34E-29		1 TP53,PTEN,APC,VHL,KRAS,ARID1A,EGFR,NF1,PIK3R1,CDKN2A,GATA3,RB1,NOTCH	INTERFERON SIGNALING%REACTOME%R-HSA-913531.2	169	-0.7541074	-2.7944484	0	0
GO:0051173	positive regulation of nitrogen compound metabolic process	1.35E-28	1.35E-28		1 TP53,PTEN,APC,VHL,KRAS,ARID1A,EGFR,PIK3R1,CDKN2A,GATA3,RB1,NOTCH1,FBS	HALLMARK_ALLOGRAFT_REJECTION%MSIGDB_C2%HALLMARK_ALLOGRAFT_REJECTION	181	-0.7311862	-2.784118	0	0
GO:0010604	positive regulation of macromolecule metabolic process	1.51E-28	1.51E-28		1 TP53,PTEN,APC,VHL,KRAS,ARID1A,EGFR,PIK3R1,CDKN2A,GATA3,RB1,NOTCH1,FBS	CELLULAR RESPONSE TO TYPE I INTERFERON%GOBP%GO:0071357	50	-0.8616462	-2.7440908	0	0
GO:1903506	regulation of nucleic acid-templated transcript	5.40E-27	5.40E-27		1 TP53,PTEN,VHL,ARID1A,EGFR,PIK3R1,CDKN2A,GATA3,RB1,NOTCH1,FBXW7,CTNN	INTERFERON GAMMA SIGNALING%REACTOME%R-HSA-877300.2	81	-0.8194266	-2.7424092	0	0
GO:2001141	regulation of RNA biosynthetic process	5.40E-27	5.40E-27		1 TP53,PTEN,VHL,ARID1A,EGFR,PIK3R1,CDKN2A,GATA3,RB1,NOTCH1,FBXW7,CTNN	ANTIGEN PROCESSING AND PRESENTATION OF EXOGENOUS PEPTIDE ANTIGEN VIA MHC CLASS I, TAP-DEPENDENT ANTIGEN PRESENTATION%GOBP%GO:0060337	72	-0.7555878	-2.7347164	0	0
GO:0009893	positive regulation of metabolic process					TYPE I INTERFERON SIGNALING PATHWAY%GOBP%GO:0060337	64	-0.8510788	-2.702983	0	0
GO:0097659	nucleic acid-templated transcription					TYPE I INTERFERON SIGNALING PATHWAY%GOBP%GO:0034340	50	-0.8616462	-2.6922164	0	0
GO:0032774	RNA biosynthetic process					RESPONSE TO TYPE I INTERFERON%GOBP%GO:0034340	54	-0.8517665	-2.6604805	0	0
GO:0019219	regulation of nucleobase-containing compound metabolic process					ANTIGEN PROCESSING AND PRESENTATION OF EXOGENOUS PEPTIDE ANTIGEN VIA MHC CLASS I%GOBP%GC ANTIGEN PRESENTATION%GOBP%GO:0060337	77	-0.7585589	-2.5932803	0	0
GO:0031325	positive regulation of cellular metabolic process					ER-PHAGOSOME PATHWAY%REACTOME%R-HSA-1236974.4	81	-0.7237594	-2.589577	0	0
GO:0006355	regulation of transcription, DNA-templated					COSTIMULATION BY THE CD28 FAMILY%REACTOME%R-HSA-388841.3	61	-0.780563	-2.5822923	0	0
GO:0051252	regulation of RNA metabolic process					PROTEASOME DEGRADATION%WIKIPATHWAYS_20200610%WP183%HOMO SAPIENS	60	-0.7425067	-2.5736508	0	0
GO:0006351	transcription, DNA-templated					REGULATION OF LEUKOCYTE PROLIFERATION%GOBP%GO:0070663	142	-0.639763	-2.567436	0	0
GO:0010468	regulation of gene expression					ANTIGEN PROCESSING AND PRESENTATION OF PEPTIDE ANTIGEN VIA MHC CLASS I%GOBP%GO:0002474	94	-0.7129676	-2.560434	0	0
GO:2000112	regulation of cellular macromolecule biosynthesis, biogenesis					ANTIGEN RECEPTOR-MEDIATED SIGNALING PATHWAY%GOBP%GO:0050851	165	-0.6937604	-2.5514965	0	0

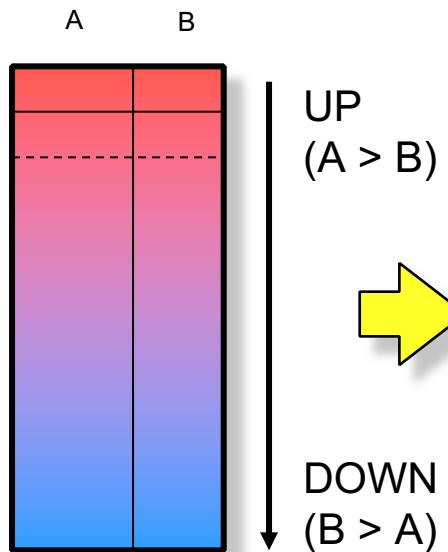
- Over the years thousands of papers have been translated into digital representations as pathways or functional annotation.
- Enrichment analysis allows us to interpret data from these papers
- But Major cognitive burden relating overlapping information

Pathway Enrichment Test: General Framework

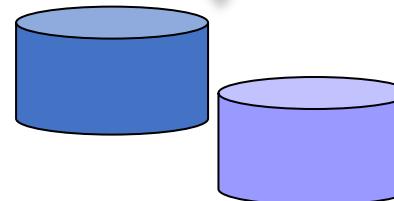


Pathway Enrichment Analysis

Ranked Gene List



GSEA



Pathways

g:Profiler

Thresholded lists (selected genes)

Enrichment in Condition A vs. B

Gene-set	Significance
Cell Cycle	0.0001
EGF Pathway	0.003
Spindle	0.007
...	...

Enrichment in Condition B vs. A

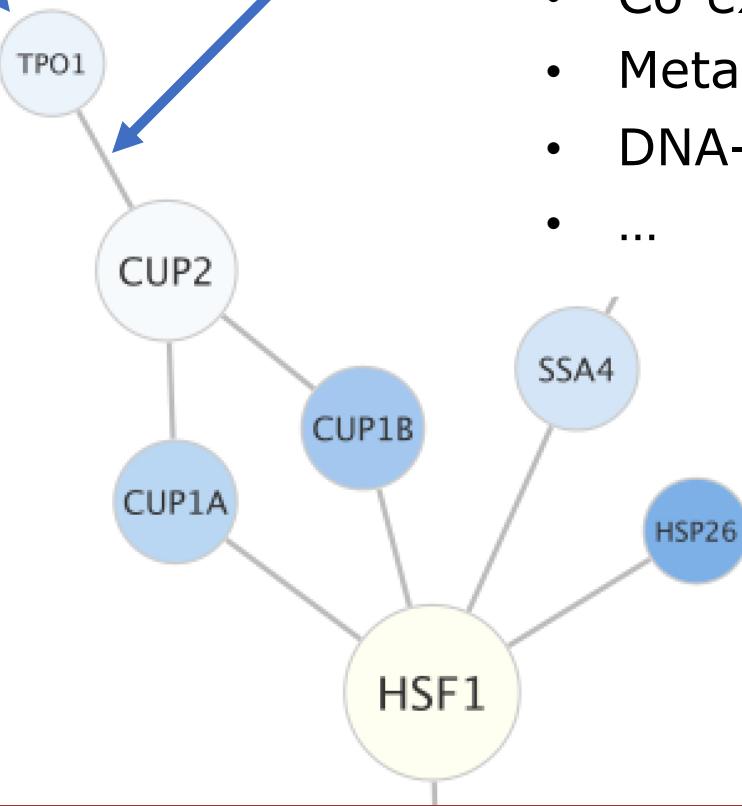
Gene-set	Significance
Proteasome	0.0002
Apoptosis	0.005
Caspase	0.009
...	...

Merico D, Isserlin R, Stueker O, Emili A, Bader GD
Enrichment map: a network-based method for
gene-set enrichment visualization and
interpretation PLoS One. 2010 Nov 15;5(11):e13984

Network Basics

Node (molecule/entity)

- Gene
- Protein
- Transcript
- Drug
- MicroRNA
- ...



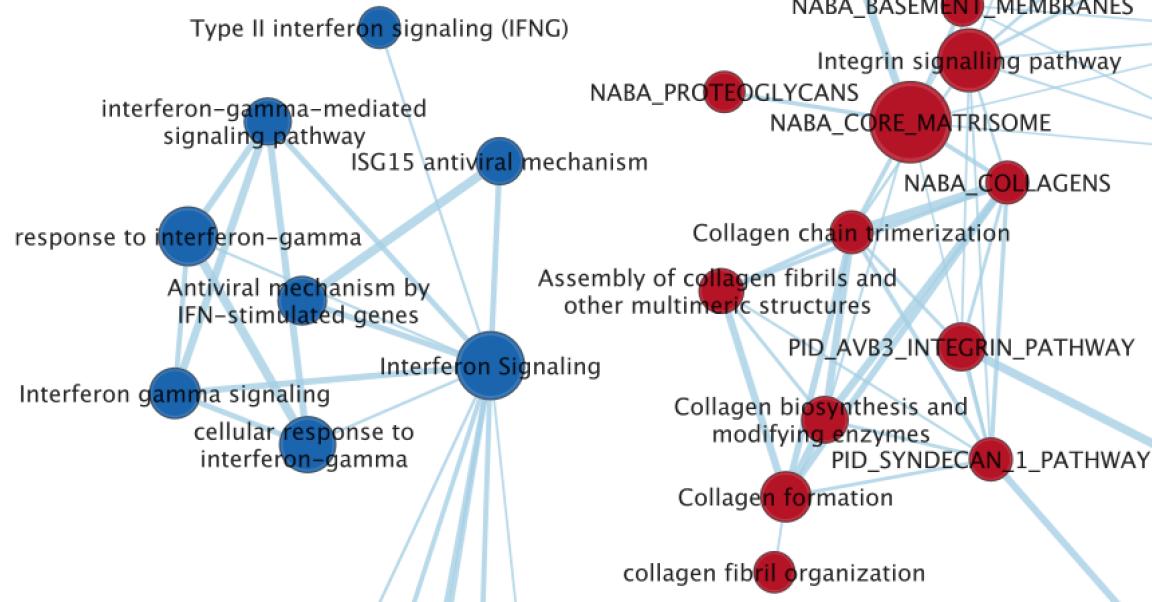
Edge (interaction/relationship)

- Genetic interaction
- Physical protein interaction
- Co-expression
- Metabolic reaction
- DNA-binding
- ...

Enrichment Map Basics

Node (molecule/entity)

- Pathway or geneset
- Size is correlated to number of genes in set
- Color indicates class in below example (for example Up/Down, classA/classB)



Edge (interaction/relationship)

- Degree of overlap between two genesets
- The more genes two pathways have in common the thicker/stronger the connection

Enrichment Map

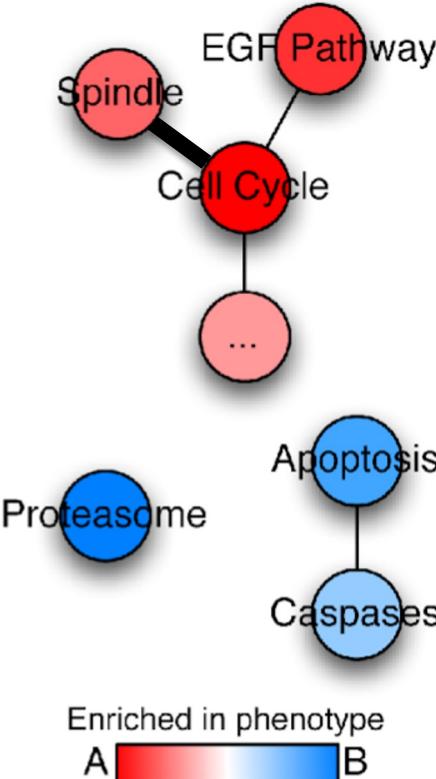
Enrichment in Condition A vs. B	
Gene-set	Significance
Cell Cycle	0.0001
EGF Pathway	0.003
Spindle	0.007
...	...

Enrichment in Condition B vs. A	
Gene-set	Significance
Proteasome	0.0002
Apoptosis	0.005
Caspase	0.009
...	...

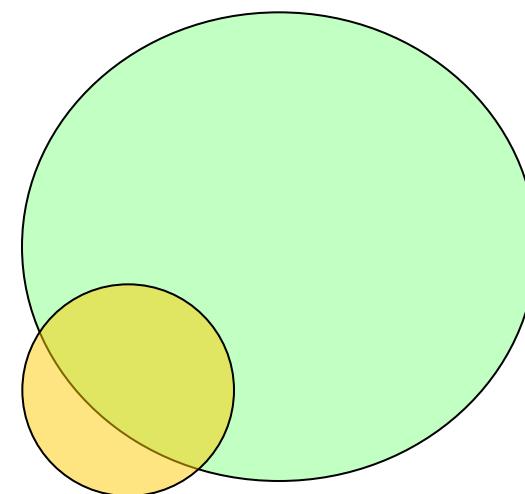
GENE-SET LIST



ENRICHMENT MAP



Overlap



$$\frac{|A \cap B|}{\min(|A|, |B|)}$$

Typical Output

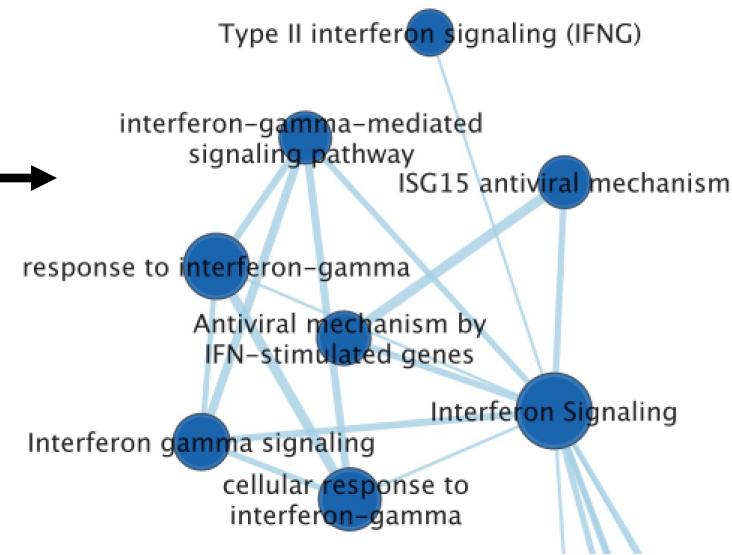
RNA HELICASE ACTIVITY%GO:GO-0003724	28	1.77	0.0041	0.0464386
MRNA SURVEILLANCE PATHWAY%KEGG%HSA03015	82	1.77	0	0.0466167
UBIQUITIN-DEPENDENT DEGRADATION OF CYCLIN D1%REACTOME%REACT_4.1	50	1.77	0.0021	0.0486015
BIOCARTA_CD40_PATHWAY%MSIGDB_C2%BIOCARTA_CD40_PATHWAY	15	1.77	0.0048	0.0483781
IGF1 PATHWAYS%PATHWAY INTERACTION DATABASE NCI-NATURE CURATED DATA%IGF1 PATHWAY	29	1.76	0.003	0.0489742
UBIQUITIN-DEPENDENT PROTEIN CATABOLIC PROCESS%GO:GO-0006511	204	1.76	0	0.0488442
PHAGOSOME%KEGG%HSA04145	147	1.76	0	0.0486164
PROTEASOME COMPLEX%GO:GO-0000502	29	1.76	0.007	0.0490215
ANTIGEN PRESENTATION: FOLDING, ASSEMBLY AND PEPTIDE LOADING OF CLASS I MHC%REACTOME%REACT_7	24	1.76	0.0041	0.0505259
ABORTIVE ELONGATION OF HIV-1 TRANSCRIPT IN THE PRESENCE OF TAT%REACTOME%REACT_6261.3	23	1.75	0	0.0529242
DNA DAMAGE RESPONSE, SIGNAL TRANSDUCTION BY P53%ASS MEDIAFOR RESULTING IN CELL CYCLE ARREST%GO:GO-0001365	67	1.75	0	0.052886
REGULATION OF MACROPHAGE ACTIVATION%GO:GO-0001330	11	1.75	0.003	0.0534709
PROTEIN FOLDING%REACTOME%REACT_16952.1	52	1.75	0.002	0.0537717
ENDOPLASMIC RETICULUM UNFOLDED PROTEIN RESPONSE%GO:GO-0006568	73	1.75	0	0.0546052
PROTEIN EXPORT%KEGG%HSA03060	24	1.75	9.75E-04	0.0548699
TRANSCRIPTION INITIATION FROM RNA POLYMERASE II PROMOTER%GO:GO-0006367	64	1.75	0.001	0.0545783
S PHASE%REACTOME%REACT_899.4	110	1.75	0	0.0546003
PROTEASOMAL PROTEIN CATABOLIC PROCESS%GO:GO-0006510	163	1.75	0	0.0550066
ATP-DEPENDENT RNA HELICASE ACTIVITY%GO:GO-0006511	20	1.74	0.0059	0.0556722
ACID-AMINO ACID LIGASE ACTIVITY%GO:GO-0016804	217	1.74	0	0.0560217
GO:GO-00722474	67	1.74	0.002	0.0565978
GO:GO-0035966	107	1.74	0	0.0562957
GO:GO-0072413	67	1.74	9.81E-04	0.05761
BIOCARTA_I4_PATHWAY%MSIGDB_C2%BIOCARTA_I4_PATHWAY	11	1.74	0.0082	0.0581508
ASSOCIATION OF TRIC CCT WITH TARGET PROTEINS DURING BIOSYNTHESIS%REACTOME%REACT_16907.2	28	1.74	0.0039	0.0581298
UBIQUITIN-DEPENDENT DEGRADATION OF CYCLIN D1%REACTOME%REACT_938.4	50	1.74	0.0029	0.057876
MODIFICATION-DEPENDENT PROTEIN CATABOLIC PROCESS%GO:GO-0019941	207	1.74	0	0.0576579
TRANSLATION INITIATION COMPLEX FORMATION%REACTOME%REACT_1979.1	55	1.74	0.0021	0.0575181
GO:GO-0001906	13	1.74	0.0117	0.0572877
G1 S TRANSITION%REACTOME%REACT_1783.2	107	1.74	0	0.0572618
GO:GO-0034620	73	1.73	0.0021	0.0576606
SIGNALING BY NOTCH%REACTOME%REACT_299.2	19	1.73	0.0069	0.0578565
RESPONSE TO UNFOLDED PROTEIN%GO:GO-0006986	102	1.73	0	0.0583864
SIGNAL TRANSDUCTION INVOLVED IN G1 S TRANSITION CHECKPOINT%GO:GO-0072404	68	1.73	0.002	0.0582213
GO:GO-0072431	67	1.73	0	0.058551
BIOCARTA_PROTEASOME_PATHWAY%MSIGDB_C2%BIOCARTA_PROTEASOME_PATHWAY	19	1.73	0.0099	0.0586655
HOT INTERACTIONS OF HIV FACTORS%REACTOME%REACT_6288.4	117	1.73	0	0.0586888
AUTOPHAGIC VACUOLE ASSEMBLY%GO:GO-0000045	13	1.73	0.0122	0.0588271
CYCLIN A:CDK2 ASSOCIATED EVENTS AT S PHASE ENTRY%REACTOME%REACT_9029.2	66	1.73	0	0.0610093
DUPLICATING REACTION IN THE NUCLEUS%REACTOME%REACT_2323.2	29	1.72	0.006	0.0610229

Each row is a gene-set (pathway).

It displays:

- a score associated with the magnitude of overlap between gene-set and gene list.
 - a pvalue that estimates the significance of the enrichment (by chance or not).
 - a corrected pvalue (FDR) that corrects for multiple hypothesis testing.

Network Visualization



Enrichment Map: use case I

Single enrichment

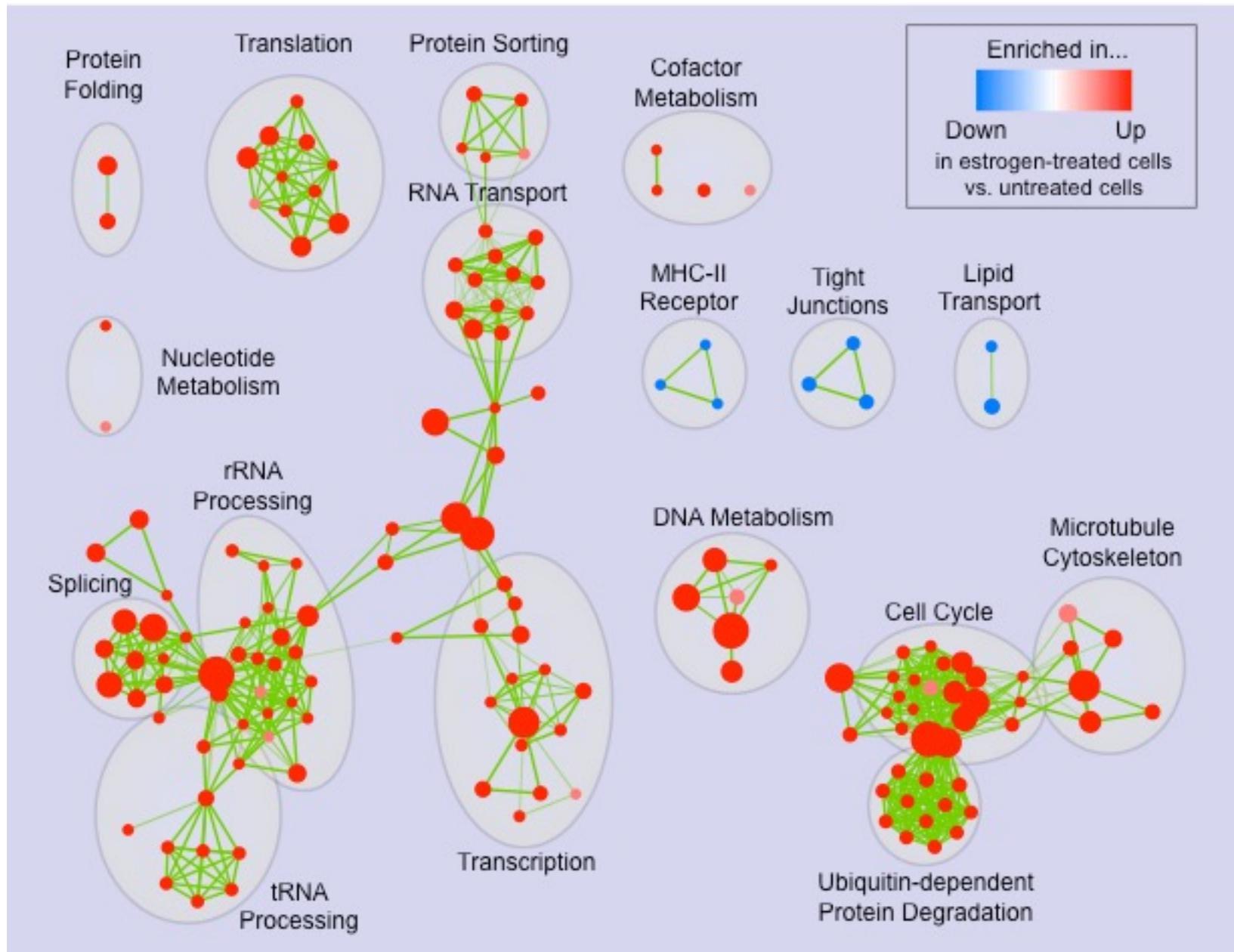
Estrogen treatment of breast cancer cells

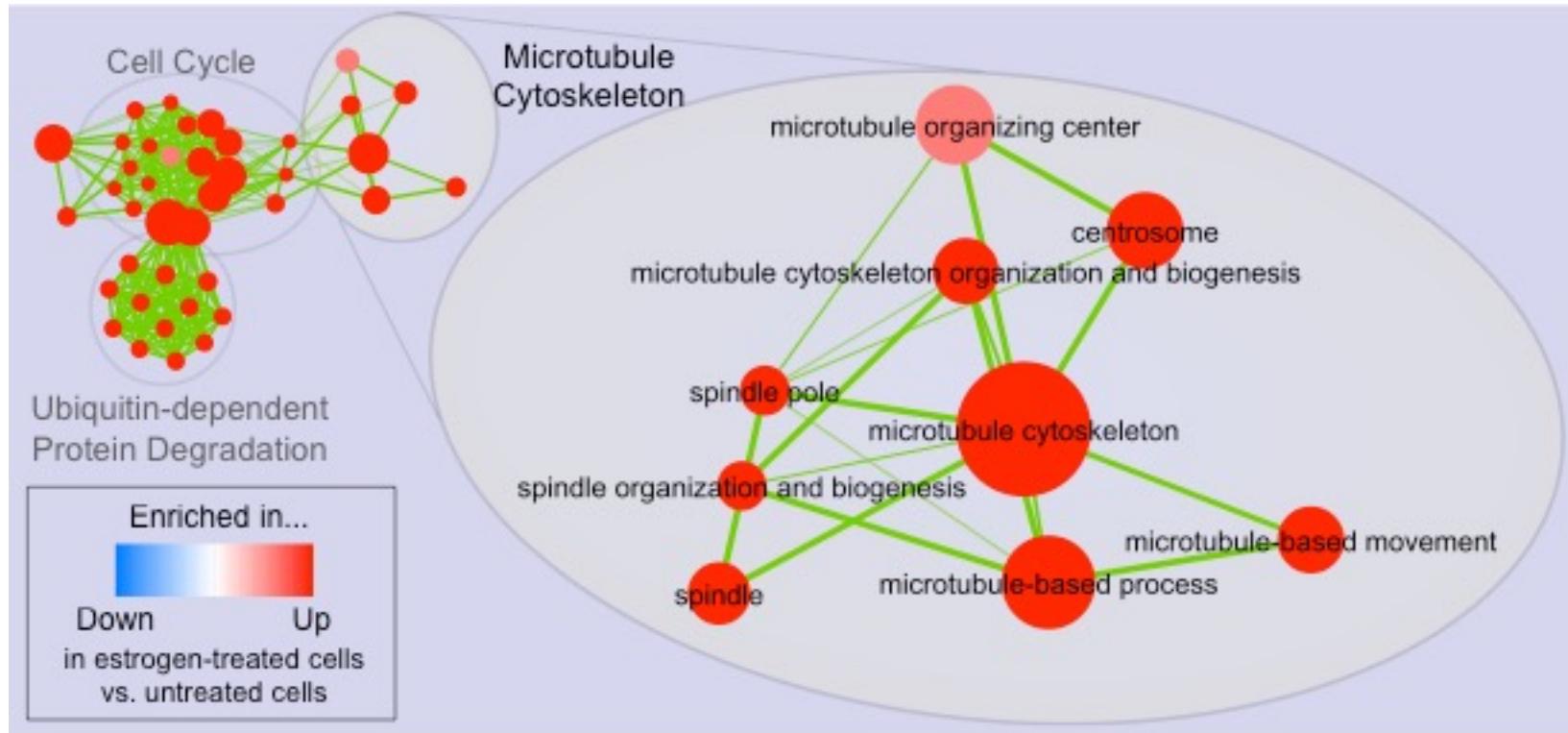
- Design:
2-time points, two-class

	12 hrs	24 hrs
Estrogen-treated	3	3
Untreated	3	3

- Gene set Database:
Gene Ontology

Lin C-Y, Vega VB, Thomsen JS, Zhang T, Kong SL, et al. (2007) Whole-genome cartography of estrogen receptor alpha binding sites. PLoS Genetics 3:e87





Enrichment Map: use case II

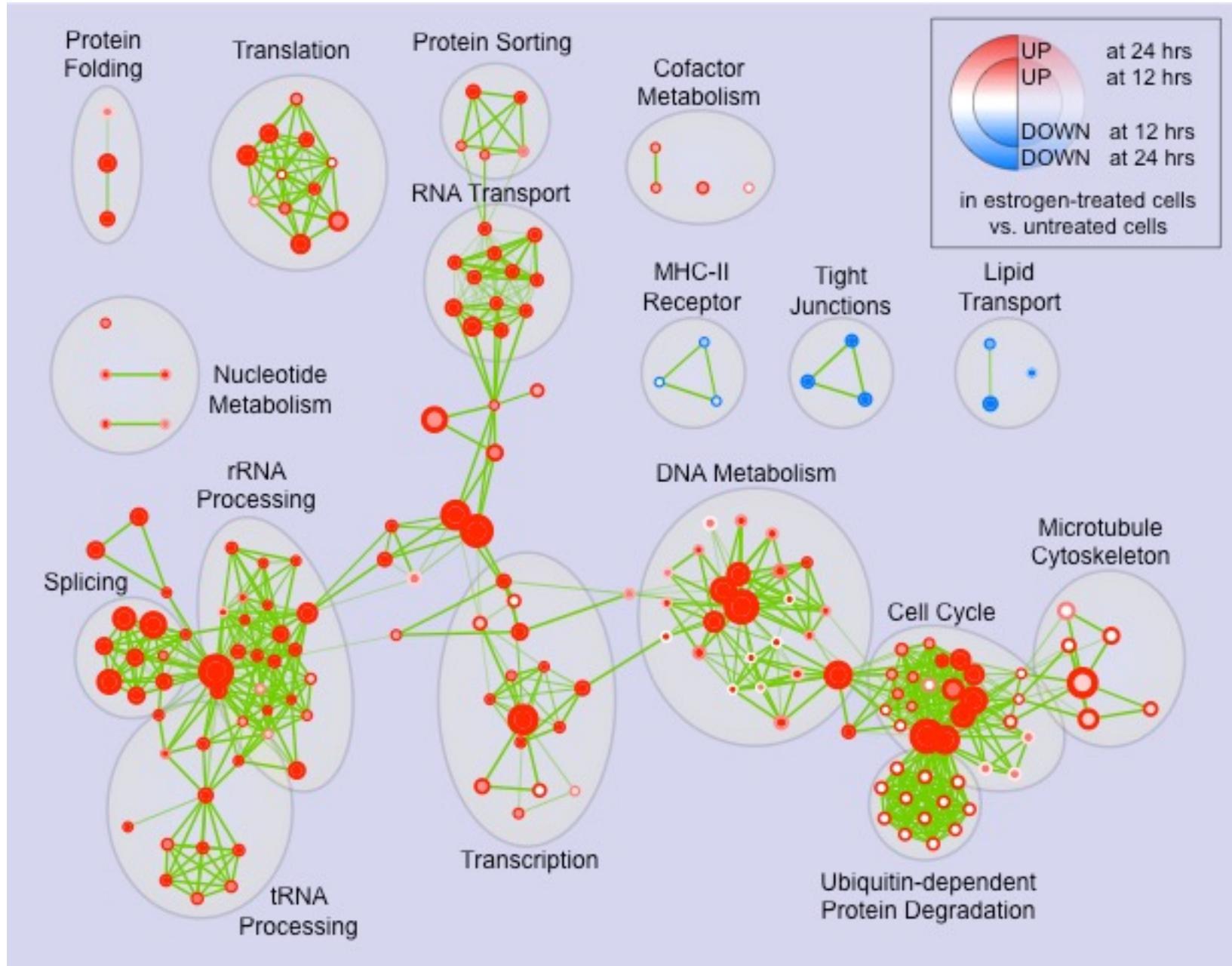
Comparison of two enrichments

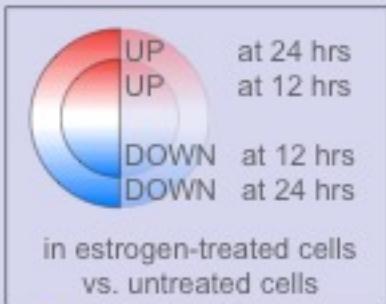
Estrogen treatment of breast cancer cells

- Design:
2-time points, two-class

	12 hrs	24 hrs
Estrogen-treated	3	3
Untreated	3	3

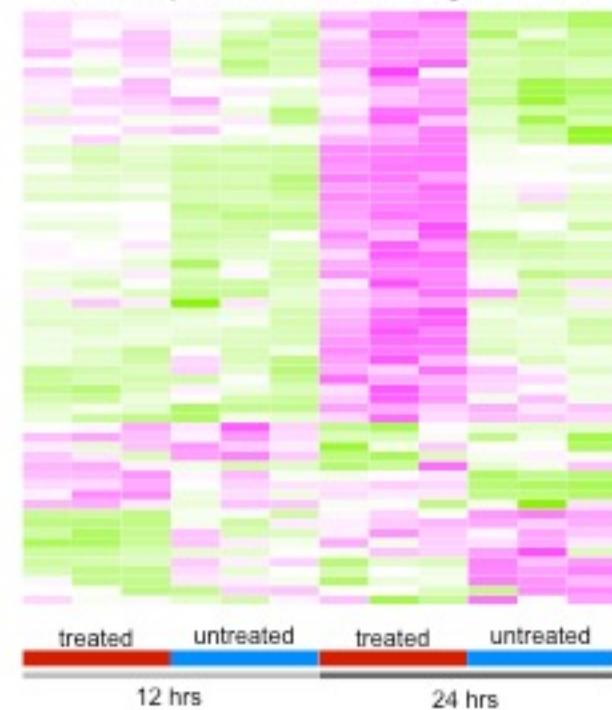
- Gene set Database:
Gene Ontology



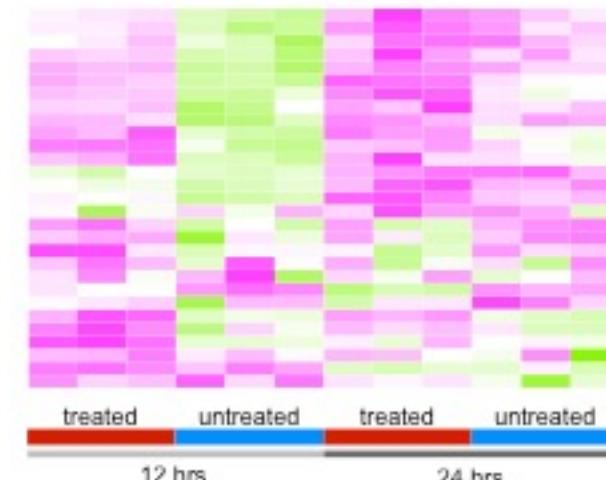


Estrogen-treated vs untreated:
■ Up
■ Down

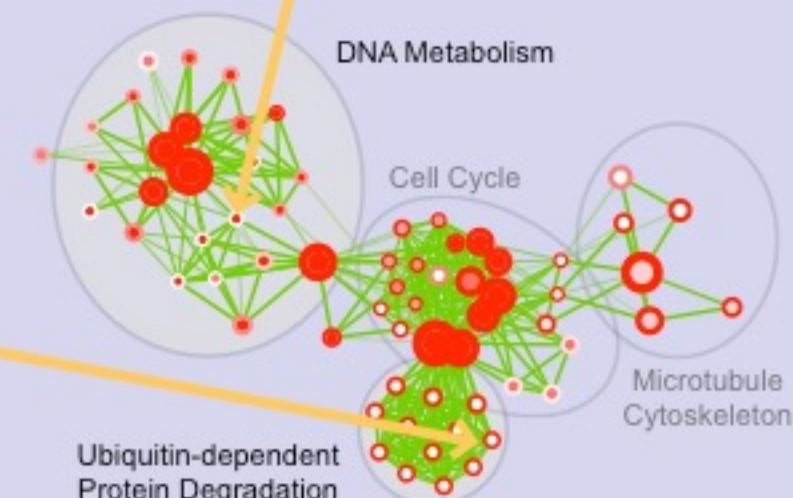
APC-dependent Protein Degradation



Replication Fork

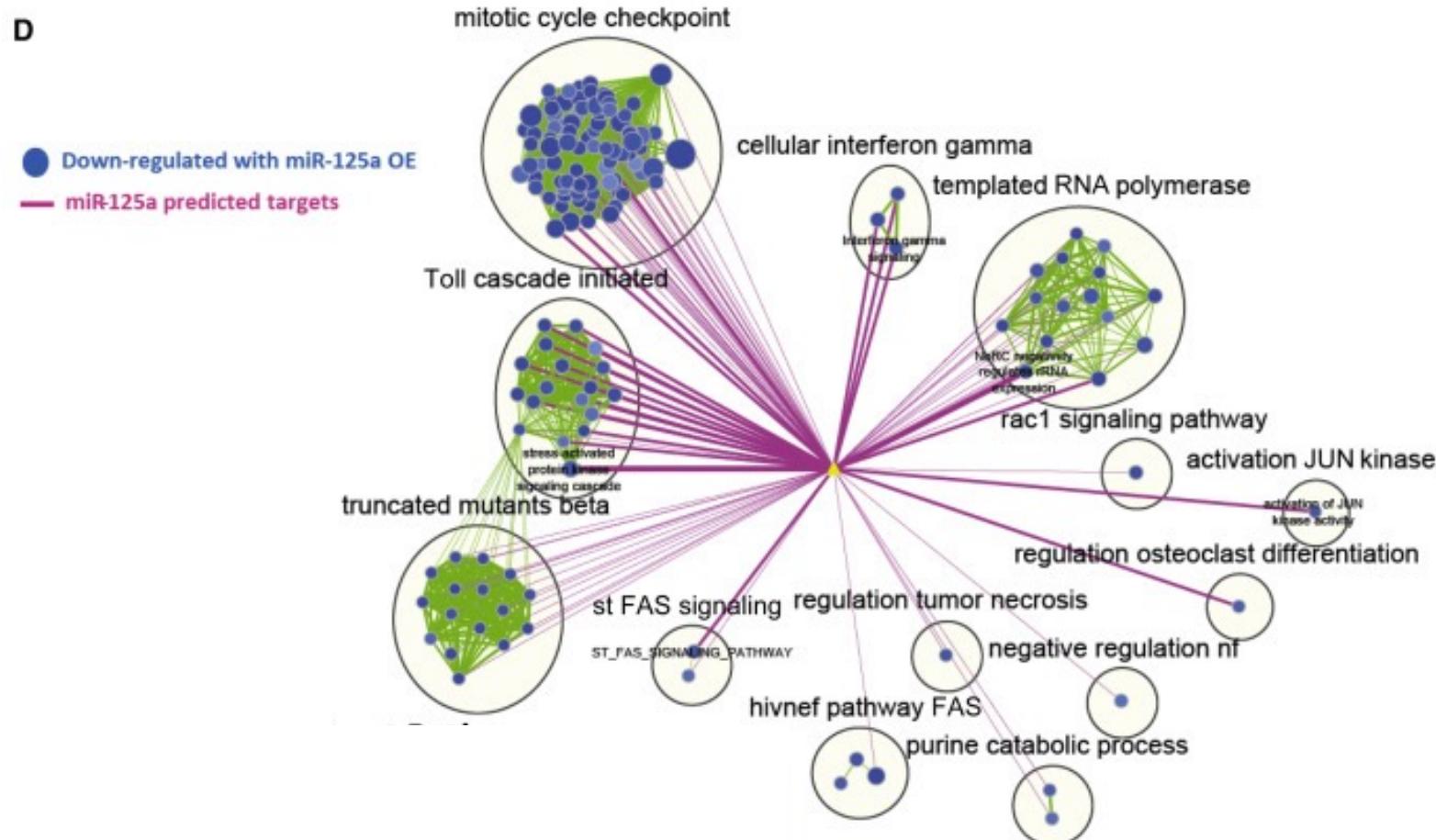


DNA Metabolism



Enrichment Map: use case III

Query Set Analysis

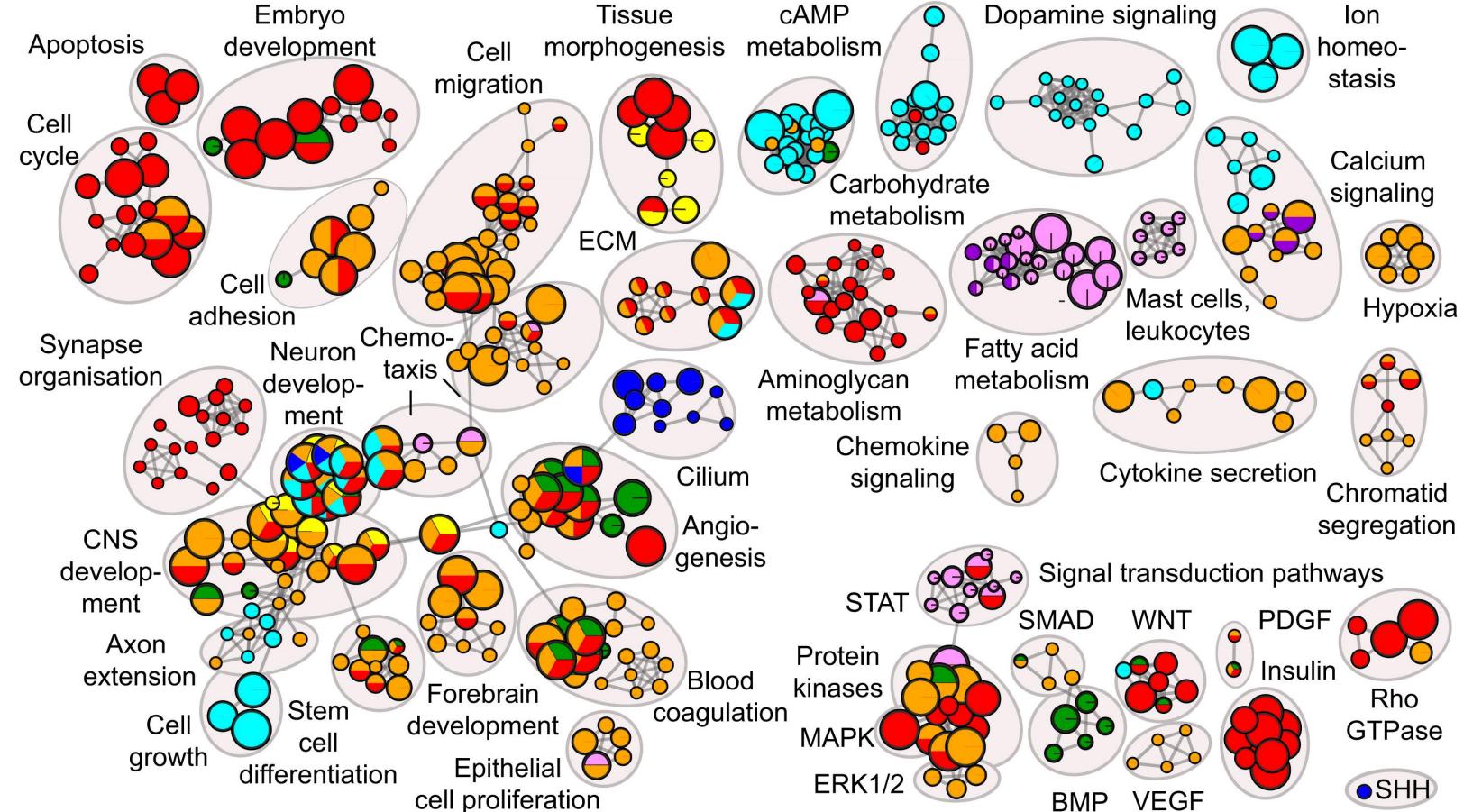


Wojtowicz EE, Lechman ER, Hermans KG, et al. Ectopic miR-125a Expression Induces Long-Term Repopulating Stem Cell Capacity in Mouse and Human Hematopoietic Progenitors. *Cell Stem Cell*. 2016;19(3):383-396

Enrichment Map: use case IV

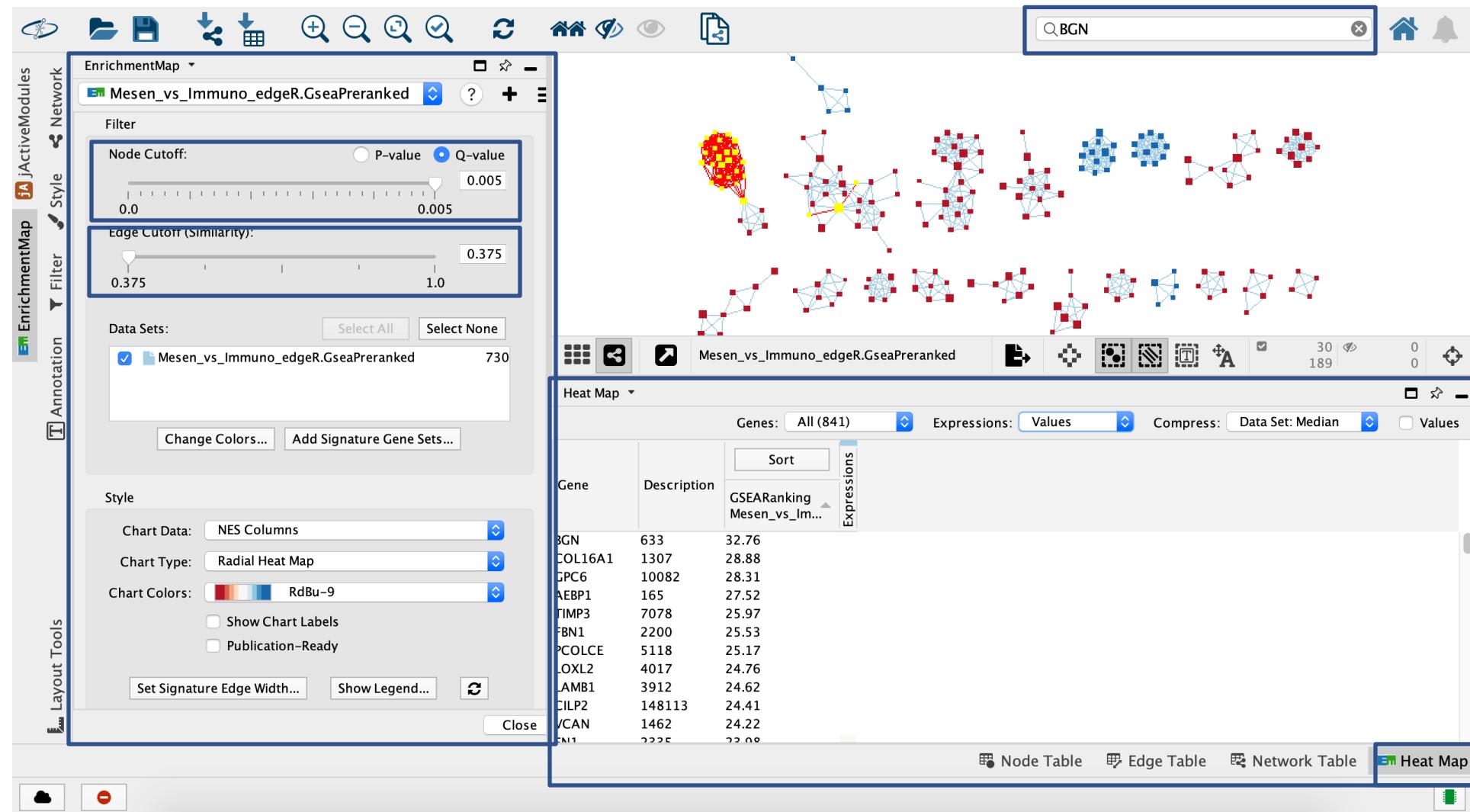
Multiple Datasets

c

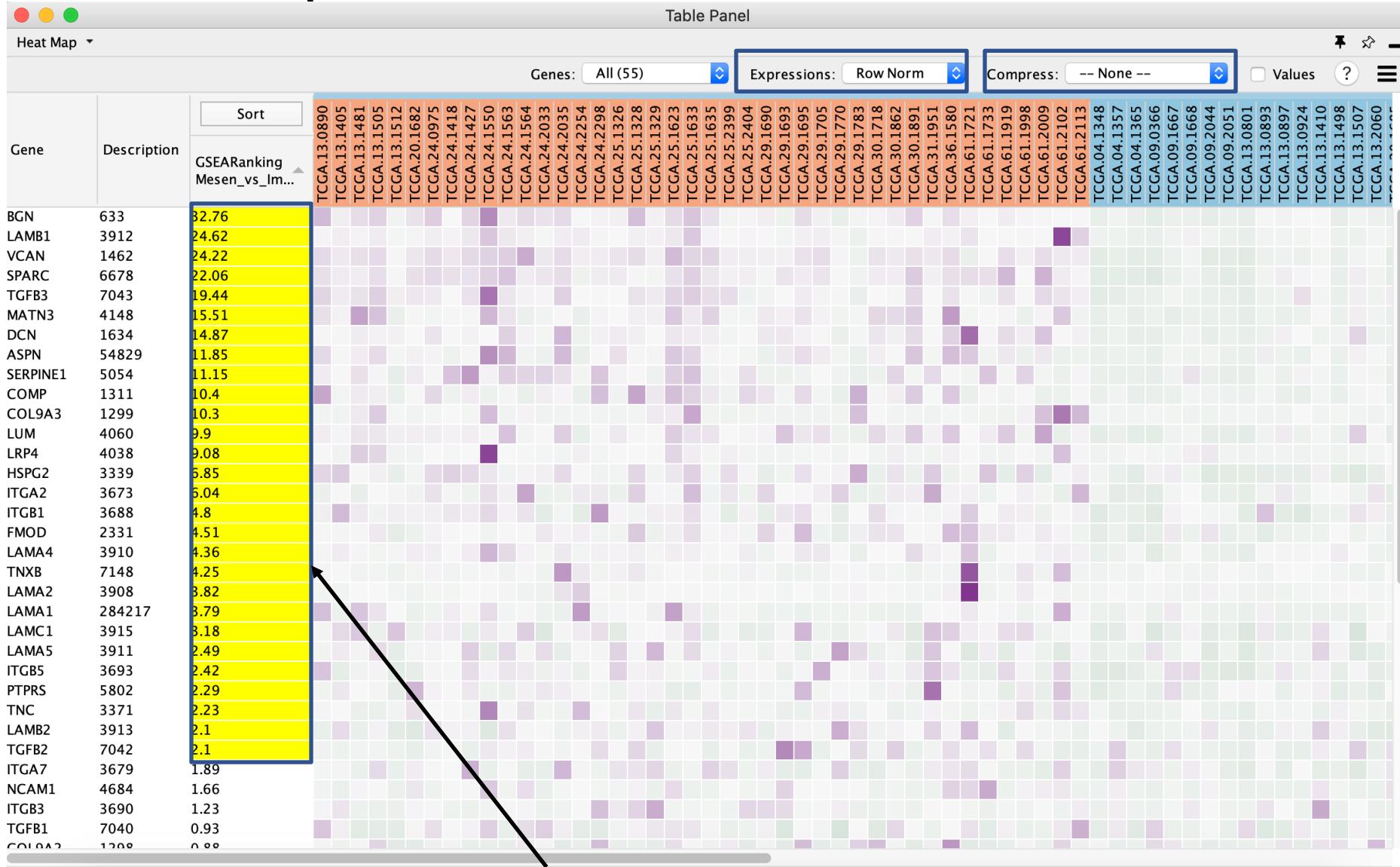


Pajtler KW, Witt H, Sill M, et al. Molecular Classification of Ependymal Tumors across All CNS Compartments, Histopathological Grades, and Age Groups. *Cancer Cell*. 2015;27(5):728-743

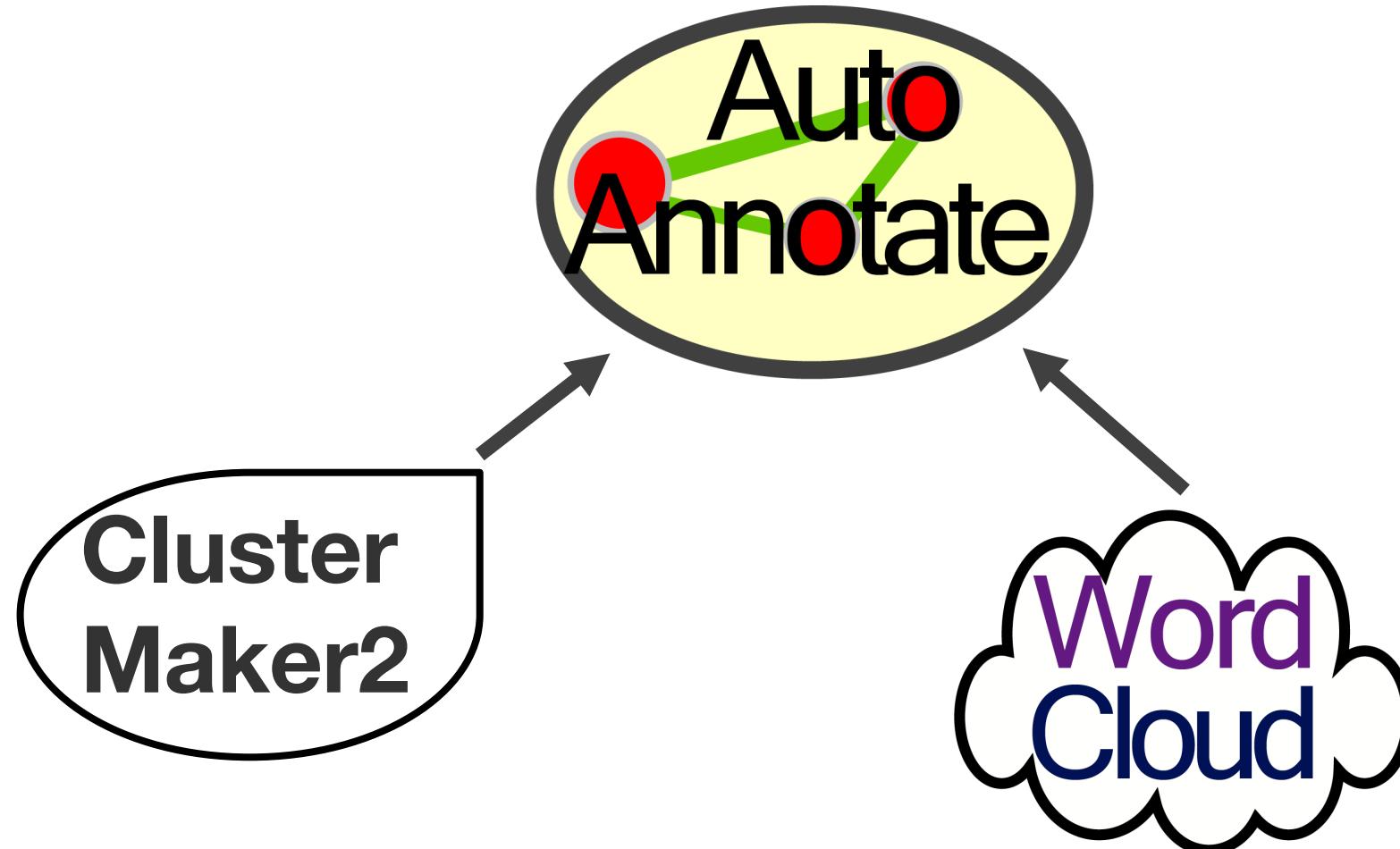
Enrichment Map Features



Heatmap Panel



Enrichment Map Features

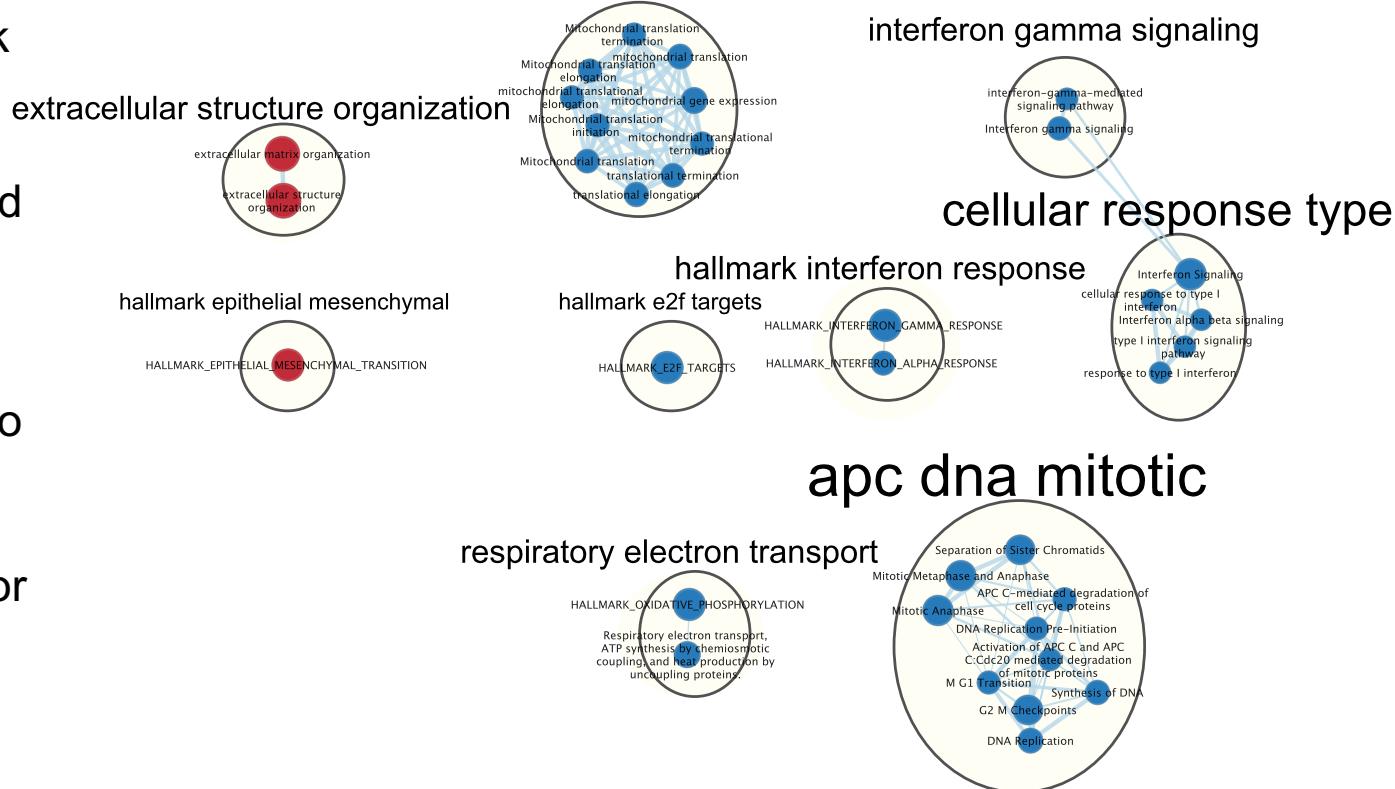


AutoAnnotate

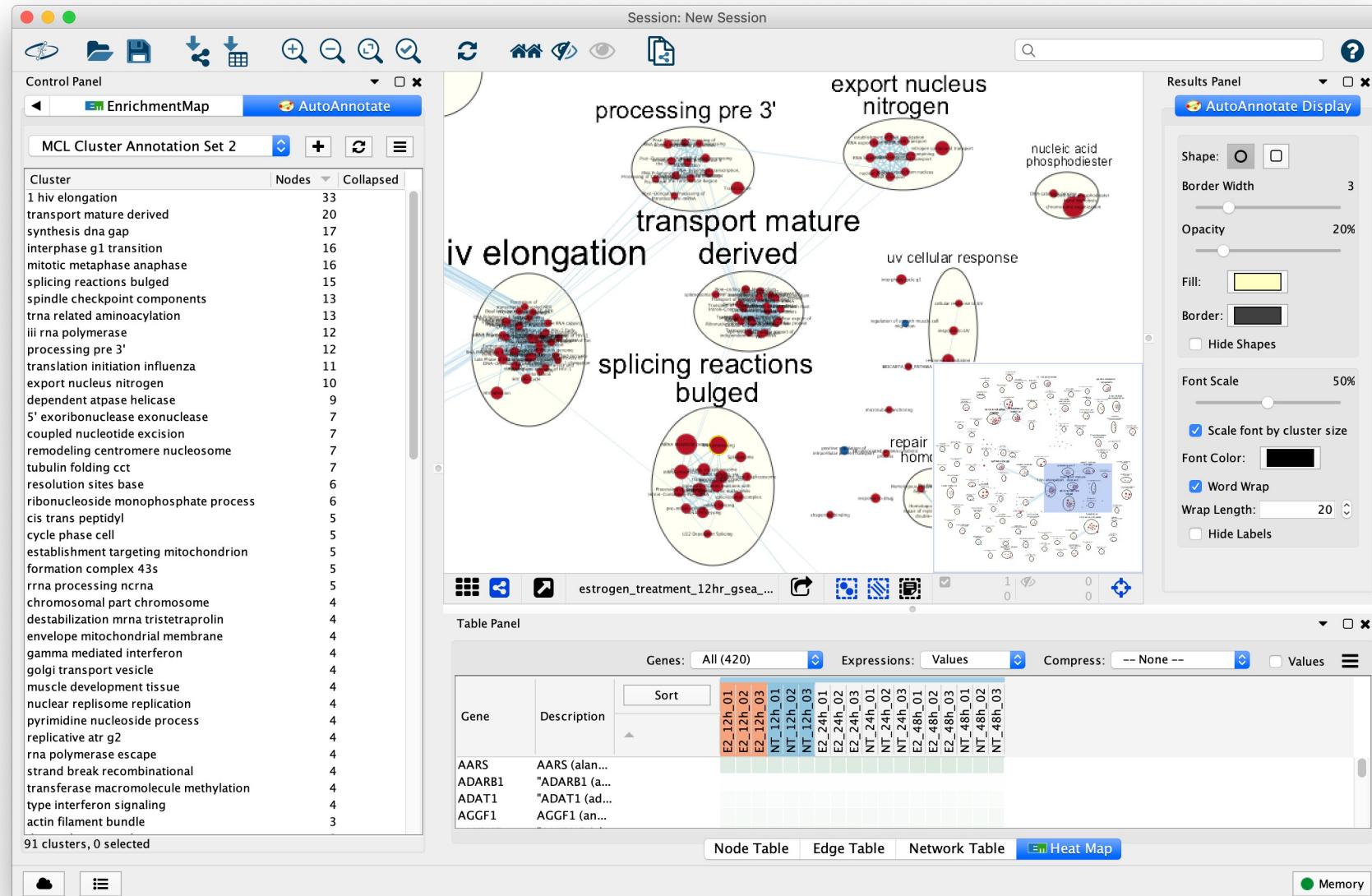
Steps:

1. Cluster the Network
2. For each cluster, find frequent words in node labels
3. Select top 3 words to create labels
4. Possibility to move or edit labels to make figure publication ready.

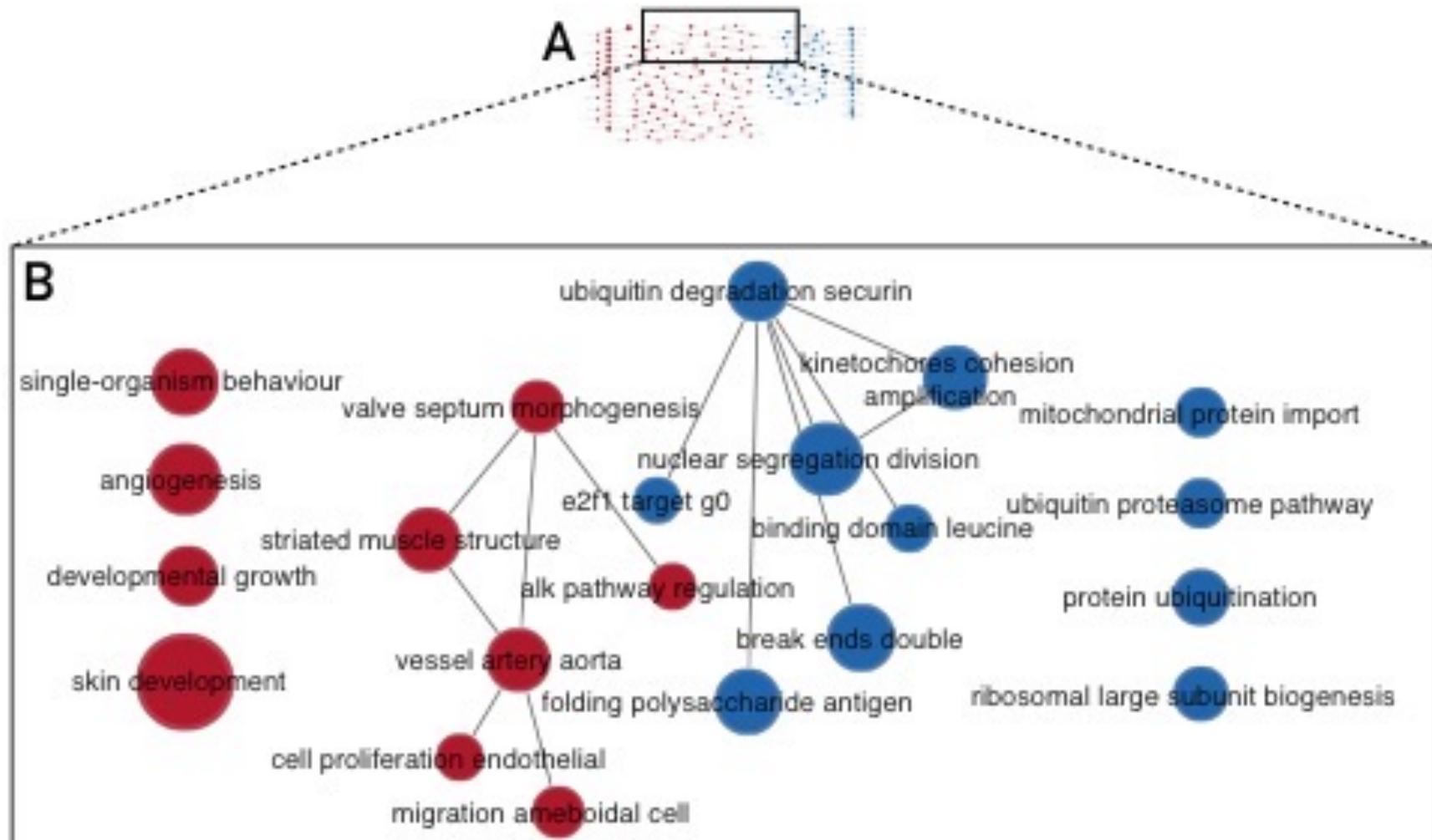
mitochondrial translational translation



AutoAnnotate



Collapsed network



Practical Lab

- Experiment with Cytoscape
- Try out Enrichment Map
- Load your g:Profiler results – and - load the GSEA enrichment analysis files that you generated in Module 2 lab
- Try out all the different Enrichment Map features
- Try AutoAnnotate to annotate your networks.

We are on a Coffee Break & Networking Session

Workshop Sponsors:



Canadian Centre for
Computational
Genomics



HPC4Health



GenomeCanada