

# Canadian Bioinformatics Workshops

[www.bioinformatics.ca](http://www.bioinformatics.ca)  
[bioinformaticsdotca.github.io](https://bioinformaticsdotca.github.io)

Creative Commons

This page is available in the following languages:

Afrikaans Български Català Dansk Deutsch Ελληνικό English English (CA) English (GB) English (US) Esperanto  
Español Castellano (AR) Español (CL) Castellano (CO) Español (Ecuador) Castellano (MX) Castellano (PE)  
Eesti Eesti keel Esperanto Español (CA) Galego 日本語 Italiano Magyar Italiano 日本語 日本語 Macedonian Malay  
Nederlands Norsk Sesotho sa Leboa polski Português română slovenski jezik cрnоsr ерsrаkи (latinica) Sotho sveneka  
中文 華語 (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

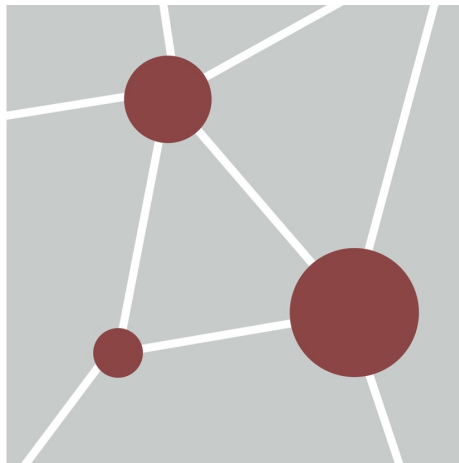
Your fair dealing and other rights are in no way affected by the above.  
This is a human-readable summary of the Legal Code (the full licence) available in the following languages:  
English French

[Learn how to distribute your work using this licence](#)

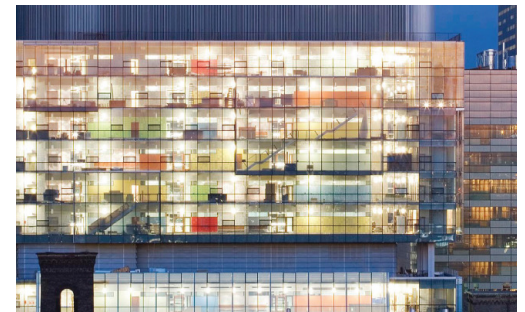
# Introduction to Pathway and Network Analysis of Gene Lists



Gary Bader  
Pathway and Network Analysis  
June 5-7, 2023

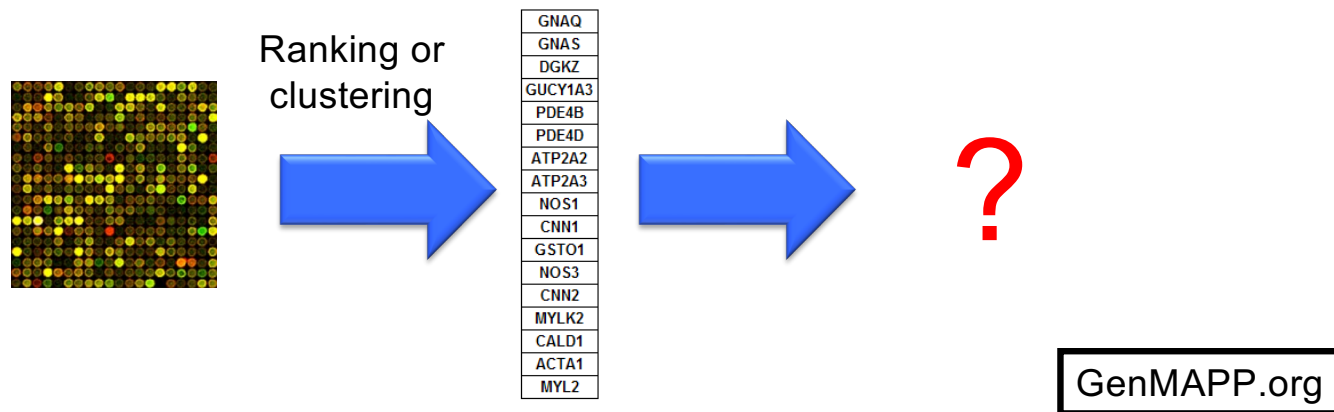


<http://baderlab.org>



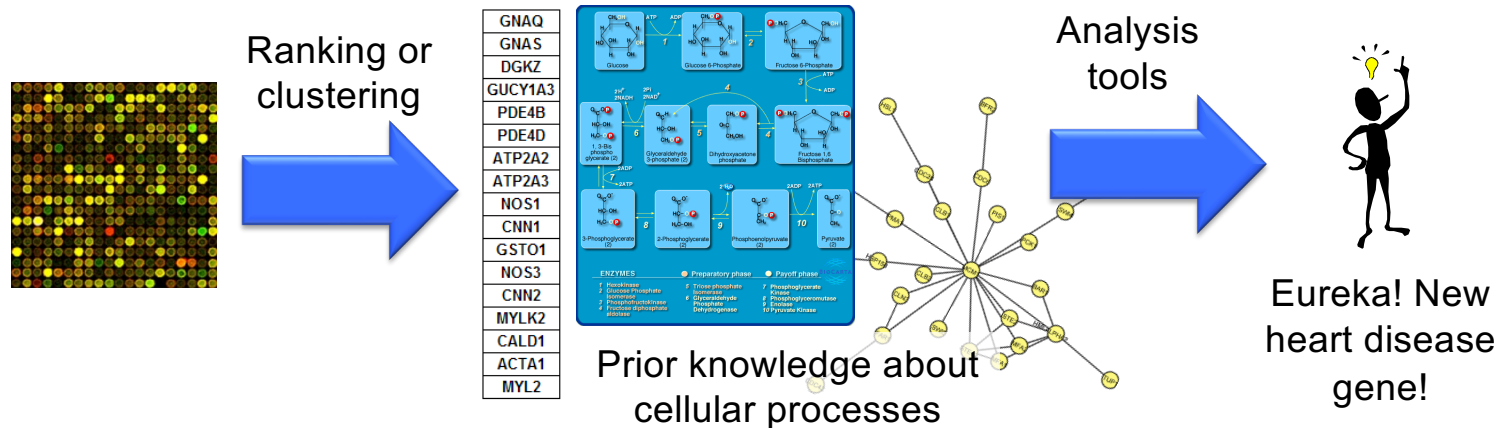
# Interpreting gene lists

- My cool new screen worked and produced 1000 hits! ...Now what?
- Genome-Scale Analysis (Omics)
  - Genomics, Proteomics
- Tell me what's interesting about these genes



# Interpreting gene lists

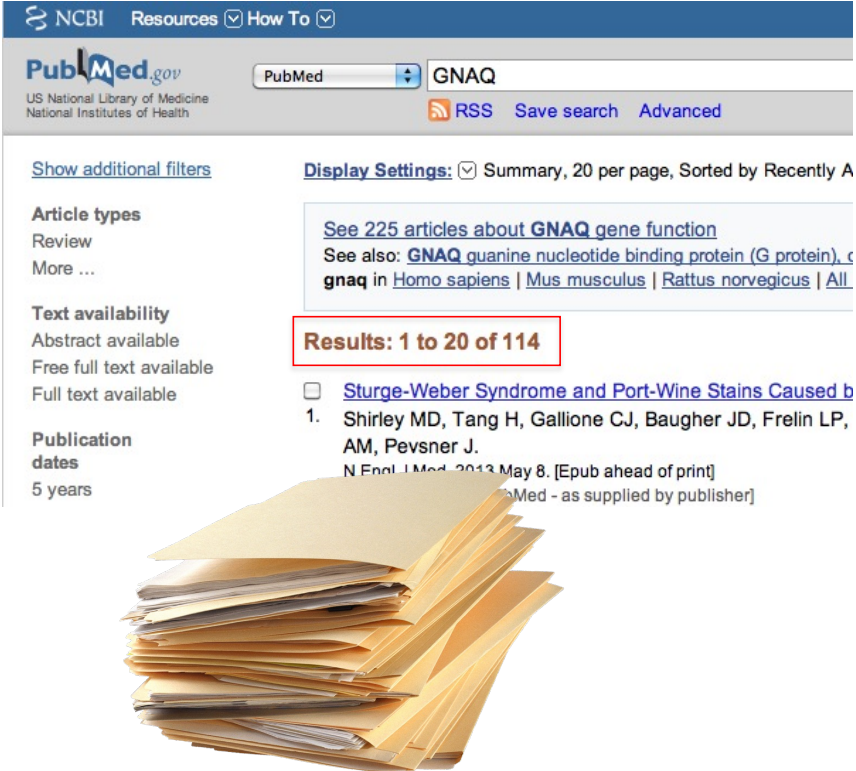
- My cool new screen worked and produced 1000 hits! ...Now what?
- Genome-Scale Analysis (Omics)
  - Genomics, Proteomics
- Tell me what's interesting about these genes
  - Are they enriched in known pathways, complexes, functions



# Pathway and network analysis

- Save time compared to traditional approach

GNAQ	→
GNAS	→
DGKZ	→
GUCY1A3	→
PDE4B	→
PDE4D	→
ATP2A2	→
ATP2A3	→
NOS1	→ my favorite gene
CNN1	→
GSTO1	→
NOS3	→
CNN2	→
MYLK2	→
CALD1	→
ACTA1	→
MYL2	→



NCBI Resources How To

PubMed.gov  
US National Library of Medicine  
National Institutes of Health

PubMed GNAQ

RSS Save search Advanced

Show additional filters

Display Settings: Summary, 20 per page, Sorted by Recently Added

Article types  
Review  
More ...


Text availability  
Abstract available  
Free full text available  
Full text available

Publication dates  
5 years

See 225 articles about **GNAQ** gene function  
See also: **GNAQ** guanine nucleotide binding protein (G protein) ...  
**gnaq** in [Homo sapiens](#) | [Mus musculus](#) | [Rattus norvegicus](#) | [All](#)

Results: 1 to 20 of 114

[Sturge-Weber Syndrome and Port-Wine Stains Caused by a Mutation in the GNAQ Gene](#)  
1. Shirley MD, Tang H, Gallione CJ, Baugher JD, Frelin LP, AM, Pevsner J.  
N Engl J Med. 2013 May 8. [Epub ahead of print]  
PubMed - as supplied by publisher



# Pathway and network analysis

- Helps gain mechanistic insight into 'omics data
  - Identifying a master regulator, drug targets, characterizing pathways active in a sample
- Any type of analysis that involves pathway or network information
- Most commonly applied to help interpret lists of genes
- Most popular type is pathway enrichment analysis, but many others are useful

## Autism spectrum disorder (ASD)

- Genetics
  - highly heritable
    - monozygotic twin concordance 60-90%
    - dizygotic twin concordance 0-10%  
(depending on the stringency of diagnosis)
  - known genetics:
    - 5-15% rare single-gene disorders and chromosomal re-arrangements
    - de-novo CNV previously reported in 5-10% of ASD cases
    - GWA (Genome-wide Association Studies) have been able to explain only a small amount of heritability

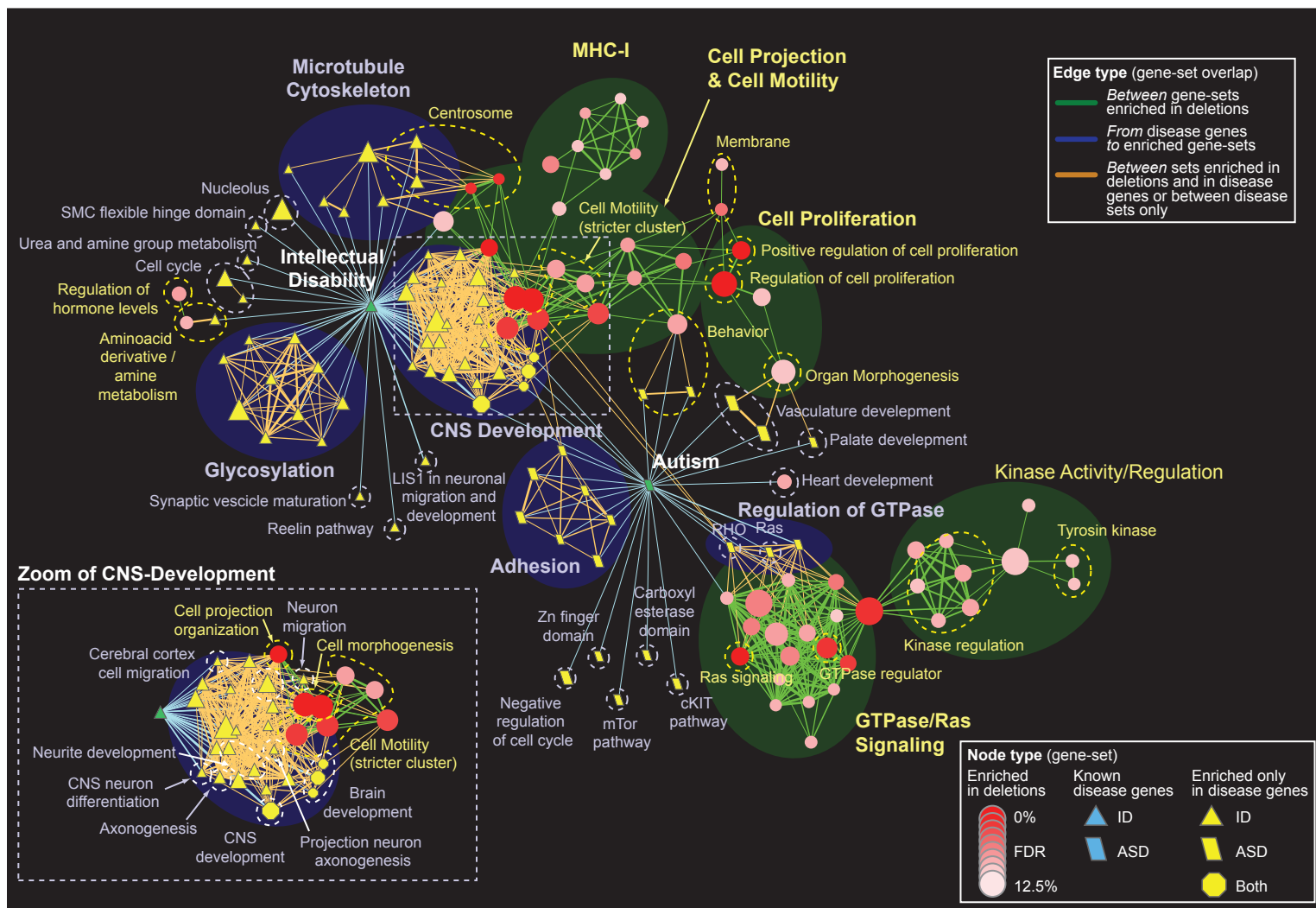
Pinto et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature. 2010 Jun 9.



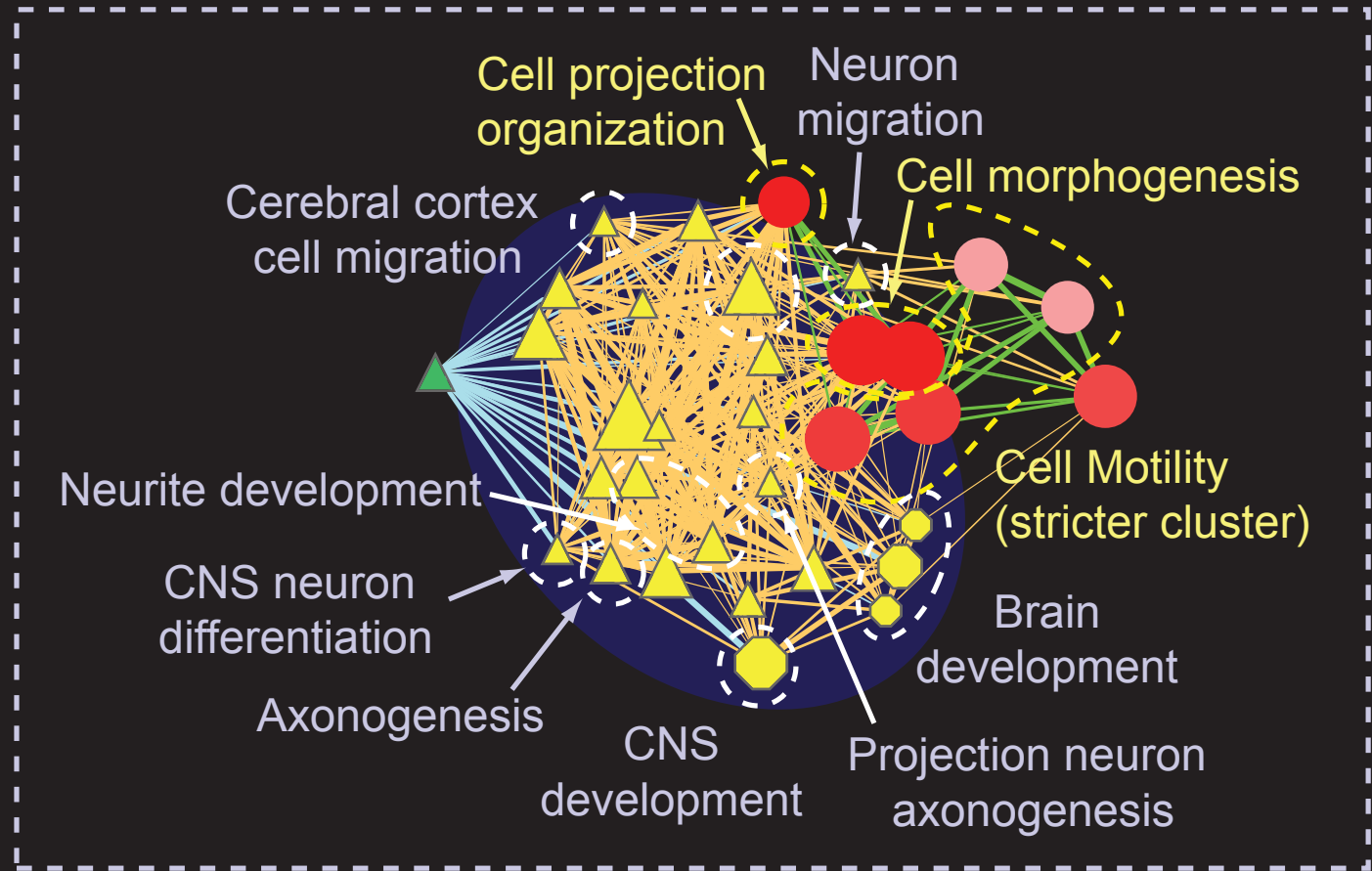
# Rare copy number variants in ASD

- Rare Copy Number Variation screening (Del, Dup)
  - 889 Case and 1146 Ctrl (European Ancestry)
  - Illumina Infinium 1M-single SNP
  - high quality rare CNV (90% PCR validation)
    - identification by three algorithms required for detection
      - QuantiSNP, iPattern, PennCNV
    - frequency < 1%, length > 30 kb
- Results
  - average CNV size: 182.7 kb, median CNVs per individual: 2
  - > 5.7% ASD individuals carry at least one de-novo CNV
  - Top ~10 genes in CNVs associated to ASD

# Pathways enriched in autism spectrum



# Zoom of CNS-Development



## Pathway analysis example 2

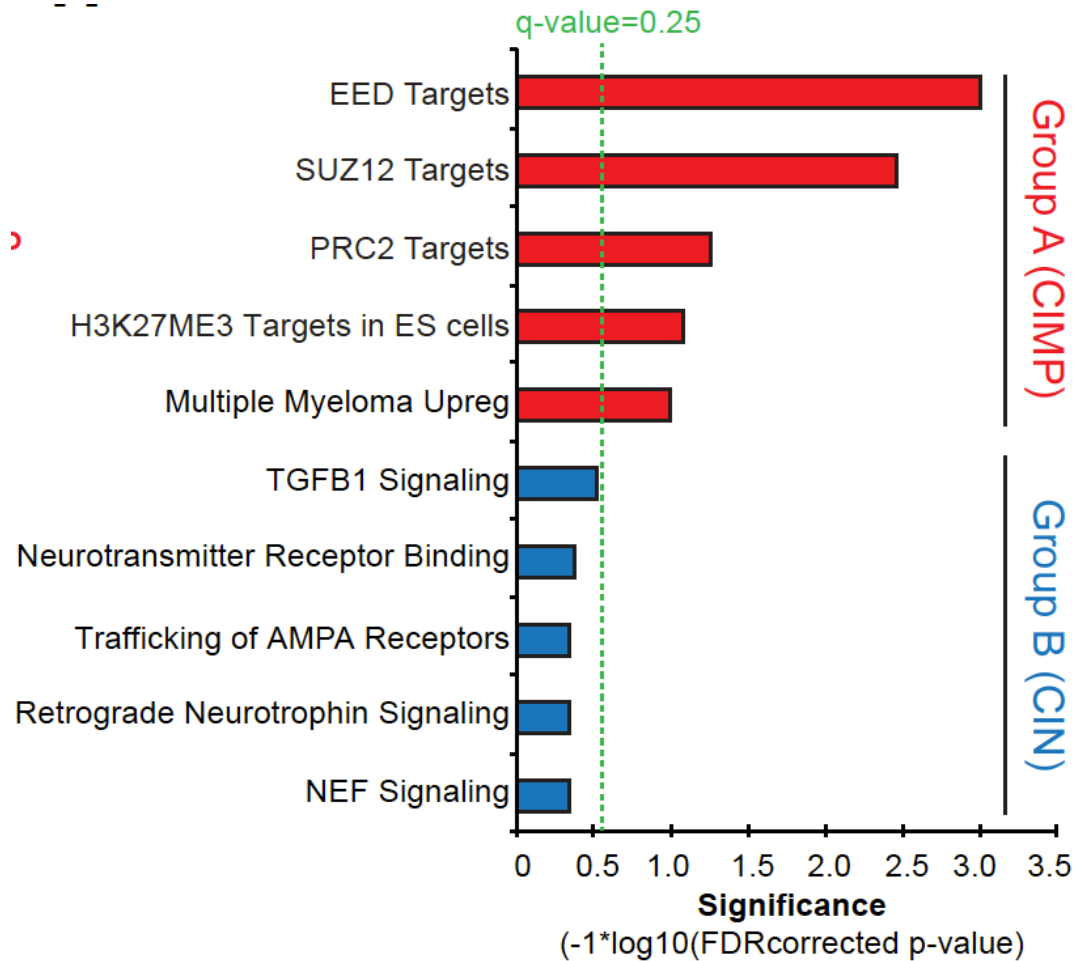
# Ependymoma pathway analysis

- Ependymoma brain cancer - most common and morbid location for childhood is the posterior fossa (PF = brainstem + cerebellum)
- Two classes: PFA - young, dismal prognosis, PFB - older, excellent prognosis. Determined by gene expression clustering.
- Exome sequencing (42 samples), WGS (5 samples) showed almost no mutations, however methylation arrays showed clear clustering into PFA and PFB (79 samples)
- PFA more transcriptionally silenced by CpG methylation

Witt et al., Cancer Cell 2011

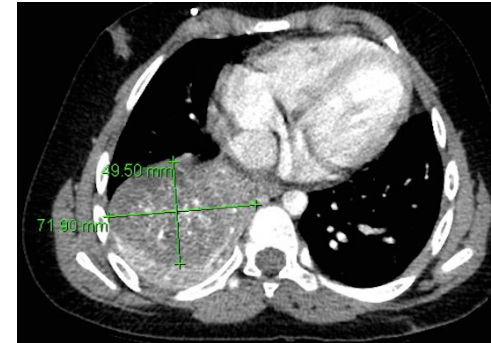
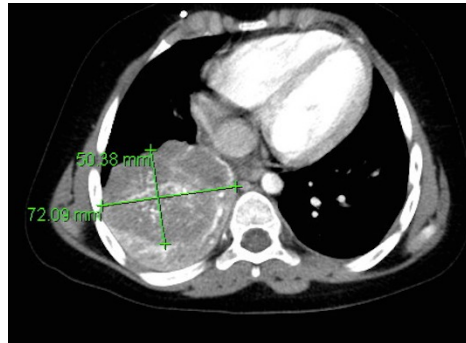
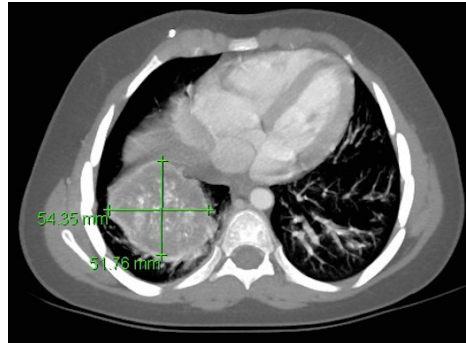
**Nature. 2014 Feb 27;506(7489):445-50**

**Steve Mack, Michael Taylor, Scott Zuyderduyn**



polycomb repressor complex 2 – inhibited by SAHA, DZNep, GSK343 – killed PFA cells  
No known treatment. Now in clinical trial

9 yo with metastatic PF ependymoma to lung  
treated with azacytidine



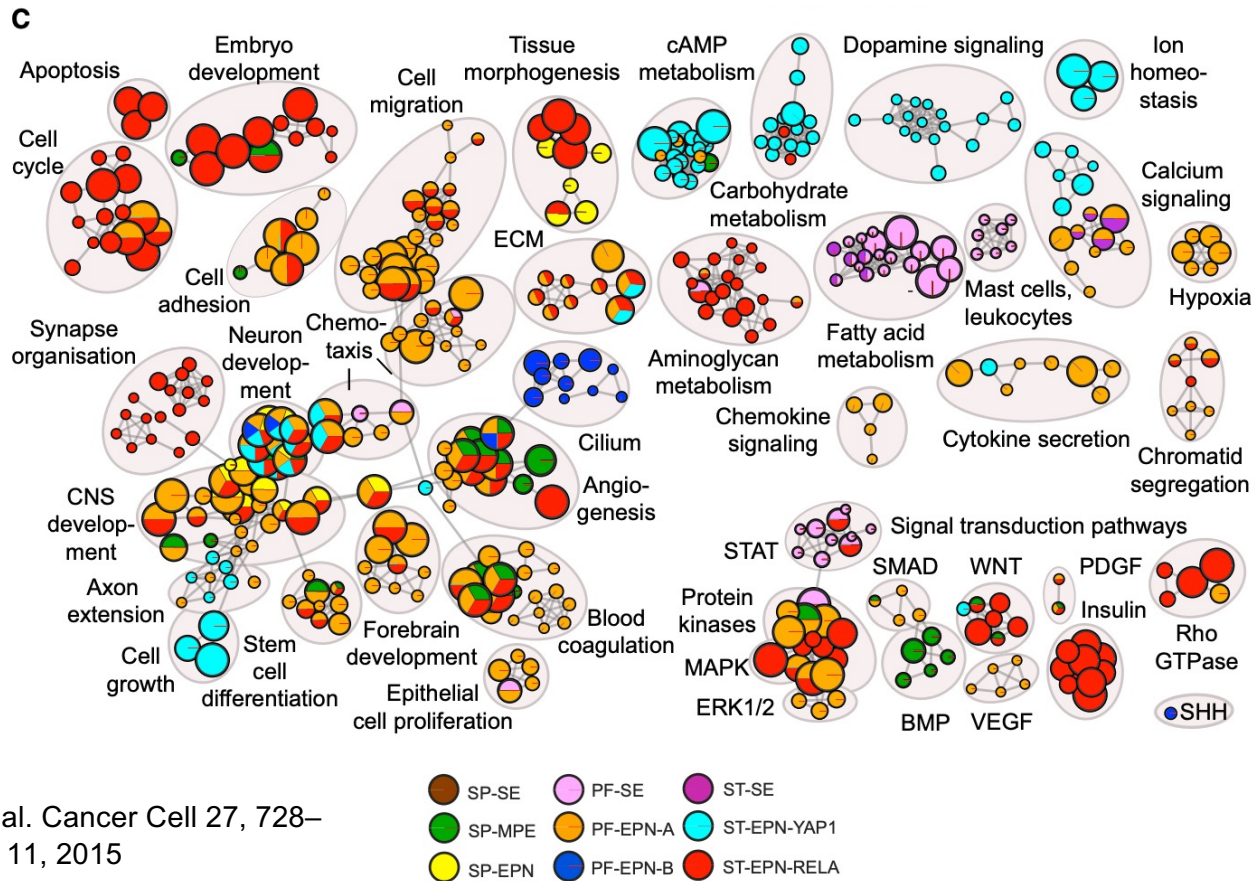
2  
months

3 months  
3 cycles  
Vidaza

Effect lasted 15 months

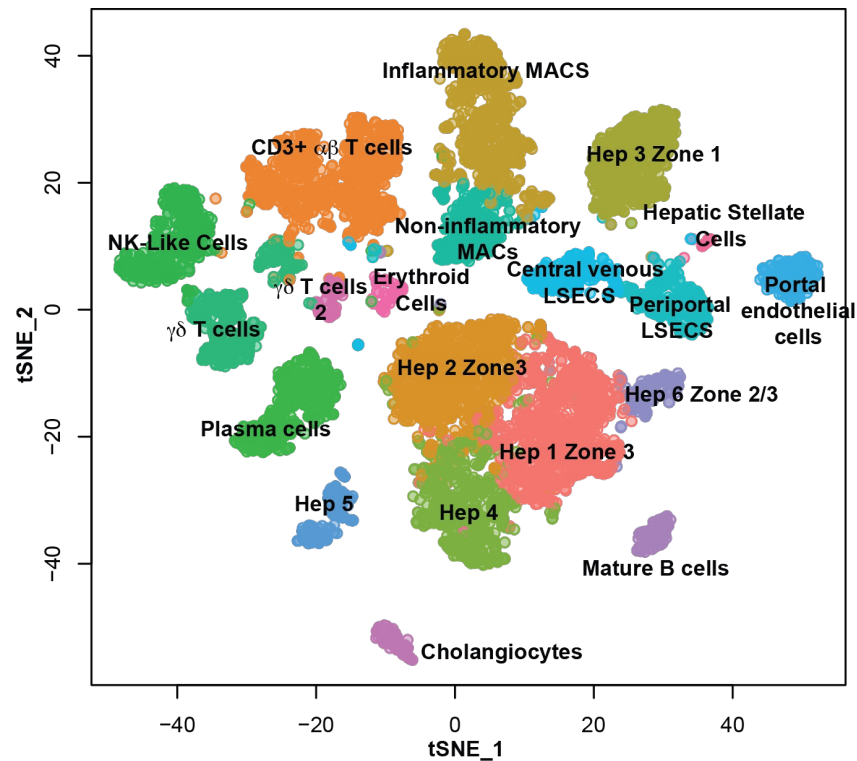
## Pathway analysis example 3

# Molecular classification of ependymal tumors

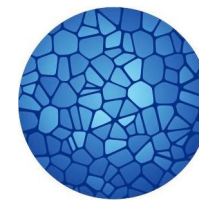


## Pathway analysis example 4

# ScRNA-seq of 5 healthy livers reveals 20 cell types



8444 single cells from five human liver samples



**HUMAN  
CELL  
ATLAS**

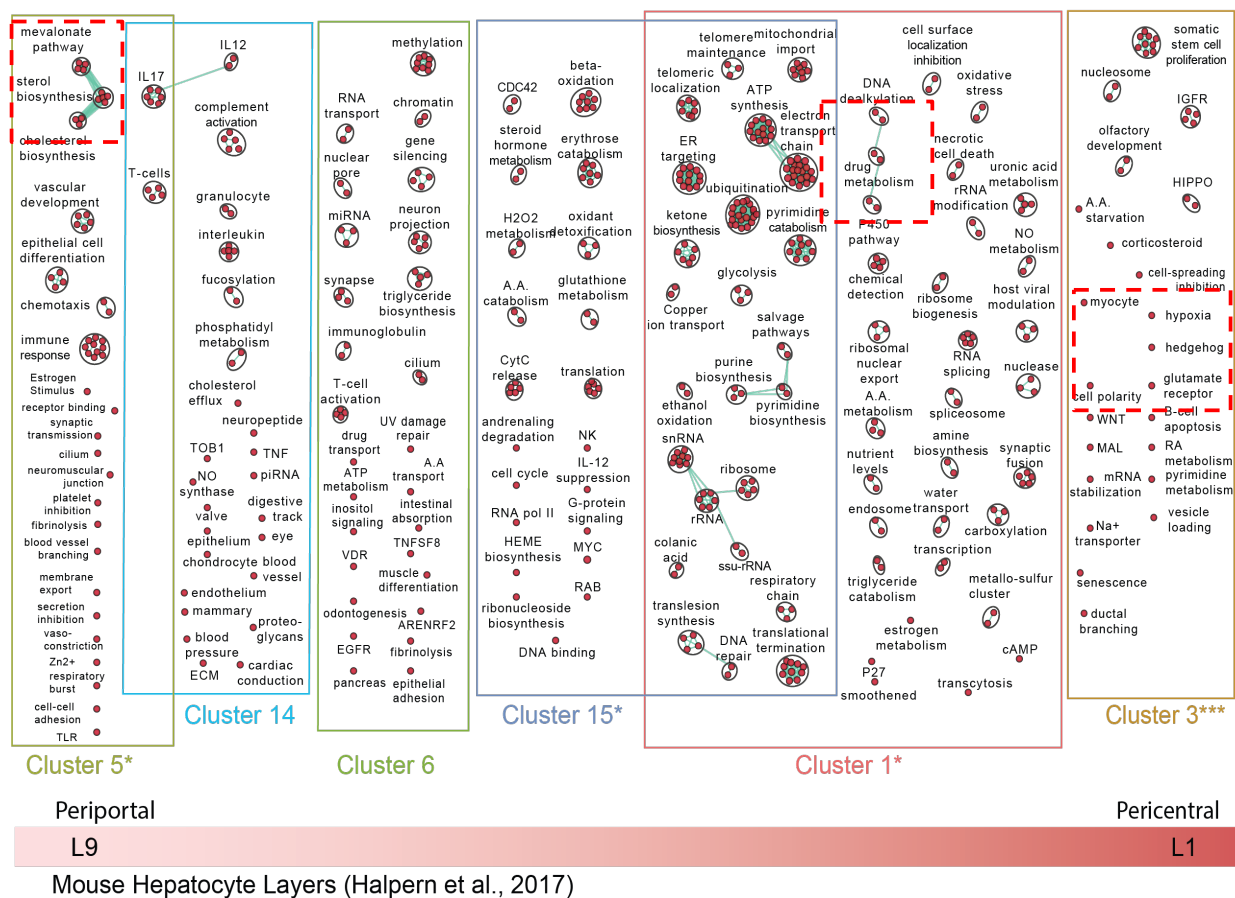
Public data, contributed to  
Human Cell Atlas  
(<https://www.humancellatlas.org>)

[shiny.baderlab.org/HumanLiverAtlas/](https://shiny.baderlab.org/HumanLiverAtlas/)

MacParland et al., Oct.22.2018,  
Nature Communications



# Pathway analysis identifies the division of labour among hepatocyte populations



# Example: Genome Wide Association Study (GWAS)

- Genotypes for 5 cases and 5 controls – ideal situation

	Cases					Controls				
SNP	1	2	3	4	5	6	7	8	9	10
A	1	1	1	1	1	0	0	0	0	0
B	0	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	1	1	1	1	1
E	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	0	0	0	0	0

Mutation A is perfectly associated with cases (5/5)

Mutation D is perfectly associated with controls (5/5)

# Example: Genome Wide Association Study (GWAS)

- More realistic situation:

	Cases					Controls				
SNP	1	2	3	4	5	6	7	8	9	10
A	1	0	0	0	0	0	0	0	0	0
B	0	1	0	0	0	0	0	0	0	0
C	0	0	1	0	0	0	0	0	0	0
D	0	0	0	1	0	0	0	0	0	0
E	0	0	0	0	1	0	0	0	0	0
F	0	0	0	0	0	1	0	0	0	0

No mutation is associated with cases or controls

# Example: Genome Wide Association Study (GWAS)

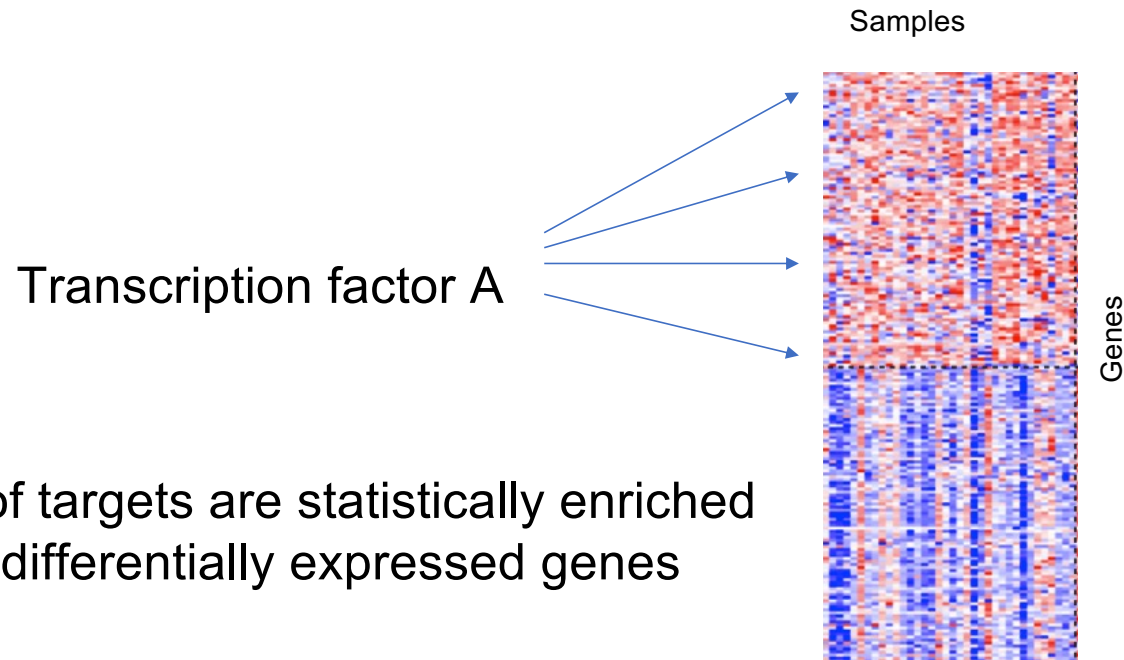
- Pathway analysis view: SNPs A-F are part of one pathway

	Cases					Controls				
Pathway	1	2	3	4	5	6	7	8	9	10
Apoptosis	1	1	1	1	1	0	0	0	0	0

- Mutations in 'apoptosis' pathway are perfectly associated with cases (5/5)
- Increased statistical power via **aggregating counts** and **reducing multiple testing**
- Generating mechanistic hypotheses (i.e. apoptosis is related to the case phenotype)

# Example: Master regulator analysis

- What do 1000 differentially expressed genes have in common?

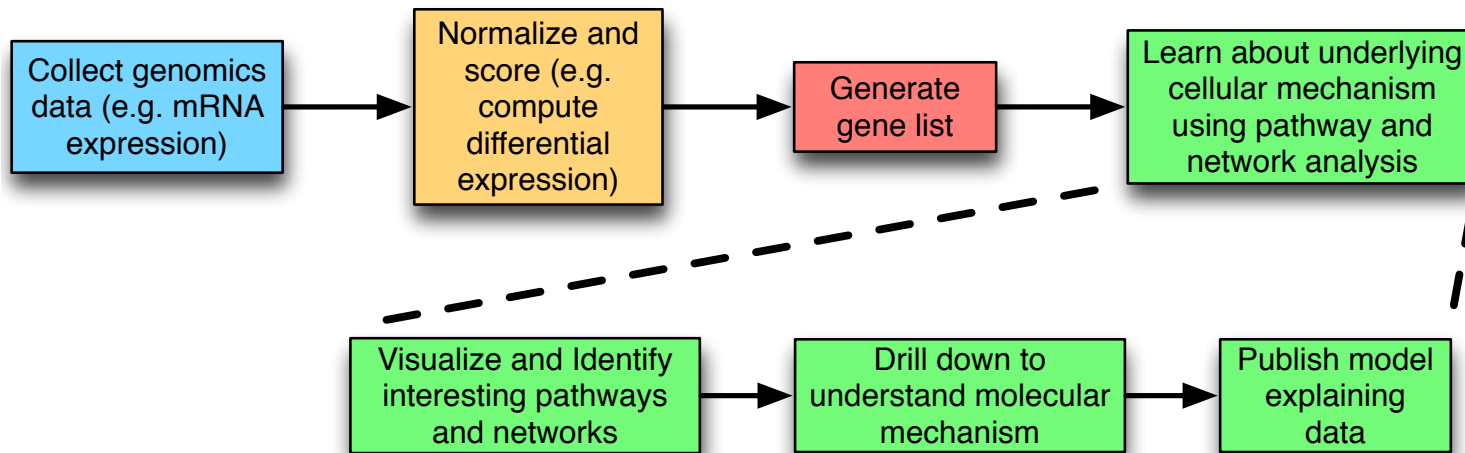


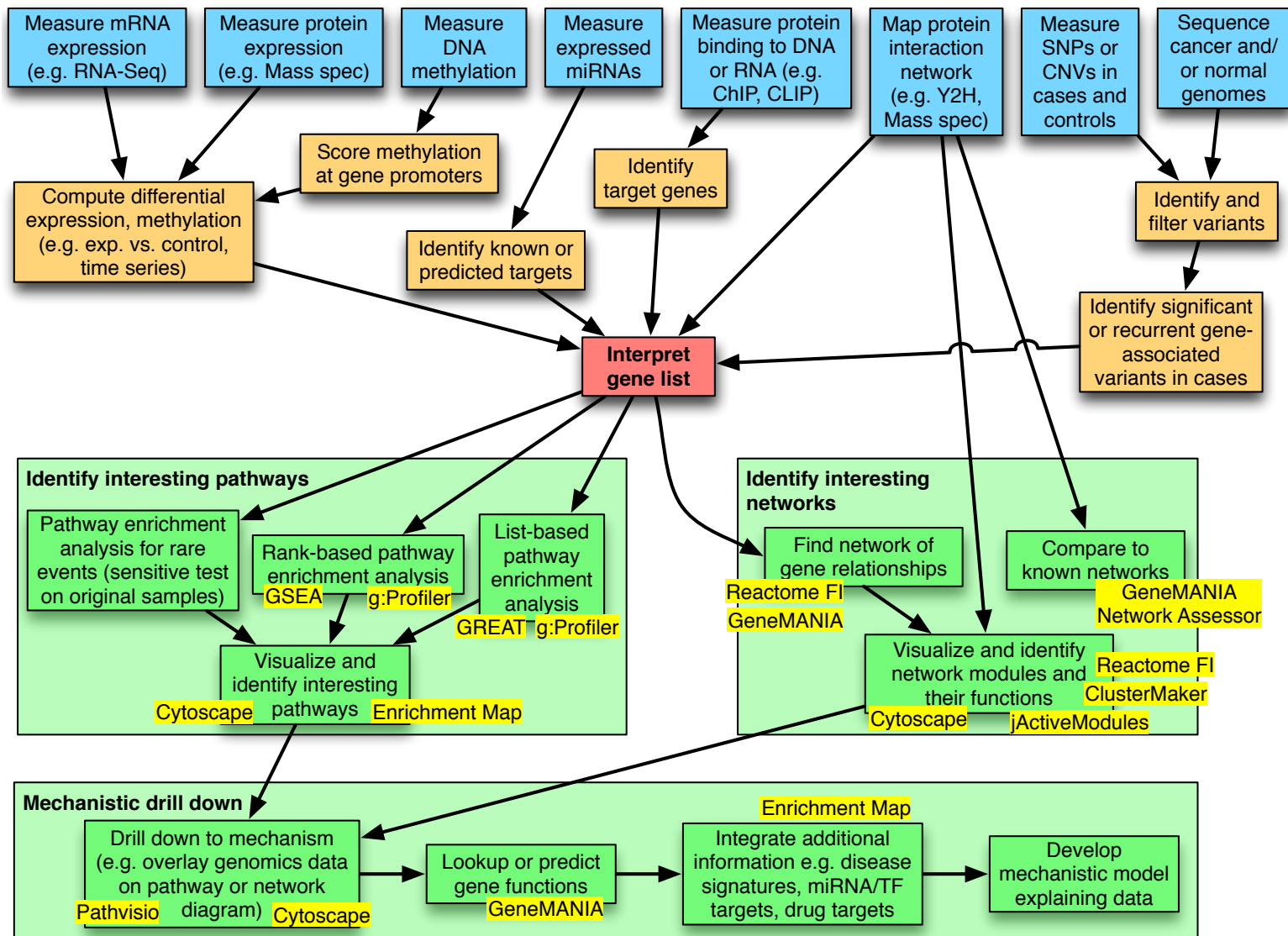
# Benefits of pathway analysis

vs. transcripts, proteins, SNPs...

- Easier to interpret
  - Familiar concepts e.g. cell cycle
- Identifies possible causal mechanisms
- Predicts new roles for genes
- Improves statistical power
  - Fewer tests, aggregates data from multiple genes into one pathway
- More reproducible
  - E.g. gene expression signatures
- Facilitates integration of multiple data types

# Pathway analysis workflow overview







# Workshop outline

- Computational analysis methods we will cover
  - Pathway enrichment analysis: summarize and compare
  - Network analysis: predict gene function, find new pathway members, identify functional modules (new pathways)
  - Regulatory network analysis: find and analyze controllers

# We are on a Coffee Break & Networking Session

Workshop Sponsors:

