

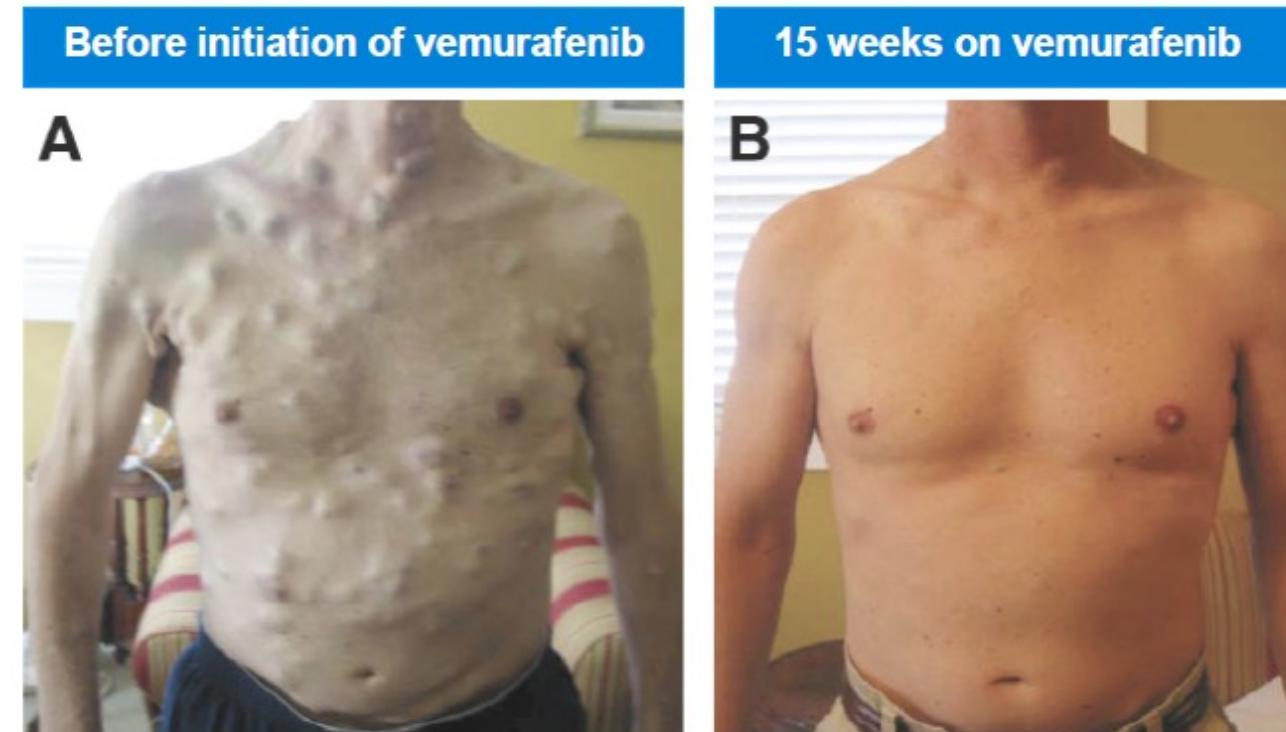


## AI IN CLINICAL CARE – WORKSHOP

Genomics for precision medicine

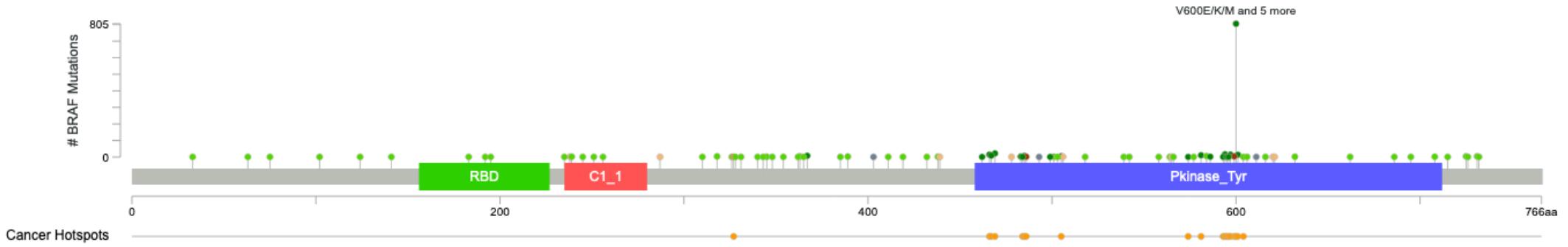
KILEY GRAIM  
February 2023





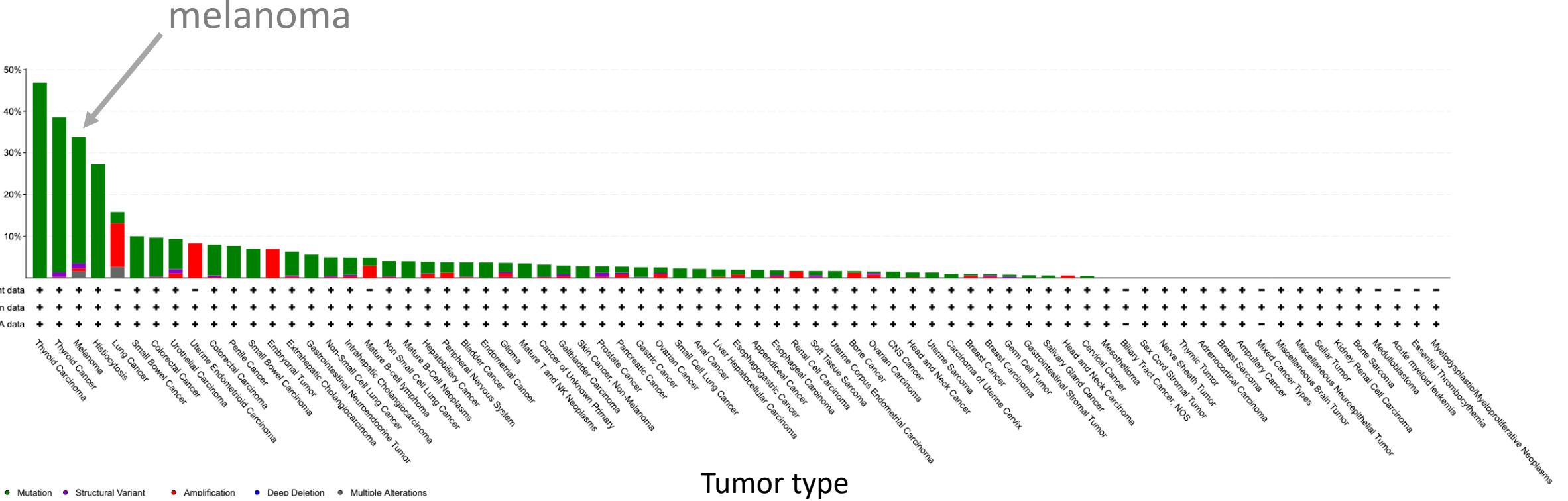
# (Genomic) Precision Medicine

In cancers driven by a single mutation, targeted drugs can give spectacular results



~50% of melanoma patients have a BRAF mutation

Alteration Frequency



melanoma

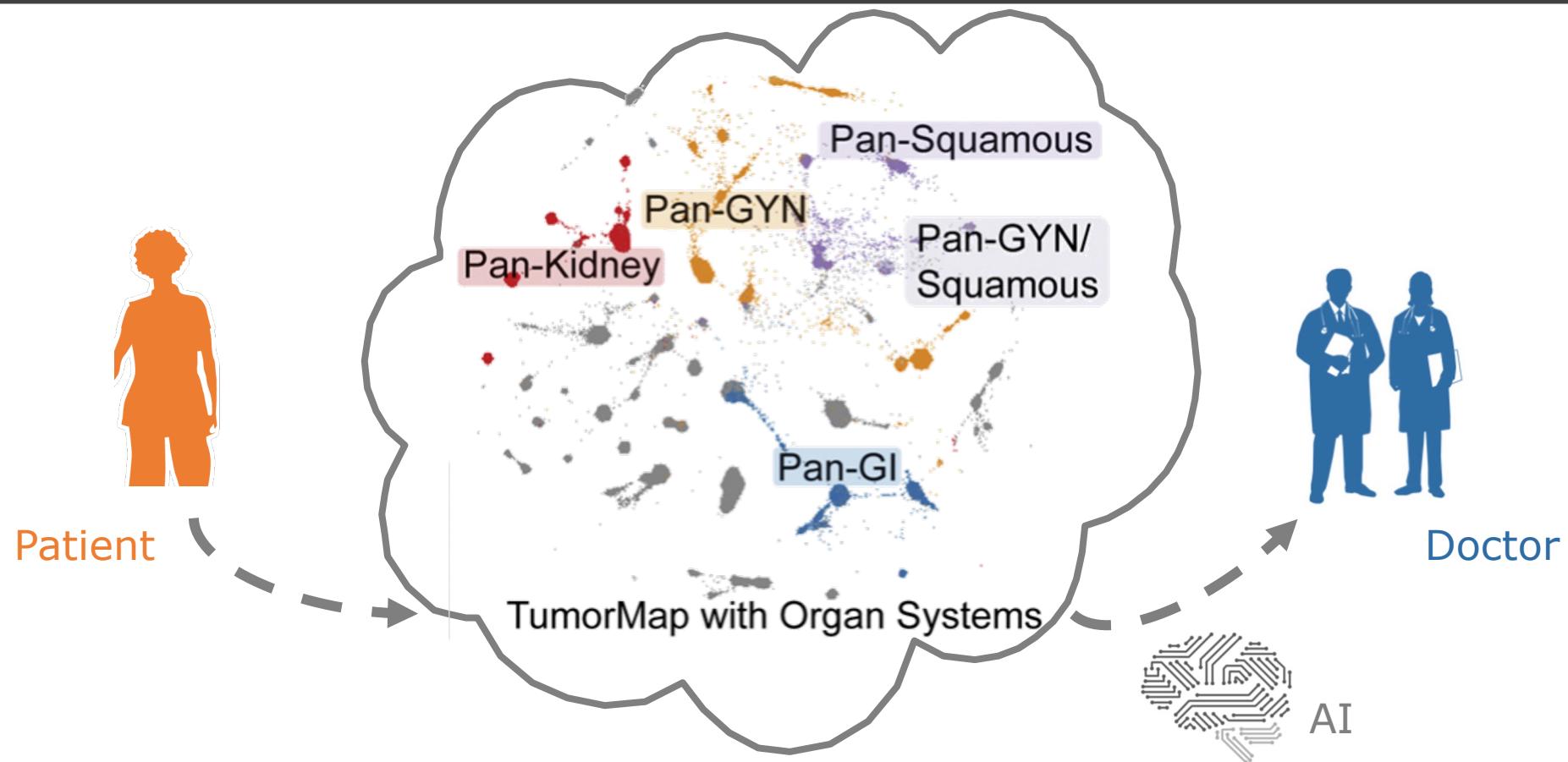
Tumor type

... but so do many other cancer patients

# Genomes are the key to the future of disease treatment



# Multi-cancer groups have clinical utility





There is a new  
data-intensive,  
molecular  
approach to  
disease

---

The challenge is now to interpret disease in the context of other molecular changes within the cell, and physiological changes in the organism. We need to better understand how genomes work.

---

Genomes give us an opportunity for basic science, and this creates a synergistic opportunity for medicine.

---

Big data will be needed for statistical power. We are building million+ genome databases.

---

We all carry genetic variations we inherited from our parents and *de novo* mutations that occur just in our genome.

# (Some)Precision medicine initiatives.

Country	Project Name	Expected Size	Common Diseases	Rare Diseases
USA	All of Us	> 1 million	✓	
USA	NHGRI Genomic Medicine	nationwide	✓	✓
EU	1+ Million Genomes Initiative	>1 million	✓	
Australia	Genomics Health Futures Mission	200,000		✓
Japan	Genome Medical alliance Japan	Nationwide	✓	✓
Saudi Arabia	Saudi Human Genom Program	100,000	✓	✓
... many more				

# What do we do?

- Identify the data
- Download the data, combine, harmonize
- Differential Expression
- Signature detection
- Enrichment

# Where is the data?

## Search for normalized transcriptome data

Search

Try searching for: [Notch](#) [Medulloblastoma](#) [GSE24528](#)



### Find the data you need

Search the multi-organism collection of genome wide gene expression data obtained from publicly available sources like GEO, ArrayExpress, and SRA. The data has been processed uniformly and normalized using a set of [standardized pipelines](#) curated by the [Childhood Cancer Data Lab \(CCDL\)](#).



### Create custom datasets

Build and download custom datasets tailored to your needs including gene expression matrices and sample metadata.

[Query](#)
[Quick Search Beta](#)
[Download](#)

### Select Studies for Visualization & Analysis:

0 studies selected (0 samples)

#### PanCancer Studies 10

#### Pediatric Cancer Studies 13

#### Immunogenomic Studies 8

#### Cell lines 3

#### Adrenal Gland 2

#### Ampulla of Vater 1

#### Biliary Tract 15

#### Bladder/Urinary Tract 19

#### Bone 2

#### Bowel 17

#### Breast 25

#### CNS/Brain 24

#### Cervix 2

#### Esophagus/Stomach 18

#### Eye 5

#### Head and Neck 15

#### Kidney 17

#### Liver 10

#### Lung 33

#### Lymphoid 21

#### Myeloid 12

#### Other 26

#### Ovary/Fallopian Tube 7

#### Pancreas 11

Quick select: [TCGA PanCancer Atlas Studies](#) Curated set of non-redundant studies

Looking for AACR Project GENIE, the largest public clinicogenomic cancer dataset? [It's available here.](#) ⓘ

#### PanCancer Studies

- MSK-IMPACT Clinical Sequencing Cohort (MSK, Nat Med 2017)
- Metastatic Solid Cancers (UMich, Nature 2017)
- MSS Mixed Solid Tumors (Broad/Dana-Farber, Nat Genet 2018)
- SUMMIT - Neratinib Basket Study (Multi-Institute, Nature 2018)
- TMB and Immunotherapy (MSK, Nat Genet 2019)
- Tumors with TRK fusions (MSK, Clin Cancer Res 2020)
- Cancer Therapy and Clonal Hematopoiesis (MSK, Nat Genet 2020)
- China Pan-cancer (OrigiMed, Nature 2022)
- Pan-cancer analysis of whole genomes (ICGC/TCGA, Nature 2020)
- MSK MetTropism (MSK, Cell 2021)

#### Pediatric Cancer Studies

- Pediatric Preclinical Testing Consortium (CHOP, Cell Rep 2019)
- Pediatric Acute Lymphoid Leukemia - Phase II (TARGET, 2018)
- Pediatric Rhabdoid Tumor (TARGET, 2018)
- Pediatric Wilms' Tumor (TARGET, 2018)
- Pediatric Acute Myeloid Leukemia (TARGET, 2018)
- Pediatric Neuroblastoma (TARGET, 2018)
- Pediatric Pan-Cancer (DKFZ, Nature 2017)
- Pediatric Pan-cancer (Columbia U, Genome Med 2016)
- Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2016)
- Acute Lymphoblastic Leukemia (St Jude, Nat Genet 2015)
- Pediatric Ewing Sarcoma (DFCI, Cancer Discov 2014)
- Ewing Sarcoma (Institut Curie, Cancer Discov 2014)
- Medulloblastoma (PCGP, Nature 2012)

#### Immunogenomic Studies

- Glioblastoma (Columbia, Nat Med. 2019)
- Metastatic Melanoma (DFCI, Science 2015)
- Melanoma (MSK, NEJM 2014)
- Metastatic Melanoma (UCLA, Cell 2016)
- Non-Small Cell Lung Cancer (MSK, Cancer Cell 2018)
- Non-Small Cell Lung Cancer (MSK, Science 2015)
- TMB and Immunotherapy (MSK, Nat Genet 2019)
- Clear Cell Renal Cell Carcinoma (DFCI, Science 2019)

#### Cell lines

0 studies selected (0 samples)

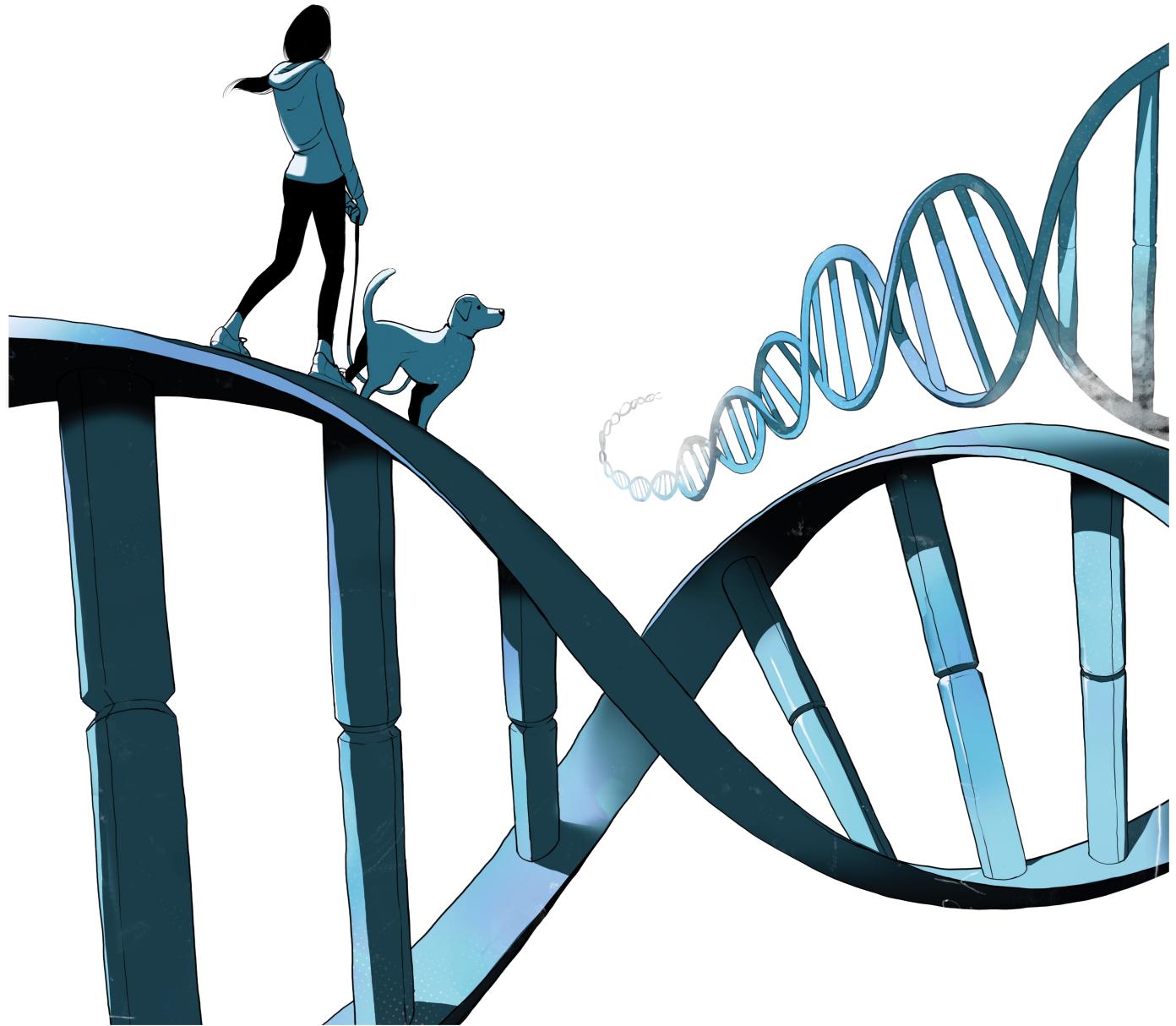
Query By Gene

OR

Explore Selected Studies

# Where is the data?

# Integrative Analysis of 2 Species





# Modeling the Molecular Development of Breast Cancer

## Using canine mammary tumors as a natural model

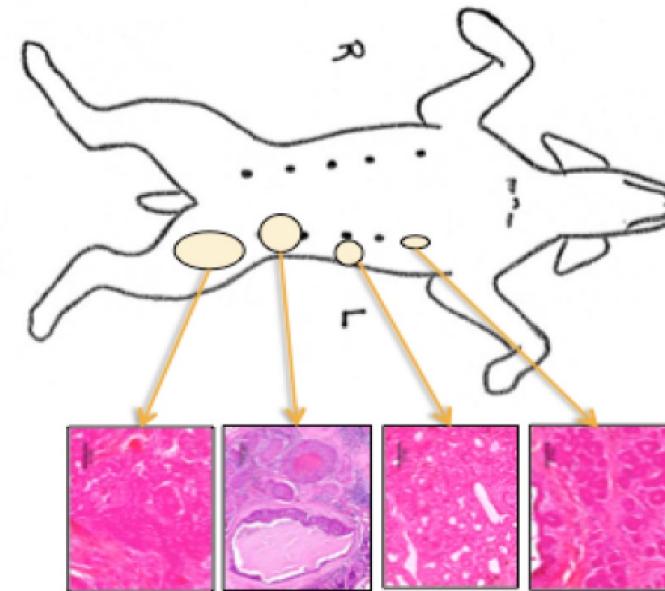
“As an intermediate, the dogs provide an intriguing option to traditional models or human studies. **[Dogs] develop disease naturally** through inheritance and interactions with our environment, yet **their breed structure creates independent and unique strains** in which heterozygosity is reduced and mapping is less complex” – Parker *et al*, 2015



# Penn Vet Shelter Canine Mammary Tumor program

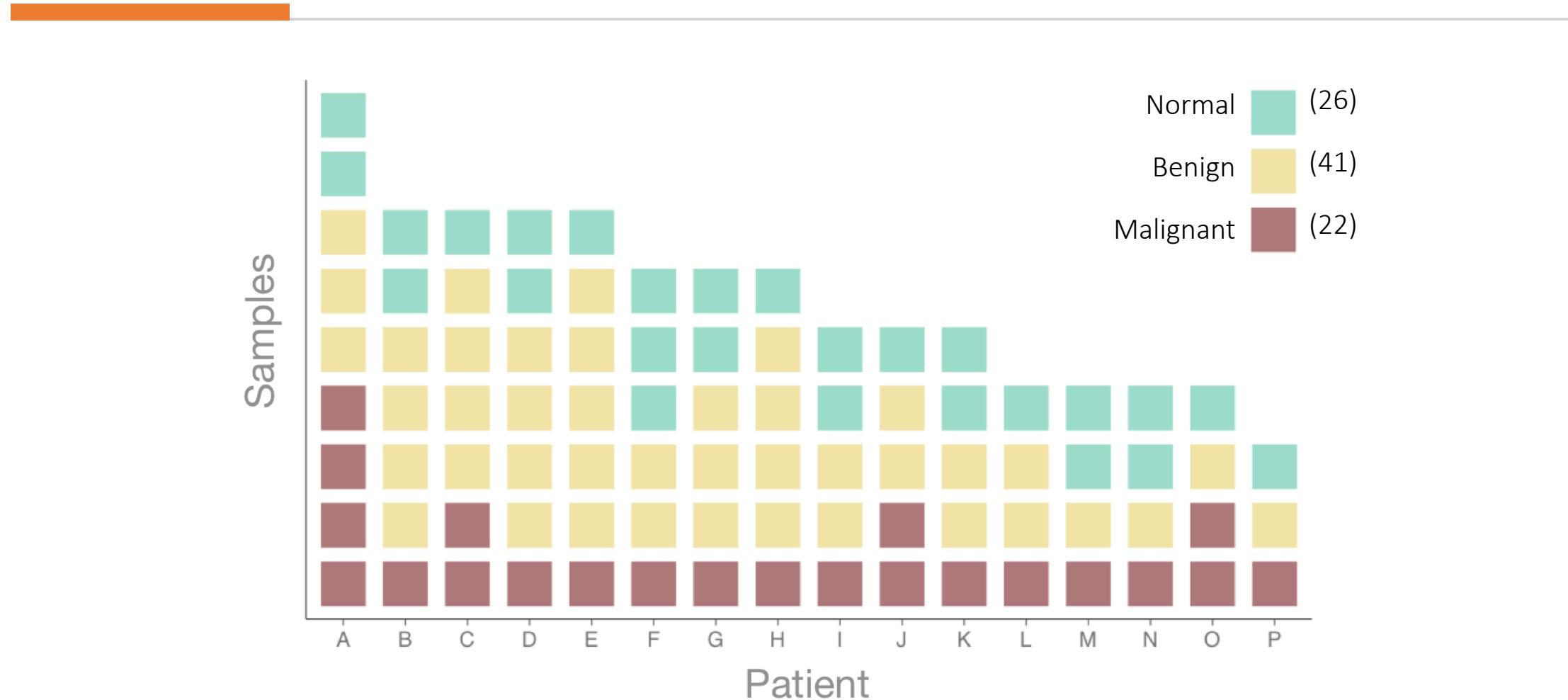
---

- Naturally occurring canine mammary tumors (CMTs)
- 2 parallel sections collected from each tumor
  - 1 sent for H&E staining
  - Other flash frozen
- 89 samples
  - 22 malignant
  - 41 benign
  - 26 normal

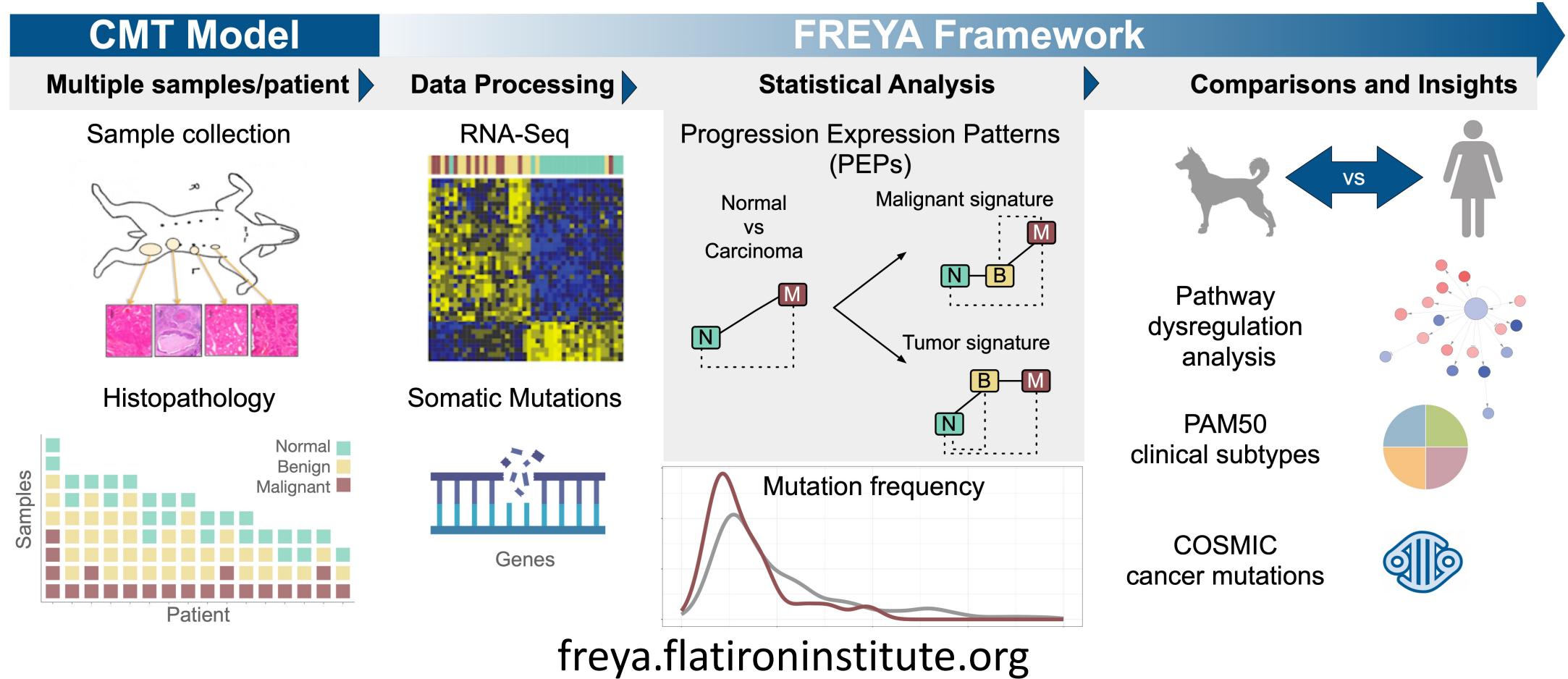


with Dr. Karin Sorenmo

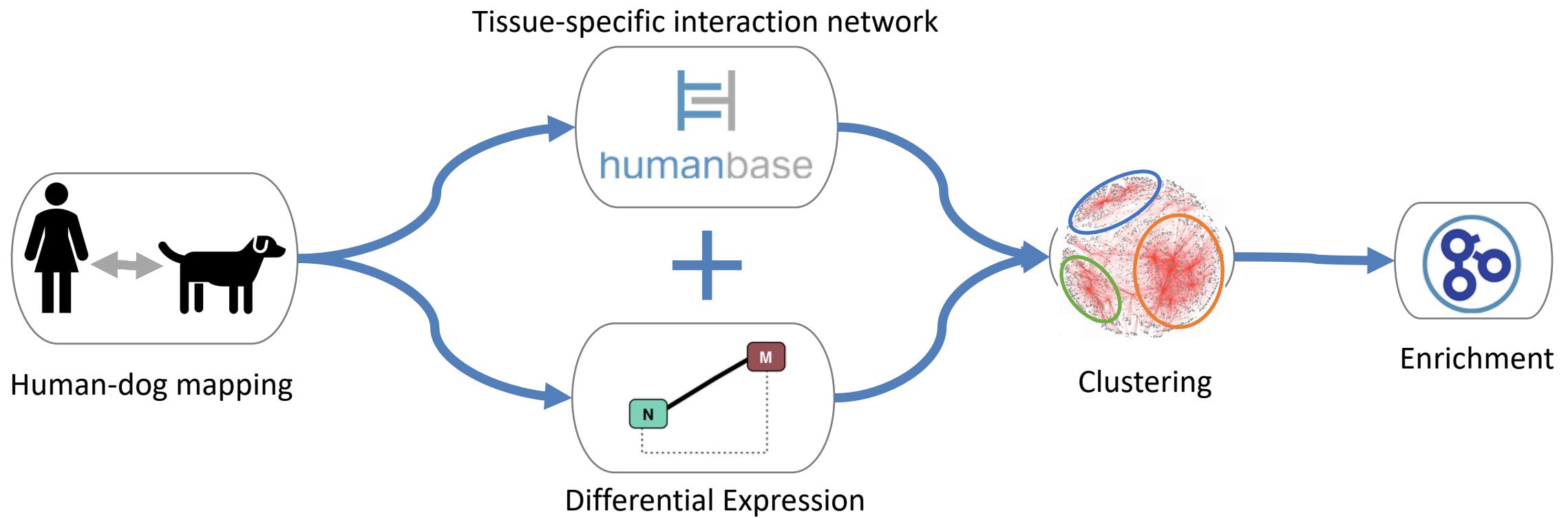
# Multiple Samples per Dog



# FREYA: FRamework for Expression analYsis Across species

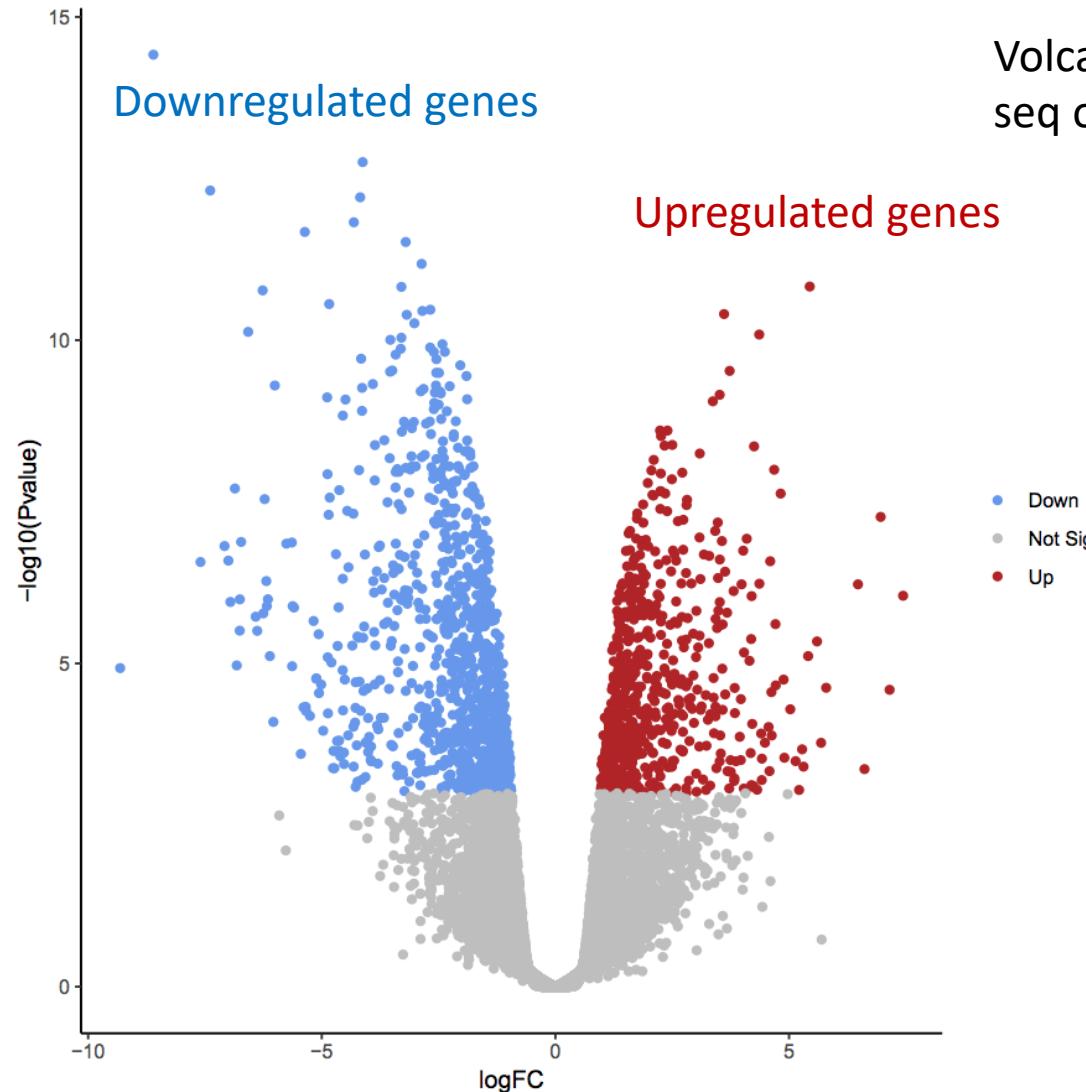


# Dog vs Human Breast Cancer



What is differential expression? Why do it?

# What is differential expression?

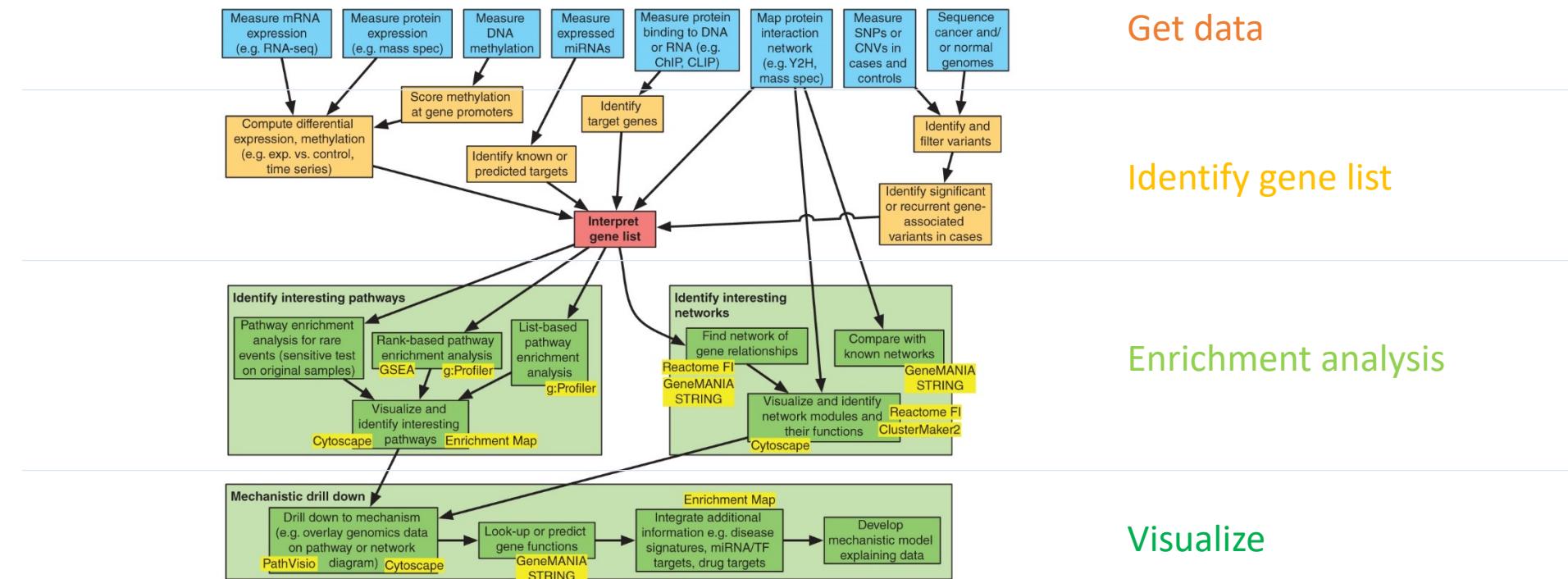


Volcano plots are commonly used to display the results of RNA-seq or other omics experiments

shows statistical significance (P value) versus magnitude of change (fold change)

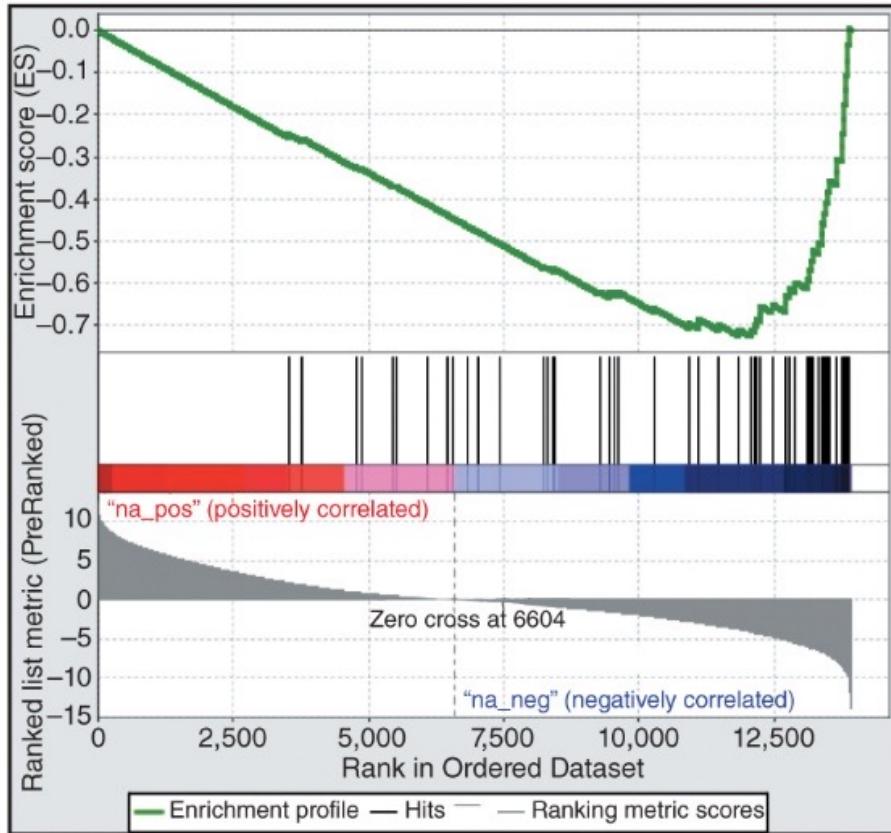
genes are colored if they pass the thresholds for FDR and Log Fold Change

# Network/Pathway analysis process

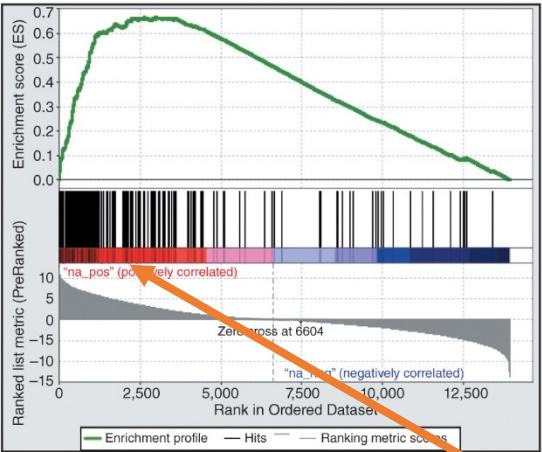


**Figure 13.19** An overview of a pathway analysis workflow, summarizing multiple tools in this chapter. The top layer in blue depicts different genomics data types. The next layer in light orange shows data processing steps required to derive a gene list from the data. The gene list is represented by a red box. The green boxes describe data analysis and interpretation steps, with pathway enrichment analysis on the left and network analysis on the right. Both parallel approaches lead to focused analysis of pathways, network regions, and genes of interest (bottom green section). Yellow rectangles highlight tools discussed in this chapter. Arrows connect boxes to show paths through the overall workflow. ChIP, chromatin immunoprecipitation; CLIP, cross-linking immunoprecipitation; CNV, copy number variant; exp., experiment; GSEA, Gene Set Enrichment Analysis; mass spec., mass spectrometry; miRNA, mitochondrial RNA; SNP, single nucleotide polymorphism; TF, transcription factor.

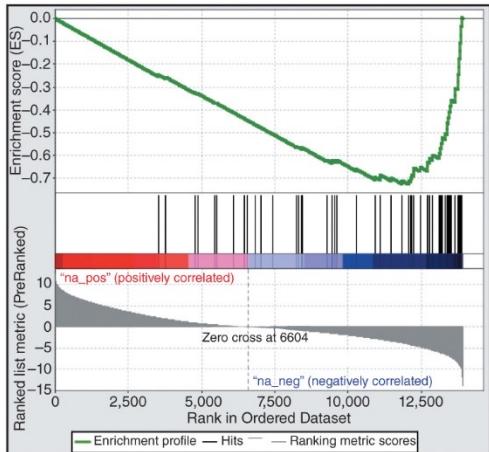
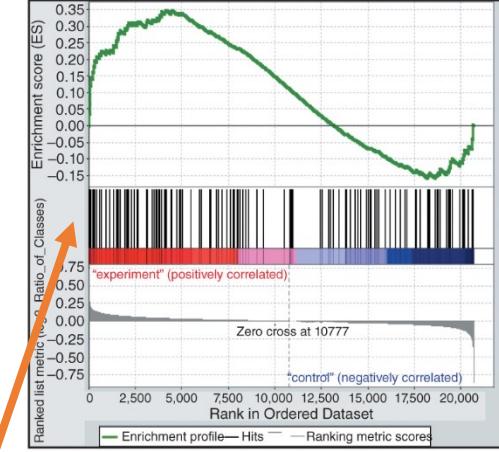
# Gene Set Enrichment Analysis (GSEA)



# Good

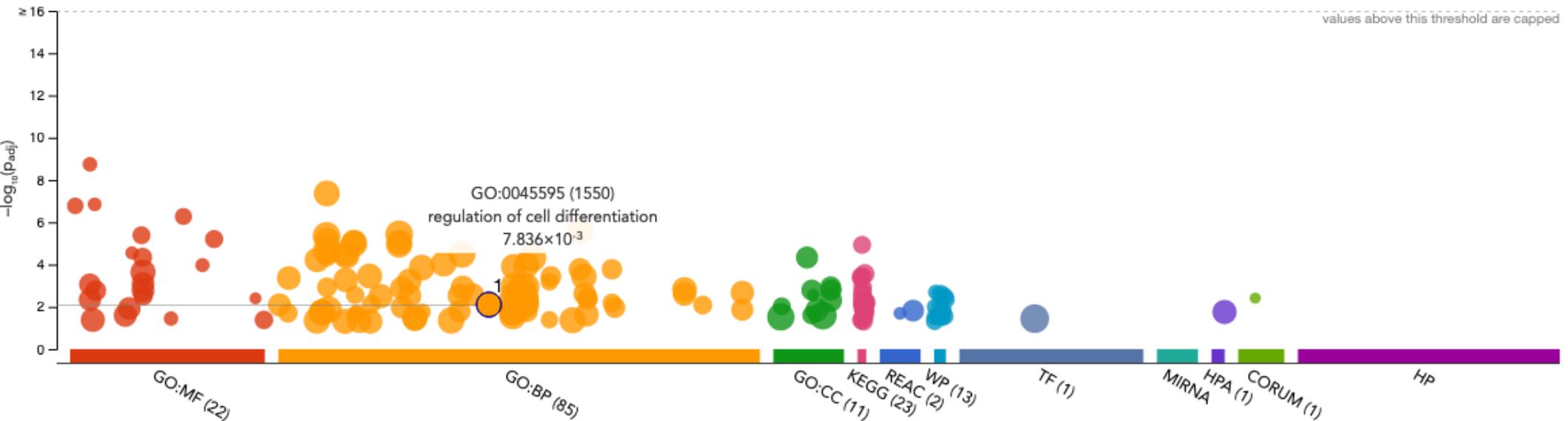


# Poor

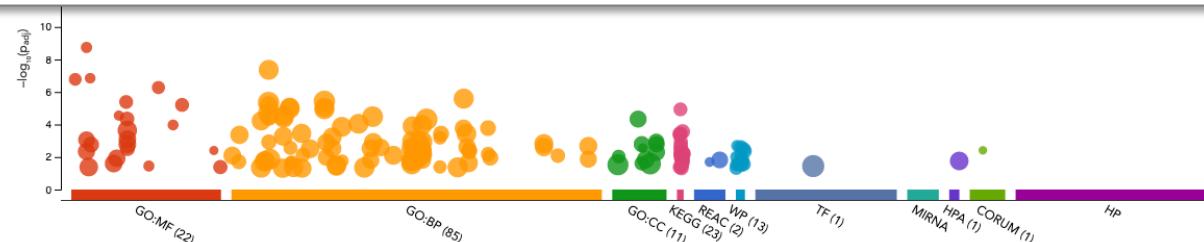


Notice the gene spread

g:Profiler has been updated with new data from Ensembl.

 Capped ? Make unselected terms transparent

ID	Source	Term ID	Term Name	padj (query_1)
1	GO:BP	GO:0045595	regulation of cell differentiation	$7.836 \times 10^{-3}$



g:Profiler

version  
date  
organisme106\_eg53\_p16\_65fc97  
9/26/2022, 3:44:09 PM  
hsapiens

g:Profiler

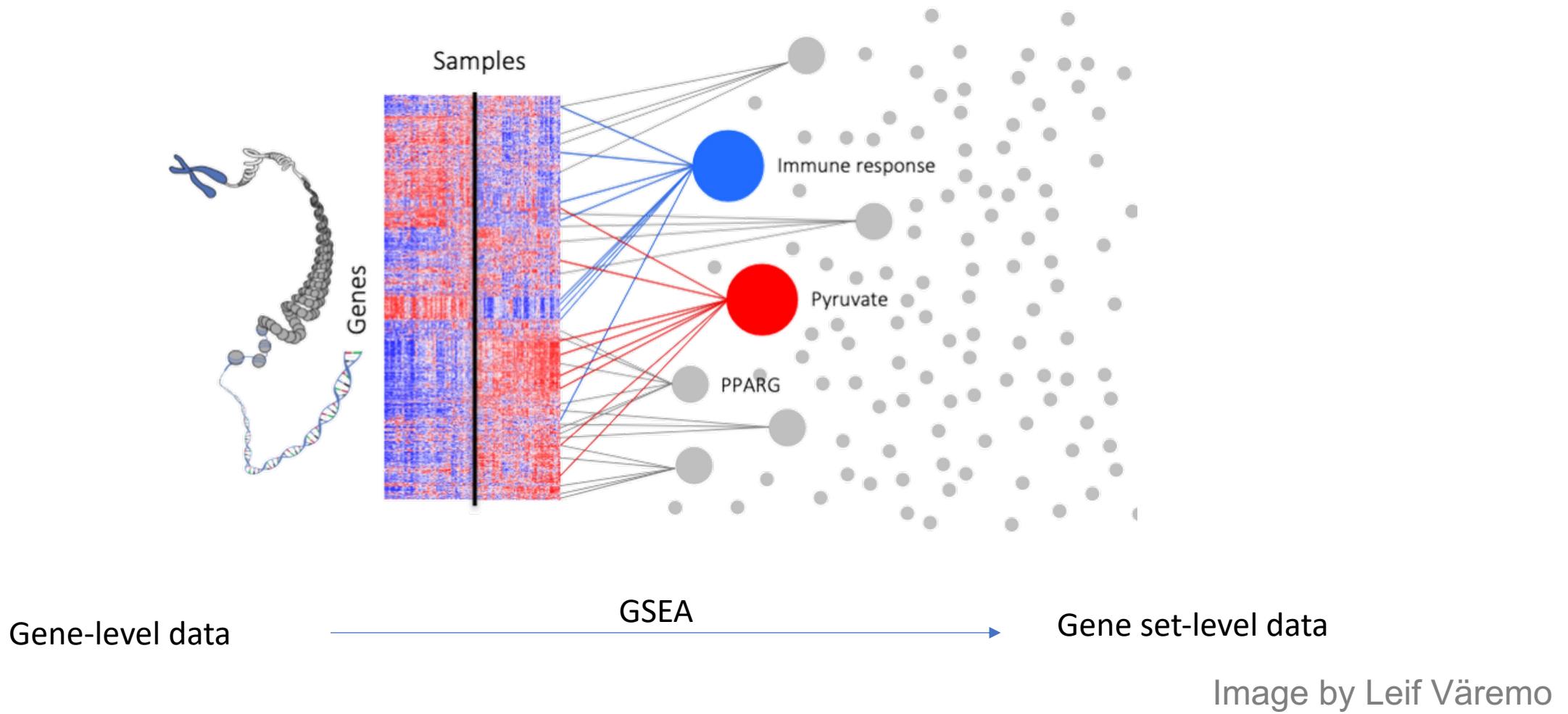
# Now what?

Figure out the functions of the differentially expressed genes



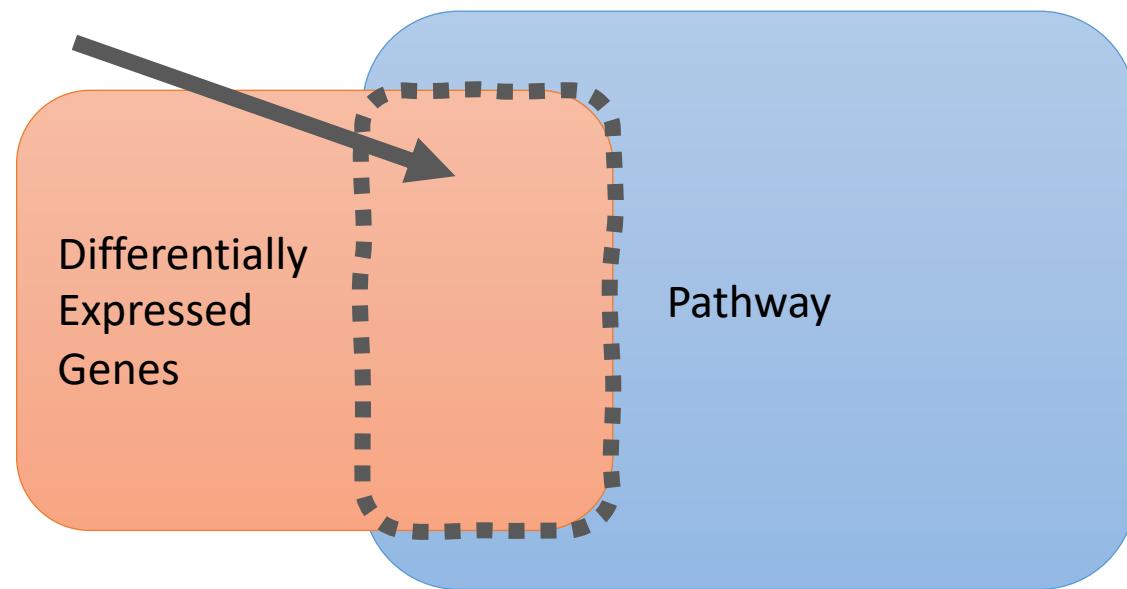
Or what diseases,etc  
they're associated with

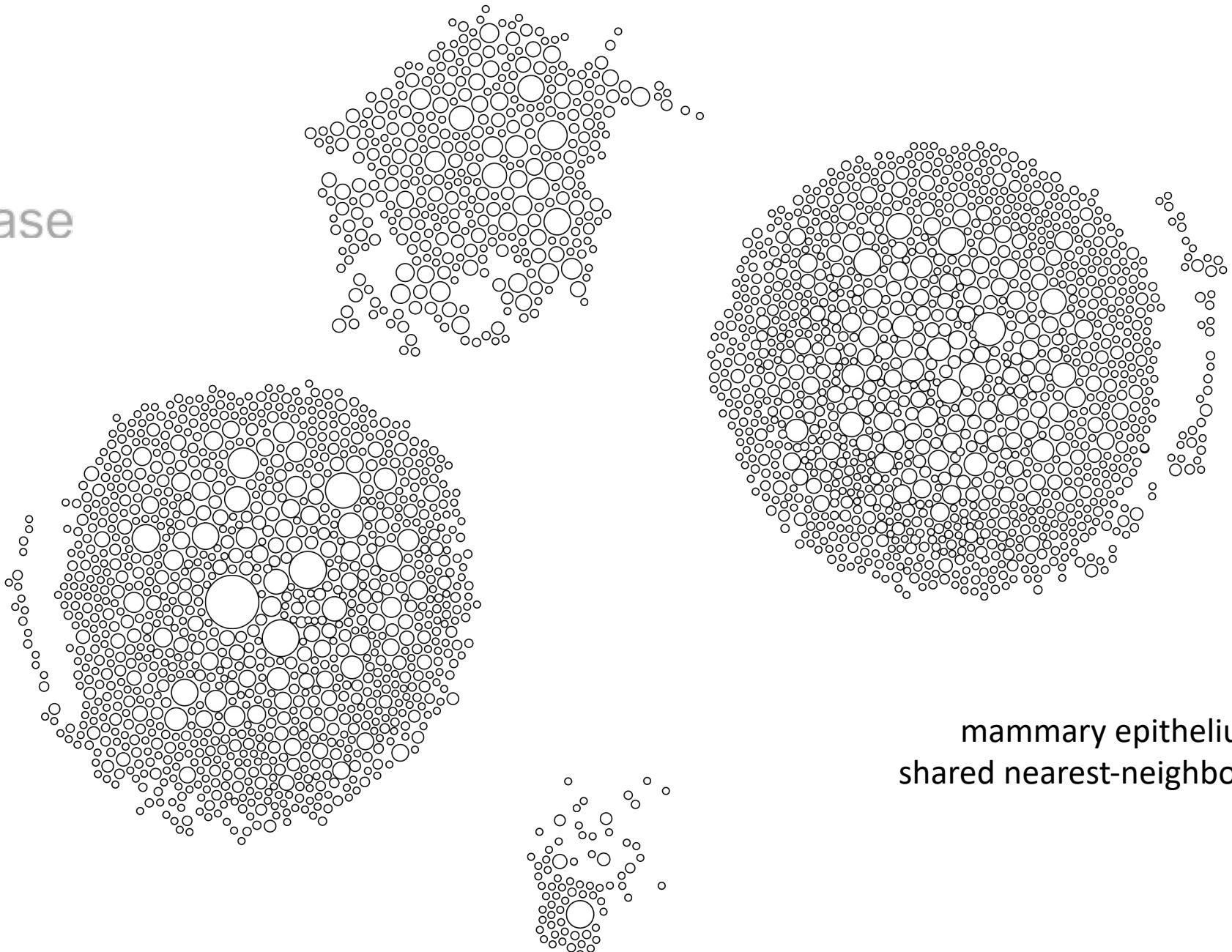
# Enrichment Analysis!



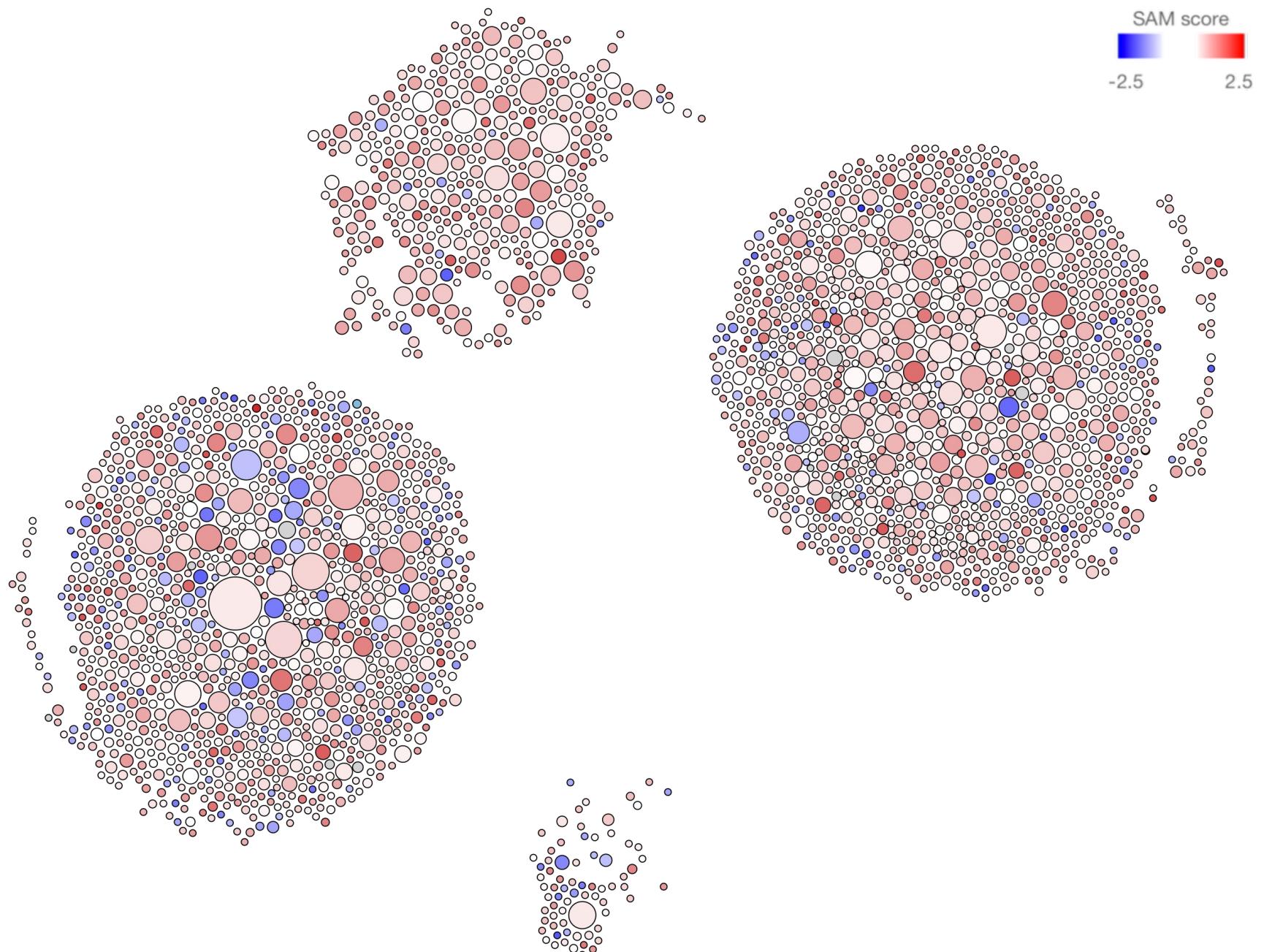
# Example: Fisher's exact test

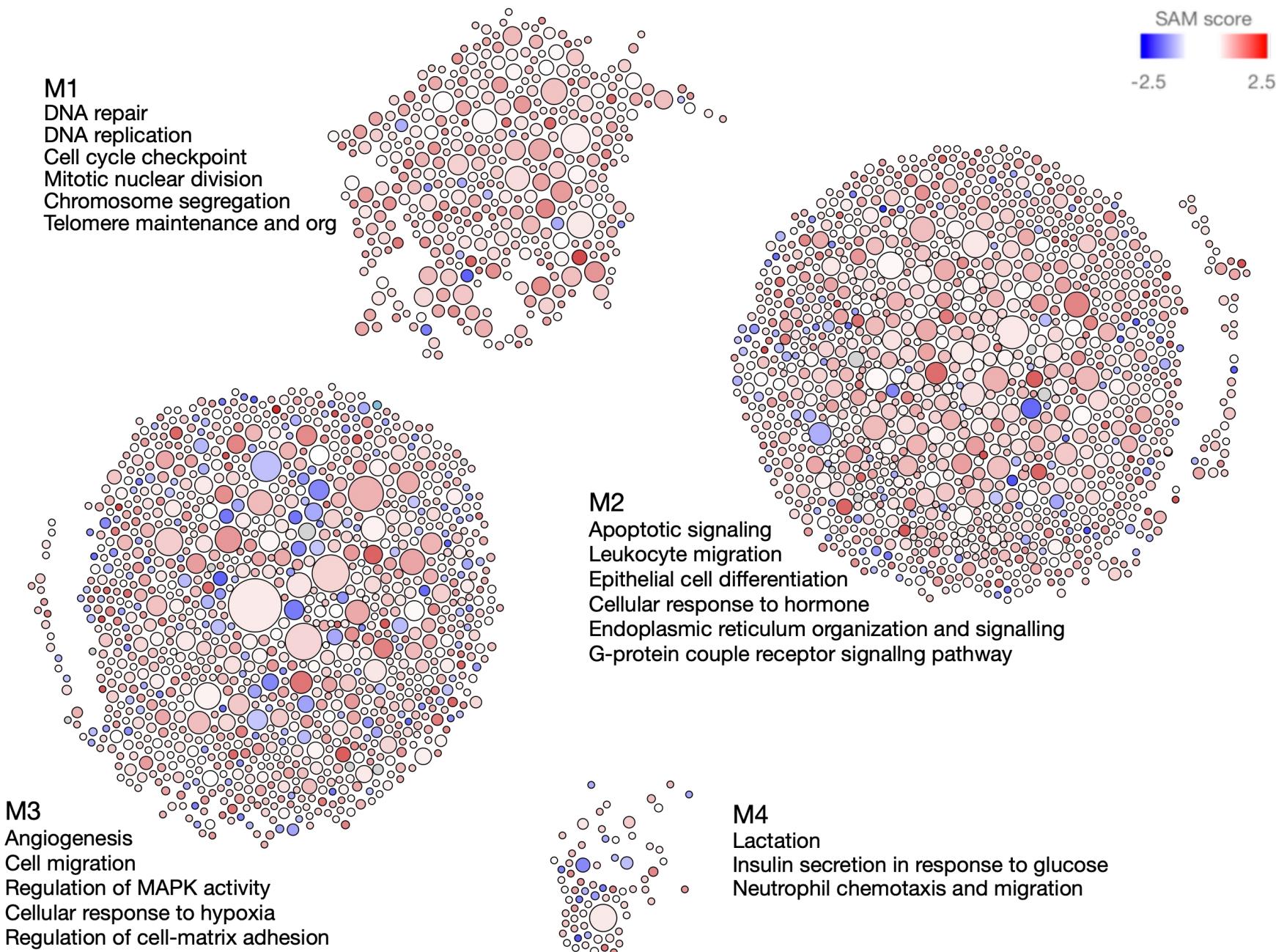
- Hypergeometric test (Fisher's exact test)
- Is overlap between DE genes and gene set term > random set?

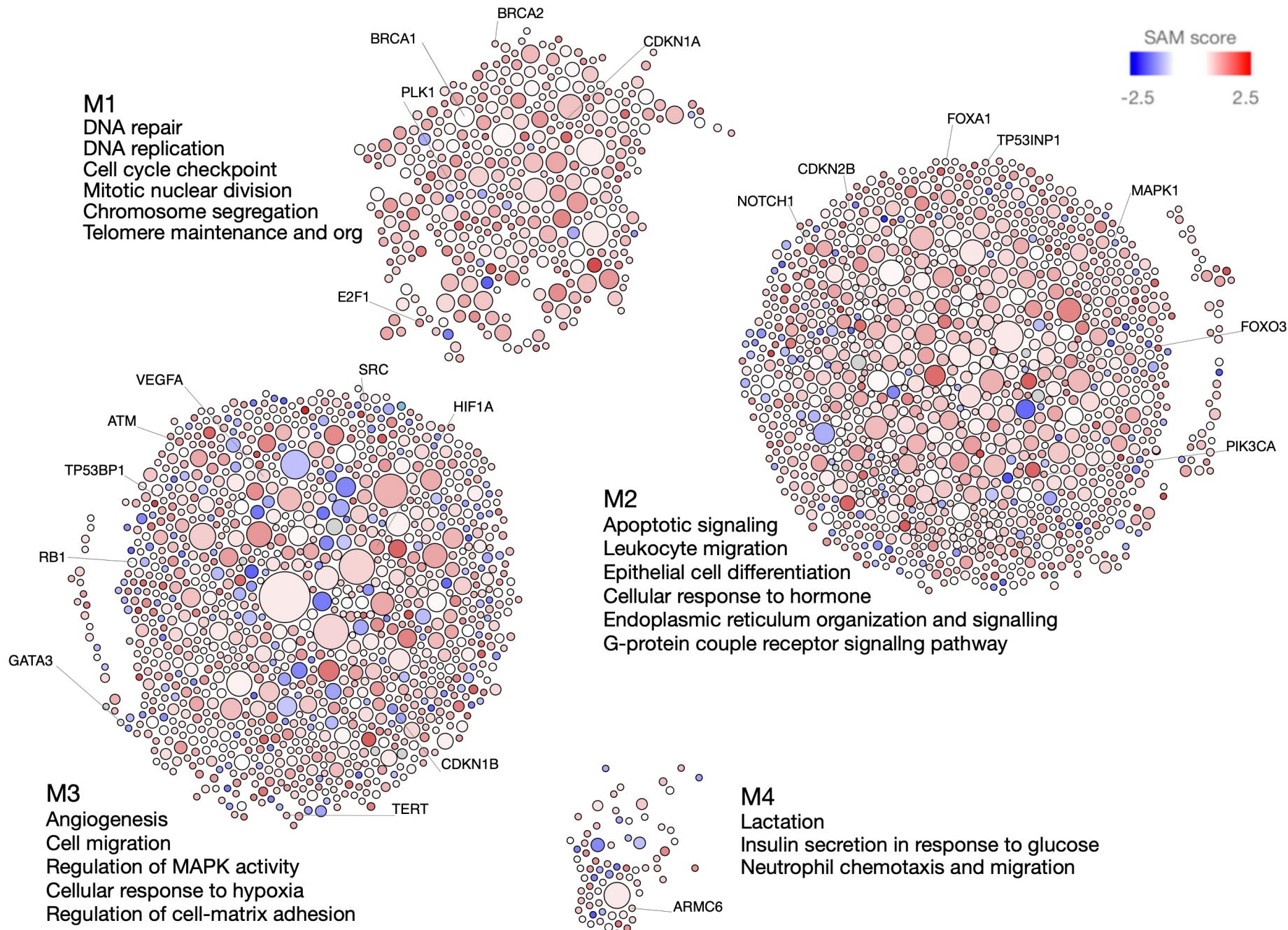




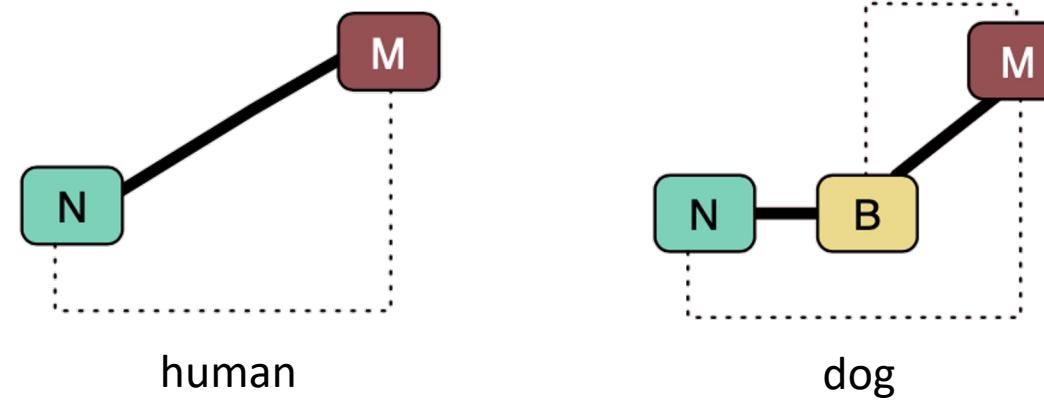
mammary epithelium network  
shared nearest-neighbor clustering



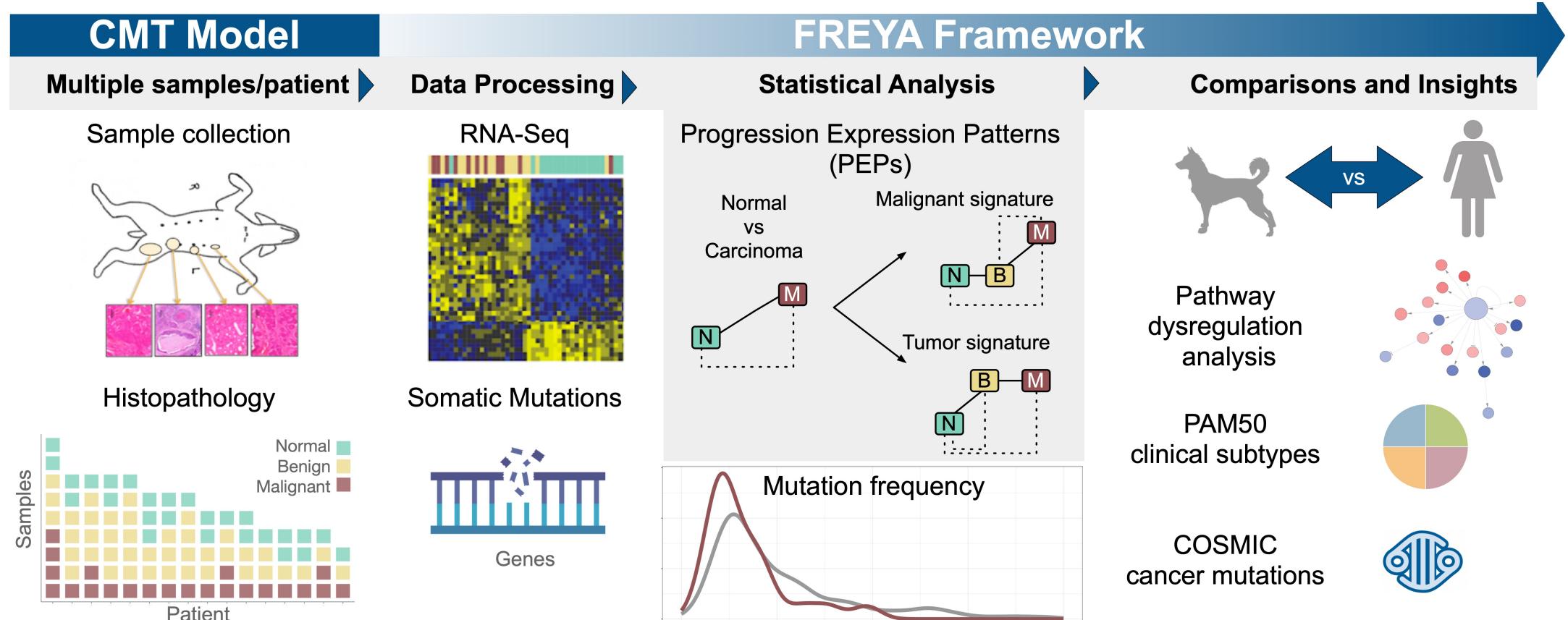




# Modeling the molecular development of human breast cancer using dogs

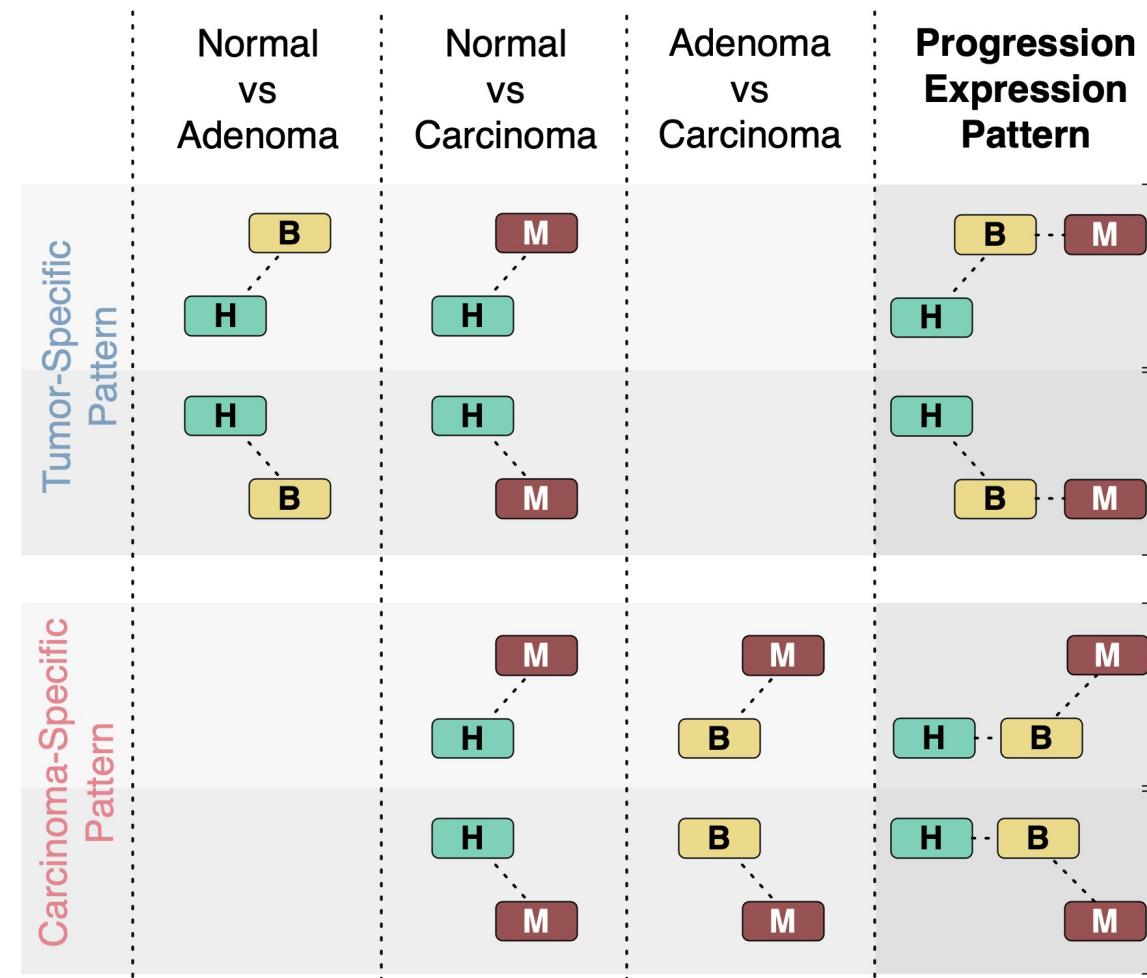


# FREYA: FRamework for Expression analYsis Across species

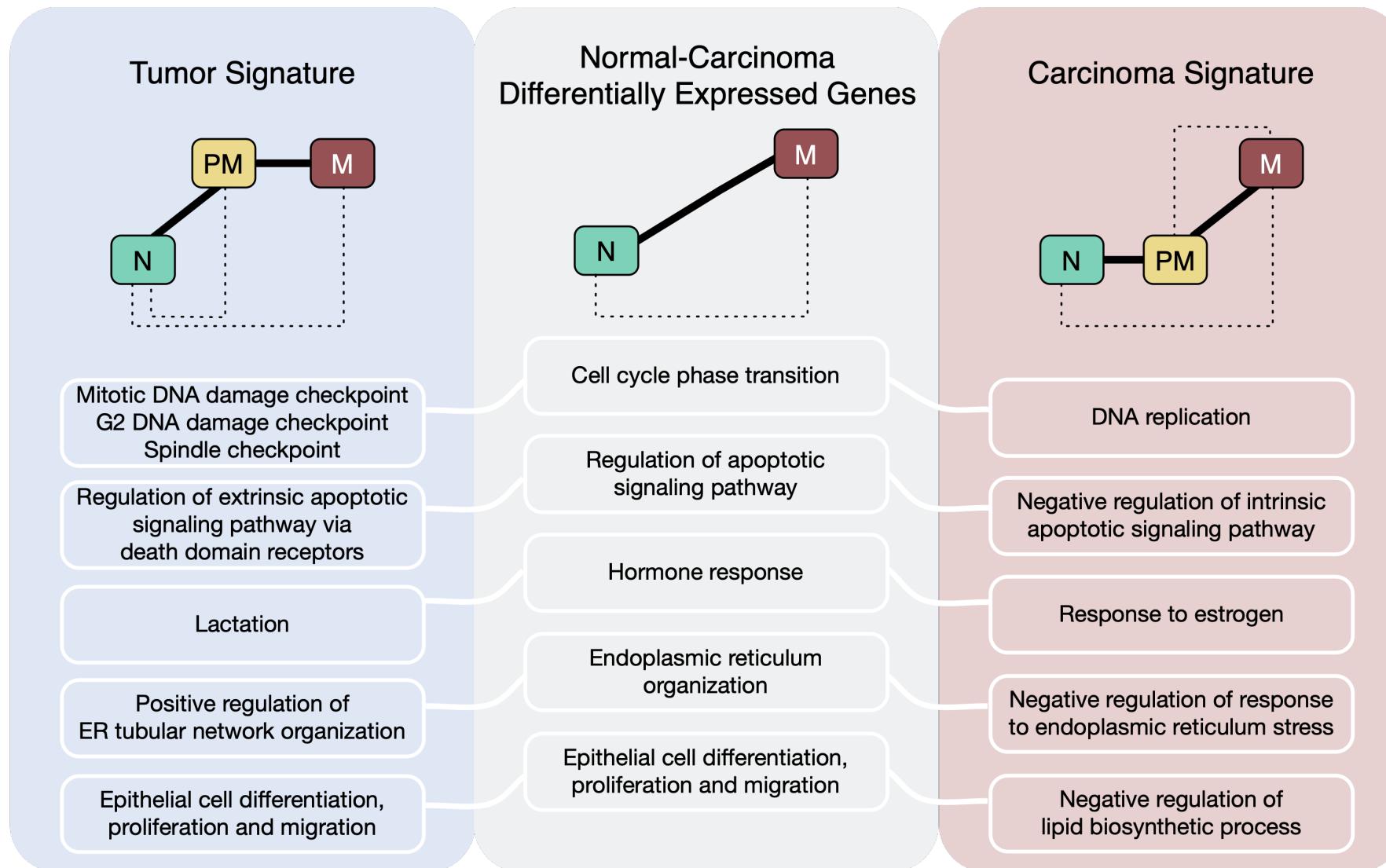


[freya.flatironinstitute.org](http://freya.flatironinstitute.org)

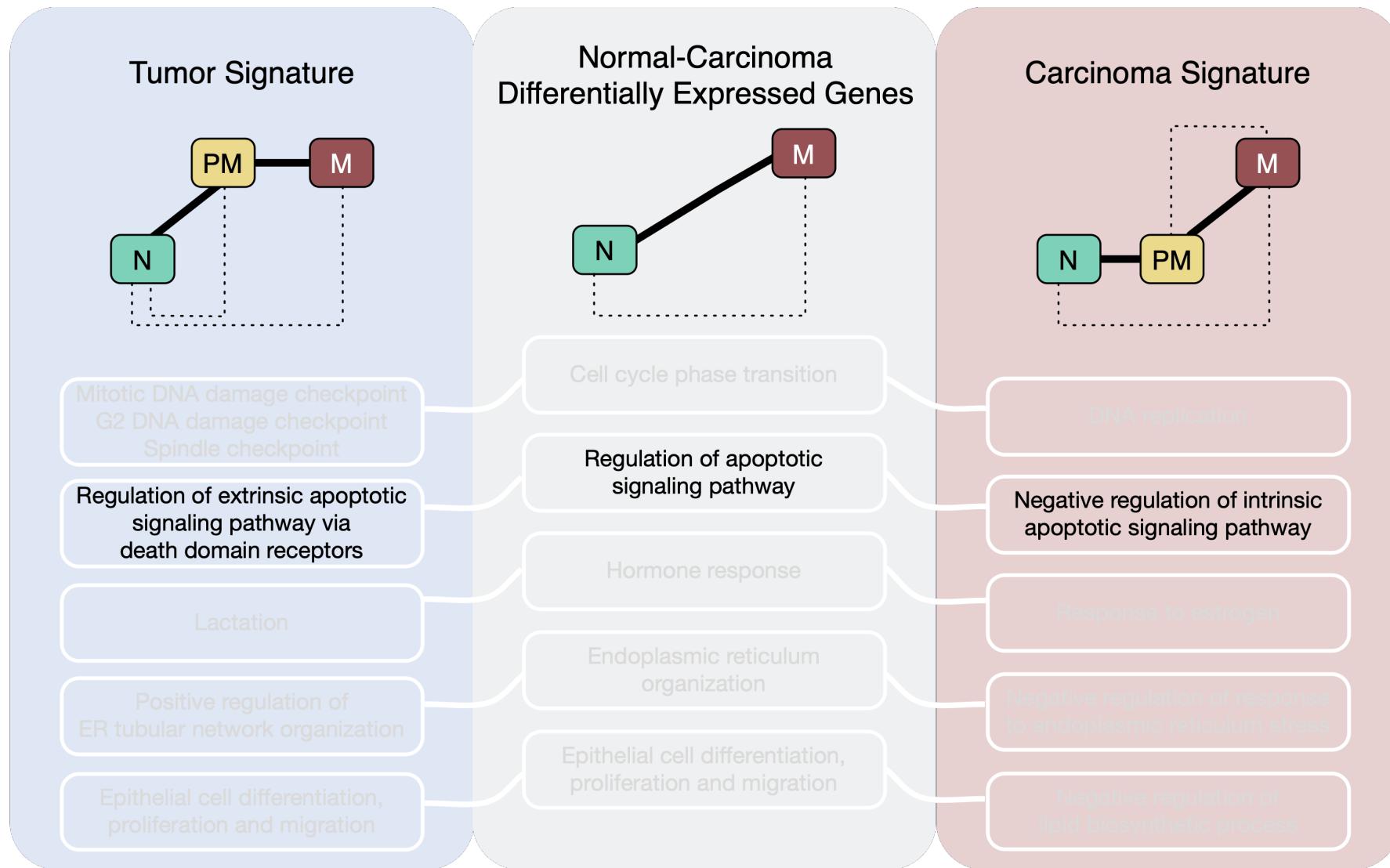
# Progression Expression Patterns (PEPs)



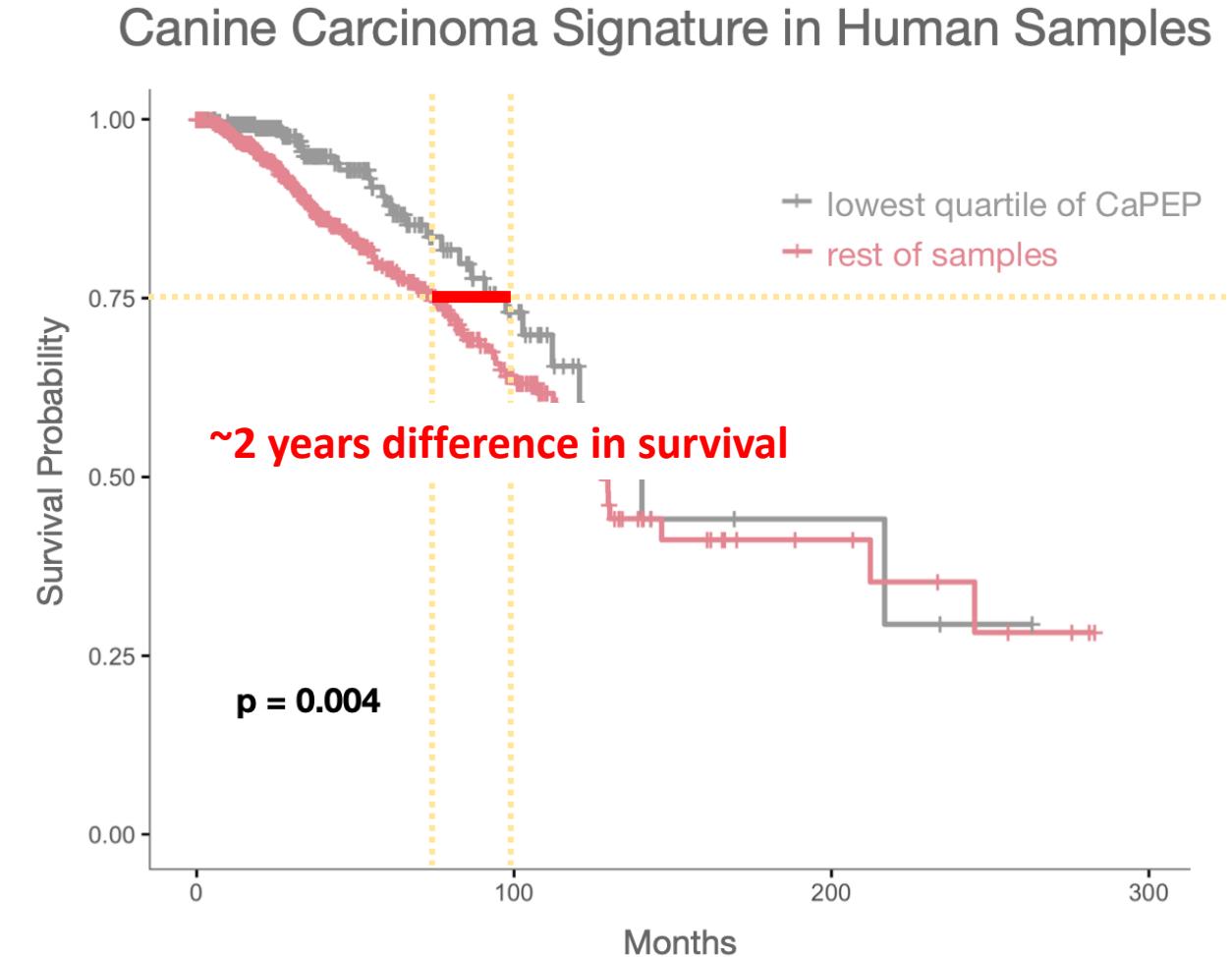
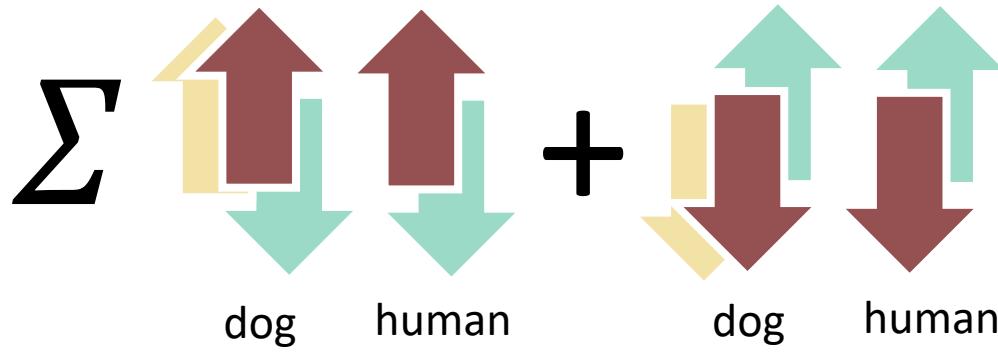
# Resolution of Hallmark Cancer Processes Helps Discern PEPs



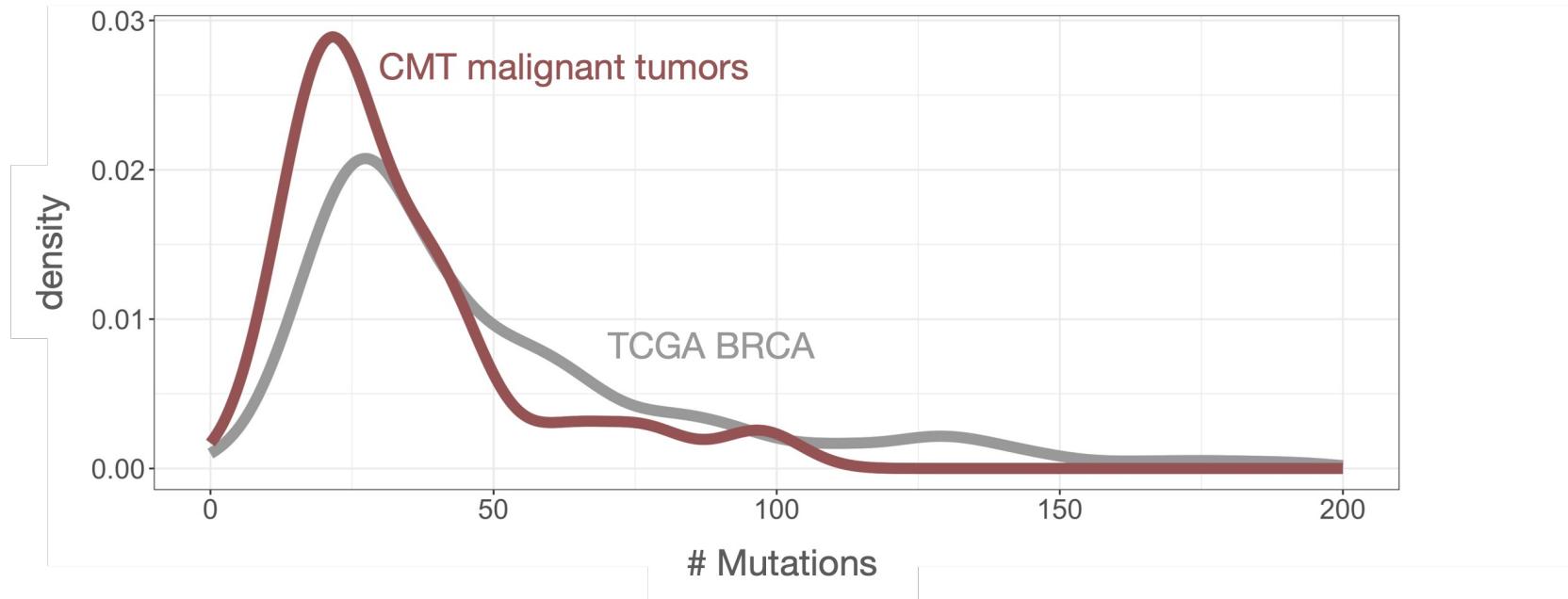
# Resolution of Hallmark Cancer Processes Helps Discern PEPs



# Calculating Human CaPEP Scores

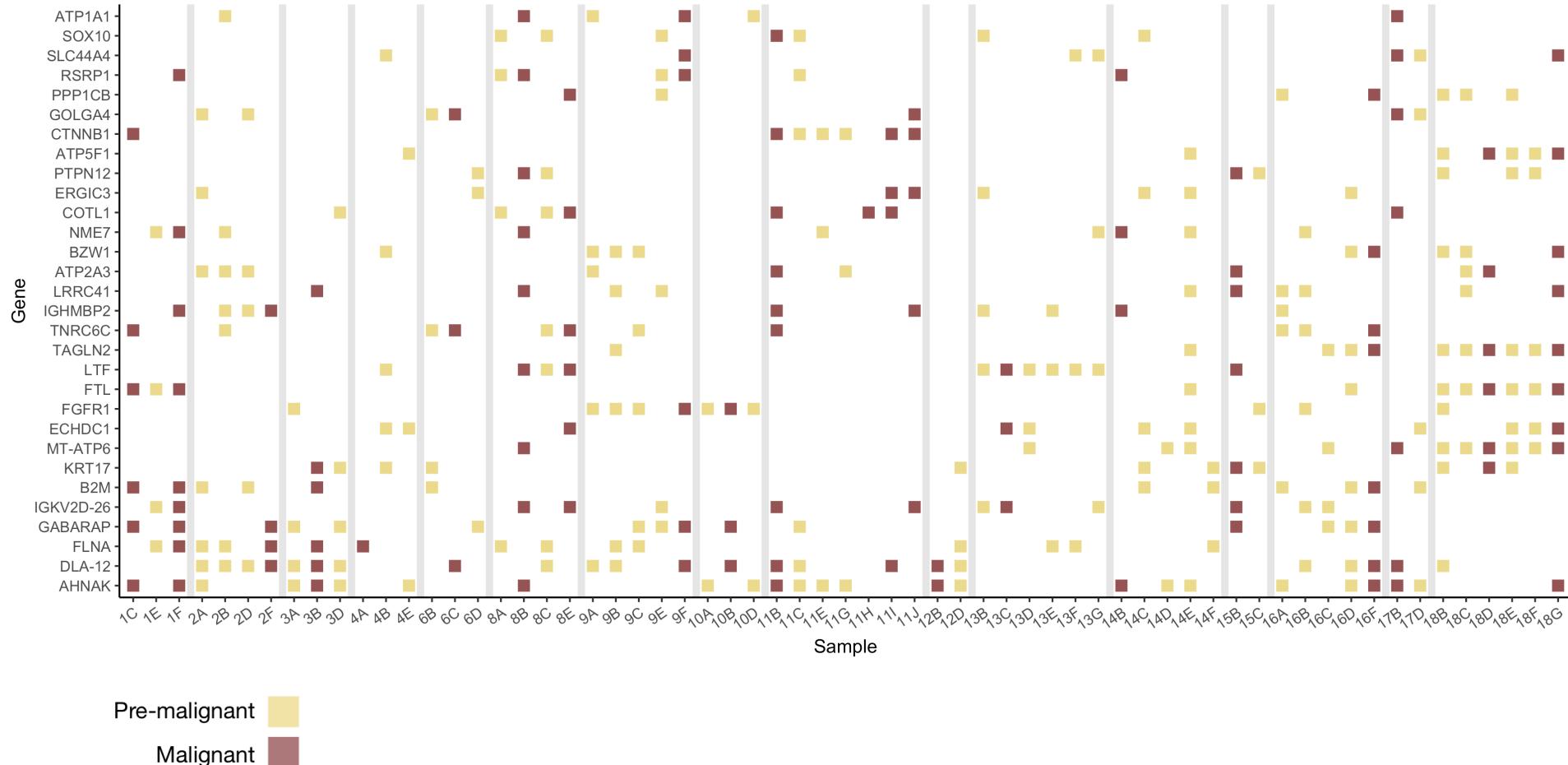


# Similar mutations rates across human and dog breast cancers

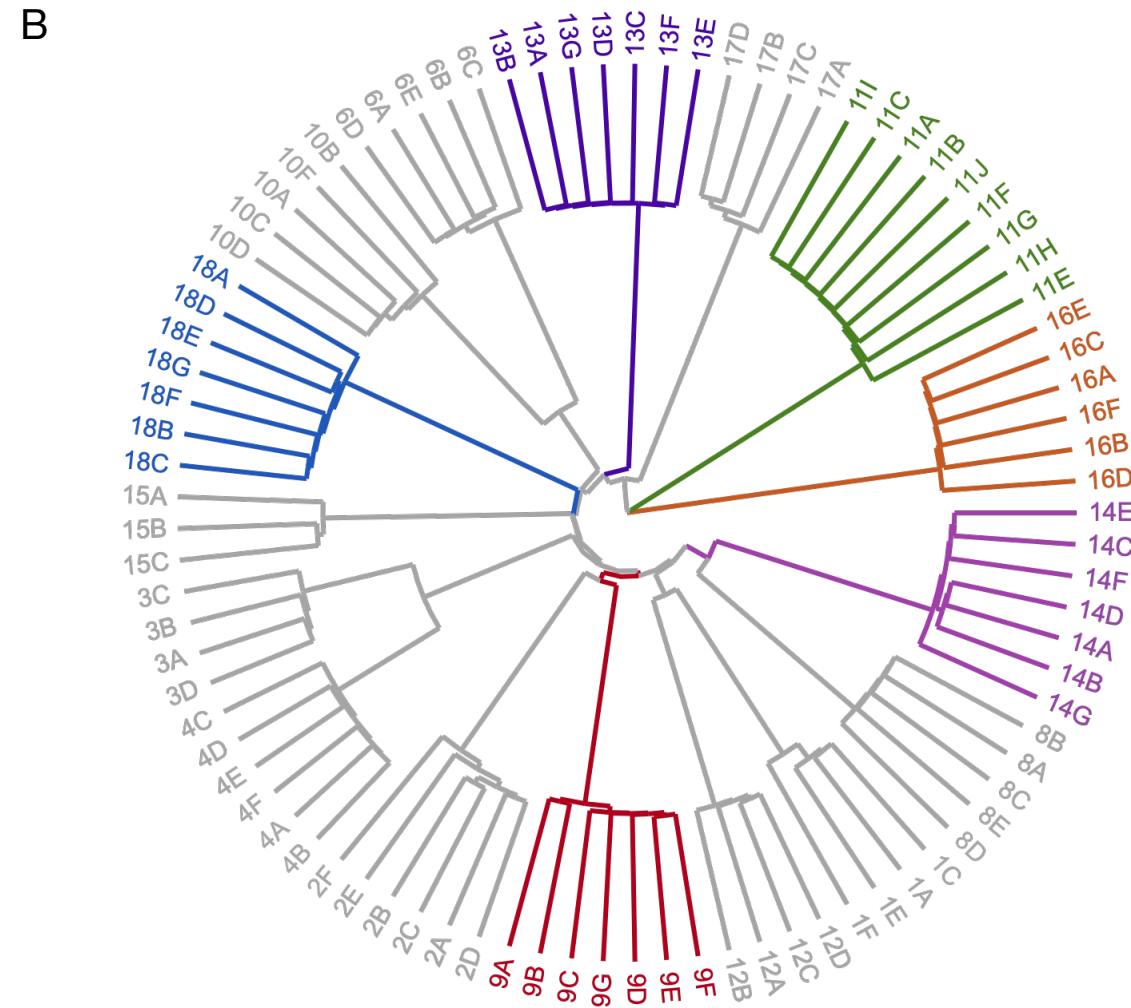
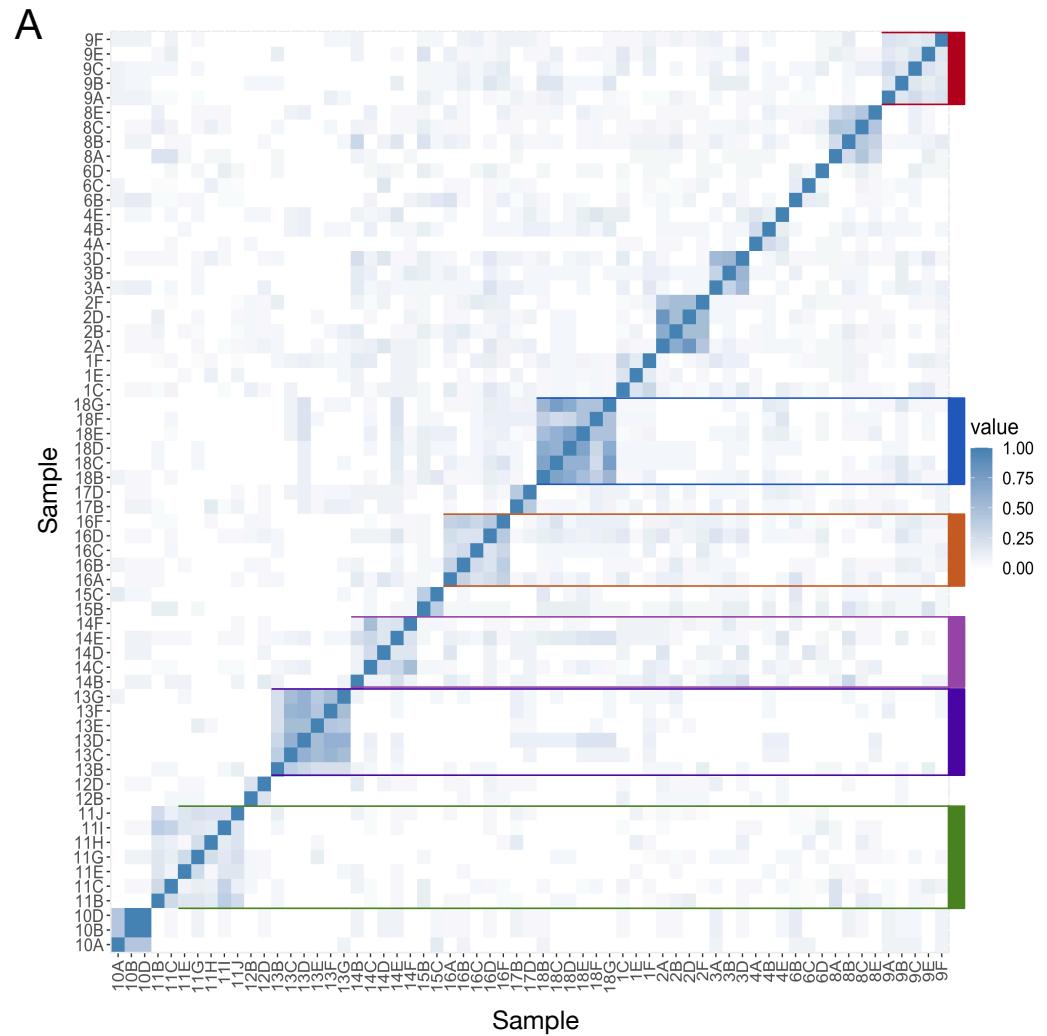


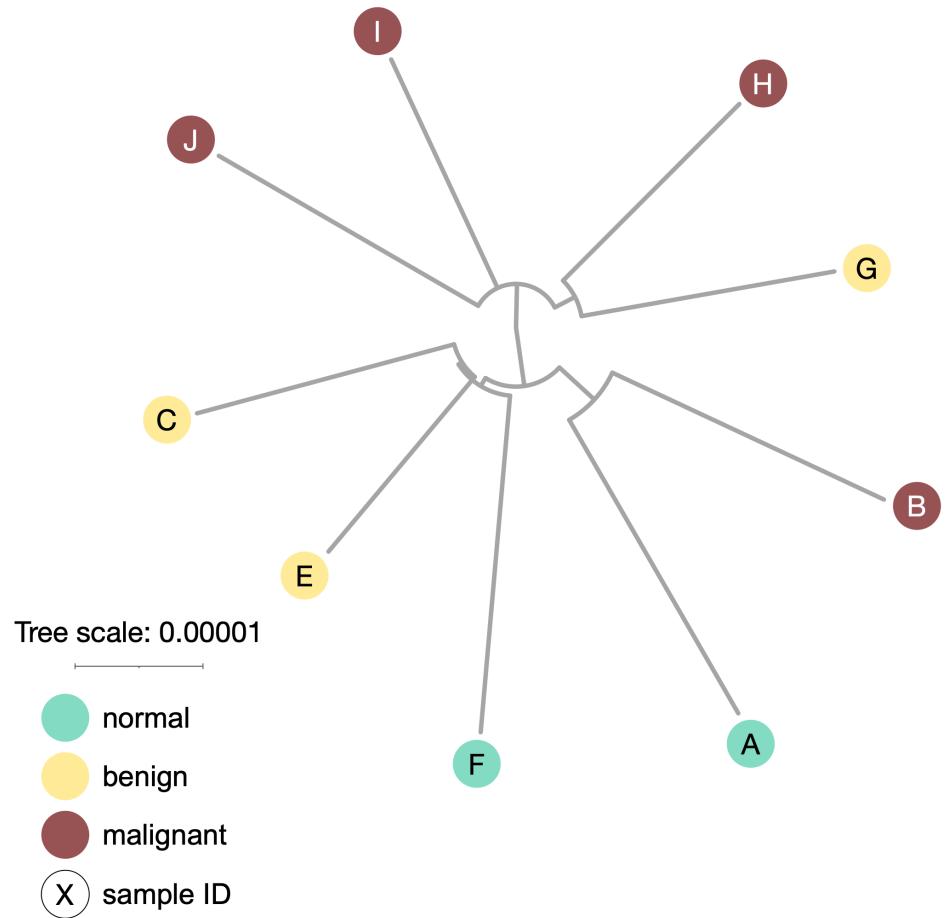
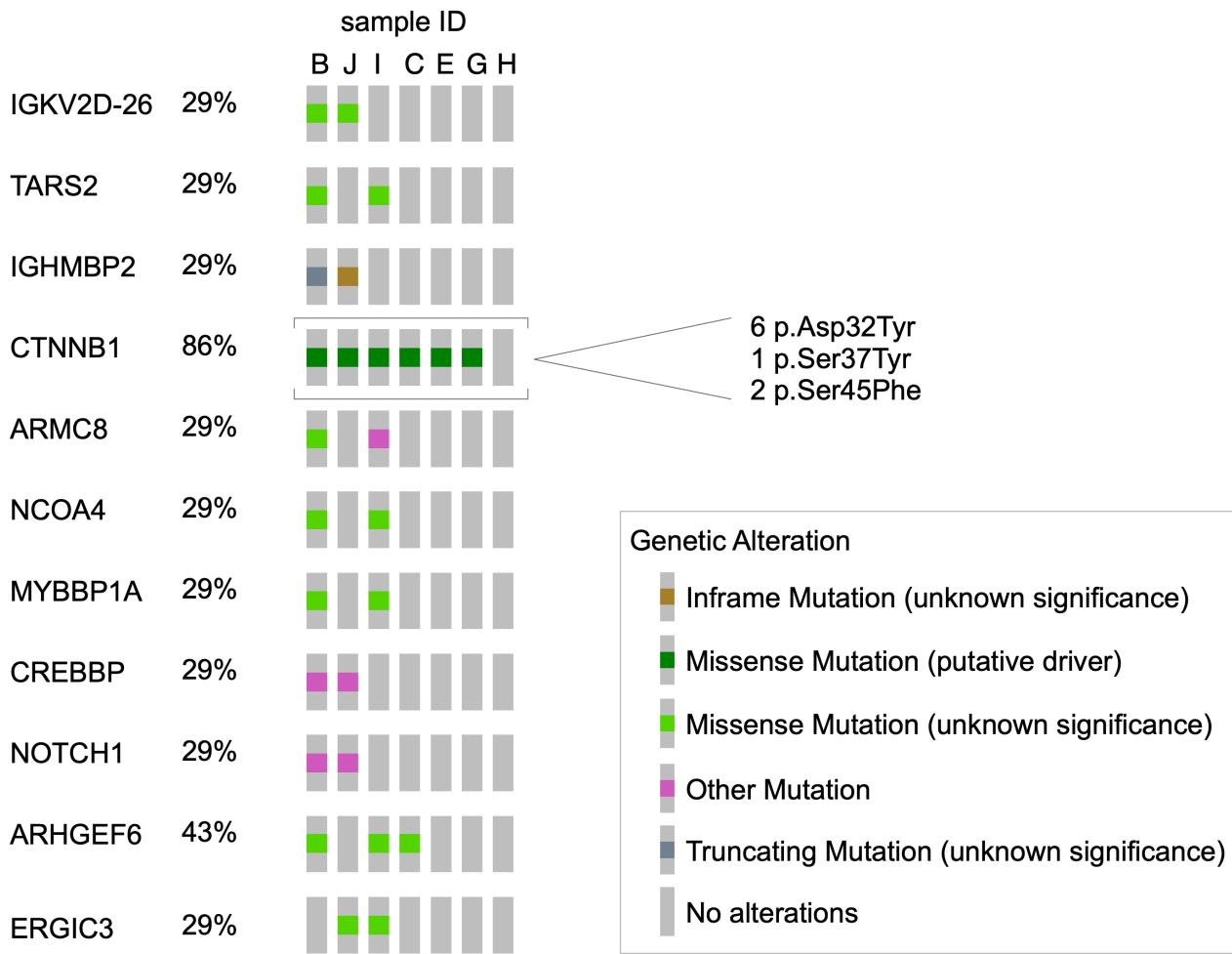
1904 mutations in 524 genes  
226 are COSMIC genes

# Mutations are Consistent within Each Individual Dog



# Most dogs have multiple primary tumors



**A****B**



# GRAIM LAB

Thanks!



Support was provided by the University of Florida Health Cancer Center Pilot Grant # AI-2022-02  
Support was provided by NCI R01 # 1R01CA265907-01

