

Pathway Fingerprinting

A functional framework for multi-platform integration

Gabriel Altschuler - Hide Lab

Harvard School of Public Health

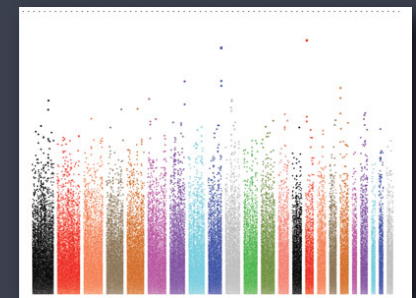
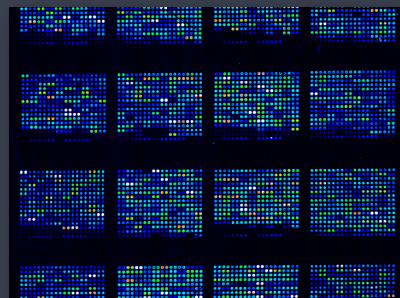
FANTOM 5

21st February 2011

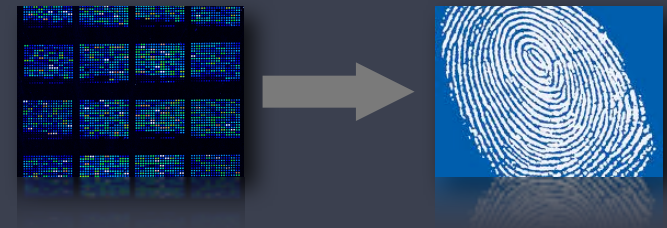
The Pathway Fingerprint

A resource for classification and discovery of biological function

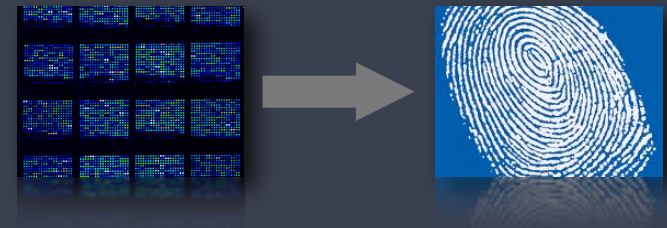
- Unified framework to analyze the genome, epigenome and transcriptome



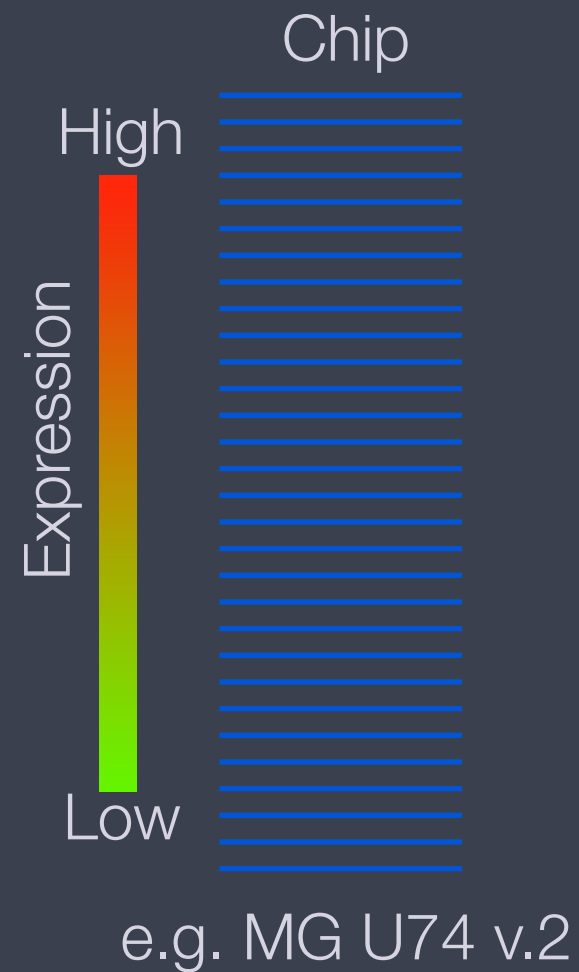
Pathway Fingerprint: Expression array



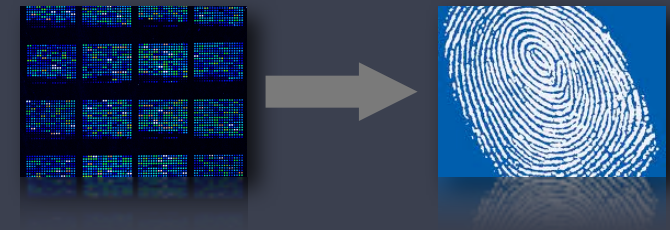
Pathway Fingerprint: Expression array



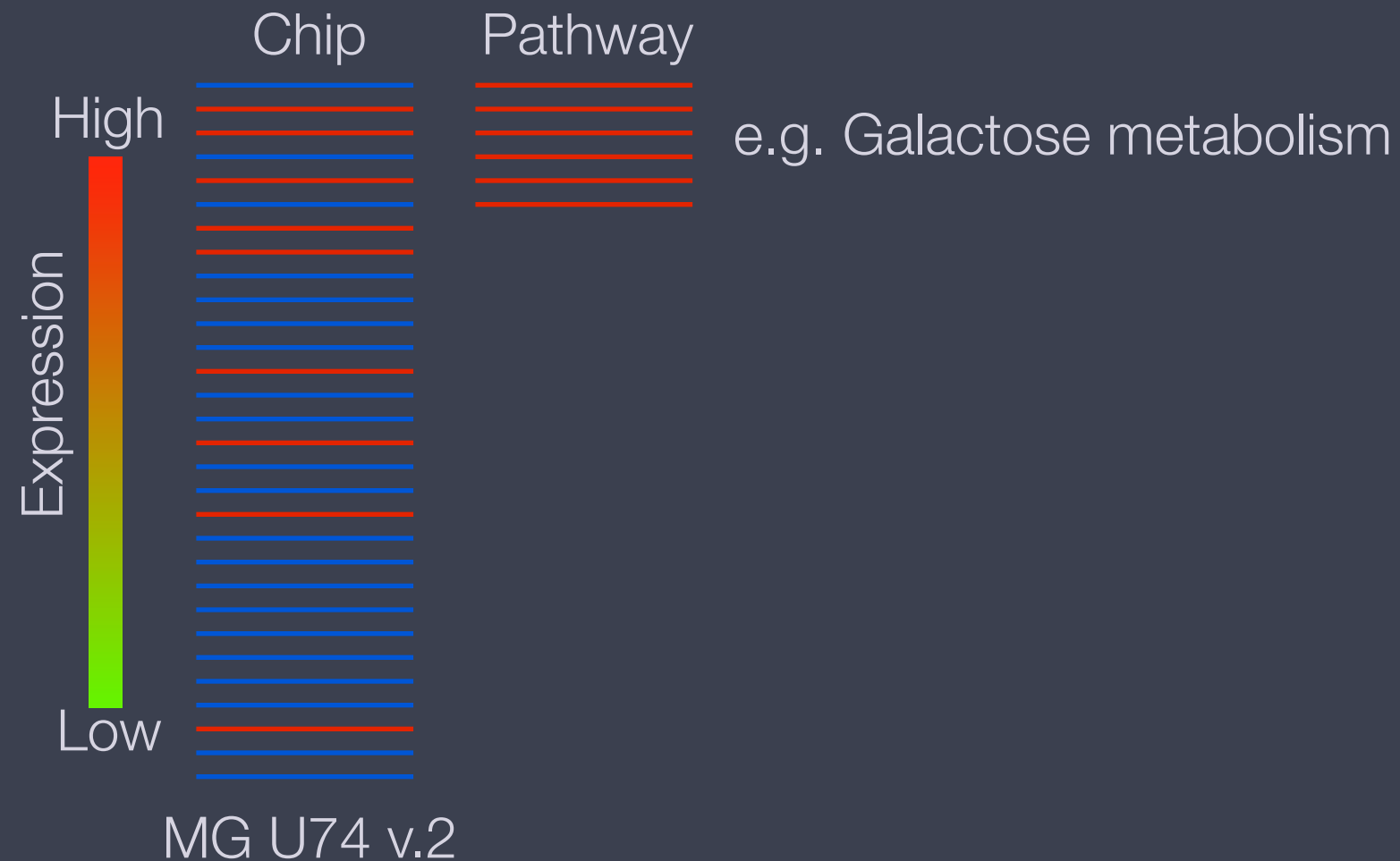
1. Genes ranked by raw expression



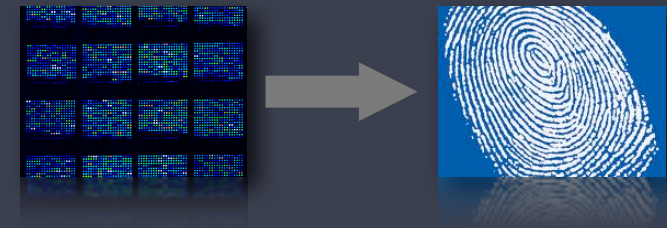
Pathway Fingerprint: Expression array



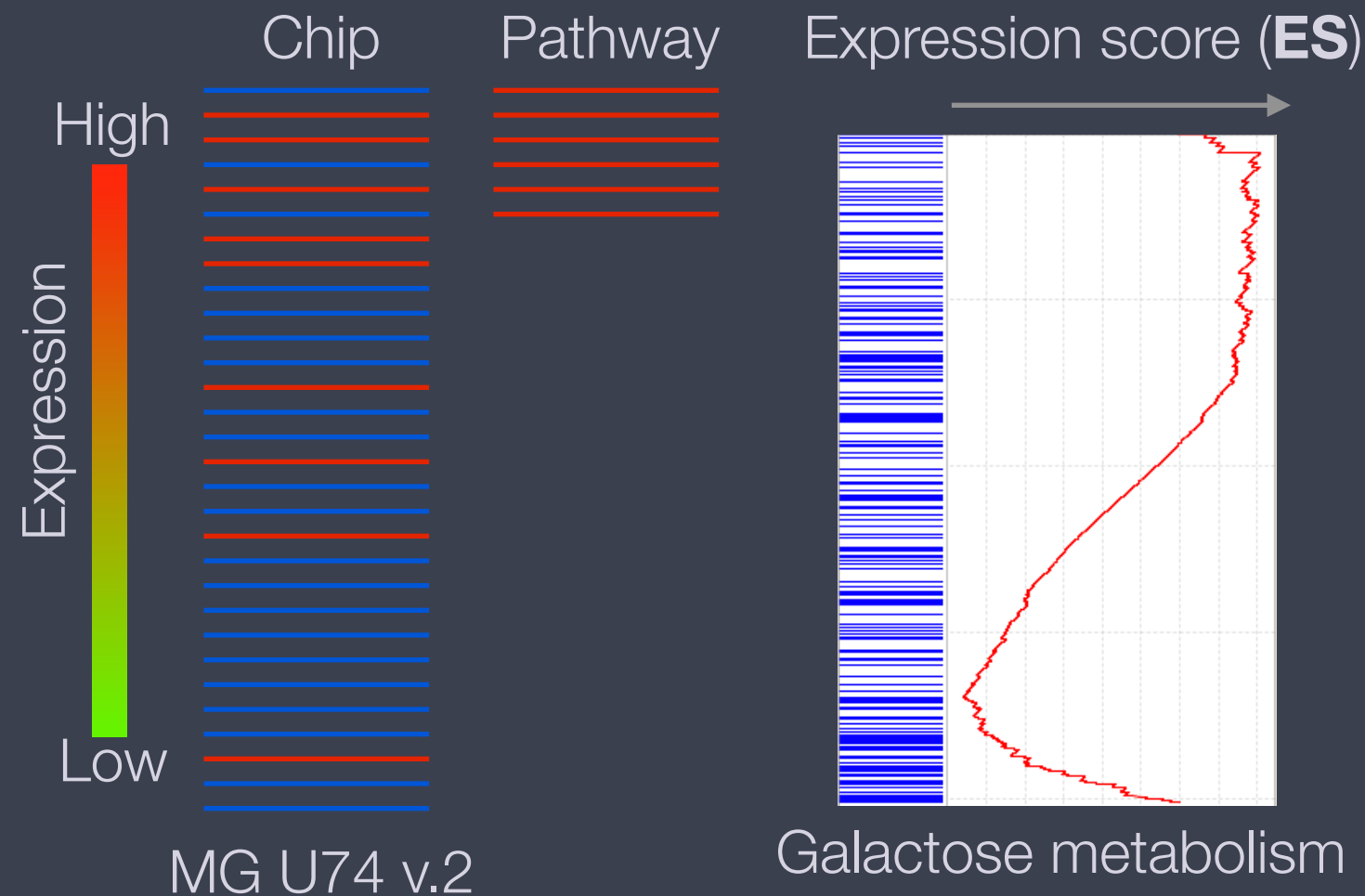
2. Identify overlap with pathway



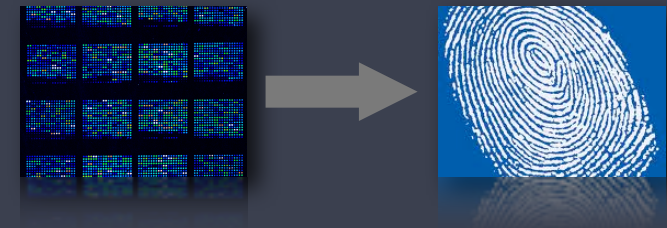
Pathway Fingerprint: Expression array



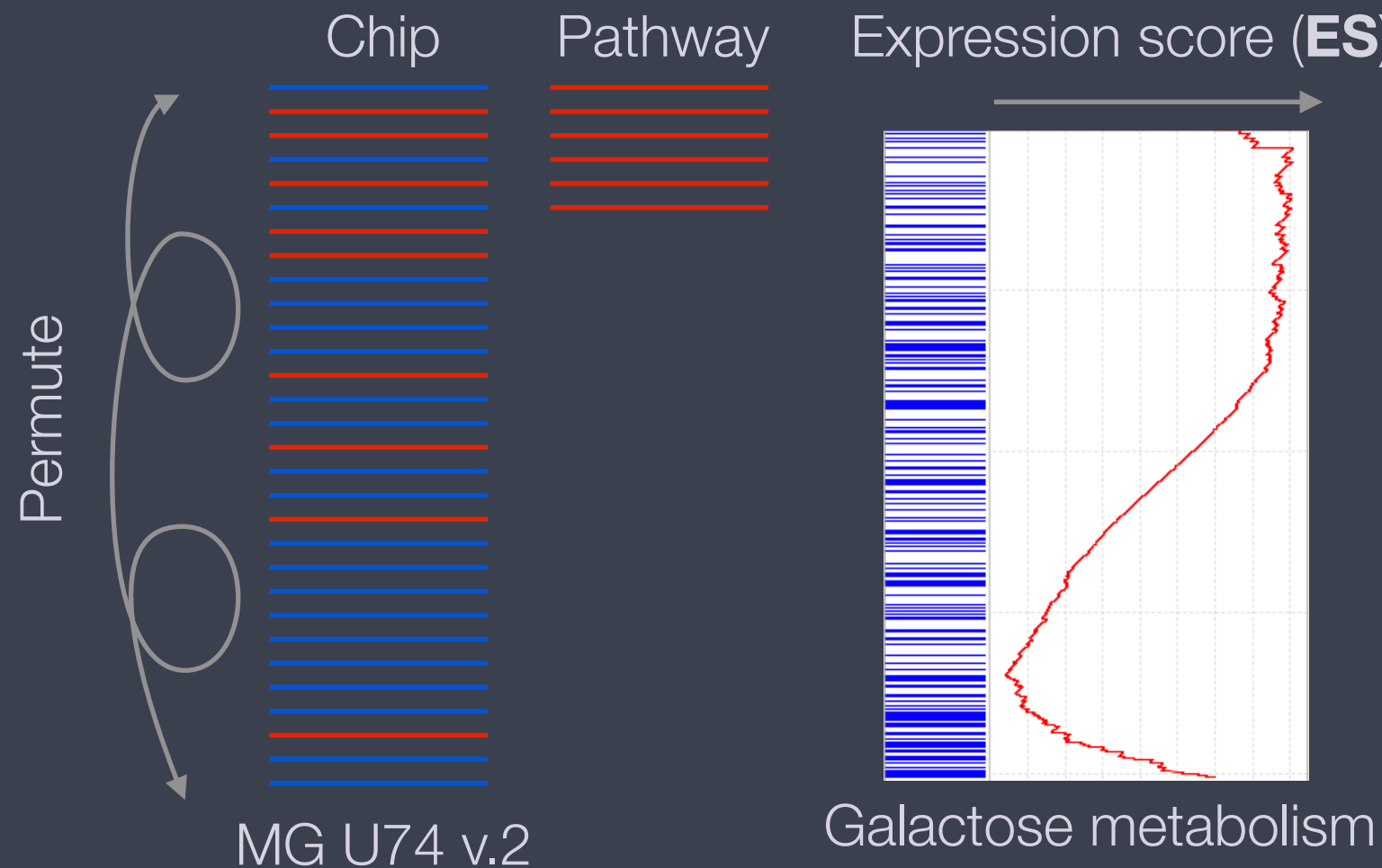
3. Score pathway by modified Kolmogorov-Smirnov statistic (GSEA type)



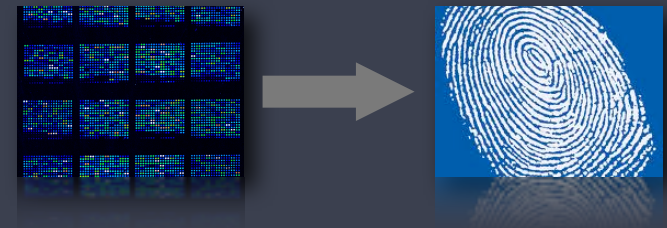
Pathway Fingerprint: Expression array



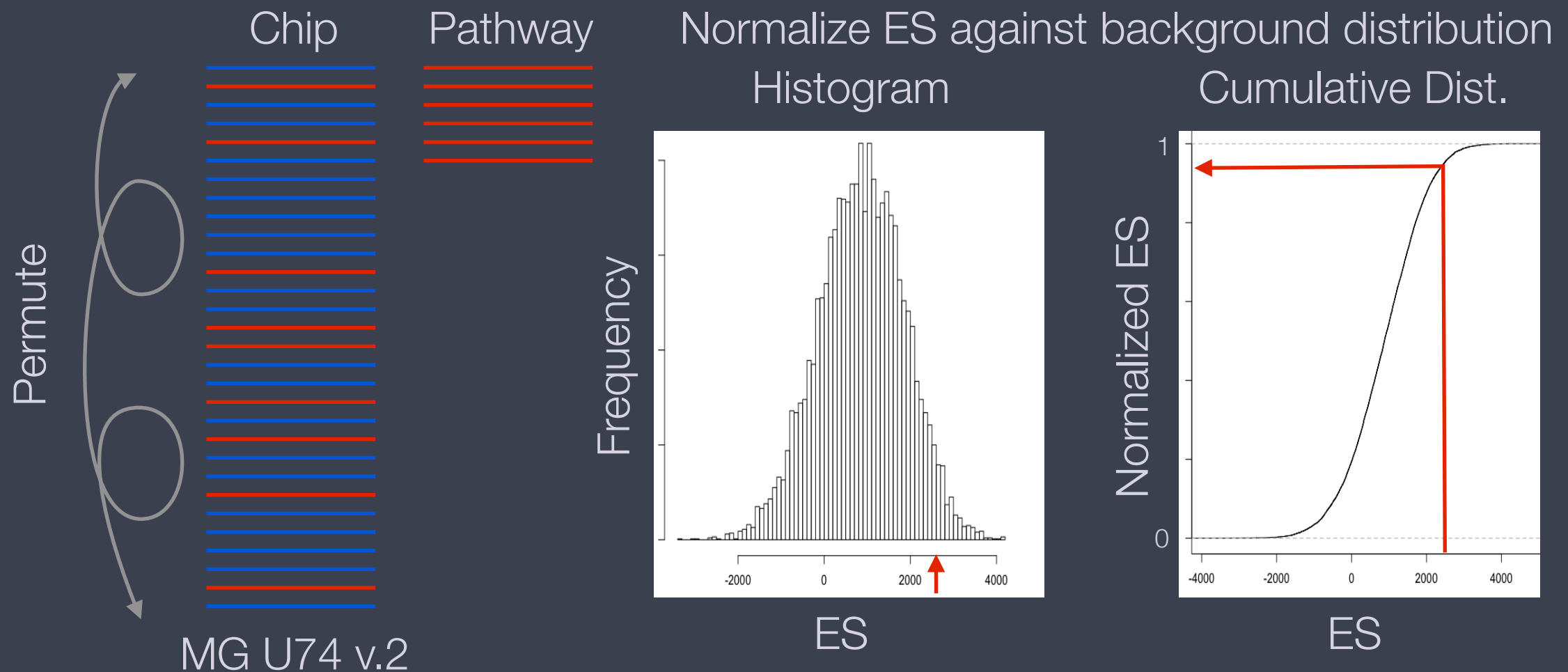
3. Score pathway by modified Kolmogorov-Smirov statistic (GSEA type)



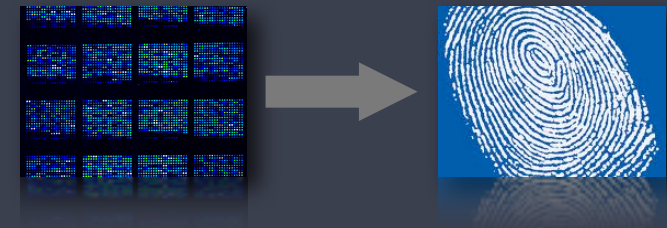
Pathway Fingerprint: Expression array



4. Permute genes - test against Q1 hypothesis

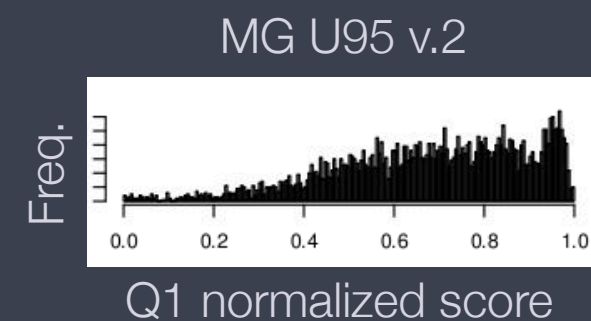
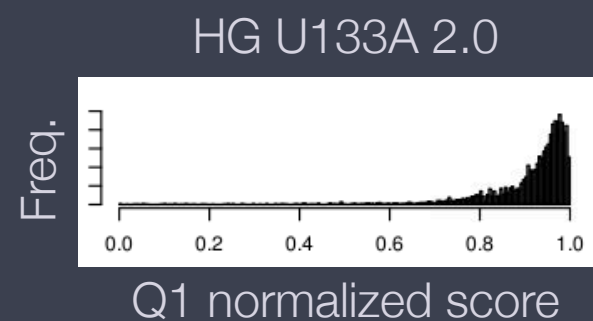
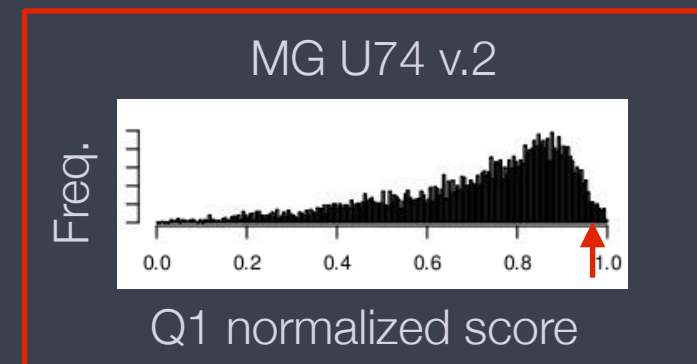
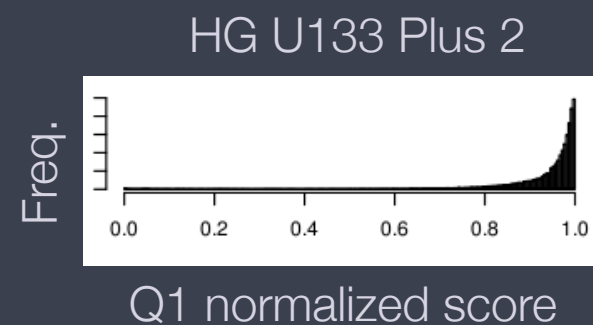


Pathway Fingerprint: Expression array



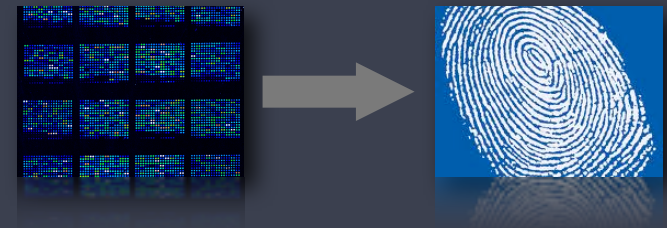
5. Permute samples, platform specific test against Q2 hypothesis

Galactose Metabolism



Background built from GEO corpus (120,000 chips)

Pathway Fingerprint: Expression array

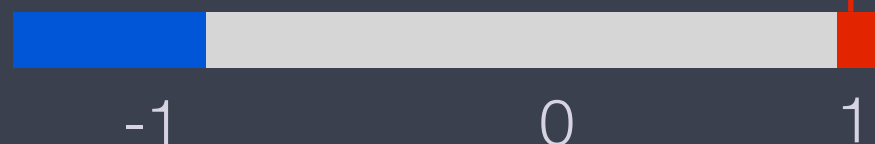


6. Represent as ternary (-1,0,1) vector, compile pathways

MG U74 v.2 - Galactose Metabolism



Threshold



Pathway fingerprint

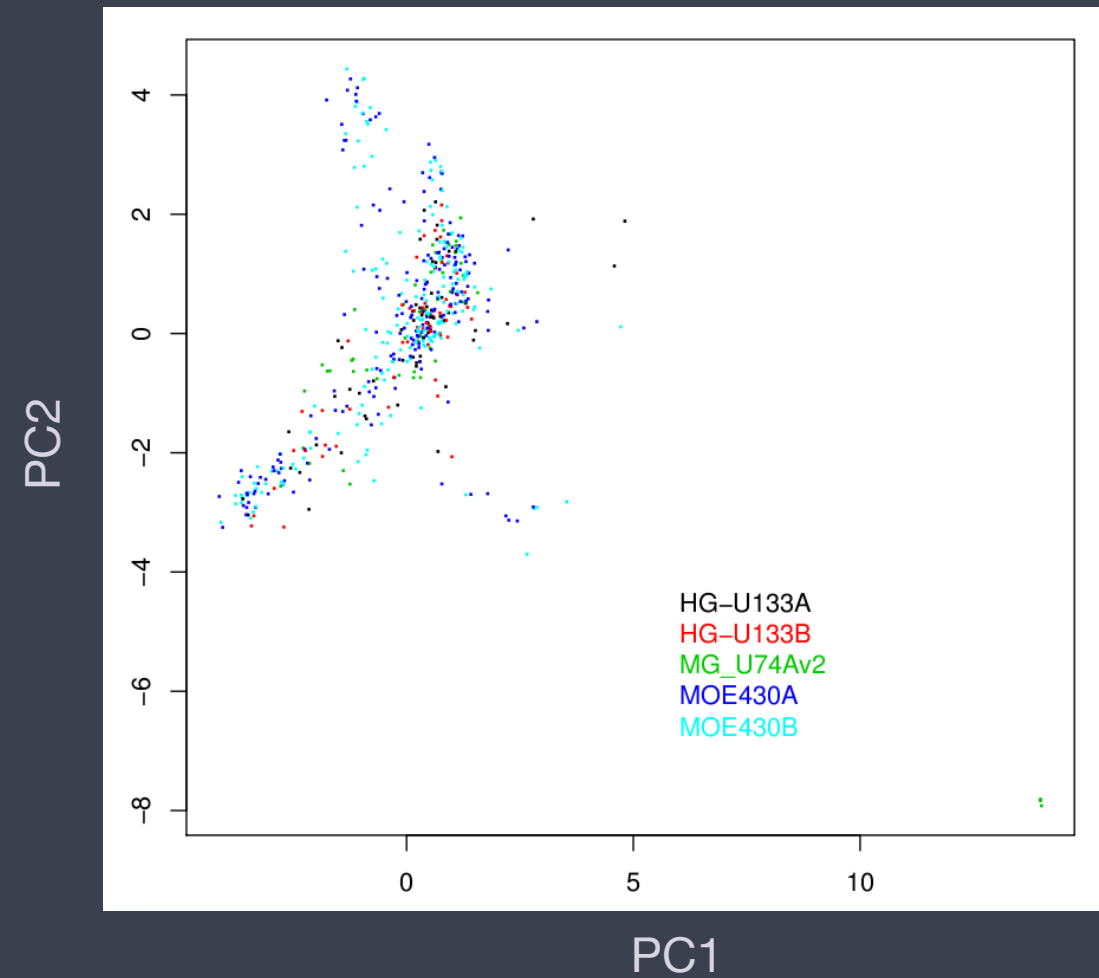
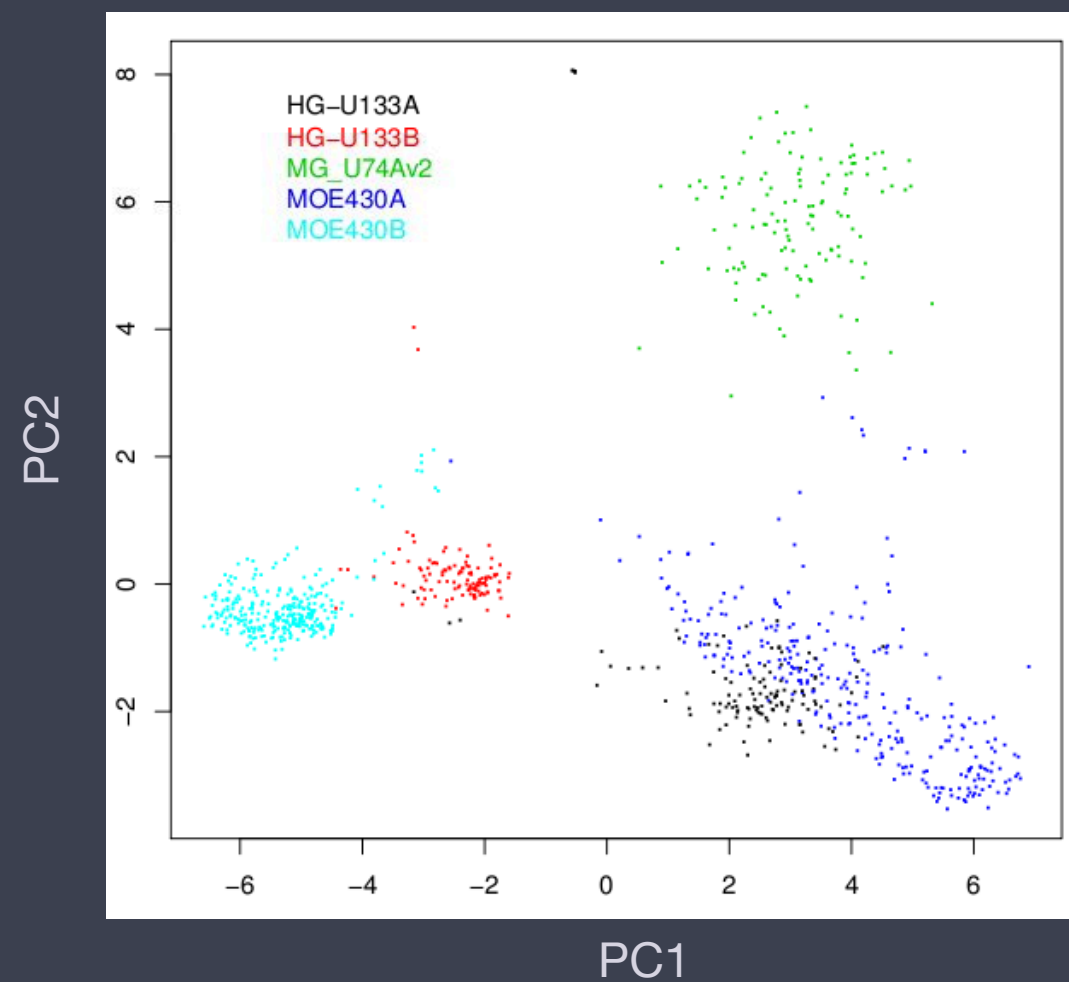
Complete suite of pathways

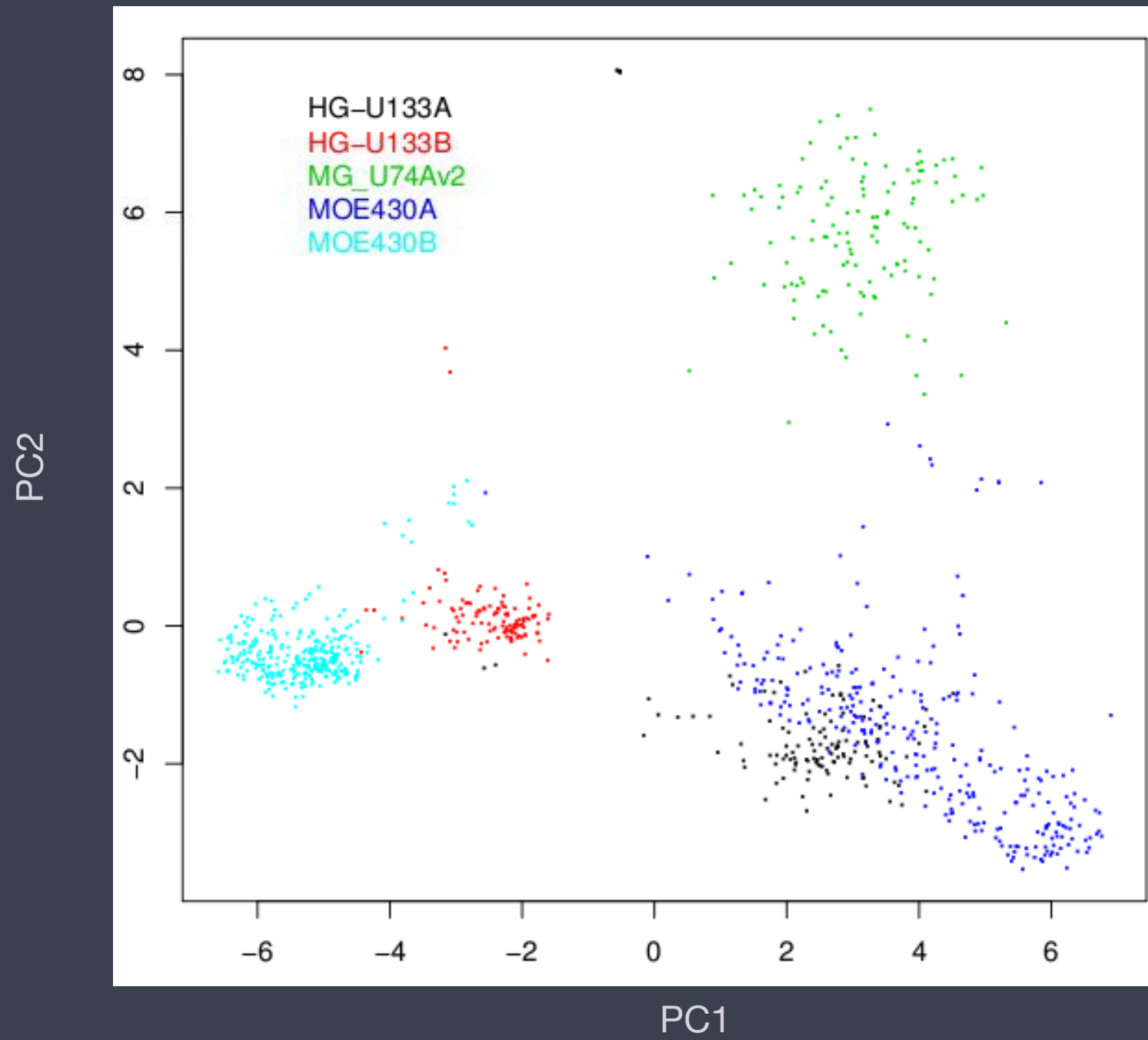


Galactose Metabolism

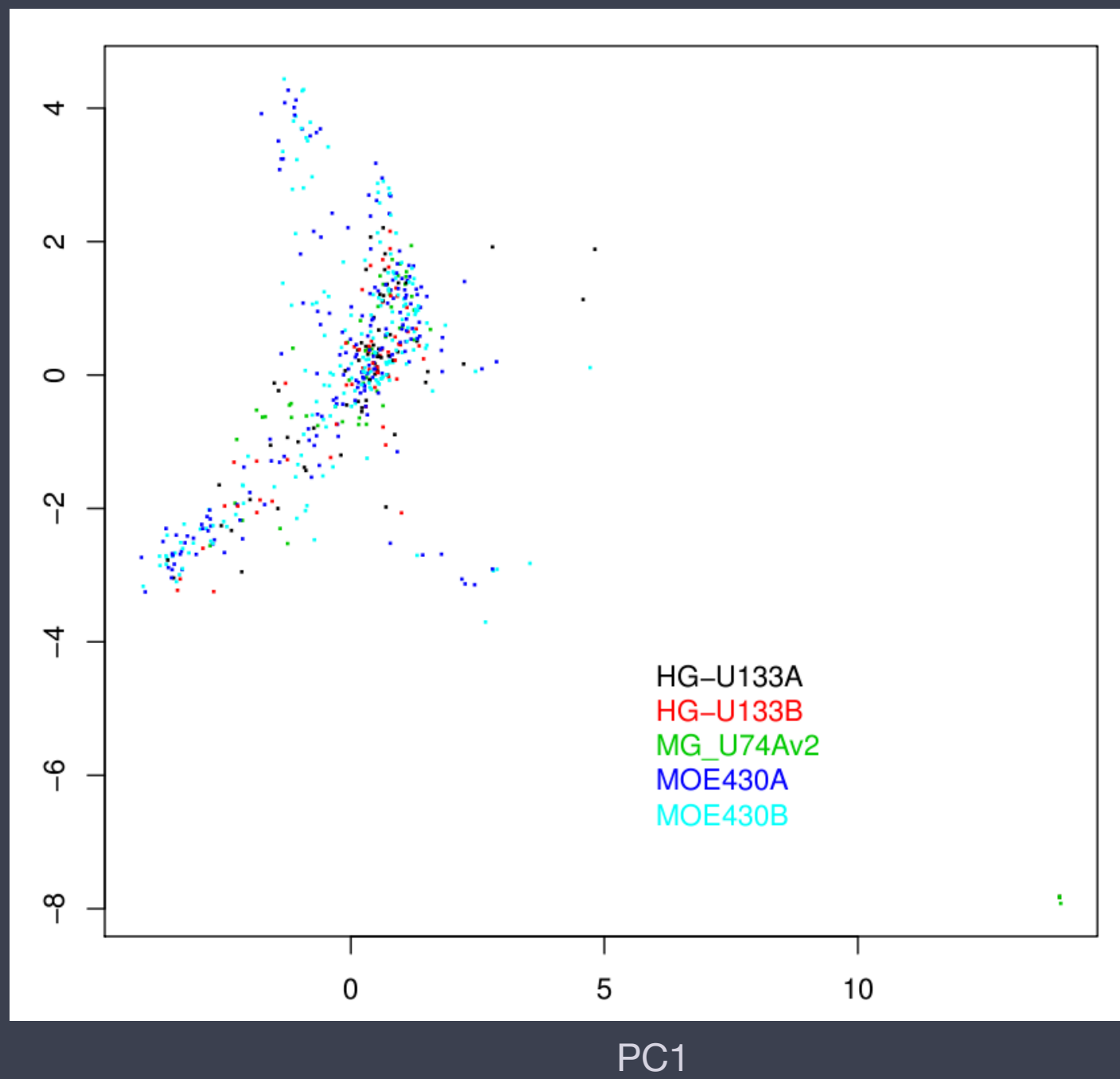
Background built from GEO corpus (120,000 chips)

Platform integration





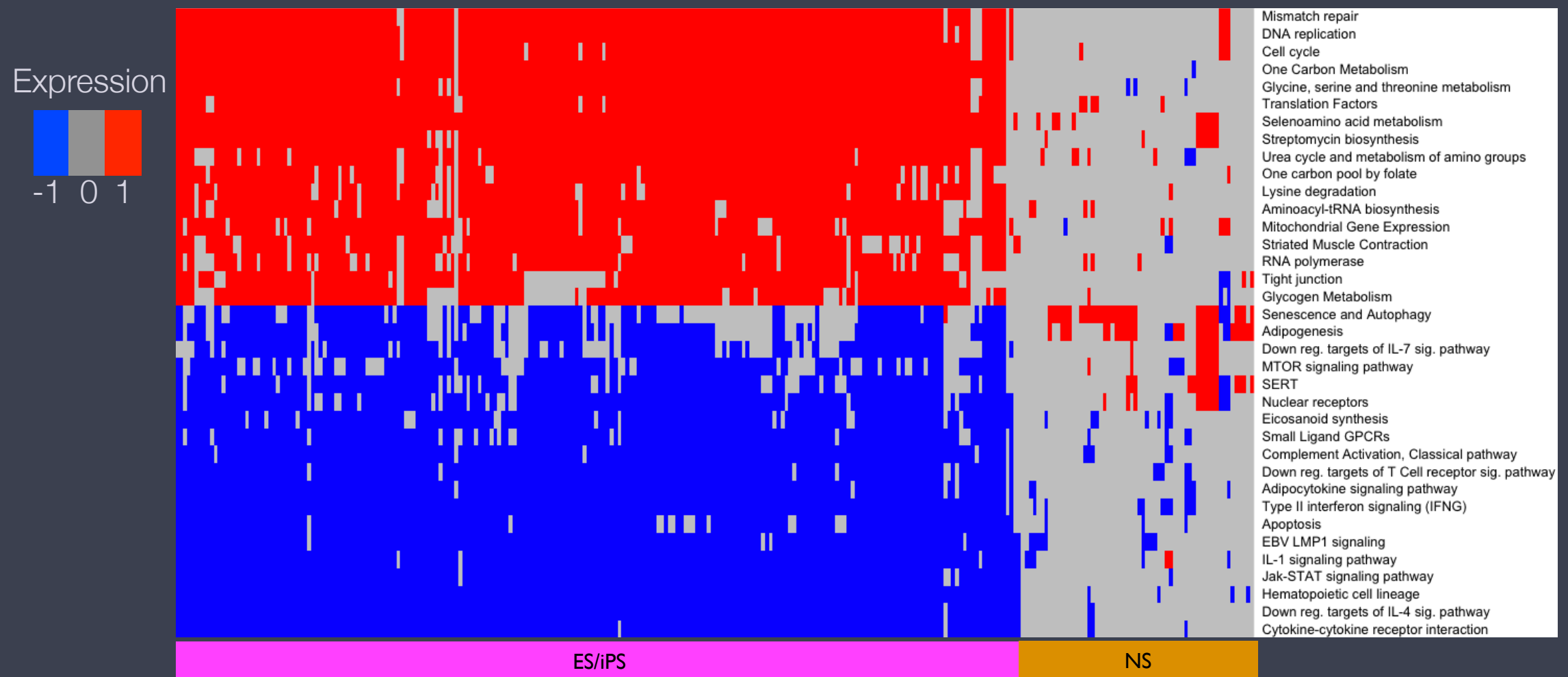
PC2



Case study 1 - Pluripotency

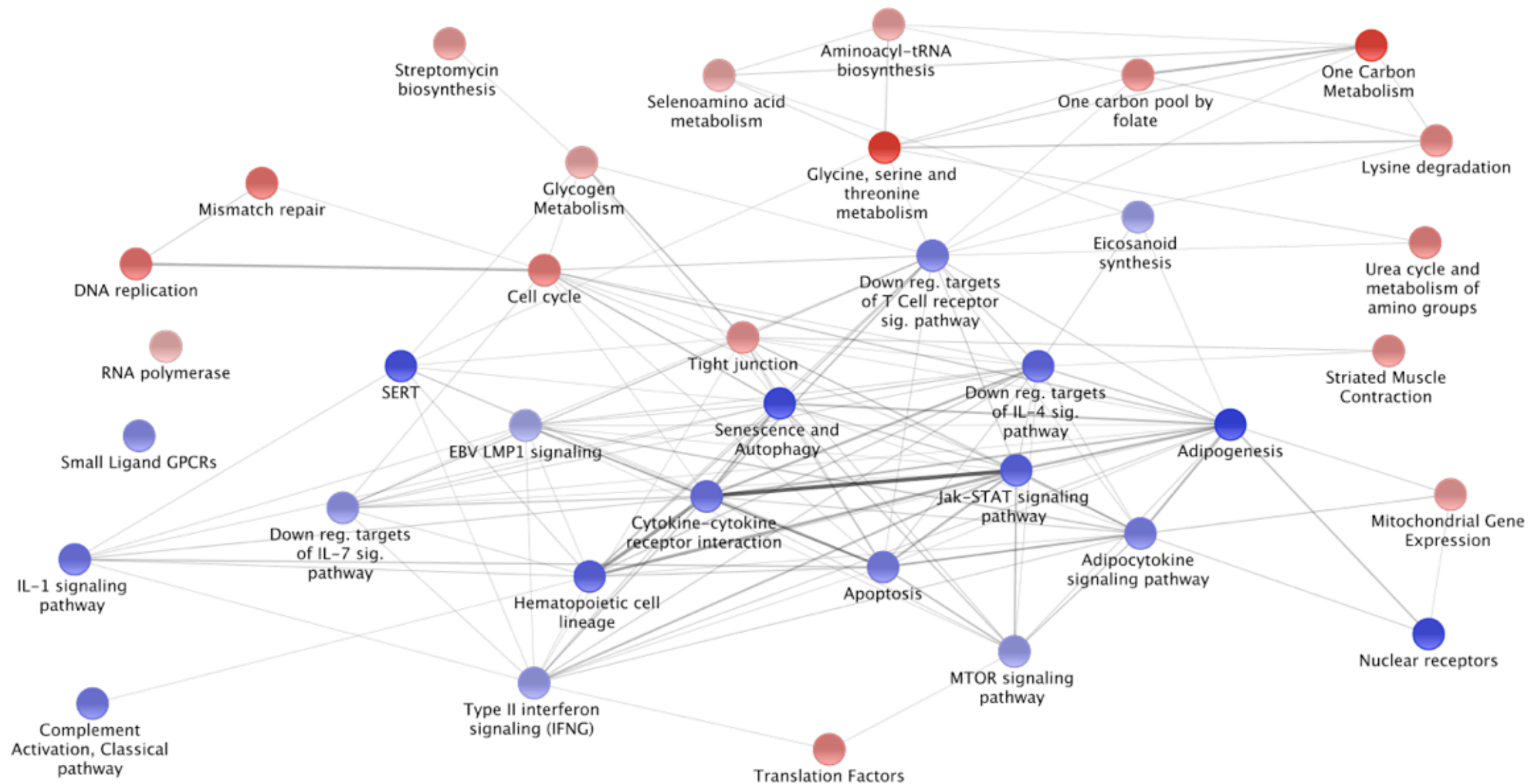
Can we use **pathway fingerprints** to determine the **key functional differences** between pluripotent cell types and their progeny?

Pluripotent pathway signature



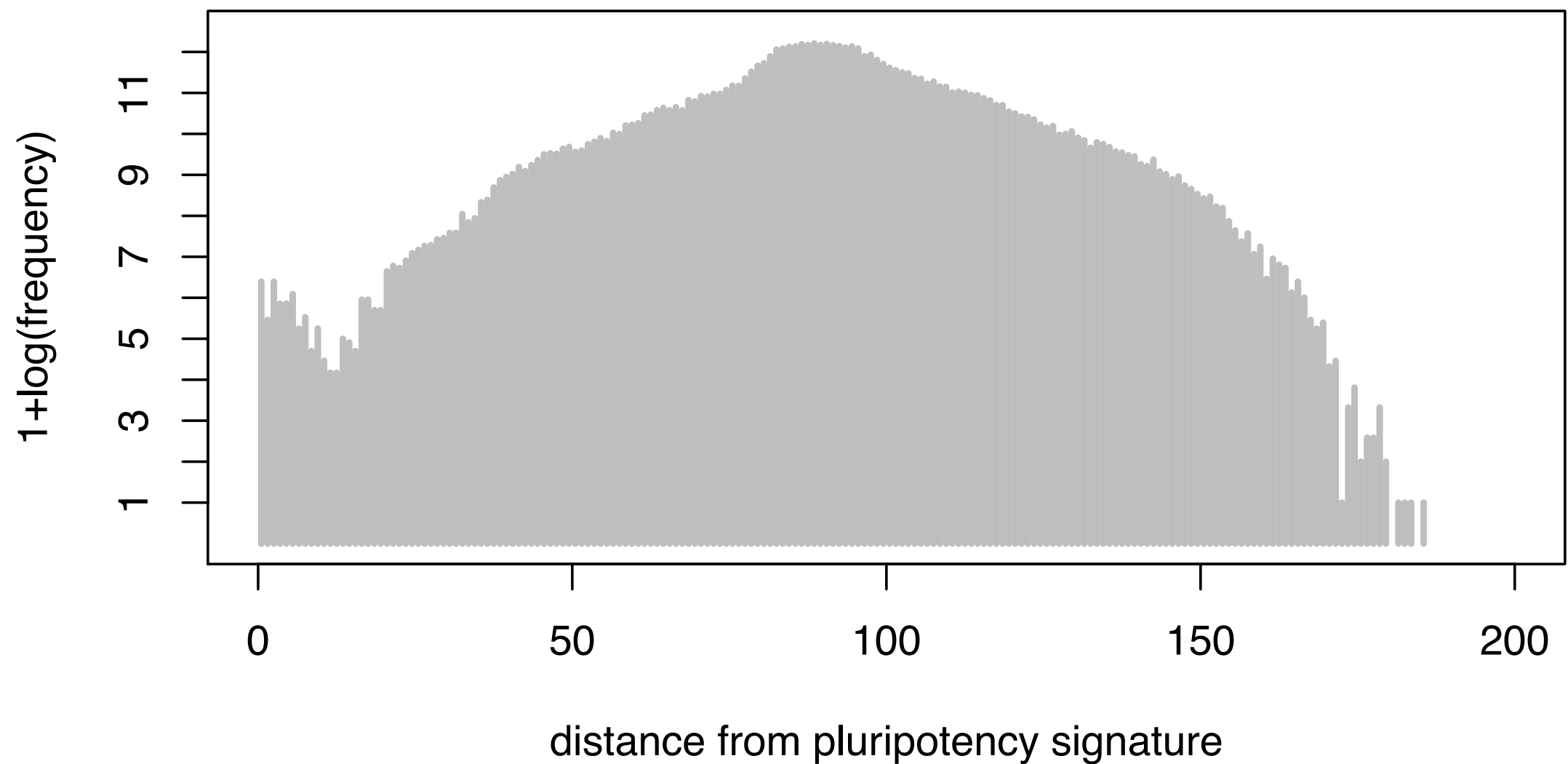
Pathways with high/low expression specific to pluripotent cells

Pluripotent pathway network



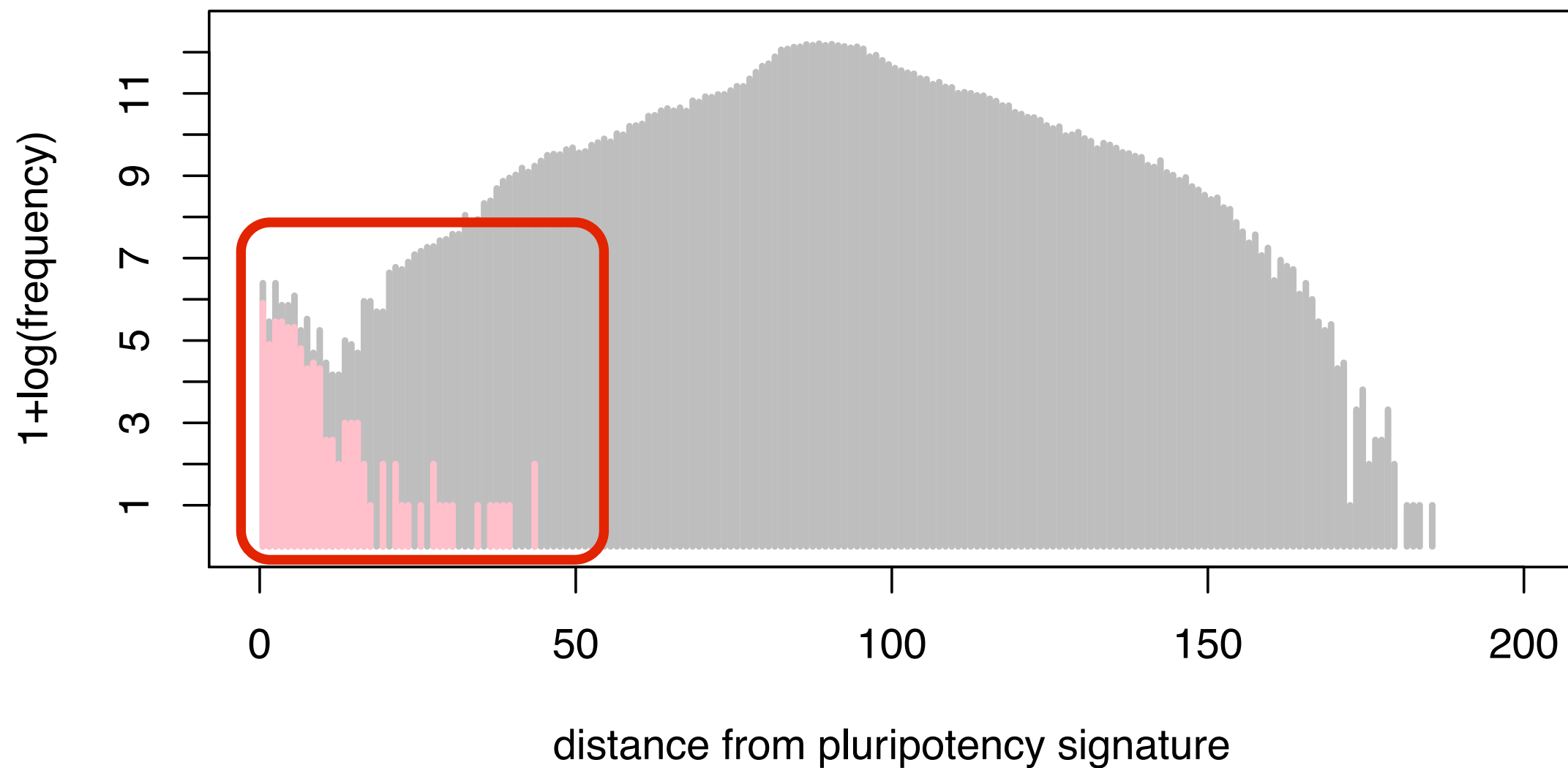
Pathways expressed at high ● and low ● level in pluripotent cells
Edge weight represents number of overlapping genes

Distance from Pluripotency



All GEO

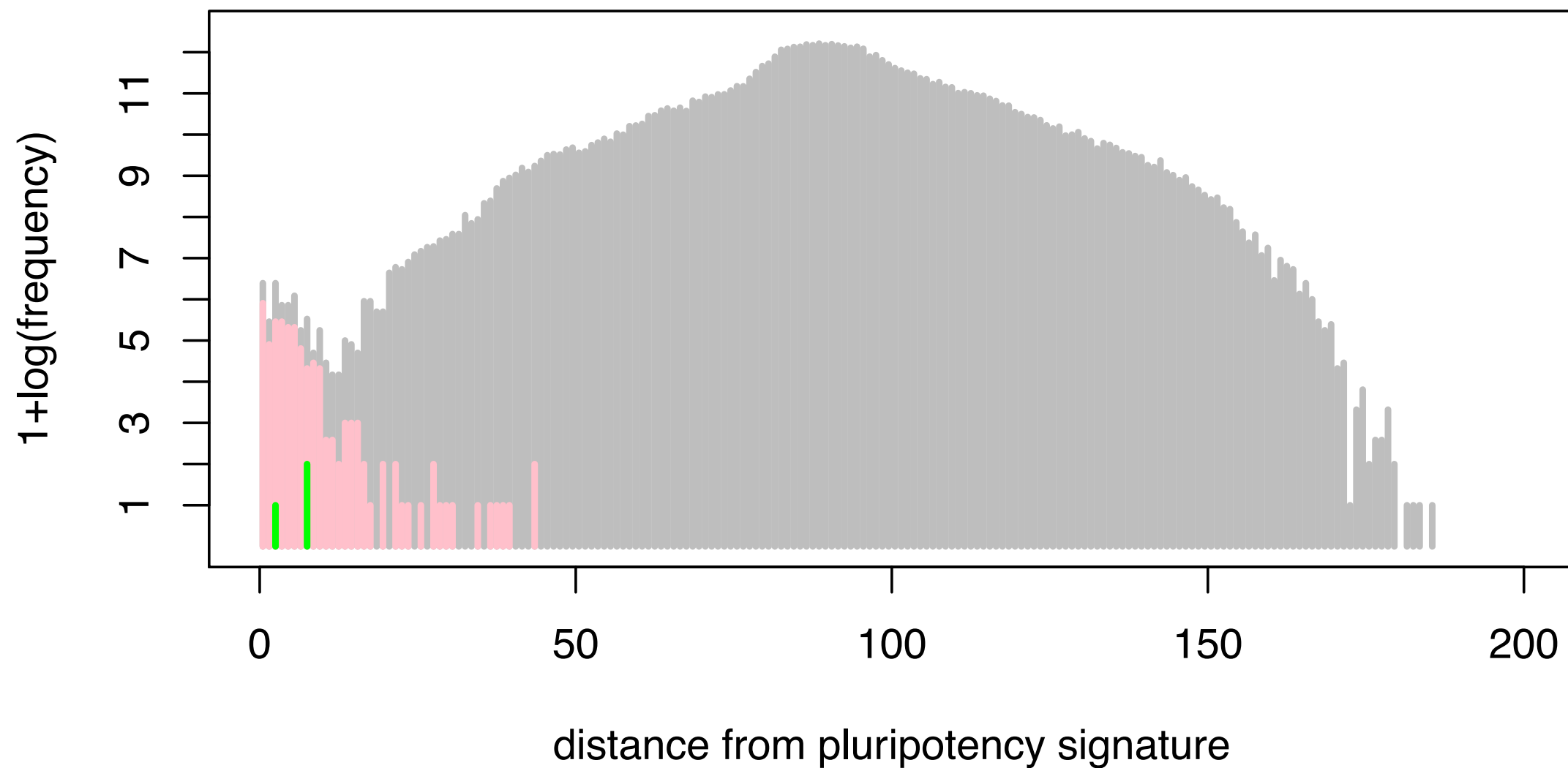
Distance from Pluripotency



All GEO

ES/iPS

Distance from Pluripotency

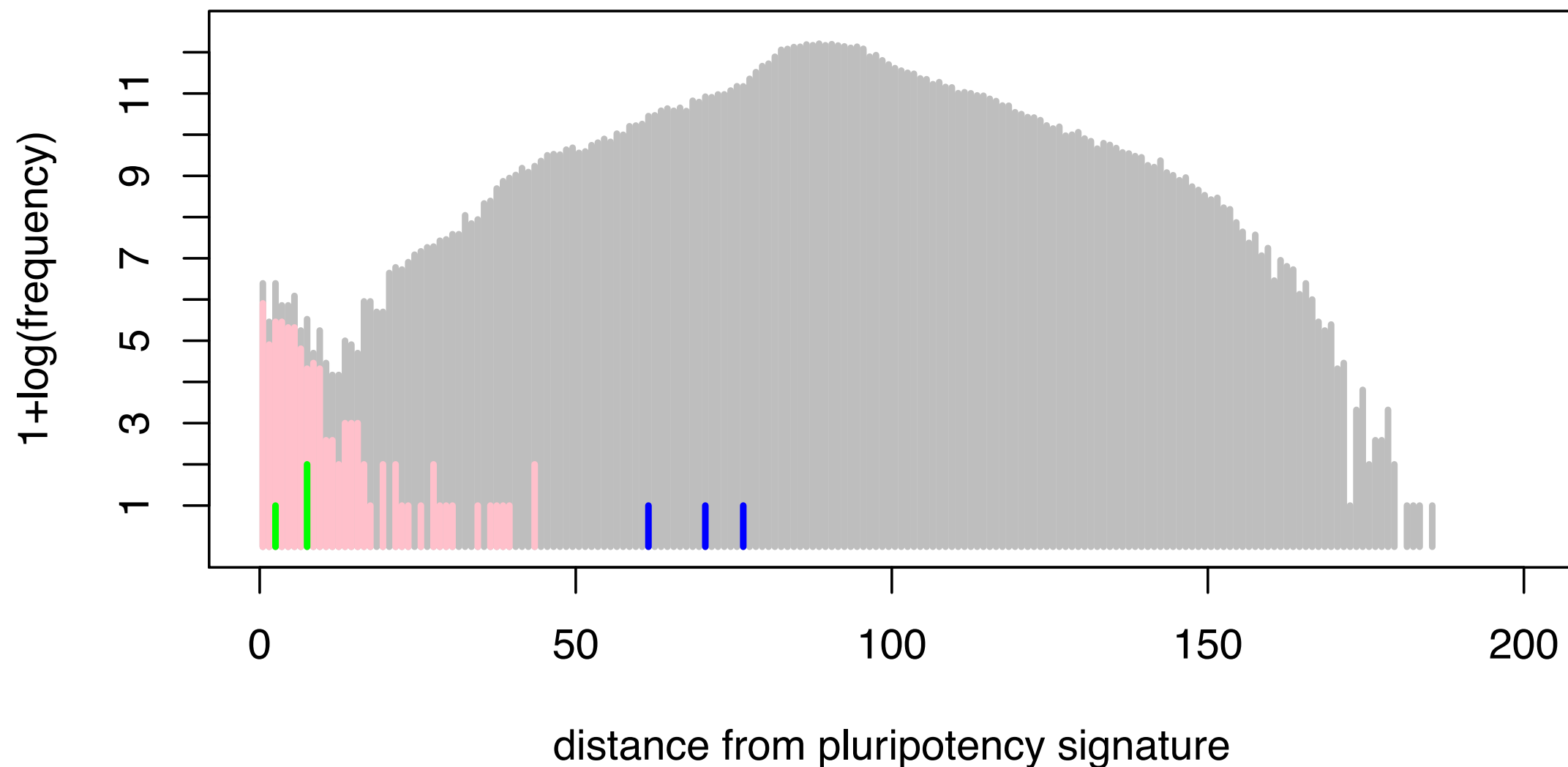


All GEO

ES (Conrad et al)

ES/iPS

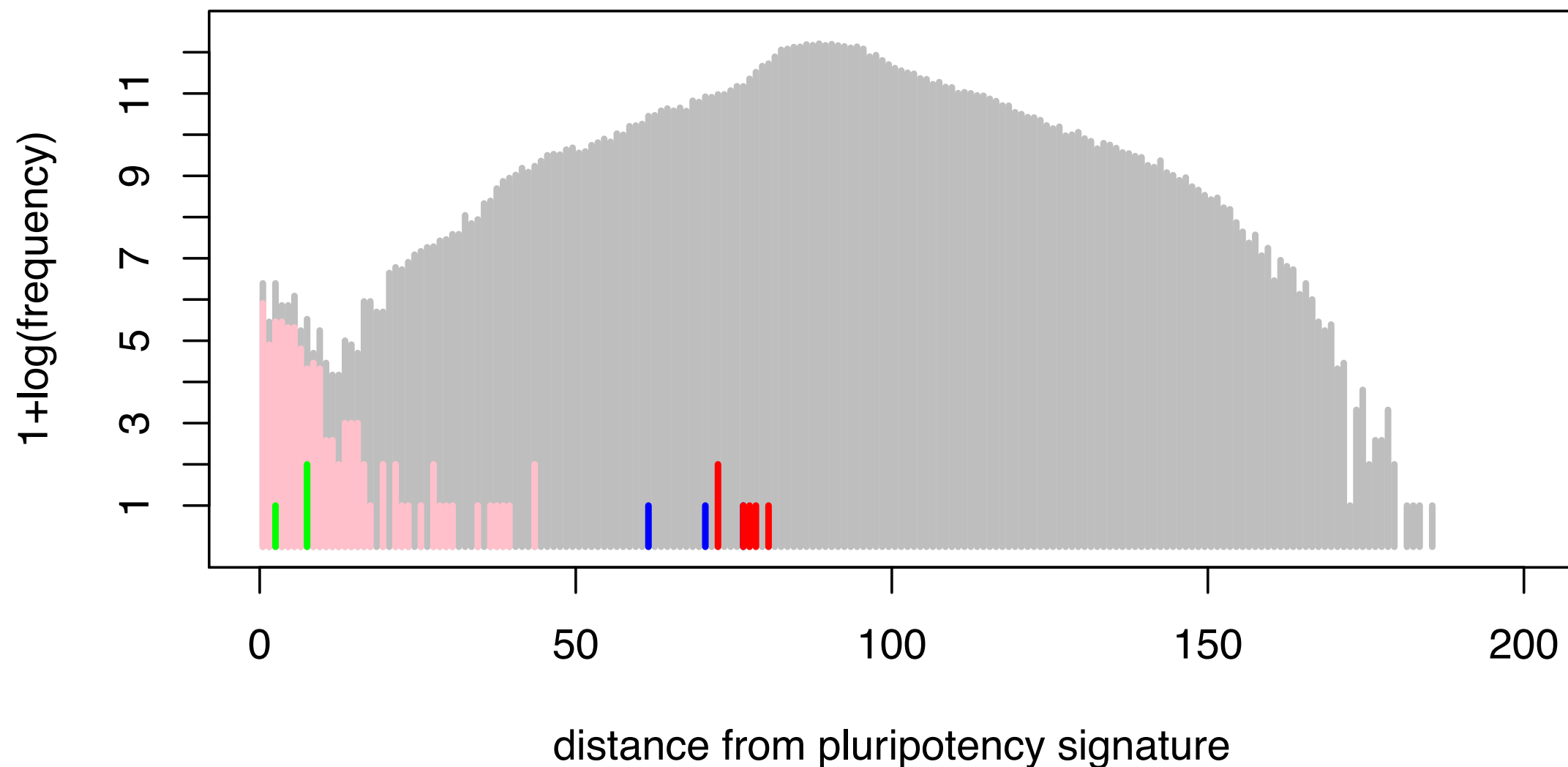
Distance from Pluripotency



All GEO ES (Conrad et al)

ES/iPS SC (Conrad et al)

Distance from Pluripotency



All GEO

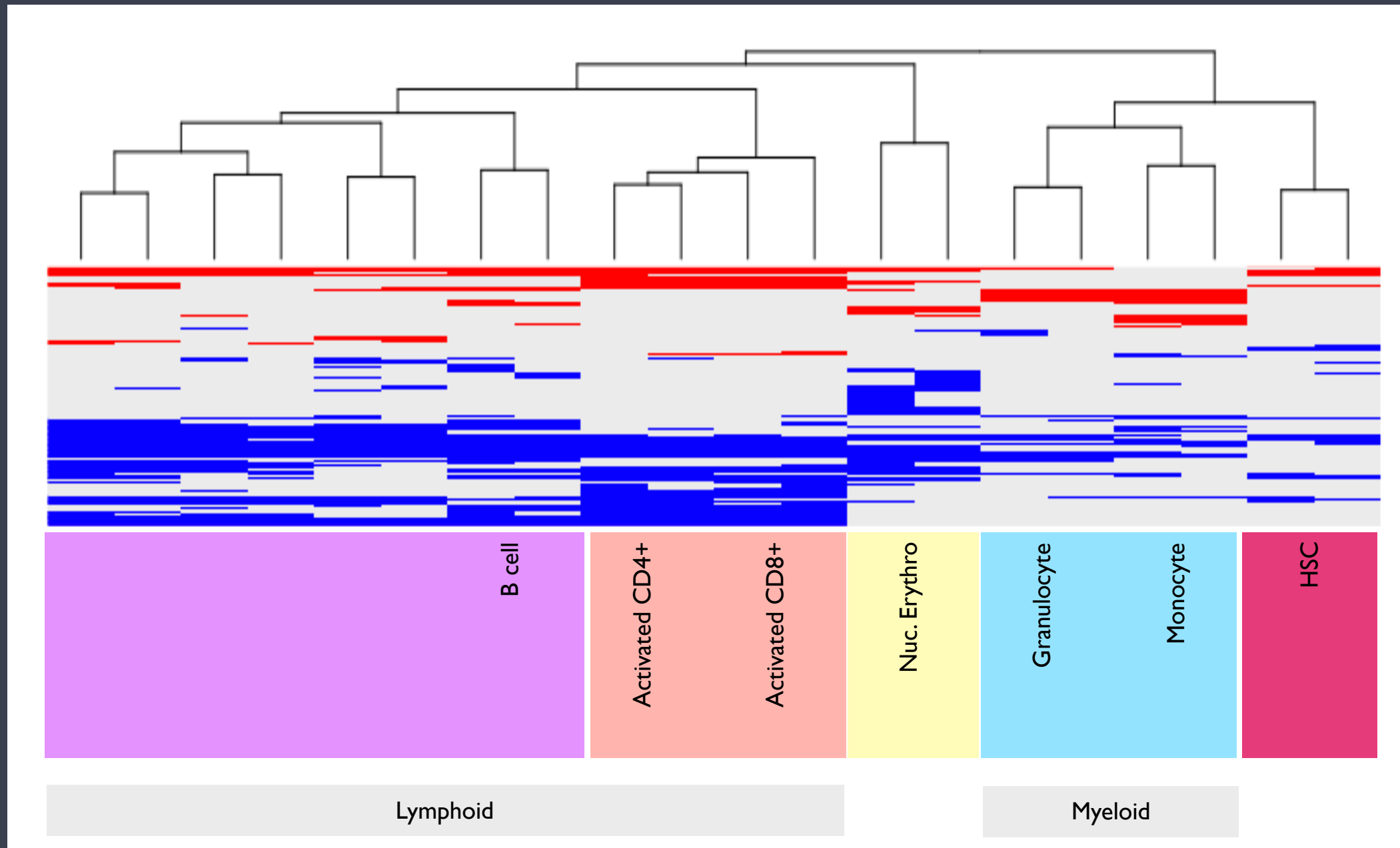
ES (Conrad et al)

haGSC (Conrad et al)

ES/iPS

SC (Conrad et al)

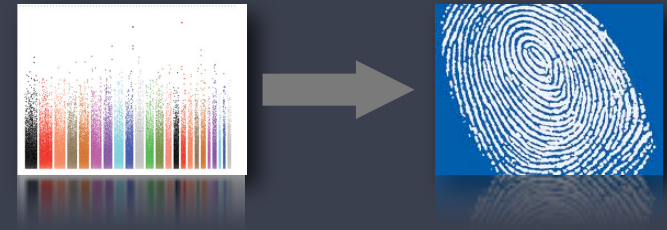
Lineage



Key pathways -> key genes

- Development of a gene-level fingerprint
- Provides a sliding scale of resolution from **pathways to genes**, while maintaining a consistent overall framework
- Identify the **key interactions** that determine the functional profile

Pathway Fingerprint: GWAS



- Mapping **SNPs to pathways** will allow systematic integration between GWA studies and **functional and environmental** context

Case study 2 - NSCLC

- Can the **GWAS-pathway fingerprint** based genetic variants associated with risk of non-small cell lung cancer (**NSCLC**) be systematically linked to **functionally relevant gene-expression studies**
- To what extent can the GEO corpus reveal **environmental** factors relevant to this GWAS.

Fingerprint Integration



CHRNAI FI network

Cytokines, inflammatory response (Wiki)

IL-5 down regulated genes (Netpath)

MLH1 FI network

NAPILI FI network

EPRS FI network

VHL FI network

Proteasome degradation (Wiki)

RPS27A FI network

Fingerprint Integration



CHRNA1 FI network

Cytokines, inflammatory response (Wiki)

IL-5 down regulated genes (Netpath)

MLH1 FI network

NAPILI FI network

EPRS FI network

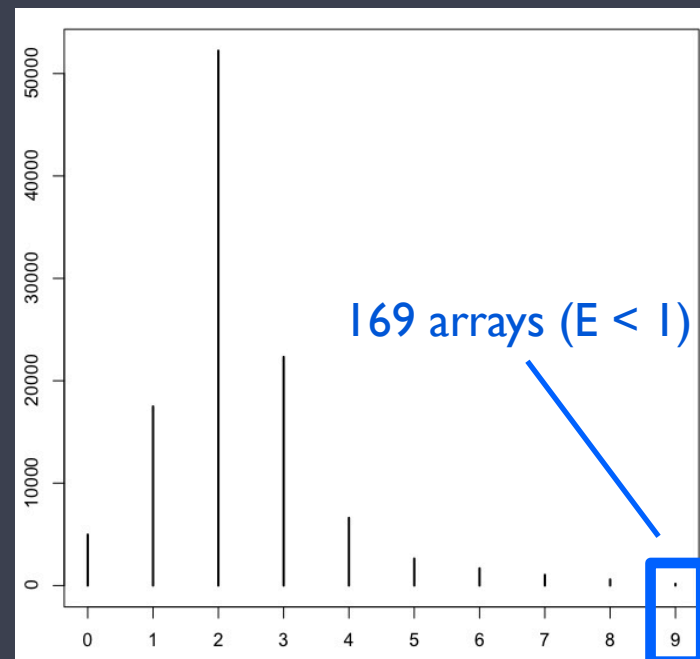
VHL FI network

Proteasome degradation (Wiki)

RPS27A FI network



N arrays in GEO



N significantly expressed
GWAS-fingerprint pathways

Fingerprint Integration



CHRNA1 FI network

Cytokines, inflammatory response (Wiki)

IL-5 down regulated genes (Netpath)

MLH1 FI network

NAPILI FI network

EPRS FI network

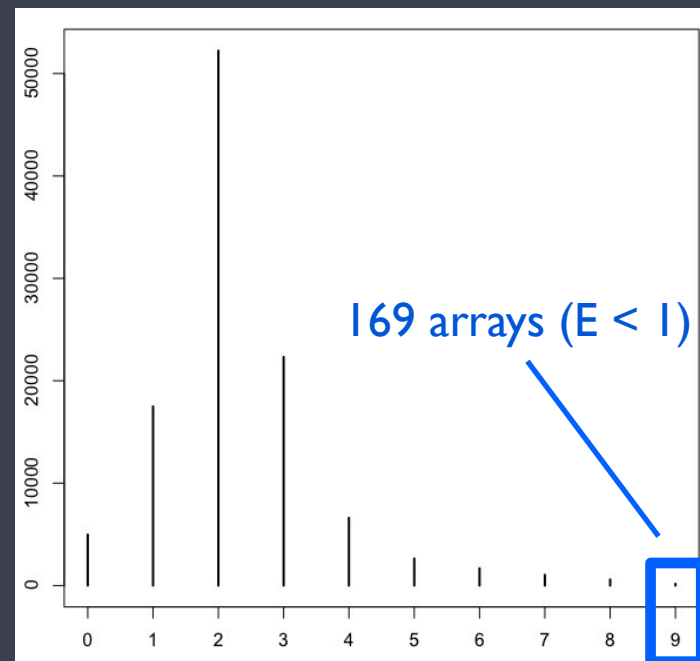
VHL FI network

Proteasome degradation (Wiki)

RPS27A FI network



N arrays in GEO



N significantly expressed
GWAS-fingerprint pathways



Disease	GEO Accession
lung adenocarcinoma	GSE17475
early stage NSCLC	GSE19188
Cell lines and tissues	
human cancer cell lines	GSE5823
lymph node biopsies	GSE6338
human pre-frontal cortex,liver, and colon tissues and colon tumors	GSE13471
Environment	
effect of alcohol on human airway epithelial cells	GSE12253
allergen exposure in mouse lung	GSE6858
Vancomycin nephrotoxicity study in mouse kidney	GSE7793

Fingerprint Integration



CHRNA1 FI network

Cytokines, inflammatory response (Wiki)

IL-5 down regulated genes (Netpath)

MLH1 FI network

NAPILI FI network

EPRS FI network

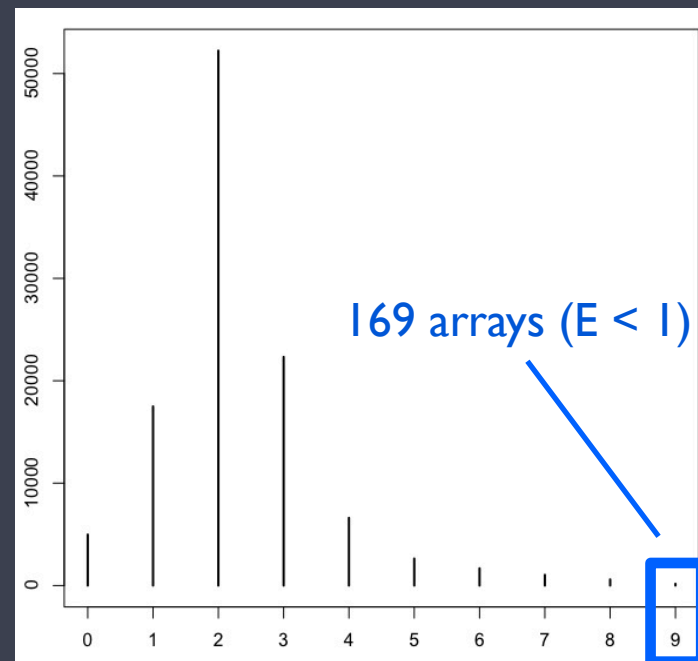
VHL FI network

Proteasome degradation (Wiki)

RPS27A FI network



N arrays in GEO



N significantly expressed
GWAS-fingerprint pathways



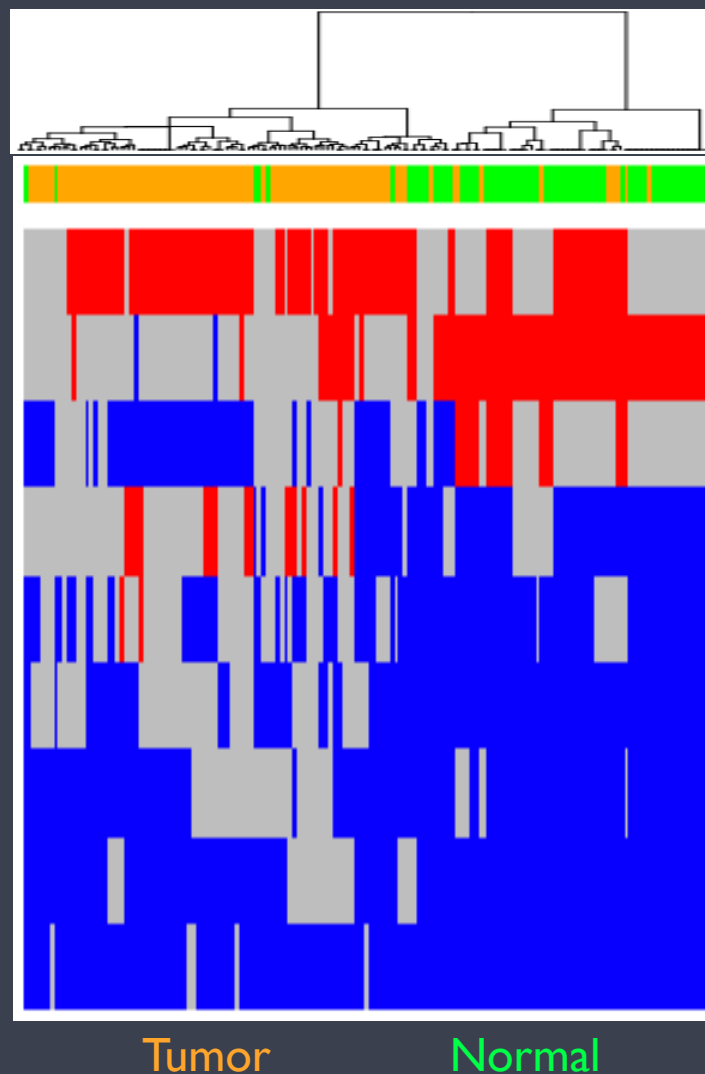
Disease	GEO Accession
lung adenocarcinoma	GSE17475
early stage NSCLC	GSE19188
Cell lines and tissues	
human cancer cell lines	GSE5823
lymph node biopsies	GSE6338
human pre-frontal cortex,liver, and colon tissues and colon tumors	GSE13471
Environment	
effect of alcohol on human airway epithelial cells	GSE12253
allergen exposure in mouse lung	GSE6858
Vancomycin nephrotoxicity study in mouse kidney	GSE7793

Fingerprint Integration



GWAS fingerprint
pathways ($p < 0.02$)

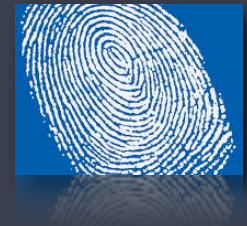
CHRNA1 FI network
Cytokines, inflammatory response (Wiki)
IL-5 down regulated genes (Netpath)
MLH1 FI network
NAPILI FI network
EPRS FI network
VHL FI network
Proteasome degradation (Wiki)
RPS27A FI network



Gene expression
fingerprint scores

GWAS pathways
distinguish normal vs
tumor arrays

Pathway Fingerprint: CAGE



We have begun to systematically map CAGE data to the pathway fingerprint to provide

- 1) a consistent quantitative **functional annotation** of cell types
- 2) a framework for **data integration** across the CAGE libraries and within the existing public data corpus

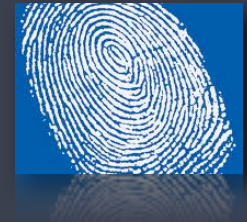
Case study 3 - Tissue specificity

FANTOM5 release 10 expression data

CTSS 500bp proximal to RefSeq
Merged to gene level

Comparison of tissue data

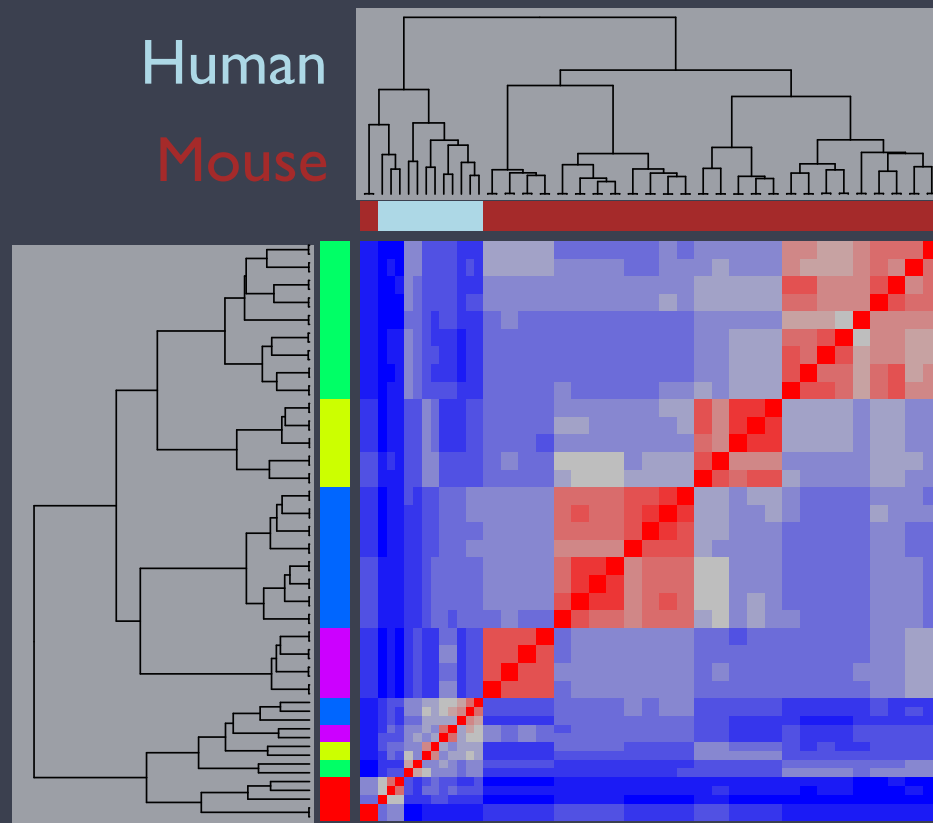
Cross-species clustering of tissues



Clustering by **genes**

Human

Mouse

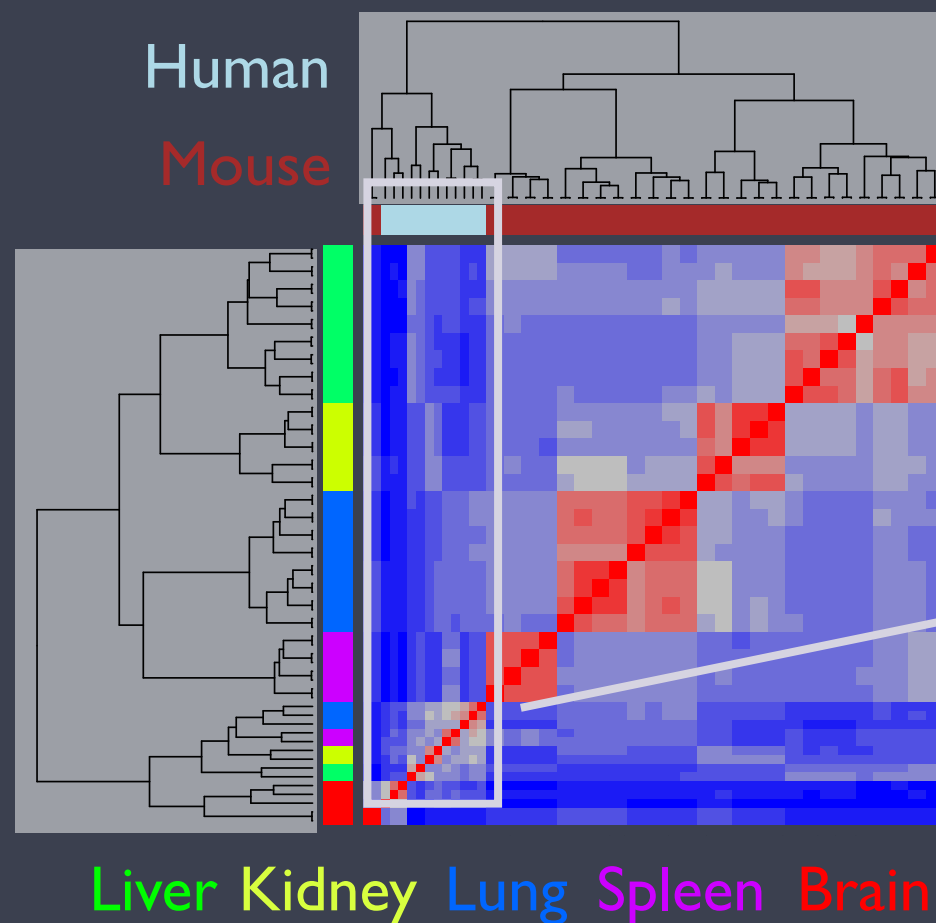


Liver Kidney Lung Spleen Brain

Cross-species clustering of tissues

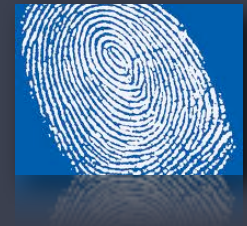


Clustering by **genes**

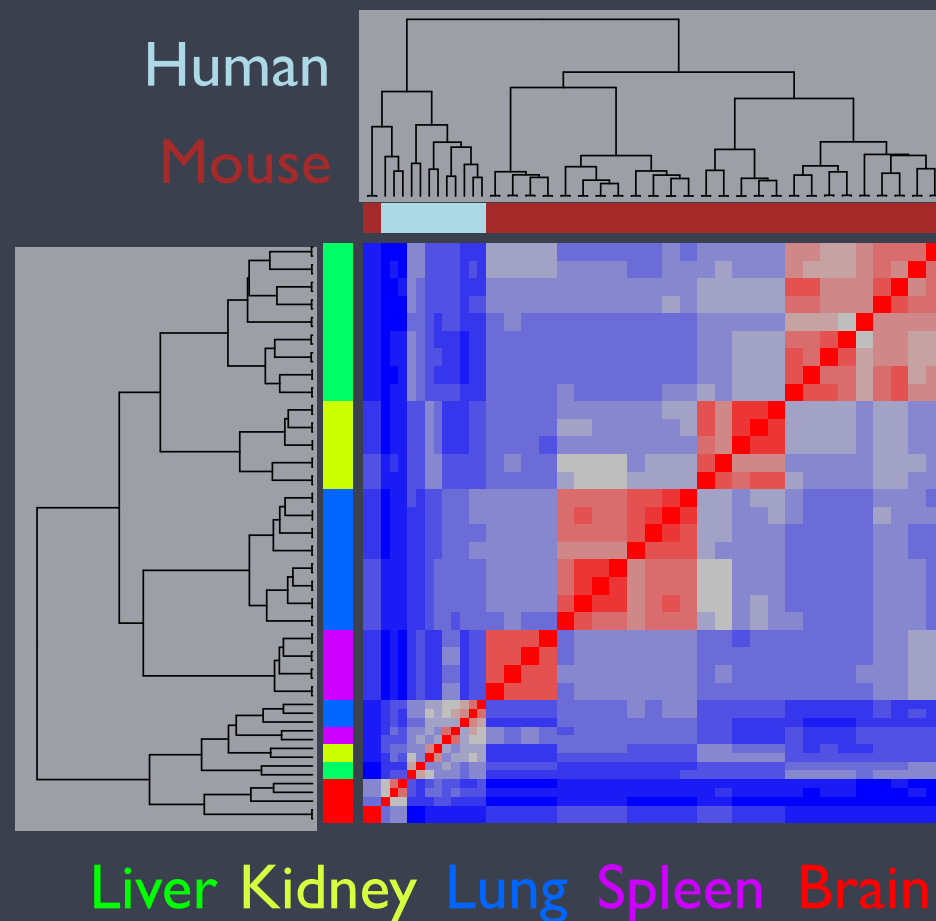


Samples cluster by **species**
rather than tissue type

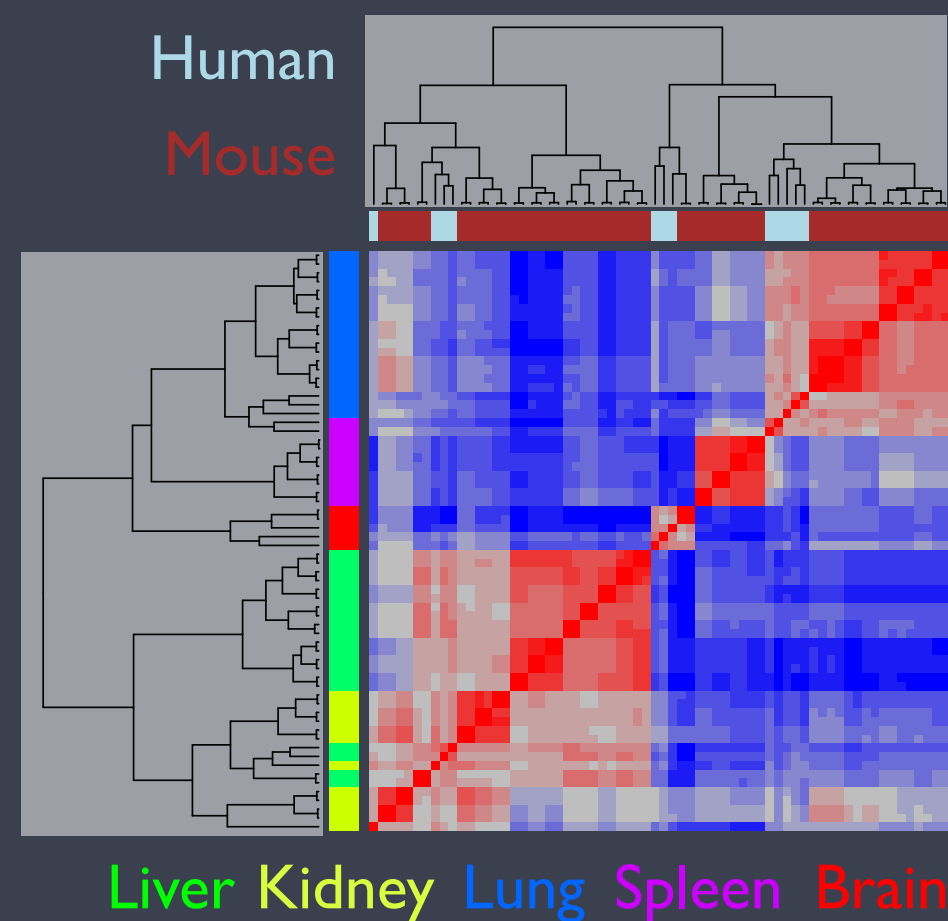
Cross-species clustering of tissues



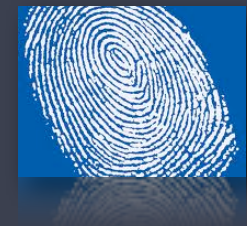
Clustering by **genes**



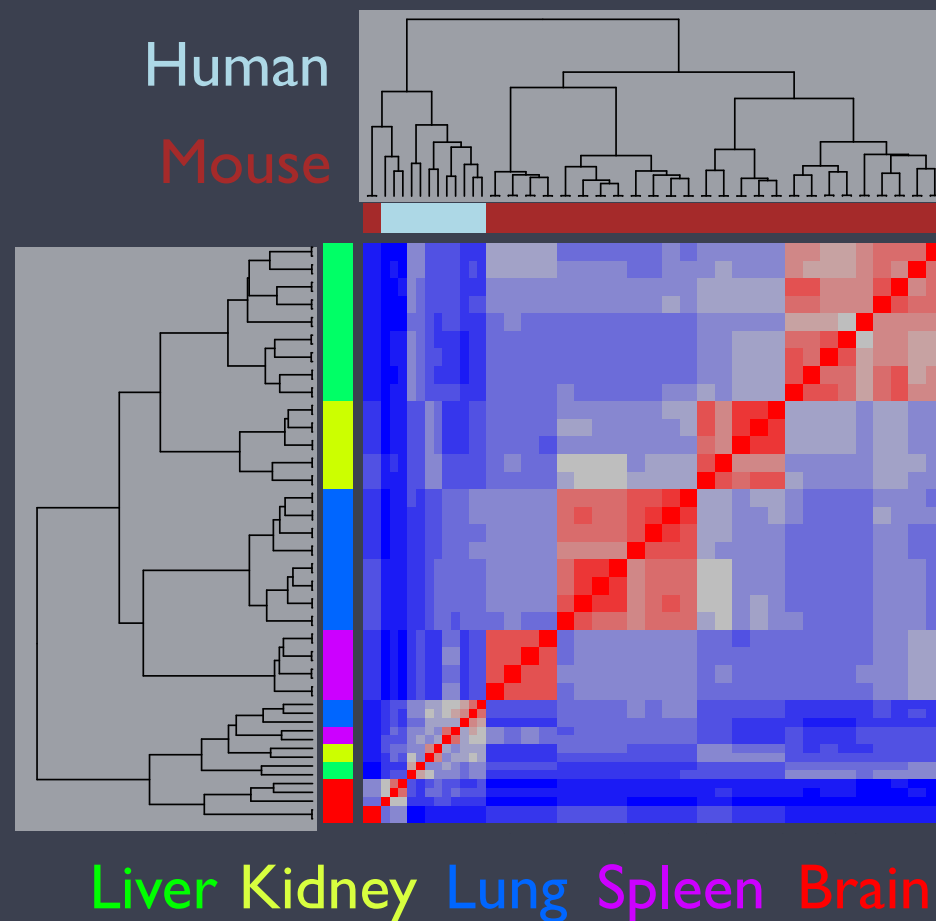
Clustering by **pathway**



Cross-species clustering of tissues

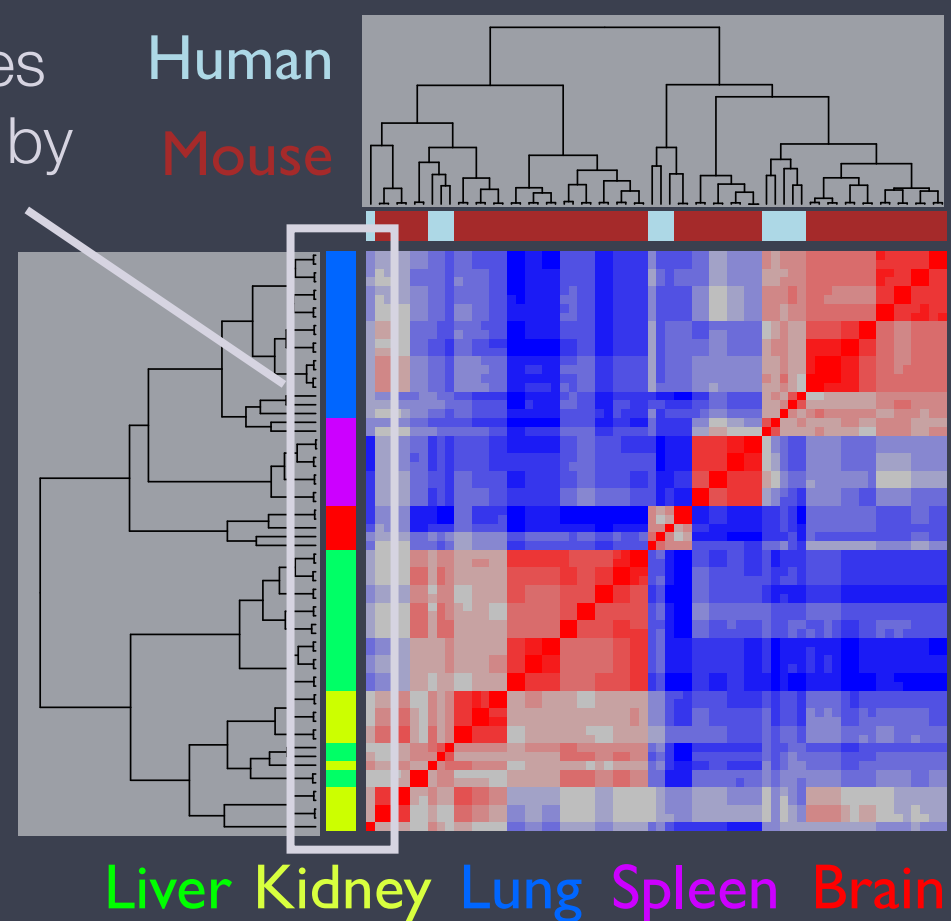


Clustering by **genes**

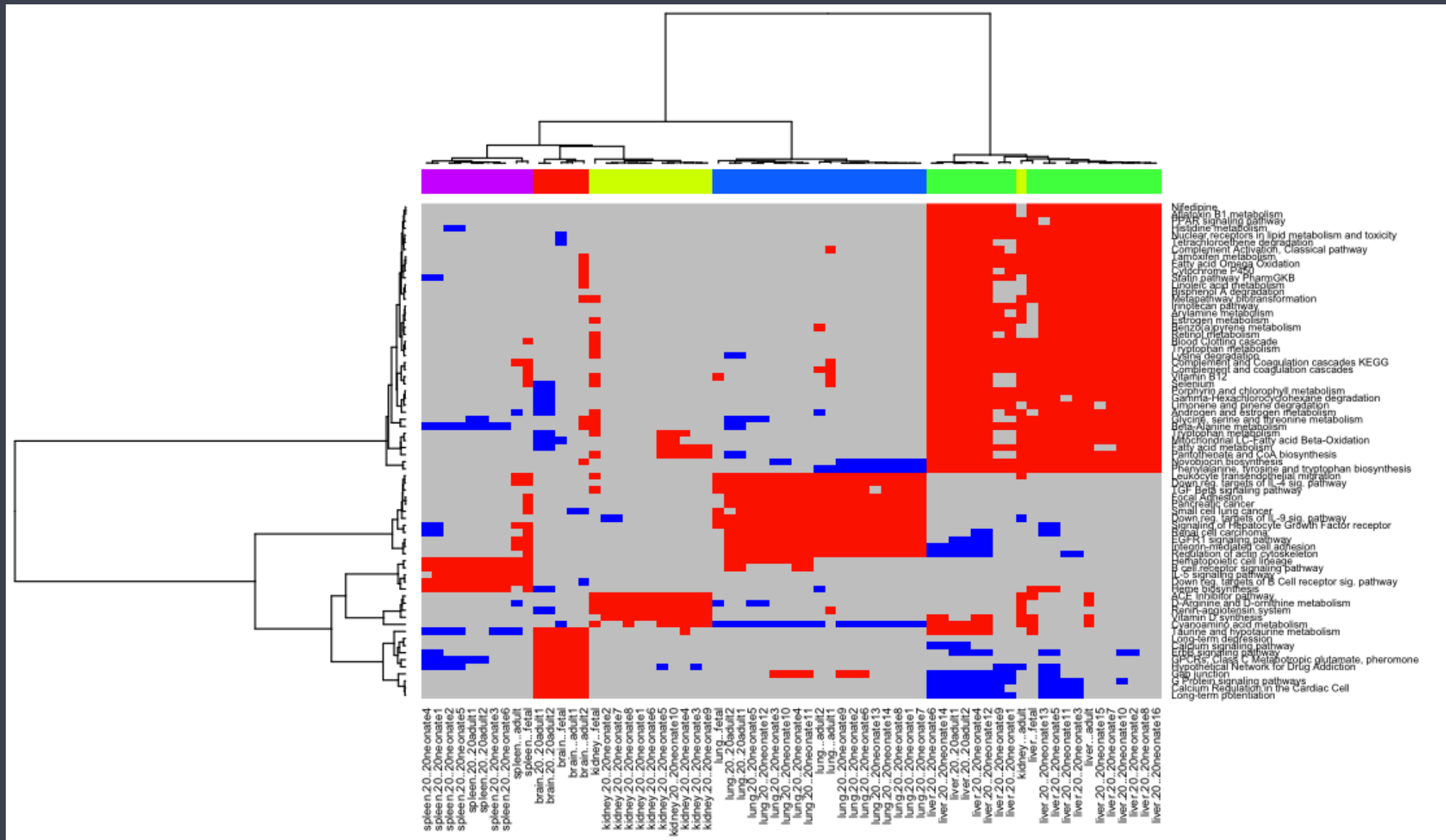


Samples
cluster by
tissue

Clustering by **pathway**



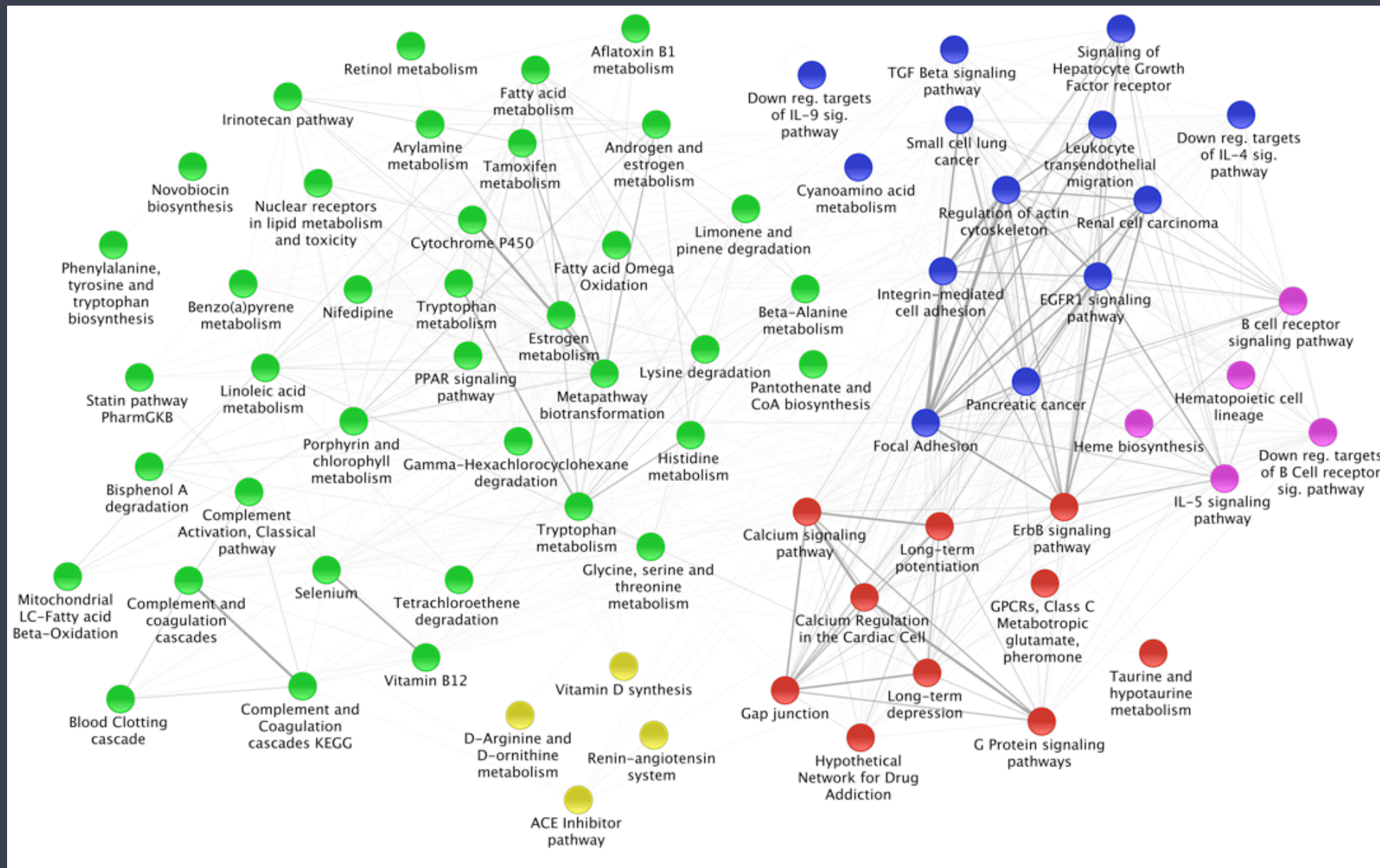
Cell-type specific pathways



Pathways

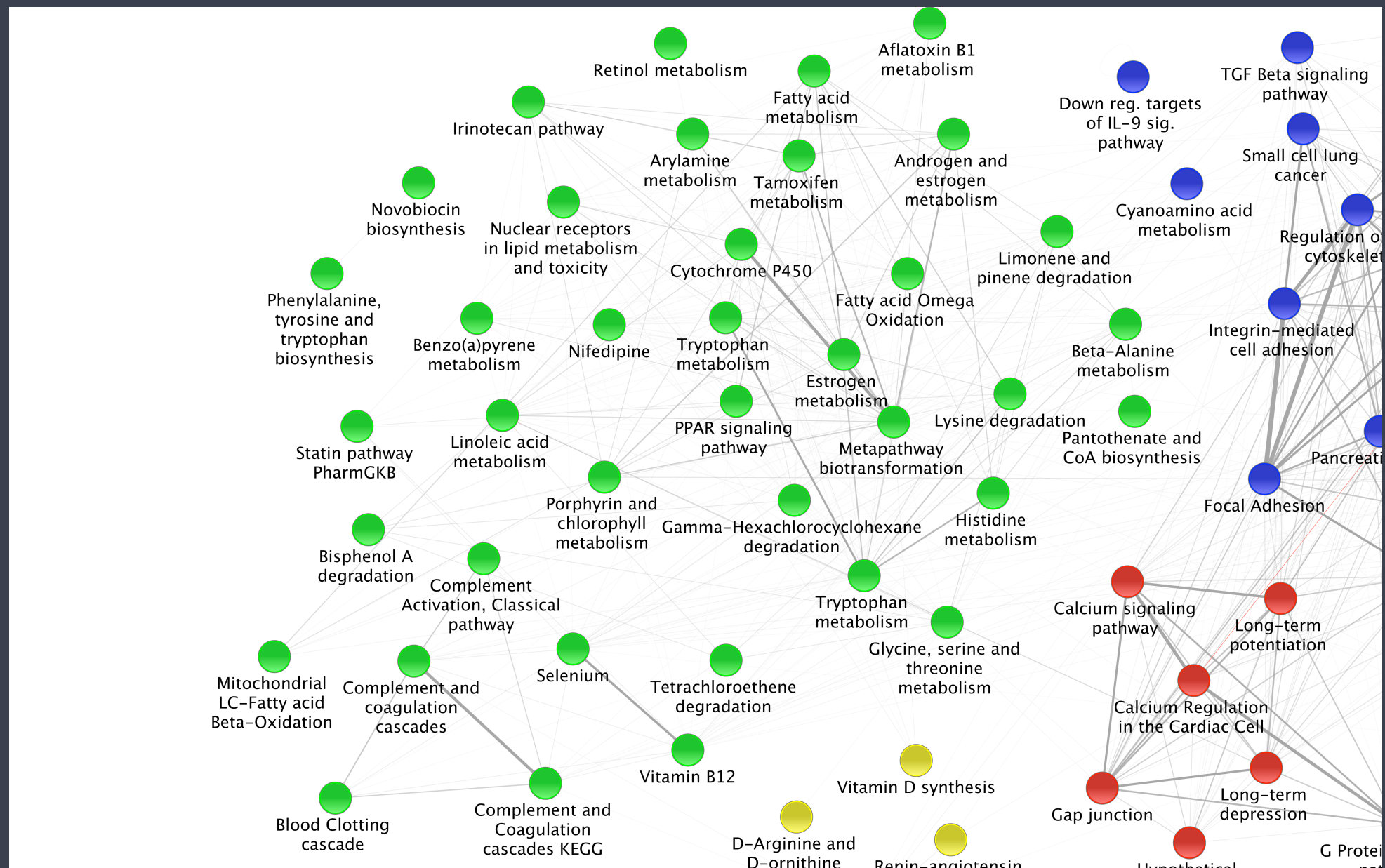
Samples **Liver** **Kidney** **Lung** **Spleen** **Brain**

Cell-type specific pathway map



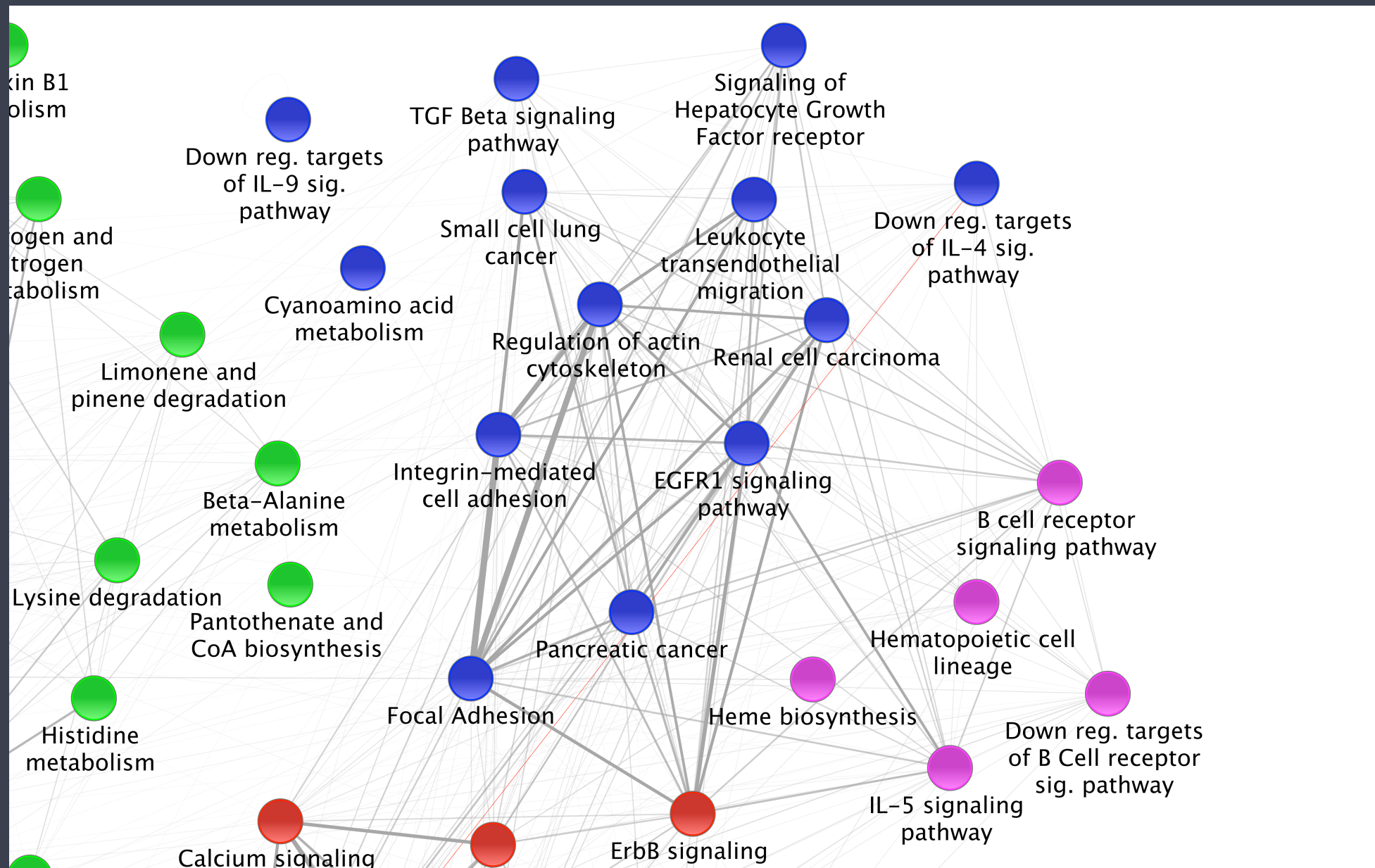
Liver Kidney Lung Spleen Brain

Cell-type specific pathway map



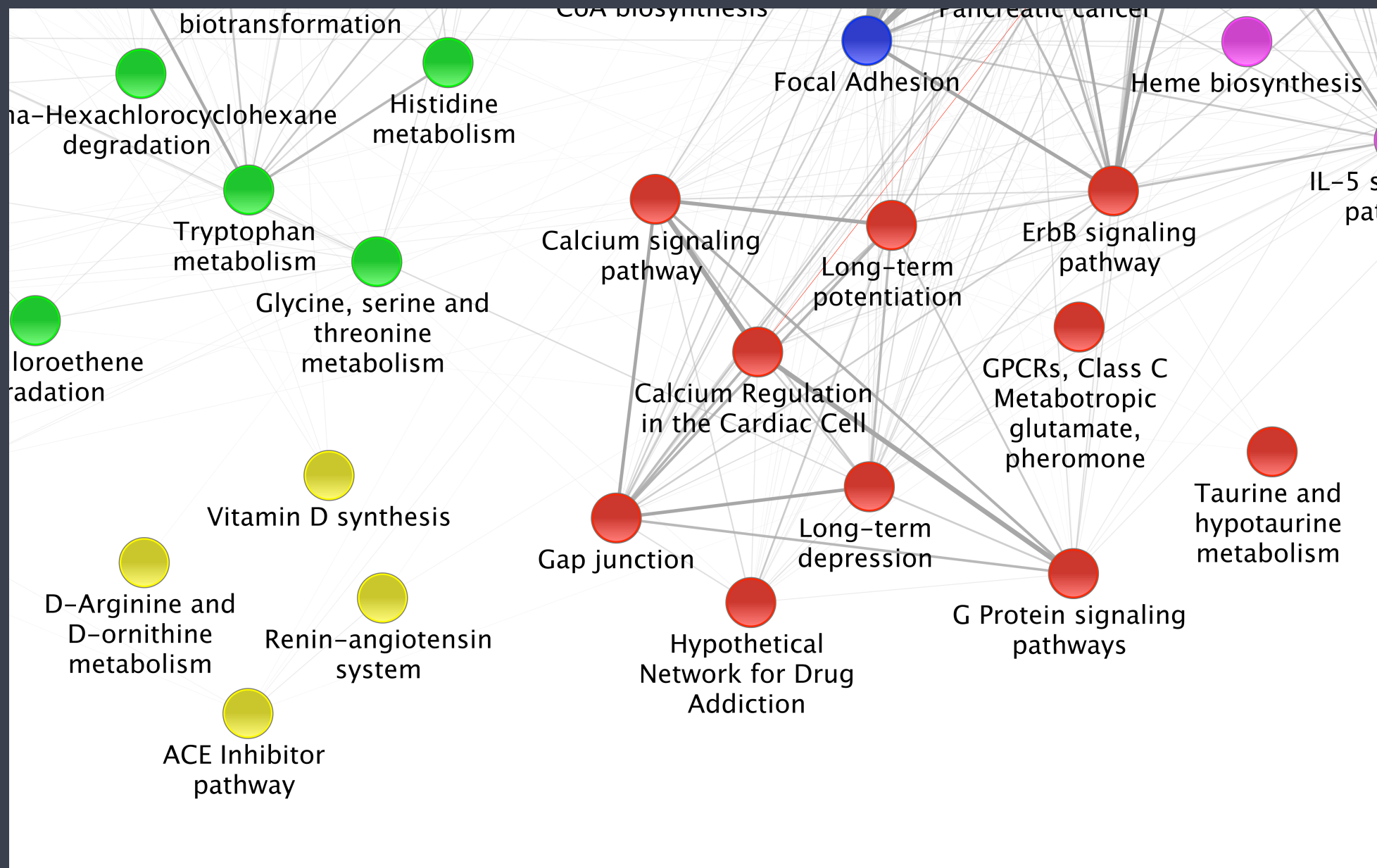
Liver Kidney Lung Spleen Brain

Cell-type specific pathway map



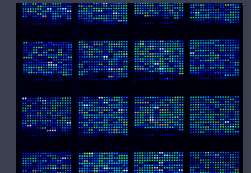
Liver Kidney Lung Spleen Brain

Cell-type specific pathway map



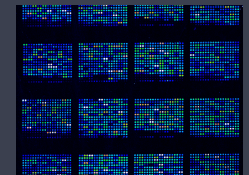
Liver Kidney Lung Spleen Brain

Integrating the expression corpus



HGU133 Plus 2.0 (n = 64)	Human CAGE (n = 12)
HGU133A (n = 103)	Mouse CAGE (n = 54)
MGU430 2.0 (n = 341)	

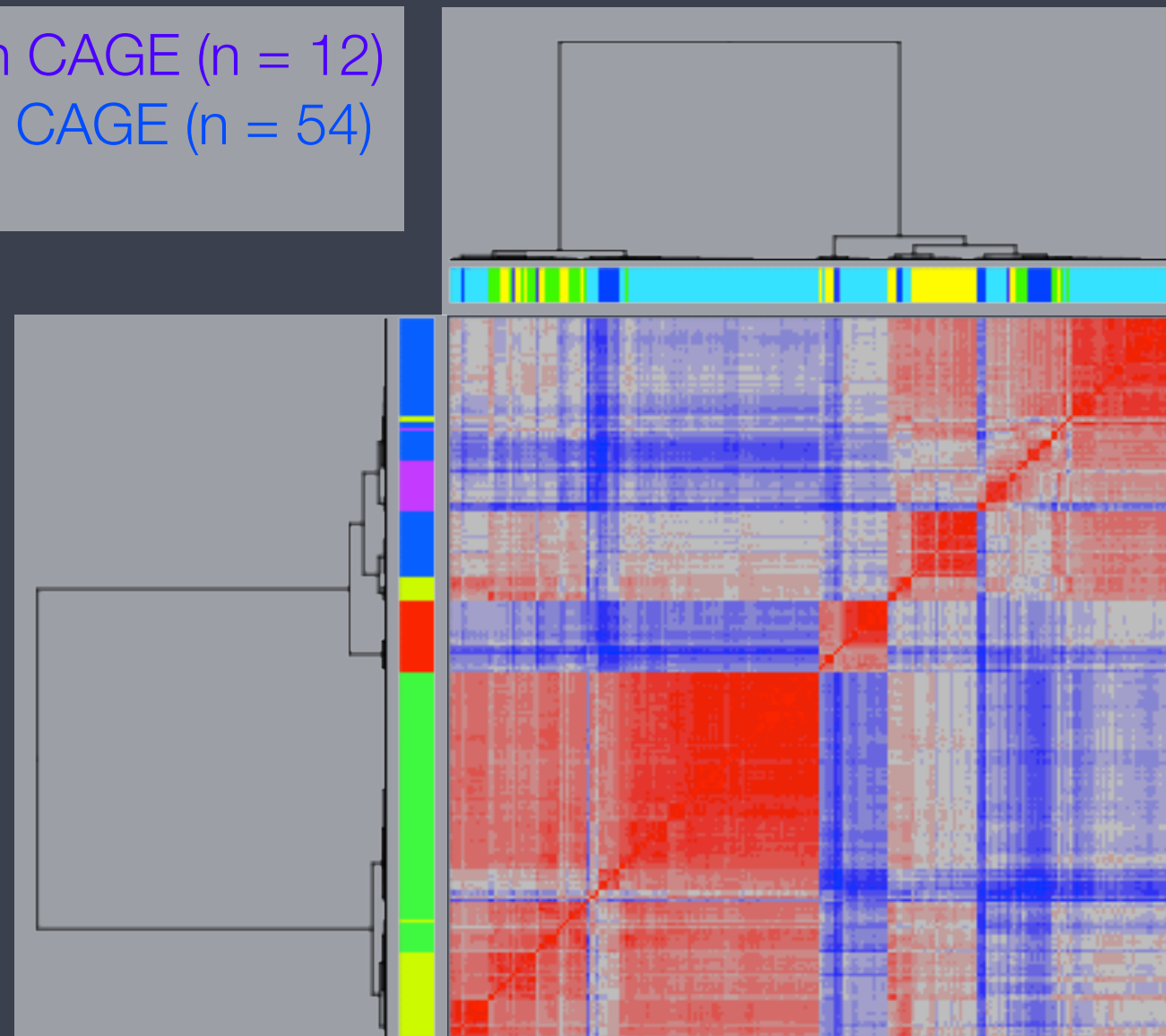
Integrating the expression corpus



HGU133 Plus 2.0 (n = 64) Human CAGE (n = 12)
HGU133A (n = 103) Mouse CAGE (n = 54)
MGU430 2.0 (n = 341)

Liver Kidney Lung Spleen Brain

CAGE and expression array data can be combined at the **functional level**



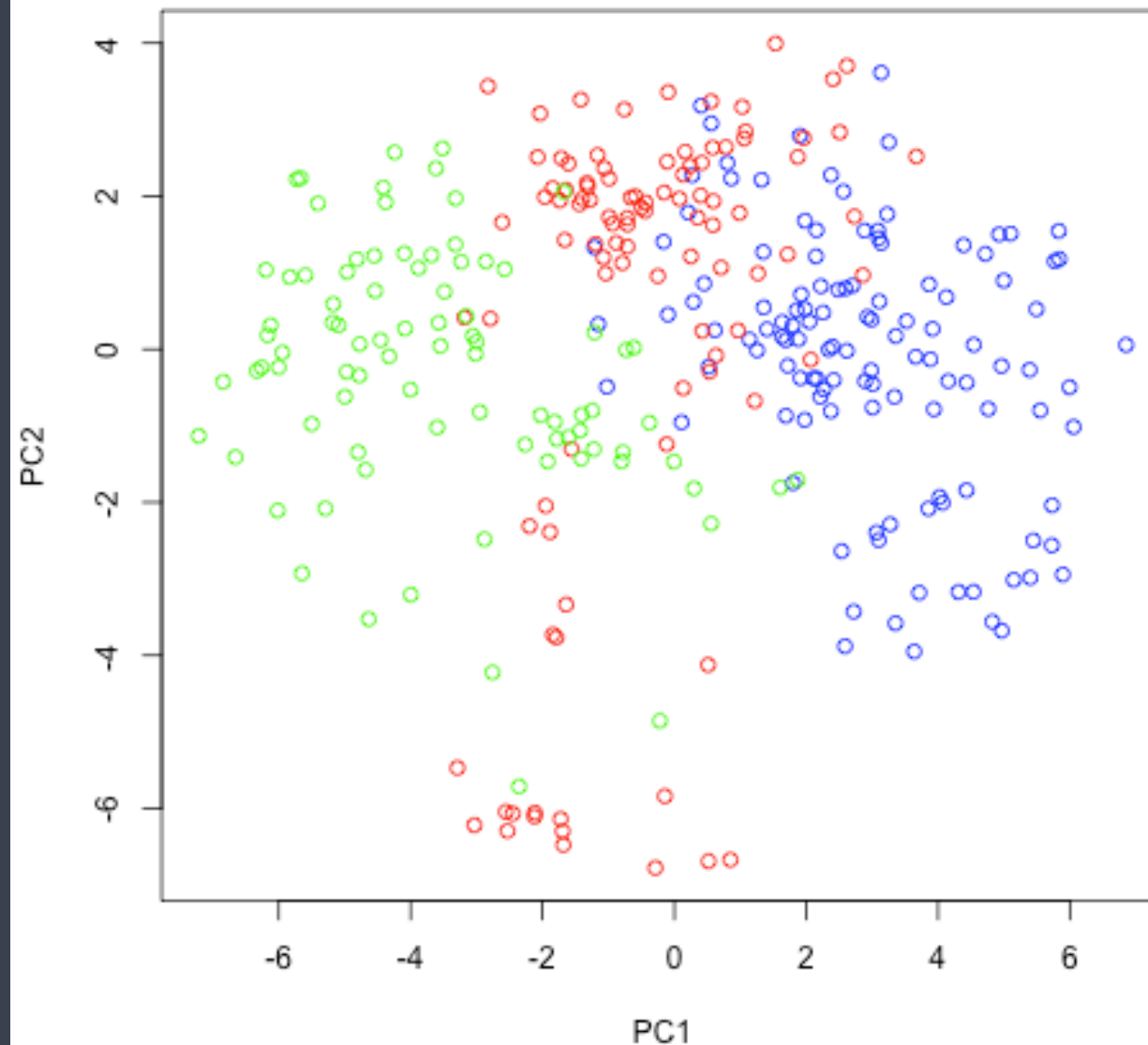
Human cell types

Pathway-based PCA
of full human release
10 dataset

Cell lines

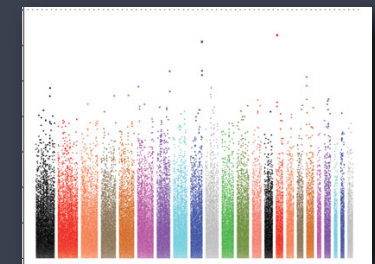
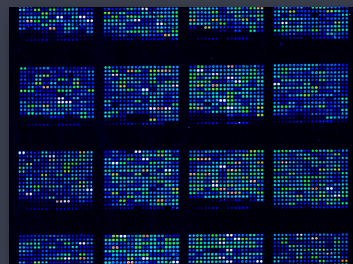
Primary cells

Tissue



The Pathway Fingerprint

Data dissemination



The Pathway Fingerprint Data dissemination



- ▶ **Functional integration:** Informative pathways for all CAGE samples
- ▶ **Model prioritization:** Contributing genes for modeling of TF networks
- ▶ **Network validation:** Include TF networks /TF targets into fingerprint

Acknowledgements

Win Hide



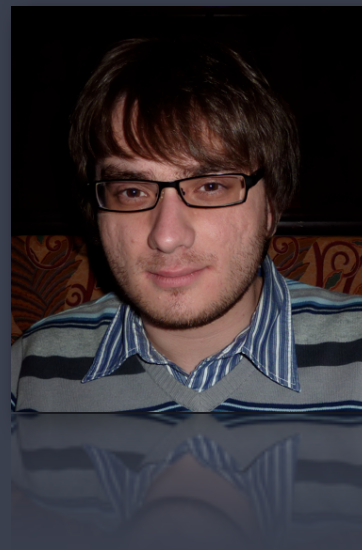
Rama Sompallae



Oliver Hofmann



Emmanuel Dimont



Shannan Ho Sui



Jieun Jeong

