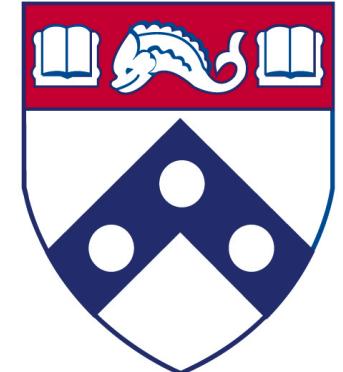
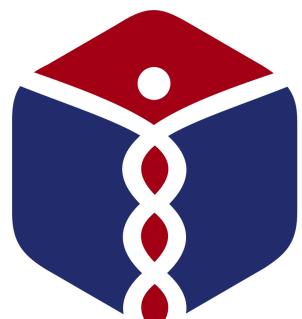


Unsupervised discovery of leukemogenesis-cooperating genes in *de novo* acute myeloid leukemia



Perelman
School of Medicine
UNIVERSITY of PENNSYLVANIA



**Institute for
Biomedical
Informatics**

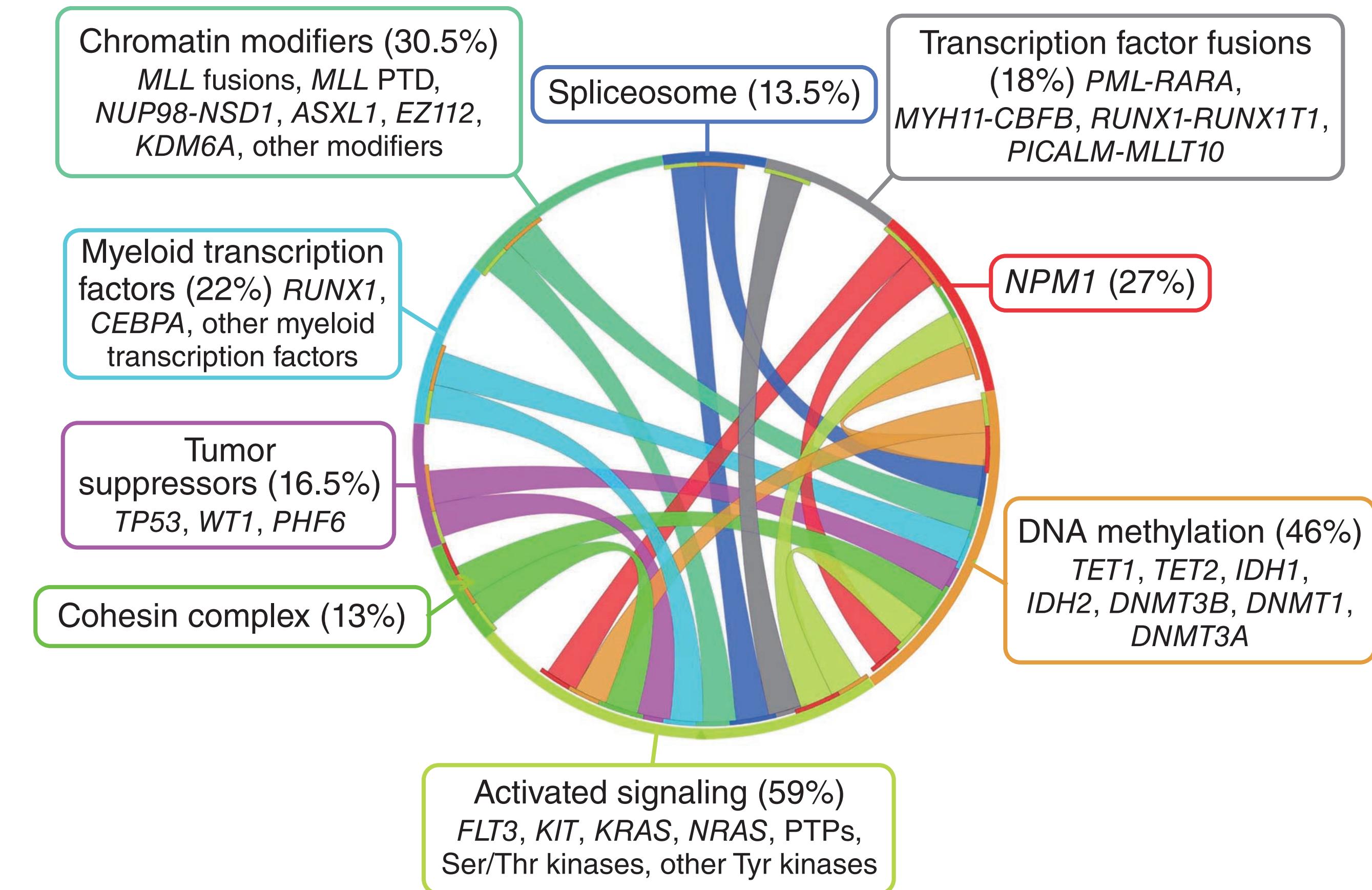
Osvaldo D. Rivera
Cancer Biology Graduate Student
Biomedical Graduate Studies

Contents

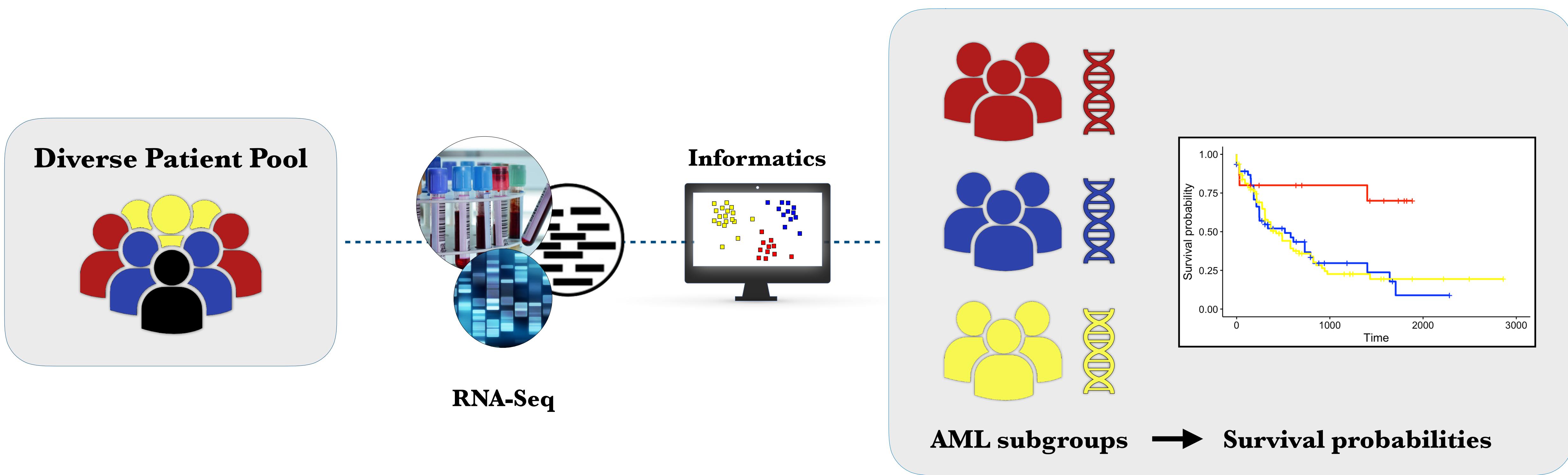
- Acute Myeloid Leukemia
- Question
- Workflow
- Results
 - Gene Expression Clustering
 - Survival Analysis
 - Feature Selection
 - Gene ontology analysis

Acute myeloid leukemia

- Acute myeloid leukemia (AML) is categorized as a highly heterogeneous disease.
- Considerable molecular diversity leads to distinct pathogenesis and clinical outcomes.
- Poor prognosis (~10,000 deaths; 5 year survival <25%)
- Very low mutation burden compared to other tumors. ~ 40 recurrent mutations in AML with only 1-3 mutations per patient • *FLT3* and *IDH1/2* treatable; not curable
- Molecular alterations have been found to potentially affect different layers of biological regulation.



Are there underlying gene expression patterns associated with higher survivability?





THE CANCER GENOME ATLAS

National Cancer Institute
National Human Genome Research Institute

**NATIONAL CANCER INSTITUTE
GDC Data Portal**

[Home](#) [Projects](#) [Exploration](#) [Analysis](#) [Repository](#) [Quick Search](#) [Manage Sets](#) [Login](#) [Cart 0](#) [GDC Apps](#)

Harmonized Cancer Datasets

Genomic Data Commons Data Portal

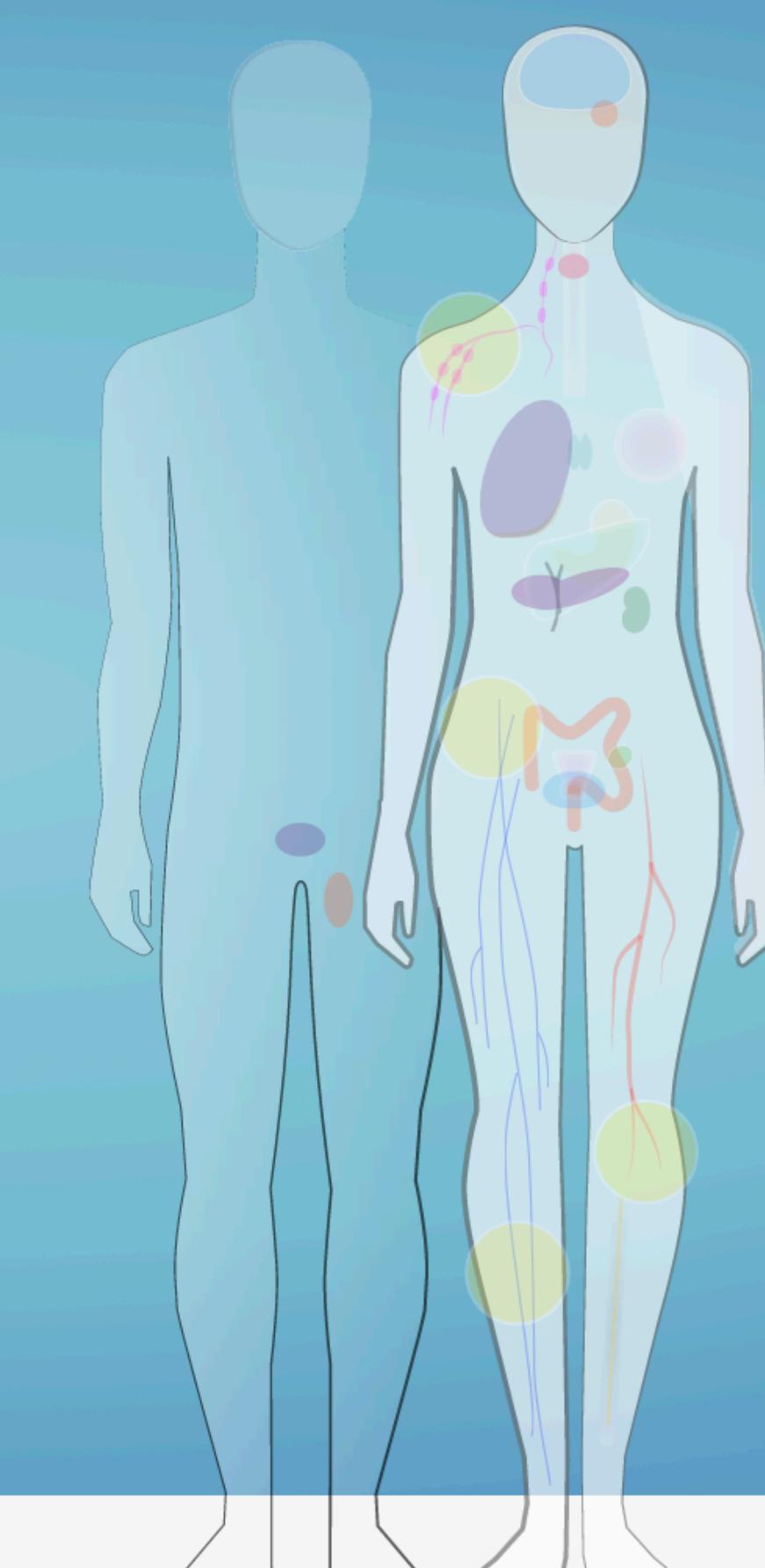
Get Started by Exploring:

[Projects](#) [Exploration](#) [Analysis](#) [Repository](#)

Q e.g. BRAF, Breast, TCGA-BLCA, TCGA-A5-A0G2

Data Portal Summary Data Release 9.0 - October 24, 2017

PROJECTS	PRIMARY SITES	CASES
40	60	32,555
FILES	GENES	MUTATIONS
310,858	22,144	3,115,606

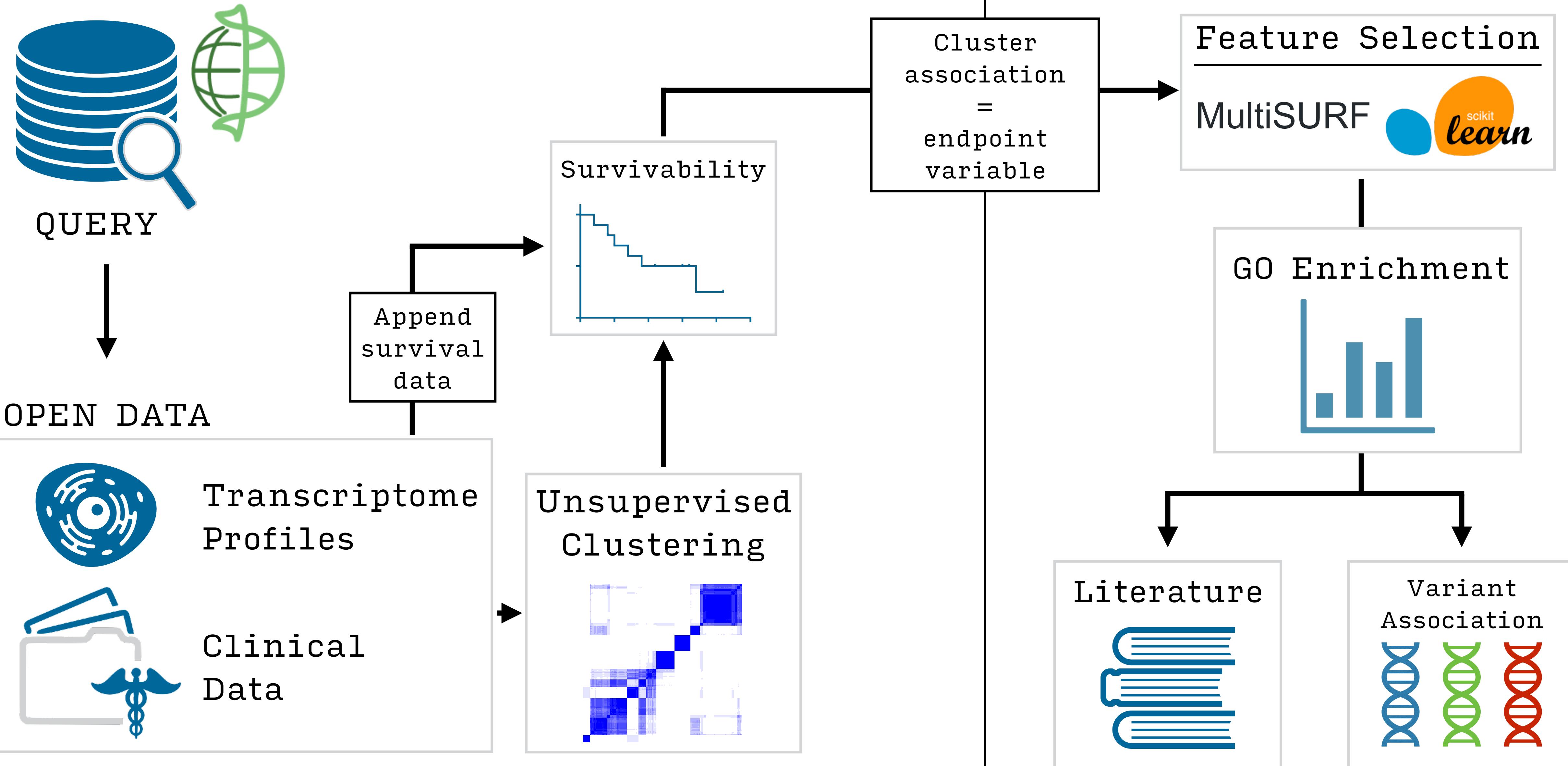


Cases by Major Primary Site

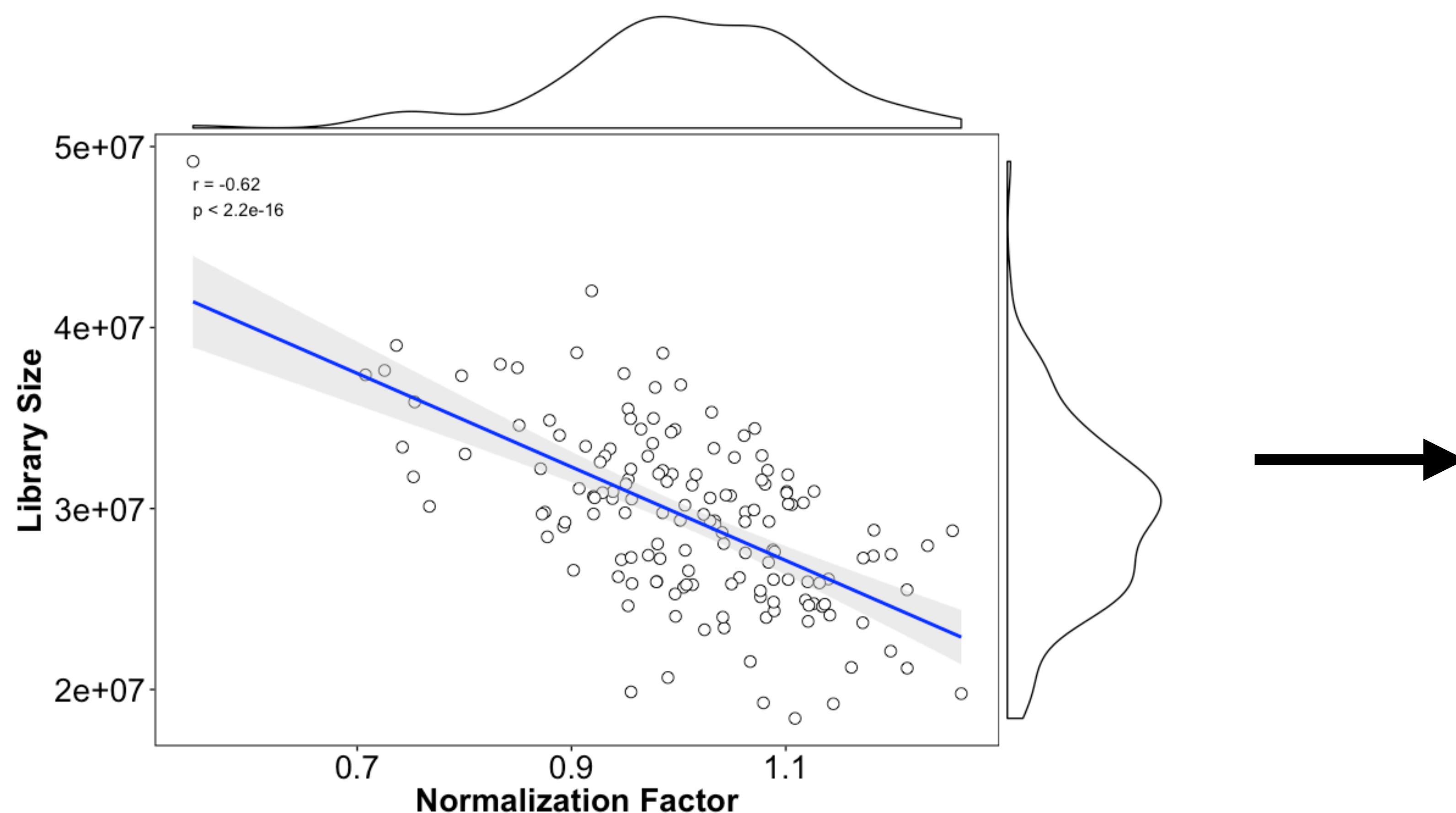
Primary Site	Cases
Adrenal Gland	~100
Bile Duct	~100
Bladder	~1,000
Blood	~1,000
Bone	~100
Bone Marrow	~100
Brain	~1,000
Breast	~3,500
Cervix	~100
Colorectal	~2,800
Esophagus	~1,000
Eye	~100
Head and Neck	~1,000
Kidney	~2,000
Liver	~1,000
Lung	~4,000
Lymph Nodes	~100
Nervous System	~2,000
Ovary	~1,500
Pancreas	~1,000
Pleura	~100
Prostate	~1,000
Skin	~1,000
Soft Tissue	~100
Stomach	~1,000
Testis	~100
Thymus	~100
Thyroid	~1,000
Uterus	~1,000

Are there underlying gene expression patterns associated with higher survivability?

Which features (genes) contribute the **most** to the observed survival-based stratification?



Library size and TMM normalization factors



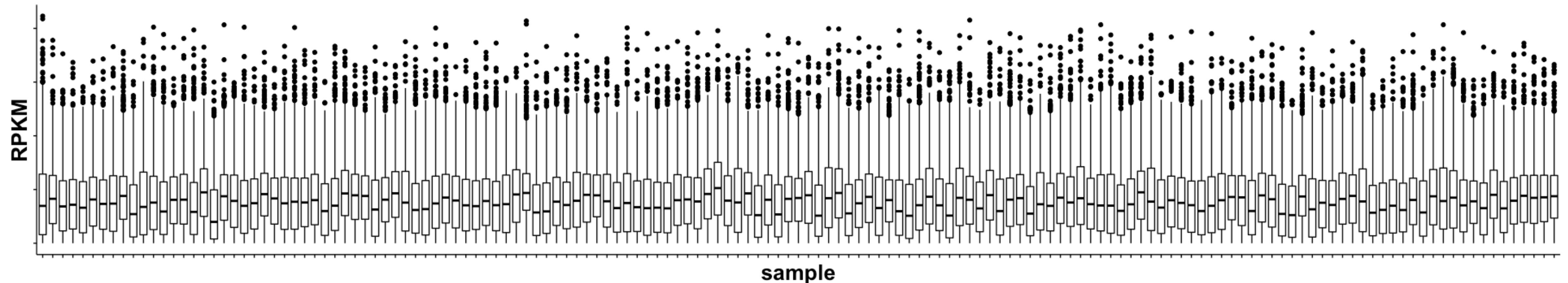
mRNA Abundance metric

RPKM

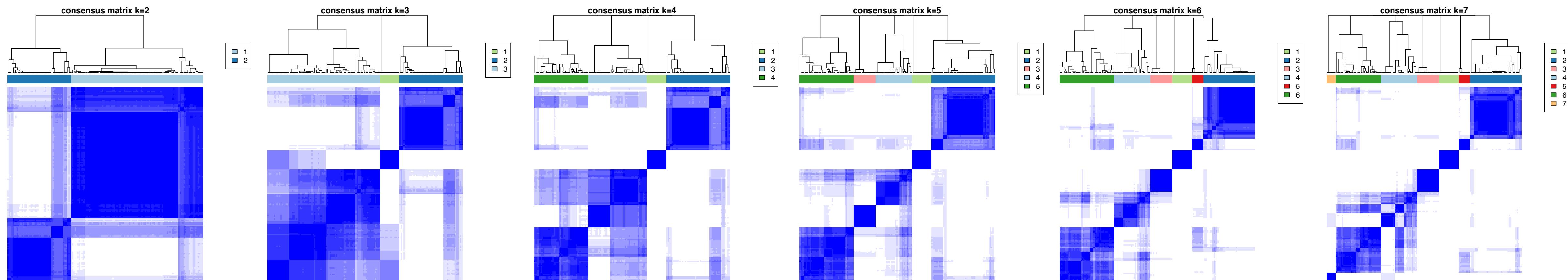
(Reads **P**er **K**ilobase of transcript per **M**illion mapped reads)

The main aim in TMM normalization is to account for library size variation between samples of interest.

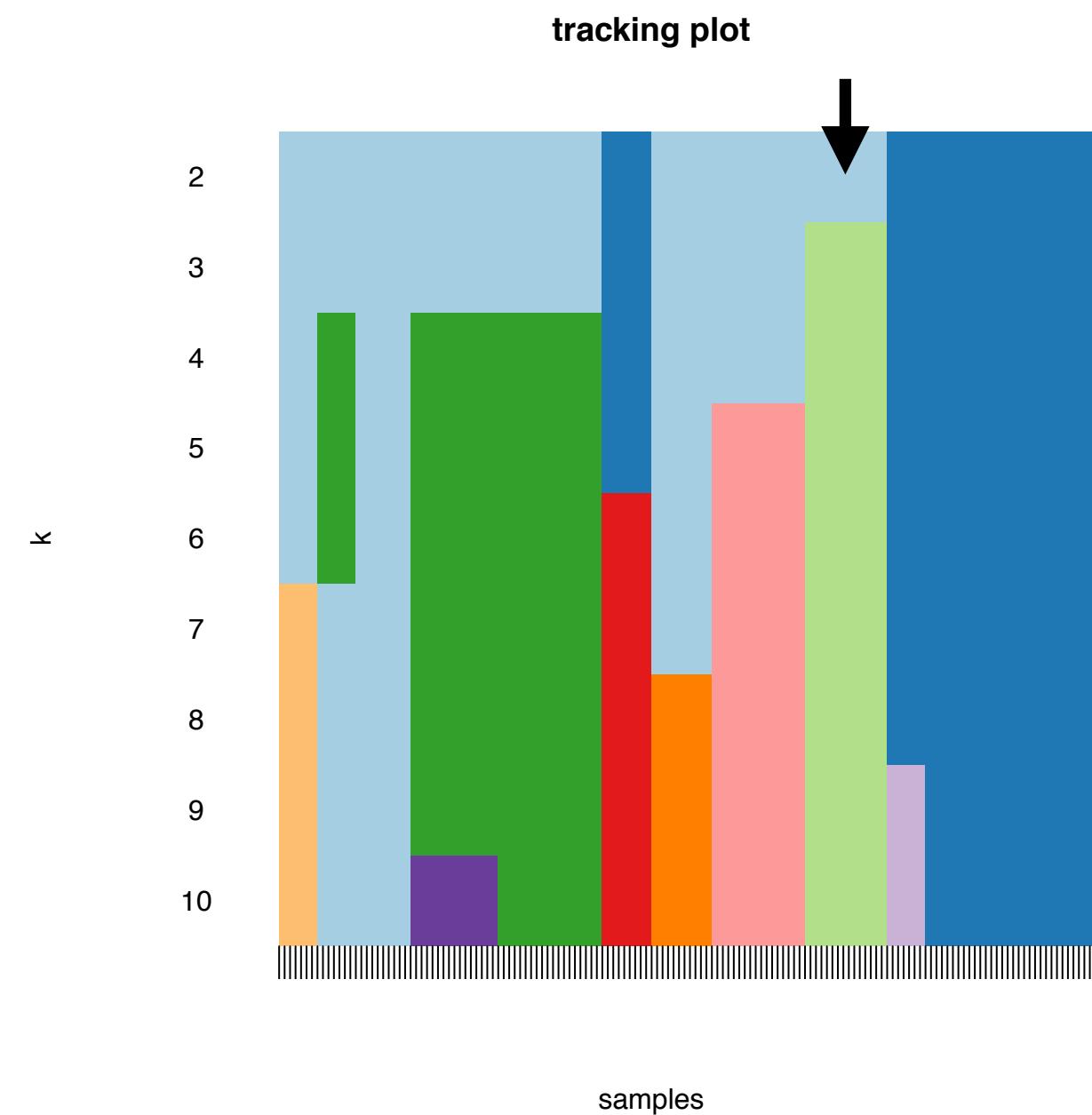
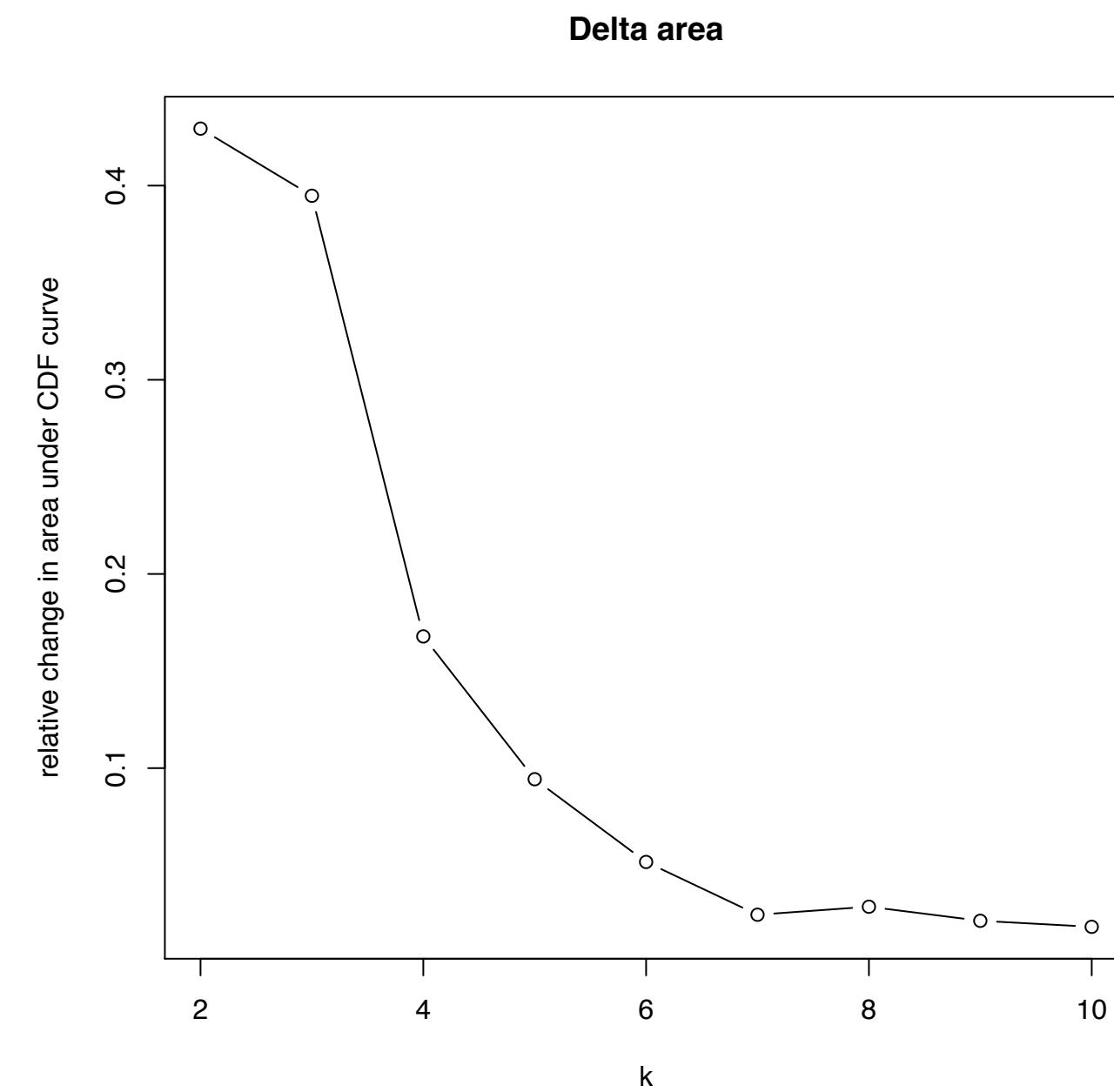
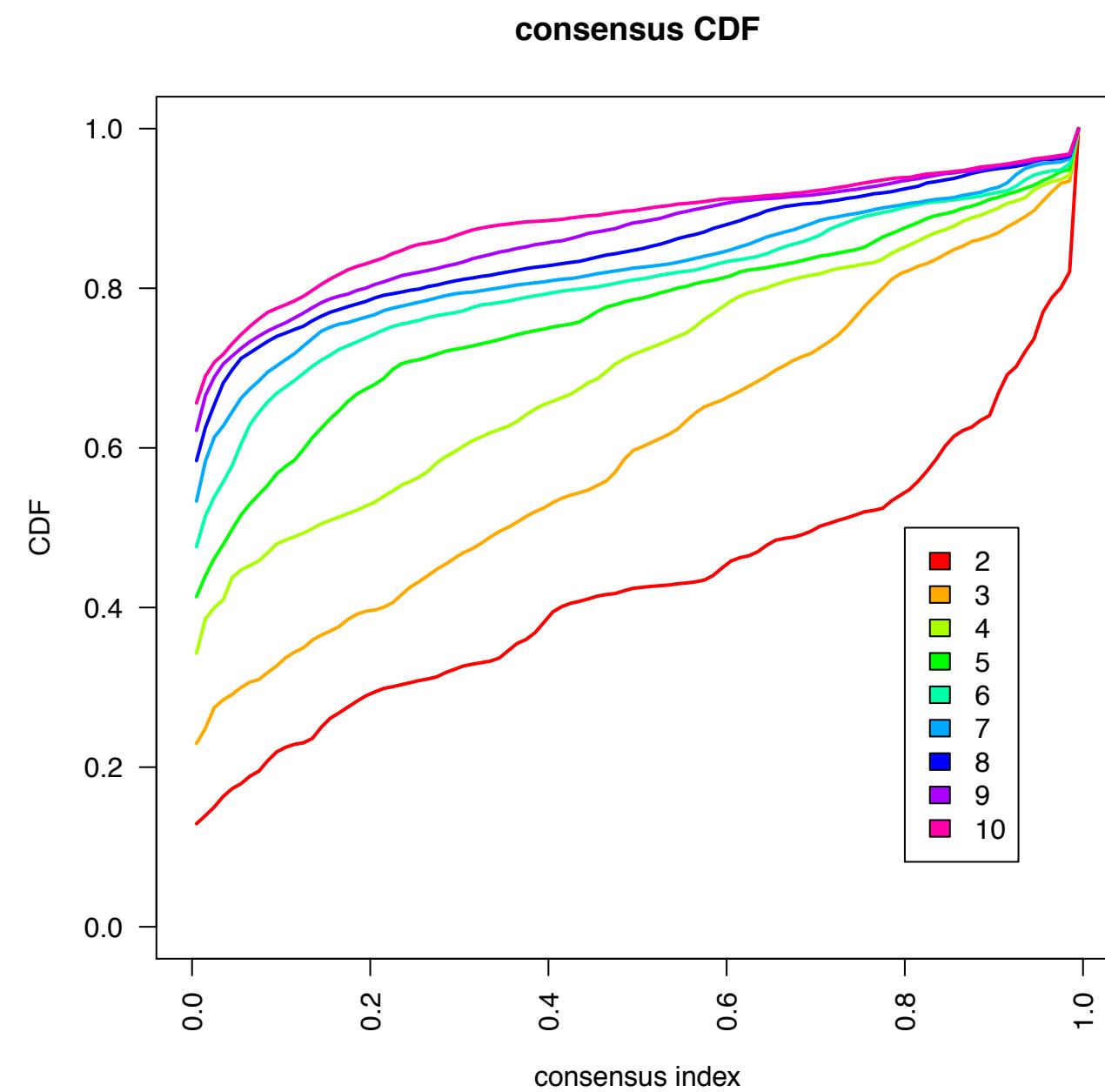
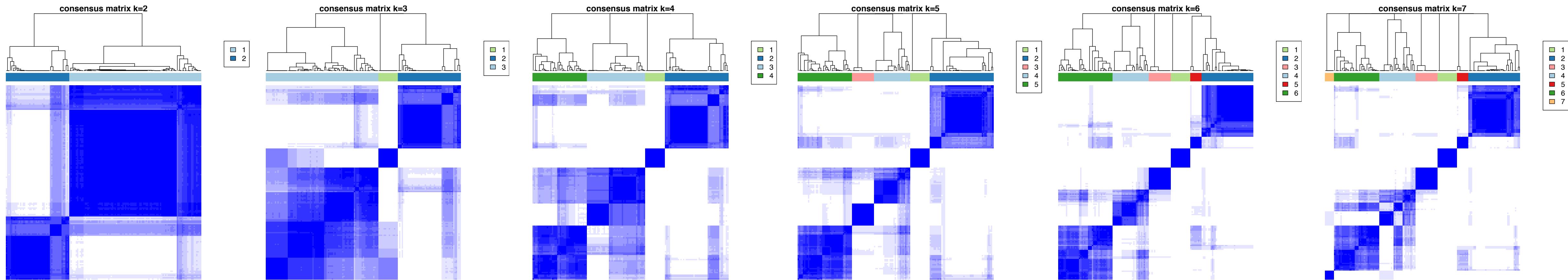
mRNA abundance most variable genes



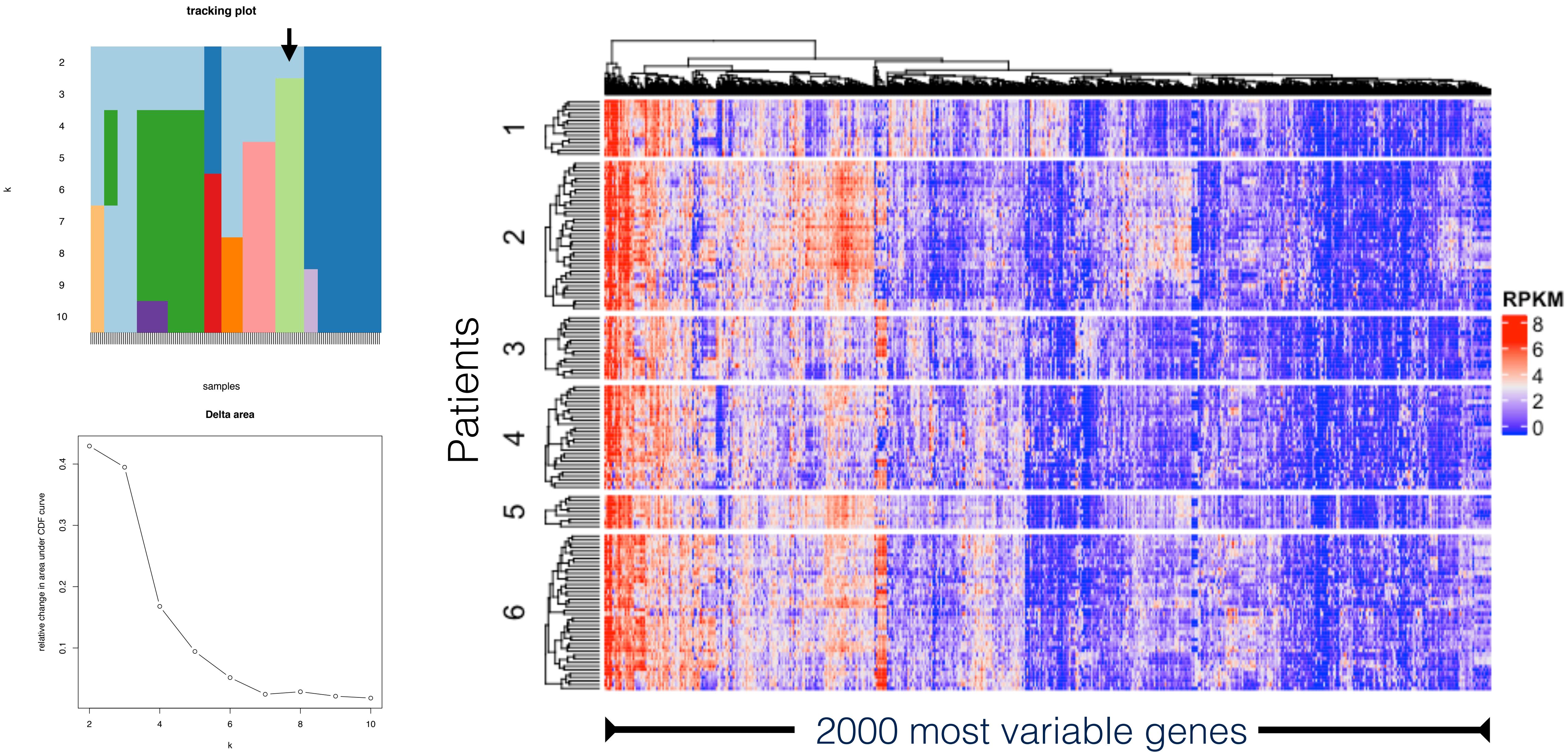
Consensus Clustering



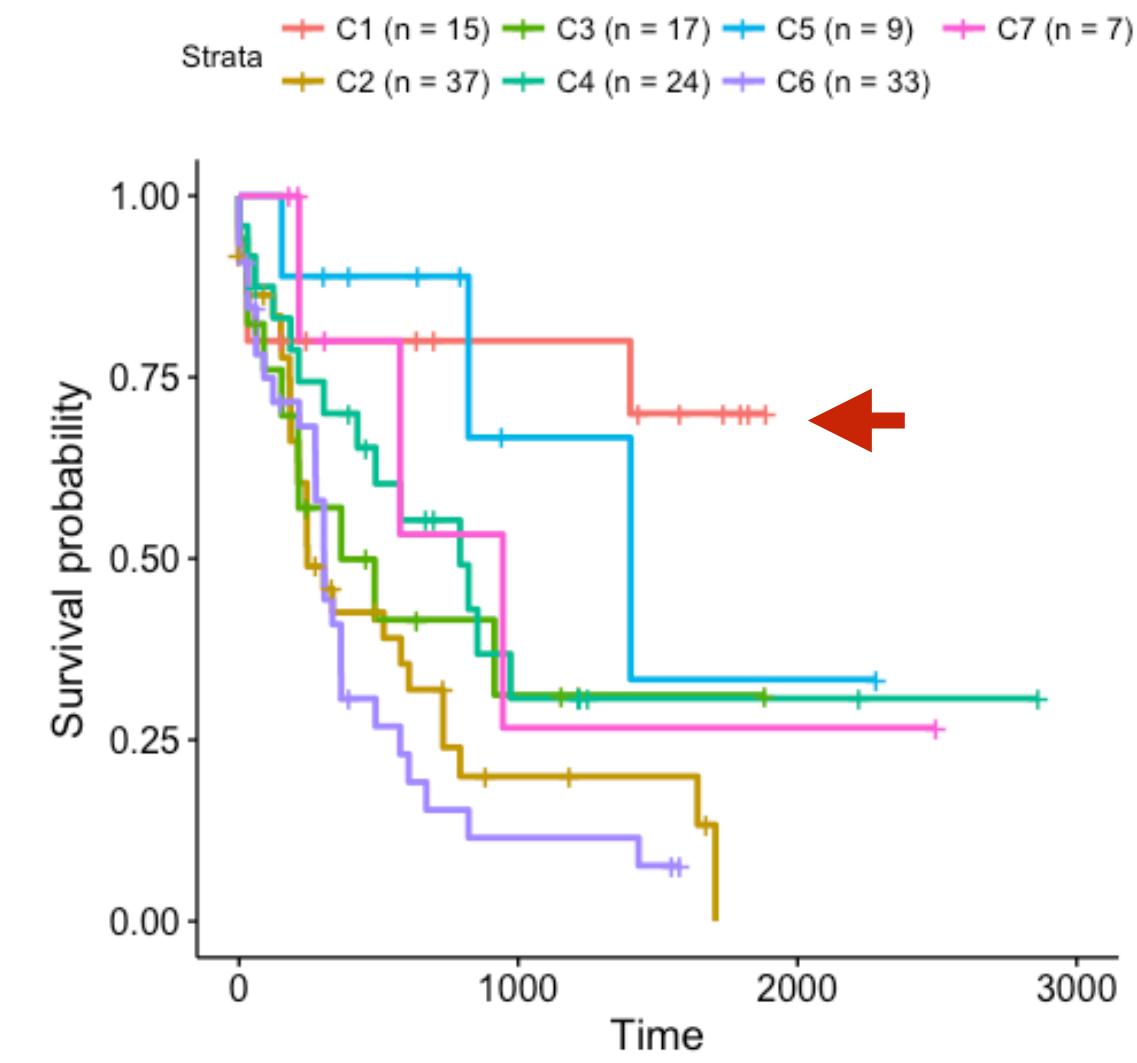
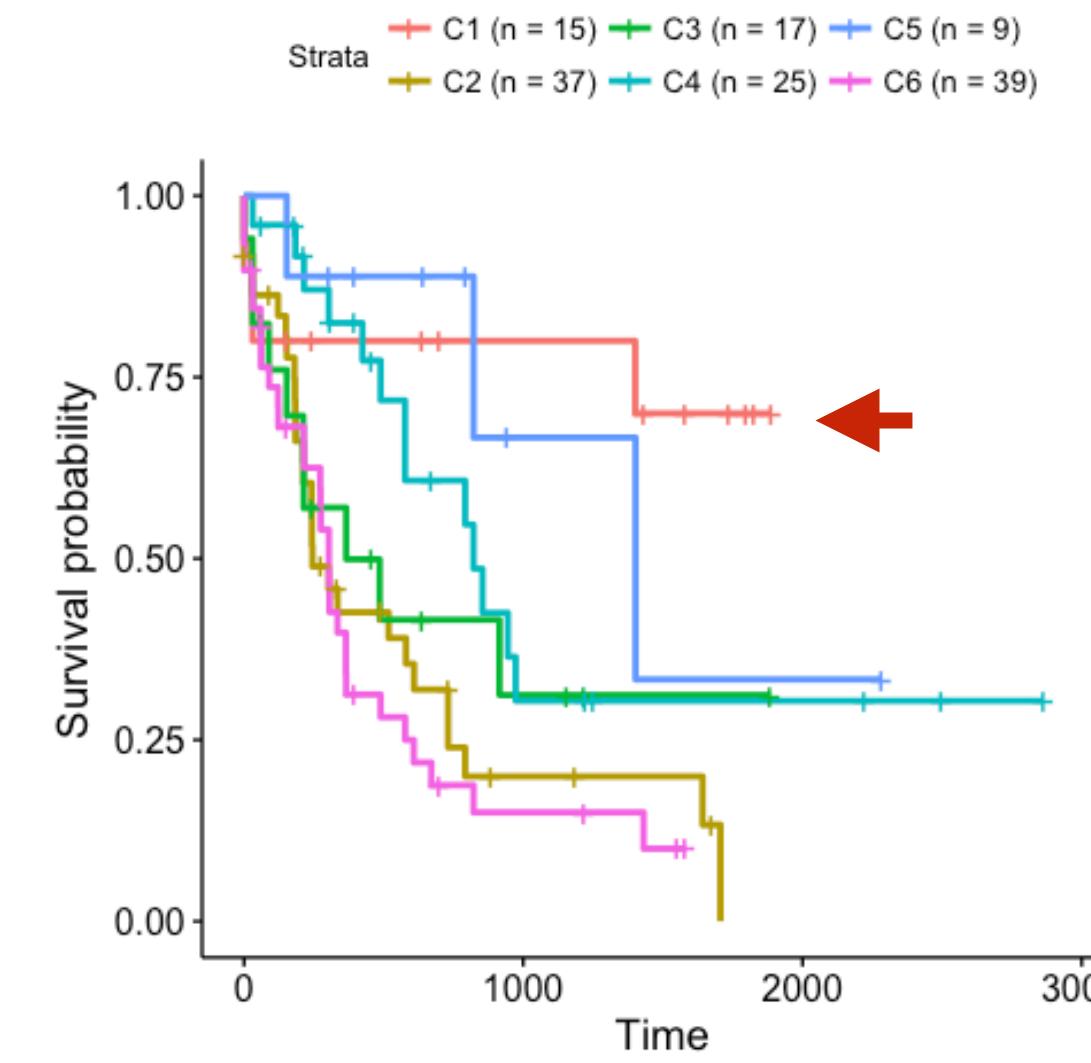
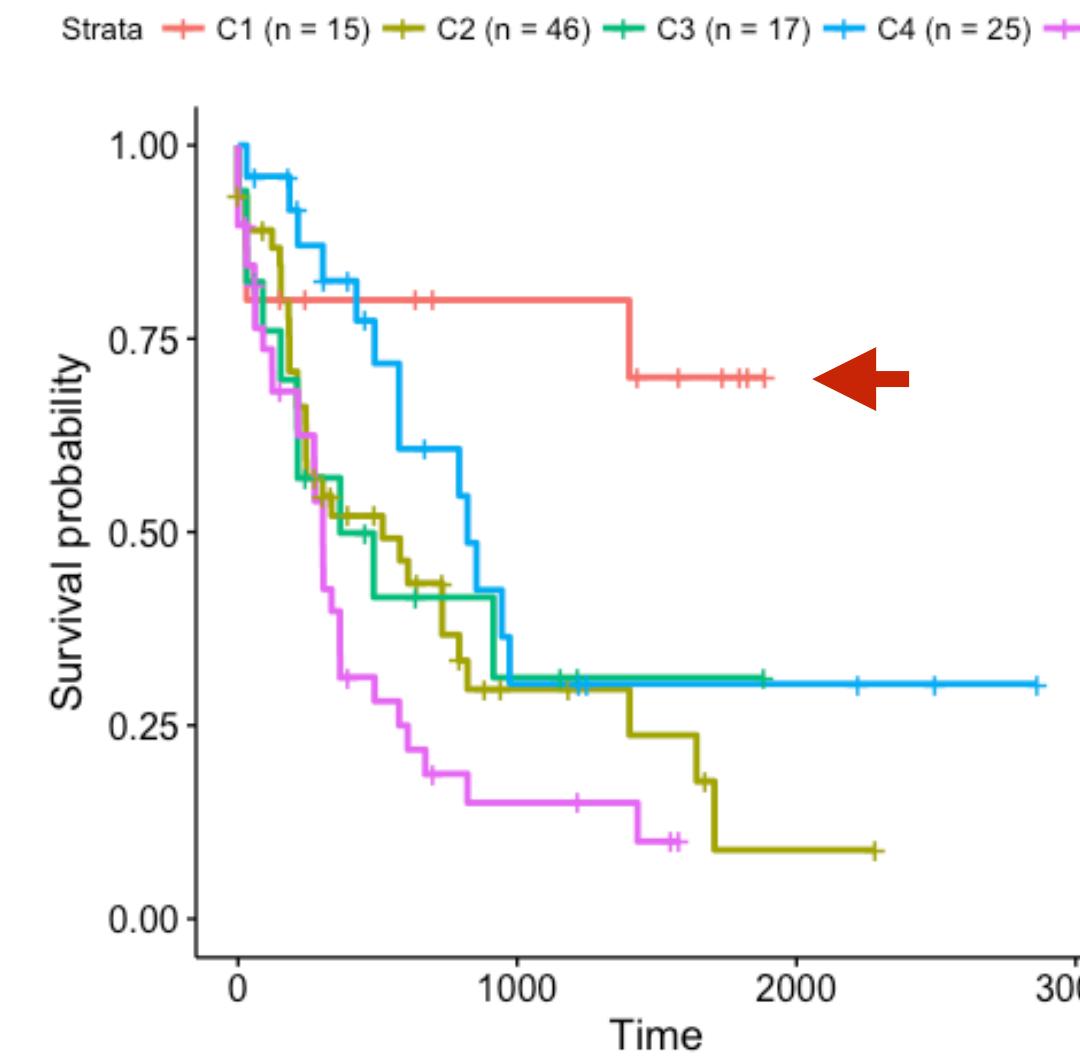
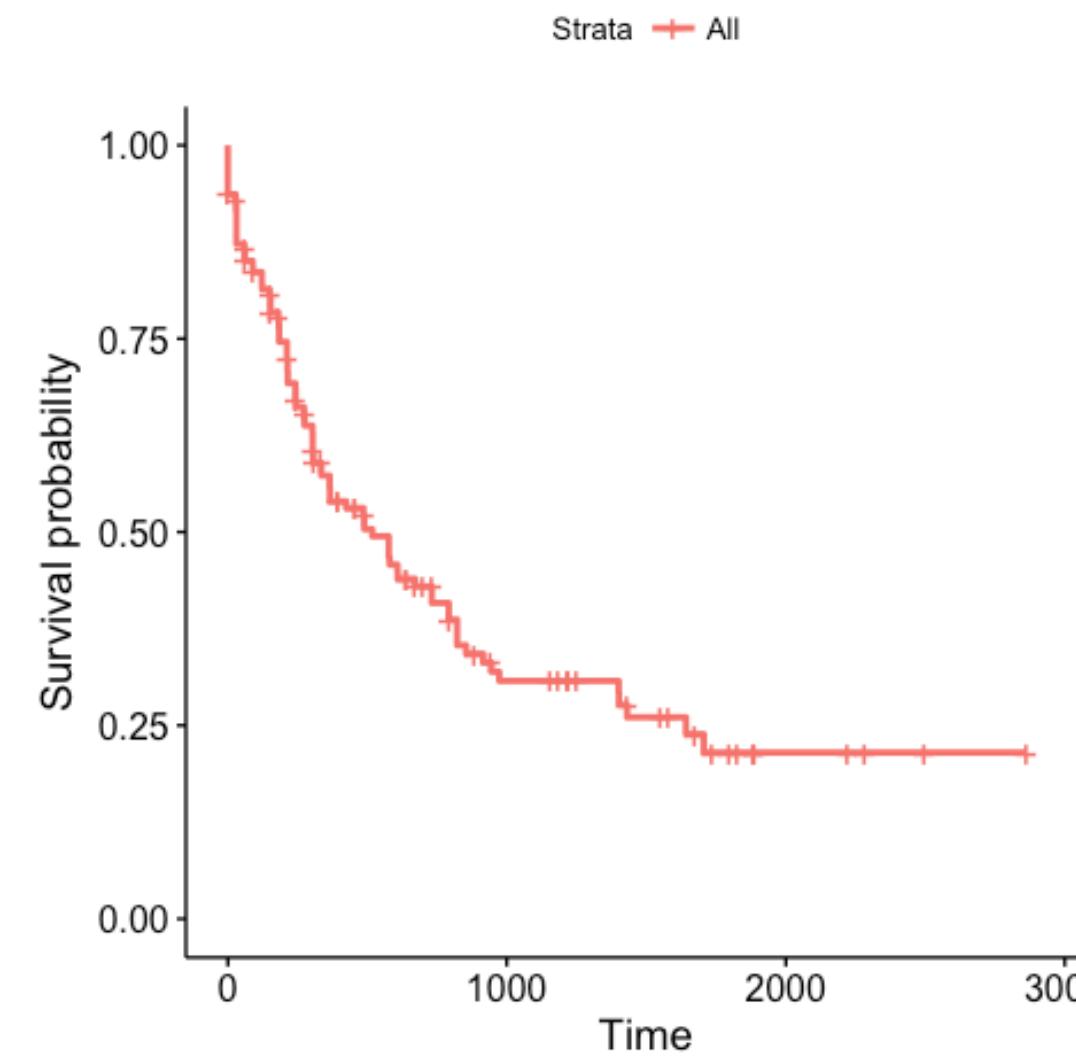
Consensus Clustering for subtype discovery



Consensus Clustering for subtype discovery

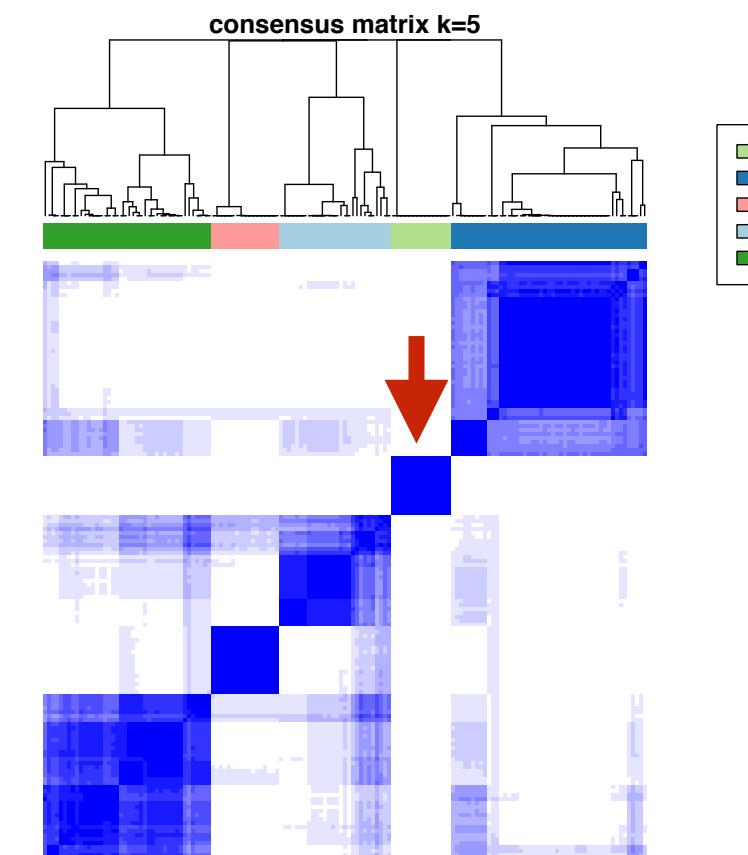


Consensus Clustering of gene expression profiles suggest association with survival probability.

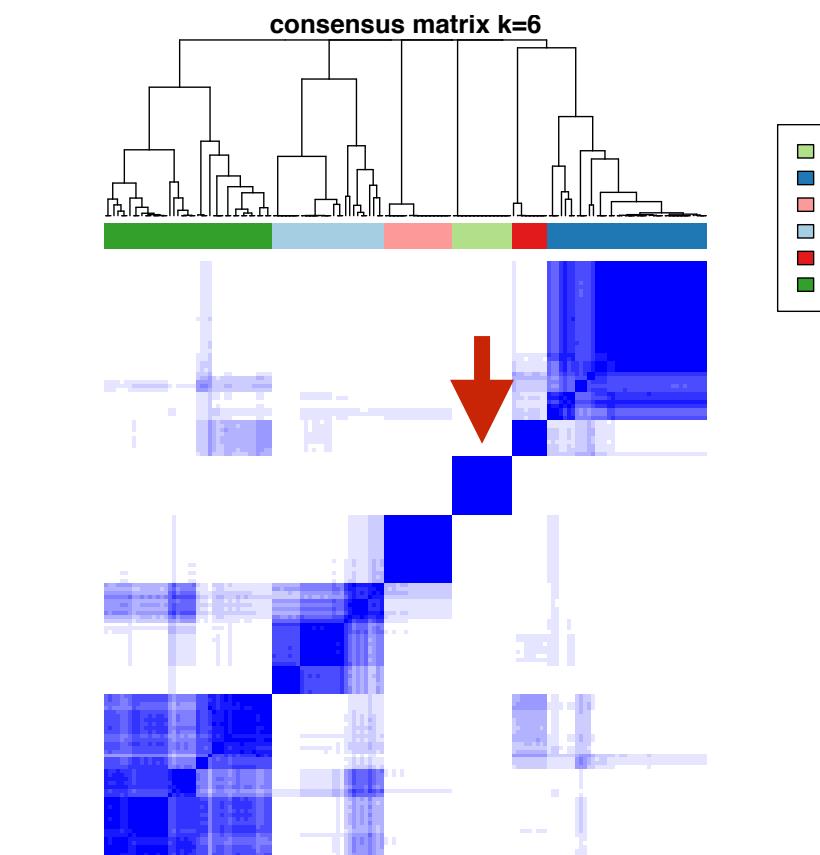


Overall Survivability
Poor Prognosis
(5 year survival <25%)
 $N = 142$

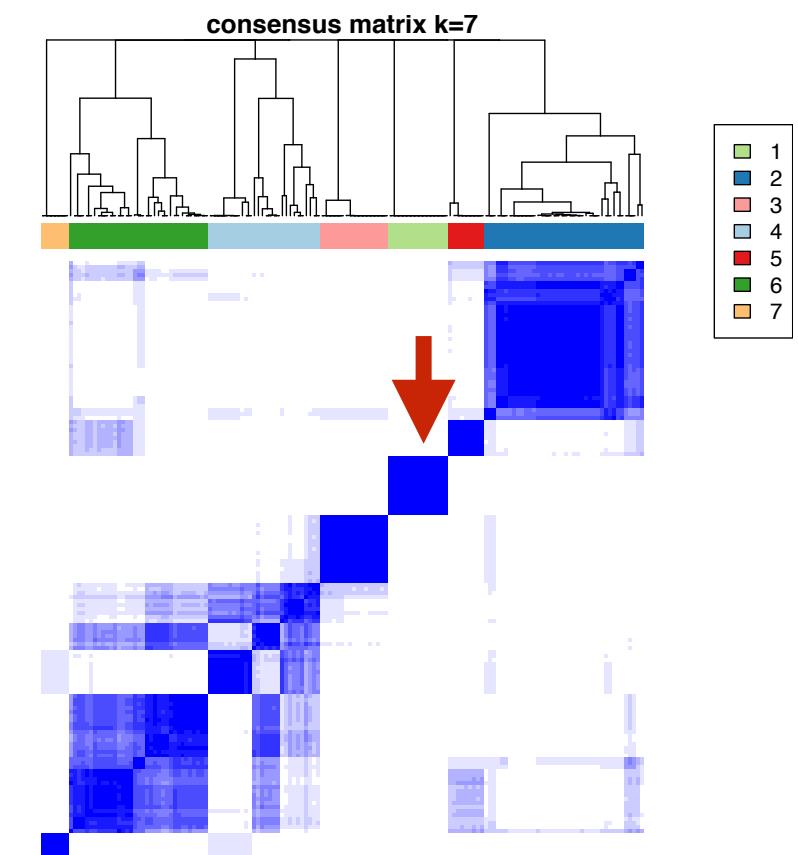
$$p = 0.0011 \\ K = 5$$



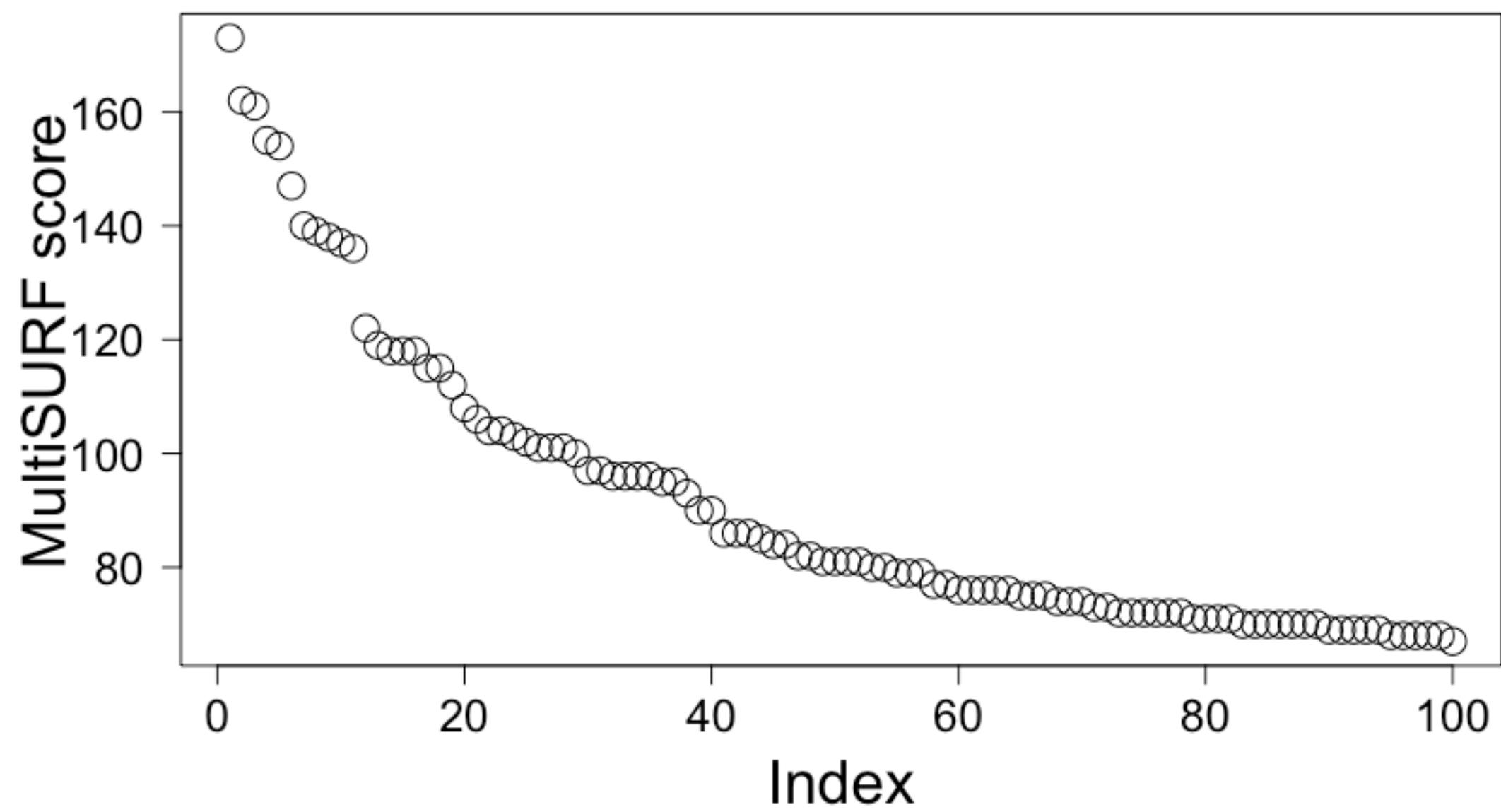
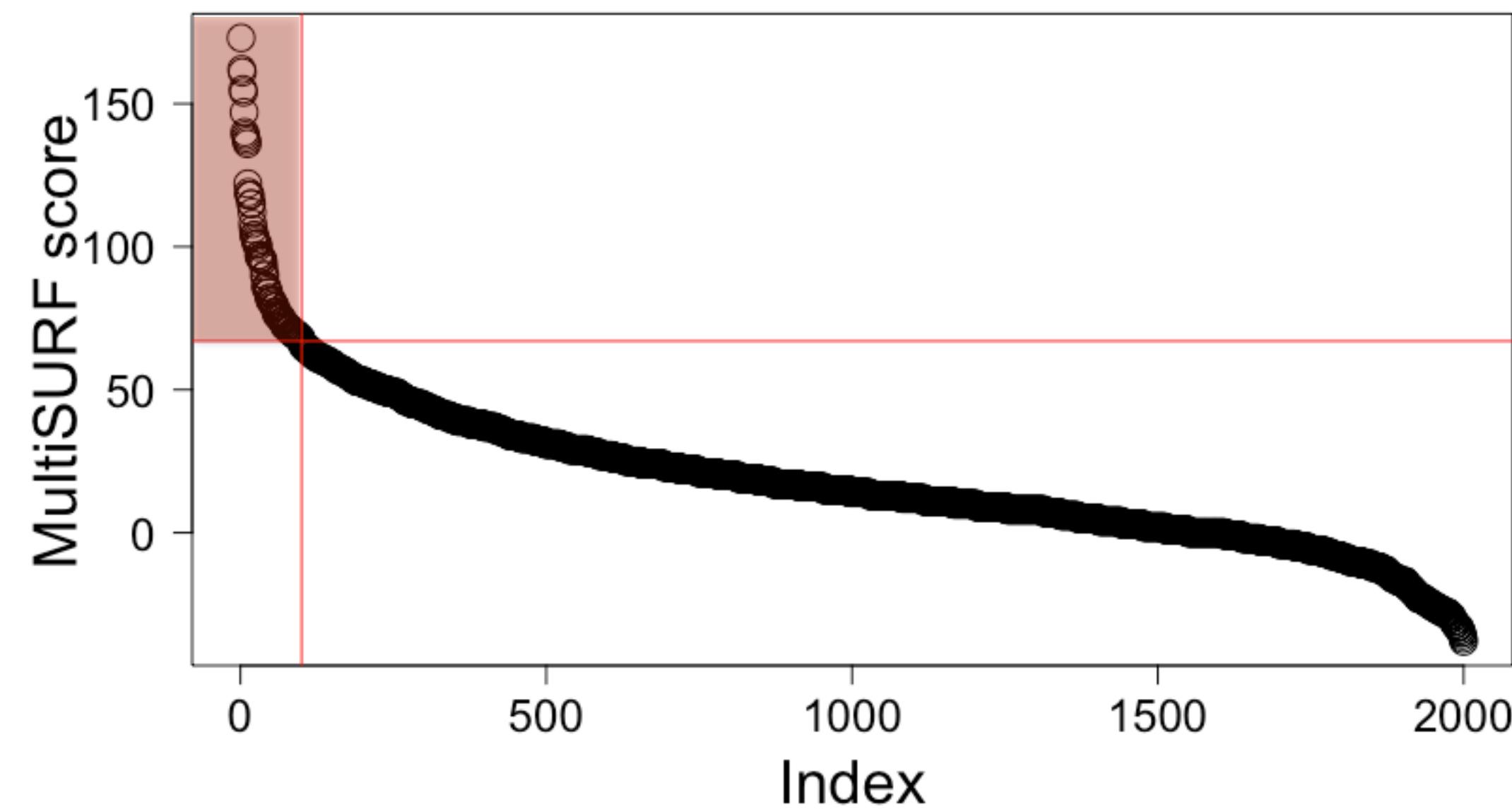
$$p = 0.00014 \\ K = 6$$



$$p = 0.00061 \\ K = 7$$

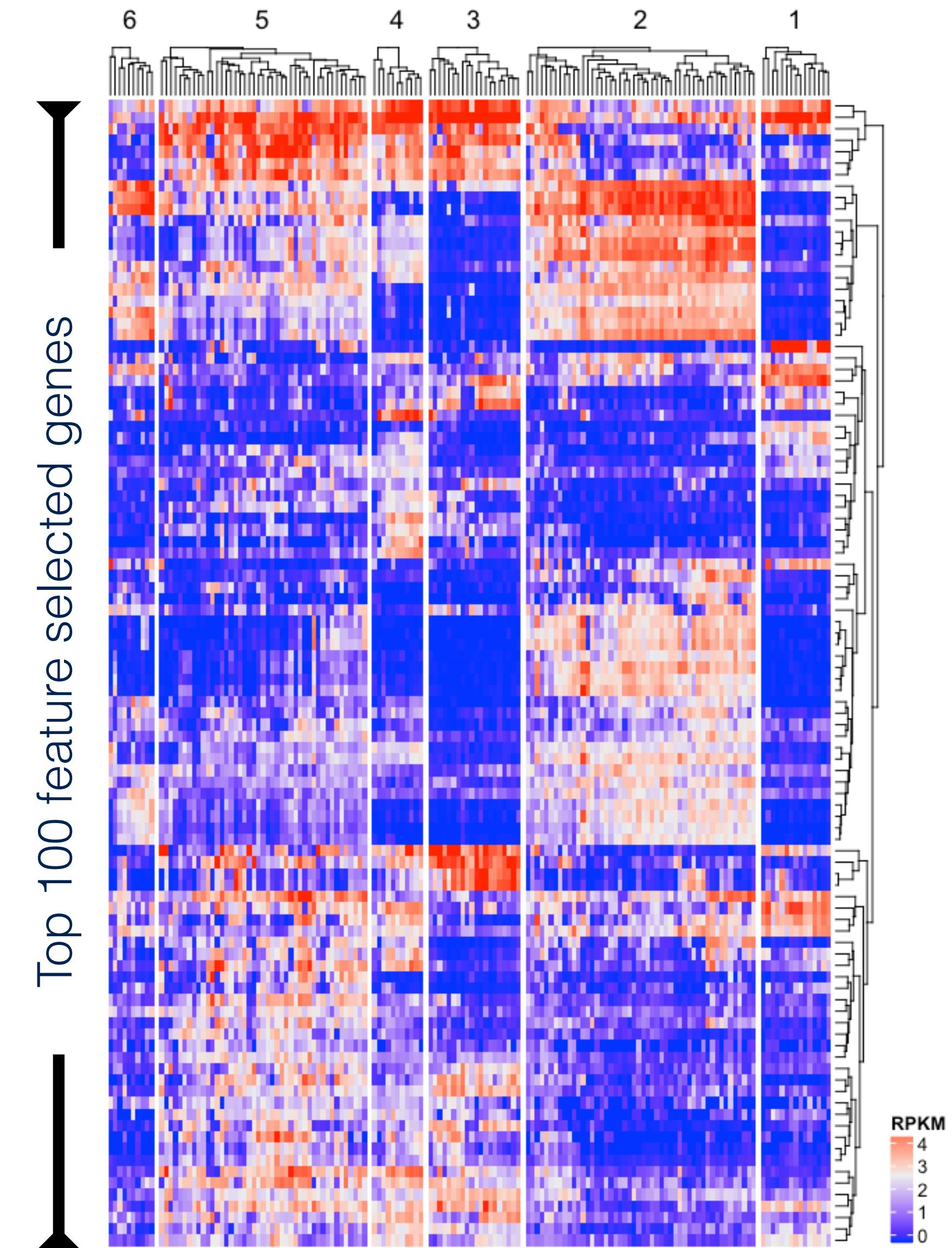
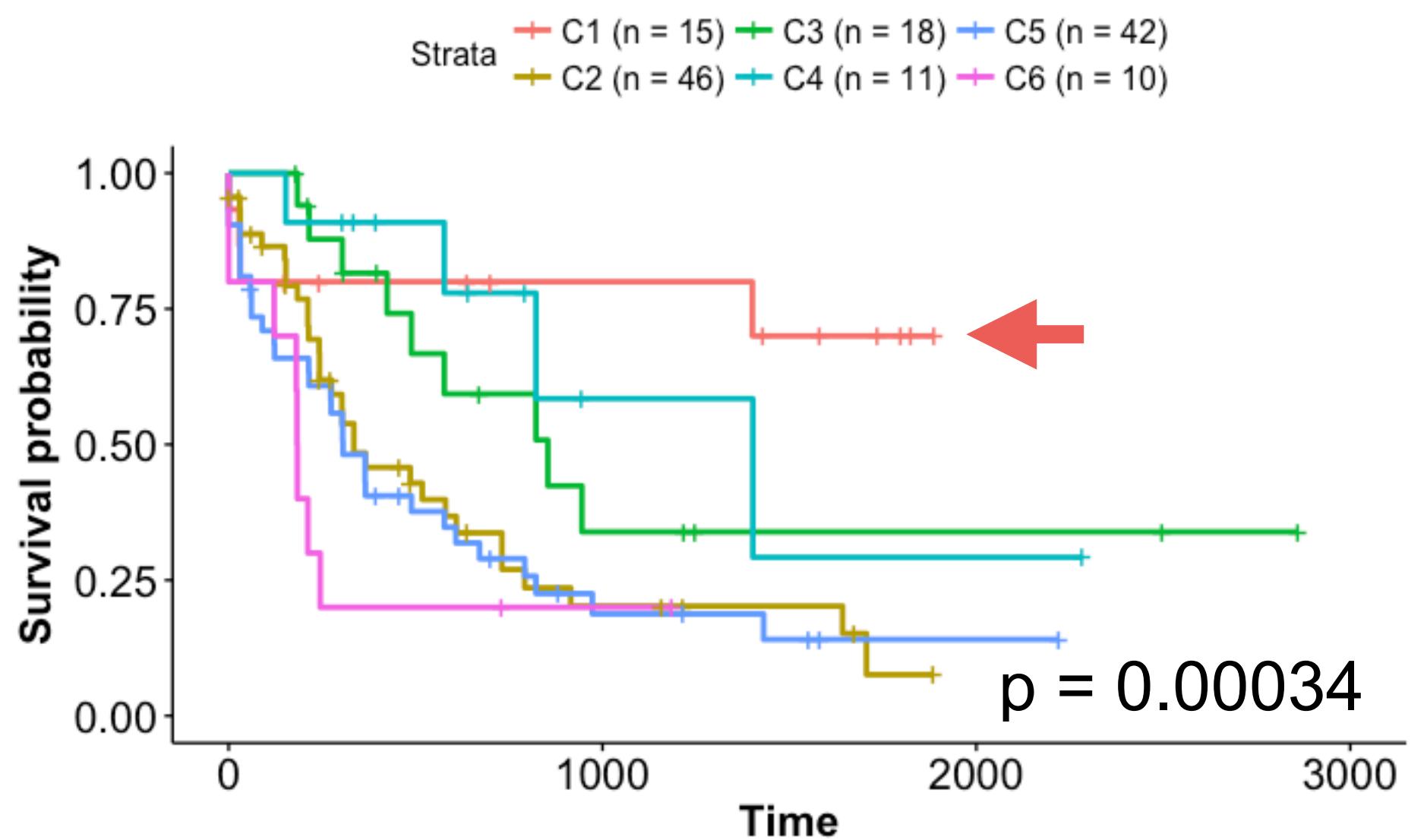
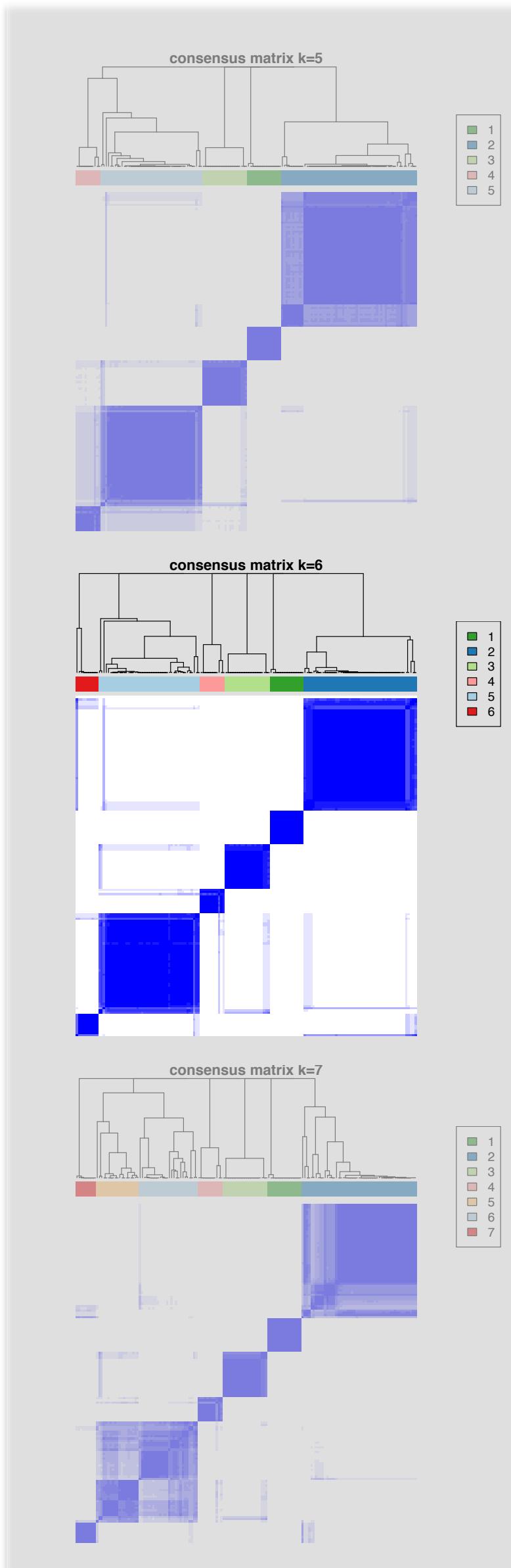


Which features (genes) contribute the **most** to the observed survival-based stratification?

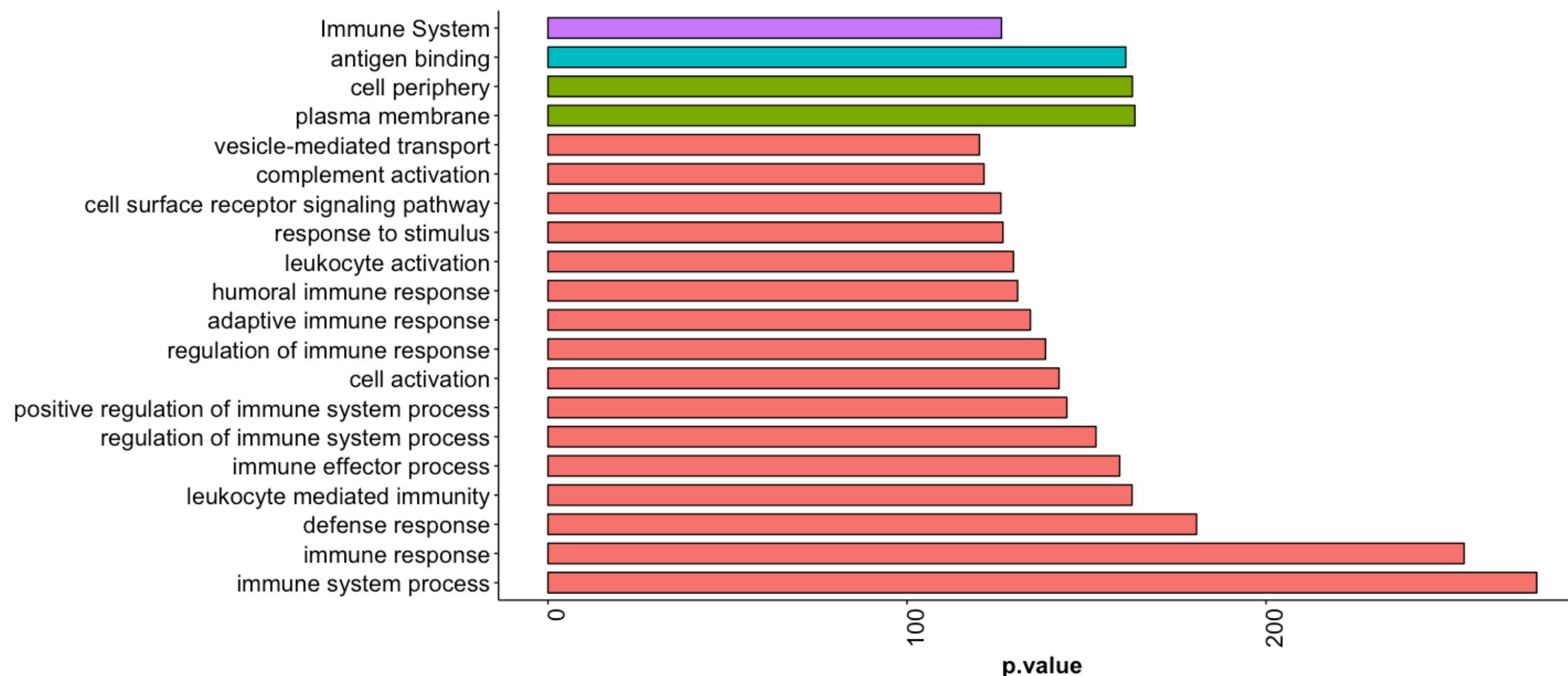


Picked top 100 genes based on MultiSurf score.

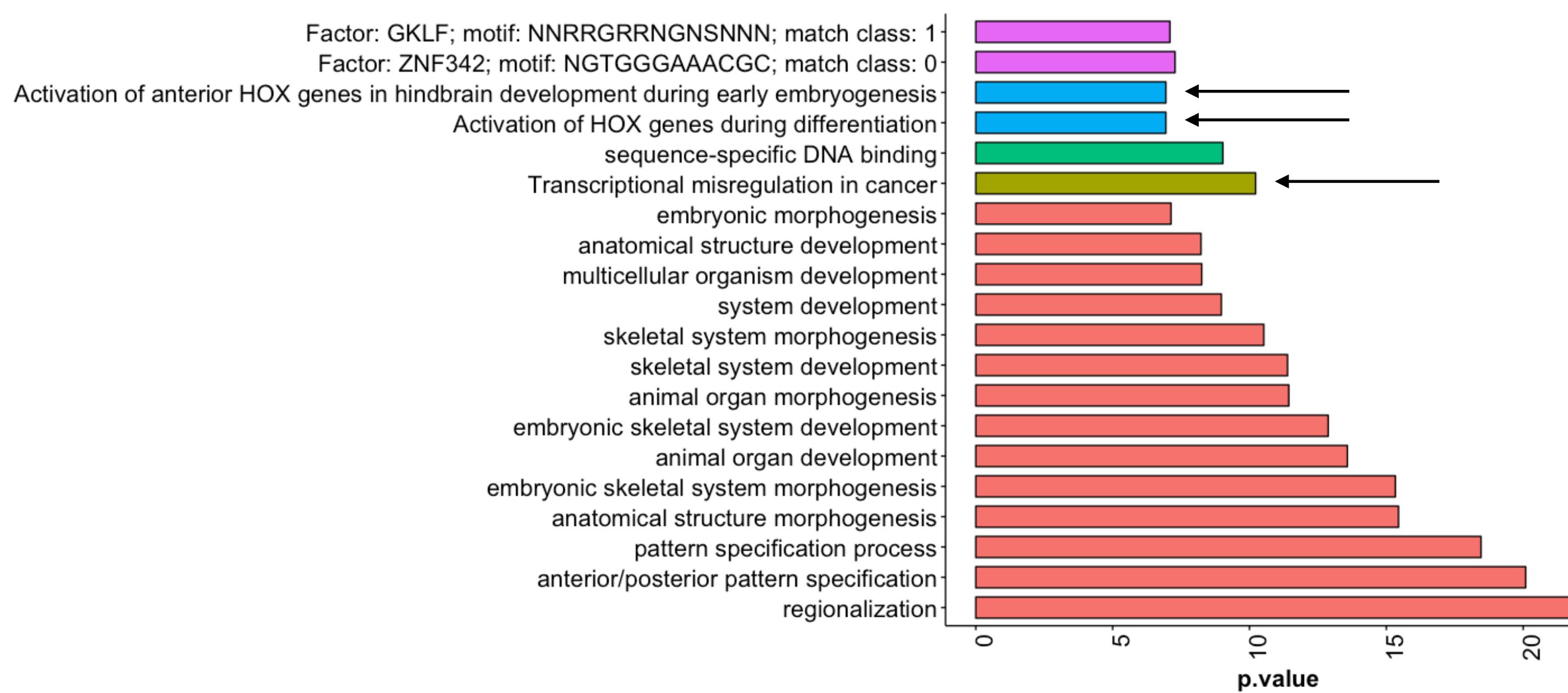
Feature selection retains stratification of patients with the highest survival probabilities.



Gene Ontology



Immune system genes

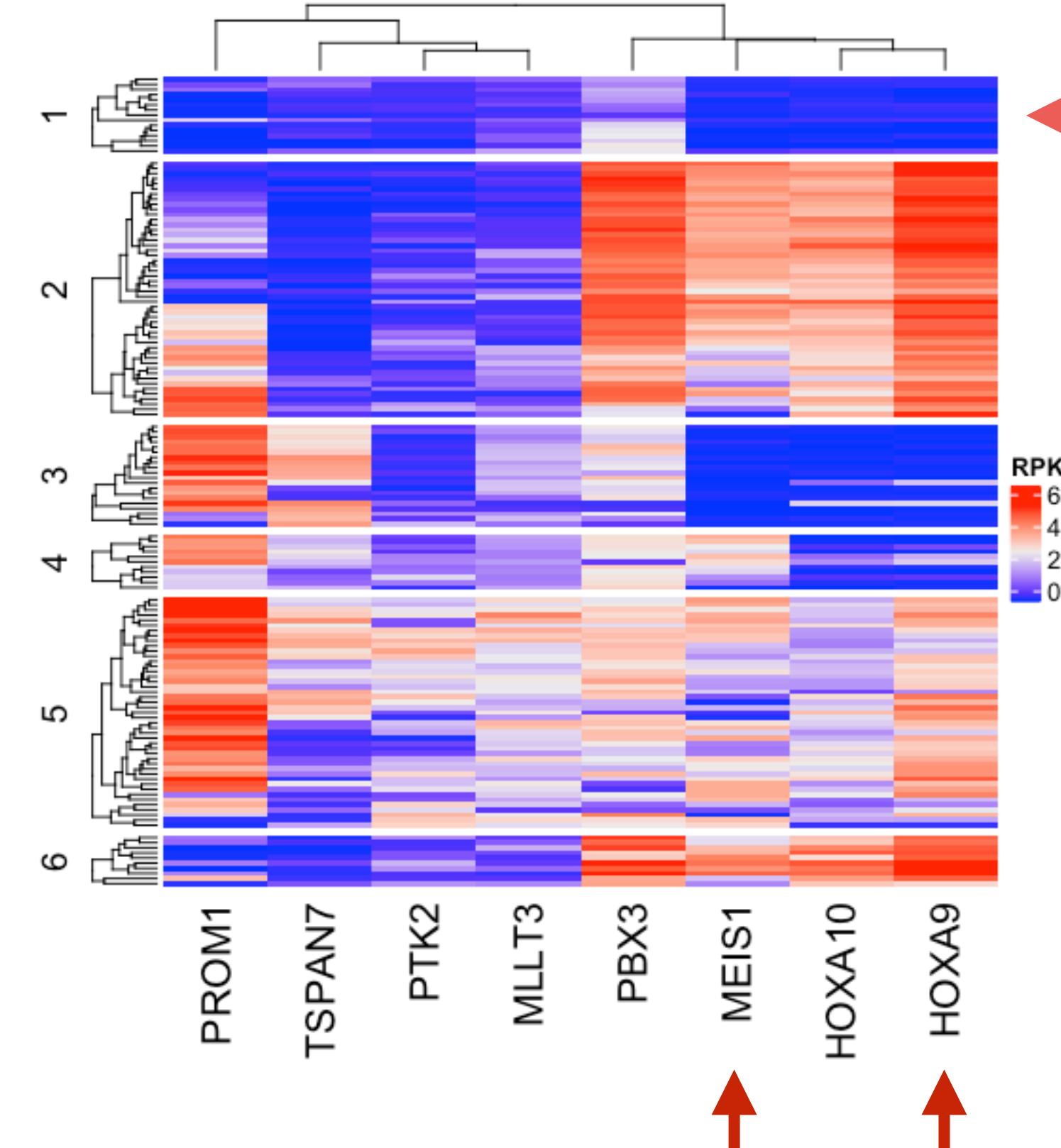


Feature Selection

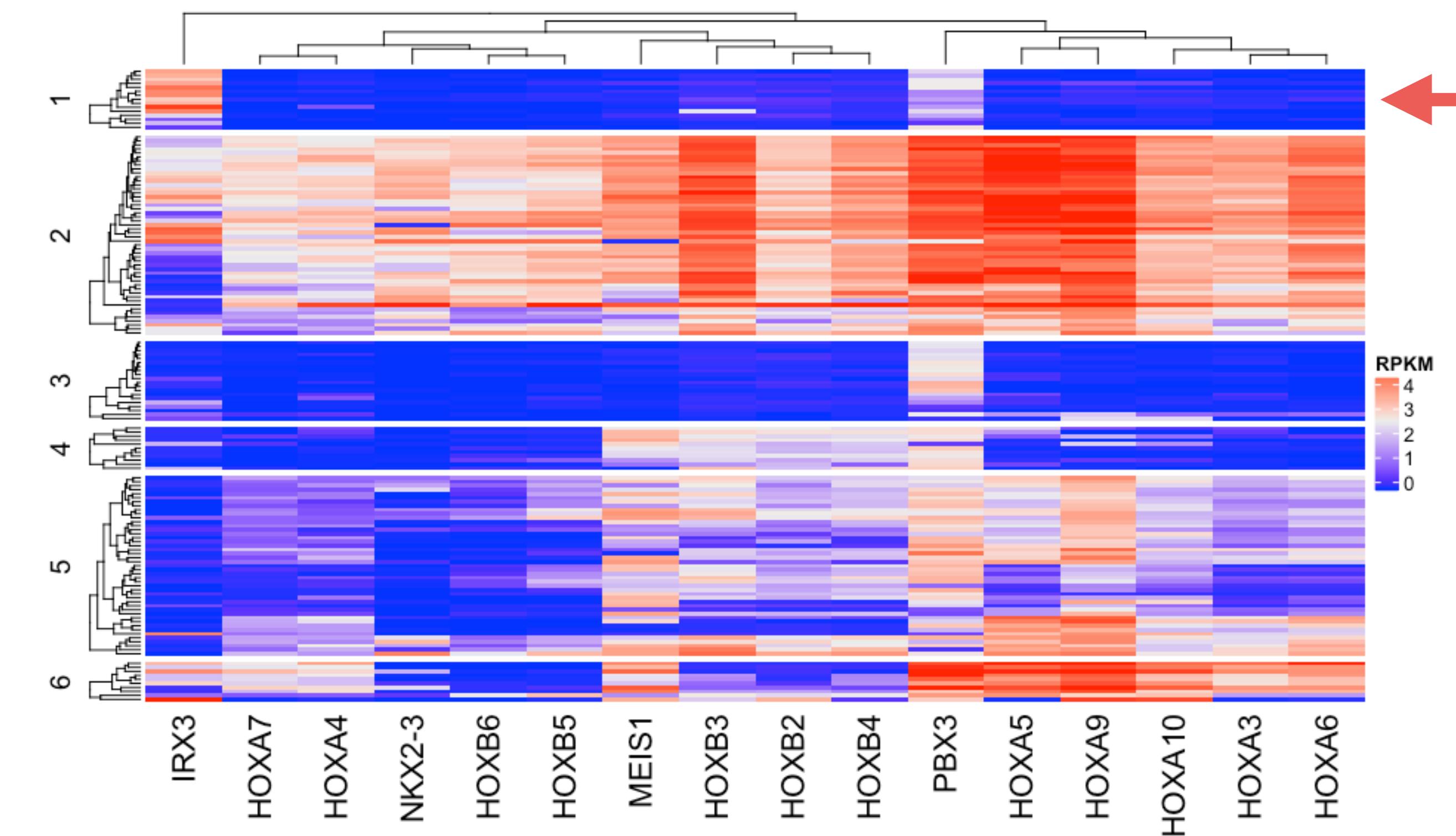
Developmental genes

MultiSURF feature selection identifies differential expression of HOX gene clusters as potential association to survival probability.

GO: Transcriptional Misregulation in Cancer



GO: HOX gene clusters



Hypothesis:

Expression of HOX gene clusters could be associated with a more aggressive and proliferative form of AML, leading to poor prognosis.

Current literature has linked HOX gene expression misregulation to AML



Comment on Morgado et al, page 4020, and comment on Jin et al, page 3998

Survival signaling in HoxA9/Meis1 AML

Gang G. Wang and Mark P. Kamps UNIVERSITY OF CALIFORNIA SAN DIEGO SCHOOL OF MEDICINE



Oncogene (2007) 26, 6766–6776
© 2007 Nature Publishing Group All rights reserved 0950-9232/07 \$30.00

www.nature.com/onc

REVIEW

Hox genes in hematopoiesis and leukemogenesis

B Argiopoulos¹ and RK Humphries^{1,2}

¹Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, BC, Canada and ²Department of Medicine, University of British Columbia, Vancouver, BC, Canada

I ran out of time to keep going...Future Works....

- **Associate clusters with somatic mutation profiles:** I was recently able to obtain Simple Nucleotide Variation files quantified via 4 different pipelines. (muse, varscan2, somaticsniper, mutect)
- **Explore how the occurrence of multiple genetic variants can converge into the gene expression profile of AML:** A large amount of AML patients express multiple somatic variants. Thus, it will be interesting to see how each of these variants contributes to the heterogeneity of the underlying expression profiles.