

PP19

*Data Cleaning and Gene Expression Visualization in Hematological Conditions

Ashwika Papasani¹ , Nasreen Bano², Micheal Atchison²

¹Downingtown STEM Academy, Downingtown, PA, USA

²University of Pennsylvania, Philadelphia , PA, USA

*The work is done at the University of Pennsylvania

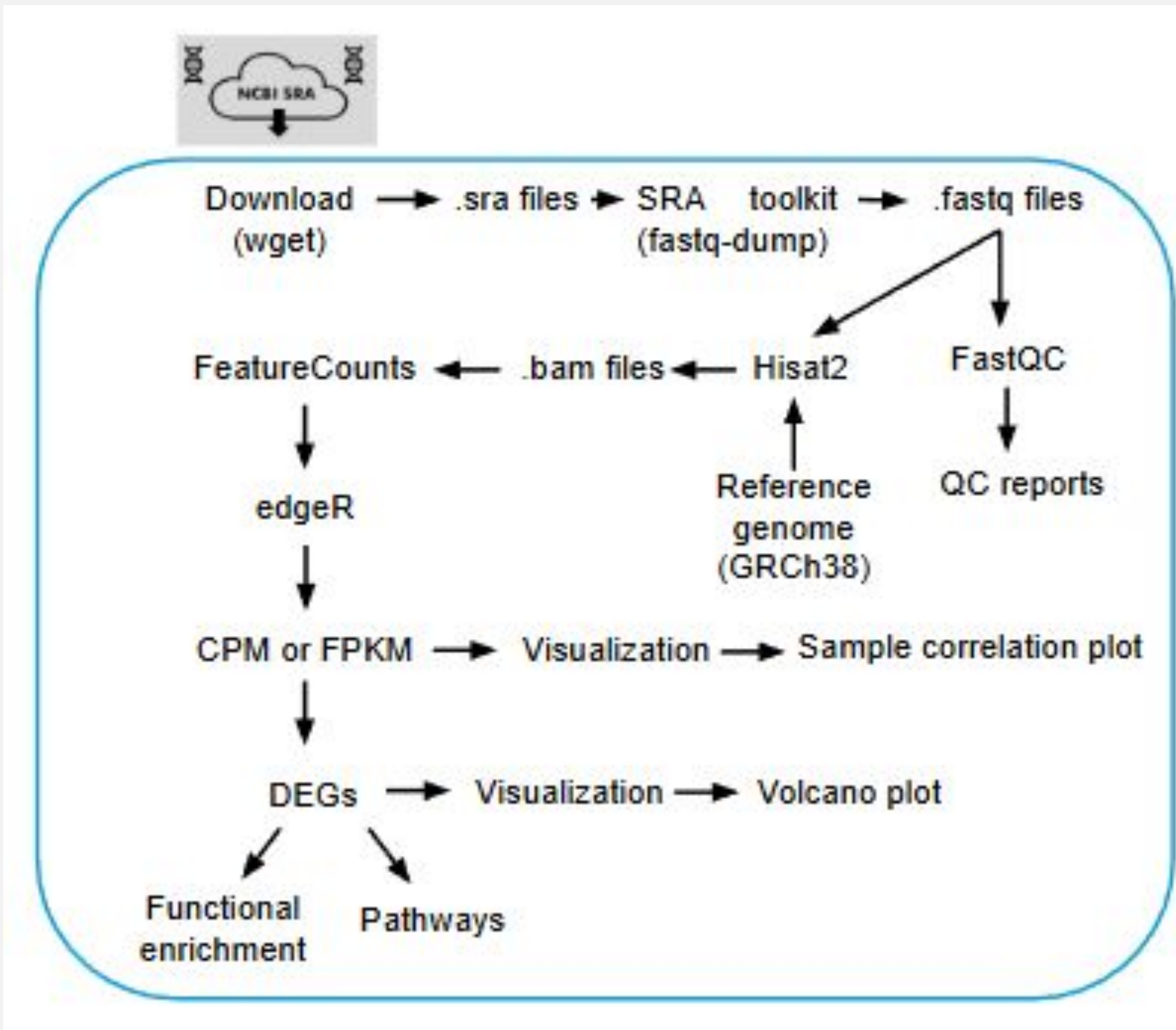
ABSTRACT

A new technology that offers a view into gene expression at the cell level or across the entire cell population is RNA sequencing. In clinical research, it is becoming increasingly crucial for diagnosing diseases and understanding how patients respond to treatment. In preclinical studies, this approach has already shown promise in assisting researchers in understanding the expression of genes and the functioning of cells. Large volumes of data are produced by this technology, necessitating the use of specialized bioinformatics tools to identify, clean, and eliminate low-quality cells, group cells according to gene expression, and produce visualizations that show variations. In order to investigate blood cell formation, we examine in this presentation how this data is collected, processed, and visualized.

INTRODUCTION

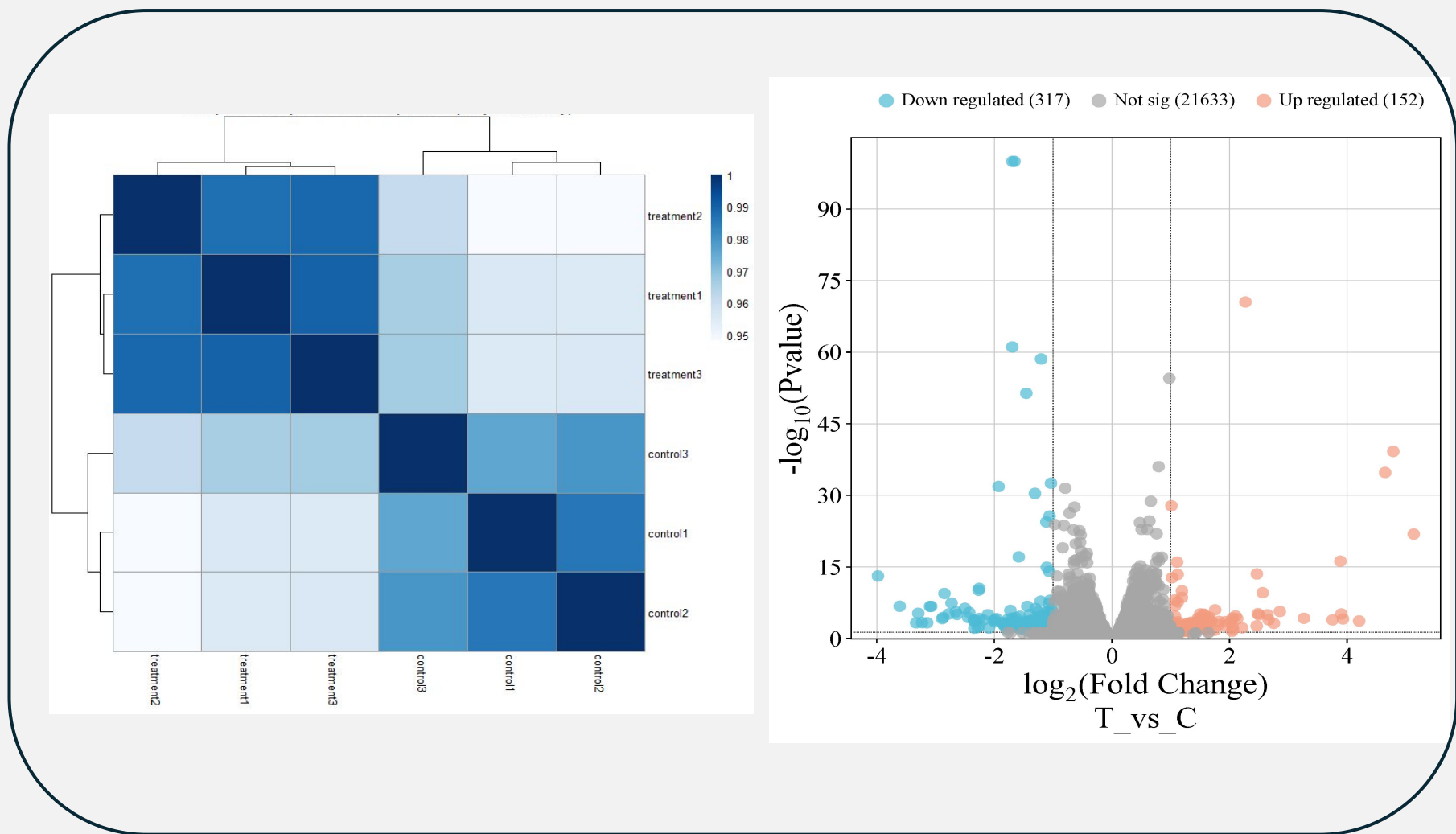
RNA sequencing (RNA-seq) is a powerful approach for profiling gene expression and revealing transcriptional changes across diverse biological systems. Applied at either the single-cell or bulk level, RNA-seq has become an essential tool in clinical and preclinical research. In this study, we showcase a pipeline using bioinformatic software or tools for bulk RNA-seq that combines quality control, alignment, read count generation and differential expression analysis between two conditions. RNA-seq analysis was performed on HCT116 cells that were divided into two groups: a control group transfected with the empty vector (pcCon) and an experimental/treated group transfected with the YY2 expression plasmid (pcYY2). Each group included three biological replicates. Comparative transcriptomic analysis between these groups was conducted to identify YY2-regulated genes associated with ferroptosis. Ferroptosis is a regulated form of cell death driven by iron-dependent lipid peroxidation and disruption of redox homeostasis. Understanding the transcriptional networks that regulate ferroptosis may therefore provide insight into both cancer biology and blood-related diseases, highlighting the broader relevance of YY2-mediated gene regulation.

METHODOLOGY

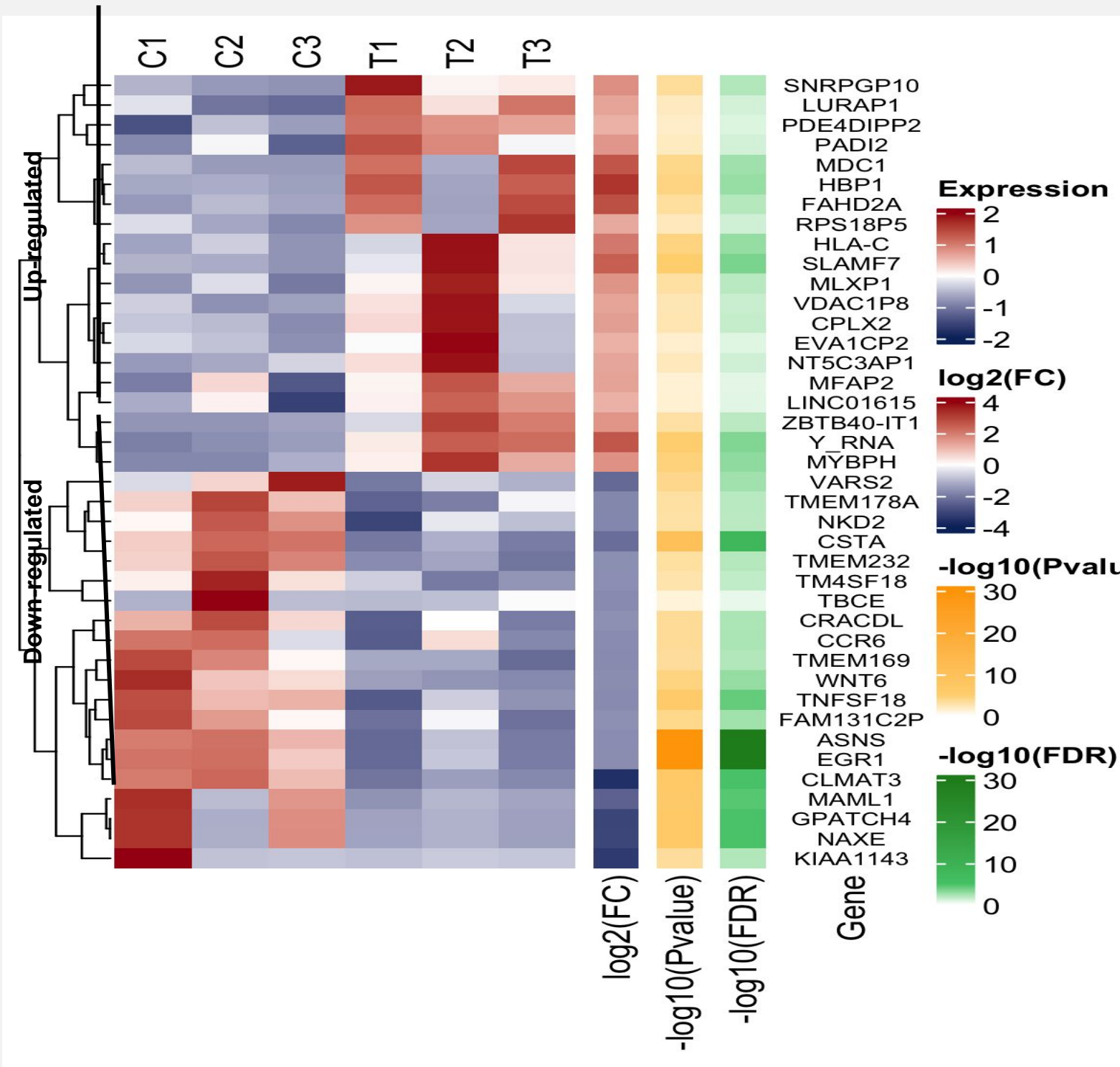


RESULTS

Correlation heatmap of RNA-seq samples and volcano plot of differentially expressed genes (DEGs)



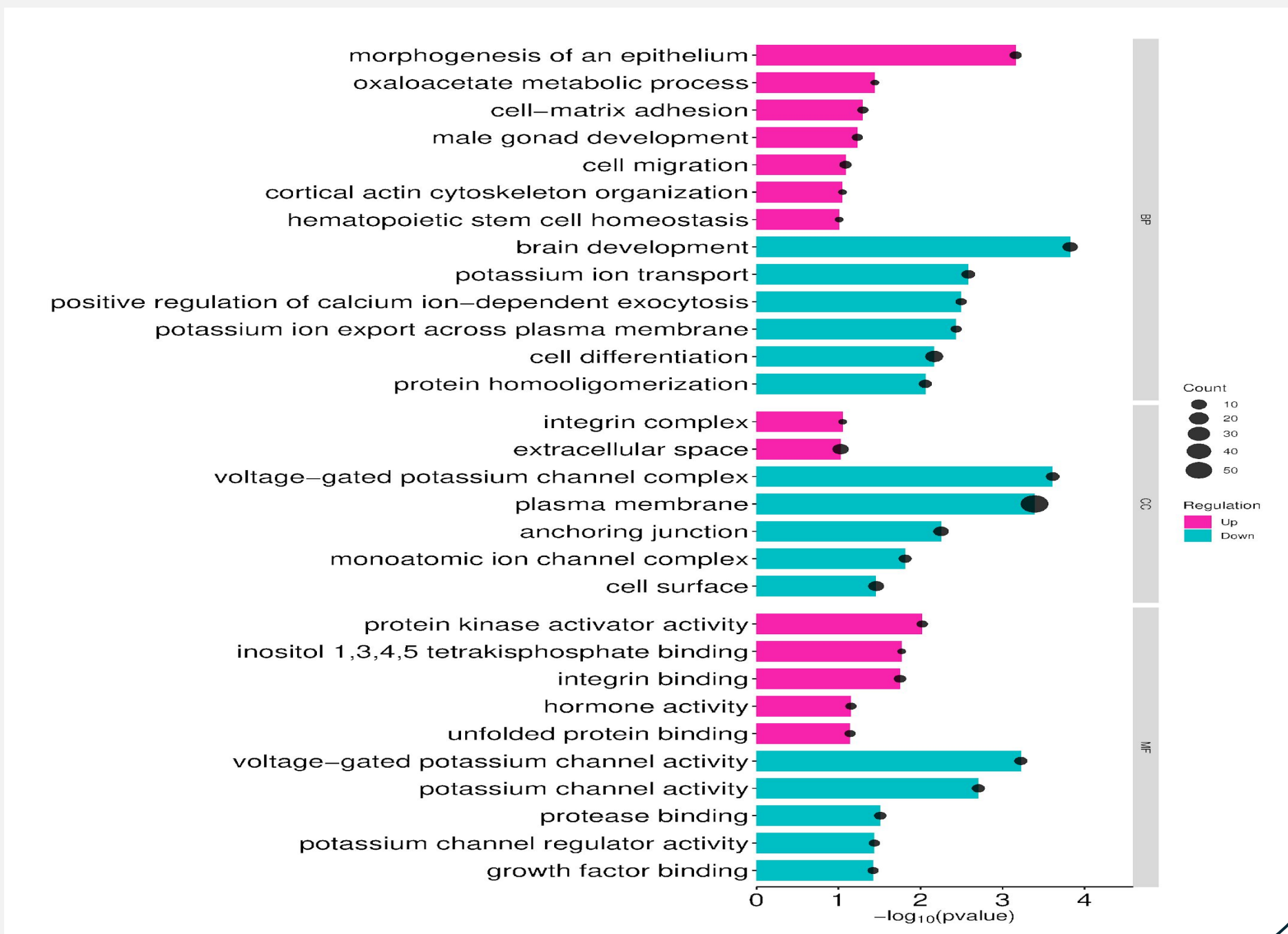
Top 20 up and down-regulated DEGs



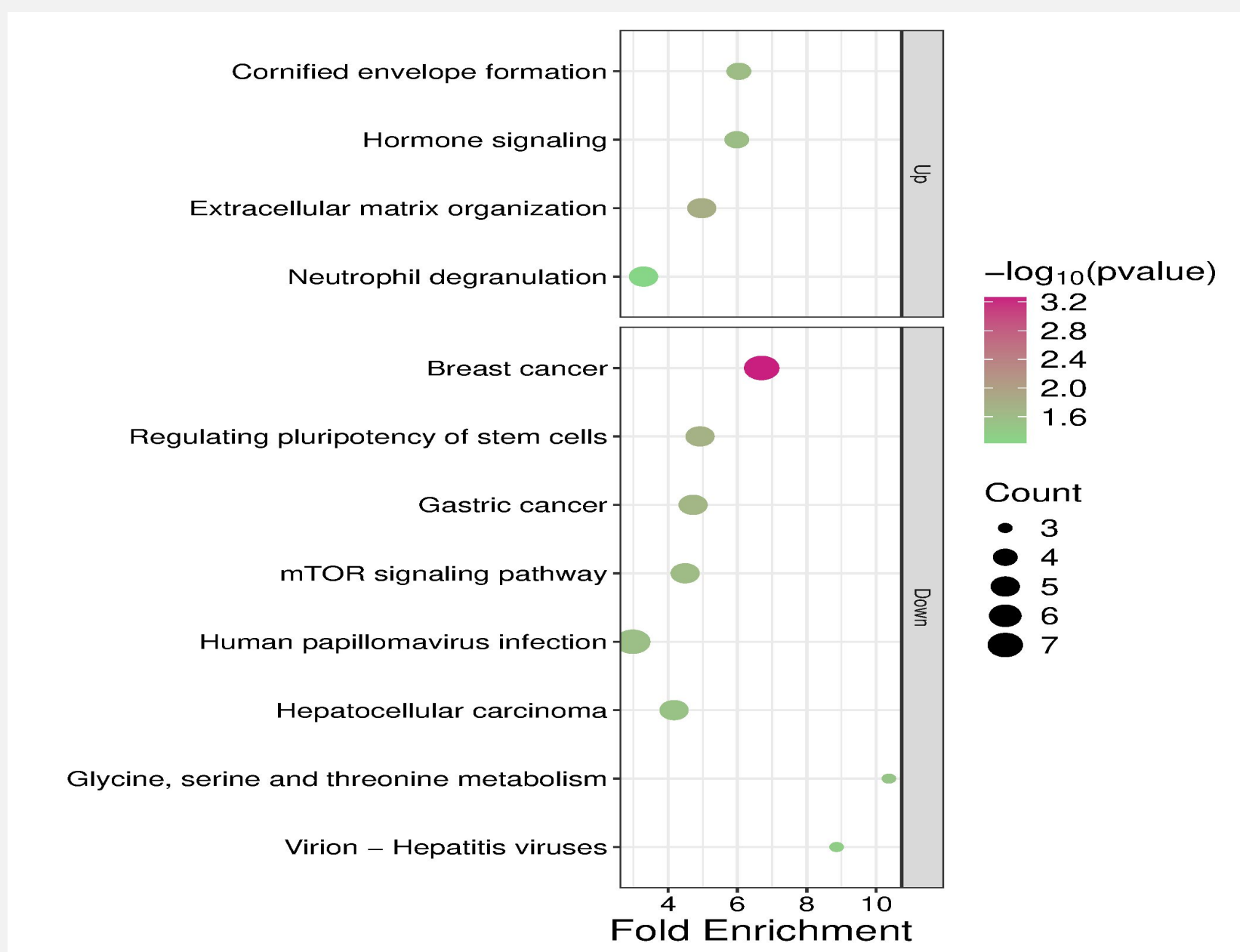
REFERENCES

- Kim, D., Langmead, B. and Salzberg, S.L., 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature methods*, 12(4), pp.357-360.
- Liao, Y., Smyth, G.K. and Shi, W., 2014. featureCounts: an efficient general-purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7), pp.923-930.
- Robinson, M.D., McCarthy, D.J. and Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *bioinformatics*, 26(1), pp.139-140.

Gene Ontology enrichment of DEGs



Pathway enrichment analysis of DEGs



CONCLUSIONS

RNA-seq analysis demonstrated high reproducibility between control and YY2-overexpressing samples and identified 152 up-regulated and 317 down-regulated genes. Several YY2-induced genes such as HBP1, MDC1, SLAMF7, HLA-C, CPLX2, and VDAC1P8, along with the lncRNAs LINC01615 and ZBTB40-IT1, and pathways like extracellular matrix remodeling and neutrophil degranulation were associated with ferroptosis, particularly through oxidative stress, iron-related processes, and microenvironmental regulation rather than direct core ferroptosis machinery. These mechanisms are highly relevant to hematological conditions, where redox imbalance, immune-microenvironment interactions, and iron dysregulation play central roles in disease progression, including leukemia, anemia, and inflammatory blood disorders. Collectively, these findings highlight YY2 as a potential modulator of ferroptosis-associated transcriptional programs and demonstrate the utility of integrated RNA-seq and bioinformatic analyses in uncovering regulatory networks relevant to both cancer biology and hematological disease mechanisms.