Big Data Processing Omics & Bio-Registries - Practical Course

Janes Nagai & Ivan G. Costa
Institute for Computational Genomics
RWTH Aachen University
www.costalab.org

Prof. Dr. med. Steffen Koschmieder
Dept. of Hematology, Oncology, Hemostaseology, SCT
RWTH Aachen University

Institute for Computational Genomics
RWTH AACHEN UNIVERSITY

Objective of the lecture

Practical course on gene expression analysis

Based on a Jupyter notebook

https://jupyter.rwth-aachen.de/

[MMDS_trans] Introduction to Transcriptomics

Cellular Complexity

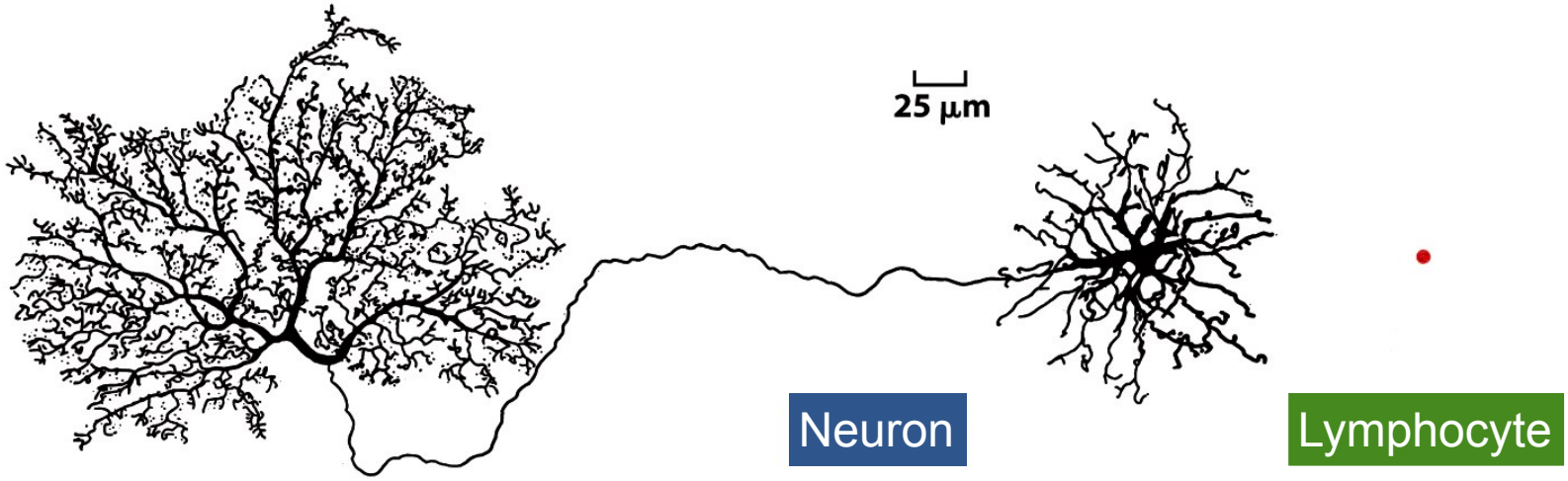


Figure 7-1 Molecular Biology of the Cell 5/e (© Garland Science 2008)

Two cells of a organism have exactly* the same DNA

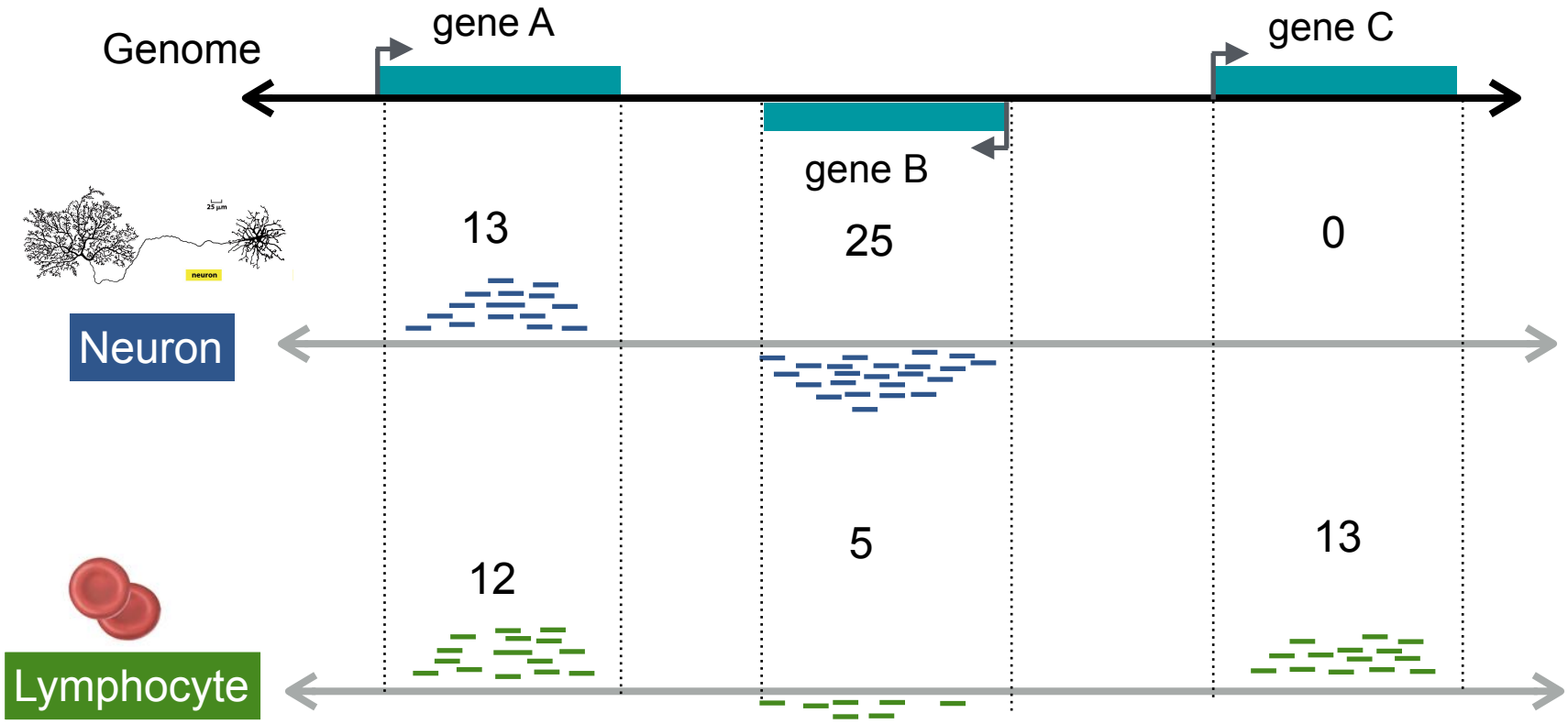
How does this differences arise?

How is cell fate remembered?

*** with exception of somatic mutations and rearrangements of immunological loci**

Gene Quantification

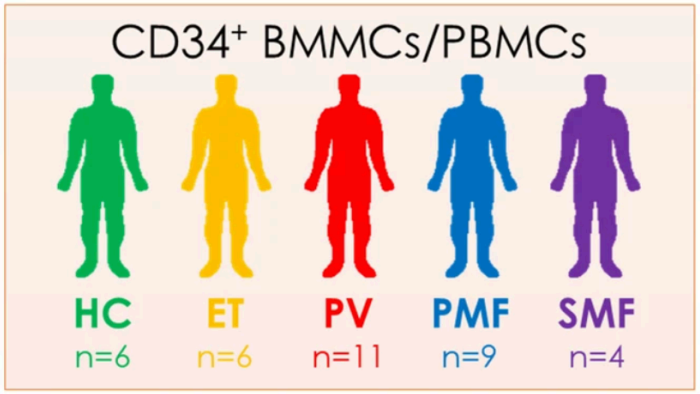
- Perform sequencing for each cell (neuron, lymphocyte)
- Align reads to genome
- Count number of reads inside genes (using known genes annotation)



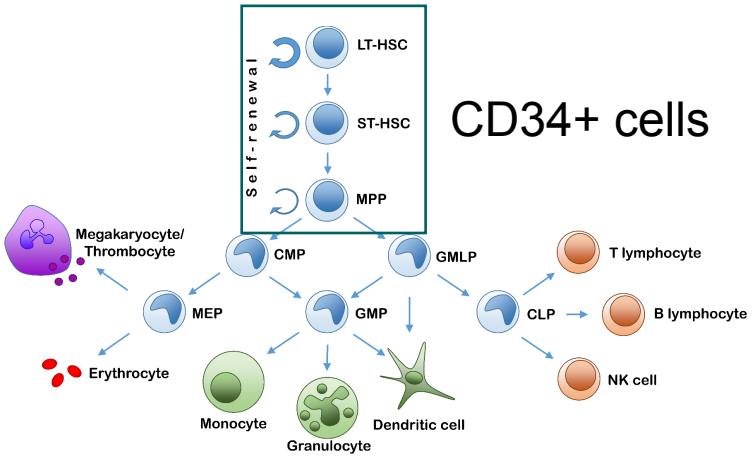
Expression Analysis

- Identify genes related to a particular MPN disease entity
  - example - Baumeister et al. 2021.

- We will consider:
  - Healthy controls (HC)
  - polycythemia vera (PV)
  - essential thrombocythemia (ET)
  - Primary and secondary myelofibrosis (PMF and SMF)

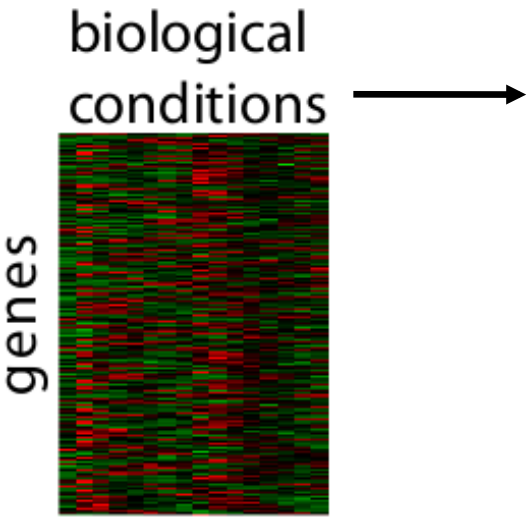


Source: Baumeister et al. 2021.

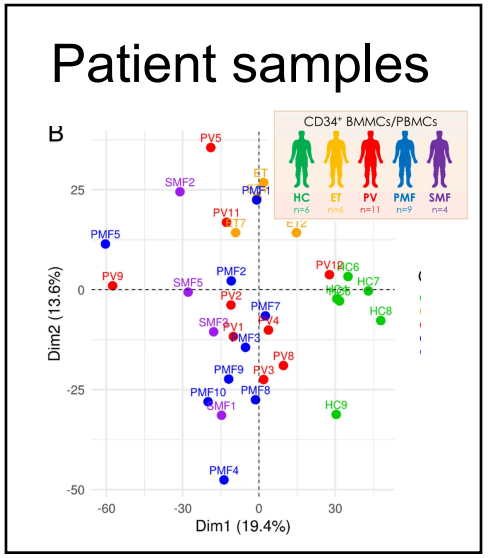


CD34+ cells

Analysis of Gene Expression



gene expression matrix



1- Which genes are up/down regulated after treatment or disease?

- differential analysis / clustering genes

2 - Which cells/patients are more similar?

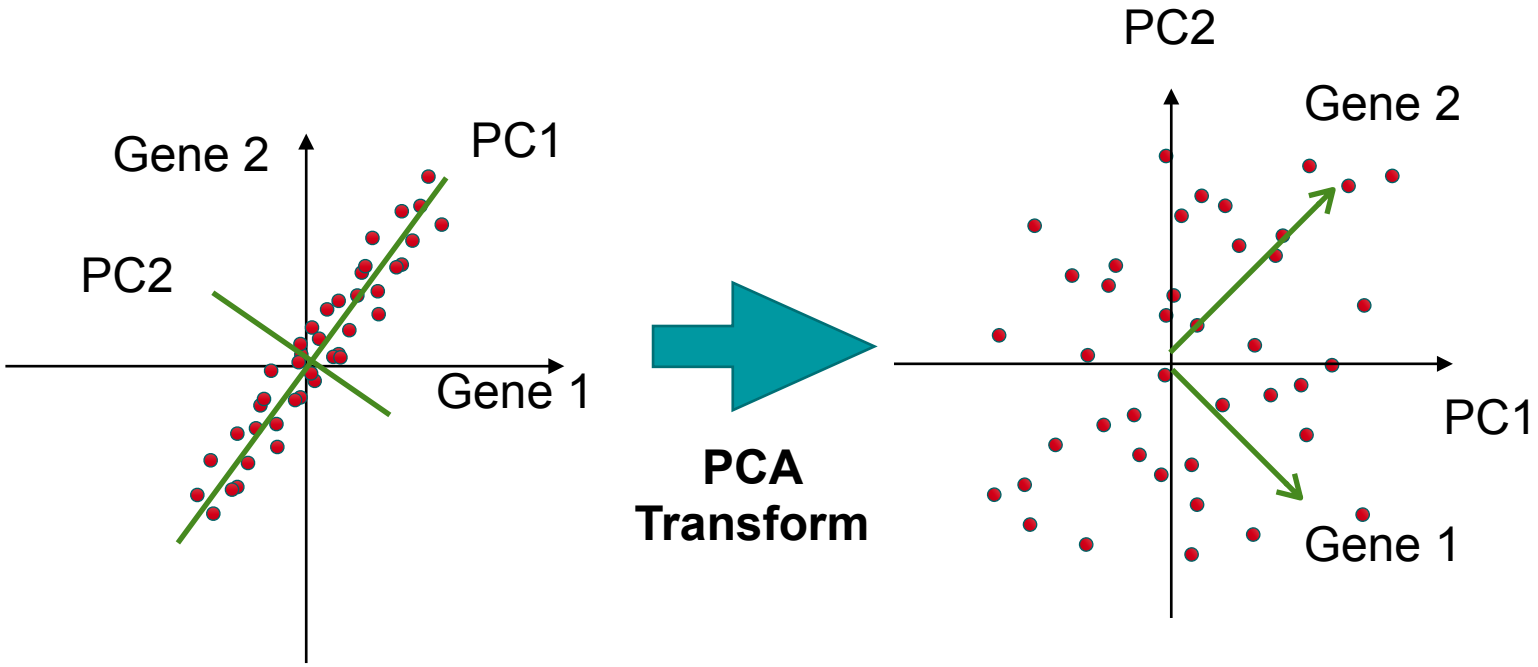
- clustering samples / PCA

3 - How to interpret large lists of genes?

- gene ontology enrichment

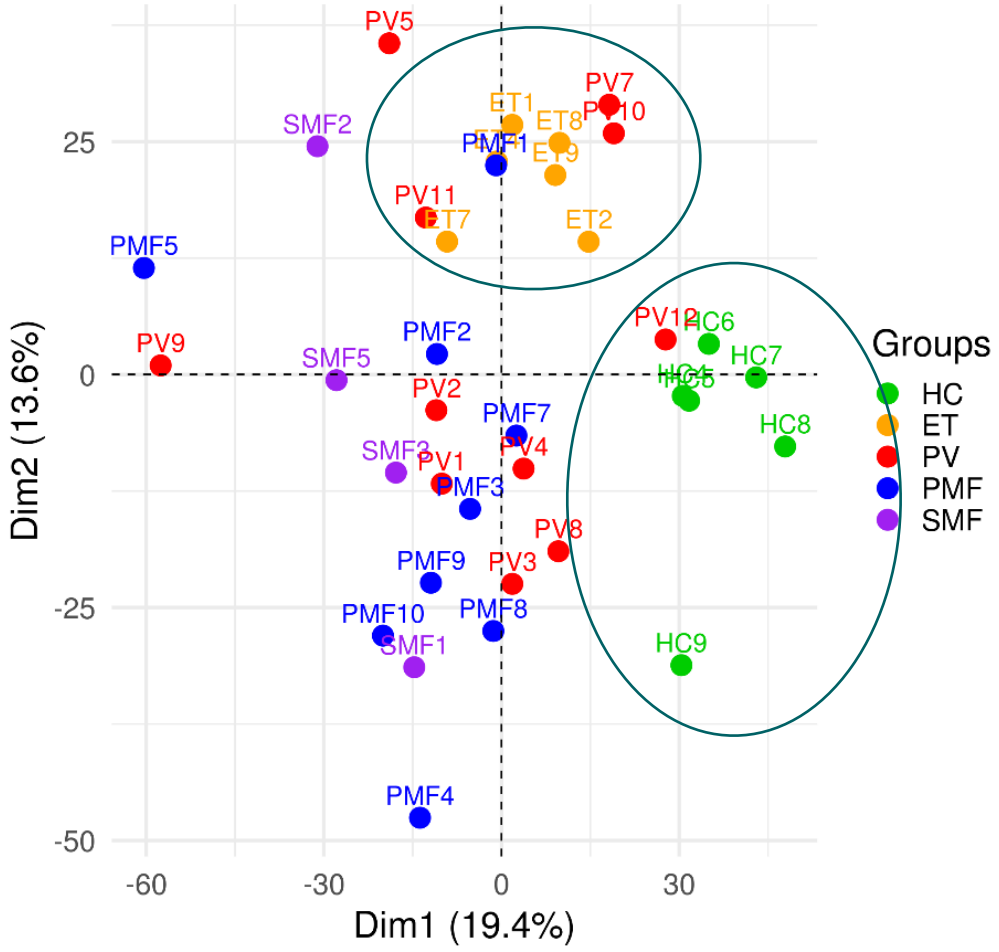
Principal Component Analysis

- **method for dimension reduction**
  - **find combination of genes explaining cells with distinct expression**
- **finding directions with highest variance**



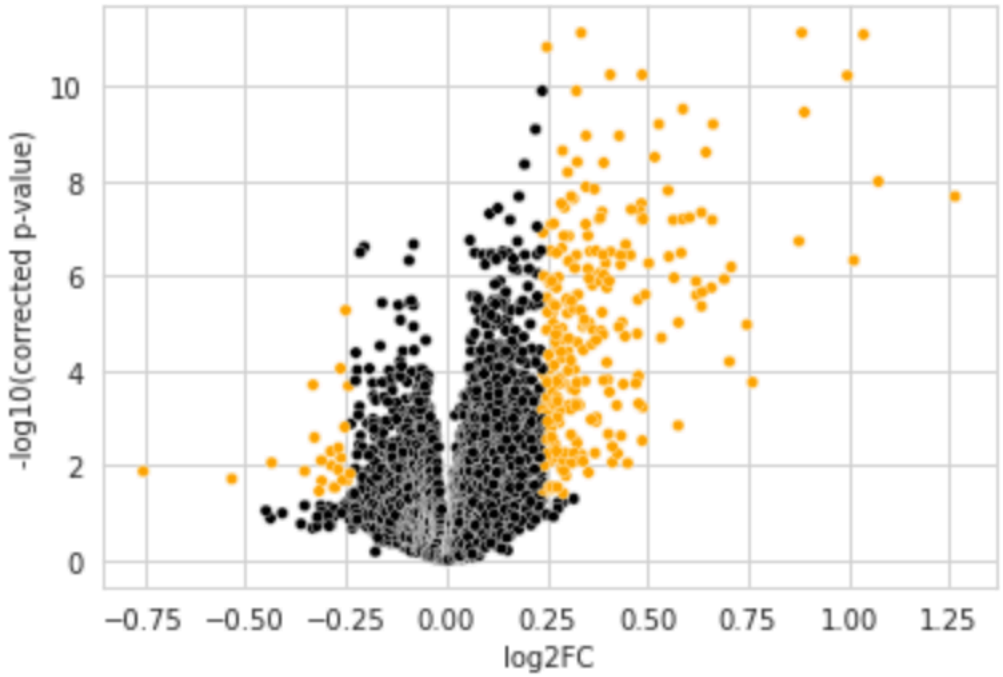
Recommended reading:
Ringner M., *Nature Biotechnology* 26, 303 - 304 (2008)

Gene Expression - PCA Example 2



Differential Gene expression

Volcano Plot - combine p-value and fold change

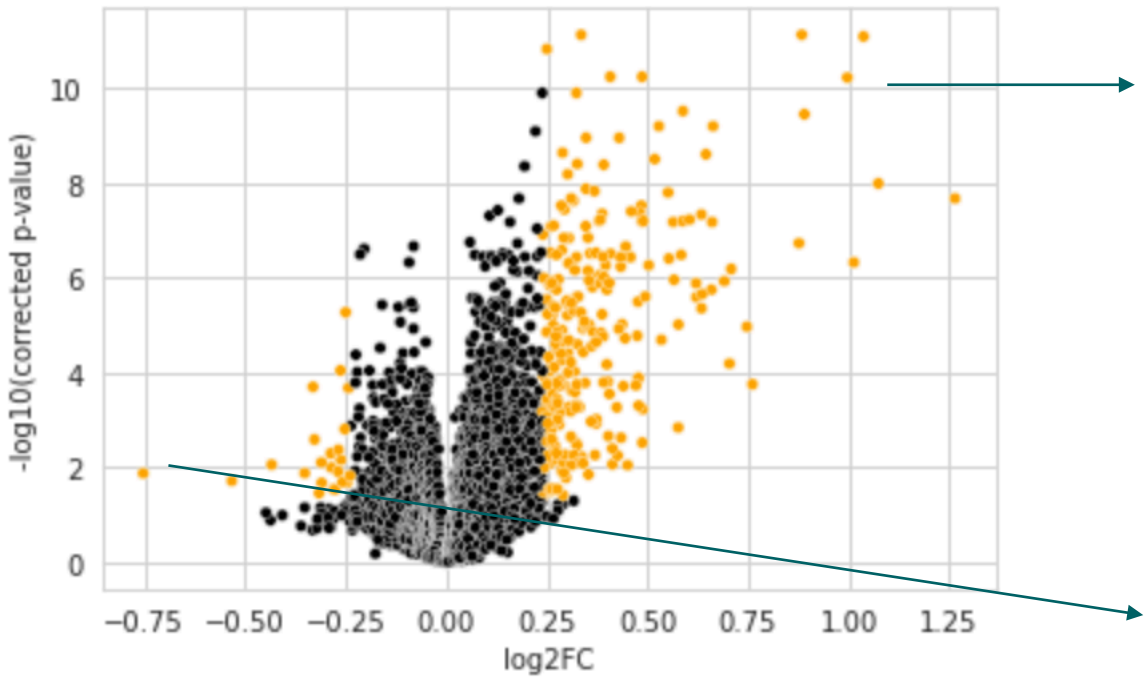


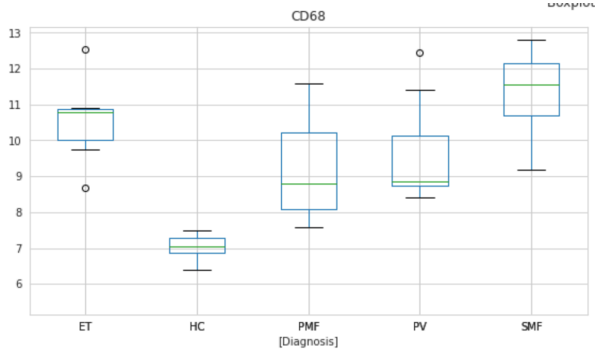
HC genes

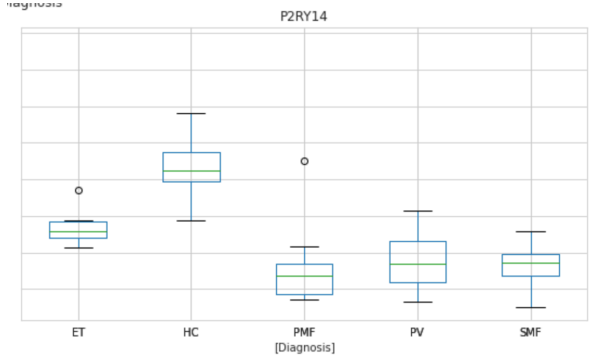
MPN genes

Differential Gene expression

Volcano Plot - combine p-value and fold change



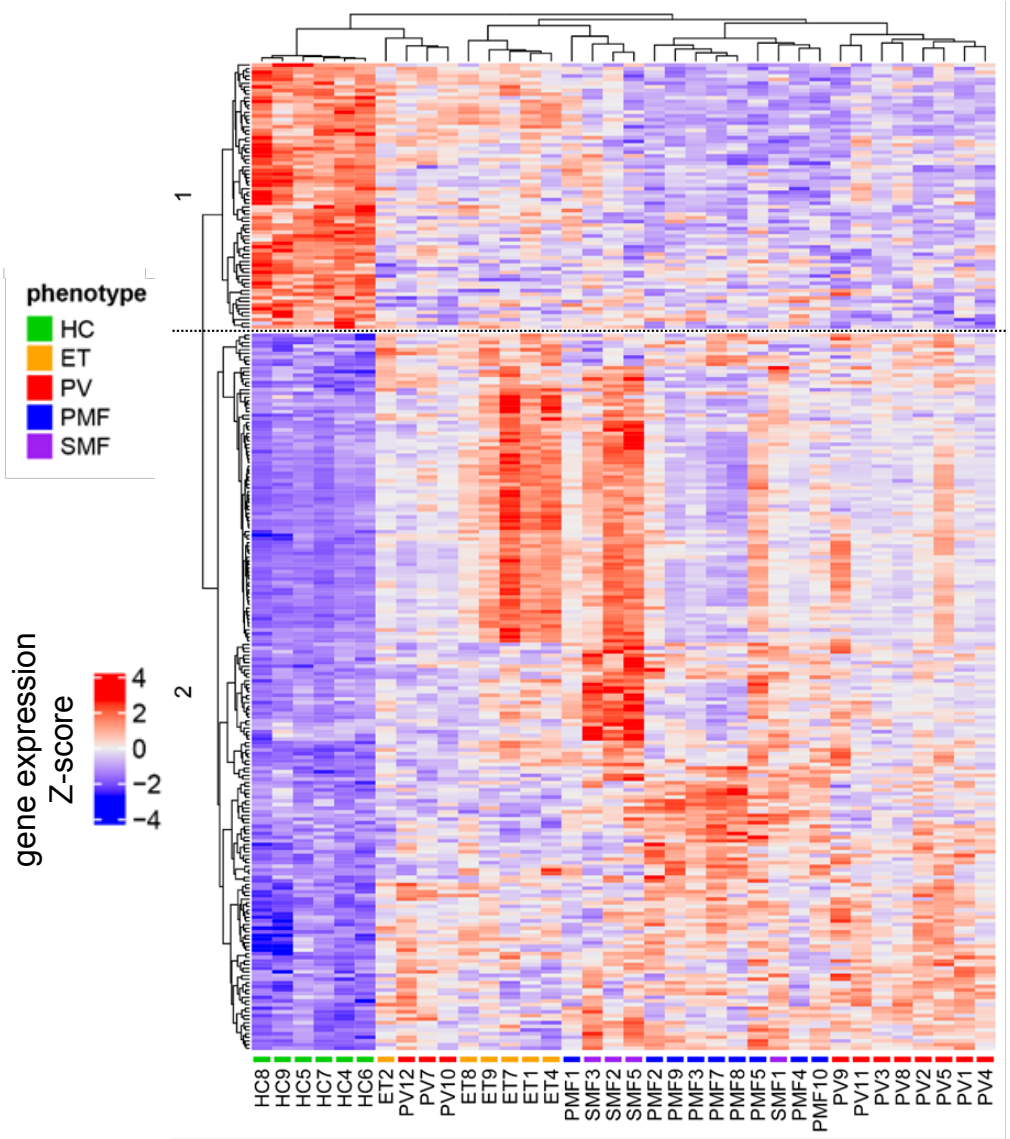




HC genes

MPN genes

Hierarchical Clustering - Average Linkage



- Two clear expression for HC vs. MPNs
- PMF+SMF group together
- Some PV samples group with ET samples

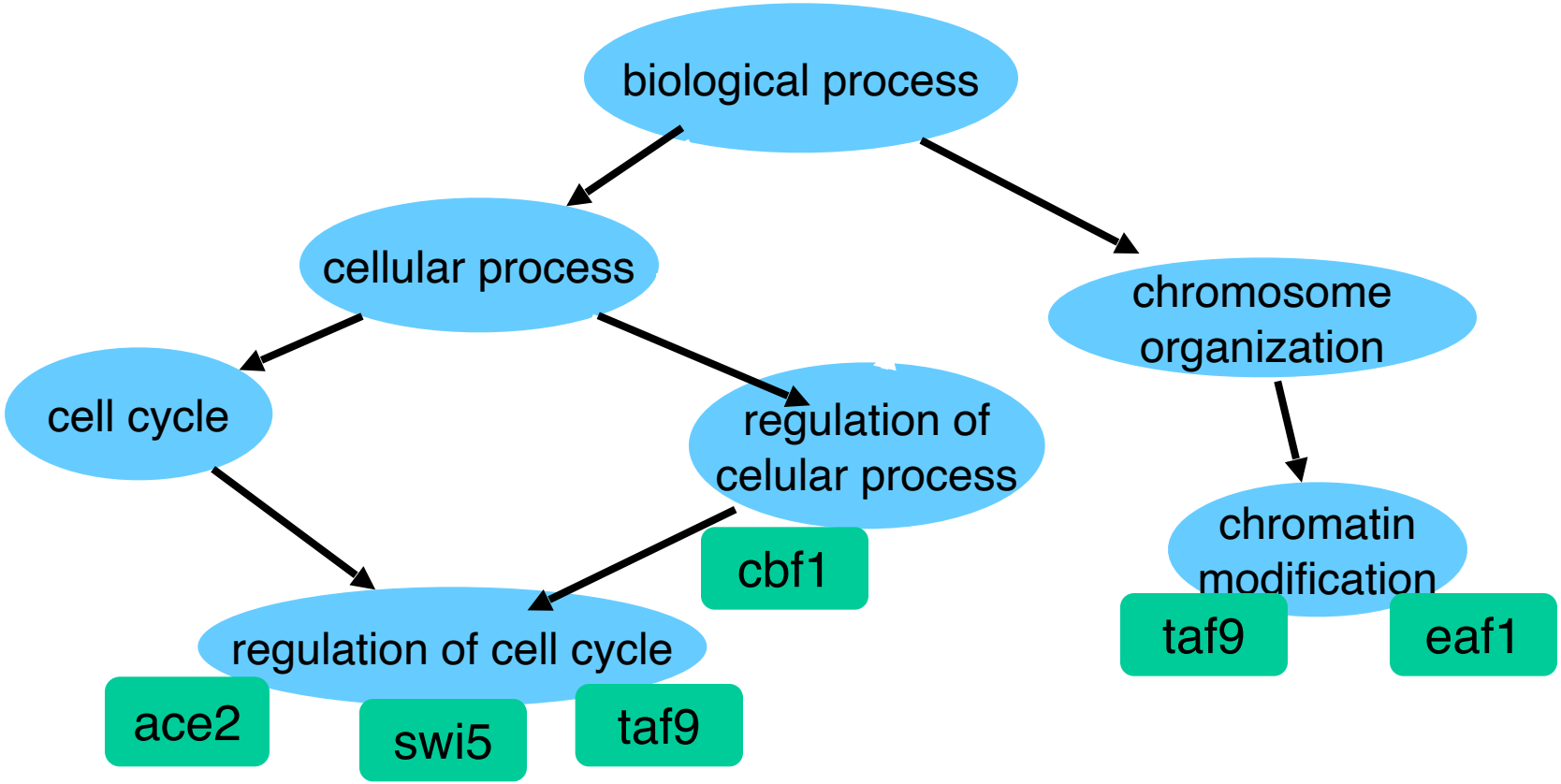
Functional Analysis

Clustering/Differential Expression (DE) returns lists of hundreds of genes How to functionally characterize these?

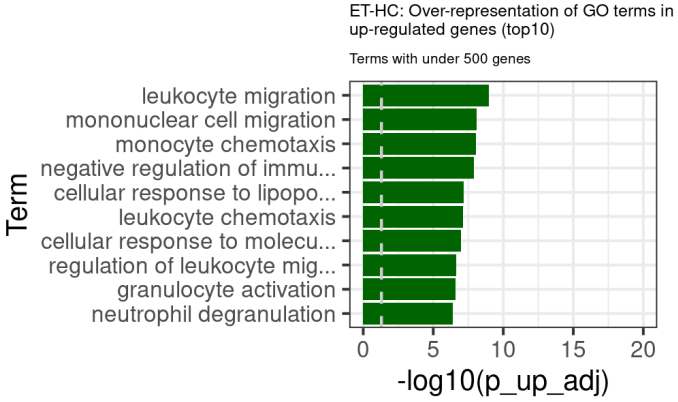
Solution - Relate these genes to annotations from databases

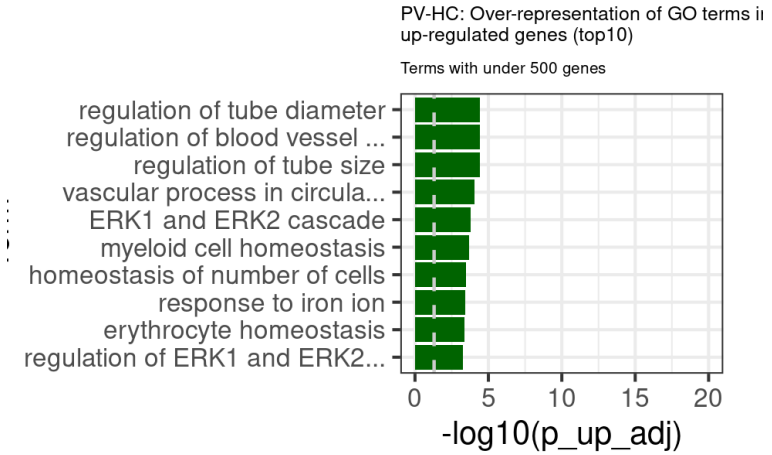
- Gene Ontology, pathways, gene sets, disease ontology, …

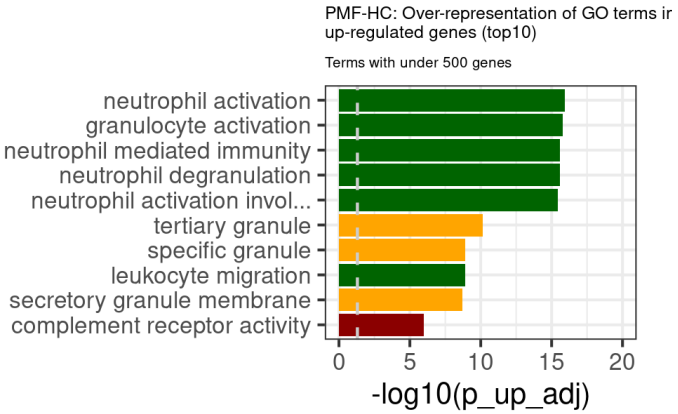
Gene Ontology

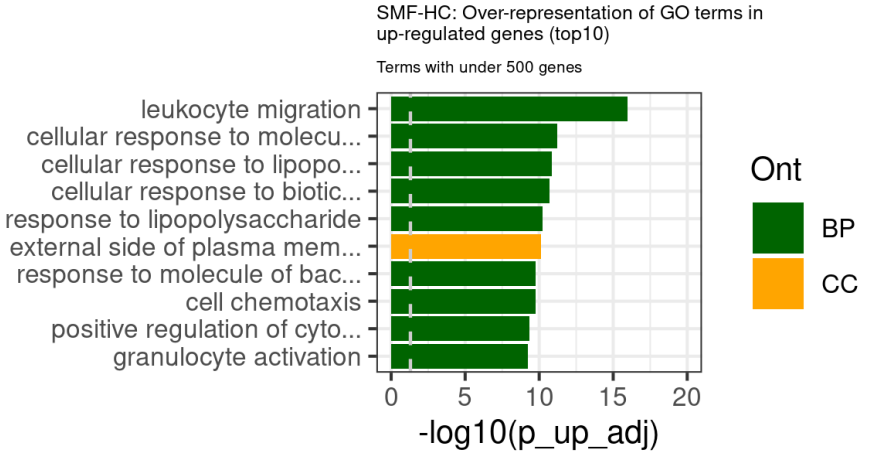


GO Analysis in MPNs









Institute for Computational Genomics

RWTH AACHEN UNIVERSITY

Resume

- **The course will be done in Jupiter lab**

https://jupyter.rwth-aachen.de/

Please select the following profile:

[MMDS_trans] Introduction to Transcriptomics

www.costalab.org

Institute for Computational Genomics

RWTH AACHEN UNIVERSITY