BIOSTATISTICS GROUP
WROCLAW, POLAND

The effect of transcriptomic annotations in breast cancer differential gene expression study

R. Stępień[1], J. Szyda[1,2], B. Czech[1], M. Mielczarek[1,2*]
[1]Biostatistics Group, Wroclaw University of Environmental and Life Sciences; Wroclaw, Poland. [2]National Research Institute of Animal Production; Cracow-Balice, Poland. *presenting author: magda.mielczarek@upwr.edu.pl

Material & Methods

- RNA-seq of MCF7 breast cancer (Cancer reads **ERR358487-88**) and normal tissues (Normal reads **ERR358485-86**)
- Illumina HiSeq 2000, 2 x 150 bp
- The library size: 73,974,766 - 97,983,949 reads

- One differential gene expression (DGE) piepeline (FastQC, Trimmomatic, Kallisto, DESeq2) → four approaches
- Gene Set Enrichment Analysis (GSEA) of DEGs
- PCA based on expression of all genes
- T-test based on expression of Naderi breast cancer prognosis down- and upregulated genes (doi.org/10.1038/sj.onc.1209920)

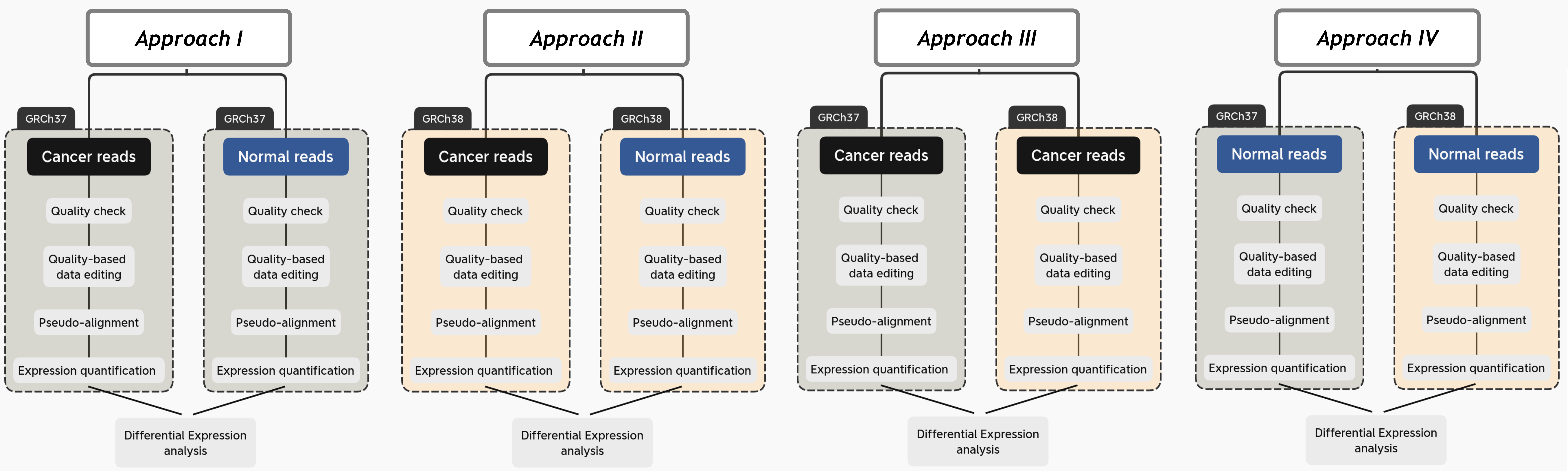
Objectives

Investigation of the transcriptome annotation effect on:

- DGE of all genes
- genes related to survival prognosis in breast cancer

Conclusions

- The overall DGE outcomes were not identical between GRCh37 and GRCh38 annotations
- …however, the transcriptome annotation had no effect on genes related to survival prognosis in breast cancer



Approach I

GRCh37 – Cancer reads: Quality check → Quality-based data editing → Pseudo-alignment → Expression quantification
GRCh37 – Normal reads: Quality check → Quality-based data editing → Pseudo-alignment → Expression quantification
→ Differential Expression analysis

Approach II

GRCh38 – Cancer reads: Quality check → Quality-based data editing → Pseudo-alignment → Expression quantification
GRCh38 – Normal reads: Quality check → Quality-based data editing → Pseudo-alignment → Expression quantification
→ Differential Expression analysis

Approach III

GRCh37 – Cancer reads: Quality check → Quality-based data editing → Pseudo-alignment → Expression quantification
GRCh38 – Cancer reads: Quality check → Quality-based data editing → Pseudo-alignment → Expression quantification
→ Differential Expression analysis

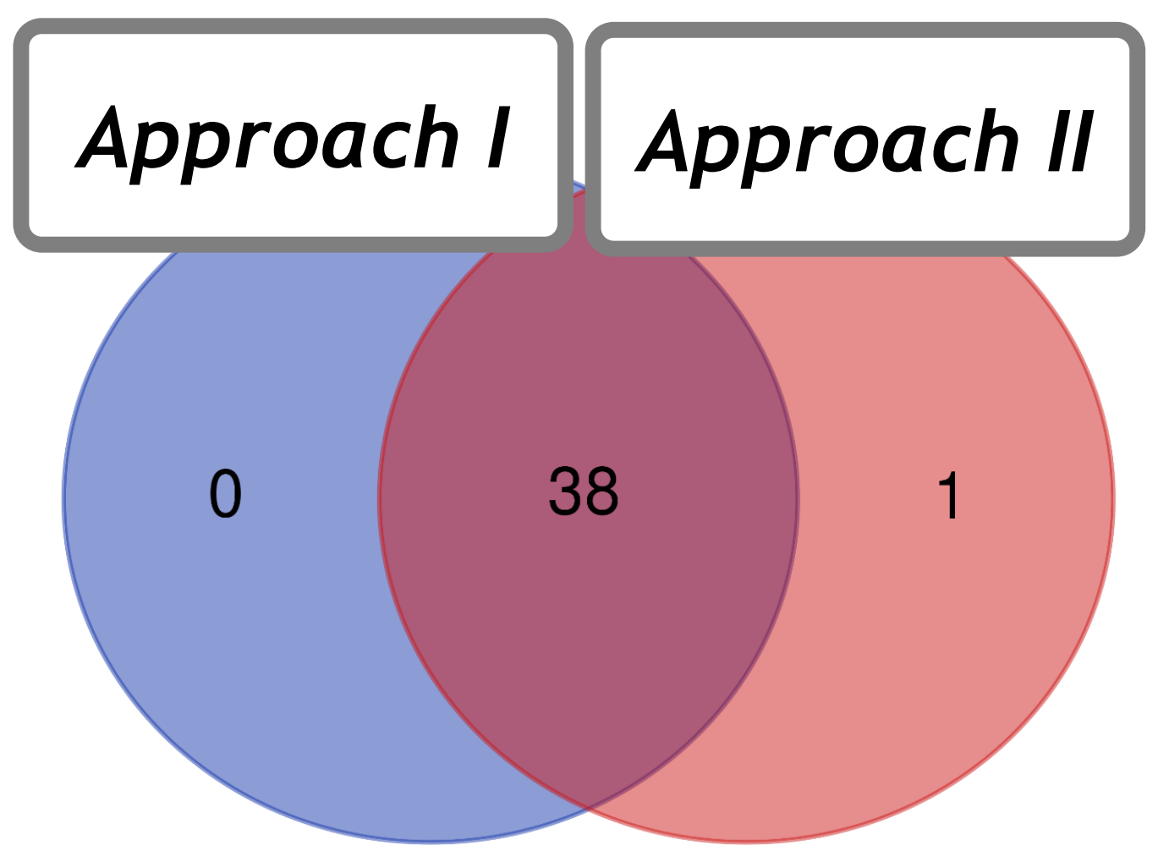
Approach IV

GRCh37 – Normal reads: Quality check → Quality-based data editing → Pseudo-alignment → Expression quantification
GRCh38 – Normal reads: Quality check → Quality-based data editing → Pseudo-alignment → Expression quantification
→ Differential Expression analysis

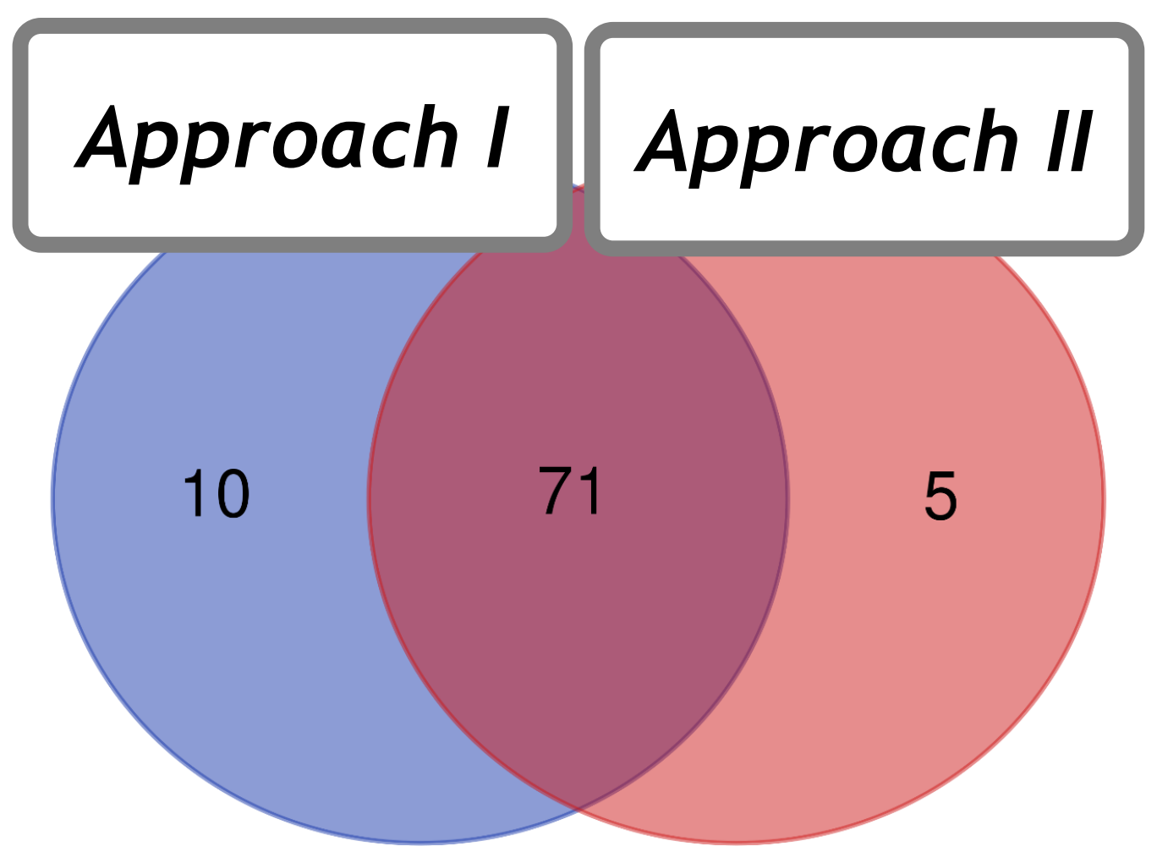
Results

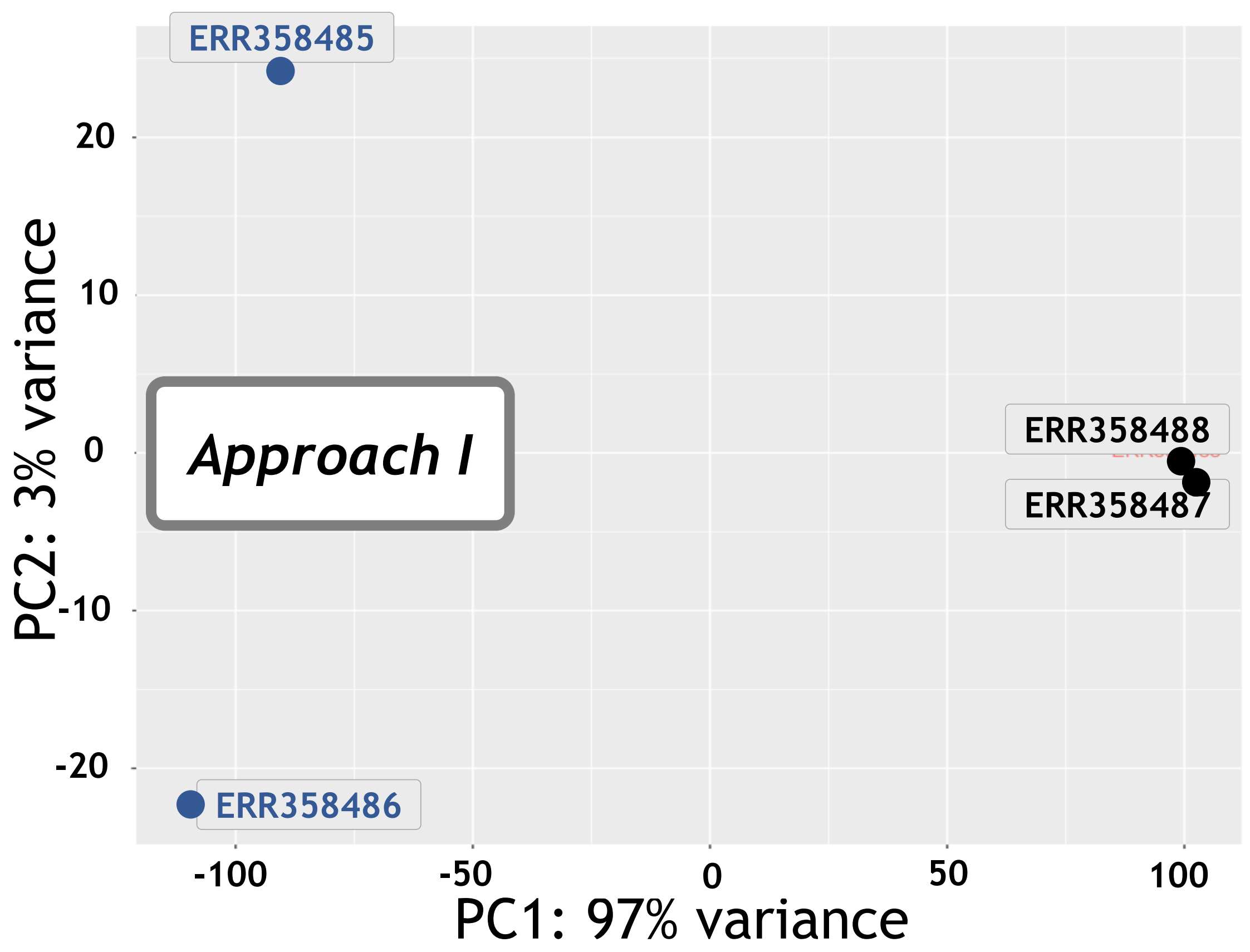
GSEA:

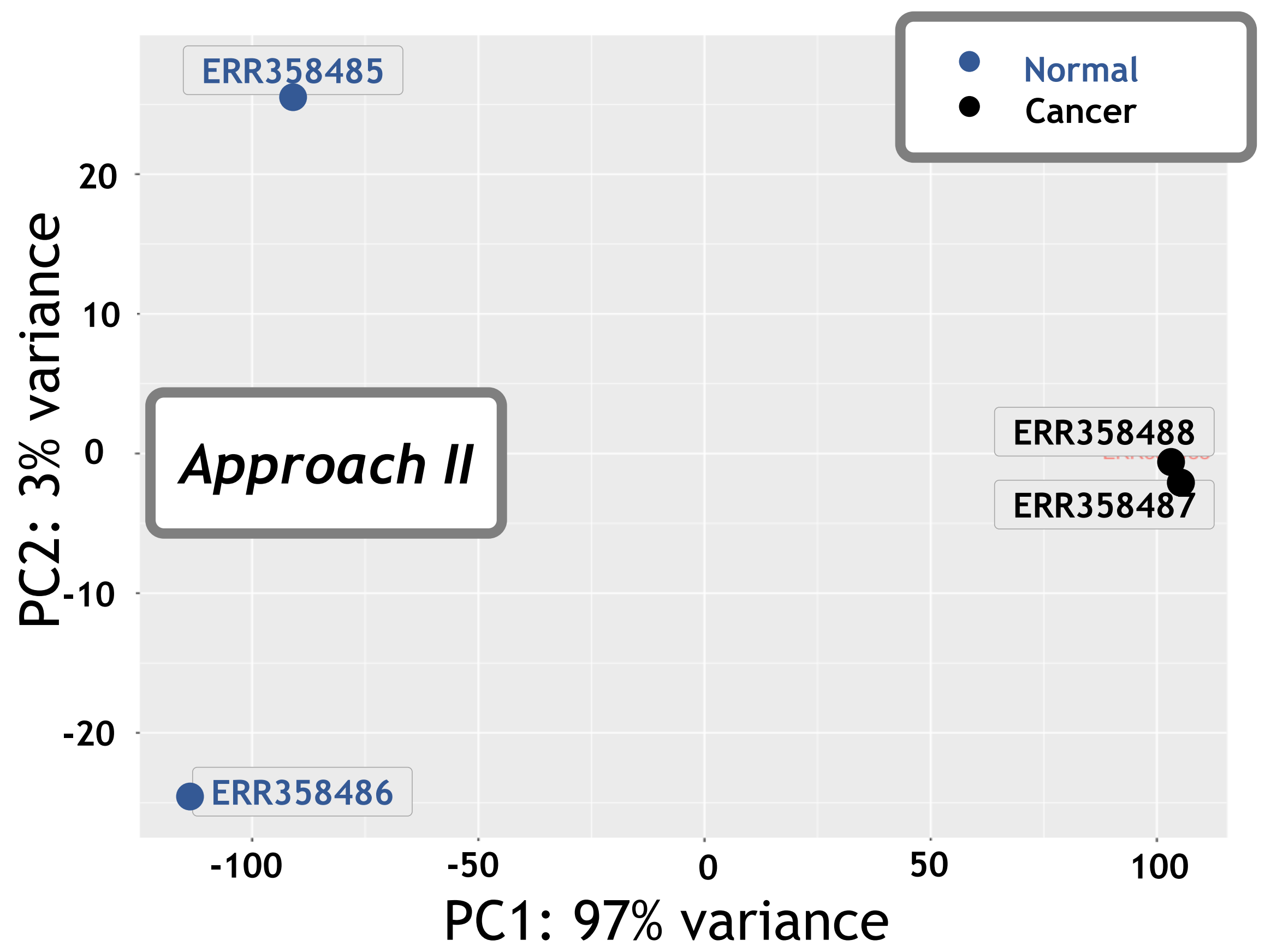
- Hallmark pathways

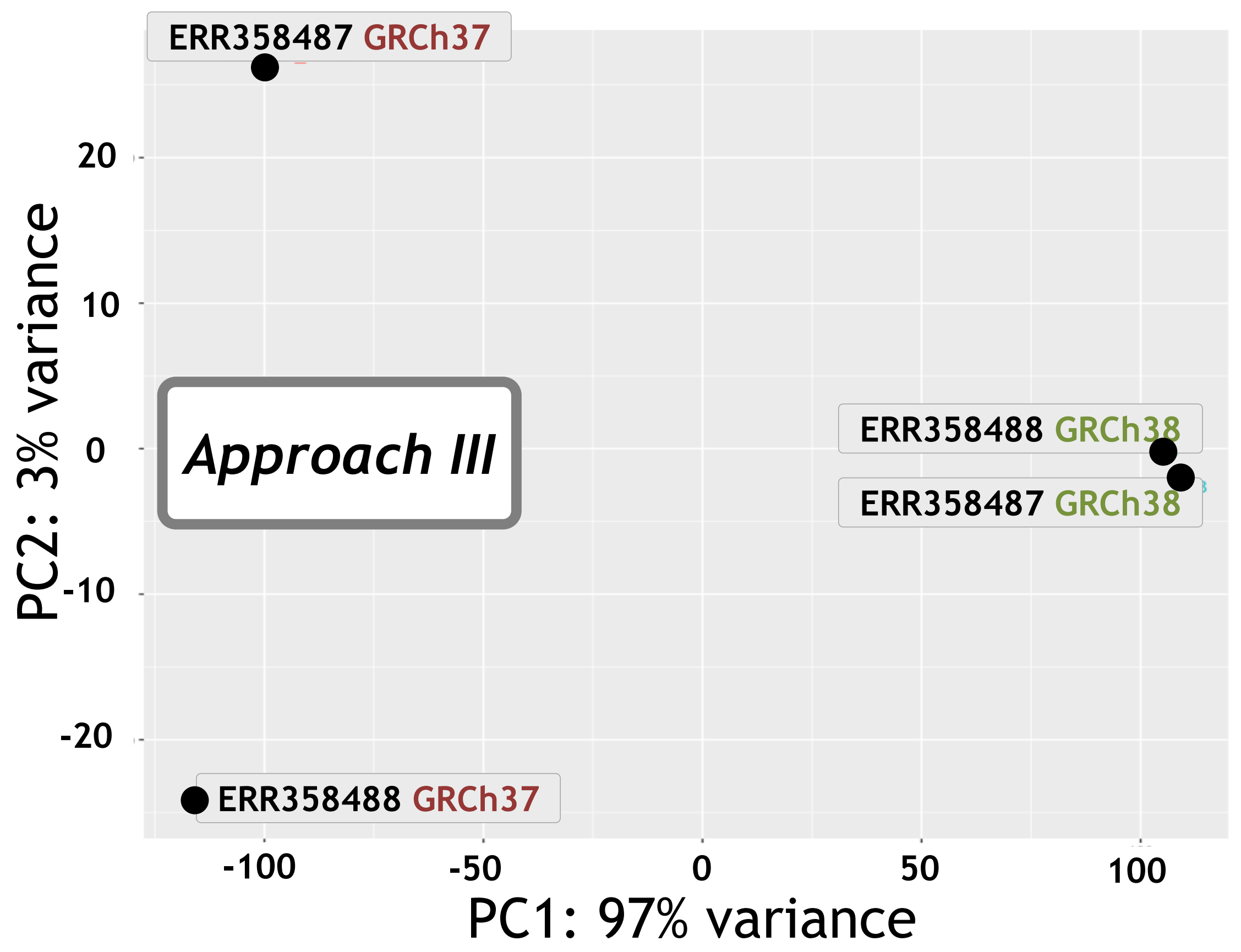


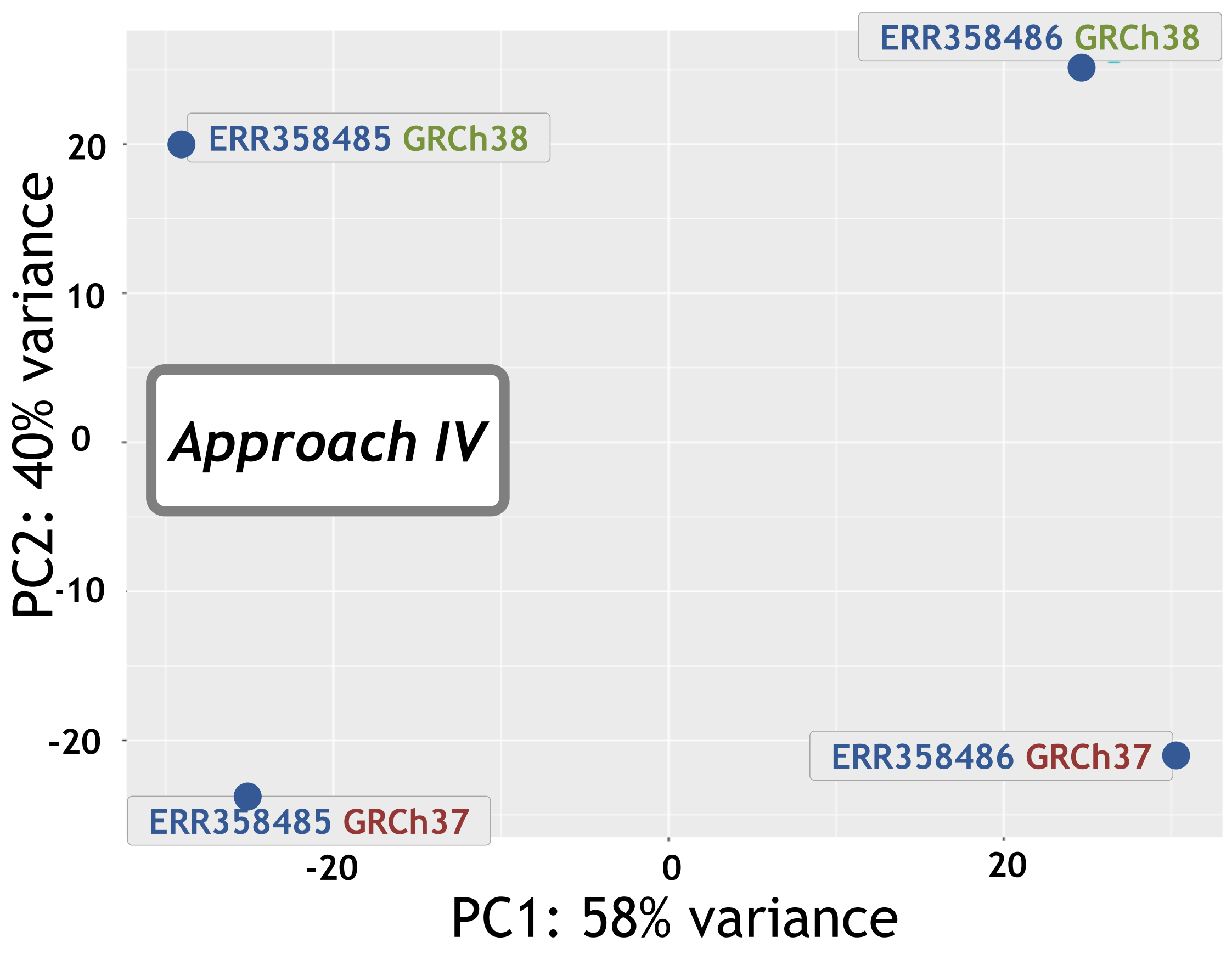
- KEGG pathways



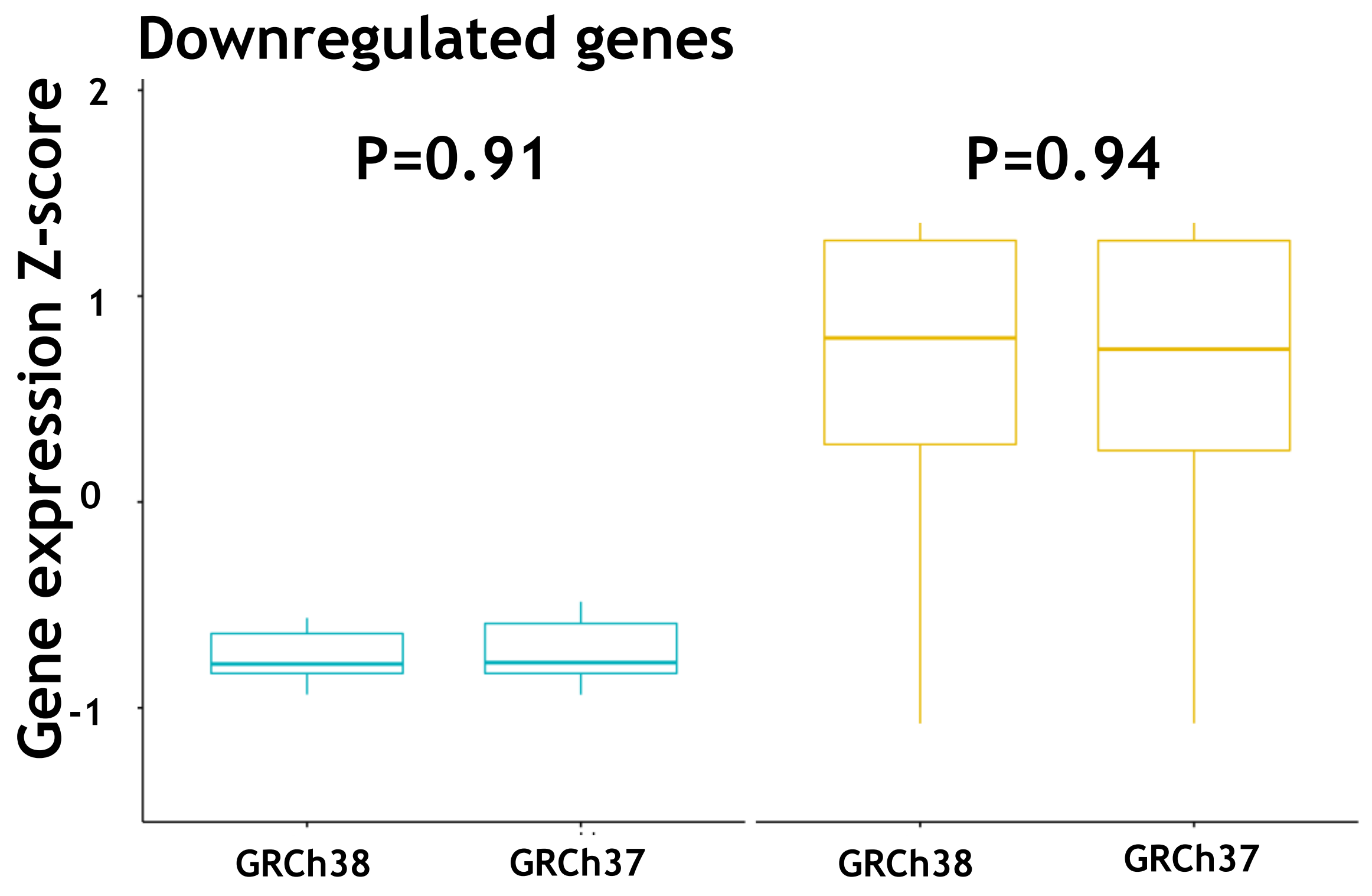


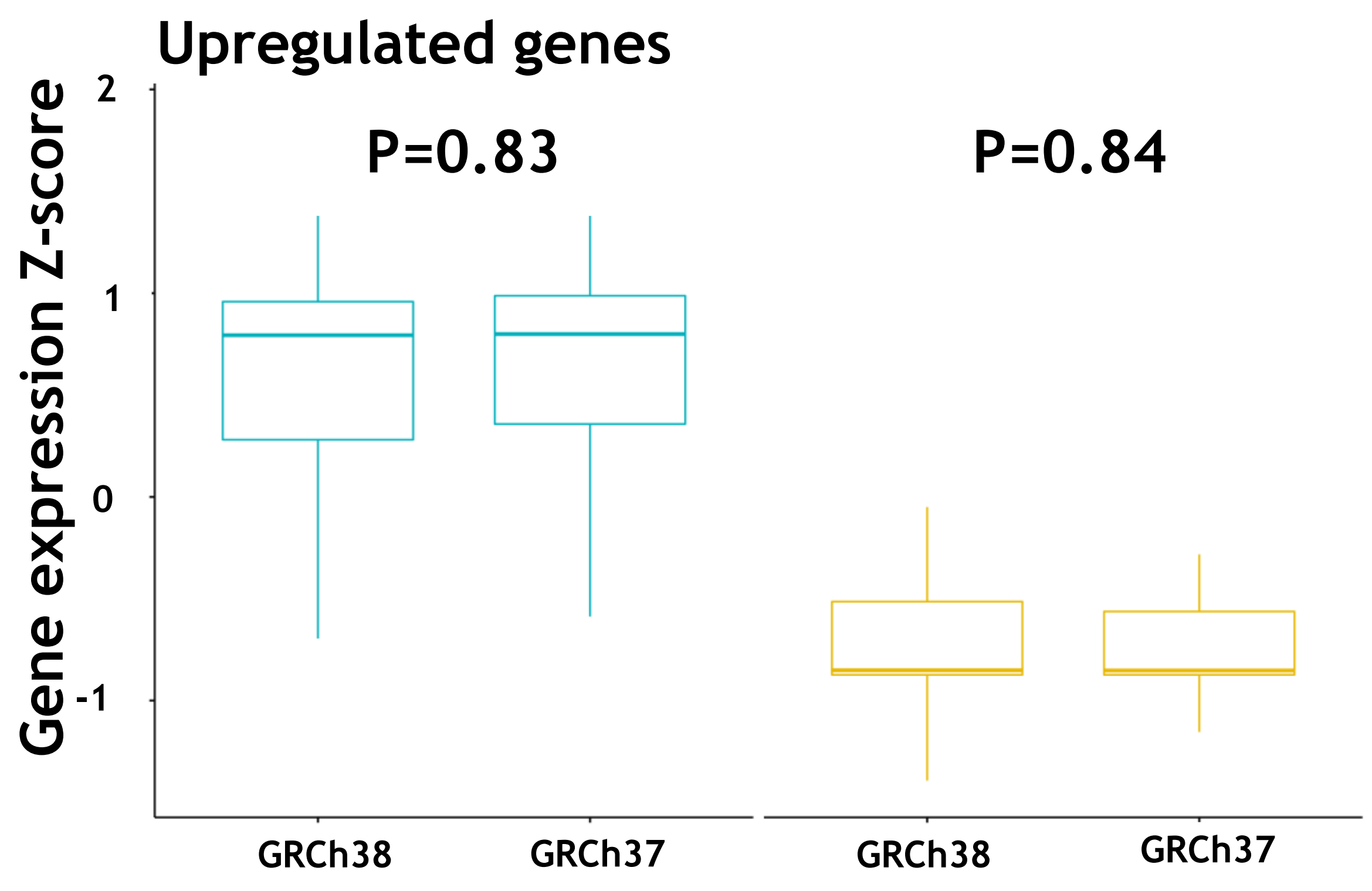






Sample 🟦 cancer 🟨 normal





Acknowledgements to Poznan Supercomputing and Networking Center

WROCŁAW UNIVERSITY OF ENVIRONMENTAL AND LIFE SCIENCES
Magda Mielczarek, PhD
THETA Biostatistics Group
Wroclaw University of Environmental and Life Sciences
magda.mielczarek@upwr.edu.pl