



THETA
Statistical Genetics Group



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

R. Stępień¹, J. Szyda^{1,2}, B. Czech¹, **M. Mielczarek^{1,2*}**

¹Biostatistics Group, Wrocław University of Environmental and Life Sciences; Wrocław, Poland

²National Research Institute of Animal Production; Cracow-Balice, Poland.

Objectives

- **Motivation**

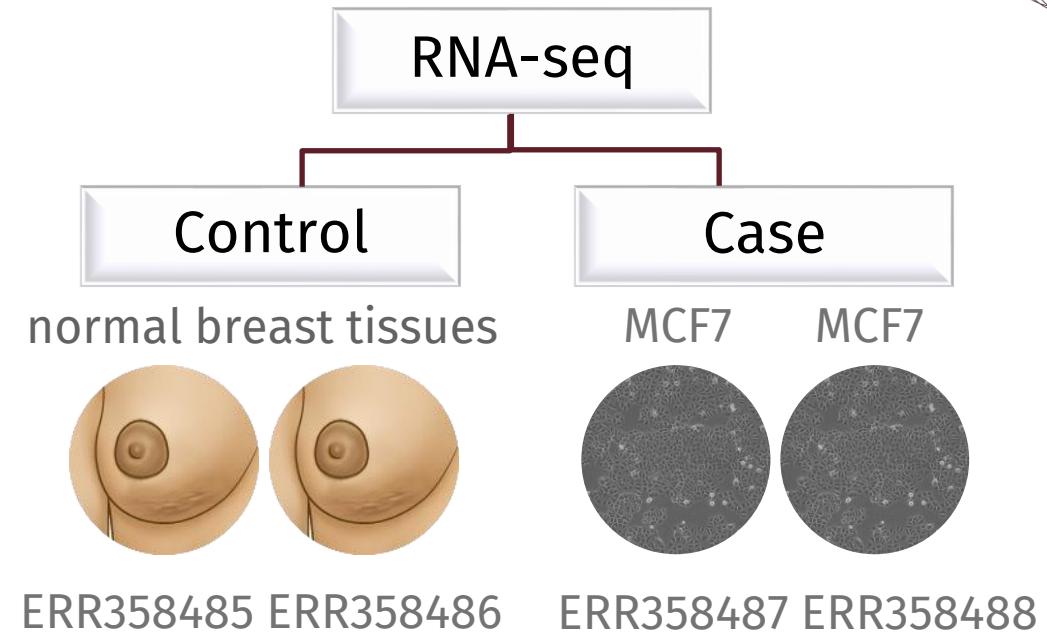
- In NGS data analysis, including differential gene expression (DGE) study, the most recent reference genome/transcriptome (**GRCh38**) is regarded as the annotation standard
- However, **GRCh37** is also considered relevant

- **Goals**

- Investigation of the transcriptome annotation version effect on:
 - DGE of all genes
 - genes related to survival prognosis in breast cancer

Material

- **RNA-seq:**
 - MCF7 breast cancer samples and normal breast tissues
 - NCBI BioProject ID: PRJEB4829
 - Illumina HiSeq 2000, 2 x 100 bp
 - Library size:
73,974,766 - 97,983,949 reads
- All known transcripts (coding and non-coding) annotated to **GRCh37** and **GRCh38**
- **Naderi breast cancer prognosis down- and upregulated genes**

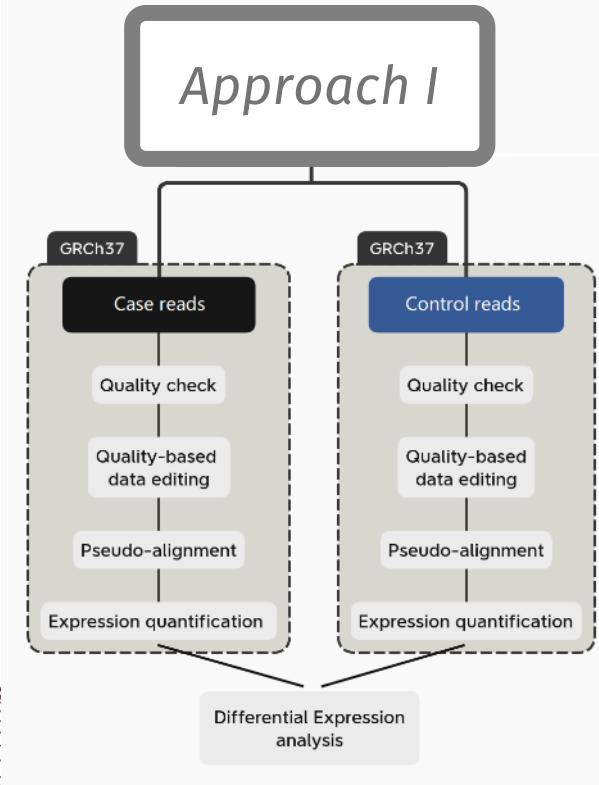


A gene-expression signature to predict survival in breast cancer across independent data sets

[A Naderi](#), [A E Teschendorff](#), [N L Barbosa-Morais](#), [S E Pinder](#), [A R Green](#), [D G Powe](#), [J F R Robertson](#),
[Aparicio](#), [I O Ellis](#), [J D Brenton](#)✉ & [C Caldas](#)✉

Methods

- Four approaches to assess DGE



Fastqc

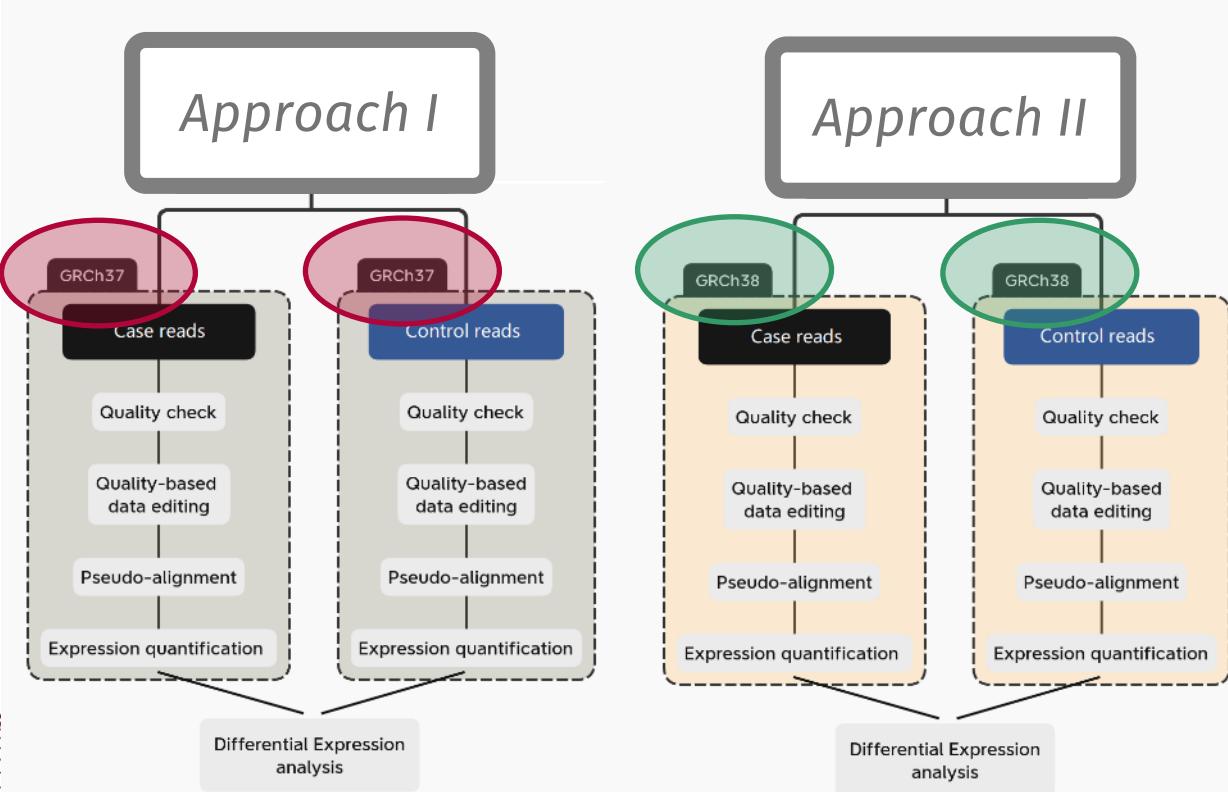
Trimmomatic

Kallisto

DESeq2

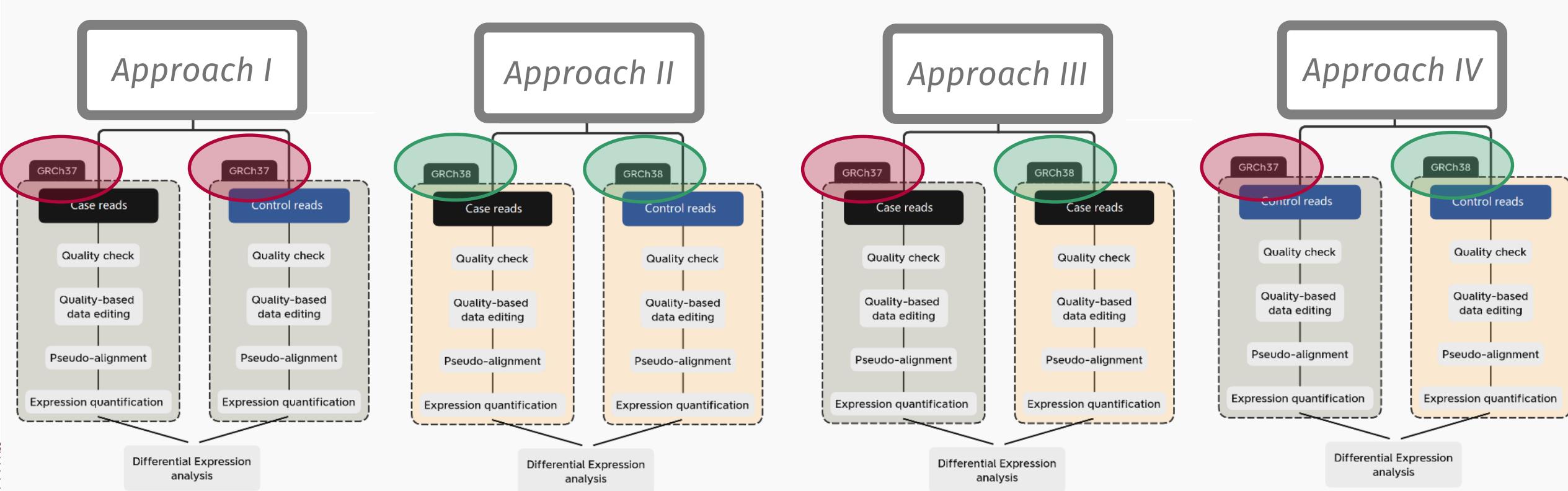
Methods

- Four approaches to assess DGE

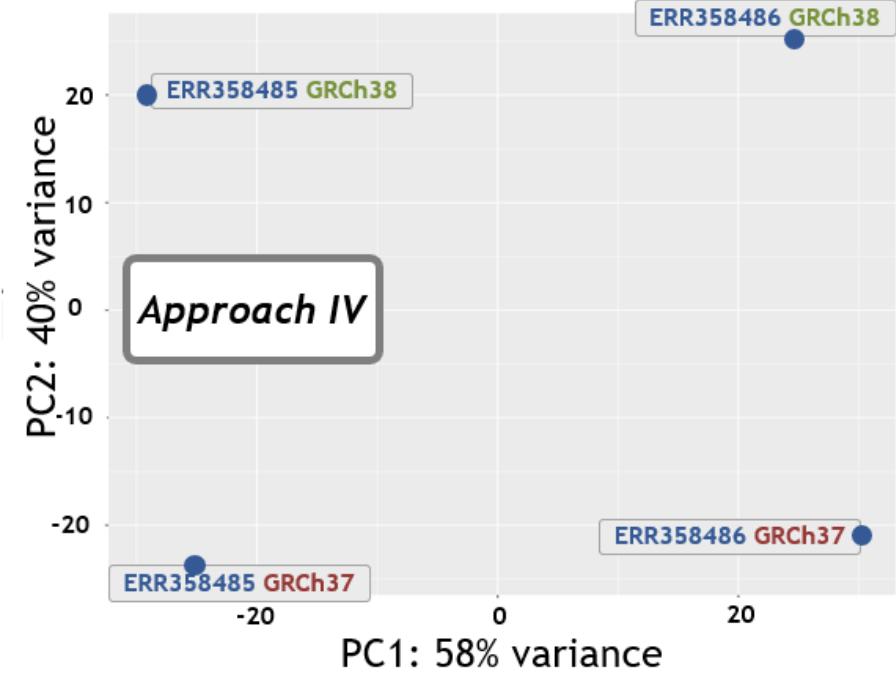
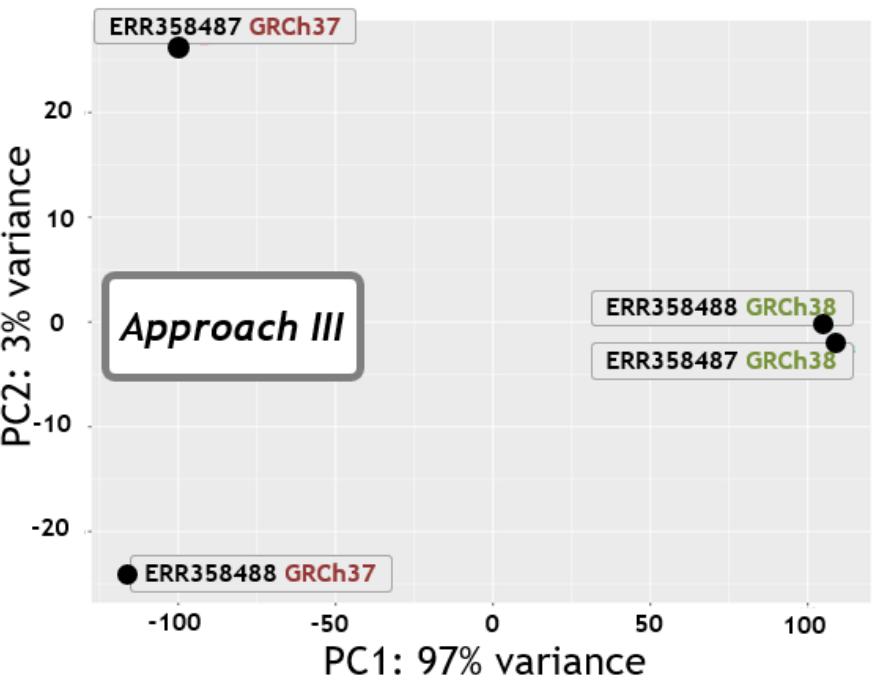
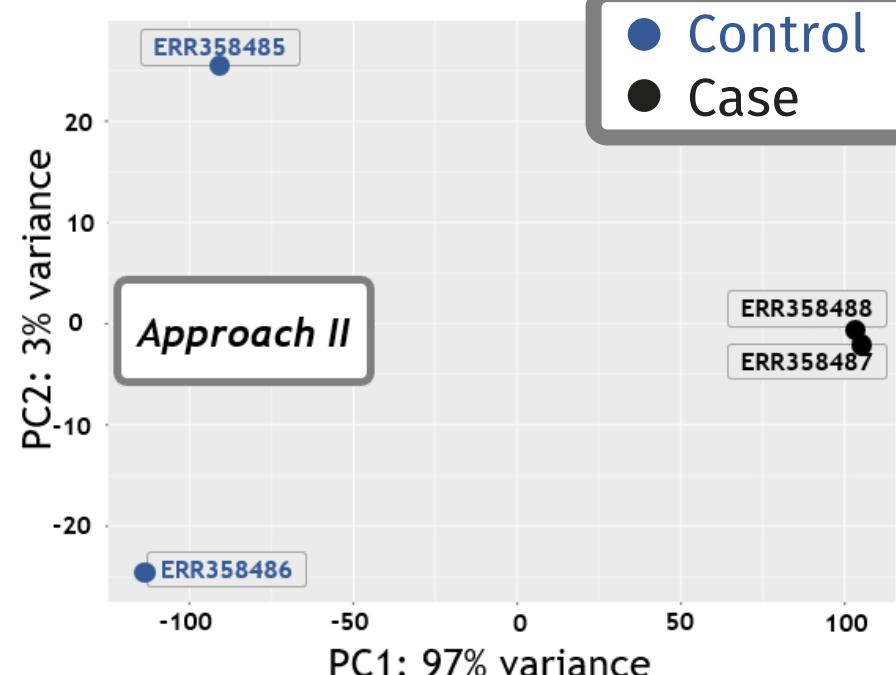
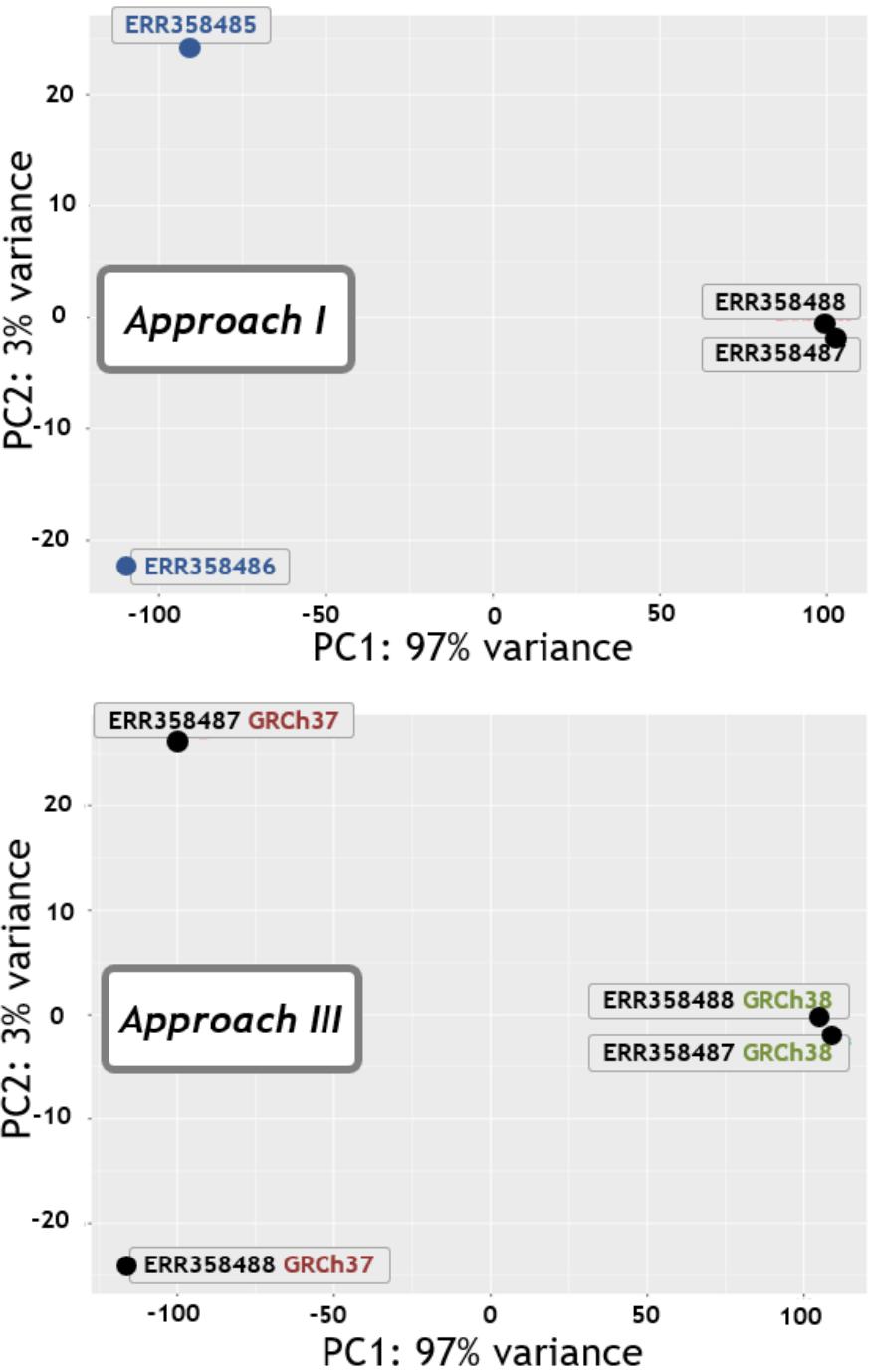


Methods

- Four approaches to assess DGE

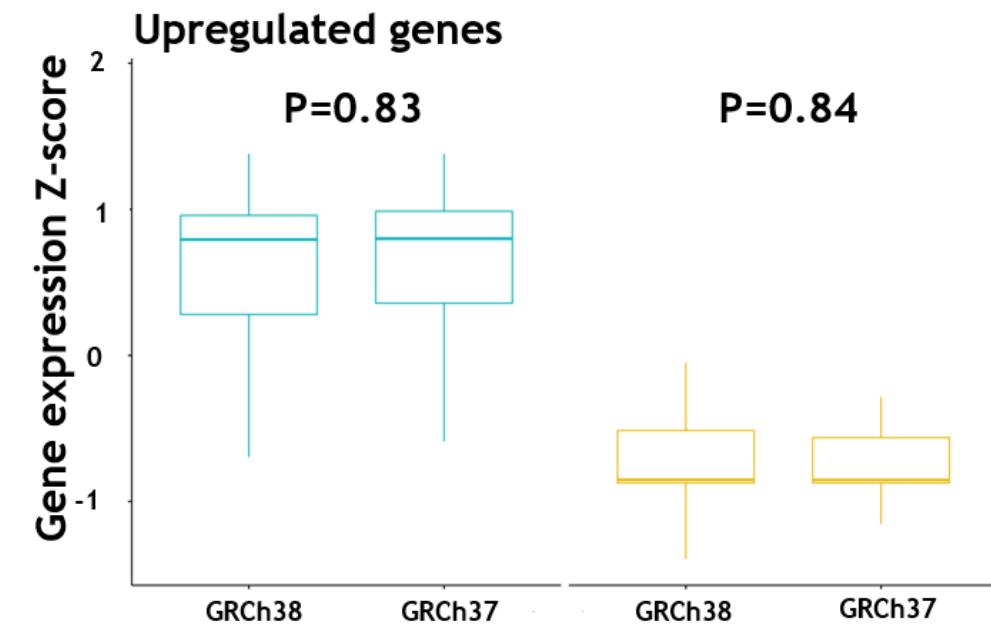
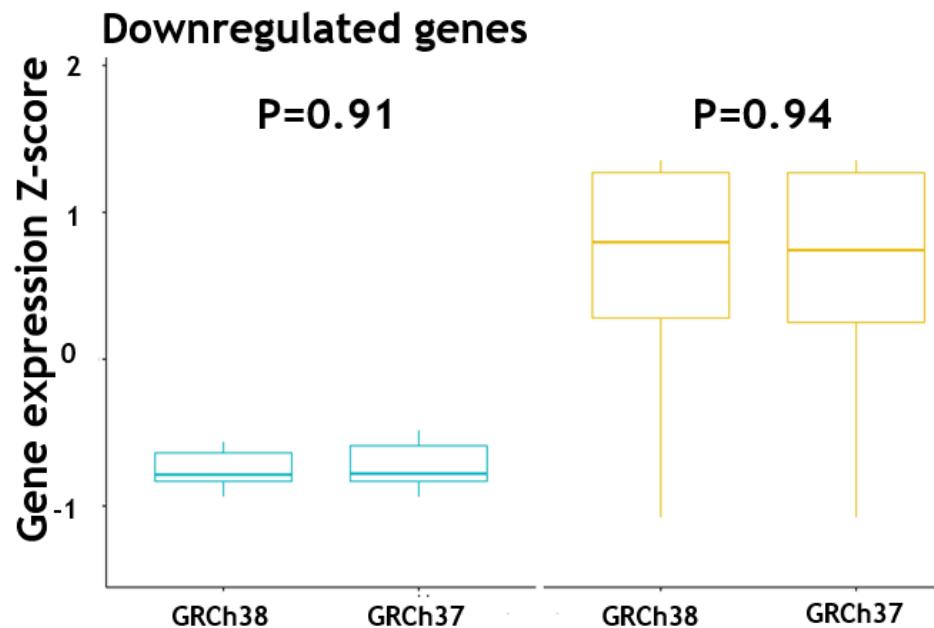


Results: PCA

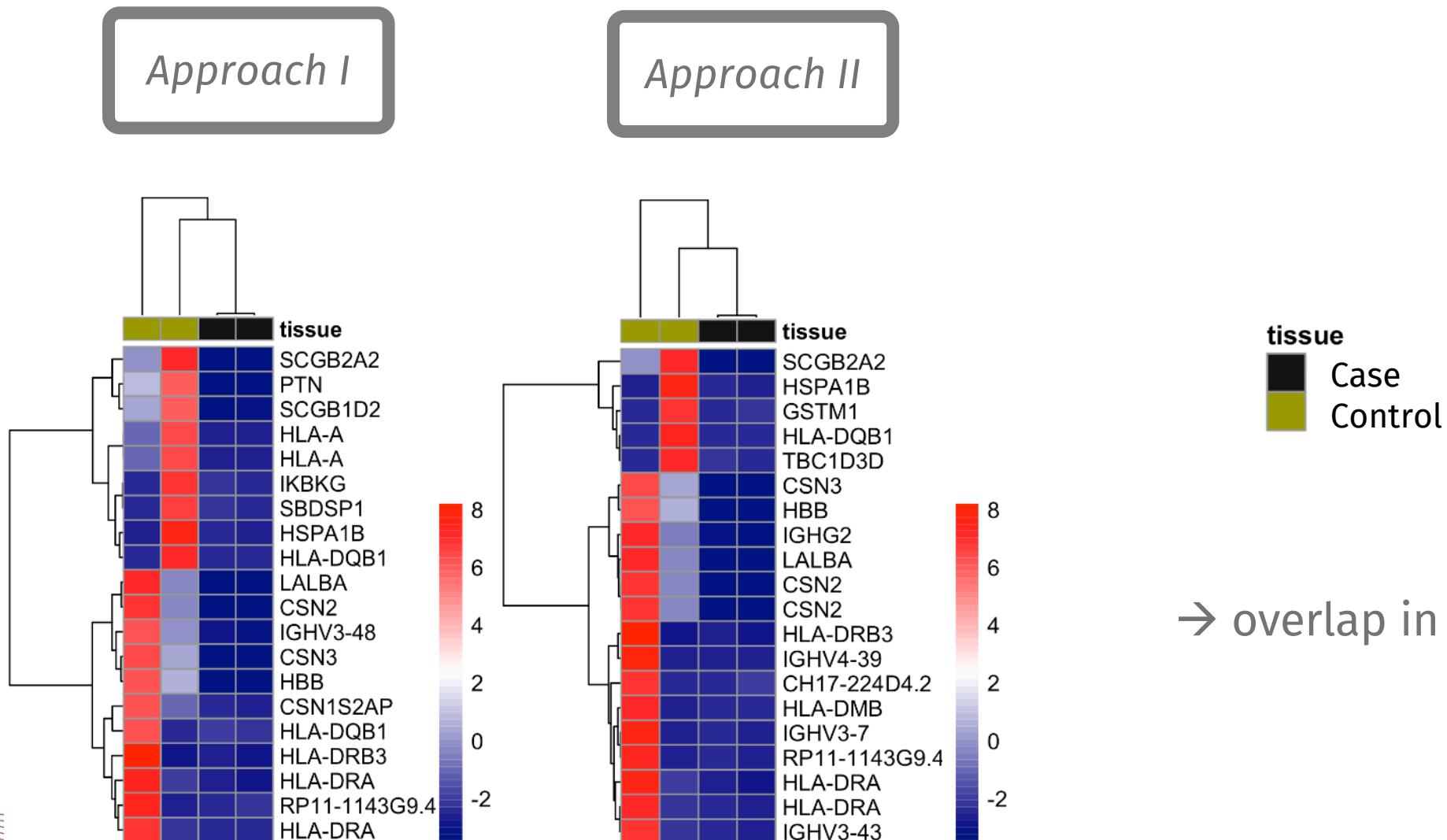


Results: T-test of Naderi genes

⇨ Control ⇨ Case

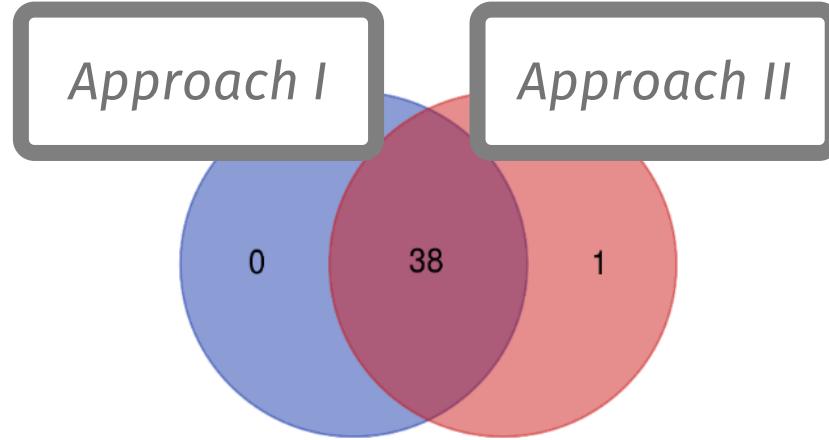


Results: Top20 DE genes

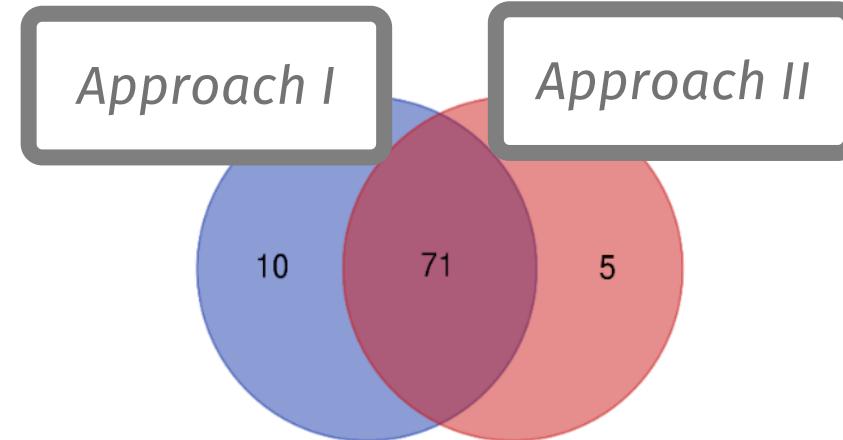


Results: GSEA of DEGs

Hallmark pathways



KEGG pathways



Conclusion

- The choice of the transcriptomic annotation version does not affect differential gene expression estimates nor the prognosis



Acknowledgements to Poznan
Supercomputing and Networking
Center

Thank you!



Magda Mielczarek, PhD
THETA Biostatistics Group
Wroclaw University of Environmental
and Life Sciences
magda.mielczarek@upwr.edu.pl

