Systems Genomics
FS22 – Group 6

Lukas Radtke
Ilya Schneider
Moritz Ullhofen

21. Dec 2022, Basel
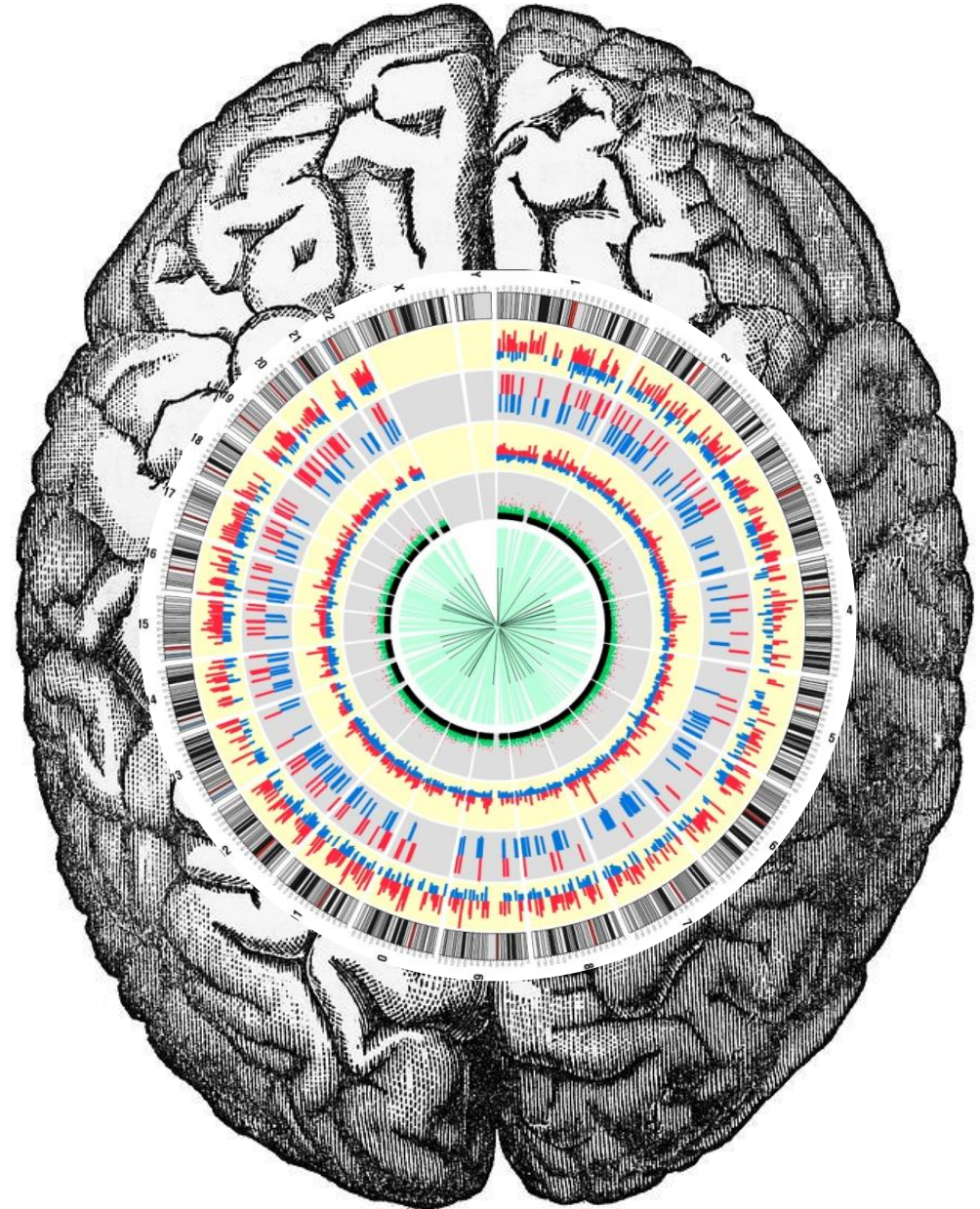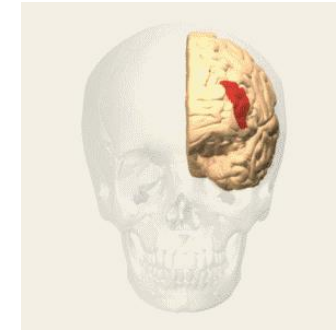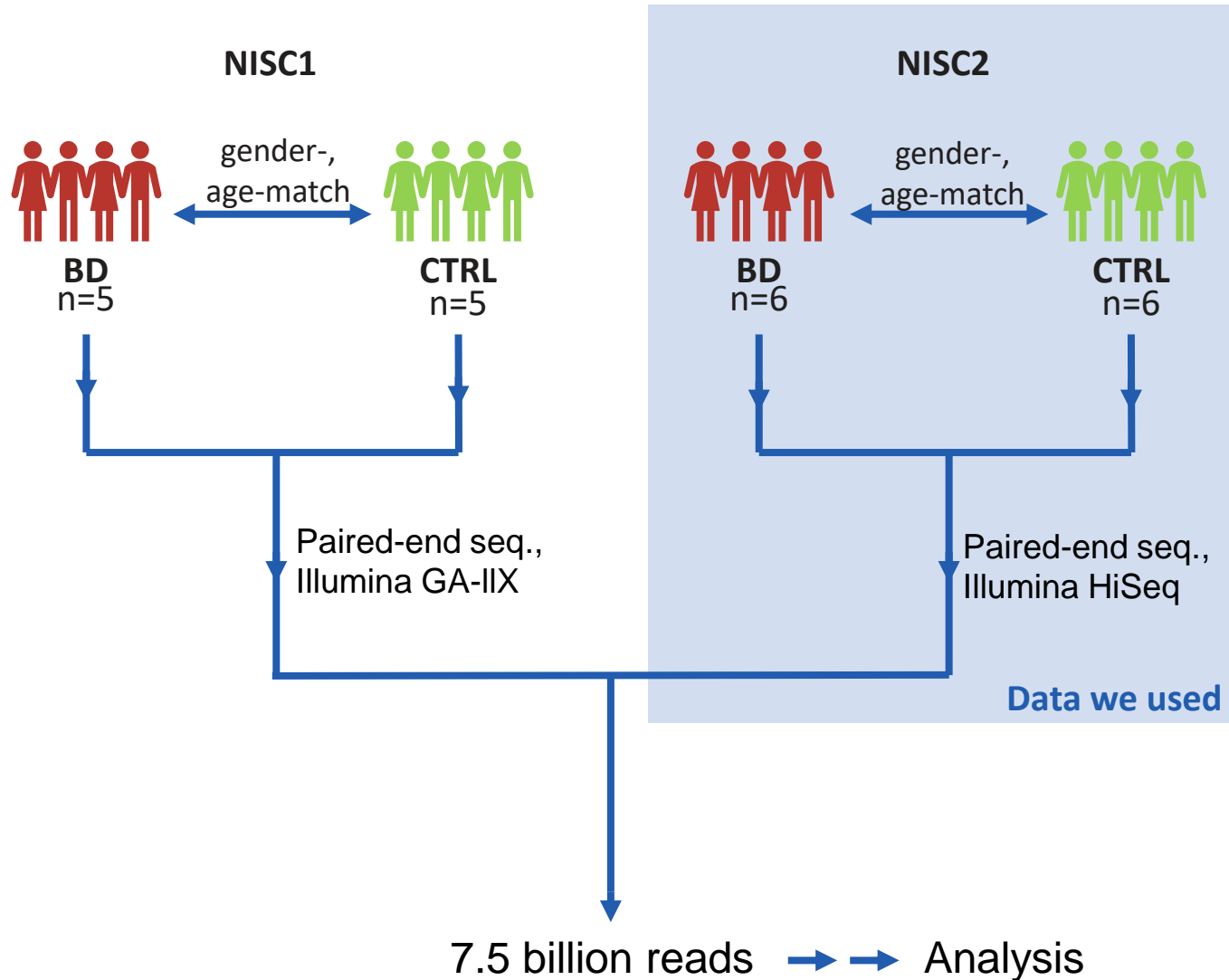
RNA-sequencing of the brain transcriptome implicates dysregulation of neuroplasticity, circadian rhythms and GTPase binding in bipolar disorder.

N Akula, J Barb, X Jiang, JR Wendland, KH Choi, SK Sen, L Hou, DTW Chen, G Laje, K Johnson, BK Lipska, JE Kleinman, H Corrada-Bravo, S Detera-Wadleigh, PJ Munson and FJ McMahon

The Paper – Sample

NISC1

gender-, age-match

BD n=5

CTRL n=5

Paired-end seq., Illumina GA-IIX

NISC2

gender-, age-match

BD n=6

CTRL n=6

Paired-end seq., Illumina HiSeq

Data we used

7.5 billion reads ➜ ➜ Analysis

Dorsolateral Prefrontal Cortex (DLPFC)

- Bipolar Disease vs. Control
- Groups are age- & gender matched
- Two cohorts (NISC1 & NISC2)
- Tissue-samples were taken post-mortem
- Bulk-RNAseq
- Gene = all exons belonging to any isoform
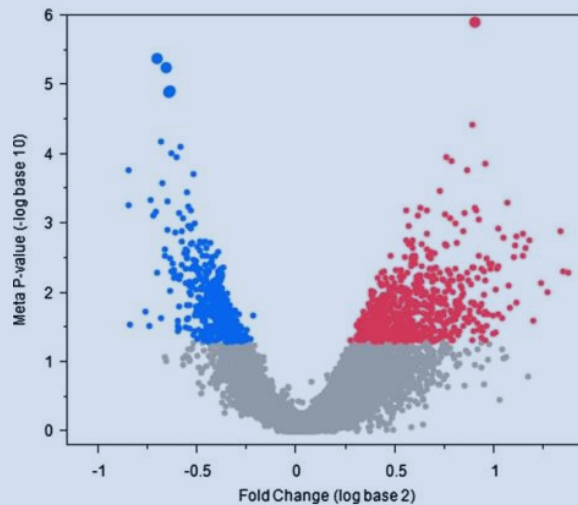- Transcript = all exons belonging to one isoform

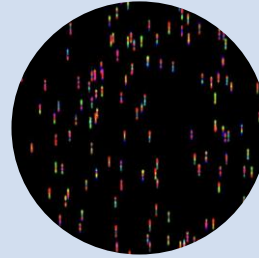# The Paper – Results

Read Depth, Mapping, Basic Stats
- 87% of reads mapped to the human genome build 19 Ensemble transcriptome
- PC1 & PC2 accounted for 67% variability

Meta Analysis:
- 1225 DEGs ($p < 0.05$)
- 5 DE genes ($q < 0.05$):
  - LINC00173 (lncRNA)
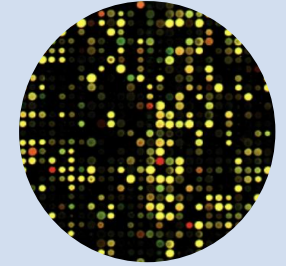 - Prominin 1
  - ABCG2
  - A2M
  - FLI1

Confirmation Methods:

Nanostring nCounter
(41 genes)

qPCR
(15 DEGs, 13 concordant)

Microarray
(7579 genes, r = 0.42 for DEGs)

Gene Ontology:
- transmembrane receptor protein phosphatase activity (> 5 fold)
- enrichment in 'regulation of transmission of nerve impulse (> 2 fold)
- GTPase binding
- regulation of cyclic nucleotide metabolic process
- cell part morphogenesis

GWAS:
- To distinguish cause and consequence
- Found enrichment of SNPs in genes of the 5 ontological categories

**ETH** zürich

The Paper – Discussion and Limitations

Discussed Genes

SRSF5

circadian regulation

RFX4

circadian regulation

OSBPL3

disrupted in sleep deprivation in model organisms

FLI1/ERGB

underexpressed → defect in neural crest development

lncRNAs

Implicated in neurological disease

GANAB

DE in DLPFC in shizophrenia, synaptic disfunction

PROM1/CD133

Underexpressed → reduced plasticity

ABCG2
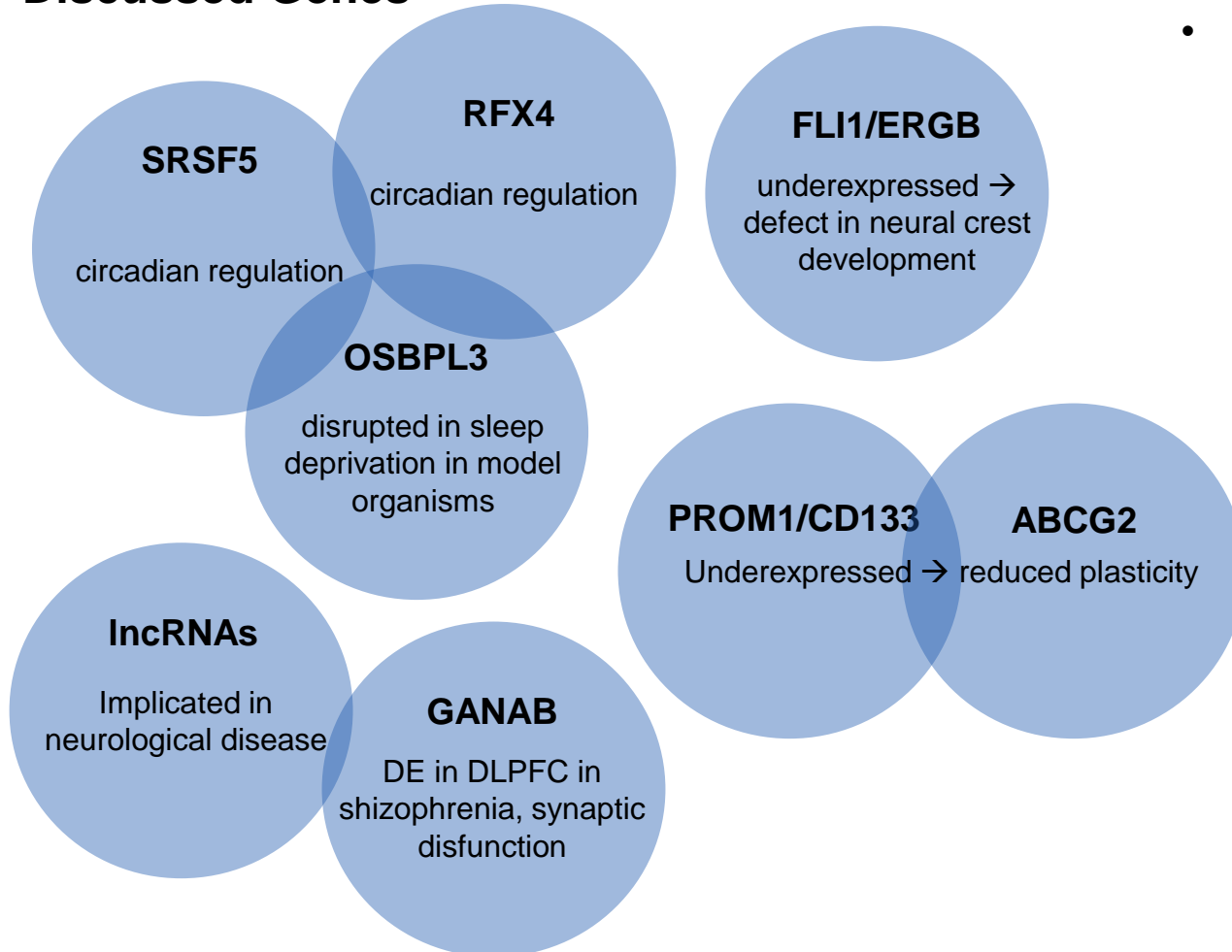
Novelty

- Constituted the first RNAseq profiling of BD brain-transcriptome

Limitations

- No exon level analysis was conducted (differential splicing)

- Subject to agonal and post-mortem variables

- Individual differences

- Medication treatment

- No attempt to differentiate RNA of different cell types in DLPFC

Preprocessing Workflow
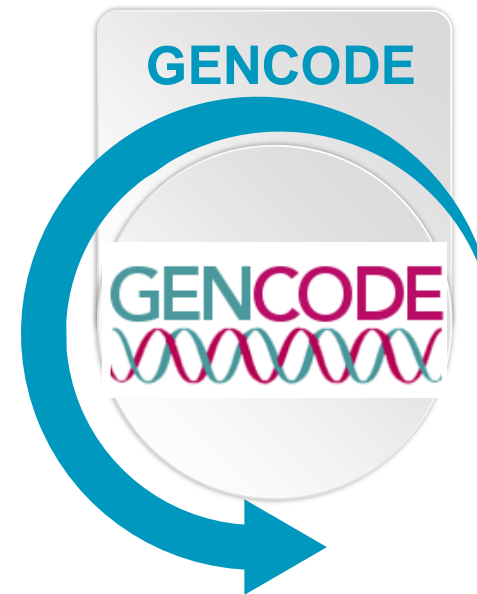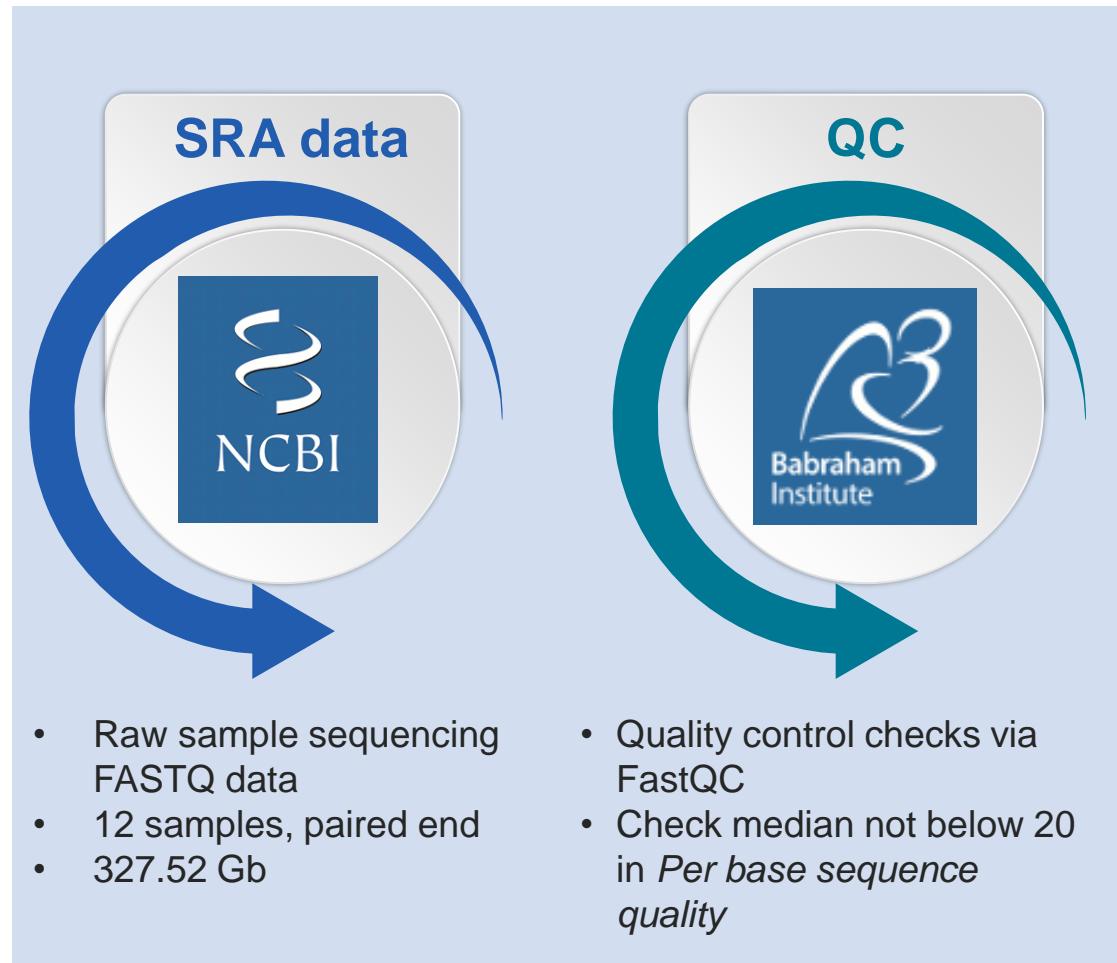
SRA data

- Raw sample sequencing FASTQ data
- 12 samples, paired end
- 327.52 Gb

QC

- Quality control checks via FastQC
- Check median not below 20 in *Per base sequence quality*

GENCODE

- Download newest version human genome
- Release 42, Genome assembly version: GRCh38.p13

STAR

- Read Mapping
- RSEM gene expression quantification

ETH *zürich*
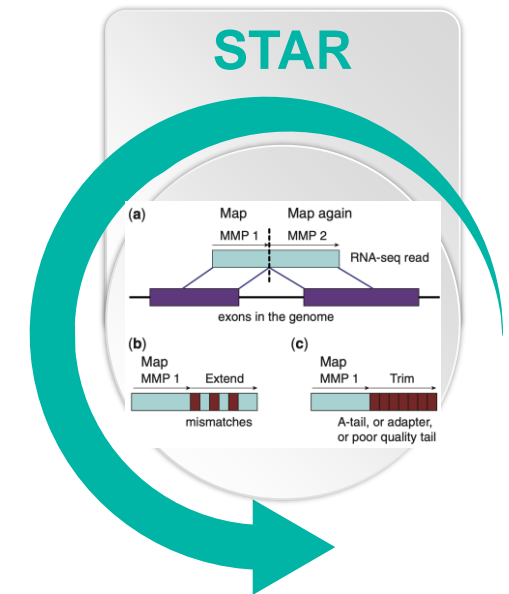
Quality Control
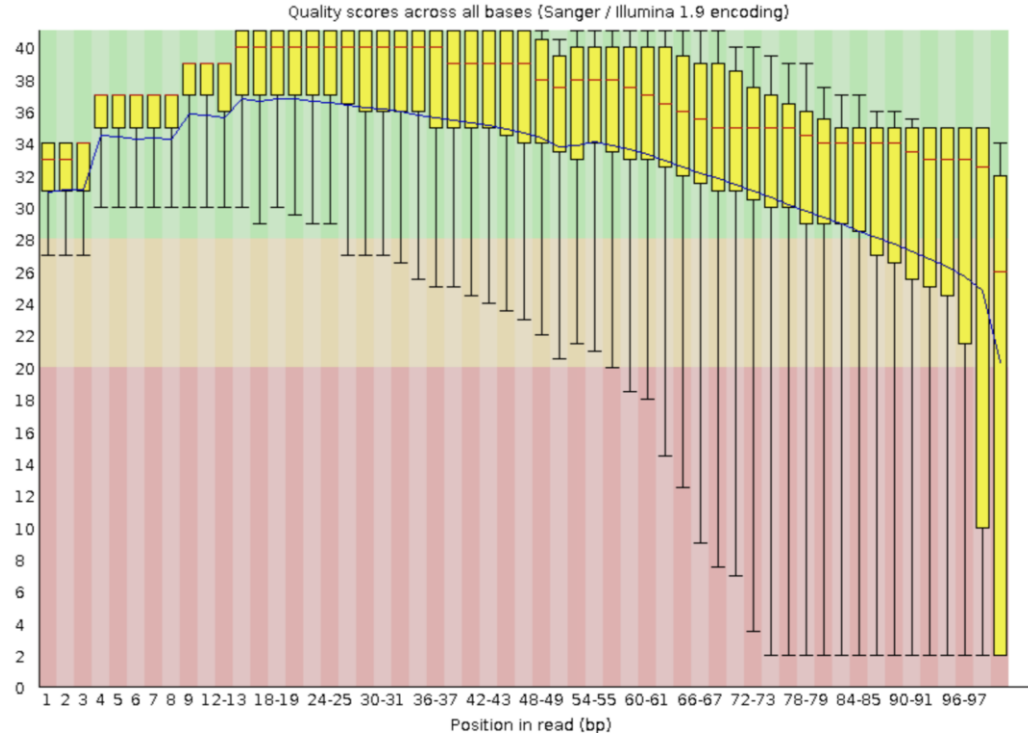
Average QC result (`SRR1047871_2`)
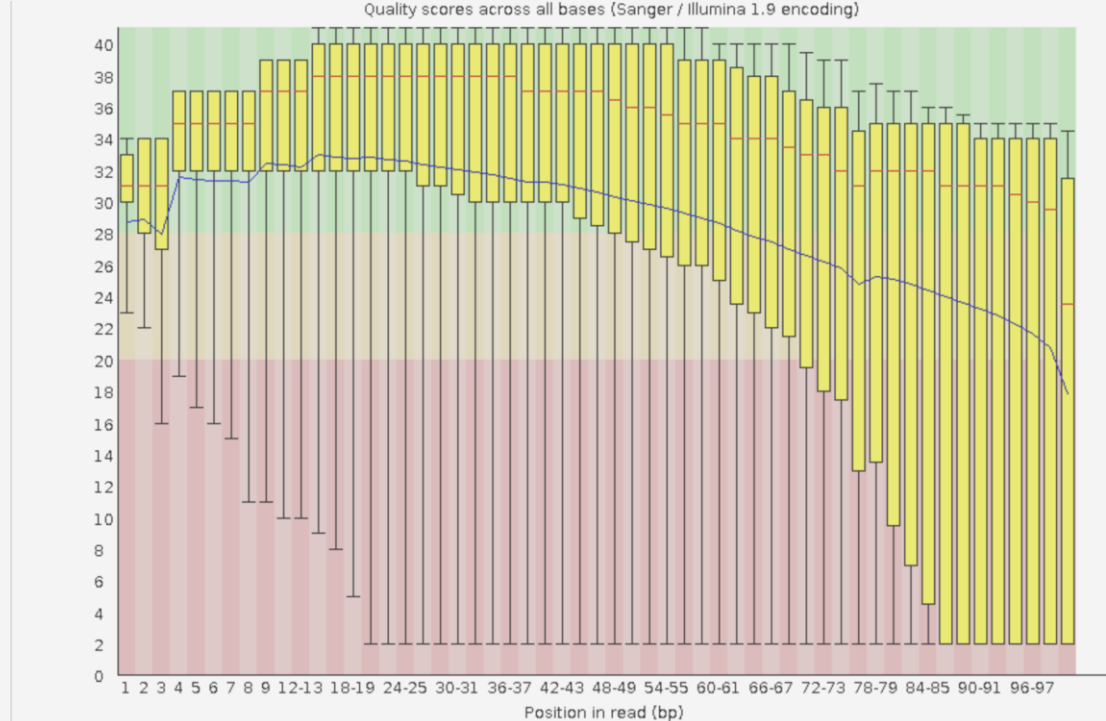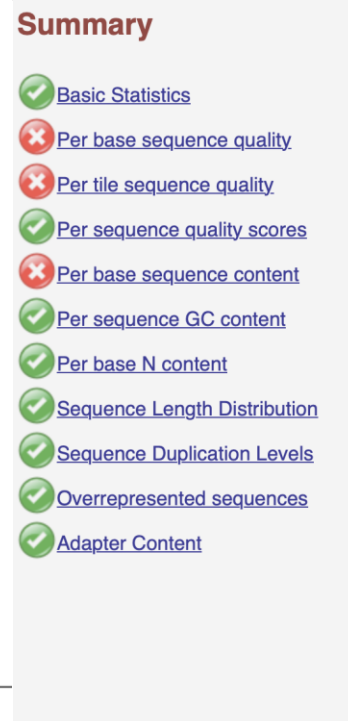
Worst QC result (`SRR1047863_1`)

- Per base sequence quality median locus never below 30, except last 1-2 bases

- Worst QC: last 15 bases low quality, but median above QC 20

- No trimming or deduplication: STAR soft-trimming

Preprocessing Workflow

SRA data

- Raw sample sequencing FASTQ data
- 12 samples, paired end
- 327.52 Gb

QC

- Quality control checks via FastQC
- Check median not below 20 in *Per base sequence quality*

GENCODE

- Download newest version human genome
- Release 42, Genome assembly version: GRCh38.p13

STAR
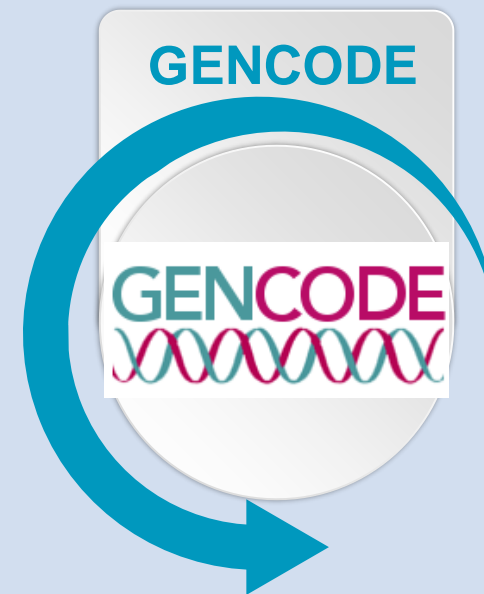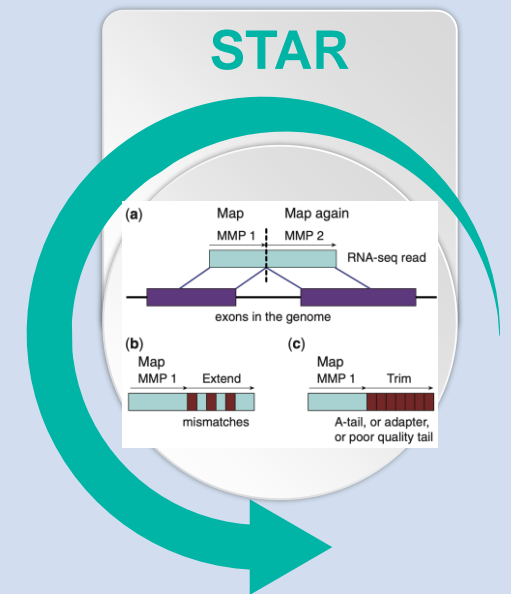
- Read Mapping
- RSEM gene expression quantification

Preprocessing - with its challenges

- **SRA dataset download**:
  - very large dataset 327.52 Gb
  - *even with multiple cores on EULER, up to 7 hours download*

- **Reference human genome**:
  - create genome index (Release 42, Genome assembly version: GRCh38.p13)
  - *problems with creating index, newest release (19.10.2022)*

- **Read mapping and RSEM gene expression quantification**:
  - map transcriptome to human genome
    - Count number of reads aligned to each exon
    - Normalization via TPM (Transcript Per Million reads)
  - Quantification of gene expression per sample (RSEM)
    - RSEM : effective gene length = weighted average of all its isoforms

- *Script for mapping and RSEM took 44 hours with 200 GB RAM on EULER; 4 tries to run completely*

Transcriptomic Profile Across Samples

Gene occurence in patients

Log-log transformed Histogram of average expression level per gene
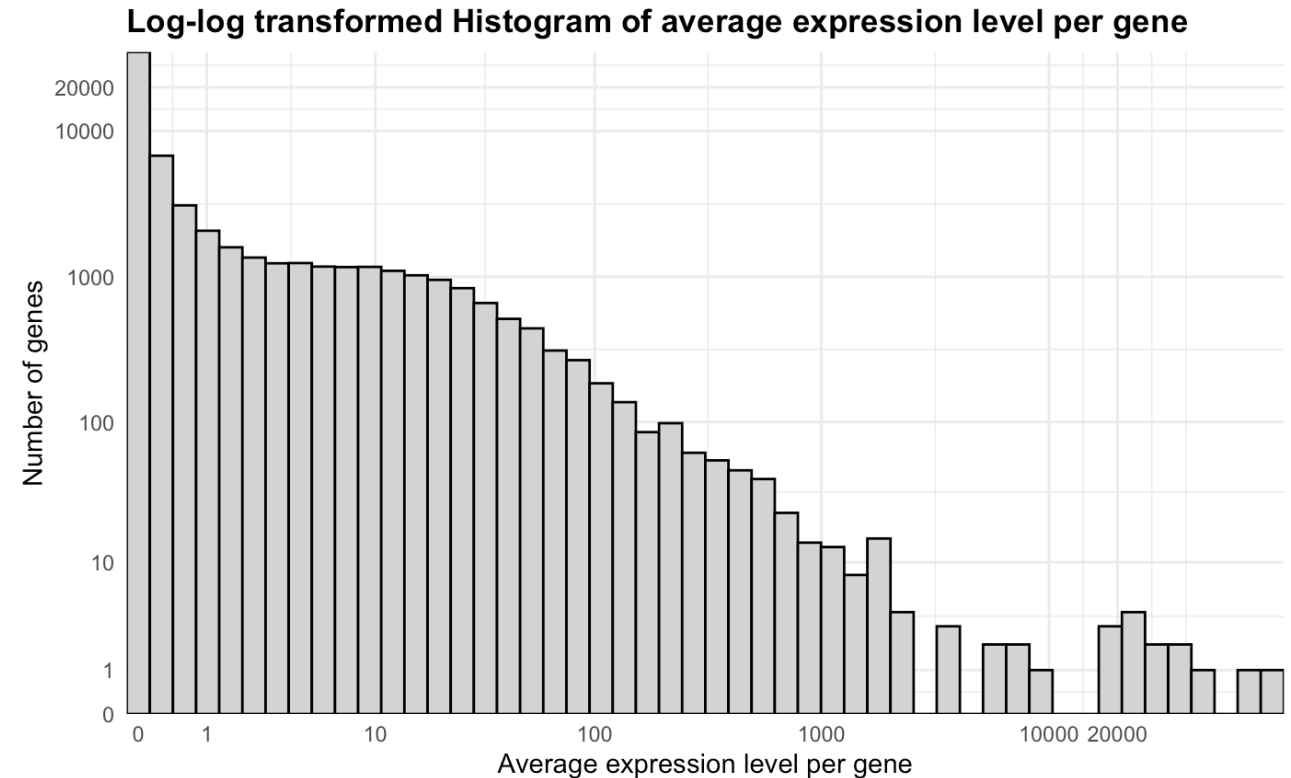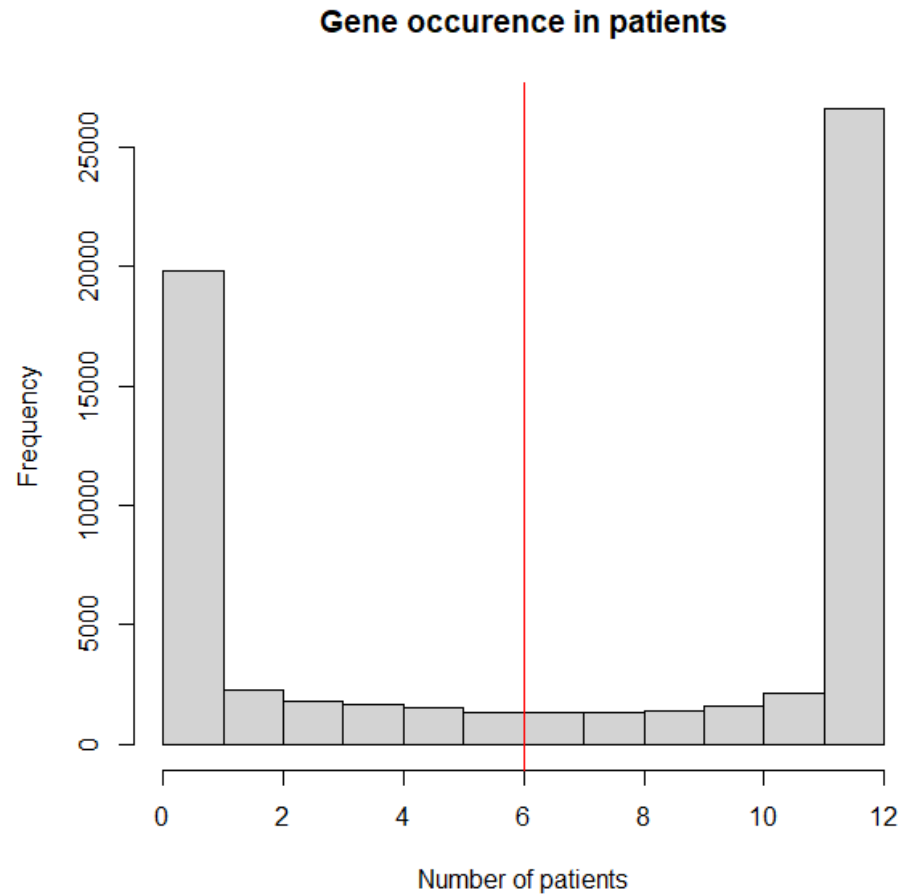
62 696 genes among 12 patients
threshhold: genes in at least half of samples & TPM ≥ 1
35 828 genes selected for further analysis

Patient clustering in terms of gene expression
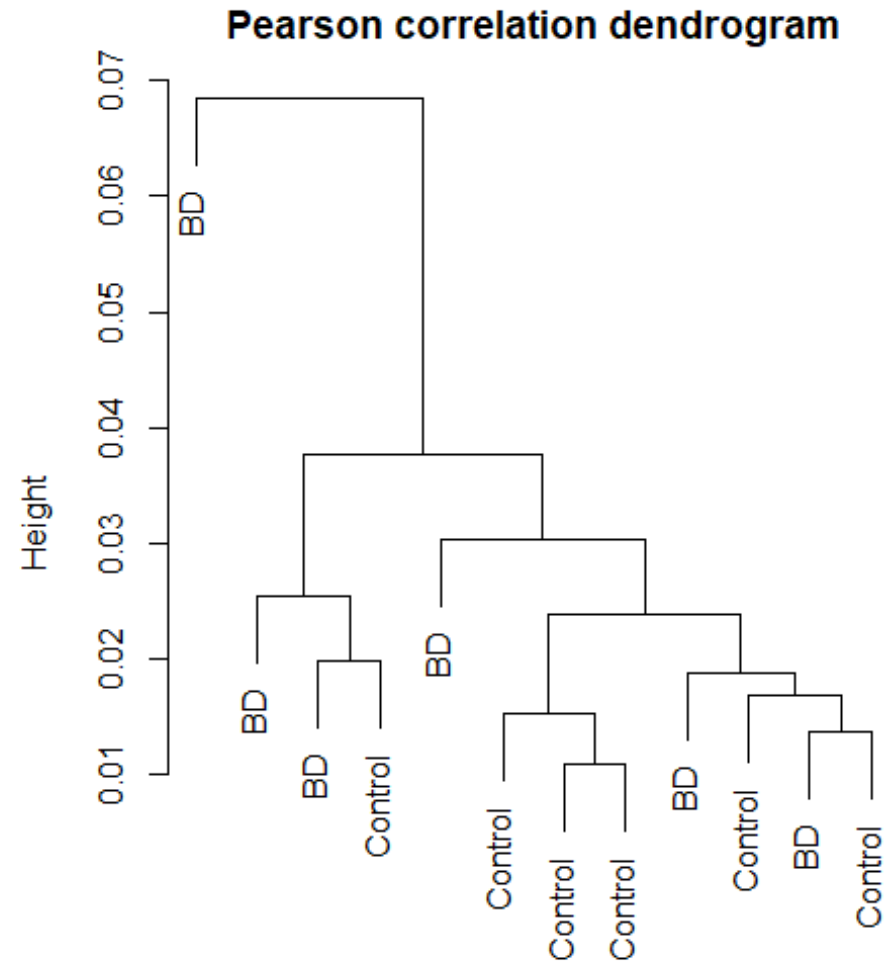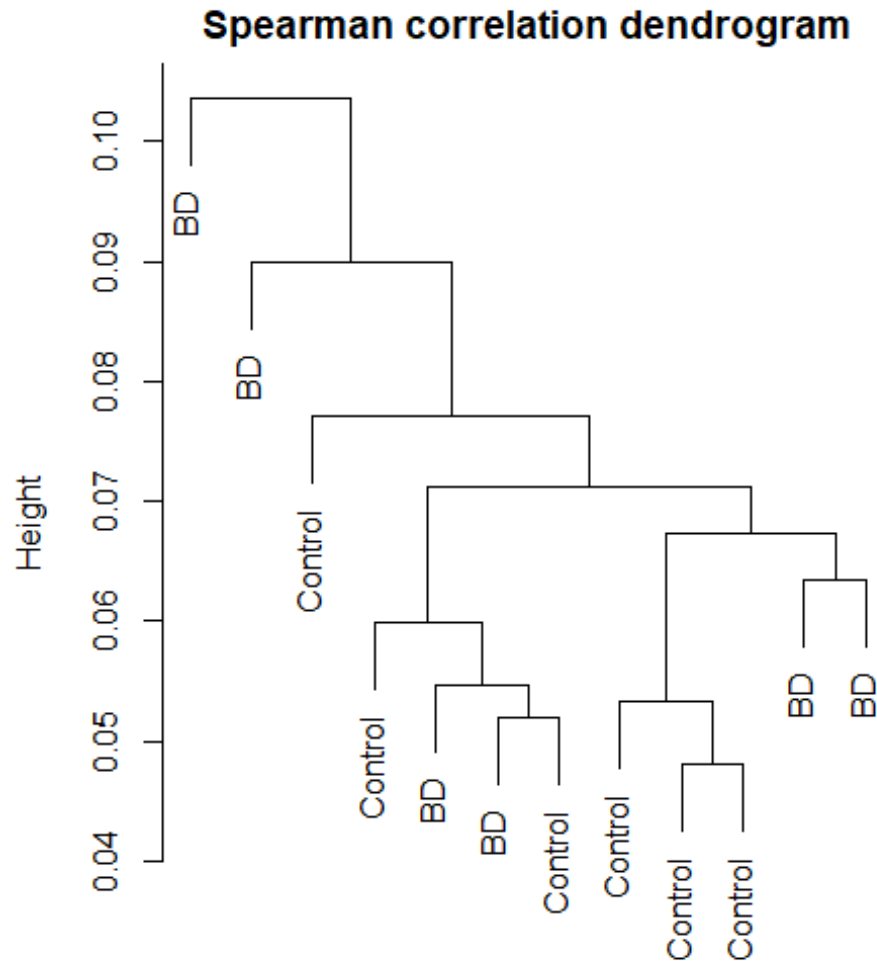
Data Analysis - PCA
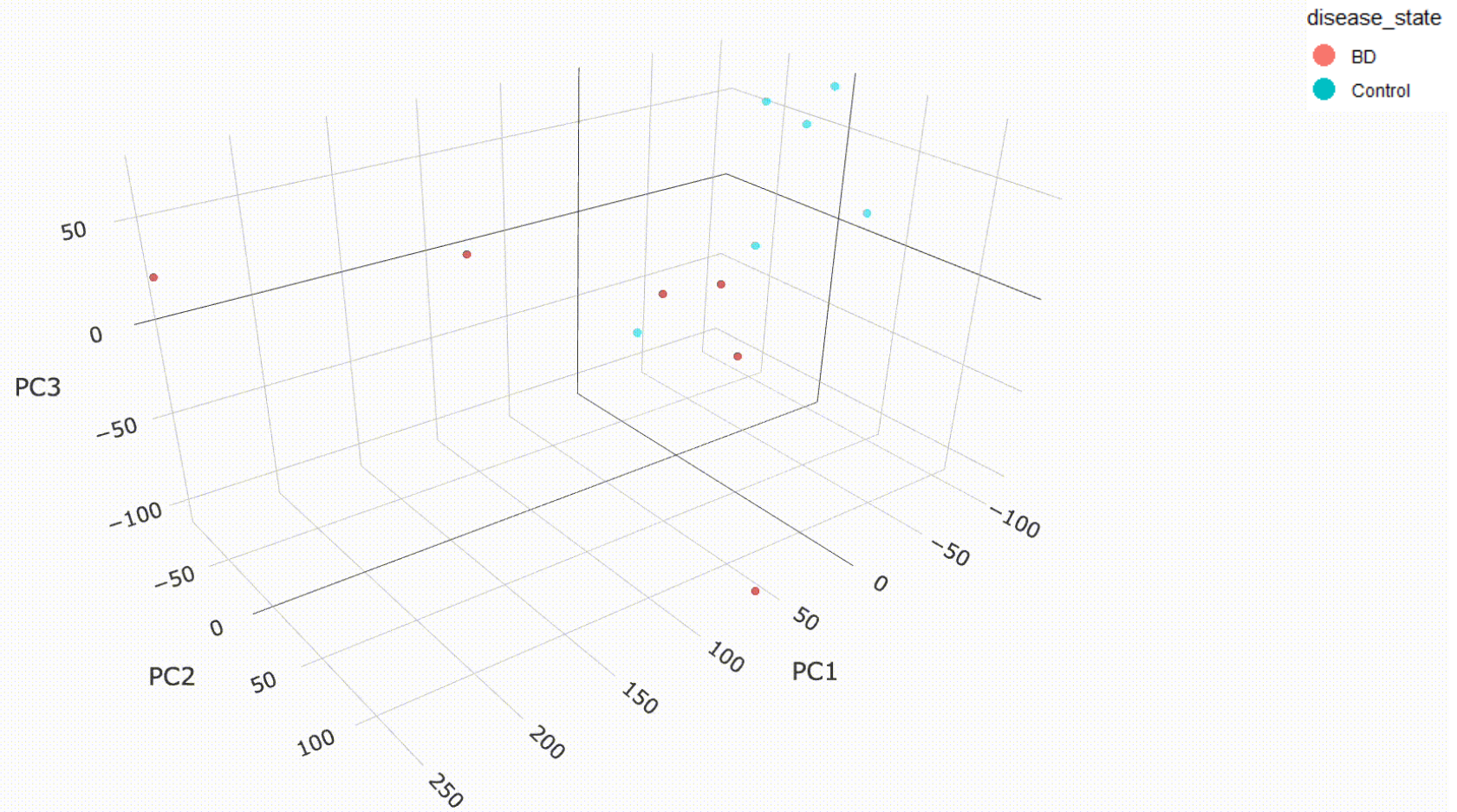
PCA with Most Variance

44.56%

43.15 %
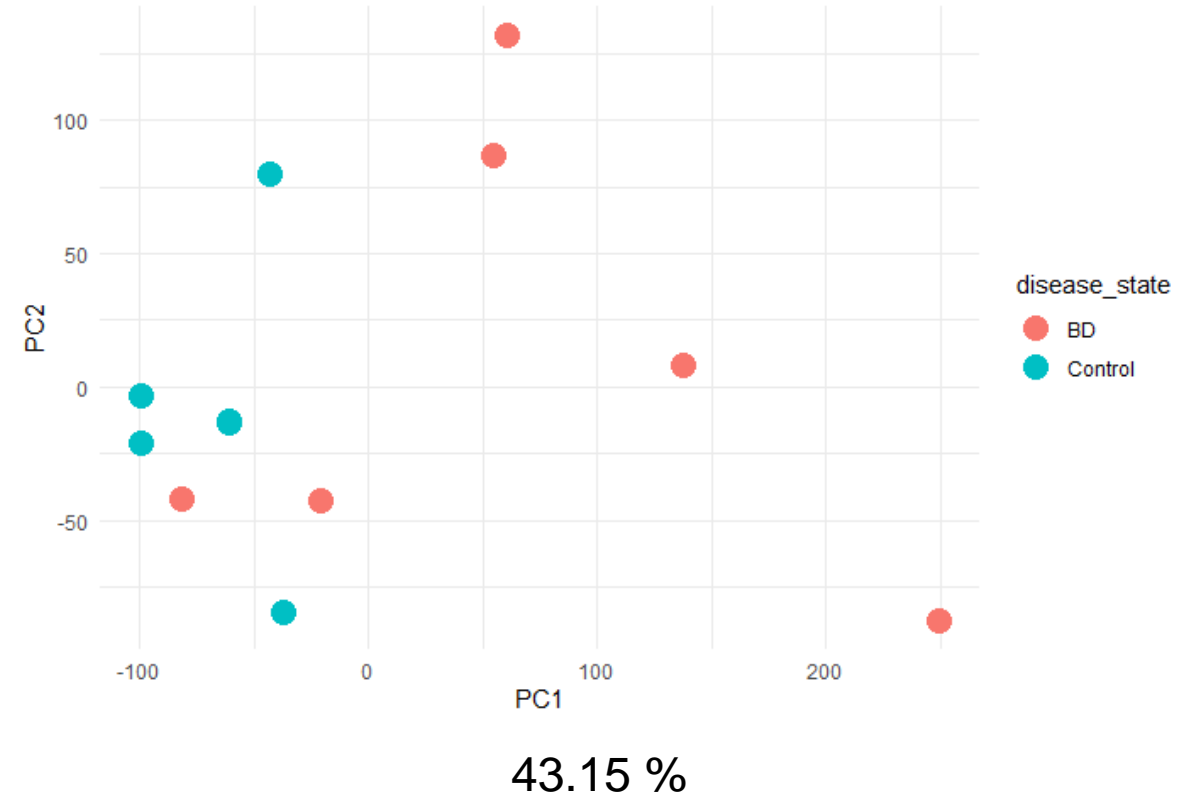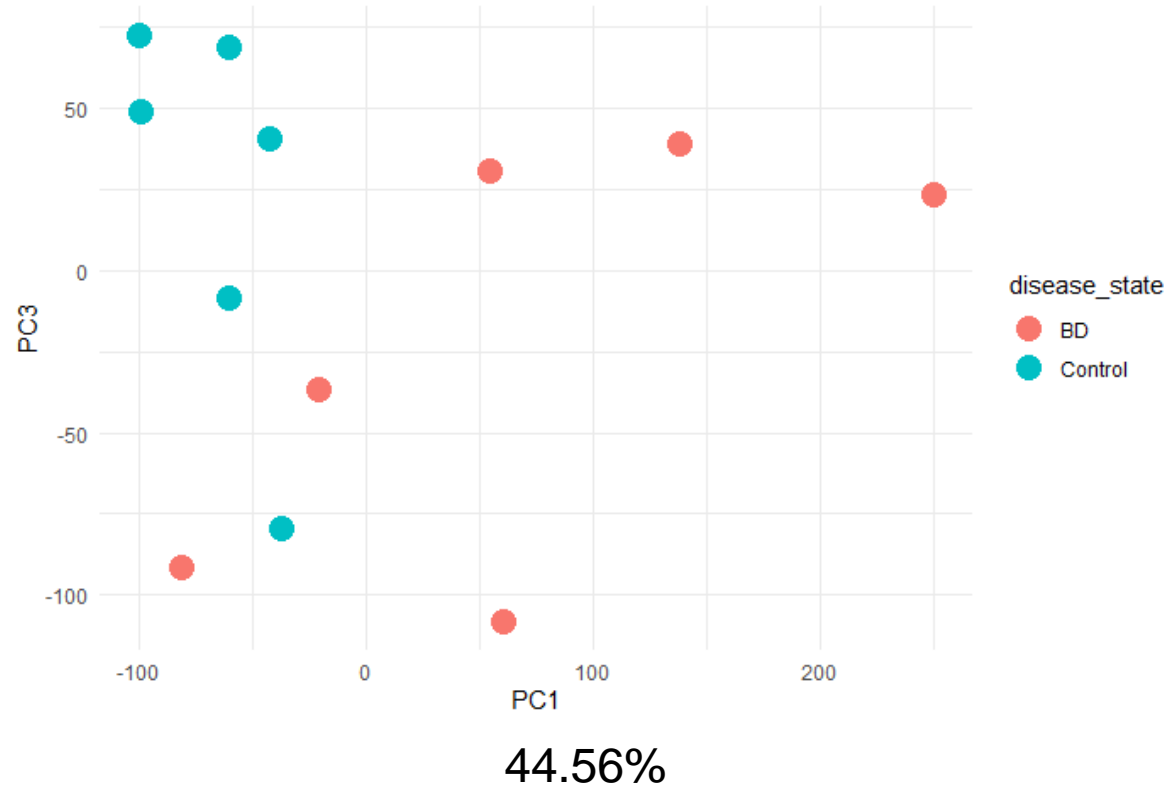
Differential Expressed Analysis

- ANOVA: 0 differentially expressed genes (DEGs)

- DeSeq2: **549 DEGs**
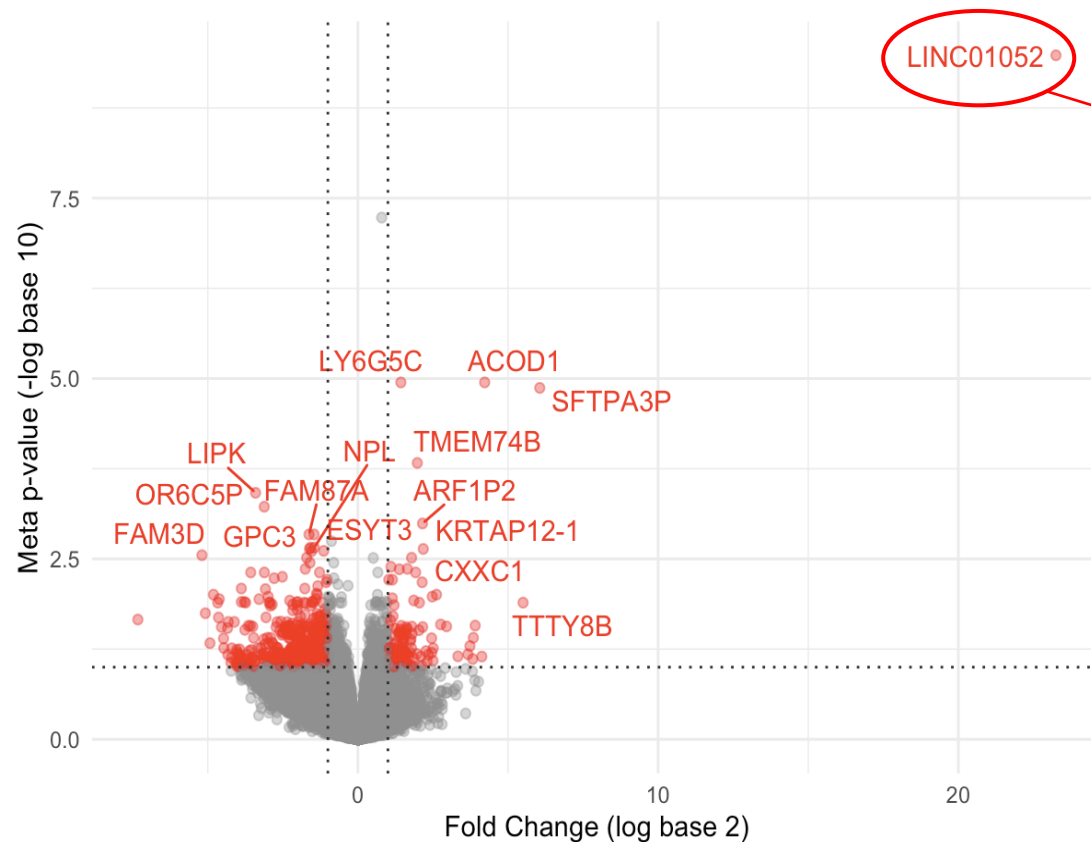
lncRNA associated with schizophrenia

Akula *et al*., 2014 emphasize importance of upregulated lncRNA in BD patients

Grouping of DEGs
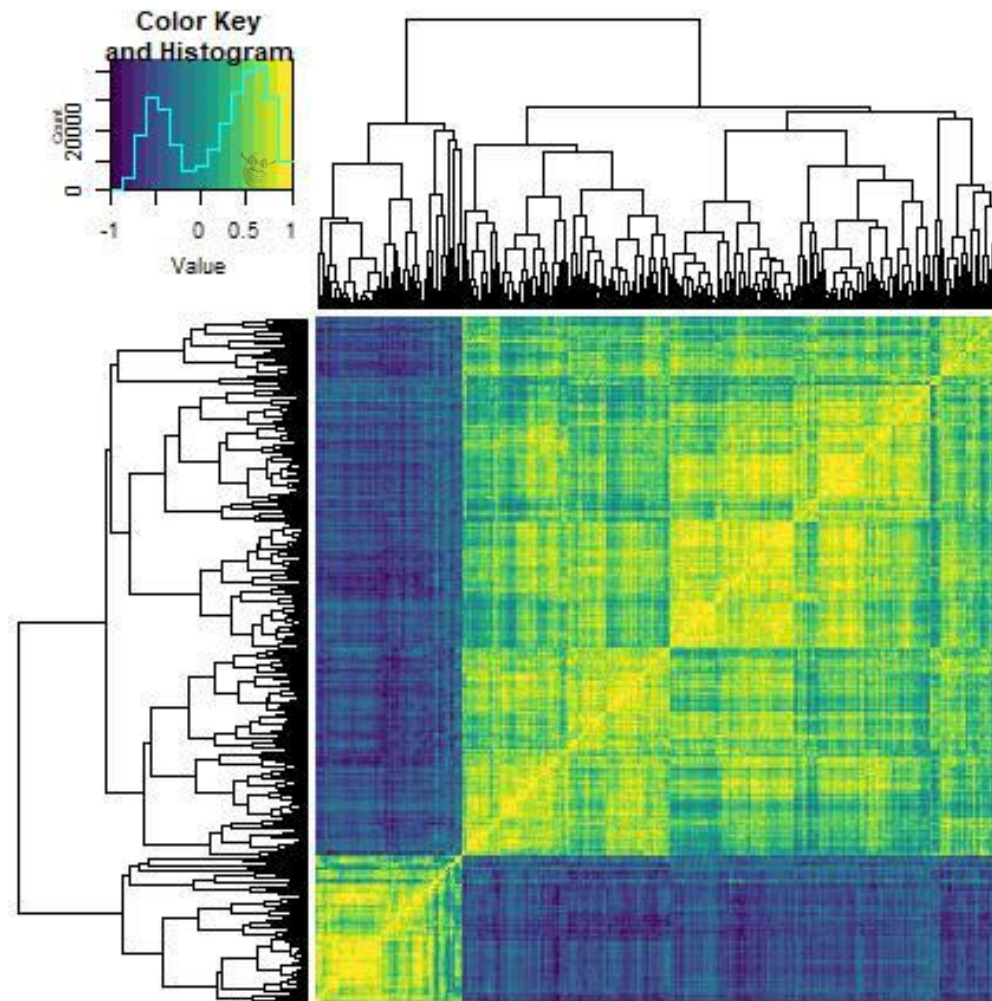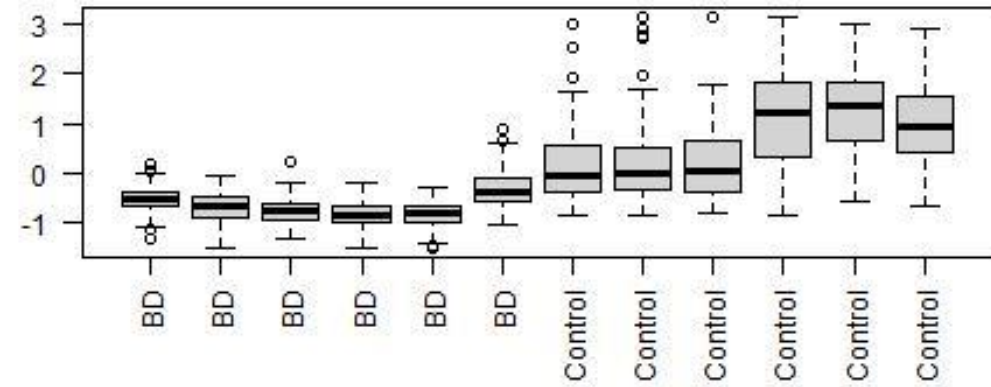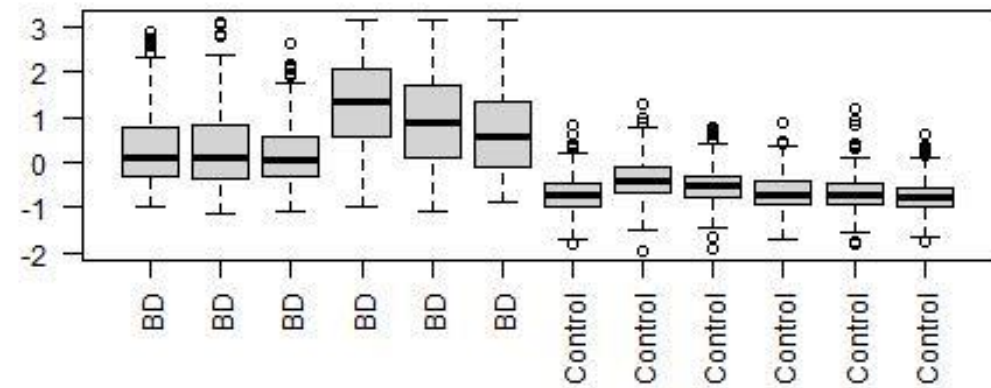
Identification of the Co-expression Modules

- Goal: identify smaller gene clusters, without specifying their number

- Weighted Gene Co-expression Network Analysis (WGCNA)
  - 499 out of 549 genes were assigned to a module
  - A total of 4 modules were identified

Cluster Dendrogram
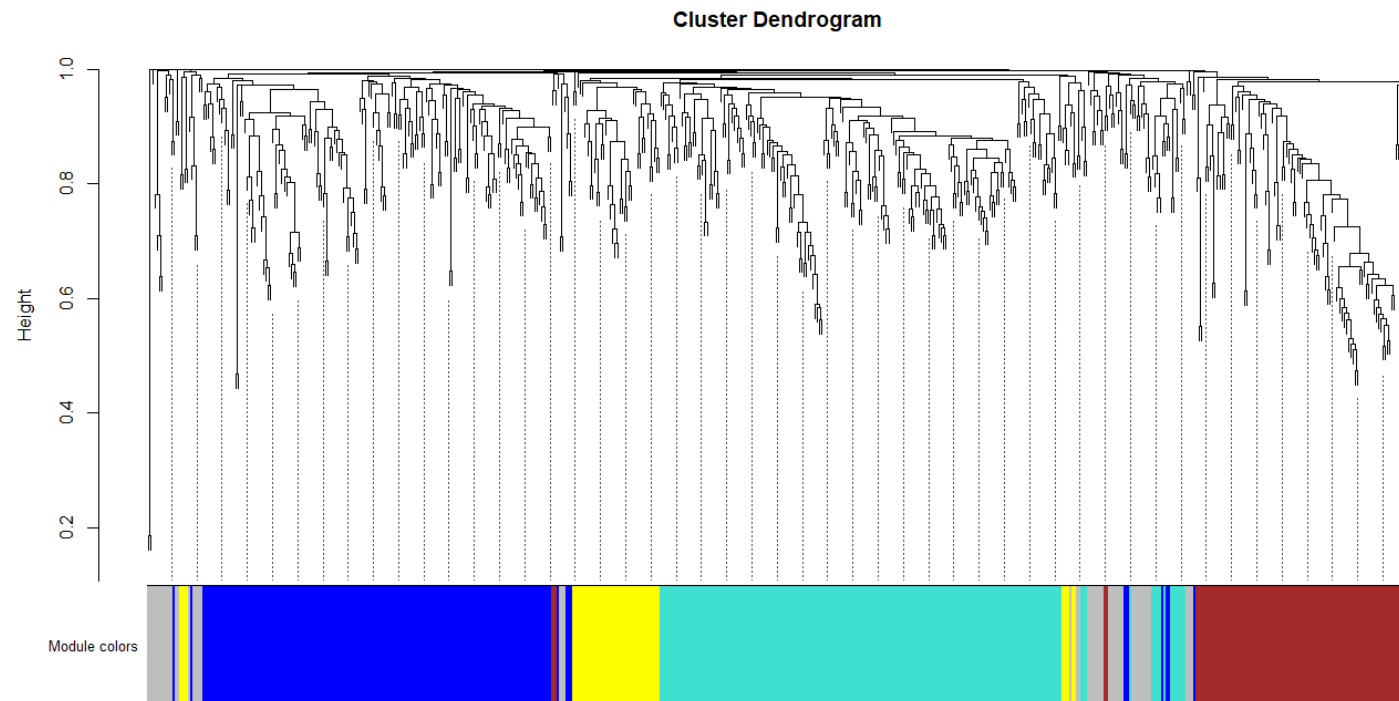
Modules display specific expression patterns for BD patients

Gene Enrichment Analysis – Frequency Comparison DAVID

Gene Enrichment Analysis – Frequency Comparison DAVID

Membrane Protein
GPCR

Membrane Protein
Glycosylation

Modules:

1 3

2 4

Membrane Protein
Cardiac / Developmental
RedOx / Metallo Protein

Cell Apendages / Structures
Ion Transport
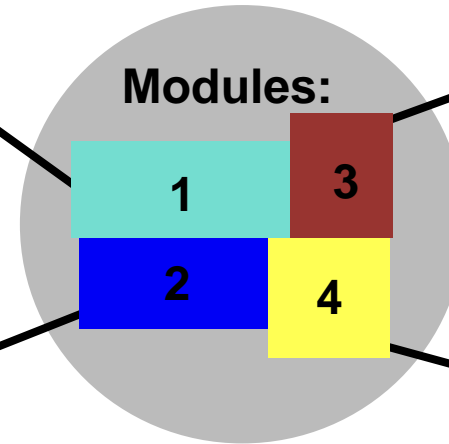
Gene Enrichment Analysis – Rank Distribution Comparison GSEA

- **Gene Set Enrichment Analysis (GSEA)**: rank-based statistical tests for enrichment analysis

- Advantages vs DAVID:
    - No arbitrary threshold for DE, gene list of interest
    - Genes ordered based on: $s_g = sign(logFC_g) \times (-log_{10}P_g)$

- Estimated significance is created with Kolmogorov-Smirnov test

- Compare normalized ES genes to curated gene subsets of Molecular Signatures Database (MSigDB)

- We chose C5: gene subsets ordered to same **Gene Ontology (GO)** term

- **Results**: GO terms that are most enriched (most genes per GO gene set)

Gene Enrichment Analysis – GSEA (GO terms)

|	pathway	padj	NES	size
1	GOCC_VESICLE_MEMBRANE	0.00	3.54	22
2	GOBP_HOMEOSTATIC_PROCESS	0.00	2.25	25
3	GOBP_VESICLE_MEDIATED_TRANSPORT	0.01	1.90	19
4	GOBP_CHEMICAL_HOMEOSTASIS	0.03	1.76	20
5	GOBP_ION_TRANSPORT	0.08	1.60	26
6	GOBP_REGULATION_OF_CELLULAR_COMPONENT_MOVEMENT	0.08	1.56	16
7	GOBP_REGULATION_OF_CELL_DEATH	0.08	1.53	20
8	GOMF_ION_TRANSMEMBRANE_TRANSPORTER_ACTIVITY	0.24	1.23	15
9	GOBP_POSITIVE_REGULATION_OF_CELLULAR_BIOSYNTHETIC_PROCESS	0.32	1.13	15
10	GOBP_PROGRAMMED_CELL_DEATH	0.38	1.10	19
11	GOBP_NERVOUS_SYSTEM_PROCESS	0.54	0.96	24
12	GOBP_ION_TRANSMEMBRANE_TRANSPORT	0.60	0.92	20
13	GOBP_CATION_TRANSPORT	1.00	-0.37	17

Membrane Protein/Transport

Vesicle

Final Remarks

- Found 549 DE genes, 3 models with upregulated, 1 module with downregulated genes in BD
- Our pipeline allows to make comparable conclusions with Akula *et al.*, 2014
- Multiple GO terms from Akula *et al.*, 2014 were identified
  - Membrane components
  - Ion and membrane transport
  - Cell death and homeostasis
- However not all GO categories Akula *et al.*, 2014 from were confirmed

- Additionally, our analysis possibly links new aspects to bipolar disorder
  - Glycosylation
  - Cardiac, developmental and RedOx processes

ETH *zürich*
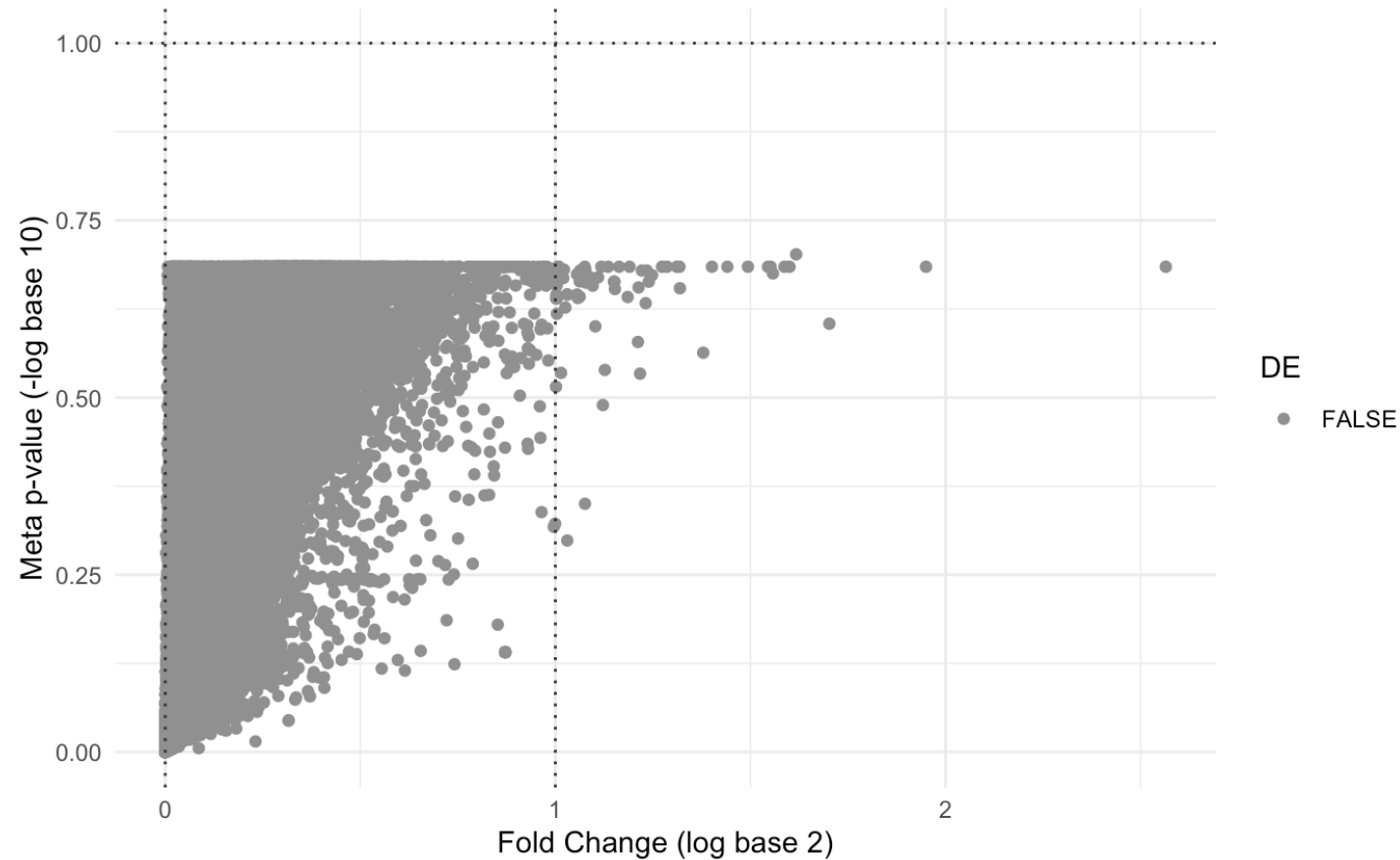
Additional: ANOVA

ANOVA results of DE analysis: fold change > 2, padj < 0.1
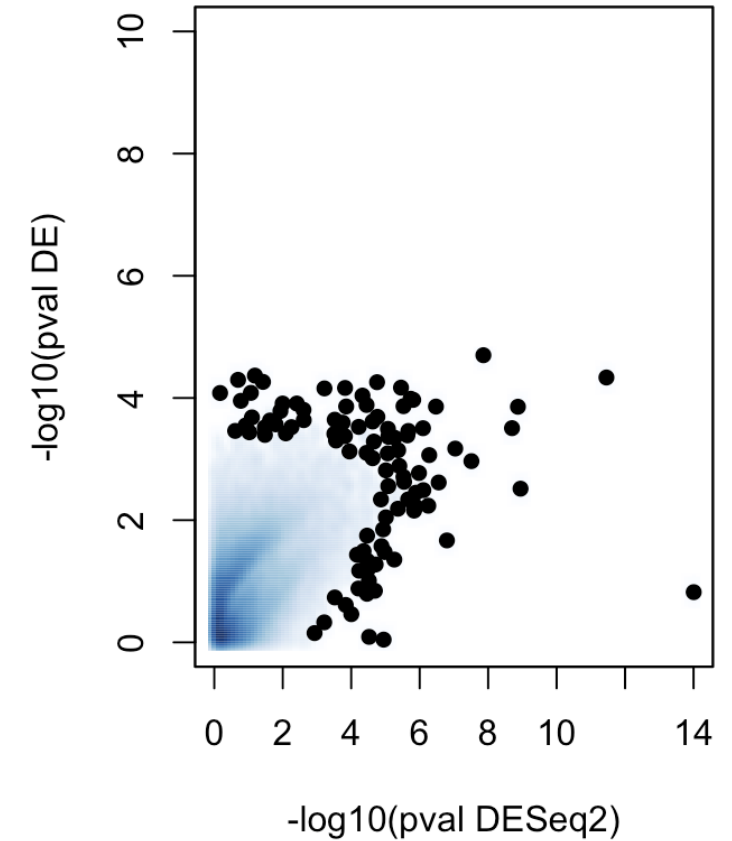
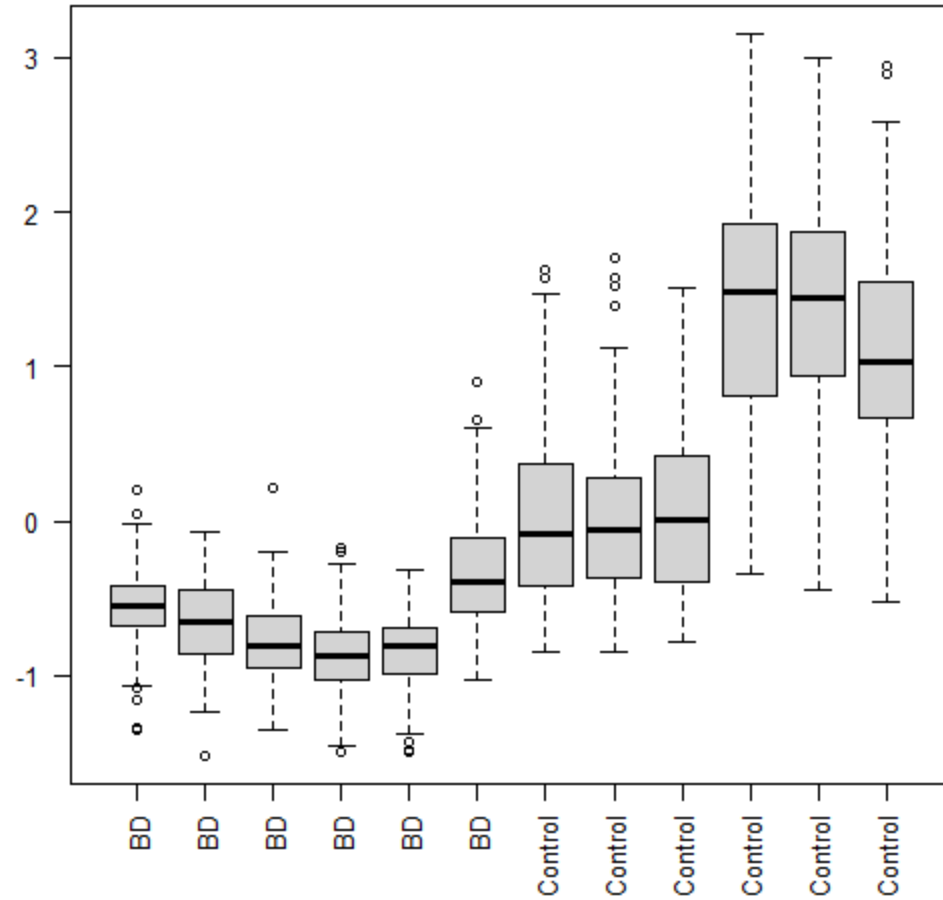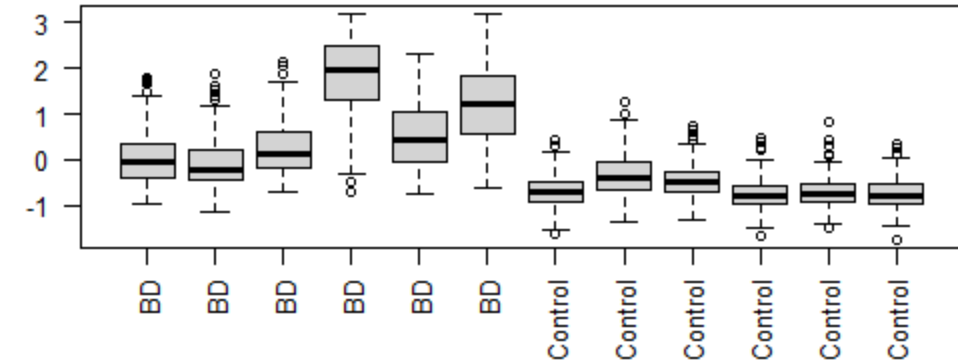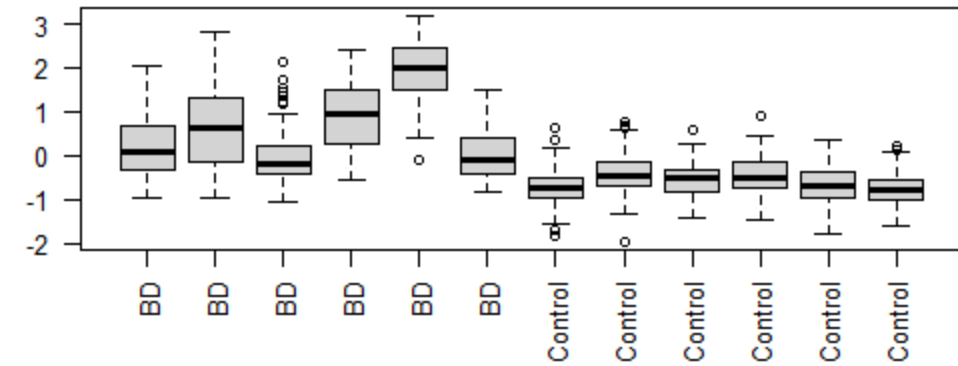Comparison DESeq2 and ANOVA

Additional: WGCNA on Separate Clusters

Module 1 (101)
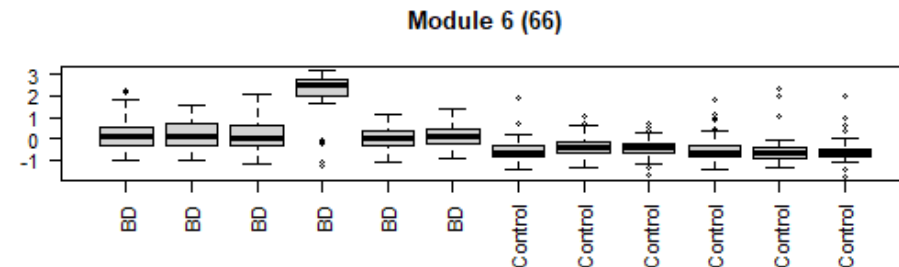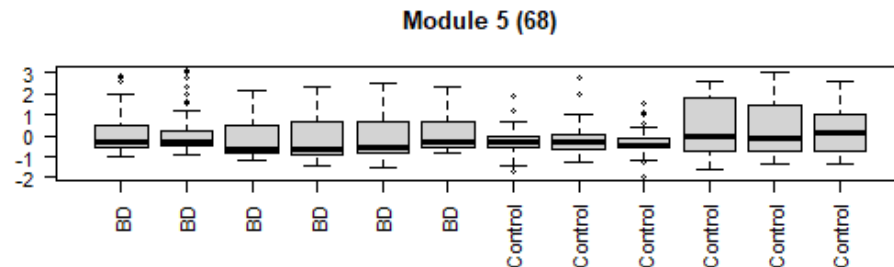
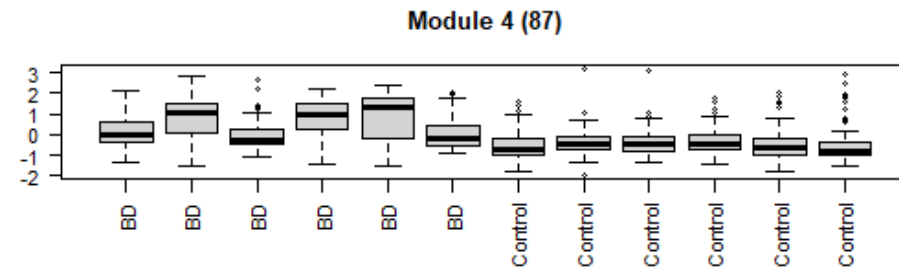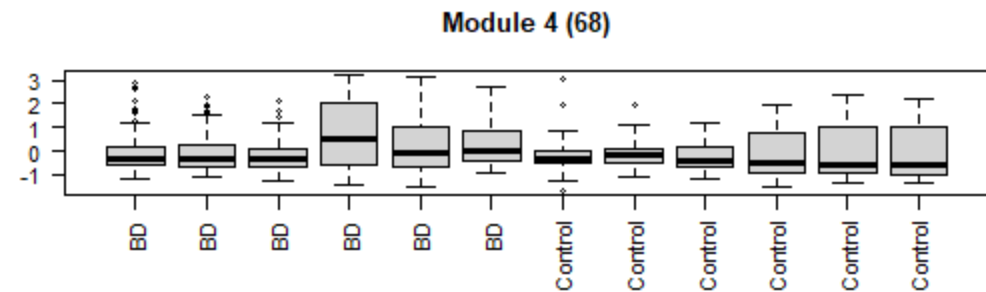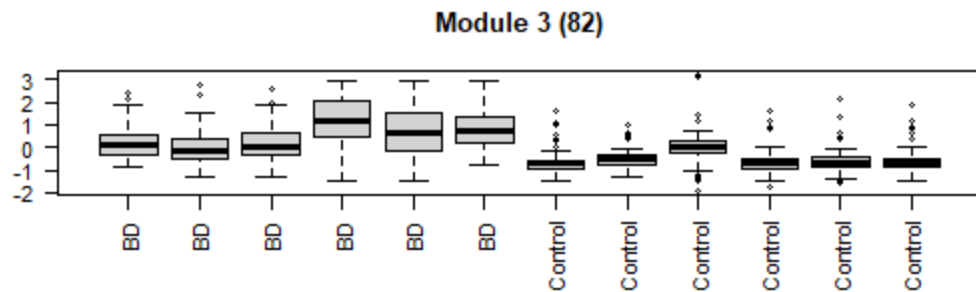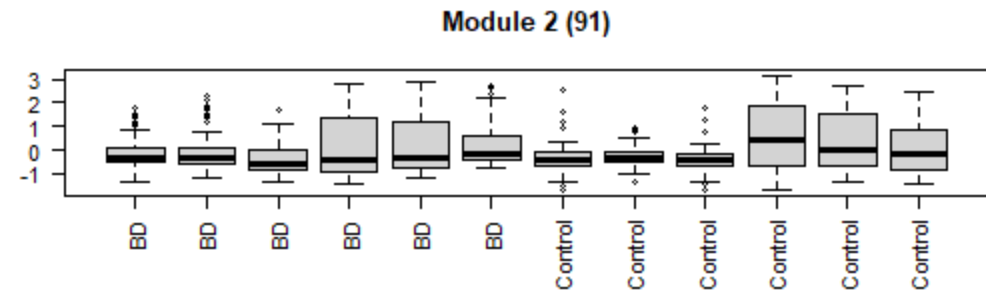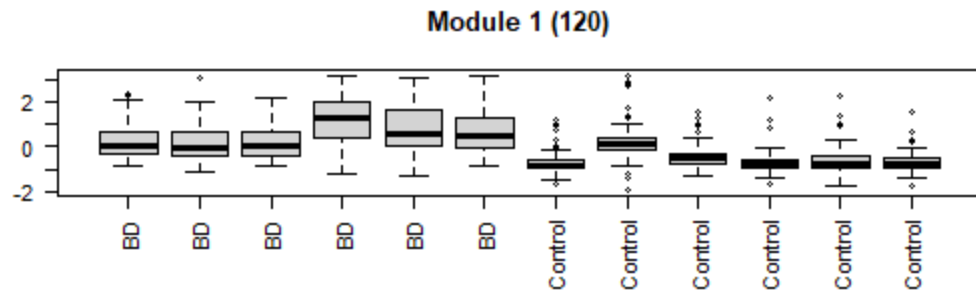Module 1 (222)

Module 2 (146)

Additional: WGCNA on Separate Patient Groups (BD)
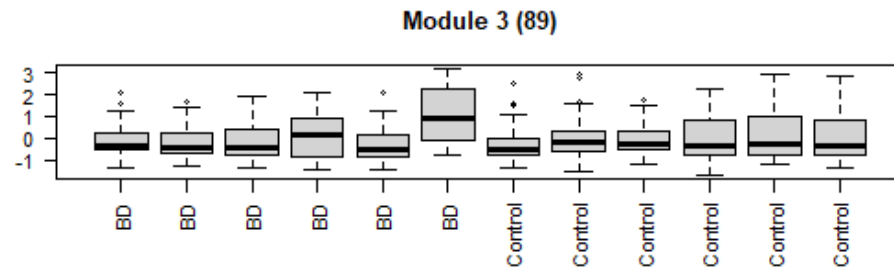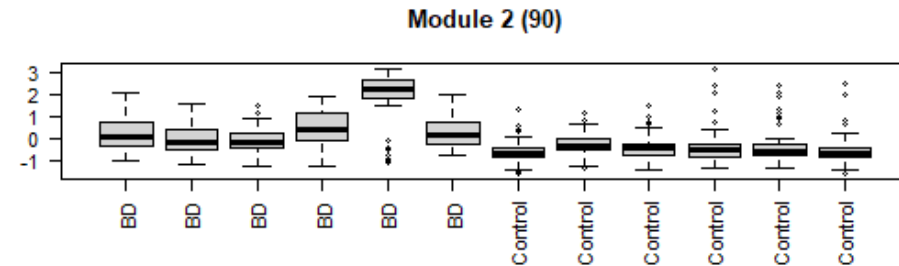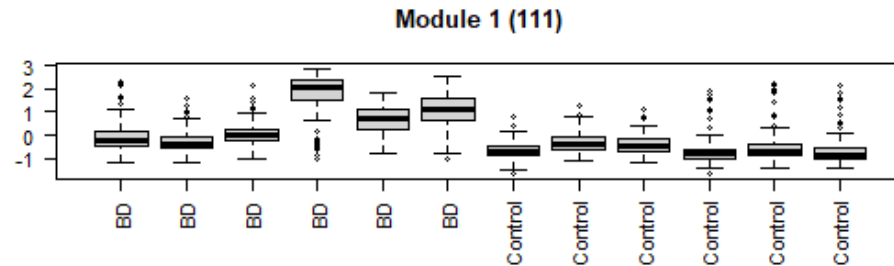
Additional: WGCNA on Separate Patient Groups (Control)

Additional: GSEA plots

ETH *zürich*