1 – Data exploration: understand transcriptomic profiles.

* 1. – Differential expression analysis

QC check for biological replicates.

How many genes are differentially expressed? What is the distribution of their expression (many genes with large change / just a few)? Do they change in the same way under different conditions (treatments/genotype)? Are there clusters of co-regulated genes?

R: edgeR, complexHeatmap, ggplot2

Expected results:

* PCA plot of sample replicates
* Volcano plot to check expression distribution
* Table with differential expression (logFC and FDR)
* Heatmap to compare differential expression profiles between different conditions
* Gene co-regulation dendrogram\*
  1. – Functional characterization of differentially expressed genes

Which biological processes are affected (GO enrichment, GSEA, KEGG)?

R: clusterprofiler, biomaRt, org.db, pathview, WeightedCluster

Expected results:

* GO enrichment and GSEA table
* GO enrichment and GSEA plots (with representative GOs/most relevant)
* KEGG enrichment plots
  1. – Direct target gene identification

QC check: peak location and distribution in gene features (promoter, utr, cds)

Which transcription factors bind to the “promoter” (we will use 3kb up + 1kn downstream of TSS) of DE genes?

R: Granges, Gviz, rtracklayer, EnrichedHeatmap

Expected results:

* Peak location and distribution plots
* Add ChIPseq/Dapseq data to DE tables
* Heatmap annotation of binding TFs

2 – Study of BRC1/HBs network