**Paper Laurent & Tina:**

**Step 1:** Data import and pre-processing

Gene-expression data: GEO:GSE42069, Caco2, SAE and THP1 cells induced with TiO2-NBs for 24h. Two concentrations; 10 ug/ml and 100 ug/ml.

Pathway databases: KEGG v7.0, Reactome v7.0 and Wikipathways-20190910

GO-terms: Apoptopic process (GO0006915), Inflammatory response (GO0006954), Cellular response to DNA damage stimulus (GO0006974), Cellular response to oxidative stress (GO0034599)

GO-terms chosen based on literature that shows ROS formation, apoptopic cell death, inflammation and DNA damage as a reaction to nanoparticles: <https://doi.org/10.1002/jat.3817>, <https://doi.org/10.1186/s40169-018-0212-7> a lot of literature can be found about this topic

Processed this data in 3 different R-scripts.

**Step 2:** Pathway selection

Selection of pathways from the three pathway databases. Selection is based on genes in GO-terms. Enricher from clusterProfiler package used to find pathways enriched with genes from GO-term

Enricher settings: Adjusted p-value cutoff: < 0.05, q-value cutoff < 0.05. Other settings were kept standard.

**Step 3:** GSEA analysis

Selected pathways will be used as the genesets for the GSEA analysis. GSEA function of clusterProfiler package was used. Gene expression data will be used for the enrichment analysis. Genes from the gene expression data will be ranked according a calculated ranking score.

Formula to calculate ranking score: signed Fold Change \* -log10(p-value)

Adjusted p-value cutoff < 1.00, so we will include all pathways which will later be filtered out based on significance. Other variables were kept standard

**Step 4:** Selection of GSEA results and creation of heatmap

Select pathway if in one of the conditions the p-value < 0.01 use these pathways to create a heatmap which depicts NES (maybe change this to ES which depicts, relatively, changes better). Cluster rows of heatmap.

NB: We are still discussing how to do GSEA. As GSEA compares genesets within the conditions and not between conditions, while we want to compare them between conditions. We want to do this statistically correct, so we are trying some things. Maybe we will switch steps 2 and 3. So first do GSEA analysis with all pathways (genesets) from the databases and then select them based on a list of significant enriched pathways from a GO-term.