

Project synopsis

2022-04-06

**Group 5: Astrid (s184248), Caroline (s183493), Christian (s165476),
Frederik (s184260) and Nils (s220672)**

Link to an overview of the GTEx files: <https://www.gtexportal.org/home/datasets>

Links for the data we will be using:

https://storage.googleapis.com/gtex_analysis_v8/annotations/GTEX_Analysis_v8_Annotations_SampleAttributesDS.txt

https://storage.googleapis.com/gtex_analysis_v8/annotations/GTEX_Analysis_v8_Annotations_SubjectPhenotypesDS.txt

https://storage.googleapis.com/gtex_analysis_v8/rna_seq_data/GTEX_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_reads.gct.gz

We will not be using all the data available, but will use a subset (e.g. a specific tissue). We will use the `sampleAttributes.txt` file to extract the relevant samples from the `gene_reads.gct` file. To ensure we do not break github, the data sets will not be on our github, but instead they will be downloaded automatically when running our script.

We plan to tidy the data using basic tidyverse functions for data wrangling. Then, we plan to investigate differential gene expressions and visualize this with plots. The DESeq2 package can help us calculate differential gene expressions if we have time for it.