# Supplementary material

## Supplementary methods

Here you might include a description of the methods used in obtaining and analysing the data before getting into this document, e.g. how sequencing data was aligned and counted.

## Supplementary tables and figures

def get\_geodata(GEO\_ID):  
 from urllib import request  
 from Bio import Geo  
 import gzip as gz  
 ftp\_url = f"https://ftp.ncbi.nlm.nih.gov/geo/series/{GEO\_ID[0:5]}nnn/{GEO\_ID}/soft/{GEO\_ID}\_family.soft.gz"  
 request.urlretrieve(ftp\_url, f"{GEO\_ID}.gz")  
 fh = gz.open(f"{GEO\_ID}.gz", 'rt')  
 records = Geo.parse(fh)  
 geo\_id\_dict = {}  
 for r in records:  
 if r.entity\_type == "SAMPLE":  
 ID = r.entity\_id  
 name = r.entity\_attributes['Sample\_source\_name\_ch1']  
 title = r.entity\_attributes['Sample\_title']  
 geo\_id\_dict[ID] = {"name": name, "title": title}  
 df = pd.DataFrame(geo\_id\_dict).T  
 return df

## Reproducibility

The code for reproducing this analysis is available in this [GitHub repo](https://github.com/NBISweden/workshop-reproducible-research/tree/main/jupyter). The repo contains:

* A Snakemake workflow for running all analysis steps
* A Conda environment file for installing all needed dependenciesx
* A Docker file for running the analysis in a well-defined and isolated system