**Data Stewardship: Gebruik van GitHub**

Tijdens dit project heb ik **GitHub** gebruikt om mijn projectbestanden gestructureerd te beheren en om te zorgen voor een overzichtelijke en reproduceerbare workflow. GitHub heeft ervoor gezorgd dat mijn scripts, data en resultaten goed vindbaar zijn en dat anderen mijn project eenvoudig kunnen begrijpen en herhalen.

**Structuur en Vindbaarheid**

In mijn GitHub repository heb ik een duidelijke mappenstructuur gebruikt:

* **/Data\_RAW**: bevat de ruwe data en de uitgepakte bestanden uit de zip-bestanden.
* **/Scripts**: bevat alle gebruikte R-scripts, inclusief het script voor het uitpakken van bestanden, het mappen van reads, het genereren van BAM-bestanden, de tellingen en de DESeq2-analyse.
* **/Resultaten**: hierin staan de outputbestanden, waaronder de resultaten van de DESeq2-analyse en de volcano plot.
* **/Documentatie**: bevat de README en uitleg over het gebruik van GitHub en Data Stewardship.
* **/Afbeeldingen**: hierin zijn de gemaakte figuren, zoals de volcano plot, opgeslagen.

Deze structuur zorgt ervoor dat data, scripts en resultaten logisch gescheiden zijn en dat het makkelijk is om specifieke onderdelen terug te vinden.

**Bestandsnamen en Versiebeheer**

Voor de bestandsnamen heb ik een duidelijke en consequente naamgeving gebruikt, bijvoorbeeld:

* Norm1.BAM, RA1.BAM voor de BAM-bestanden.
* count\_matrix.txt voor de tellingen.
* ResultatenWC3.csv voor de analyse-uitkomsten.
* VolcanoplotWC.png voor de visualisatie.

Door deze consistente naamgeving is direct duidelijk welke data bij welk onderdeel van het project hoort.

Met **GitHub versiebeheer** kan ik alle aanpassingen in mijn scripts en data volgen. Iedere wijziging heb ik opgeslagen via commits met duidelijke beschrijvingen, zodat het proces van begin tot eind inzichtelijk is. Dit zorgt ervoor dat ik altijd kan teruggaan naar eerdere versies indien nodig.

**Reproduceerbaarheid**

Mijn GitHub-pagina is volledig ingericht zodat anderen mijn onderzoek eenvoudig kunnen reproduceren:

* De scripts zijn voorzien van duidelijke commentaren die stap voor stap uitleggen wat er gebeurt.
* De mappen en bestanden zijn logisch gestructureerd.
* In het README-bestand staat een handleiding waarin uitgelegd wordt hoe je het project kunt opstarten en uitvoeren.

Een andere onderzoeker kan hierdoor gemakkelijk mijn analyse opnieuw uitvoeren en eventueel uitbreiden.

**Samenwerking en Open Science**

Door mijn project openbaar te maken op GitHub draag ik bij aan **open science**. Anderen kunnen gebruik maken van mijn scripts en resultaten. Als iemand bijvoorbeeld een vergelijkbaar onderzoek wil doen, kan mijn GitHub-pagina direct als basis dienen.

GitHub biedt ook de mogelijkheid om samen te werken via branches en pull requests, wat ik in dit project nog niet heb gebruikt, maar wat in toekomstige projecten goed toepasbaar is.