



GROUP 3: CODE AND DATA

CAROLINE FORSYTHE, JULIA LAPUCHA & ASTGHIK SARUKHANYAN

DATA

Raw RNA-Seq data	<ul style="list-style-type: none">• https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE134056
Raw MBD-Seq data	<ul style="list-style-type: none">• https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE134052
Sample Information	<ul style="list-style-type: none">• manually assembled, for 38 samples: https://git.imp.fu-berlin.de/lapuj92/group3/-/blob/main/data/sample_info.txt• manually assembled, for 77 samples: https://git.imp.fu-berlin.de/lapuj92/group3/-/blob/main/data/allMethylationSample_info.txt
Annotated Methylation Data	<ul style="list-style-type: none">• https://git.imp.fu-berlin.de/lapuj92/group3/-/blob/main/data/annotatedMethylationCounts_allsamples.csv
Significant DEGs/DMRs	<ul style="list-style-type: none">• DEGs: https://git.imp.fu-berlin.de/lapuj92/group3/-/blob/main/data/sig_005.rds• DMRs: https://git.imp.fu-berlin.de/lapuj92/group3/-/blob/main/data/sig_DMR.rds
ML Input	<ul style="list-style-type: none">• Labels: https://git.imp.fu-berlin.de/lapuj92/group3/-/blob/main/data/sample_labels.csv• DEGs: https://git.imp.fu-berlin.de/lapuj92/group3/-/blob/main/data/sig_DEG_counts_37samples.csv• DMRs: https://git.imp.fu-berlin.de/lapuj92/group3/-/blob/main/data/transposedMethylationCounts.csv

GROUP 3: CODE AND DATA

Caroline Forsythe, Julia Lapucha & Astghik Sarukhanyan

CODE (GIT AND WHITEBOARD)

MethylationAnnotation	<ul style="list-style-type: none">• Language: R Input: raw methylation Data Output: annotated methylation data with ensemblids, boxplot for all 77 samples Function: Annotated chromosomal regions to unique ensemblids
Diff_Analysis_mbdseq	<ul style="list-style-type: none">• Language: R Input: Annotated Methylation Dataset Output: rds file with significant DMRs, transposed Table for ML, volcano and boxplot for DMRs Function: Preprocessing, differential methylation analysis
maseq	<ul style="list-style-type: none">• Language: R Input: raw RNA data, sample info (38 samples) Output: rds with DEGs, boxplot, volcanoplot Function: Differential Expression Analysis, genomic annotation.
GoTermEnrichment	<ul style="list-style-type: none">• Language: R Input: raw Data, DEGs, DMRs Output: Plots, tables Function: Functional enrichment Analysis using several packages and documentation of online tools
ML_model1	<ul style="list-style-type: none">• Language: Python Input: Labels, transposed DEGs and DMRs Output: Confusion Matrices and ROC curves for all classifiers, comparison table at the end Function: Trains a logistic regression classifier on the individual data sets and then builds a metamodel based on their averaged probabilities.
ML_model2	<ul style="list-style-type: none">• Language: Python Input: Labels, transposed DEGs and DMRs Output: Confusion Matrices and ROC curves for all classifiers, comparison table at the end Function: combine the Datasets and train 3 classifiers (logistic regression, decision, tree, SVM) and then stack them to an ensemble classifier.

REFERENCES

- Picture:
- <https://www.bmfsfj.de/bmfsfj/aktuelles/alle-meldungen/ueber-endometriose-aufklaeren-und-die-gynaekologie-dafuer-sensibilisieren-201906> (Access: 12.07.2024, 10:30)