PROJECT 3
SOFTWARETOOLS IN BIOINFORMATICS

GROUP
Accession code for your data set
at ArrayExpress
          Link to the corresponding paper

GROUP 1
E-MTAB-2086

DOI: 10.1111/acel.12234
DOI: 10.1126/sciadv.1500882

GROUP 2
E-MTAB-2533

DOI: 10.1182/blood-2014-02-553792

GROUP 3
E-MTAB-1560

DOI: 10.3892/ijo.2017.3893

GROUP 4
E-MTAB-2101

DOI: 10.1186/s12964-015-0096-8

GROUP 5
E-MTAB-4785

DOI: 10.1542/peds.2013-2552

GROUP 6
E-MTAB-5439

DOI: 10.1016/j.kint.2016.10.030

GROUP 7
E-MTAB-4609

DOI: 10.1152/physiolgenomics.00064.2016

GROUP 8
E-MTAB-2122

DOI: 10.1182/blood-2013-01-478420

GROUP 9
E-MTAB-6512

DOI: 10.3109/09553002.2014.905724

An end to end workflow for differential gene expression using Affymetrix microarrays
In this project you shall follow the workflow of an Affymetrix microarray differential expression (DE) workflow using Bioconductor packages from Bernd Klaus and Stefanie Reisenauer (EMBL Heidelberg).

Here and here you will find detailed step by step **instructions** together with the short methods descriptions. Both resources are complimentary and to small details same.
Web page: https://f1000research.com/articles/5-1384/v2

Your task is
I. Data source.
   Check your datas' page in Array Express. Study which information is available. Get accustomed with this data resource.
II. Scientific paper.
   Work through the paper and data assigned to your group. Read scientific paper. Get its message and the use-case for the DE analysis. Carefully read the corresponding methods section.
III. Replicate workflow.
   Follow and replicate examples in the workflow. Get accustomed with the code, data handling.
IV. Apply workflow to your assigned dataset.
   ! Your data will have different characteristics/columns/features you shall map them to the study design done in the reference paper to correctly group the samples for analysis.
   I would recommend you schedule the meeting with me to settle your study design at this point.
   Run the analysis workflow from above on your data. Document your logic, results, and findings.

.

DOCUMENTATION & CODE

I. Data source. (Points 10/100)
   - Overview on the ArrayExpress. Which information can be found/uploaded there? How? Which functionality it provides? Etc. (5/10)
   - Brief condense review on the data based on the information found on your E-MTAB-XXX ArrayExpress page. (5/10)
II. Scientific paper. (Points 30/100)
   - One+ DIN A4 (be humble with the font choice;) page critical review essay on the referenced paper assigned to your group. (15/30)
   - Condense information on the methods and tools used for the (differential) gene expression assay analysis. (5/30)
   - [write after completing Task III] Compare the methods and analysis authors of the paper have done to the capabilities of the workflow we focus on. (10/30)
III. Replicate workflow. (Points 10/100)
   - Provide print screens of the generated figures and reports with the short summary on the applied method and figure description. (10/10)
IV. Apply workflow to your assigned dataset. (Points 50/100)
   - Same as for III (30/50)
   - One+ DIN A4 page report to summarise your findings and compare them to the results published in the referenced article, scientific paper assigned to you. (20/50)

HOW TO SUBMIT

Per email at
daria.gaidar@bioinformatik.uni-saarland.de
Documentation in .pdf, code in .rtf

TILL WHEN TO SUBMIT

Your submission shall reach me by 13:00 of
14.02.2019 ?

USFULL INFO AND RESOURCES

R studio – Environment to run R

R cheat sheet – to have the syntaxis in hand

You are not alone – when you face some trouble with R function, e.g. ggplot(), there has most probably been someone who faced it before and others who helped to solve it. So, formulate your trouble well and google it. Or browse directly on StackOverflow.

```
"maEndToEnd" installation
tipps and tricks
```

Versions matter.
One of the dependencies, DO.db, might prefer to be installed under R.3.5 and Bioconductor.3.8. First install it, then change to R developer version 3.6 and Bioconductor.3.9 to proceed with a bulk dependencies installation for "maEndToEnd" as described in the workflow material.
You can swap between several versions of R on you machine under Tools-Global Options-R version in RStudio.