Interactive & Collaborative Assignments

Kristel Van Steen, PhD^2 (*)

kristel.vansteen@uliege.be

(*) GIGA-R Medical Genomics, Systems Genetics Lab, University of Liège, Belgium

Systems Medicine Lab, KU Leuven, Belgium

Name	**Topic**	**Main goal**
Leonor	GWAIS confounding: PRS	Rerun existing GWAIS analysis protocol but for a disease trait that has been adjusted for PRS. Compare old and new runs
Laura	GWAIS methods: EDCF	Perform GWAIS analysis with the indicated algorithm. Compare with results from fellow students who also perform GWAIS
Nicolas	GWAIS methods: GWGGI	Perform GWAIS analysis with the indicated algorithm. Compare with results from fellow students who also perform GWAIS
Ba-Thien	GWAIS methods: RF	Implement RF algorithms that (without extra programming) rank pairs of SNPs. Compare between different implementations or with results from fellow students
Robin	GWAIS input	Assess impact of two new recoding for SNPs on GWAIS results: RFselect and EDGE encoding
Christophe	GWAIS interpretation	Given GWAIS results, use STRING network and network propagation tools to enlarge the discovery set of genes. Assess overlap with in-house discovered gene-pairs or via GWAIS results obtained via fellow students
Yansen	GWAIS heterogeneity	Rank-based approaches to compare GWAIS results

GWAIS confounding: PRS

1. Take a provided extended SNP set and compute PRS by following the recommendations in the provided tutorial Choi et al. 2020.
2. Compare with the PRS scores already computed in-house
3. Rerun the analysis pipeline provided by Diane Duroux, who also provides an additional supporting doc with her results. The difference between her and your results will be that the disease trait in the analysis of Diane was not a priori corrected for PRS scores, but only a posteriori.
4. How do the final results differ? What are your conclusions?
5. [Take CD cases and controls as training data and test on UC cases; Take UC cases and controls as training data and test on CD cases; discuss in view of the tutorial guidelines]

GWAIS methods

For all who implement a GWAIS method:

- Take the two provided case-control datasets 1 and 2. Both sets have been cleaned and LD pruned. Hence, if you need data with strong LD (r2>0.75) still present, then contact us.
- If your method requires complete data (so no missing SNPs), then use the provided complete data instead. Otherwise, use the incomplete data.
- Both datasets 1 and 2 (possibly completed as needed, see previous bullet point) need to be analyzed and results compared
- This is the minimum analysis workflow
- Extended workflow is given next, for each method separately

GWAIS methods: EDCF

1. Perform GWAIS as explained in general
2. Optimize parameters of EDCF via the real-life data
    a. Use insights from the EDCF developers
    b. Define “optimal”
3. Discuss the impact of different parameter settings on EDCF results

GWAIS methods: RF

1. Perform GWAIS as explained in general, using “default” options for parameters
2. Explicitly compare predictive performance of RF with MB-MDR:
   a. Via the provided reference Gola et al. 2019 (code available: https://github.com/imbs-hl/MBMDRClassifieR)
   b. Via the Le et al 2020 https://lelaboratoire.github.io/rethink-prs-ms/; code available)
3. Discuss findings. You may potentially improve the predictive performance of RF by optimizing parameters.

GWAIS methods: GWGGI

1. Perform GWAIS as explained in general, using “default” options for parameters or as recommended
2. If you have chosen TAMW before, than also consider using LRMW (or vice versa)
3. Discuss differential findings.

GWAIS input

1. Take one of the two provided case-control datasets. Both provided sets have been cleaned and LD pruned. For convenience, consider the completed data (so no missing data on SNPs).
2. Implement the RFselect way of recoding SNP pairs (no need to perform RF at this point)
3. Implement the EDGE encoding of recoding SNPs

   ----
4. Perform two GWAIS on the data with the encodings developed in 2 and 3. You can choose which analysis method (f.i. classical RF with variable importance scores in case of encoding 2. and RF with SNP pair evaluation scores in case of encoding 3.; OR any other ML approach)

Table 2. Recessive, additive, dominant, codominant and EDGE encoding schemes (Hall et al. 2021 – under review).

Biological Action	Homozygous Referent	Heterozygous	Homozygous Alternate
Recessive	0	0	1
Additive	0	0.50	1
Dominant	0	1	1
Codominant (Het)	0	1	0
Codominant (HA)	0	0	1
EDGE	0	**α**	1

A. $Y \sim \beta_{Het}SNP_{Het} + \beta_{HA}SNP_{HA}$

B. $\alpha = \beta_{Het} / \beta_{HA}$

GWAIS interpretation

1. Consider in-house generated gene-based GWAIS results (provided)
2. Per GWAIS analysis: Superimpose the genes on a molecular network (such as STRING), select at least one propagation method to enlarge the provided set of genes by explicitly exploiting the "known" molecular interactions from the chosen molecular network
3. The result per GWAIS analysis will be an enlarged set of genes:
   a. Do you find evidence that the new genes are linked to inflammatory bowel disease? (e.g. DisGeNet searches)
   b. Is the overlap between enlarged sets across GWAIS larger than the overlap between original GWAIS result sets?
4. For the gene-based GWAIS networks considered as similarity graphs, find clusters of nodes (e.g., via eigenvectors of the corresponding Laplacian matrices)

Updates

For all:

- Softwares PLINK, MBMDR and PRSice:

  /massstorage/URT/GEN/BIO3/Student2021/softwares
- PLINK example:

  /massstorage/URT/GEN/BIO3/Student2021/softwares/plink_example
- In-house generated GWAIS results:

  /massstorage/URT/GEN/BIO3/Student2021/Data/inHouse_GWAIS_results

Specific:

- Ba-Thien: Last version of R will be available on the cluster next week, with packages sent to Diane already installed.
- Leonor:
    - Original data to compute the PRS: */massstorage/URT/GEN/BIO3/Student2021/Data/originalData*
    - GWAIS analysis pipeline: */massstorage/URT/GEN/BIO3/Student2021/Leonor/SNPtoGene_pipeline*
- Christophe:
    - string network: */massstorage/URT/GEN/BIO3/Student2021/Data/GeneInformation/String*
    - Packages already sent by email will be installed in R