Supporting docs & instructions

- Lecture 1_ HW Organization
  - At the end of the year, each group should have selected minimally 1 Genetics Literature Style homework and minimally 1 Bioinformatics Literature Style homework. Depending on HW1 choices, you may have limited choice for HW2 (X)

Group	Assignment	Genetics		Bioinformatics	
		Q&A	Literature	Q&A	Literature
1	1	X		X	
	2		X		X
2	1	X			X
	2		X	X	
3	1		X		X
	2	X		X	
4	1		X	X	
	2	X			X

- Literature Style Homework:
  - Lecture 1_Critical Evaluation of a Paper
  - Duration of group presentations will depend on the number of groups who selected "Literature Style"
- Communicate your selection for Genetics and for Bioinformatics to kristel.vansteen@uliege.be before 24th of October 2020
- Send your completed homework assignments for Genetics to fdequiedt@uliege.be and for Bioinformatics to kristel.vansteen@uliege.be

---

Homework Assignment Bioinformatics

- **Bioinformatics Q&A style**

Q1. (3 pts)

Download test data named “gwas_a.bed”, “gwas_a.bim” and “ gwas_a.fam” from the course website.

Q1.1 How many SNPs crossed the threshold of MAF < 0.01?

Q1.2 How many SNPs are in LD with window size “100”, “200”?

(Hint: PLINK or R)

Q1.3 Develop the plots using ggplot2 R library.

Q2. (3pts)

Q2.1 Give chromosomal and functional details of identified genes?

Q2.2 How many of these SNPs are showing regulatory motif alterations?

(Hint: Convert SNPs into gene ID and use FUMA or other webserver to make interpretations)

1	rs4468290
2	rs11201609
3	rs4933212
4	rs701546
5	rs1241901
6	rs8087497
7	rs2409457
8	rs1666559
9	rs12943387
10	rs2036660

Q3. (3pts)

Q3.1 Define the chromosomal locations for the given gene list.

Q3.2 Identify the tissue specific expression of the following genes and assess whether these are differentially expressed in a specific tissue or not.

1	A2M	8	KCNAB2
2	A4GALT	9	KCNH2
3	A4GALT	10	KCNH7
4	KALRN		
5	KBTBD10		
6	KARS		
7	KCNA5		

Q3.3 Explain FUMA, GTex in more detail and how these can be used to “move from association to function”

Q4. (2pts)

Why was the first genome sequence considered a milestone? Please choose:

a) It proved that viruses contained nucleic acids.

b) It proved that DNA could be isolated from a virus.

c) It showed the coding sequences for all the proteins produced by a virus.

d) It proved that the coding sequences within a genome could be located and identified.

Q5. (2pts)

What is true about variation within the human genome?

a) There is more variation among regional groups than between individuals.

b) There is tremendous variation among individual human genomes.

c) There is comparatively little variation between individuals.

d) Nothing is yet known about the variation within the human genome.

Q6. (2pts)

The hypothesis of a GWAS is that

a) common haplotypes carry rare mutations

b) the odds rations of associated alleles is greater than 1

c) exons show an increased frequency of alleles associated with disease

d) there is a difference in allele frequency between affected and non-affected individuals

- **Bioinformatics Literature style**

Select one of the following papers. Topics are complementary to the classes.

- **Exploratory data analysis applied to GWAS**:
  Yik Y. Teo, Exploratory data analysis in large-scale genetic studies, *Biostatistics*, Volume 11, Issue 1, January 2010, Pages 70–81, https://doi.org/10.1093/biostatistics/kxp038

- **Testing/modelling approaches in GWAS**:
  Manduchi, E., Orzechowski, P.R., Ritchie, M.D. *et al.* Exploration of a diversity of computational and statistical measures of association for genome-wide genetic studies. *BioData Mining* **12,** 14 (2019). https://doi.org/10.1186/s13040-019-0201-4

- **Visualization approaches useful to GWAS:**
  Grace, C., Farrall, M., Watkins, H. *et al.* Manhattan++: displaying genome-wide association summary statistics with multiple annotation layers. *BMC Bioinformatics* **20,** 610 (2019). https://doi.org/10.1186/s12859-019-3201-y
  Combined with either
  Westreich, S.T., Nattestad, M. & Meyer, C. BigTop: a three-dimensional virtual reality tool for GWAS visualization. *BMC Bioinformatics* **21,** 39 (2020). https://doi.org/10.1186/s12859-020-3373-5
  or
  https://cran.r-project.org/web/packages/CMplot/CMplot.pdf

- **Pathway analysis following testing/modelling in GWAS and DNA-seq – a protocol:**
  White MJ, Yaspan BL, Veatch OJ, Goddard P, Risse-Adams OS, Contreras MG. Strategies for Pathway Analysis Using GWAS and WGS Data. Curr Protoc Hum Genet. 2019 Jan;100(1):e79. doi: 10.1002/cphg.79. Epub 2018 Nov 2. PMID: 30387919; PMCID: PMC6391732.

- **Multiplicity 1: Pleitropy - a single gene influencing two or more distinct phenotypic traits**
  Gittu George, Sushrima Gan, Yu Huang, Philip Appleby, A S Nar, Radha Venkatesan, Viswanathan Mohan, Colin N A Palmer, Alex S F Doney, PheGWAS: a new dimension to visualize GWAS across multiple phenotypes, *Bioinformatics*, Volume 36, Issue 8, 15 April 2020, Pages 2500–2505, https://doi.org/10.1093/bioinformatics/btz944

- **Multiplicity 2: Multilocus GWAS – joint testing with multiple SNPs at a time (application to methylation traits)**
  Zhang, H., Shi, J., Liang, F. *et al.* A fast multilocus test with adaptive SNP selection for large-scale genetic-association studies. *Eur J Hum Genet* **22,** 696–702 (2014). https://doi.org/10.1038/ejhg.2013.201 [check out p-value determination and multiple testing correction]

- **Multiplicity 3 – Multivariate gene-based**
  Chung, J., Jun, G.R., Dupuis, J. *et al.* Comparison of methods for multivariate gene-based association tests for complex diseases using common variants. *Eur J Hum Genet* **27,** 811–823 (2019). https://doi.org/10.1038/s41431-018-0327-8

- **Population structure in GWAS**
  Hellwege JN, Keaton JM, Giri A, Gao X, Velez Edwards DR, Edwards TL. Population Stratification in Genetic Association Studies. *Curr Protoc Hum Genet*. 2017;95:1.22.1-1.22.23. Published 2017 Oct 18. doi:10.1002/cphg.48 [you may focus on PCA based methods only]

- **Genetic associations with rare variants – a protocol:**
  Dering C, König I, Ramsey L, Relling M, Yang W, Ziegler A. A comprehensive evaluation of collapsing methods using simulated and real data: excellent annotation of functionality and large

Alternatively, you may be interested in a practical GWAS analysis. In this case, select GWAS from the last 5 years. How to discover such studies?

- Go to https://www.disgenet.org/ and the SEARCH button
- Go to "General links" and "Summary of All Variant-Disease Associations"
- Click on ADD/REMOVE FILTER
- Type in the DISEASE of your choice (make sure you use the convention suggested by the data bases)
- Select in addition SOURCE GWASCAT if you wish to have GWAS studies from the GWAS catalogue; alternatively select SOURCE GWASDB
- Take most recent studies "Last Ref" and/or those with a good number of evidences via "N PMIDs"
- GO!

Homework assignment Genetics

- **Genetics Q&A style**

Q.1 (5pts)
Going into as much detail as possible, explain which experimental results were instrumental for Watson and Crick when establishing their model for the DNA molecule. Explain the experiments and underline the main conclusions that Watson and Crick used in their original 1953 Nature paper. Try and come up with an alternative model for the DNA molecule that fulfills as many of the same experimental data as possible.

Q.2 (5pts)
Going into as much detail as possible, explain what SNP are and what are they useful for. In case of human medicine, what are the current models that explain that SNP can be associated with disease and pathological states (when addressing this part, consider the location of the SNP with regards to important functional genes or genetic elements)?

Q.3 (5pts)
Scientists have discovered a new organism living in the depths of the ocean. This creature is fascinating because it produces light. Scientists would now like to characterize the genome (1 billions of pb) of this creature and identify the gene responsible for light production. You are in charge of this project and are allocated (almost) infinite ressources. Going into as much detail as possible, describe a protocol, that would allow the sequencing of the genome and the identification of the gene of interest. Justify your experimental choices (sequencing technology, etc..).

- **Genetics Literature style**

Select one of the following papers.

- Genome editing retraces the evolution of toxin resistance in the monarch butterfly
  Nature. 2019 Oct;574(7778):409-412. doi: 10.1038/s41586-019-1610-8. Epub 2019 Oct 2.
  https://www.nature.com/articles/s41586-019-1610-8

- Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study
  Lancet Infect Dis. 2020 Jul 14;S1473-3099(20)30562-4. doi: 10.1016/S1473-3099(20)30562-4.
  https://www.thelancet.com/pdfs/journals/laninf/PIIS1473-3099(20)30562-4.pdf

- Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands.
  Nature Medicine volume 26, pages1405–1410(2020)
  https://www.nature.com/articles/s41591-020-0997-y

- Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior
  Science 30 Aug 2019: Vol. 365, Issue 6456,
  https://science.sciencemag.org/content/365/6456/eaat7693
