

Computational Genomics 2025

Homework 2: Iranome

PLEASE NOTE THE FOLLOWING INSTRUCTIONS:

- 1. You are to complete this assignment alone. The assignment is open book, so you are allowed to use any books or information available online, your own notes and your previously constructed code, etc. HOWEVER YOU ARE NOT ALLOWED TO COMMUNICATE OR IN ANY WAY ASK ANYONE FOR ASSISTANCE WITH THIS EXAM IN ANY FORM.*
- 2. Please pay attention to instructions and complete ALL requirements for ALL questions, e.g. some questions ask for code, plots, AND written answers.*
- 3. A complete answer to this assignment will include a single pdf file named in this format: 'your_name'.compG_HW2.pdf. This file should include all of your commands in linux from installation to the end, your codes, and your resulting plots.*
- 4. The exam must be sent to **mehrmohamadi@ut.ac.ir** by 11pm on the 15th of Ordibehesht, 1404. It is your responsibility to make sure that it is received by then and no excuses will be accepted.*

The Iranome database (www.iranome.com) was established by performing whole exome sequencing on 800 individuals from eight major ethnic groups in Iran in the first phase of the project. The groups included 100 healthy individuals from each of the following ethnic groups: Arabs, Azeris, Balochs, Kurds, Lurs, Persians, Persian Gulf Islanders and Turkmen. In the second phase of the project, 400 healthy individuals from additional four ethnic groups including Gilaki, Mazani, Sistani and Zartoshti (Zoroastrian) were added. These 1200 individuals from 12 different ethnic groups represent about 90 million Iranians and to some degree half a billion individuals who live in the Middle East, a region with rapid population growth expectations for the future. These ethnic groups are among the most underrepresented populations in currently available human genomic variation databases.

Use the Iranome browser to find **5 SNPs** with considerable allele frequency (AF) differences among **2 of the above ethnic groups**. For this you will need to search human gene names of your choosing (e.g. HLA-B), in the search box. Next, go through the SNP list shown and click on each SNP to be directed to a page where group-wise AFs are provided.

- 1) Describe all criteria you used for selection of your final 5 SNPs. For each of the 5, provide AF table across ethnic groups and highlight the two ethnic groups that you are considering.
- 2) Now assume you have genomic data on a new individual from Iran and you would like to estimate their ethnic admixture status using the Iranome data. Assume that your new individual has ancestry only from the two ethnic groups that you chose when selecting your SNPs, and assume **that he is heterozygous for all 5 of your SNPs**. Use a model-based inference approach to estimate the fractional ancestry for this individual between the two ethnic groups. Provide your answer including the alpha value that maximizes this individual's genotype probability, your code, your plot of probability vs. alpha values, and your interpretation of the results.