BIOINFORMATICS INSTITUTE

Building a perfect human

May 27, 2024

Homework number 5

by Sapozhnikov N.A.

1 Introduction

Genetic variation is the foundation of biological diversity and an important driver of evolutionary processes. Understanding these variations can provide insights into an individual's ancestry, physical traits, and predisposition to certain diseases. The advent of genotyping technologies, such as genotyping chips and CRISPR-Cas9, has revolutionized genetics research and personalized medicine. Genotyping chips allow for the high-throughput analysis of single nucleotide polymorphisms (SNPs) across the genome, while CRISPR-Cas9 offers precise genome editing capabilities.
This lab report explores the analysis of genetic data from a 23andMe raw dataset, focusing on SNP identification, ancestry determination, and phenotype prediction. We can use SNPs to predict likelihood of having some phenotypic trait, and, more important, likelihood of disease. And this information may require actions - changing in lifestyle, or even medical intervention.

2 Methods

The raw data was obtained from a 23andMe genetic test, provided in a text file format containing SNP information. Plink2 was used for converting 23andMe raw data into the standard VCF format. Then mtDNA and Y-DNA haplogroup were identified vie International Society of Genetic Genealogy (ISOGG) tools. To visualize SNPs Integrated Genomics Viewer (IGV) was used. The annotation of SNPs with functional information and comparing them with ClinVar and GWAS databases was performed with SnpEff/SnpSift. And finally, VEP (Variant Effect Predictor) was used to identify clinically significant variants.

3 Results

The raw 23andMe data was successfully converted to VCF format, containing all analyzed SNPs. SNPs corresponding to deletions and insertions were removed.
Maternal Haplogroup (mtDNA): Determined to be Haplogroup H (T152C), common in Western Europe.
Paternal Haplogroup (Y-DNA): Determined to be Haplogroup R1a1a, prevalent in Europe.

Phenotypic Traits:
Sex: Male, indicated by the presence of Y-chromosome SNPs.
Eye Color: Predicted to be brown based on SNPs associated with eye color traits (rs12896399, rs16891982, rs12913832).
Clinically Relevant SNPs Several SNPs were identified as clinically significant, with annotations indicating their potential impact on health: rs4680(G), rs8176746(T), rs6265(T), rs16890979(T).

4 Discussion

Variants found were annotated and from this annotation their function and molecular mechanisms can be assumed.

1. COMT gene (rs4680): The COMT gene codes for the COMT enzyme, which breaks down dopamine in the brain's prefrontal cortex. The wild-type allele is a (G), coding for a valine amino acid; the (A) substitution polymorphism changes the amino acid to a methionine. This alters the structure of the resultant enzyme such that its activity is only 25% of the wild type. As a result, A allele carriers have more dopamine in their prefrontal cortex, which may be responsible for many of the neuropsychological associations.

2. ABO blood group system (rs8176746): Determines blood type and has been linked to various health outcomes, including susceptibility to certain infections and cardiovascular disease. rs8176746 has a status 'affects' in clinvar.

3. BDNF gene (rs6265): This SNP is associated with brain-derived neurotrophic factor, which plays a role in neuronal growth and survival. Variants in this gene have been linked to differences in cognitive function, mood disorders, and susceptibility to psychiatric conditions like depression and anxiety. Variant is marked as benign and risk factor in clinvar, although there were no clear evidence in the reference article.

4. SLC2A9 gene (rs16890979): Linked to uric acid metabolism and risk of gout. Variants in this gene have been associated with differences in serum urate levels and susceptibility to gout. rs16890979 has been found to be associated with gout in several independent studies. It may be a variation in the SLC2A9 gene, which is more commonly known as GLUT9. So here we have a binign variant that is associated with 1.7x risk of gout according to SNPedia.

5 Conclusion

The identified haplogroups provide insights into the subject's ancestral origins, aligning with known historical migrations and genetic distributions. The phenotypic predictions, such as sex and eye color, were consistent with the expected outcomes based on SNP data. Clinically relevant SNPs offer valuable information for potential health risks, enabling personalized healthcare strategies.