**Bioinformatics for Ukraine**

**Genomics, Phylogenetics, and Evolutionary Biology (Oakland University) Module**

**Assignment 1 Genomic Browsers and Databases**

Using USCS browser (https://genome.ucsc.edu/), NCBI(https://www.ncbi.nlm.nih.gov/), ENSEMBL (https://ensembl.org/), and SNPedia(snpedia.org) investigate the Human genome for information. Fill in the tables below.

If asked – please select GRCh38/hg38 Human Reference Genome for answers.

1. **Genes:**

Use any database or combination of them to investigate the following genes and provide answers.

|  |  |  |  |
| --- | --- | --- | --- |
| **Gene name** | **Сhrom** | **Position** | **Impact\role\significance** |
| BRCA1 (Breast Cancer 1) | Chr 17 | q21.31 | BRCA1 is a human tumor suppressor gene. Some of variations can increase risk of developing breast and/or ovarian cancer at an earlier age |
| TP53 (Tumor Protein p53) | Chr 17 | p13.1 | TP53 encodes a tumor suppressor protein. Mutations in this gene are associated with a variety of human cancers, including hereditary cancers such as Li-Fraumeni syndrome. |
| APOE (Apolipoprotein E) | Chr 19 | q13.32 | The protein encoded by this gene is a major apoprotein of the chylomicron. Mutations in this gene result in familial dysbetalipoproteinemia, or type III hyperlipoproteinemia (HLP III), in which increased plasma cholesterol and triglycerides are the consequence of impaired clearance of chylomicron and VLDL remnants. |
| CFTR (Cystic Fibrosis Transmembrane Conductance Regulator) | Chr 7 | q31.2 | This gene encodes a member of the ATP-binding cassette (ABC) transporter superfamily. Mutations in this gene cause cystic fibrosis, the most common lethal genetic disorder in populations of Northern European descent. |
| HBB (Hemoglobin Subunit Beta) | Chr 11 | p15.4 | The HBB gene encodes beta globin. Mutations in the HBB gene may lead to several conditions, the best known of which are sickle cell anemia and beta-thalassemia, as well as beneficial aspects such as resistance to malaria. |

1. **Variation function:**

Use any database or combination of them to investigate the following SNPs(single nucleotide polymorphisms) and provide answers.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variant name** | **Gene** | **Reference** | **Alternative** | **Suggested Impact\function** |
| rs1801133 | MTHFR | G | A  C | Encodes an enzyme involved in folate metabolism. |
| rs429358 | APOE | T | C | The combination of two C alleles has a strong influence on the risk of Alzheimer's disease. Varied combination 3x increased risk for Alzheimer's; 1.4x increased risk for heart disease. And two T is a normal. |
| rs5219 | KCNJ11 | T | A  C  G | Represents risk for type-2 diabetes. |
| rs1800562 | HFE | G | A | Accounts for ~85% of all cases of hemochromatosis, a disorder whose symptoms include cirrhosis of the liver, diabetes, hypermelanotic pigmentation of the skin, and heart failure. |

1. **Population differences:**
   1. Use ENSEMBL to investigate the population distribution of the following variants

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variant name | Reference | Alternative | Frequency in AFR | Frequency in EAS | Frequency in EUR |
| rs4988235 | G | A | G: 97%  A: 3% | G: 100% | G: 49%  A: 51% |
| rs1805007 | C | T | C: 100% | C: 100% | C: 93%  T: 7% |
| rs1799971 | A | G | A: 99%  G: 1% | A: 61%  G: 39% | A: 84%  G: 16% |
| rs1051730 | G | A | G: 91%  A: 9% | G: 97%  A: 3% | G: 63%  A: 37% |

* 1. Investigate the function of variation from table 3.1, and write a single paragraph about the implications of different frequencies in different populations.

**Example:**

The **rs1801282** SNP, located in the **PPARG** (Peroxisome Proliferator-Activated Receptor Gamma) gene, is associated with insulin sensitivity and the **risk of developing Type 2 diabetes**. This SNP shows significant differences in allele frequencies between populations, being **more common in East Asian (EAS)** populations compared to European (EUR) and African (AFR) populations, which may contribute to increased susceptibility to insulin resistance and Type 2 diabetes in East Asian populations.

The rs4988235, located in the MCM6 (Minichromosome Maintenance Complex Component 6) gene, is associated with lactose intolerance. This SNP shows significant differences in allele frequencies between populations: European populations are divided into two almost equal groups who are likely to be lactose intolerant as an adult and who are likely to be able to digest milk as an adult rather than Africans and East Asians. The latter are almost all likely to be lactose intolerant as an adult.

The rs1805007, located in the MC1R (Melanocortin 1 Receptor) gene, is associated with red hair color. This SNP shows that Europeans have a higher risk of melanoma. For African and East Asian people, allele which carries a red hair-associated variant is uncommon, so they are exposed to a low risk of melanoma.

The rs1799971, located in the OPRM1 (Opioid Receptor Mu 1) gene, is associated with the risk of alcoholism. This SNP shows that African populations have the lowest risk of being alcoholic. European populations have only 16% frequency with alleles which codes stronger cravings for alcohol and East Asian populations frequency twice as much as European.

The rs1051730, located in the CHRNA3 (Cholinergic Receptor Nicotinic Alpha 3 Subunit) gene, is associated with increased risk of lung cancer. This SNP shows that European populations have the biggest allele frequency which increase the risk of lung cancer.