

Syllabus for Database Independent Studies for BPG students during Summer semester instructed by Dr. Alexei Fedorov

The main goal for this study is to investigate numerous human mutations using public Databases. Students will try to evaluate possible effects of these mutations on phenotypes. Population Genetics rules will be applied and discussed. Several GWAS studies will be evaluated.

During the first part of this project, students will intensively work with one of the largest databases – “1000 Genomes” using on-line tools and also as downloadable *vcf* text files of “1000 Genomes” in local UT computers. Students will learn how to analyze structure of SNP haplotypes in human genes and evaluate the frequency of these haplotypes across populations.

During the second part of this project, the students will work with the next-generation sequencing mRNA databases – so-called the Sequence Read Archive (SRA), using on-line NCBI browser. The goal of this research is to analyze the expression of a particular genomic region in different human tissues and at different conditions.

During one month Dr. Fedorov will supervise the students in this project in his laboratory. The time is flexible. One credit hour for this course corresponds to 5 weeks of LABs.

Students will be expected to take the MOOC online Database course (<http://www.coursetalk.com/stanford-online/sql>) in order to gain SQL experience. Additional credit may be earned for this activity.

L1. Introduction to Databases **1h14m** May 16th.

L2. SNP analysis, Gene-arrays and chips. Finding a mutation that caused Retinitis Pigmentosa. **1h7m** May 18th.

L3. SRA database of RNA-seq public data. **1h2m** May 23rd.

L4. NONCODE Database of lincRNAs. Ensemble Database of human-mouse whole genome alignments. Errors in database interpretations. README files. **1h1m** May 25th.

L5. SNP Databases (dbSNP, 1000Genomes, SNPedia, ClinVar, GTEEx Portal). **1h27m** May 30th.

L6. 1000 Genomes Project, online databases and tools. Lecture. **1h17m** June 1st.

L7. Analysis of GWAS databases **59m** June 6st.

L8. Discussion of HWs #3, 5, 6 (41m) + Online lecture about VCF format, which is most common for genomic databases (40m) June 8th.

L9. Analysis of FTO gene (structure, functioning and evolution) using NCBI on-line databases and tools 1h10m June 13th.

L1. Essential questions about FTO gene. 1h4m June 15th.

L11. Final project for RNA-Seq data interpretations. 1h30m June 20th.