IDENTIFICATION OF KAZAKH SPECIFIC GENOMIC VARIANTS USING COMPARATIVE GENOMICS ANALYSIS

A. Molkenov[1], A. Daniyarov[1], A. Sharip[1], A. Seisenova[1], D. Karabayev[1], U. Kairov[1]

[1]Laboratory of Bioinformatics and Systems Biology, Center for Life Sciences, National Laboratory Astana, Nazarbayev University, Nur-Sultan, Kazakhstan

Keywords: bioinformatics, genomics, biomedicine, complete human genome, bioinformatics, genomics, human population genetics.

Introduction: The modern development of high-performance genomic technologies opens up new possibilities for studying the human genome. Large-scale genomic research generates huge amounts of data, the active development of bioinformatics with the availability of modern methods and approaches of analysis makes it possible to create detailed databases and comprehensively study genomic data. One of contemporary task is to study and identify specific genomic variants of population by detailed analysis of complete genome and complete exome data comparison with open large-scale population datasets.

Materials and methods: Materials of the study are 14 complete genomes and 125 complete exomes of Kazakhstani individuals. Our dataset was replenished with data from large whole genome population datasets (SGDP, PRJEB26349, HGDP and 1000 Genomes) for comparative population genomics and to search and identify specific genomic variants. The data in the raw format was mapped and aligned on a single reference genome hg19, then genomic variants were searched and an individual map of the found variants was formed for each dataset in the VCF format. For replenished datasets formed a general map of all variants, which were then excluded from the total number variants found for of Kazakh sampling to search for specific genomic variants. Then the filtered variants were annotated and interpreted.

Results: For Kazakp whole exomes were found 9 heterozygous or mutant variants unique among formed genomic databases. 7 variants located on the intron region, 1on the upstream and the last variant frameshift deletion on exonic region.

For the Kazakh whole genomes were found 4732heterozygous or mutant variants, 517 variants presented among all Kazakh samples and 144 variants were completely mutant. Only 8 SNVs are located at exonic region: 4 synonymous SNV, 3 nonsynonymous SNV, and 1frameshift deletion.

Conclusion: We have discovered unique several genomic variants specific for now to the kazakh individuals. These results can serve as a basis for the creation of a Kazakh reference genome, subsequent research and comparative analysis of Kazakh individuals with various populations of the world.

Grant references: AP05135430; MES RK.