GENETIC ARCHITECTURE OF KAZAKH POPULATION

U.Kairov*, A.Molkenov, M. Zhabagin, A.Askapuli, A.Abilmazhinova, A.Akhmetova, D.Yerezhepov, Zh.Abilova, S.Rahimova, Zh.Zhumadilov, A.Akilzhanova

Department of Genomic and Personalized Medicine, Center for Life Sciences, Nazarbayev University, Astana, Kazakhstan;
*ulykbek.kairov@nu.edu.kz

INTRODUCTION.

Kazakhstan is a unique country located in the middle of Central Asia, lying on the ancient Great Silk roads. Kazakh populations have been strongly influenced by the nomadic lifestyle, and long history of migration has led to admixture of western and Asian populations, which has moulded the genetic architecture. Thus it is crucial to comprehend the genetic background of ethnic Kazakhs to properly investigate the genetic basis of common diseases or traits in Kazakh populations.

AIM AND OBJECTIVES.

The aim of the project is to define genetic architecture of Kazakh population by sequencing and analyzing whole exomes and whole genomes of Kazakhs living in different regions of Kazakhstan. To achieve this aim we have set the following objectives:

- recruit 100 healthy ethnic Kazakhs, including males and females aged 20 to 50 who reside in different regions of Kazakhstan (from different tribes) and 10-20 healthy elderly ethnic Kazakhs (aged 80+), collect health information;
- sequence 100 whole exomes and 10 whole genomes of Kazakhs to determine their genetic architecture;
- identify disease genes in observed Kazakh population data and perform comparative analysis with preexisting international data;

MATERIALS AND METHODS.

DNA was extracted from peripheral blood from among 37 samples with using of Promega Wizard Genomic DNA Preparation Kit.

Whole genome sequencing was performed on Illumina HiSeq2000 sequencer using TruSeq SBS Kit v3 - HS (200 Cycles). *.bcl files were simultaneously converted and demultiplexed using bcl2fasta application. Mapping and alignment of sequence reads performed using bowtie2 and samtools to human b19 reference genome. Sorting and indexing of reads, removing of intermediate files, *.bam files assembling, and marking duplicates were performed using PicardTools package. To perform raw data processing and running program scripts Java Runtime Environment and R Bioconductor package were installed.

RESULTS AND DISCUSSION.

Standard operational procedures for shotgun sequencing of DNA on the next-generation sequencing platform Illumina HiSeq2000 were prepared. 50 participants were recruited for the project, and blood collected (7 members for sequencing of whole genome and 42 participants for exome sequencing). DNA was extracted from collected samples for further preparation of DNA libraries and cluster generation on cBot.

Pilot launch were performed. All steps for performing the next-generation sequencing have been worked out.

Seven whole genomes sequencing, further alignment and mapping procedures on reference genome HG19 were completed. Sequencing and processing of exomes from 30 samples ongoing.

ACKNOWLEDGMENTS.

Genome Medicine Institute, Institute Curie, Harvard Medical School.