
WHOLE-GENOME SEQUENCING AND GENOMIC VARIANT ANALYSIS OF KAZAKH INDIVIDUALS

Ulykbek Kairov¹, Askhat Molkenov¹, Arang Rhie¹, Saule Rakhimova¹, Seong-Keun Yoo², Ulan Kozhamkulov¹, Aigul Sharip¹, Ainur Akhmetova¹, Aliya Abilmazhinova¹, Madina Zhalbinova¹, Dauren Yerezhepov¹, Maxat Zhabagin¹, Jeongeon Shin², Peter Kharchenko³, Jong-Il Kim², Joseph Lee⁴, Jeong-Sun Seo², Ainur Akilzhanova¹ and Zhaxybay Zhumadilov¹

¹Center for Life Sciences, National Laboratory Astana, Nazarbayev University, Kazakhstan

²Genomic Medicine Institute, Seoul National University, Korea.

³Harvard Medical School, USA.

⁴Joseph Lee, Columbia University, USA.

ulykbek.kairov@nu.edu.kz

Introduction: Kazakhstan is the multi-ethnic country with the dominant Kazakh ethnic group in the middle of the Eurasian continent. All existing human genetic diversity datasets are only partial representations of worldwide human diversity. The information about genetic diversity of Kazakh population is lacking in these datasets. Our study aims to perform whole-genome sequencing of five healthy Kazakhs to provide insights into genetic structure and diversity of Kazakh populations in Kazakhstan.

Materials and Methods: Genomic DNA of five healthy Kazakhs was isolated from peripheral blood and sequenced with Illumina HiSeq2000 to target 30-fold coverage. Reads were aligned and assembly to the human reference genome (NCBI GRCh37, hg19) and reference mitochondrial DNA rCRS (NC_012920) using Burrows-Wheeler Aligner. GATK and haplotype caller procedure has been used for genomic variants calling. All the found genomic variants were then further annotated by SIFT, PolyPhen2, SNPedia and ClinVar using ANNOVAR.

Results: We sequenced and analysed the whole-genomes of ethnic Kazakh individuals with the coverage (28X-32X). From 98.85 to 99.58 % base pairs were totally mapped with properly mapped 99.06 % in average. Het/Hom and Ti/Tv ratios for each whole genome ranged from 1.35 to 1.49 and from 2.07 to 2.08, respectively. We have identified the novel variants that previously not catalogued in 1000 Genomes Project, ExAC or NCBI Reference Assembly dbSNP (snp138, avsnp138, avsnp150, hg19).

Conclusion: We showed high genetic admixture of Kazakh genomes on autosomal level and similar complex heterogeneity of Central Asian populations. This whole-genome sequences information of healthy Kazakh individuals may be an important reference for biomedical studies investigating disease associations and population-wide genomic studies of ethnically diverse Central Asian region.

Acknowledgments: Work was supported by grant projects #AP05134722, #AP05135430 and #AP05136106 from the Committee Science and Ministry of Education and Science at the Republic of Kazakhstan.