DOI 10.18699/BGRSSB-2018-068

Exome-wide survey of the Siberian Caucasian population

A.A. Yurchenko[1], N.S. Yudin[1, 2], M.I. Voevoda[1, 3]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Institute of Internal and Preventive Medicine – Branch of Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
** e-mail: andreyurch@gmail.com*

Key words: exome sequencing, population structure, associations, Siberia, Russia

Motivation and Aim: Population structure is a very important factor in medical genetic association studies which can compromise modern genomic methods not being properly accounted for. In this study, we identified exome genetic variants for 39 individuals from Novosibirsk and compared them with the previously published genome-wide data and exomes from the 1000 Genomes Project to understand the extent of the population stratification and compared allele frequencies in our sample and European dataset for medically and pharmacogenetically important variants.
Methods and Algorithms: SNVs and indels were identified using GATK pipeline according to the GATK Best Practices workflow with the sensitivity filter equal to 99.9 in the 39 samples. The variants were used for the PCA with the European and previously published Russian populations, ADMIXTURE analysis with the European dataset from 1000 Genomes Project and pairwise Fst estimation. We tested medically (ClinVar) and pharmacogenetically (PharmGKB) relevant variants for the differences in allele frequencies between the populations with PLINK software.
Results: A total of 136276 SNVs and 14464 indels were identified in the target regions of the Agilent SureSelect V5. The PCA demonstrated an intermediate emplacement of the Novosibirsk population between the Finnish and other European populations confirming the full congruity of the exome samples and previously published microarray data of the Novosibirsk population and Siberian Starovers. The results of the ADMIXTURE analysis showed a higher Finnish component in the Novosibirsk population than in other European groups and clearly distinguished the Novosibirsk population from the others at K = 4. We identified a highly significant albeit low (Fst = 0.005–0.009) level of the genetic differentiation between the Novosibirsk and other European Non-Finnish (ENF) populations. Among the 452 pharmocogenetically and 210 medically important variants we found 3 and 7 variants respectively which showed significant allele frequency differences between the Novosibirsk and the ENF population after the multiple testing correction. The most significant differences in allele frequencies were attributed to such genes as FCGR3B, TYR, OCA2, FABP6 and SLC4A1.
Conclusion: The Caucasian Novosibirsk population is quite homogeneous and significantly differentiated from other European populations from 1000 Genomes Project demonstrating a higher Finnish component and genetic congruence with the previously published Russian dataset including partially isolated Siberian Starovers.
Acknowledgements: This study was supported by the grant from the Russian Science Foundation (project no. 16-15-00127).