

Expanding the Russian allele frequency reference via cross-laboratory data integration: insights from 6,096 exome samples

Yury A. Barbitoff^{1,3,4,✉}, Darya N. Khmelkova², Ekaterina A. Pomerantseva², Aleksandr V. Slepchenkov³, Nikita A. Zubashenko², Irina V. Mironova², Vladimir S. Kaimonov², Dmitrii E. Polev¹, Victoria V. Tsay^{1,5}, Andrey S. Glotov^{1,4}, Mikhail V. Aseev^{1,4}, Oleg S. Glotov^{1,4,5}, Arthur A. Isaev², and Alexander V. Predeus^{3,✉}

1 - CerbaLab Ltd., St. Petersburg, Russia

2 - Genetics and Reproductive Medicine Center GENETICO Ltd., Moscow, Russia

3 - Bioinformatics Institute, St. Petersburg, Russia

4 - Dpt. of Genomic Medicine, D.O. Ott Research Institute of Obstetrics, Gynaecology and Reproductology, St. Petersburg, Russia

5 - FGBE "Children's Scientific and Clinical Center for Infectious Diseases of the Federal Medical and Biological Agency", Saint Petersburg, Russia

Supplementary Information

Supplementary Figures

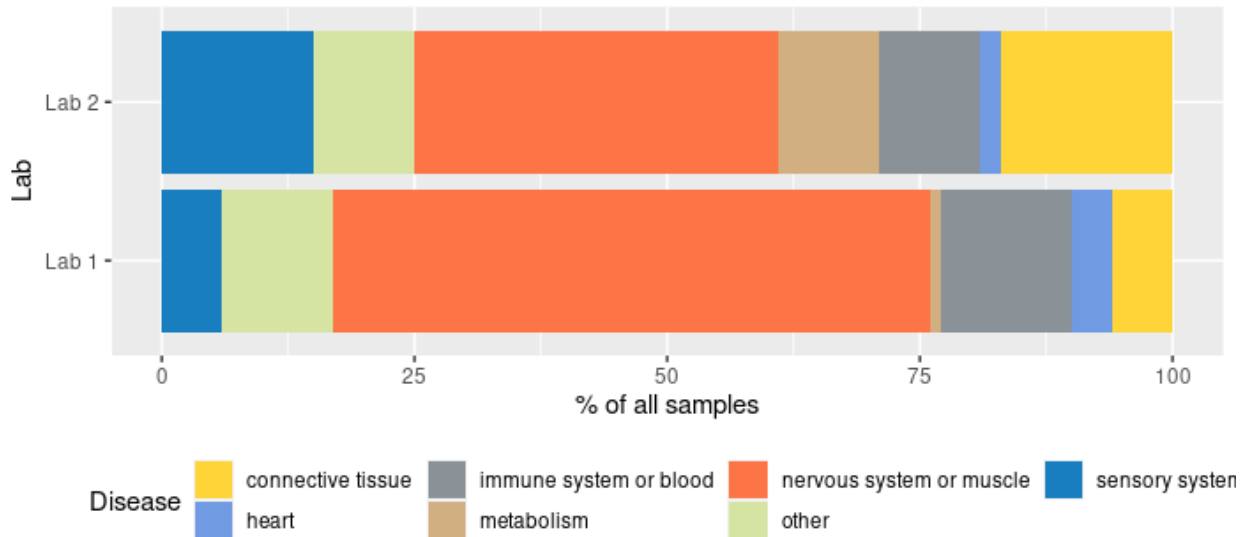


Figure S1. Approximate distribution of disease groups in the subset of diseased individuals included into the analysis.

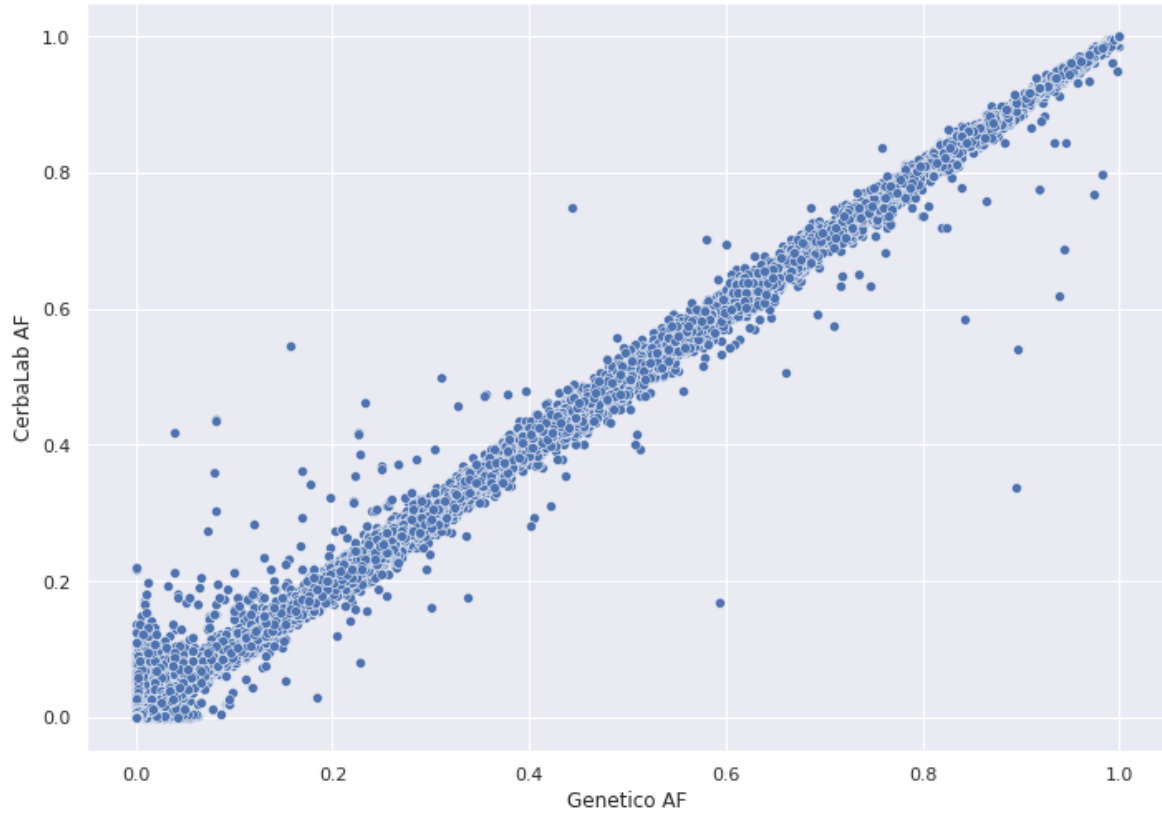


Figure S2. A scatterplot showing the correspondence between alternative allele frequencies for variants with call rate greater than 80% in samples from both participating laboratories. Pearson's r^2 for the frequencies is 0.999.

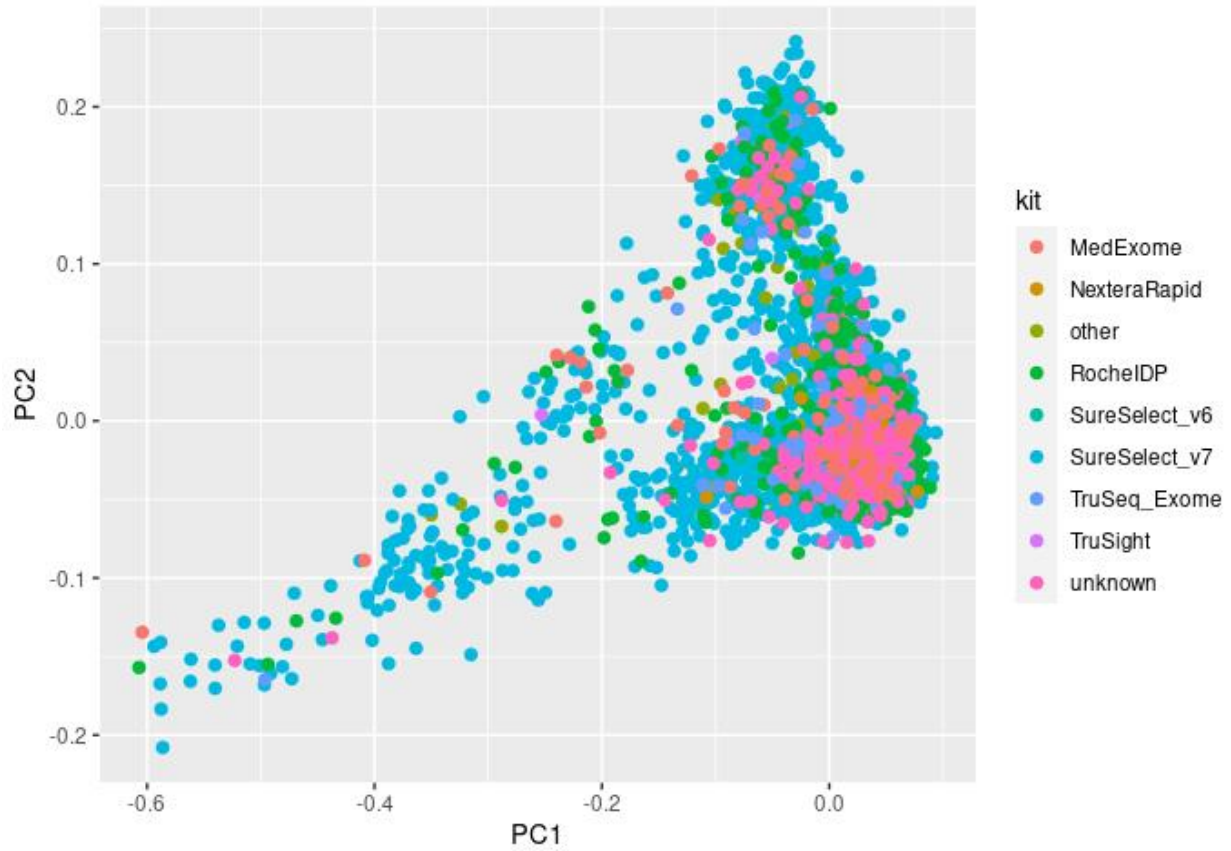


Figure S3. A scatterplot showing the results of principal component analysis of individual genotypes coloured by capture probes used for library preparation (A).

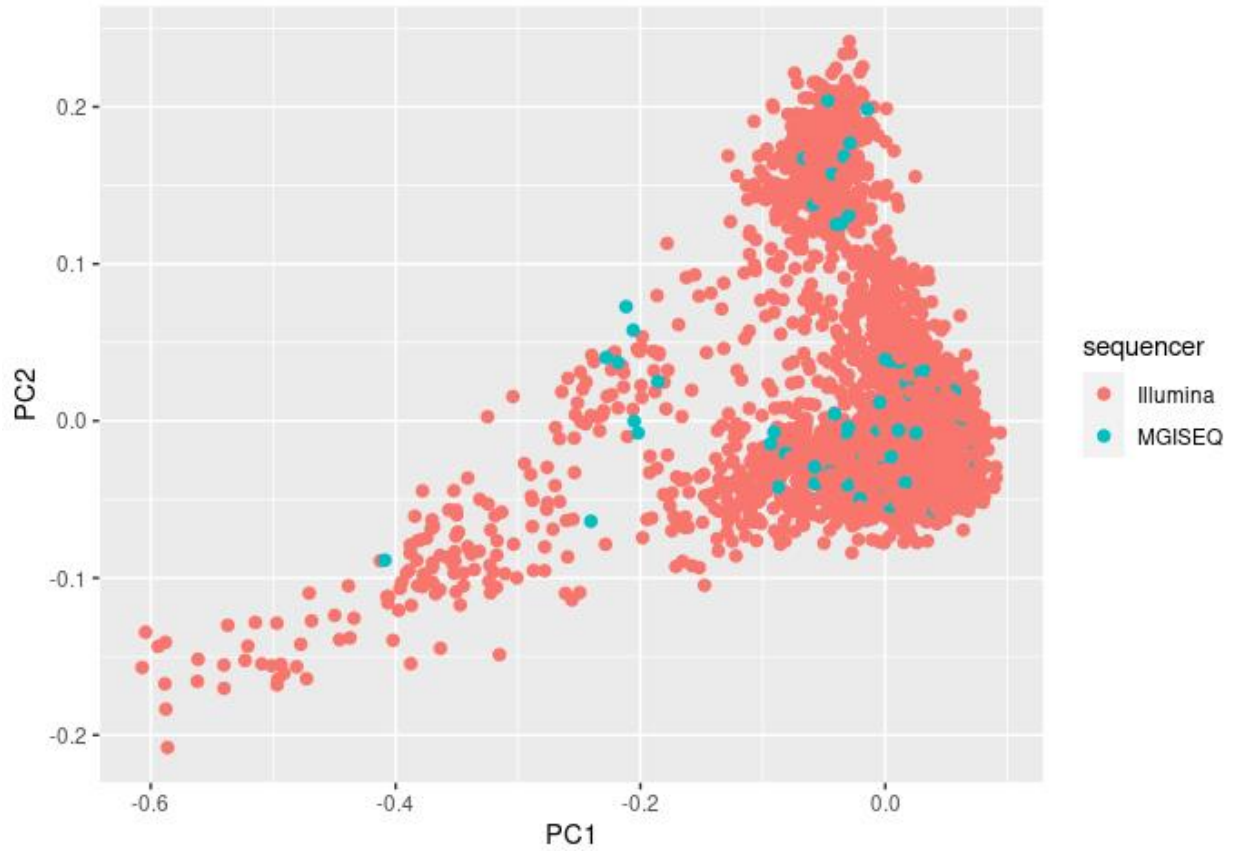


Figure S4. A scatterplot showing the results of principal component analysis of individual genotypes coloured by sequencing machine used.