Systematic dissection of biases in whole-exome and whole-genome sequencing reveals major determinants of coding sequence coverage

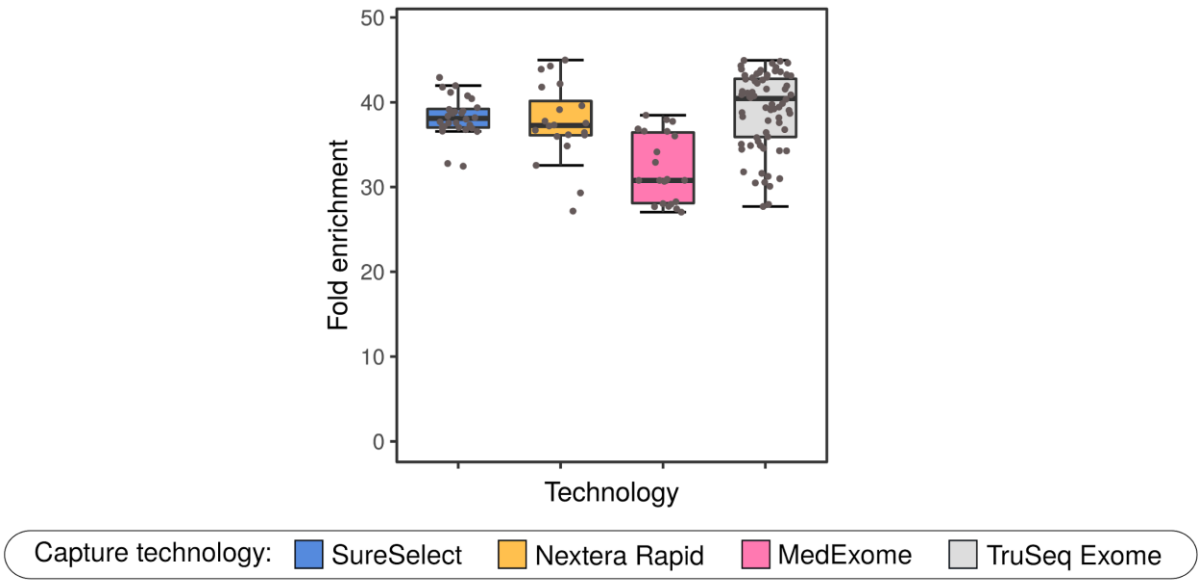
Yury A. Barbitoff[1,2,3,4], Dmitrii E. Polev[2], Irina V. Shcherbakova[2], Artem M. Kiselev[5], Andrey S. Glotov[2], Elena A. Serebryakova[2], Anna A. Kostareva[5], Oleg S. Glotov[3,6] , and Alexander V. Predeus[1]*

[1]Bioinformatics Institute, Saint Petersburg, Russia,
[2]Biobank of the Research park, Saint-Petersburg State University, Saint Petersburg, Russia,
[3]Department of Genetics and Biotechnology, Saint-Petersburg State University, Saint Petersburg, Russia,
[4]Institute of Translational Biomedicine, Saint-Petersburg State University, Saint Petersburg, Russia
[5]Almazov National Medical Research Centre, Saint Petersburg, Russia,
[6]City Hospital №40, Saint Petersburg, Russia

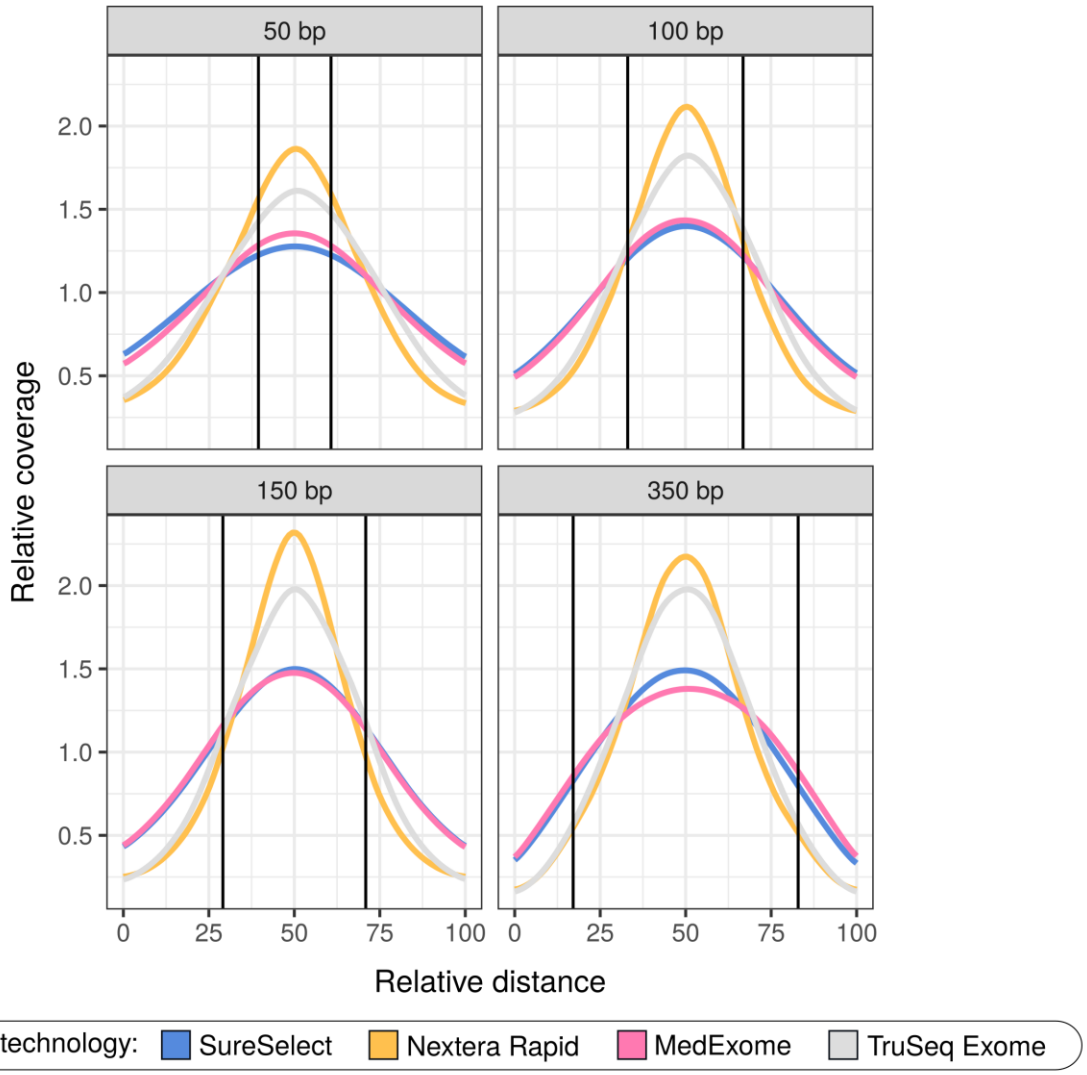
*To whom correspondence should be addressed: predeus@bioinf.me, Bioinformatics Institute, Kantermirovskaya st. 2A, St. Petersburg, 199034, Russia.

Table of contents

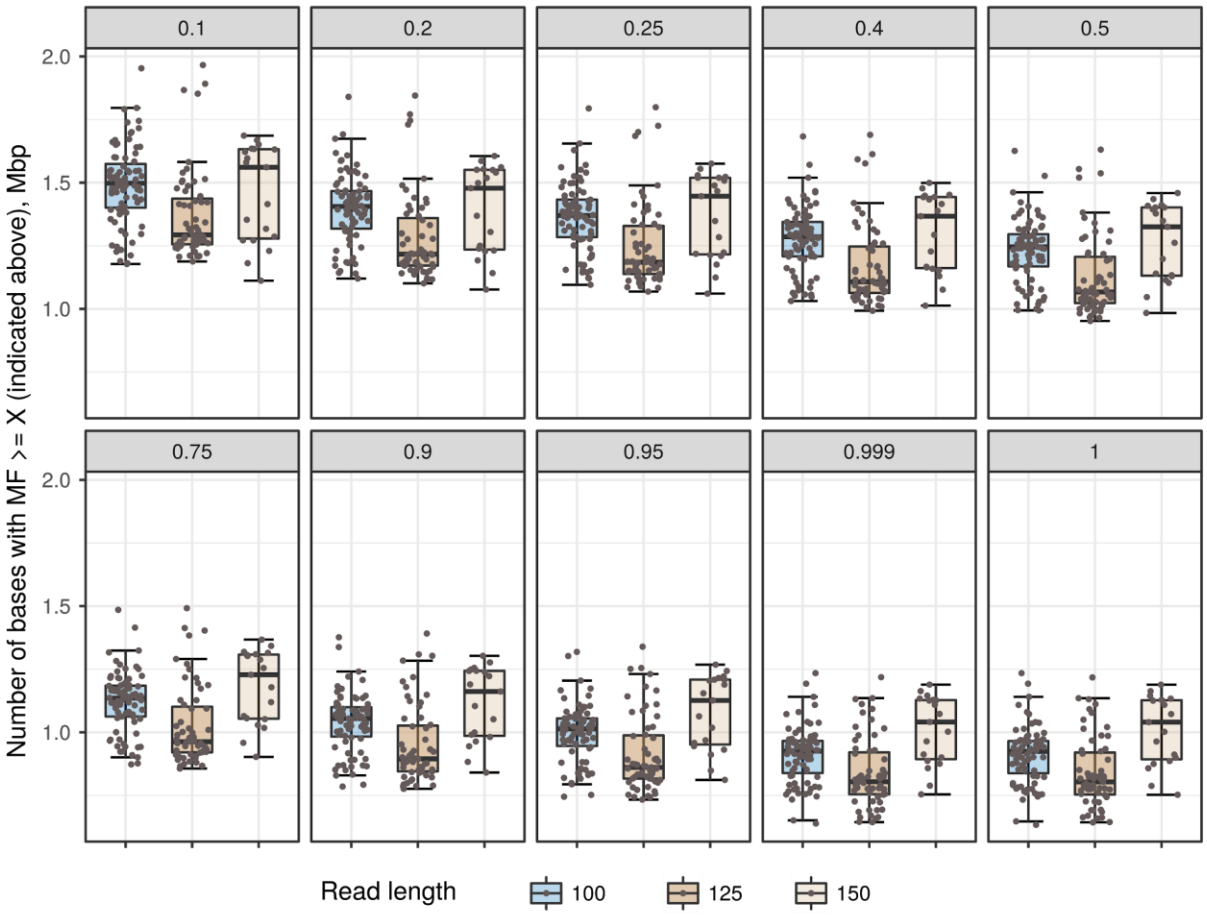
Supplementary Figures



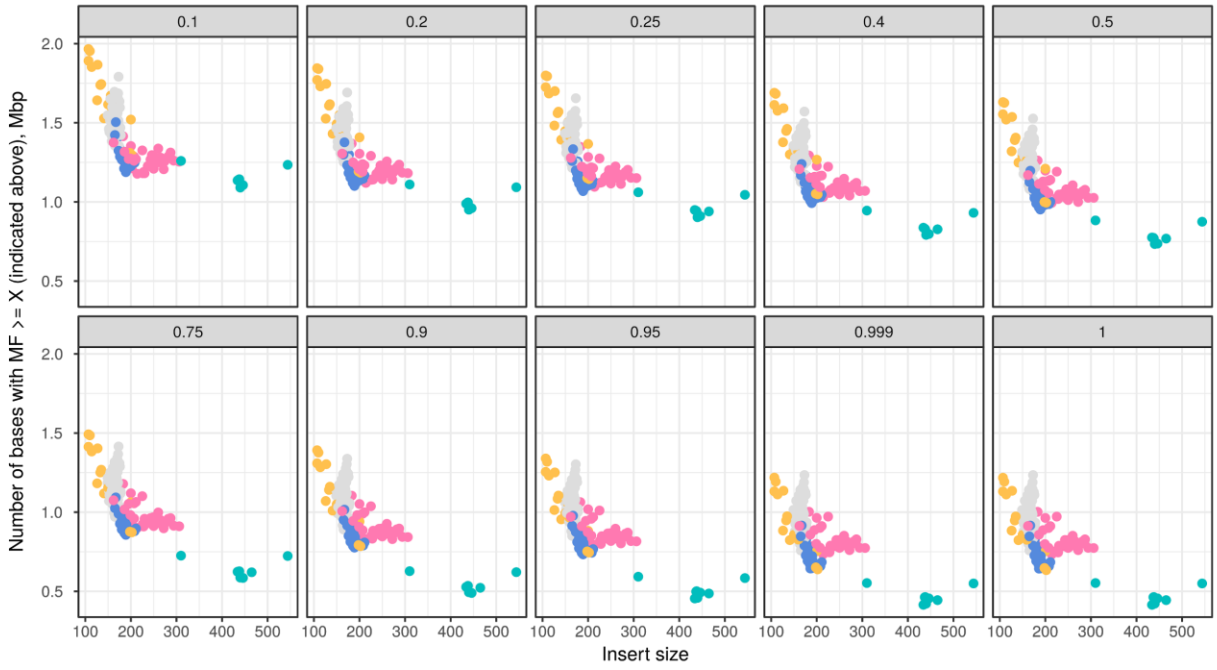
Supplementary Figure 1. Fold enrichment of CDS regions for WES samples included in the study.



Supplementary Figure 2. Profiles of relative coverage within exons divided into 4 quartiles according to the length of an interval. 100 bp of flanking bases are included; solid lines delineate CDS margins.



Supplementary Figure 3. The number of bases having an MF >= X (indicated above each plot) for WES and WGS samples having indicated paired end read lengths.



Supplementary Figure 4. The number of bases having an MF = X (indicated above each plot) for all samples plotted against the library mean insert size.