Cell Genomics, Volume *2*

Supplemental information

Decoding the diversity of killer immunoglobulin-like receptors by deep sequencing and a high-resolution imputation method

Saori Sakaue, Kazuyoshi Hosomichi, Jun Hirata, Hirofumi Nakaoka, Keiko Yamazaki, Makoto Yawata, Nobuyo Yawata, Tatsuhiko Naito, Junji Umeno, Takaaki Kawaguchi, Toshiyuki Matsui, Satoshi Motoya, Yasuo Suzuki, Hidetoshi Inoko, Atsushi Tajima, Takayuki Morisaki, Koichi Matsuda, Yoichiro Kamatani, Kazuhiko Yamamoto, Ituro Inoue, and Yukinori Okada

Supplemental Materials for

Decoding the diversity of killer immunoglobulin-like receptors by deep target sequencing and a high-resolution imputation method.

Saori Sakaue*, Kazuyoshi Hosomichi, Jun Hirata, Hirofumi Nakaoka, Keiko Yamazaki, Makoto Yawata, Tatsuhiko Naito, Nobuyo Yawata, Junji Umeno, Takaaki Kawaguchi, Toshiyuki Matsui, Satoshi Motoya, Yasuo Suzuki, Hidetoshi Inoko, Atsushi Tajima, Takayuki Morisaki, Koichi Matsuda, Yoichiro Kamatani, Kazuhiko Yamamoto, Ituro Inoue, Yukinori Okada*.

*Corresponding author:

Saori Sakaue, MD, PhD

Address: Department of Statistical Genetics, Osaka University Graduate School of Medicine, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan.

Tel: +81-6-6879-3971

E-mail: ssakaue@broadinstitute.org

Yukinori Okada, MD, PhD

Address: Department of Statistical Genetics, Osaka University Graduate School of Medicine, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan.

Tel: +81-6-6879-3971

E-mail: yokada@sg.med.osaka-u.ac.jp

Table of Contents:

Table S1-S12 are provided by a separate excel file.

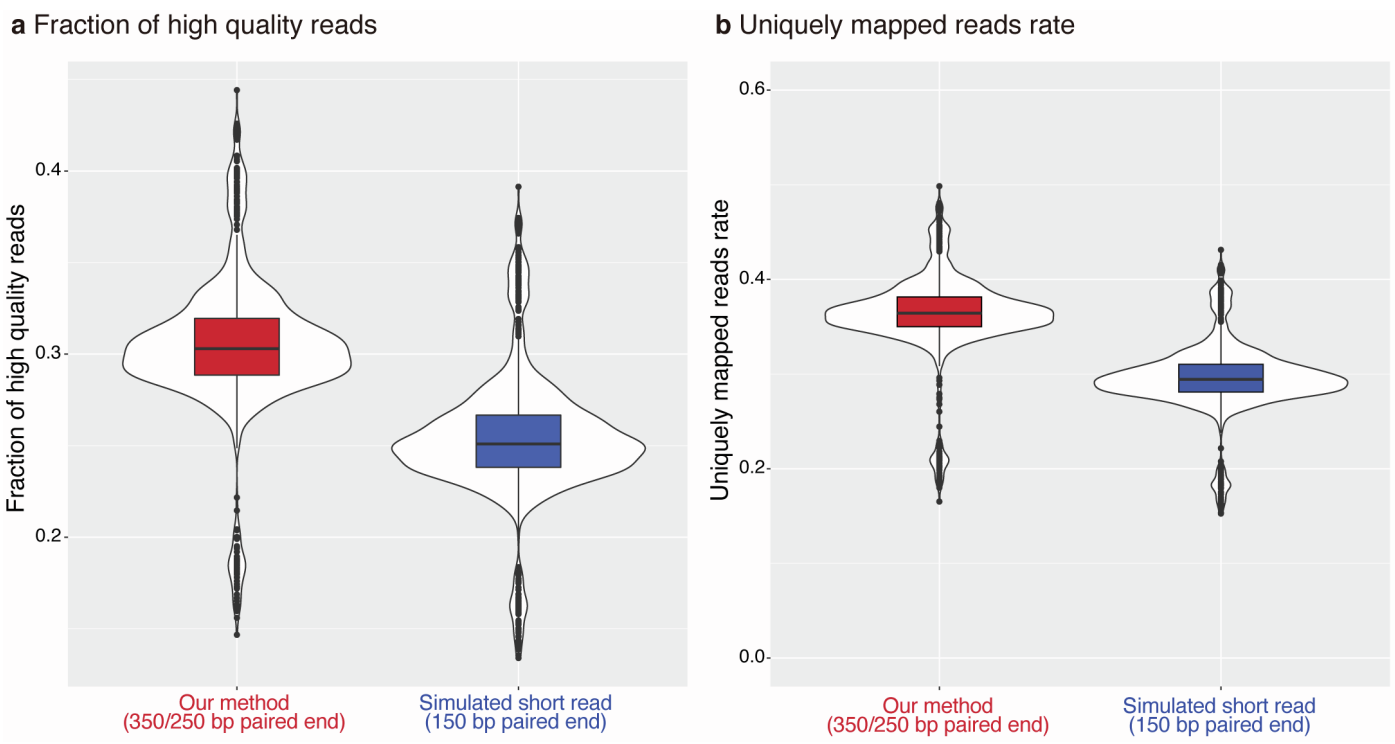

Figure S1. Comparison of mapping quality and uniquely mapped reads between 350/250 bp strategy and simulated short reads, Related to the STAR Methods.

The violin plots show the distribution of (**a**) fraction of high-quality reads and (**b**) fraction of uniquely mapped reads per individual after mapping to the reference when we use our strategy (350/250 bp paired end; red) or simulated conventional short reads (150 bp paired end; blue).

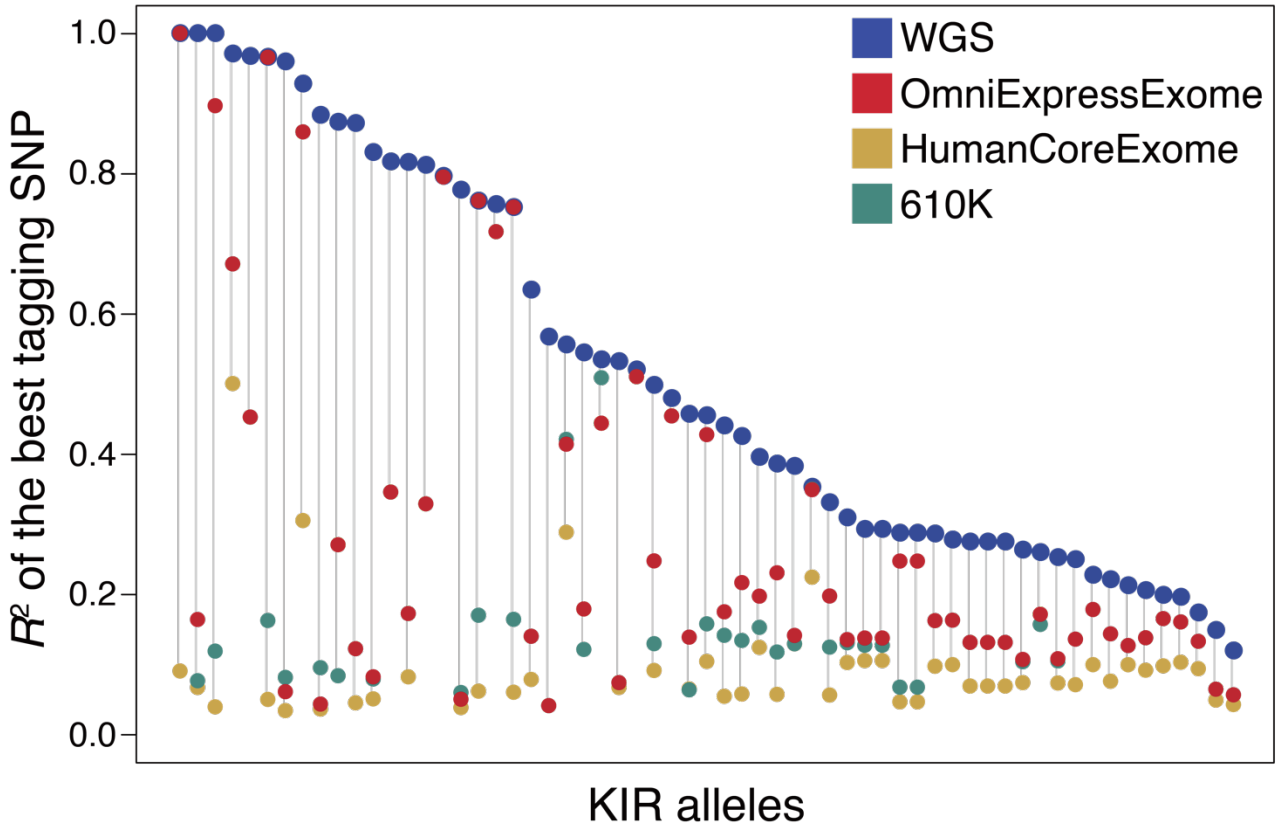

Figure S2. A comparative illustration of R^2 values of LD between KIR alleles and the best tagging SNPs, according to the genotyping arrays, Related to the STAR Methods.

The R^2 values of LD between KIR alleles and the best tagging SNPs according to the different genotyping platforms are shown. Blue plots are the R^2 values of LD between KIR alleles and the best tagging SNPs on whole genome sequencing (WGS) data, sorted by their R^2 values from higher to lower on the x axis. For each blue plot, we evaluated and plotted the R^2 values of LD between the same KIR alleles and the best tagging SNPs when we restricted the SNPs on each of the different genotyping arrays (i.e., OmniExpressExome in red, HumanCoreExome in yellow, and Illumina 610K in green), and connected the correspondence of those values from the same KIR alleles with grey lines.

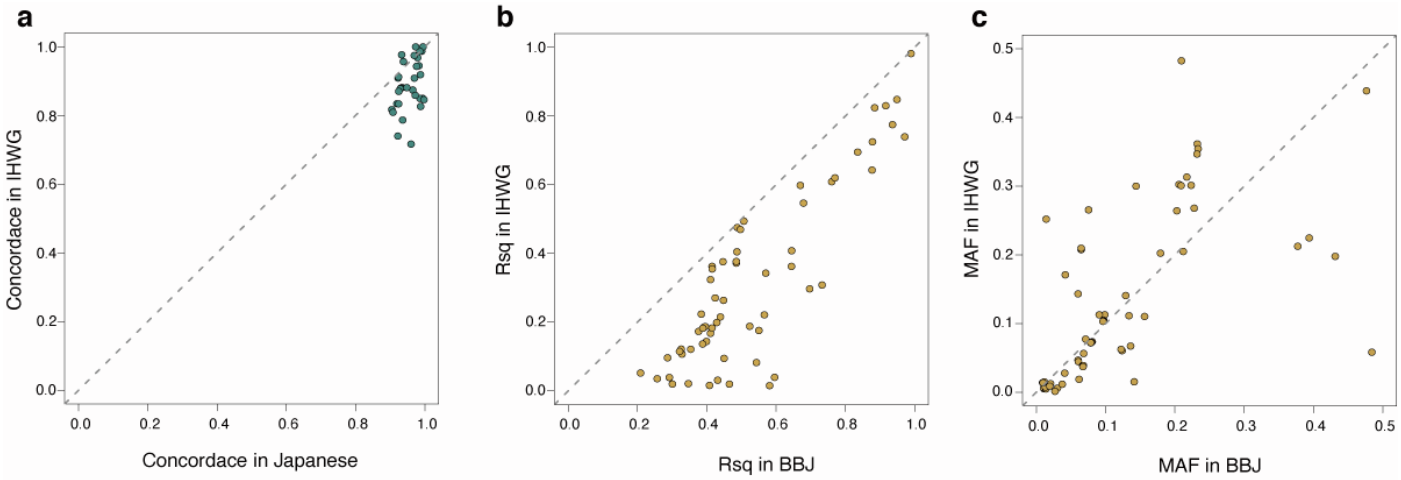

Figure S3. The imputation performance in IHWG dataset, Related to the STAR Methods.
(a) The mean concordance of imputed KIR genes and alleles with the KIR typing result. The concordance in Japanese (n_{target}=394) is shown on the x-axis, whereas the concordance in IHWG dataset (n_{target}=39) is shown on the y-axis. The dotted line indicates y = x. (b) The post-imputation Rsq of each of imputed KIR genes and alleles from minimac3 software in BioBank Japan (x-axis) and in IHWG dataset (y-axis). The dotted line indicates y = x. (c) The minor allele frequency of each of imputed KIR genes and alleles in BioBank Japan (x-axis) and in IHWG dataset (y-axis). The dotted line indicates y = x.

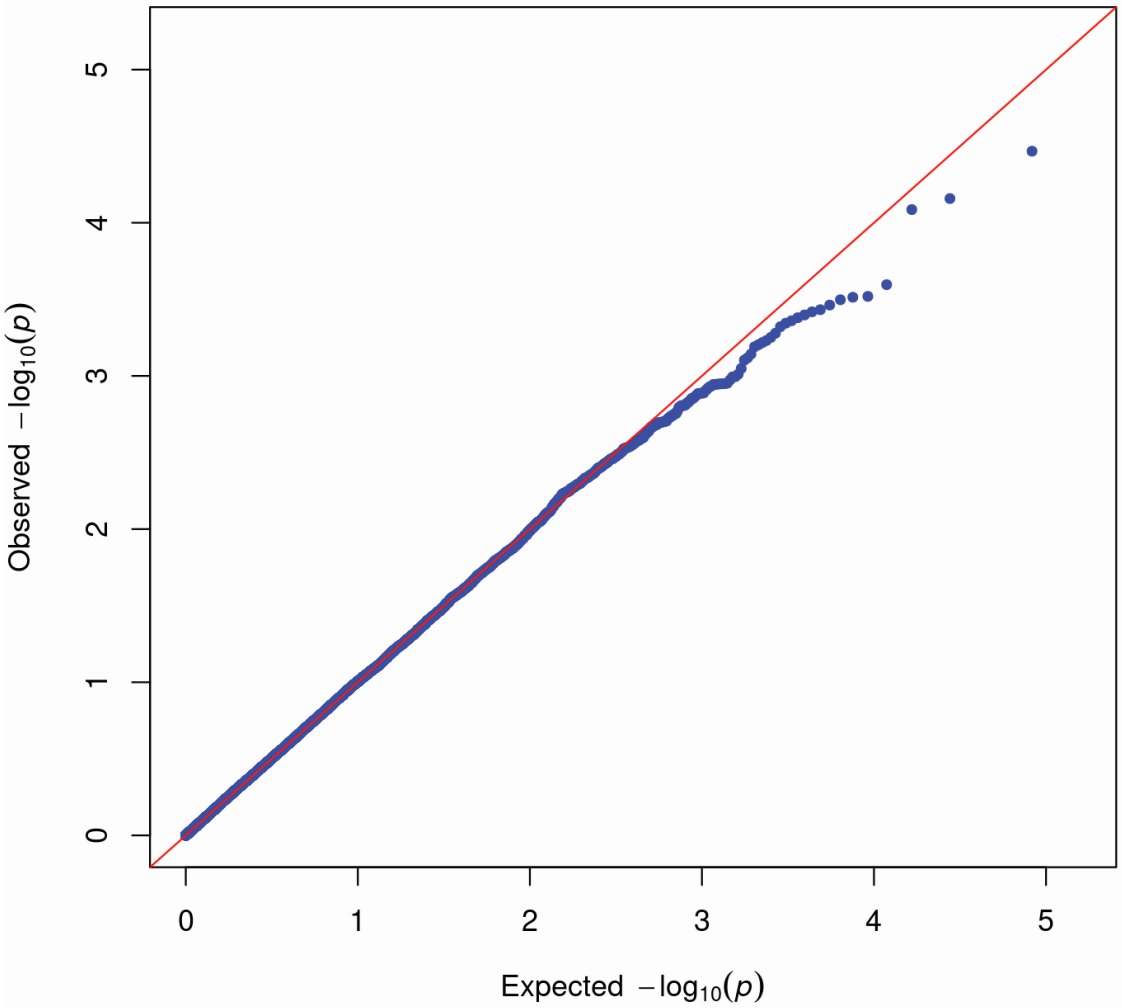

Figure S4. The quantile-quantile (QQ) plot of the interaction effect between KIR alleles and HLA class I alleles (4-digit) across 85 complex traits, Related to Figure 4.

Shown is the Quantile-quantile (QQ) plot for the *P* values in the $-\log_{10}$ scale for the interaction effects between KIR alleles and HLA alleles across 85 traits. The plot displays the relationship between the observed *P* values (vertical axis) to the expected *P* values of a null distribution (horizontal axis). The interaction effects between KIR alleles and HLA class I alleles are plotted in blue, and those between KIR alleles and other HLA alleles are plotted in red.

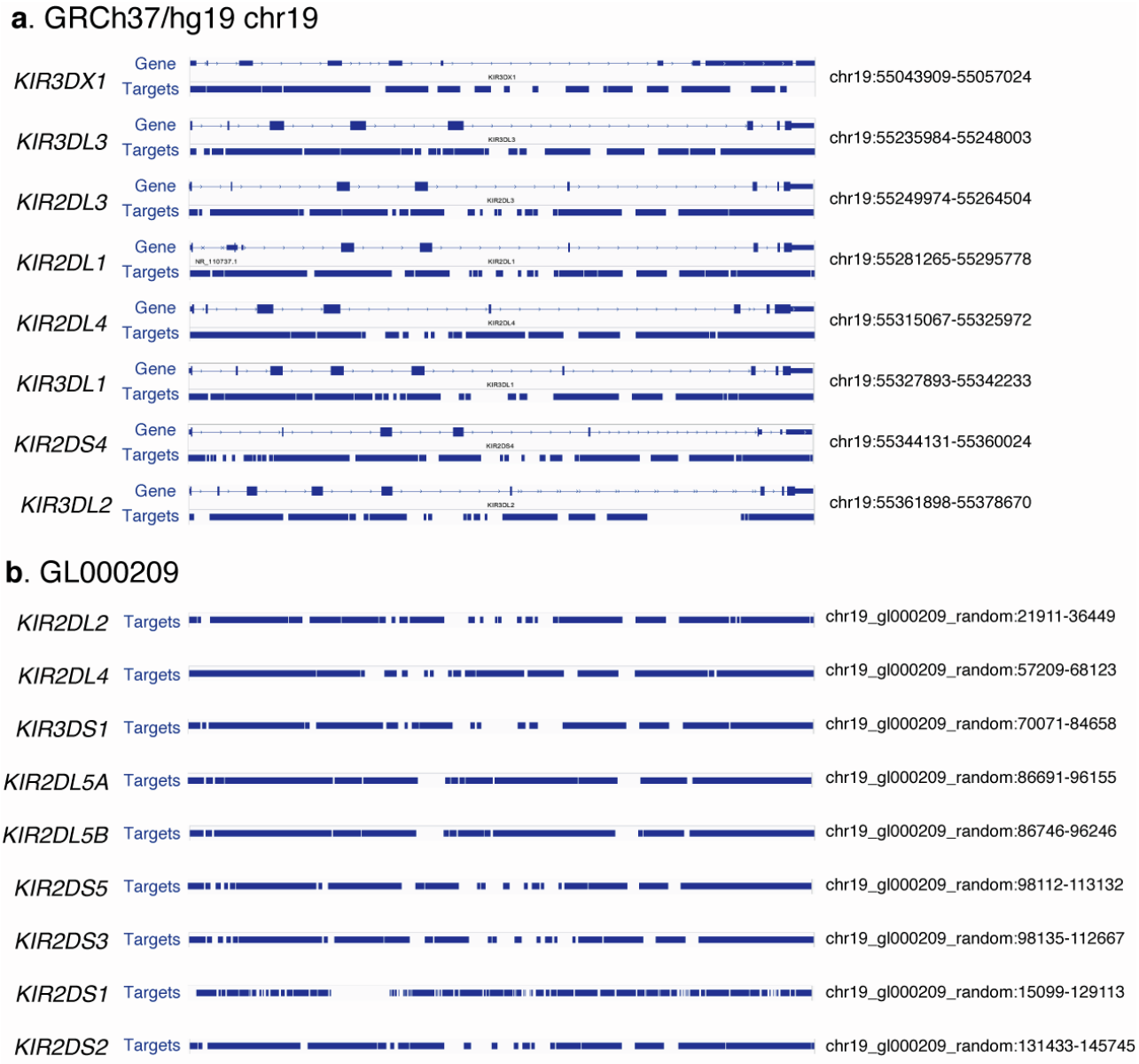

Figure S5. The schematic overview of target regions for the KIR capture method, Related to the STAR Methods.

Each row shows the exonic regions (top) and the covered regions of our KIR capture targets (bottom) in each of the KIR genes, using the IGV viewer.

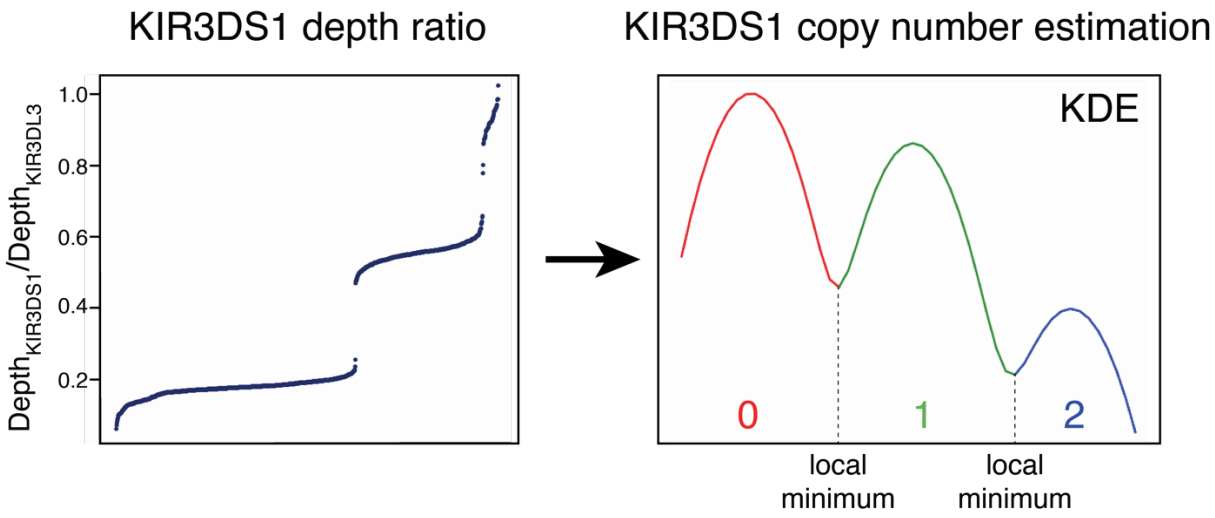

Figure S6. An example illustration of gene copy number assignment using kernel density estimation of read depth distributions, Related to the STAR Methods.

To illustrate an example of copy number assignment, *KIR3DS1* gene is picked up here. The ratio of reads mapping to a *KIR3DS1* to those mapping to *KIR3DL3* are shown and sorted by the ratio in the left panel. To assign discrete number of *KIR3DS1* copy number to an individual from these ratios, we computed the kernel-density estimates (KDE) of the ratio distribution (right panel). There were two local minimums, and thus we assigned 0, 1, 2 copies based on the relative position on the KDE.

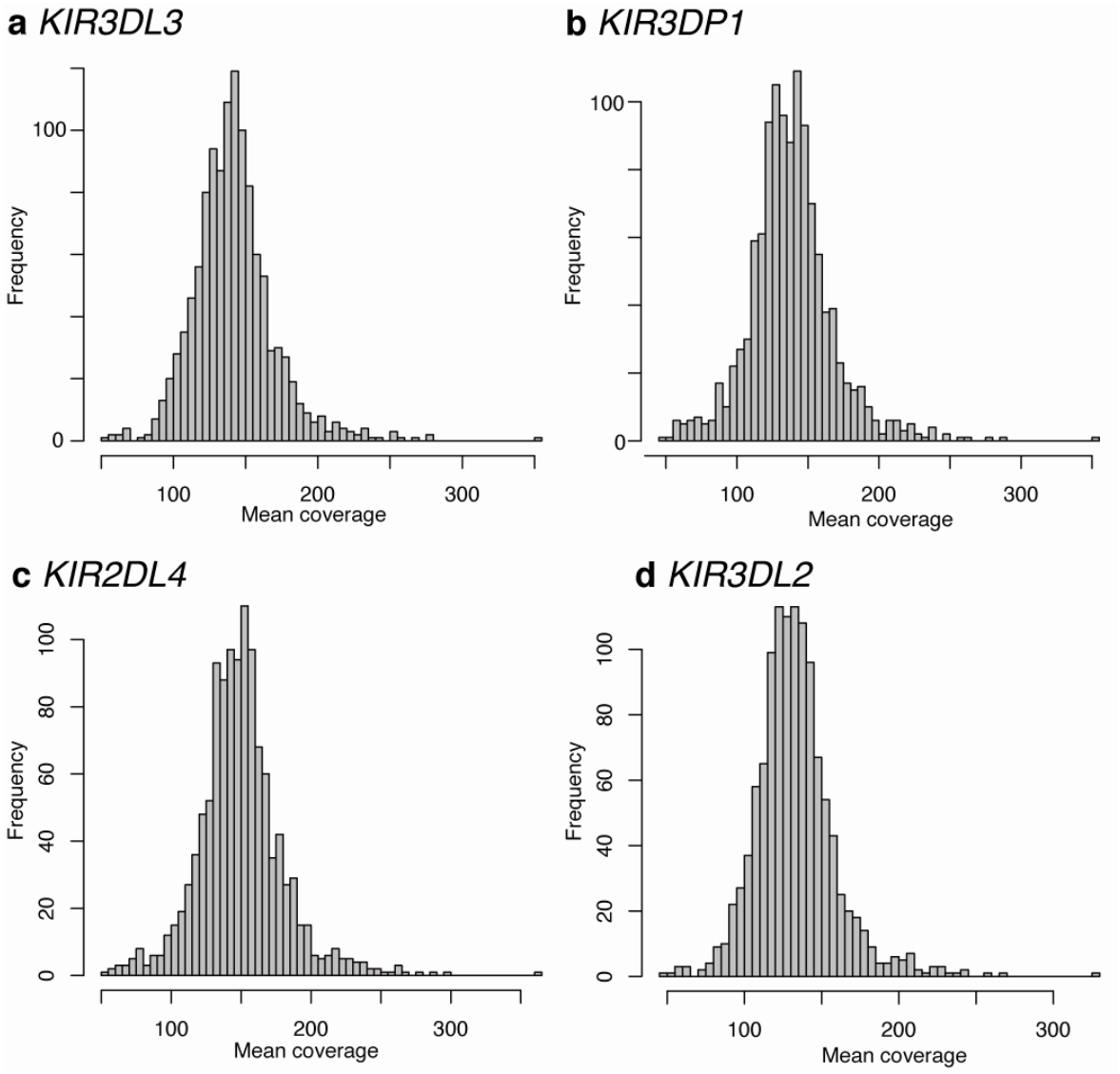

Figure S7. The histogram of the coverage in each framework gene across samples, Related to the STAR Methods.

Shown are the histograms of the mean coverage of each of the four framework genes across all samples.

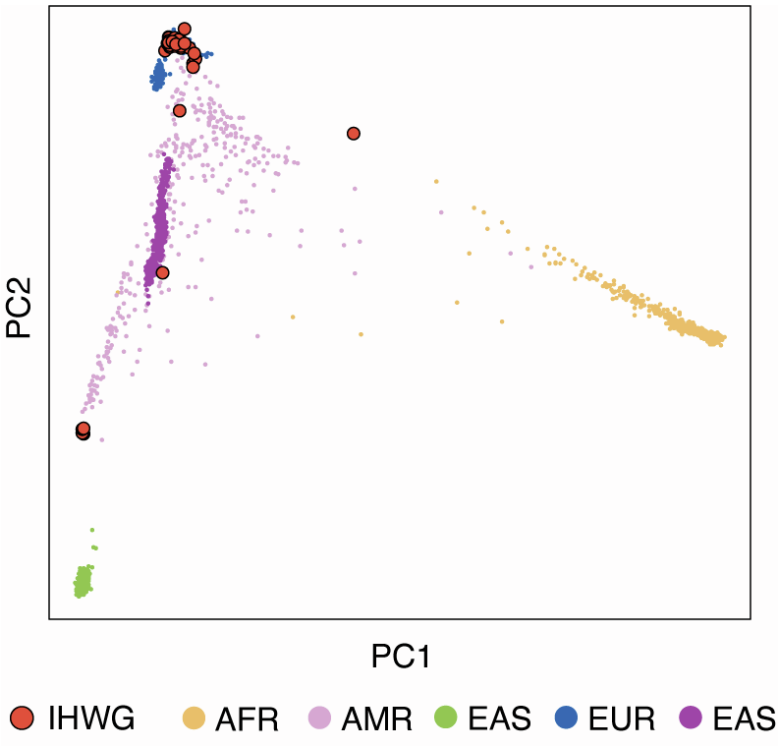

Figure S8. Principal component analysis of the IHWG genotyped individuals with 1000 Genomes Project individuals, Related to the STAR Methods.

The first two principal components of the IHWG individuals (in orange) are plotted anchored to the 1000 Genomes Project project individuals of four ancestry. Individuals are colored based on the legend.

AFR; African. AMR; Admixed American. EAS; East Asian. EUR; Europoan. SAS; South Asian.