# Supporting Information

## Lessard et al. 10.1073/pnas.1714640114

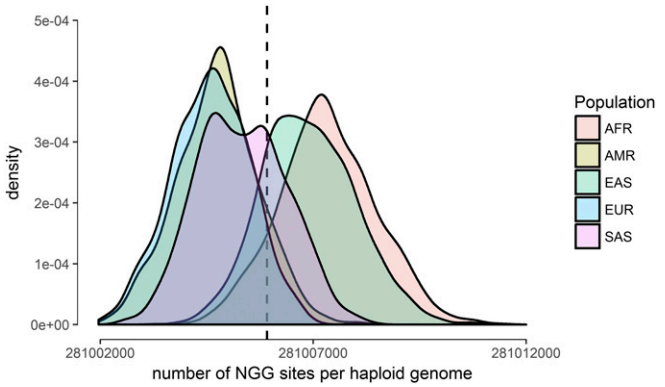


**Fig. S1.** Number of NGG PAMs within and across populations. The number of NGG PAMs in a population was determined by identifying the number of PAMs per haploid genome within the 1000G dataset after inserting the alternative allele at sites carried by each sample. The vertical line shows the number of sites in the reference genome ($n$ = 281,005,914).

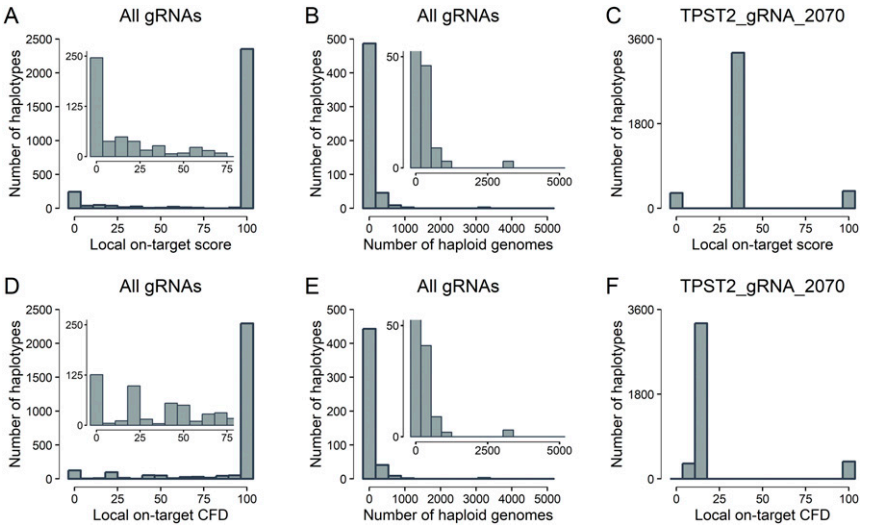


**Fig. S2.** Effect of variants on gRNA efficiency in the French Canadian dataset. (*A*) Distribution of on-target scores for human-genome–targeting gRNAs for each possible target haplotype. (*B*) Distribution of samples/individuals carrying haplotypes predicted to be targeted with a score of <100%. (*C*) Distribution of scores for the gRNA *TPST2*_gRNA_2070. (*D*) Distribution of on-target CFDs for human-genome–targeting gRNAs for each possible target haplotype. (*E*) Distribution of samples/individuals carrying haplotypes predicted to be targeted with a CFD of <100%. (*F*) Distribution of CFDs for the gRNA *TPST2*_gRNA_2070. *Inset* plots with a restricted *y*-axis range are shown for *A*, *B*, *D*, and *E* for easier visualization of data.

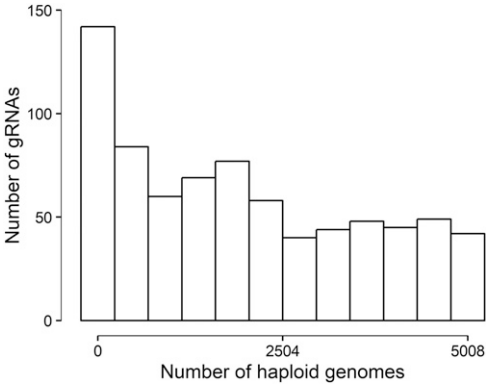


**Fig. S3.** Distribution of the number of haploid genomes from 1000G for each gRNA with an aggregate off-target score of >0. The histogram represents the distribution of gRNAs based on the number of haploid genomes with a Δ aggregate off-target score of >0. For example, from the total number of gRNAs (Fig. 3*C*), there are ∼140 gRNAs for which no haploid genomes have a Δ aggregate off-target score >0 and there are ∼40 gRNAs for which all 5,008 haploid genomes have a Δ aggregate off-target score >0.

**Dataset S1. Median autosomal variant sites from the 1000 Genomes Project database. Reprinted with permission from ref. 20**

Dataset S1

Acronyms used: AFR, Africa; AMR, the Americas; CNVs, copy number variants; EAS, East Asia; EUR, Europe; GWAS, genome-wide association studies; HGDM-DM, Human Gene Mutation Database disease mutations; indels, insertions/deletions; k, thousand; LoF, loss of function; M, million; MEI, mobile insertion elements; SAS, South Asia; SNP, single-nucleotide polymorphism; TFBS, transcription factor binding sites; UTR, untranslated region.

**Dataset S2. List of human-genome–targeting, viral-genome–targeting, and nontargeting gRNAs as well as their associated on- and off-target scores from the reference genome**

Dataset S2

Off-target mutagenesis predictions include off target score (range, 0–100), perfect matches, one mismatch, two mismatches, three mismatches, four mismatches, and total mismatches (sum of perfect matches, one, two, three, and four mismatches). On-target efficiency predictions include Fusi score (range, 0–100), Chari score (range, 0–100), Xu score (range, −2 to 2), Doench score (range, 0–100), Mor.-Mateos score (range, 0–100), Wang score (range, 0–100), out-of-frame score (range, 0–100), and Housden score (range, 0–10).

**Dataset S3. List of human-genome–targeting, viral-genome–targeting, and nontargeting gRNAs and their associated off-target scores in ambiguous genomes derived from the 1000 Genomes Project, French Canadian, and the gnomAD datasets**

Dataset S3

CFD, cutting frequency determination; MMs, mismatches.

**Dataset S4. List of on-target sites with variants in the 1000 Genomes Project and their local scores**

Dataset S4

The haplotype column shows whether each variant corresponds to the reference (0) or alternative (≥1) allele. CFD, cutting frequency determination; indel, insertion/deletion; MMs, mismatches; PAM, protospacer adjacent motif.

**Dataset S5. List of on-target sites with variants in the French Canadian dataset and their local scores**

Dataset S5

The haplotype column shows whether each variant corresponds to the reference (0) or alternative (≥1) allele. CFD, cutting frequency determination; indel, insertion/deletion; MMs, mismatches; PAM, protospacer adjacent motif.

**Dataset S6. List of off-target sites with variants in the 1000 Genomes Project and their local scores**

Dataset S6

The haplotype column shows whether each variant corresponds to the reference (0) or alternative (≥1) allele. CFD, cutting frequency determination; indel, insertion/deletion; MMs, mismatches; PAM, protospacer adjacent motif.

**Dataset S7. List of off-target sites with variants in the French Canadian dataset and their local scores**

Dataset S7

The haplotype column shows whether each variant corresponds to the reference (0) or alternative (≥1) allele. CFD, cutting frequency determination; indel, insertion/deletion; MMs, mismatches; PAM, protospacer adjacent motif.

**Dataset S8. Number of gRNAs with new two to four mismatch sites using the 1000 Genomes Project and French Canadian datasets**

Dataset S8

**Dataset S9. Aggregate off-target score statistics by population from the 1000 Genomes Project**

Dataset S9

The last columns show different percentile levels (%).