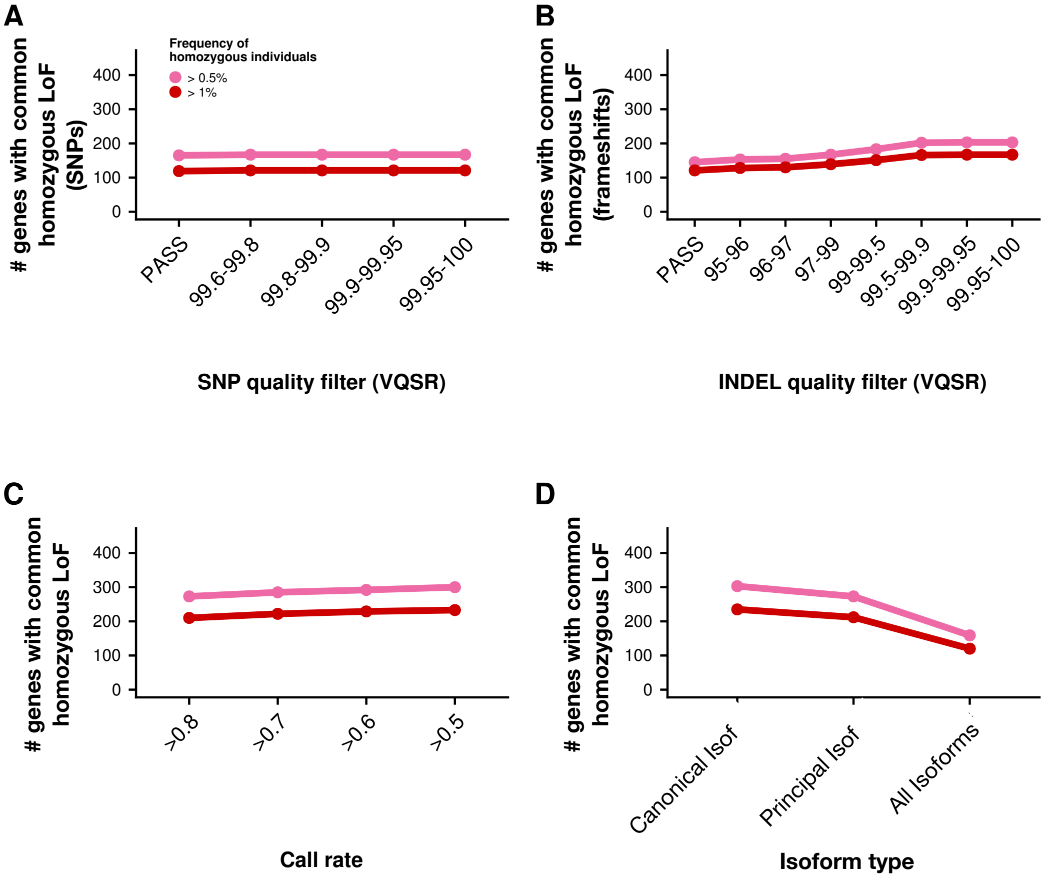
Rausell, Antonio, Luo, Yufei, Lopez, Marie, Seeleuthner, Yoann, Rapaport, Franck, Favier, Antoine, Stenson, Peter D., Cooper, David N. , Patin, Etienne, Casanova, Jean-Laurent, Quintana-Murci, Lluis and Abel, Laurent 2020. Common homozygosity for predicted loss-of-function variants reveals both redundant and advantageous effects of dispensable human genes. Proceedings of the National Academy of Sciences 117 (24) , pp. 13626-13636. 10.1073/pnas.1917993117

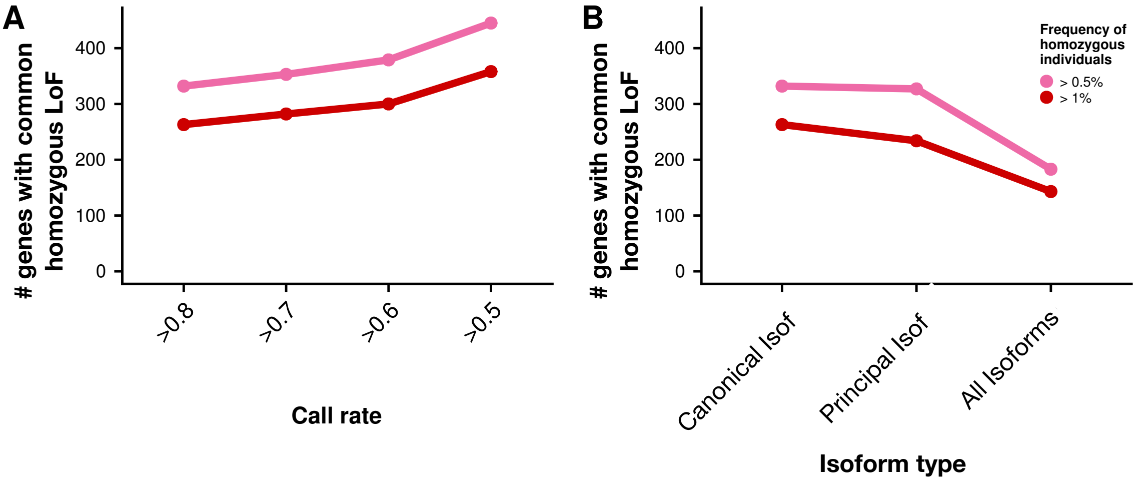
Publishers page: http://dx.doi.org/10.1073/pnas.1917993117

Supplementary Figures



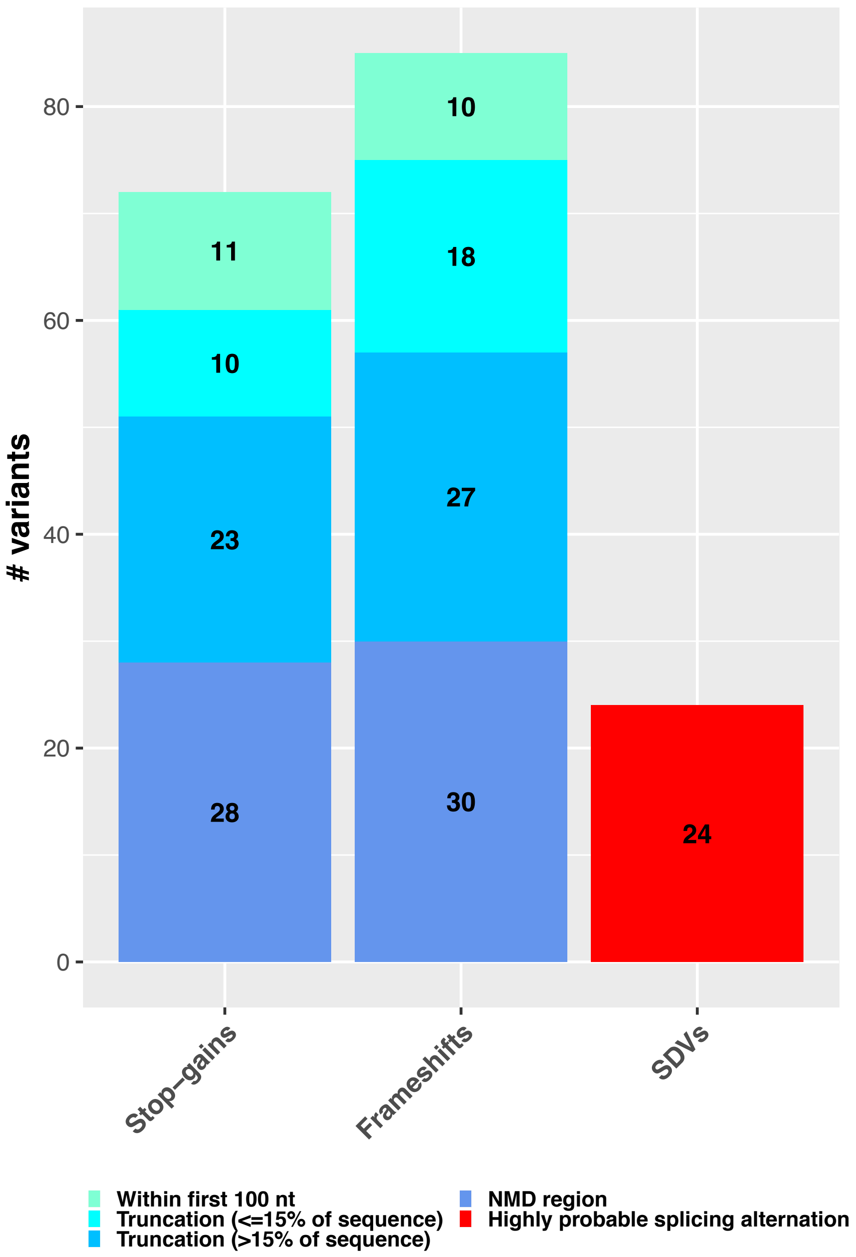
Supplementary Figure 1. Impact of the different filtering criteria on the final number of dispensable genes detected in the ExAC database.

A. Number of genes with common homozygous LoF caused by Single Nucleotide Polymorphisms as a function of the variant quality score recalibration (VQSR) threshold. **B.** Number of genes with homozygous LoF caused by frameshifts as a function of the variant quality score recalibration (VQSR) threshold. **C.** Number of genes with common homozygous LoF caused by SNPs and frameshifts as a function of the call rate threshold. **D**. Number of genes with common homozygous LoF caused by SNPs and frameshifts depending on whether LoF variant affects (i) the canonical isoform of a gene (as defined by Ensembl pipeline), (ii) the canonical isoform of a gene that represents the principal isoform of a gene, as defined by APPRIS system (corresponding to the selected criteria in our study), and (iii) all isoforms of a gene (the LoF variant is constitutive of all alternative transcripts; **Methods**).



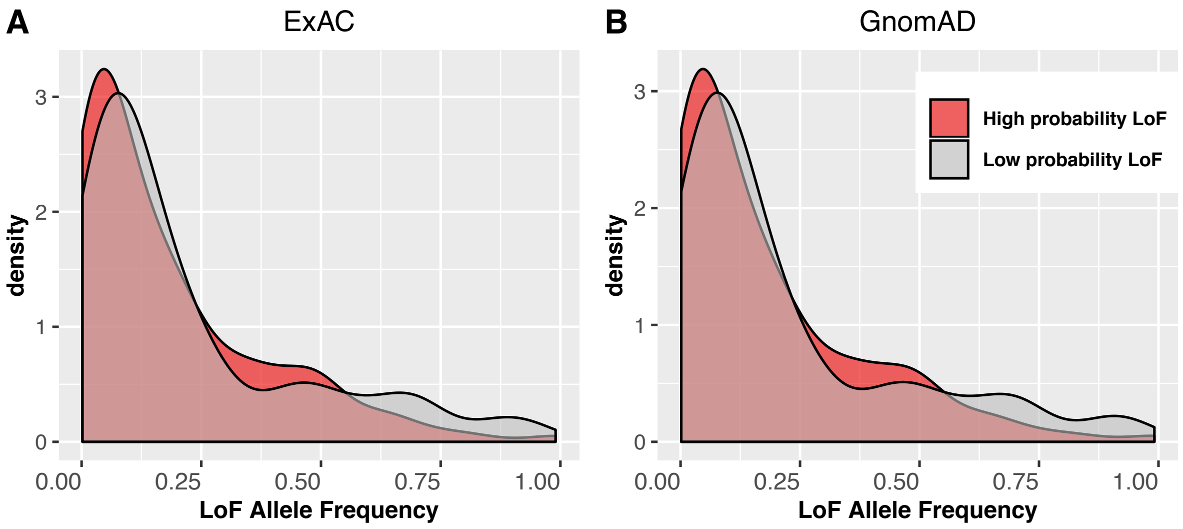
Supplementary Figure 2. Impact of the different filtering criteria adopted on the final number of dispensable genes detected in the GnomAD database.

A. Number of genes with common homozygous LoF caused by SNPs and frameshifts as a function of the call rate threshold. **B**. Number of genes with common homozygous LoF caused by SNPs and frameshifts depending on whether the LoF variant affects the canonical isoform, the principal isoform or all isoforms (variant constitutive of all isoforms). It should be noted that VQSR scores were not used in the GnomAD database, thus panels analogous to **Supplementary Figure 1 A** and **B** could not be drawn.



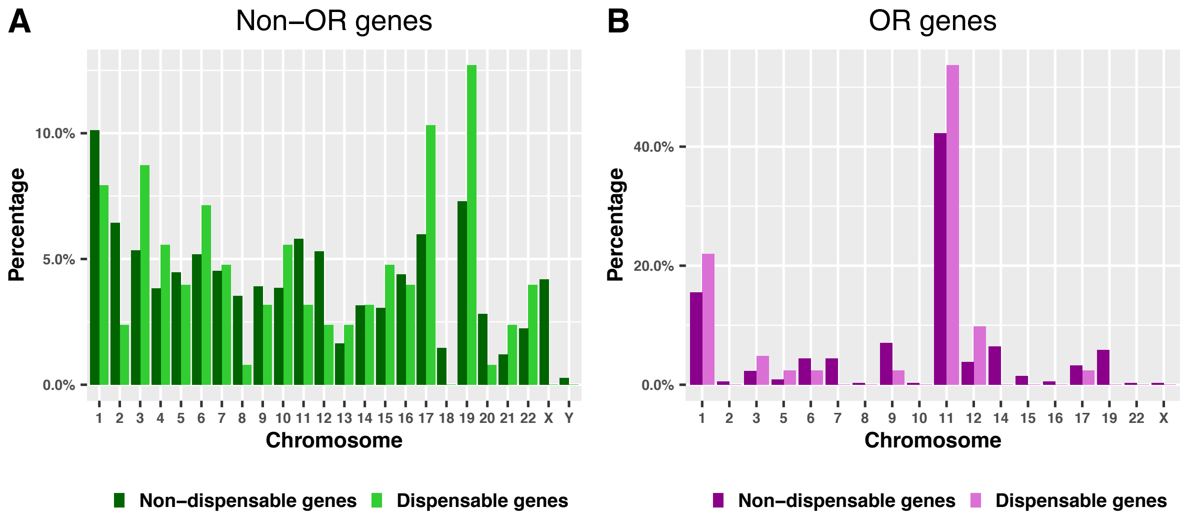
Supplementary Figure 3. Predicted functional impact of LoF variants defining the set of dispensable protein coding genes.

Bar plots show the distribution of LoF variants that define the set of dispensable genes according to their molecular consequences (stop-gains, frameshifts and splice-disrupting variants, SDV) and the predicted type of functional impact, according to the following categories: In the case of stop-gains and frameshifts, LoF variants are classified among those i) mapping to the first 100 nucleotides of the associated transcript, ii) potentially triggering NMD, or iii) truncating more, or less or equal than 15% of the affected protein sequence. In the case of putative splice-disrupting variants (SDVs), severity was computationally predicted as unknown, very low, intermediate or high impact (**Methods**).



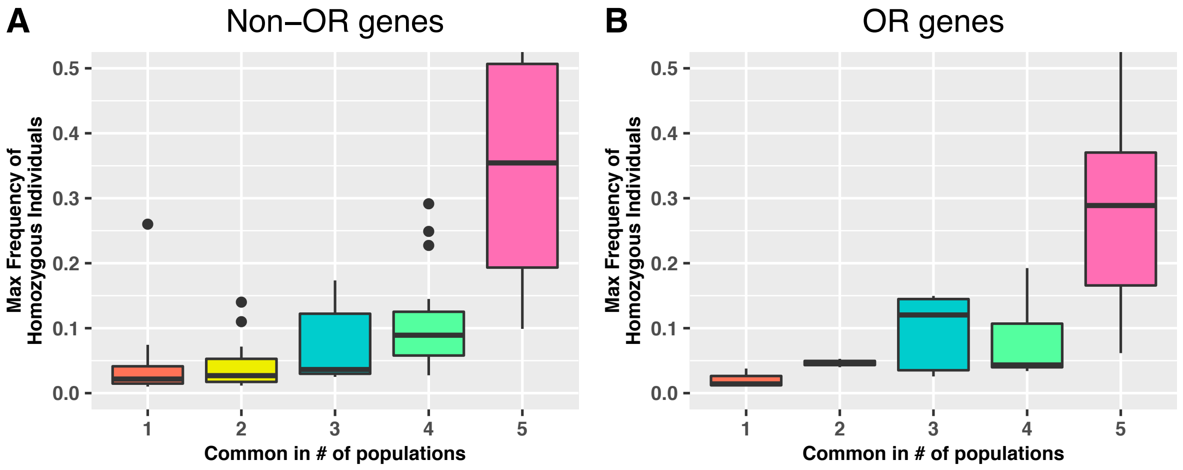
Supplementary Figure 4. Allele Frequency distribution of LoF variants defining the set of dispensable protein-coding genes.

Allele Frequency of LoF variants is represented separately for low probability LoF (light grey) and high probably LoF (dark red) variants. Allele frequencies from ExAC were used in **Panel A**, whereas GnomAD data were used in **Panel B**.

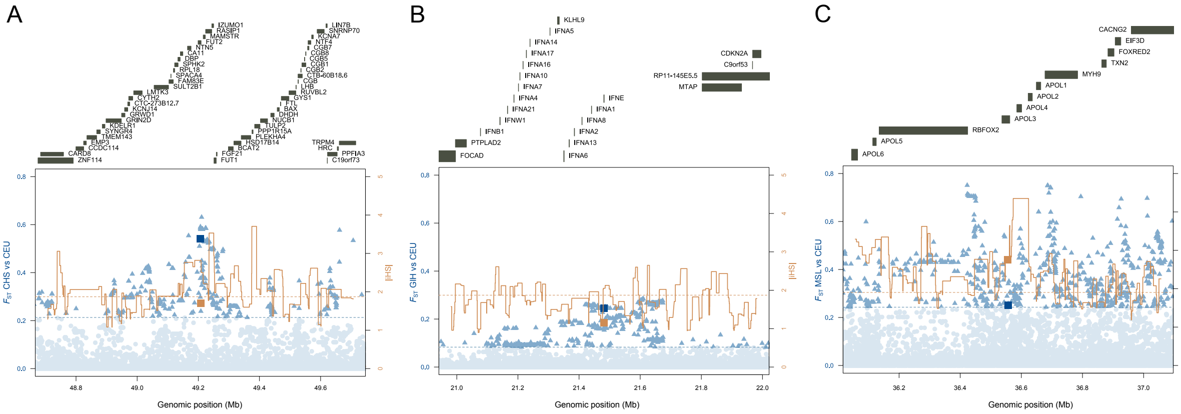


Supplementary Figure 5. Distribution of dispensable and non-dispensable genes across chromosomes.

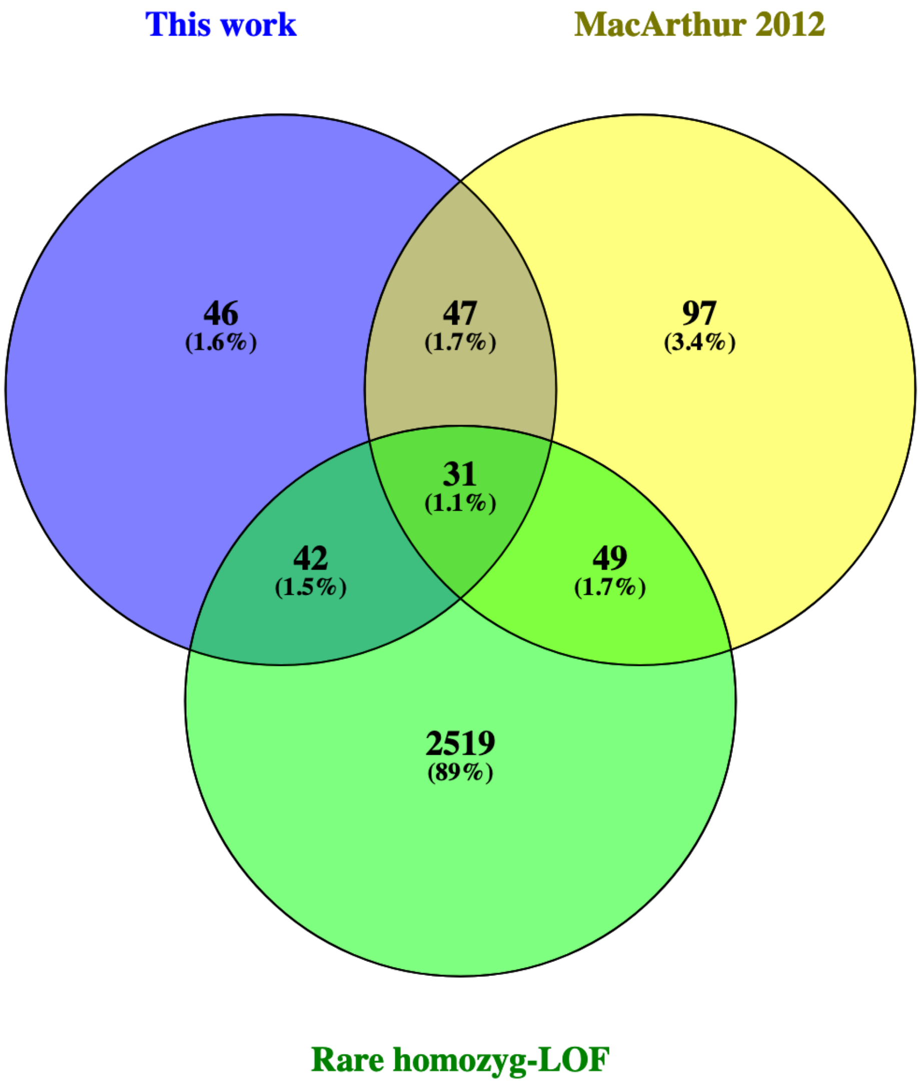
Barplots display the percentage of genes across human chromosomes of the following 4 gene sets, each adding to 100%: (**A**) dispensable non-OR genes (light green) and non-dispensable non-OR genes (dark green), and (**B**) dispensable OR genes (light purple) and non-dispensable OR genes (dark purple).



Supplementary Figure 6. Distribution of the maximum frequency of homozygous individuals for dispensable non-OR genes (**A**) and dispensable OR genes (**B**) as a function of the number of populations in which they were found to be dispensable. As in Figure 3, the homozygous LoF variant frequencies were taken from the GnomAD dataset.



Supplementary Figure 7. Selective sweep signals. Local genomic signatures of positive selection in 1Mb regions around LoF mutations located in (**A**) *FUT2* (chr19:49206674) for CEU, (**B**) *IFNE* (chr9:21481483) for GIH and (**C**) *APOL3* (chr22:36556768) for MSL. Blue and orange squares indicate F_{ST} and |iHS| values respectively at the LoF allele. Blue dots and triangles indicate SNP F_{ST} percentiles and the blue dashed line indicate 95th percentile of F_{ST} values genome-wide. Orange solid line indicate the maximum |iHS| value in sliding windows of 50 SNPs and the orange dashed line indicate the 95th percentile of |iHS| values genome-wide.



Supplementary Figure 8. Overlap of the dispensable genes detected in this work with those identified in previous studies. The figures show the Venn diagrams representing the overlap of the 166 putatively dispensable genes detected in this work with 253 genes apparently tolerant to homozygous rare LoF variants reported in MacArthur et al. (9) and a total list of 2641 presenting homozygous rare LoF variants reported from bottlenecked or consanguineous populations (10,11,13,14).