The American Journal of Human Genetics, Volume *107*

Supplemental Data

Natural Selection Shapes Codon Usage in the Human Genome

Ryan S. Dhindsa, Brett R. Copeland, Anthony M. Mustoe, and David B. Goldstein

Supplementary data

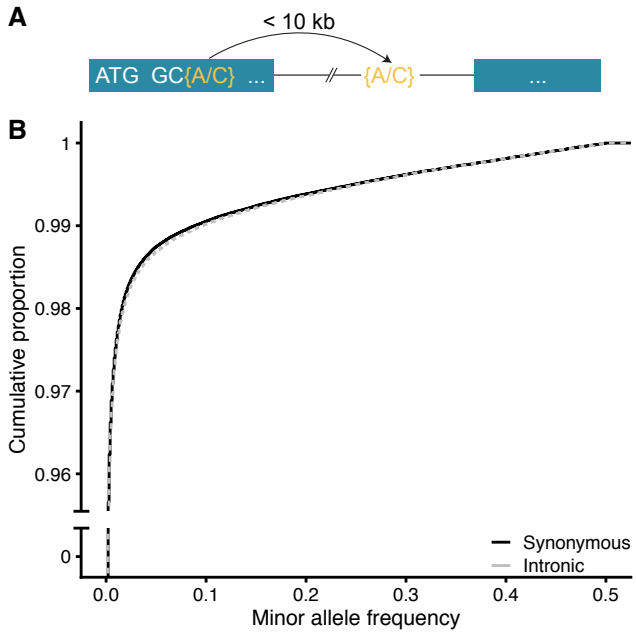

Figure S1. Demonstration of variant matching scheme and baseline SFS. (A) Illustration of our variant matching scheme for the SFS analyses. Each observed synonymous variant was matched to an observed intronic variant within 10kb with the same reference and alternate allele. We excluded all variants occurring in the first and last codon of an exon and intronic variants within 10 basepairs of splice junctions. **(B)** Site frequency spectrum of synonymous and intronic variants without accounting for codon bias.

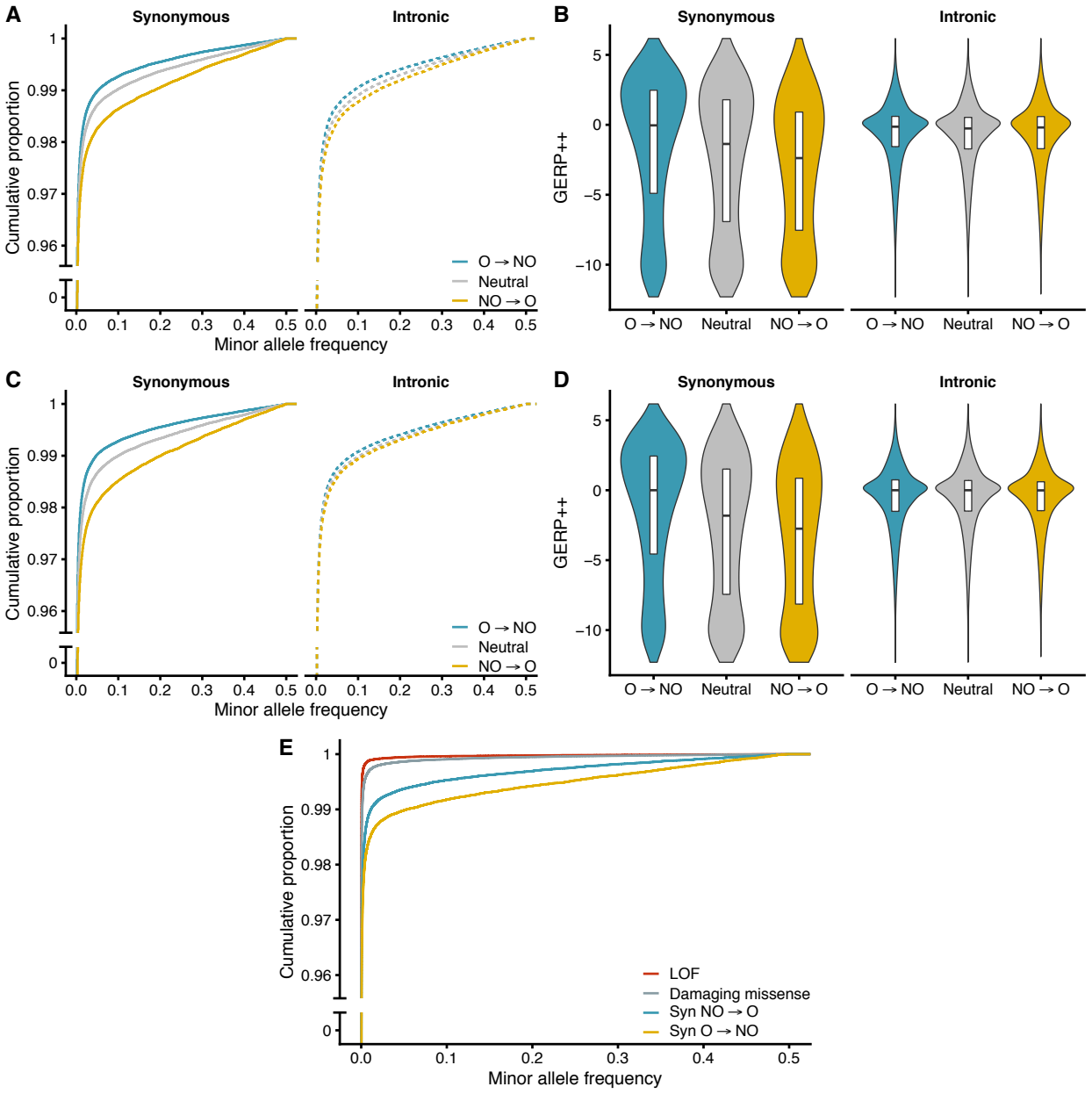

Figure S2. SFS of synonymous and intronic variants matched for 5' and 3' nucleotide content (A) Site frequency spectrum of variants matched for trinucleotide context. T-test *p*-values: Synonymous O → NO vs synonymous neutral ($p = 3.2 \times 10^{-34}$); synonymous O → NO versus intronic O → NO ($p = 5.4 \times 10^{-27}$); synonymous NO → O versus intronic NO → O ($p = 6.2 \times 10^{-4}$); and synonymous NO → O vs synonymous neutral ($p = 9.1 \times 10^{-16}$). **(B)** GERP++ distributions of the reference alleles for the matched variants included in (A). **(C)** SFS of the original matched synonymous and intronic variants using RSCU-defined codon optimality. **(D)** GERP++ distribution of reference alleles for the RSCU-annotated variants included in (C). **(E)** SFS of gnomAD loss-of-function, missense damaging (i.e. PolyPhen "probably" or "possibly" damaging), and codon optimality-altering synonymous variants.

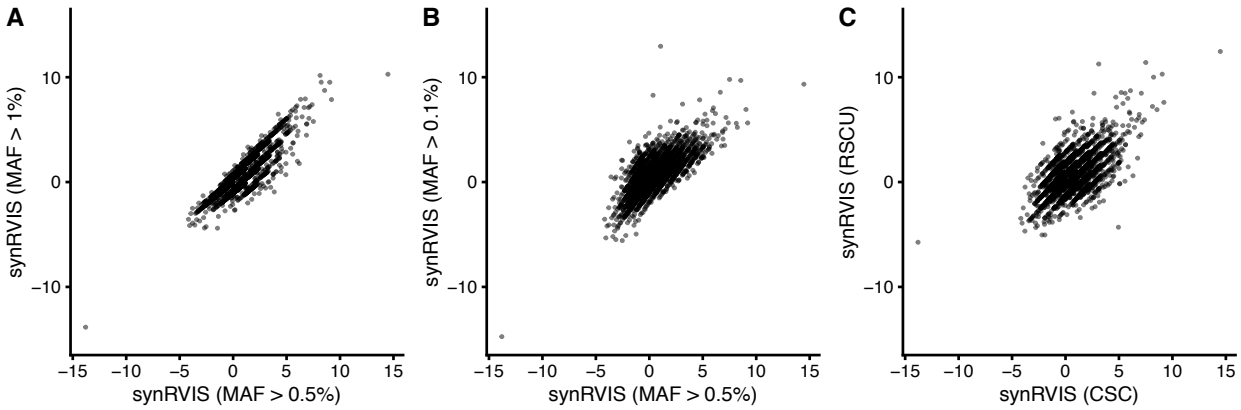

Figure S3. Comparisons of alternative synRVIS derivations. (A) Comparison of synRVIS scores calculated using a 1% MAF cutoff rather than 0.5% MAF cutoff for defining common O → NO synonymous variants (Y). **(B)** Comparison of using a MAF cutoff of 0.1% rather than 0.5% for (Y). **(C)** Comparison of CSC-defined codon optimality versus RSCU-defined codon optimality (MAF cutoff of 0.5% for both).

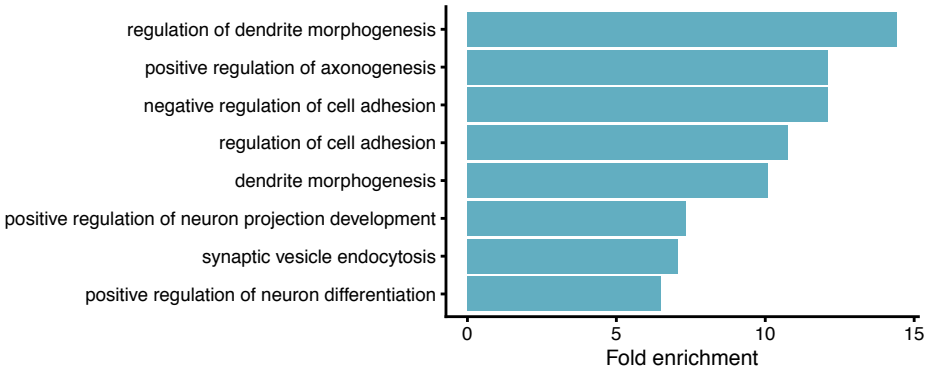

Figure S4: GO enrichments of genes tolerant to synonymous variation but intolerant to loss-of-function variation. Top gene ontology categories enriched for genes that fall in the bottom 25th percentile of LOEUF scores but top 25th percentile of synRVIS scores.

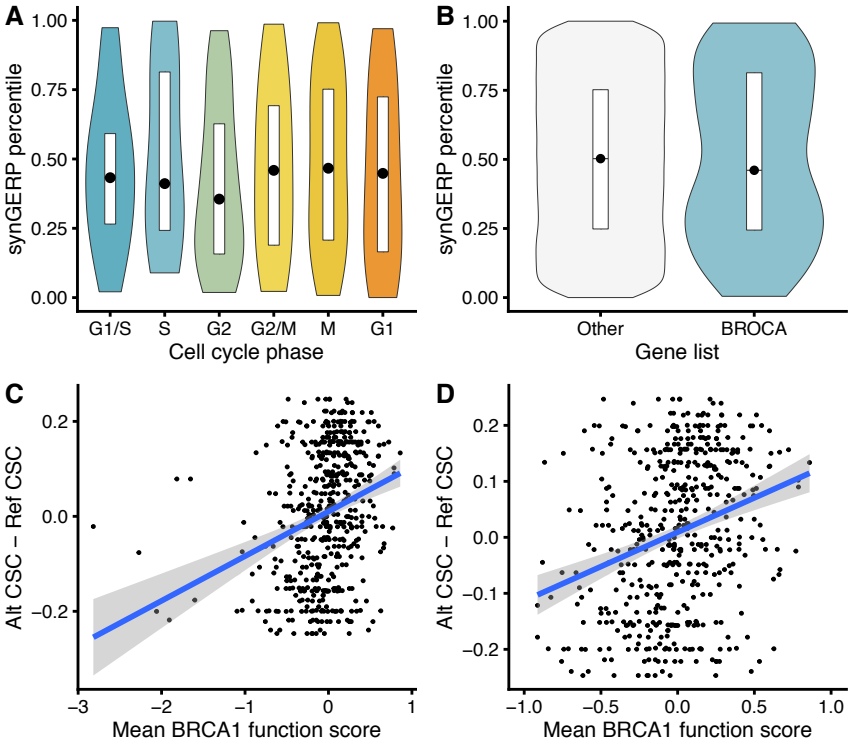

Figure S5. synGERP distributions of cell cycle expressed genes and BROCA list genes. (A) synGERP distributions of genes periodically expressed during the cell cycle. **(B)** synGERP distribution of genes contained in the BROCA panel versus all other protein-coding genes. **(C)** Scatter plot of CSC scores versus function scores for synonymous variants in *BRCA1*. **(D)** Same as (C) with outliers removed.

Supplementary tables

Table S1: List of genes with their synRVIS and synGERP scores

Table S2: Gene lists used for enrichment tests

Table S3: GO enrichment results

Table S4: Annotated *BRCA1* variants from Findlay et al. (2018)[74].