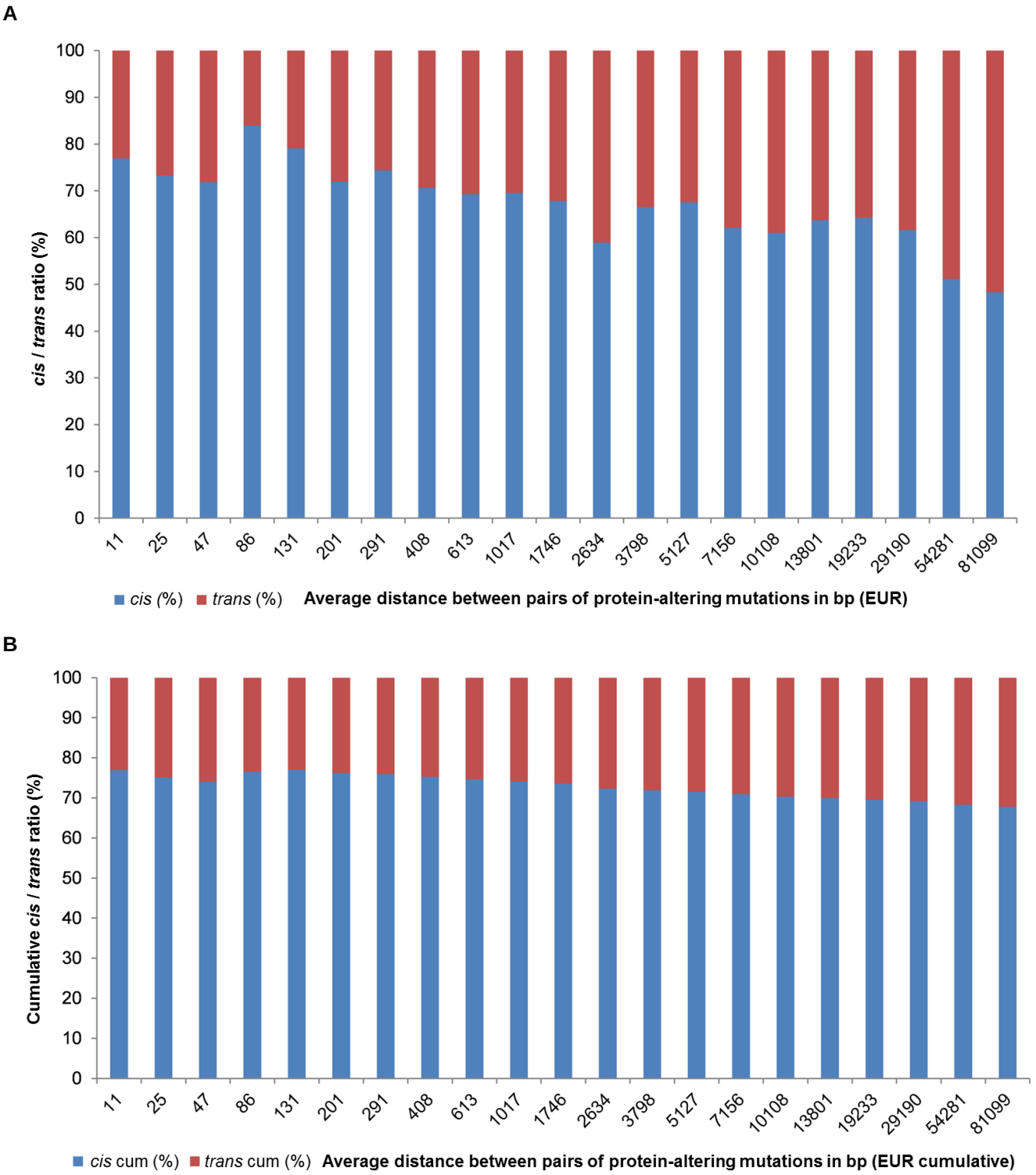
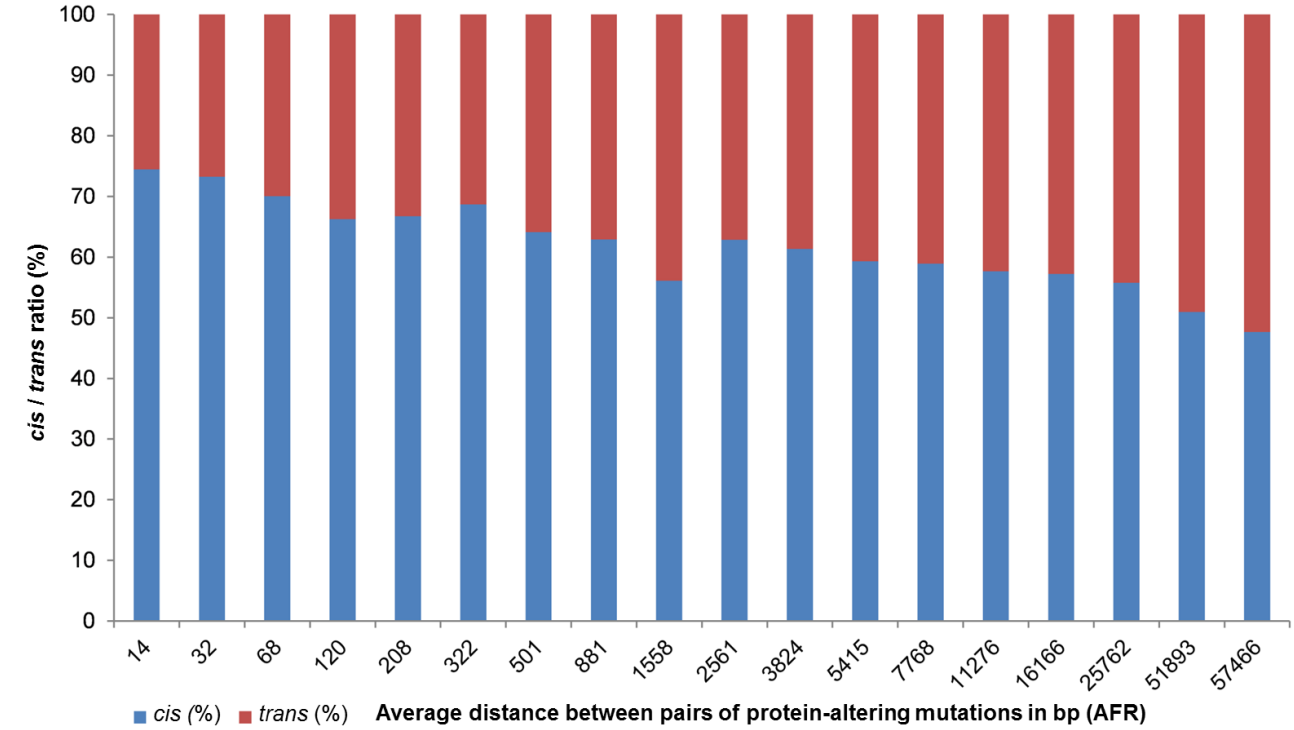


Figure S1. Relationship of inter-mutation distance with *cis*/*trans* ratio

(A) Relationship of inter-mutation distance with *cis*/*trans* ratio in EUR. *X*-axis**:** Average distance between pairs of protein-altering mutations per bin; the different pairs of mutations were sorted by distance (bp), i.e. the difference between the genomic coordinates of each mutation, then distributed into approximately 20 bins; for each bin, the average mutation distance was calculated; *y*-axis: *cis*/*trans* ratios, with the relative fraction of *cis* configurations (%) in blue, and the relative fraction of *trans* configurations (%) in red, both fractions are complementary, in sum 100%.

(B) Cumulative *cis*/*trans* ratios.

C



D

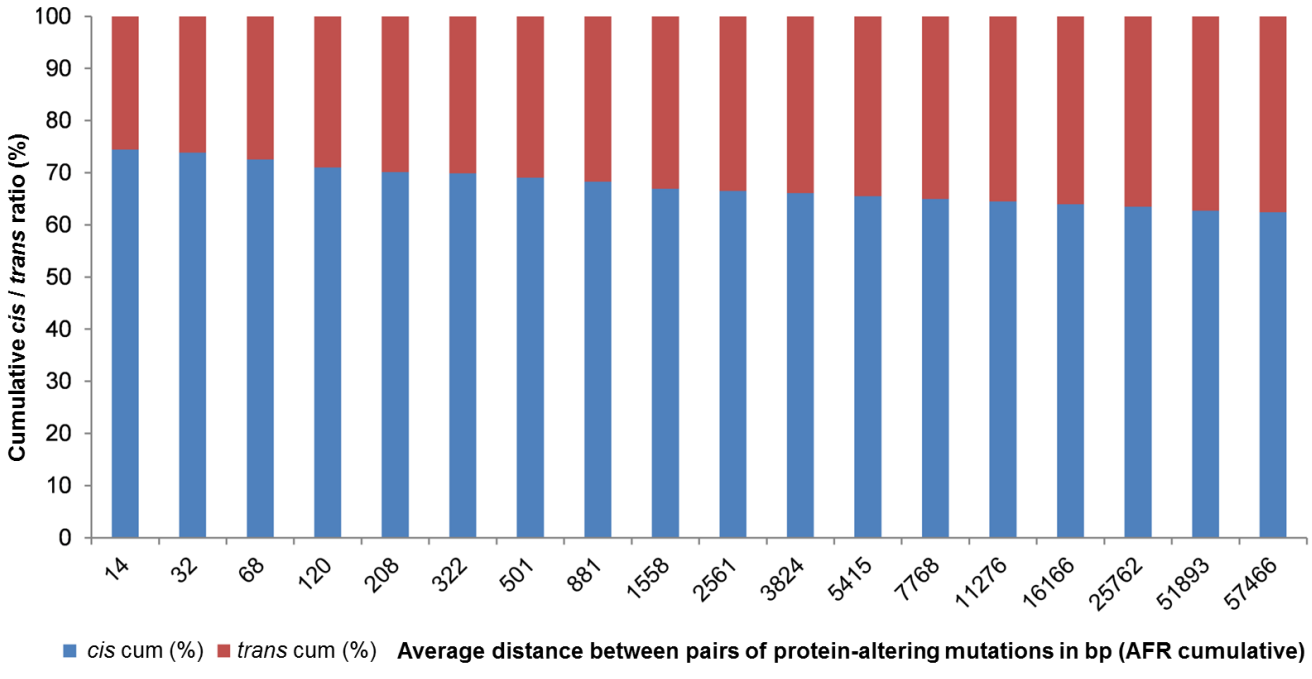
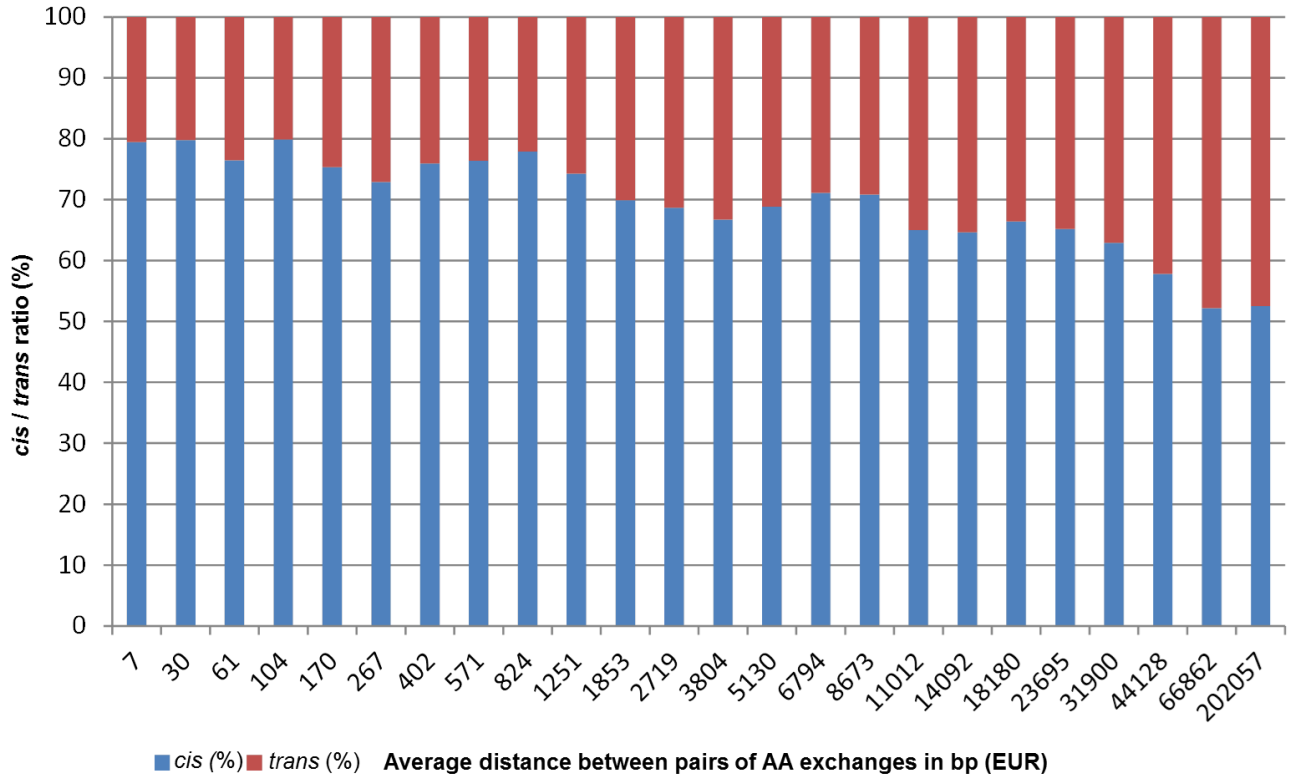


Figure S1. Relationship of inter-mutation distance with *cis*/*trans* ratio
(C) Relationship of inter-mutation distance with *cis*/*trans* ratio in AFR, analogous to *(A)*.
(D) Cumulative *cis*/*trans* ratios, analogous to *(B).*

E



F

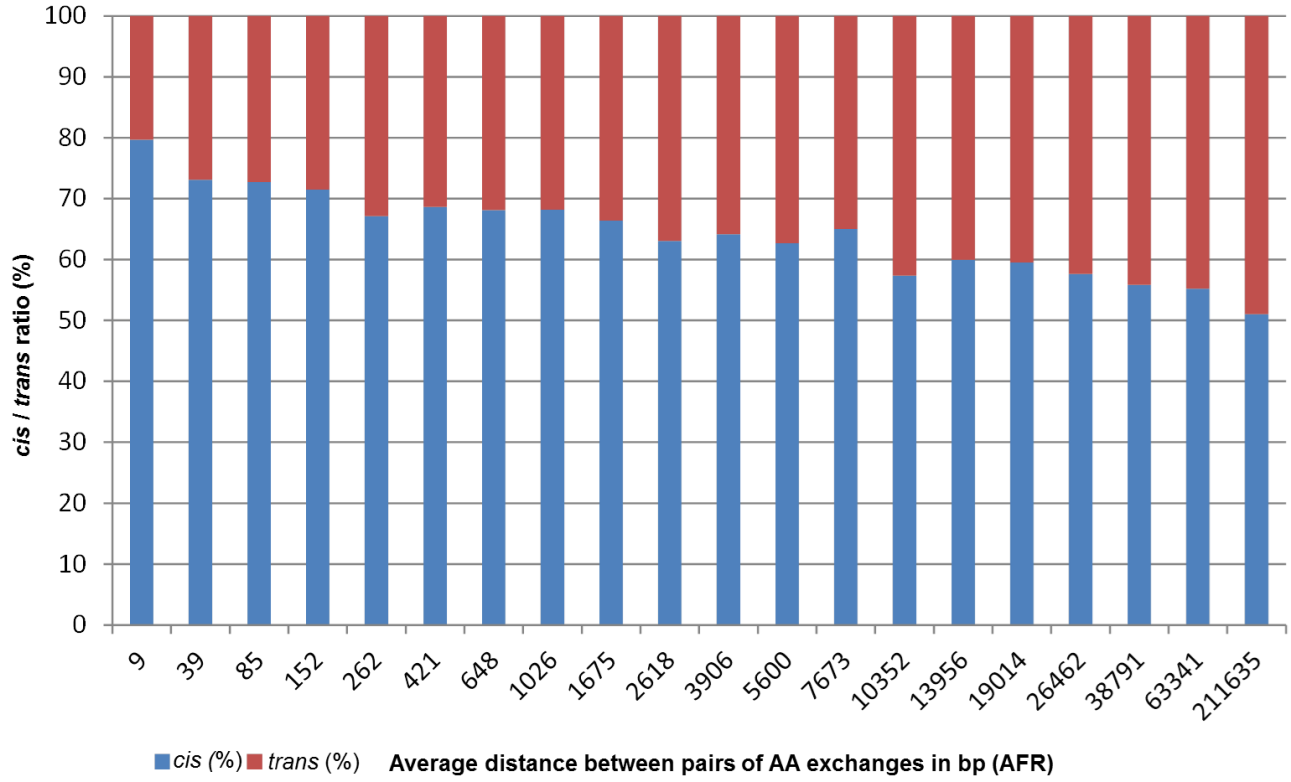
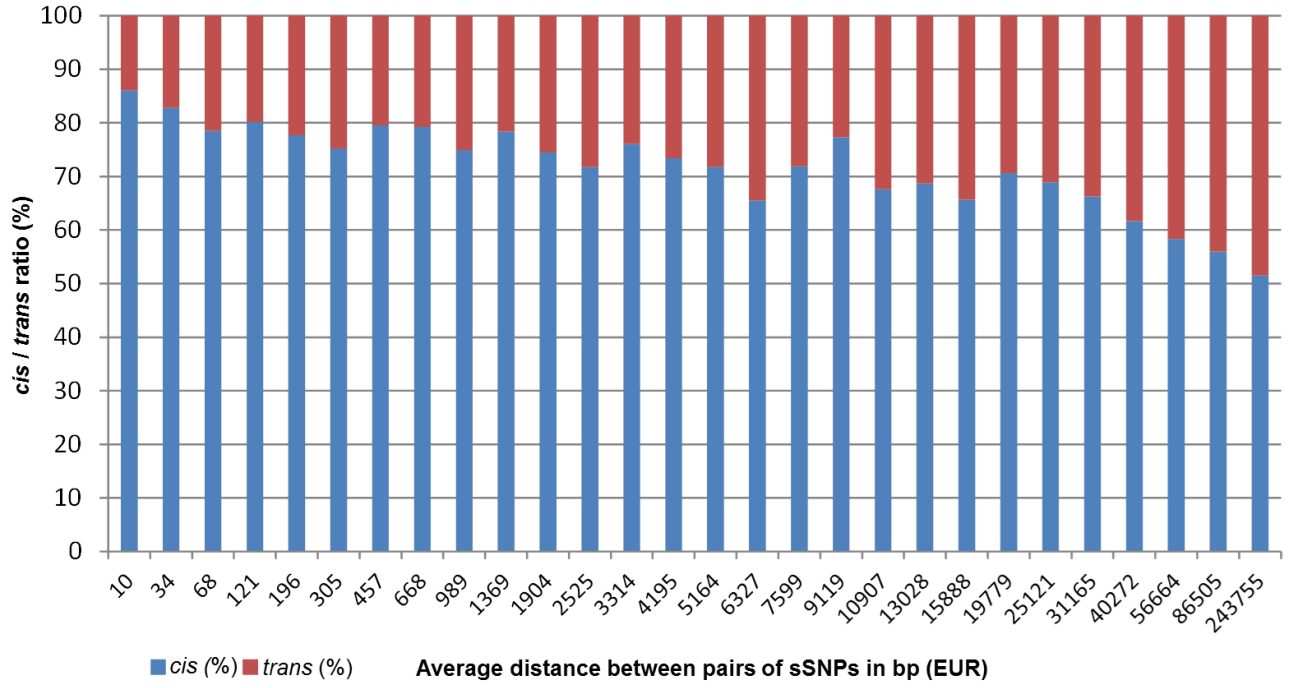


Figure S1. Relationship of inter-mutation distance with *cis*/*trans* ratio
(E) Relationship of inter-mutation distance with *cis*/*trans* ratio for pairs of amino acid (AA) exchanges in EUR, analogous to *(A)*.
(F) Relationship of inter-mutation distance with *cis*/*trans* ratio for pairs of AA exchanges in AFR, analogous to *(A)*.

G



H

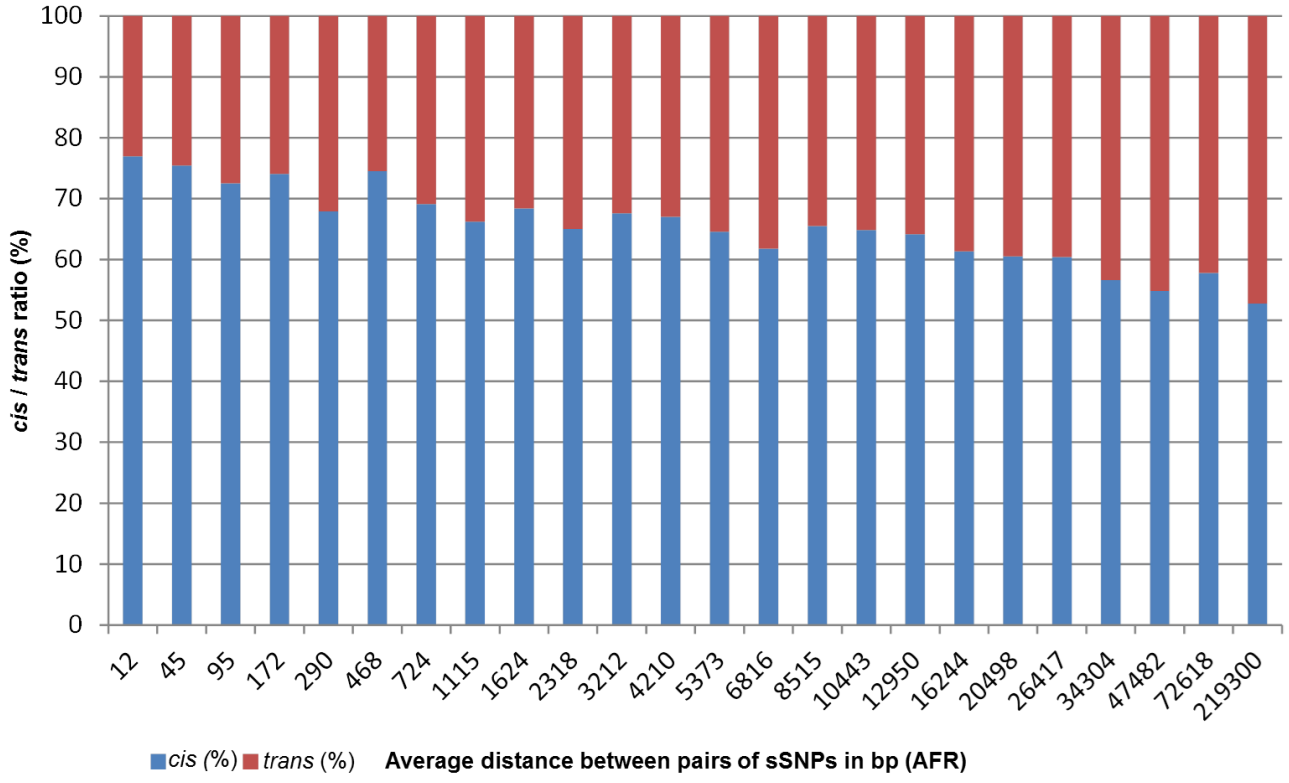


Figure S1. Relationship of inter-mutation distance with *cis*/*trans* ratio
(G) Relationship of inter-mutation distance with *cis*/*trans* ratio for pairs of synonymous SNPs (sSNPs) in EUR, analogous to *(A)*.
(H) Relationship of inter-mutation distance with *cis*/*trans* ratio for pairs of sSNPs in AFR, analogous to *(A)*.

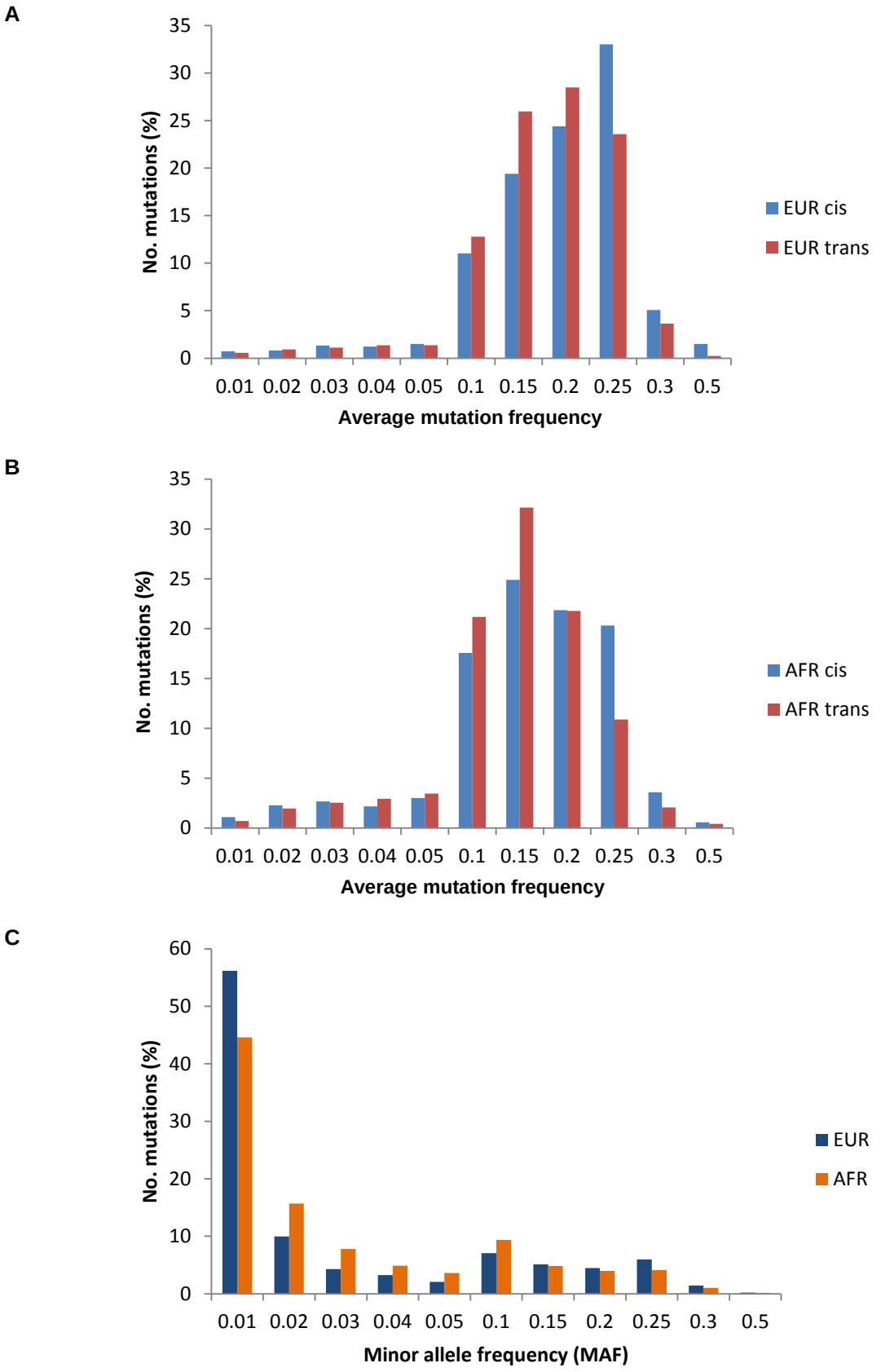
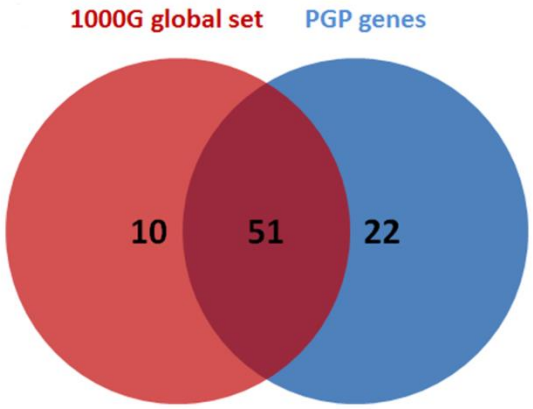


Figure S2. Average mutation frequency spectrum of pairs of mutations in *cis* and *trans*

(A) Average mutation frequency spectrum of pairs of protein-altering mutations in *cis* or *trans* in EUR (1000G); *x*-axis: average mutation frequency per bin (for each different pair of mutations in *cis* or *trans*, an average frequency is derived from the minor allele frequencies (MAFs) of each of the two mutations provided by the 1000G database (Abecasis et al. 2012) for each ancestry group). *Y*-axis: fraction of mutation pairs in *cis* or *trans* (%) relative to the total number of different pairs of protein-altering mutations in *cis* or *trans* (100%); blue bars indicate pairs of mutations in *cis,* red bars pairs in *trans*. For instance, 33% of all different mutation pairs in *cis* have an average mutation frequency > 0.2 and ≤ 0.25. For instance, 28% of all different mutation pairs in *trans* have an average mutation frequency > 0.15 and ≤ 0.2. *(B)* Analogous to *(A)*, average mutation frequency spectrum of pairs of protein-altering mutations in *cis* or *trans* in AFR. *(C)* Minor allele frequency (MAF) spectrum of all protein-altering mutations in EUR and AFR. All protein-altering mutations contained in the entirety of autosomal protein-coding genes (RefSeq) in the 1000G database, for each of the ancestry groups. *X*-axis: MAF per bin, presented separately for EUR (orange-brown) and AFR (blue); *y*-axis: the fraction of mutations (%) relative to the total number of mutations (100%), which are contained in each of the specified MAF bins.

A



B

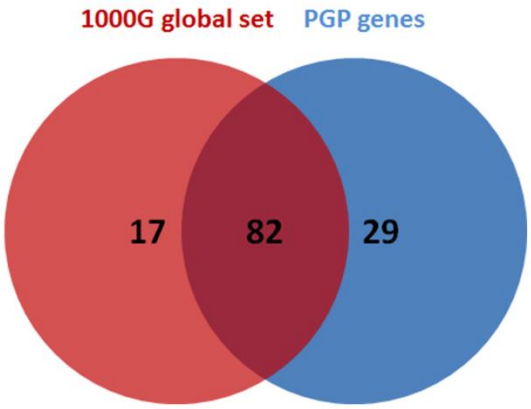


Figure S3. Pathways and GO terms shared between 1000G global set and PGP

(A) VENN diagram showing the overlap of pathways, which were significantly enriched (*P* < 0.01) in the global set of 2,402 phase-sensitive genes (1000G) (red circle) and the set of 1,627 phase-sensitive genes (*P* < 0.01), which PGP shared with 1000G (blue circle). *(B)* VENN diagram showing the overlap of GO terms, which were significantly enriched (*P* < 0.001) in the global set of 2,402 phase-sensitive genes (1000G) (red circle) and the set of 1,627 phase-sensitive genes (*P* < 0.001), which PGP shared with 1000G (blue circle).

The overlap of the sets can be quantified with Sorensen's similarity index, $S = \frac{2ab}{a+b}$, where *a* is the number of genes in the first set, *b* the number of genes in the second set and *ab* the number of genes shared by the two sets. This results in $S = 0.76$ for the similarity between pathways (~76%) and $S = 0.78$ for the similarity between GO terms (~78%).

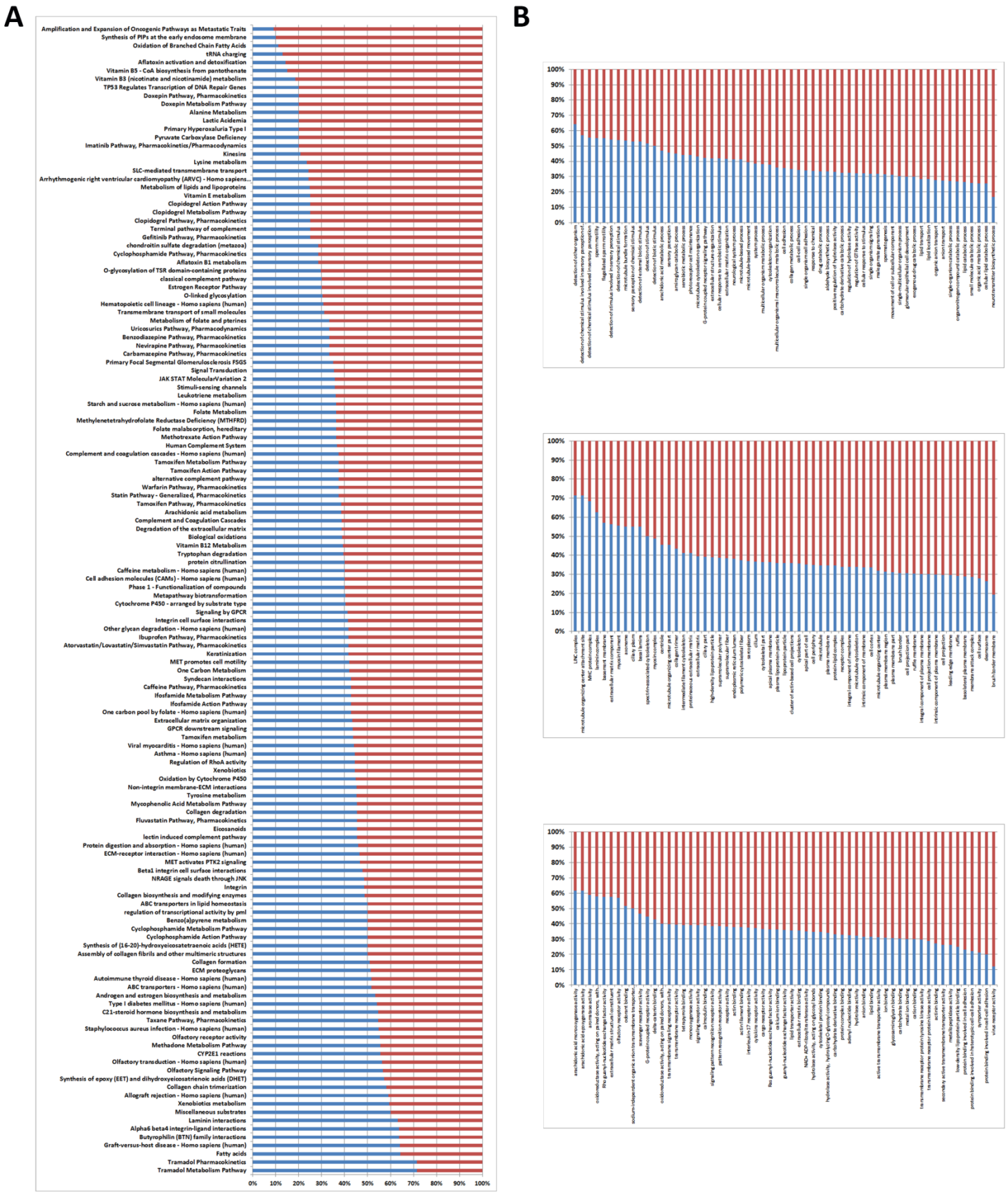


Figure S4. Enrichment of global sets of phase-sensitive and variable genes with pathways and GO terms
(A) Enrichment of the global set of 7,524 variable genes (which have ≥ 1 protein-altering mutations in at least one genome in each of the four ancestry groups in 1000G) with pathways. The 138 most significantly enriched pathways ($P < 0.01$) are shown. Blue bars: relative proportion (%) of the subset of genes with ≥ 2 protein-altering mutations, i.e. the genes from the global set of phase-sensitive genes (1000G); red bars: relative proportion (%) of the genes with one mutation. *(B)* Enrichment of the global set of 7,524 variable genes (1000G) with GO terms. The 177 most significantly enriched GO terms ($P < 0.001$) are shown. Blue bars: relative proportion/contribution (%) of the subset of genes with ≥ 2 protein-altering mutations, i.e. the genes from the global set of phase-sensitive genes (1000G); red bars: relative proportion/contribution (%) of the genes with one mutation. Top: GO terms related to 'biological process'; middle: GO terms related to 'cellular component'; bottom: GO terms related to 'molecular function'.

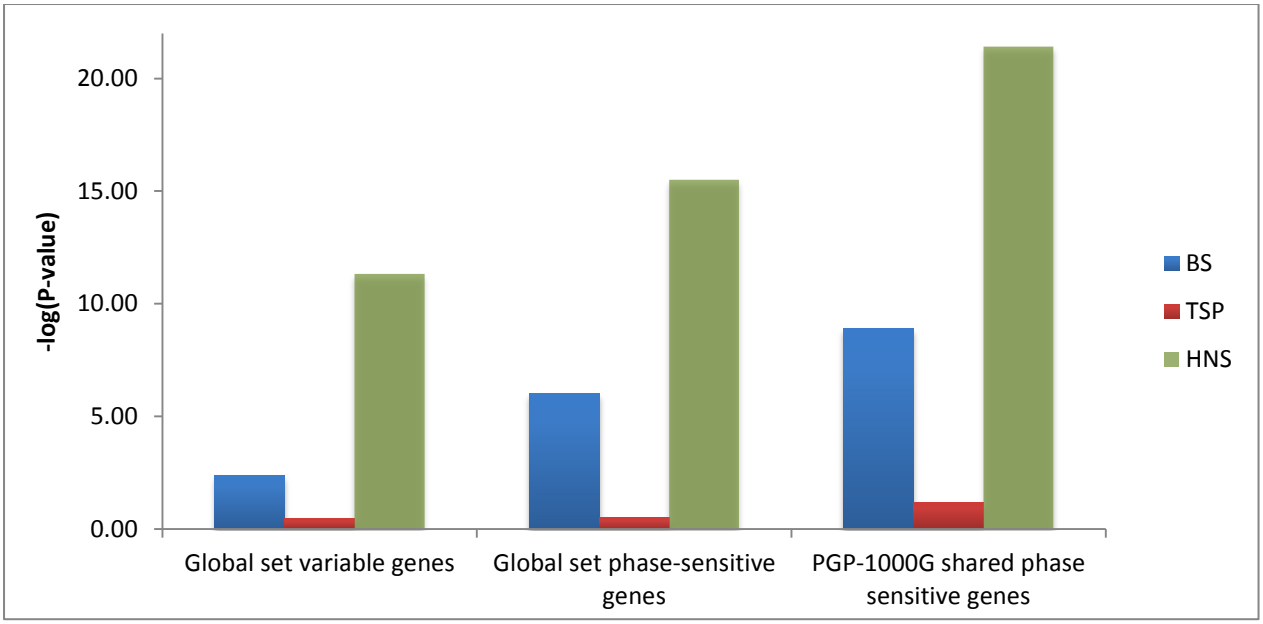


Figure S5. **Enrichment of 1000G and PGP variable genes for gene sets of evolutionary significance** Panel, left: Blue bar indicates significance of enrichment of the global set of variable genes (1000G), i.e. genes with ≥ 1 mutations, with a set of 226 genes reported to evolve under balancing selection (BS); red bar significance of enrichment with a set of 60 genes with any evidence of human-chimpanzee trans-species polymorphisms or haplotypes (TSPs), and green bar enrichment with a set of 104 genes harboring at least one ancient protein-coding SNP or haplotype shared between humans and Neanderthals (HNS); *y*-axis: negative log of the enrichment *P*-value computed with Fisher's exact test. Panel, center: significance of enrichment of the global set of phase-sensitive genes (1000G), i.e. genes with ≥ 2 mutations, with these gene sets. Panel, right: significance of enrichment of the set of 1,627 (cross-validated) phase-sensitive genes which PGP shares with the global set of phase-sensitive genes (1000G). The gene sets of potential evolutionary significance are described in Savova et al. (2016).

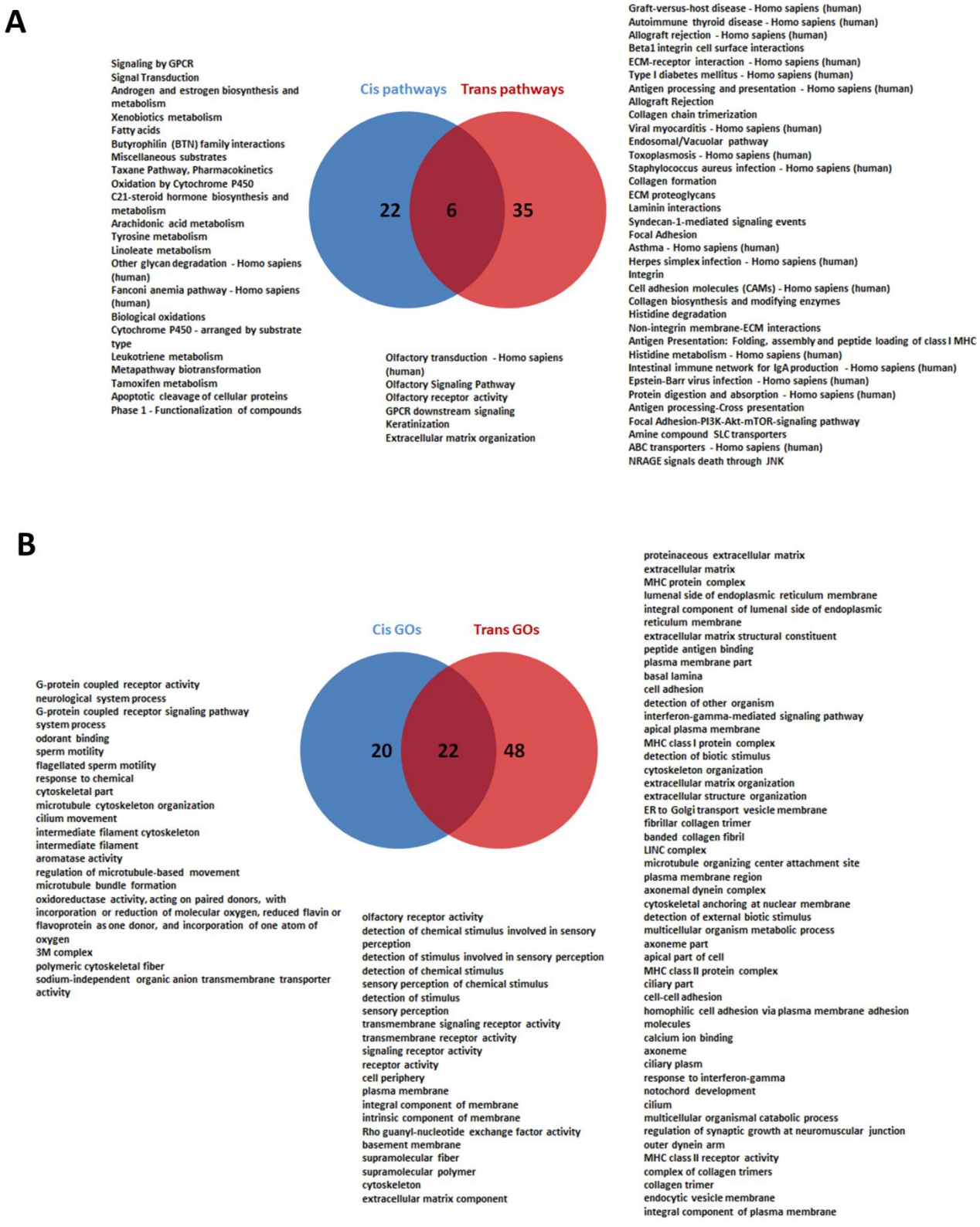


Figure S6. Differential enrichment of *cis*- and *trans*-abundant genes with pathways and GO terms
(A) VENN diagram showing the intersection of the pathways which are enriched in either *cis*- (left) or *trans*-abundant genes (right); pathways listed in-between are enriched in both gene categories. *(B)* Intersection of the GO terms enriched in either *cis*- (left) or *trans*-abundant genes (right); GO terms listed in-between are enriched in both gene categories.

A

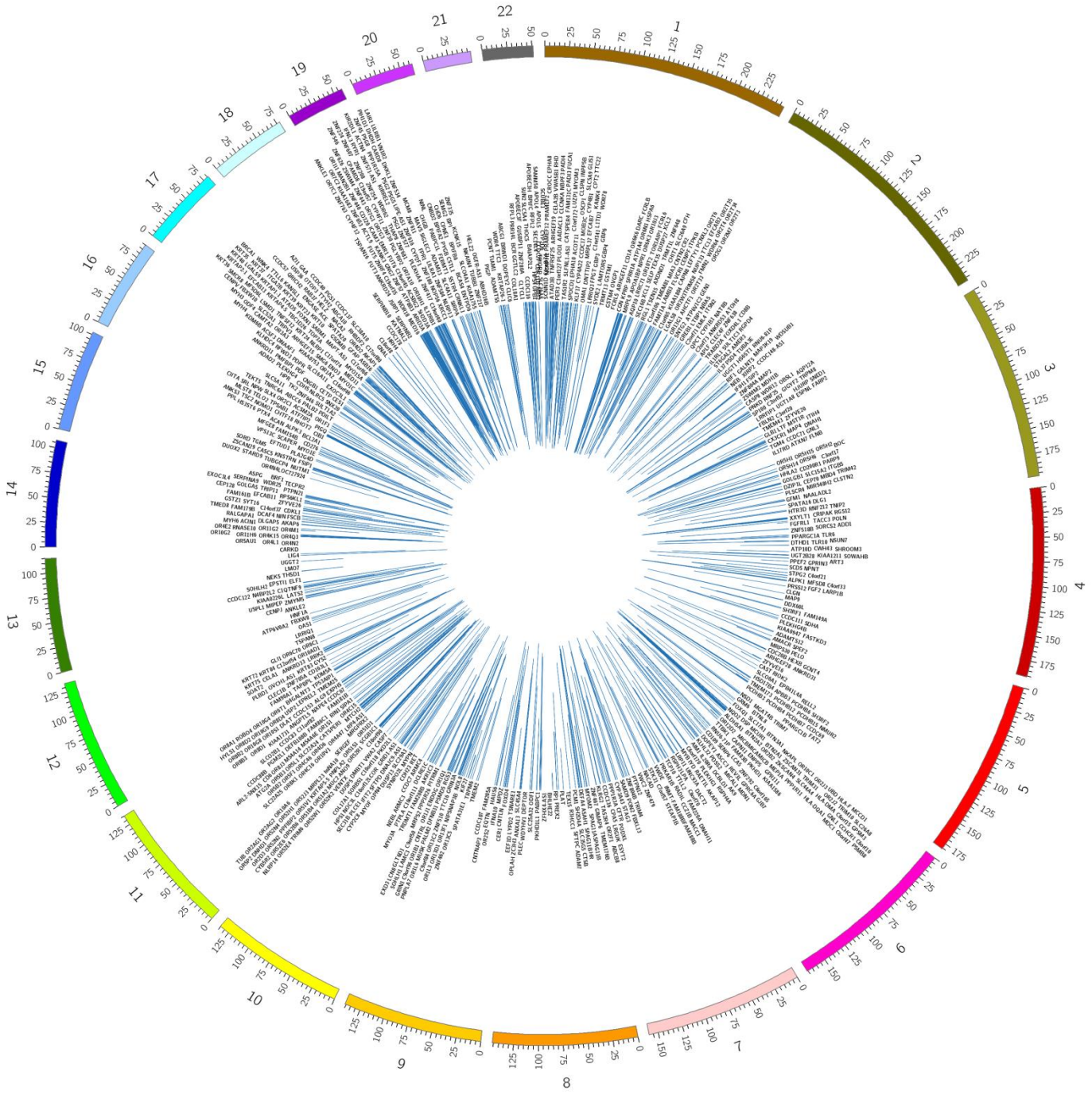


Figure S7. Distribution of *cis*- and *trans*-abundant genes across the autosomes
(A) Distribution of *cis*-abundant genes.

B

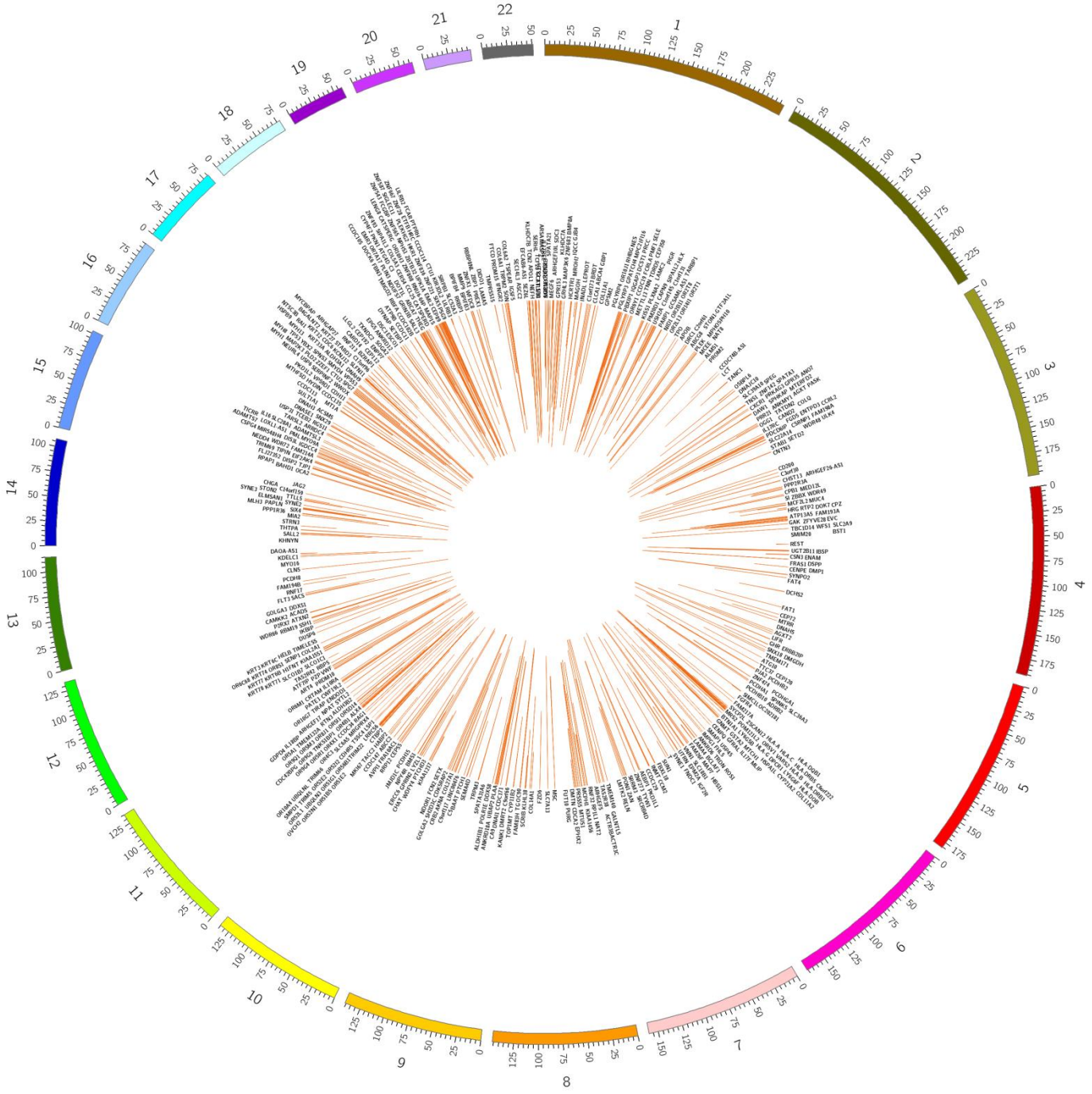


Figure S7. Distribution of *cis*- and *trans*-abundant genes across the autosomes
(B) Distribution of *trans*-abundant genes.