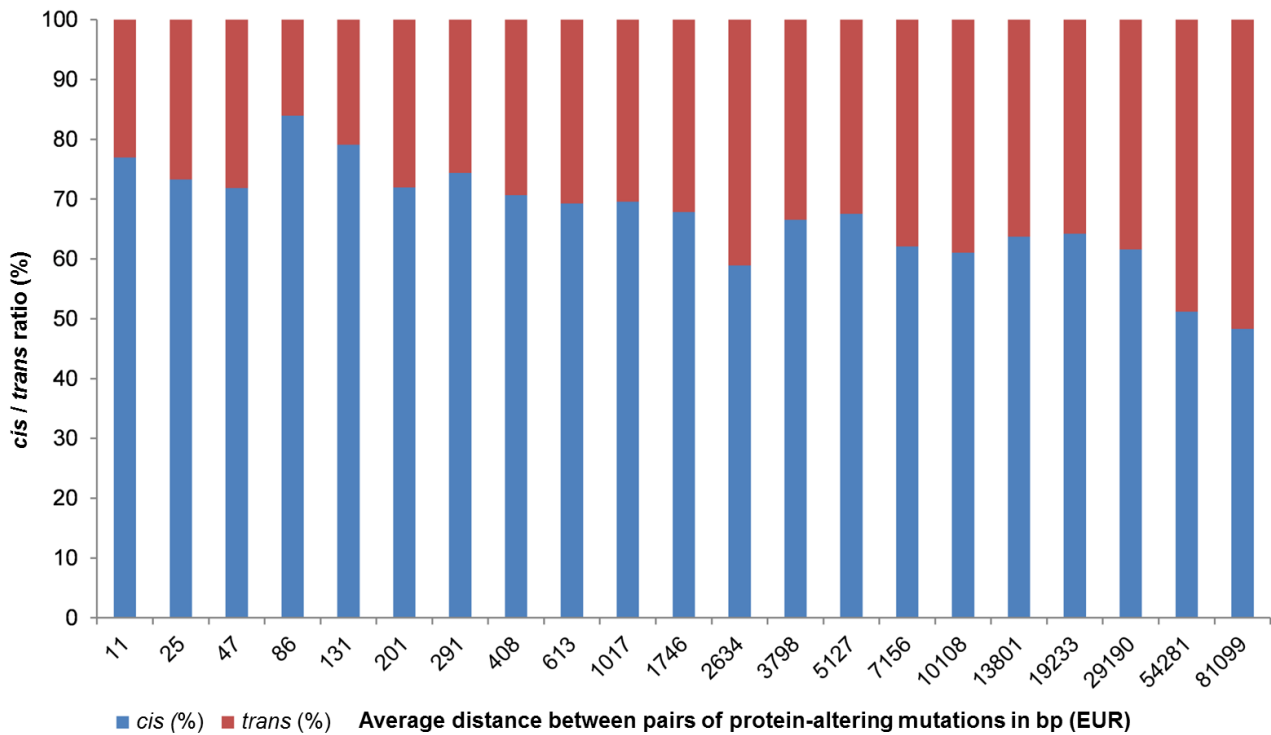
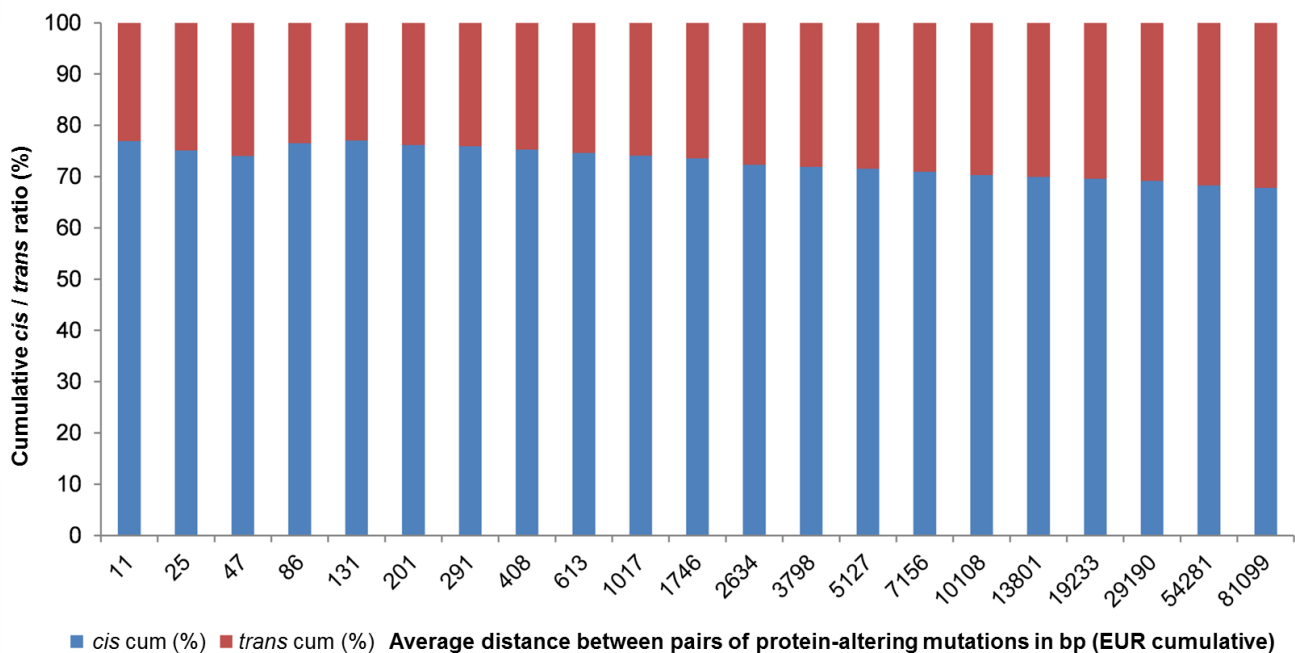
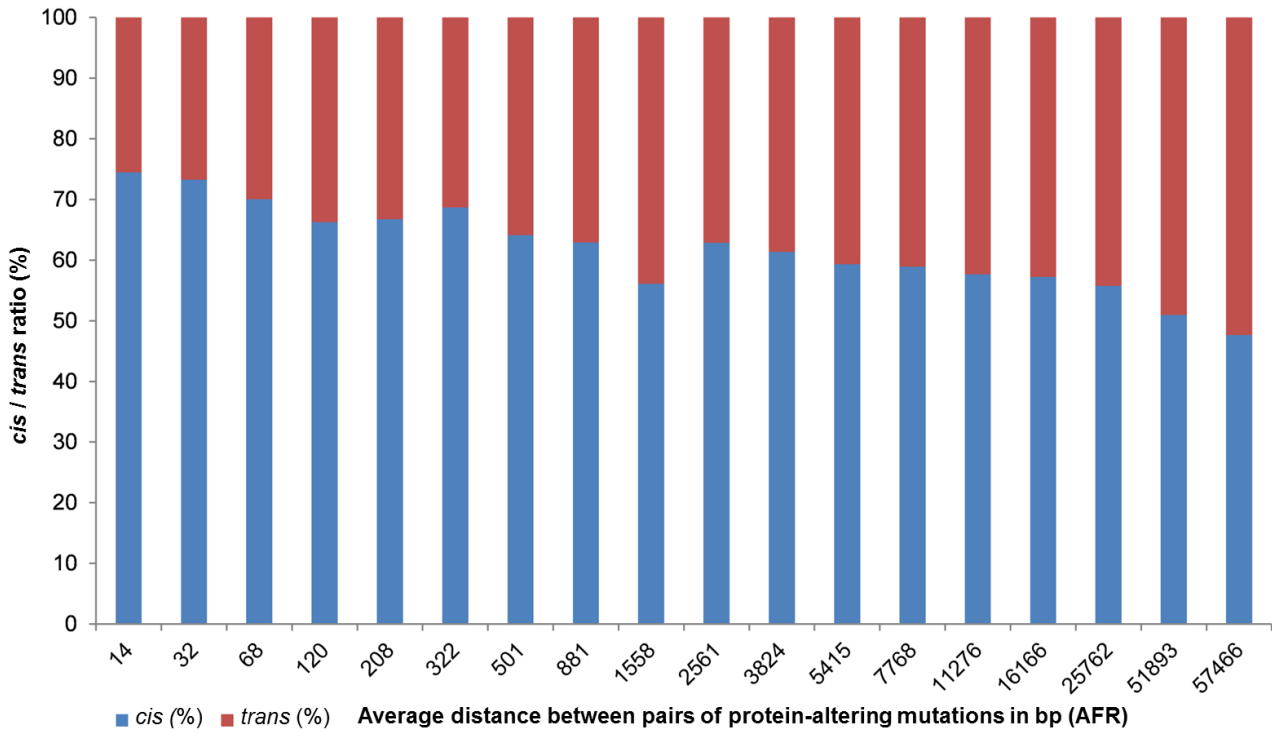
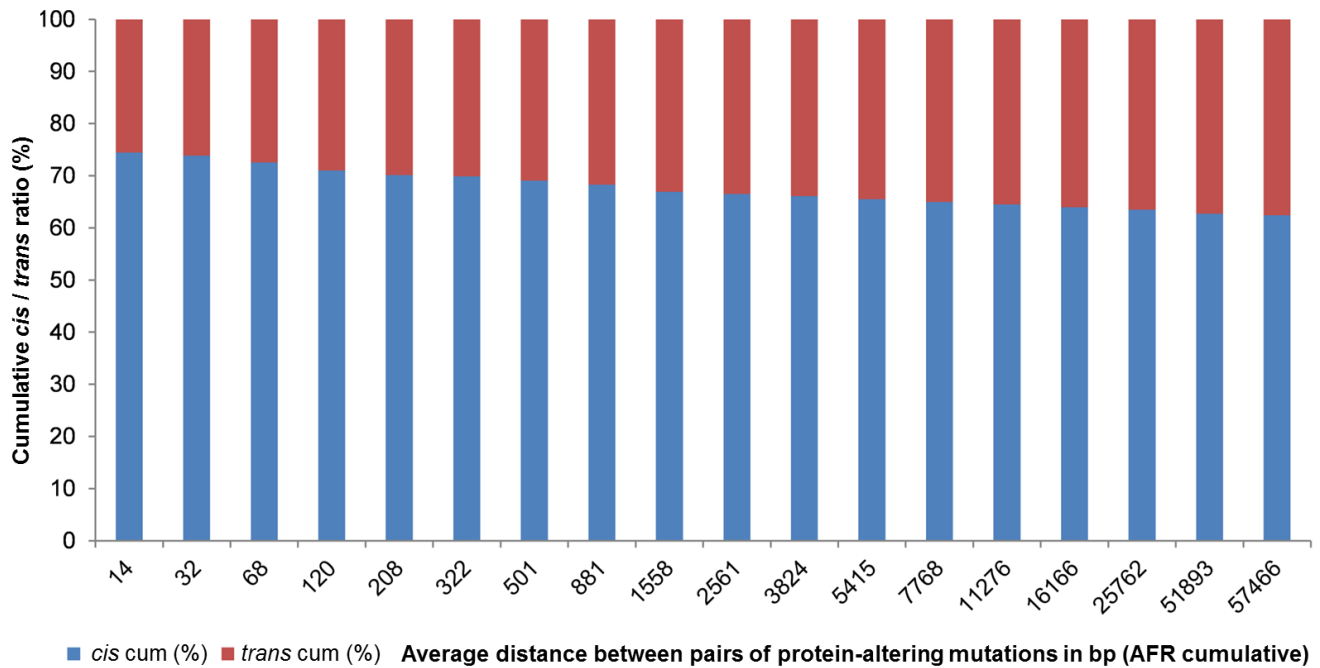


A**B****Figure S1. Relationship of inter-mutation distance with *cis/trans* ratio**

(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)*.

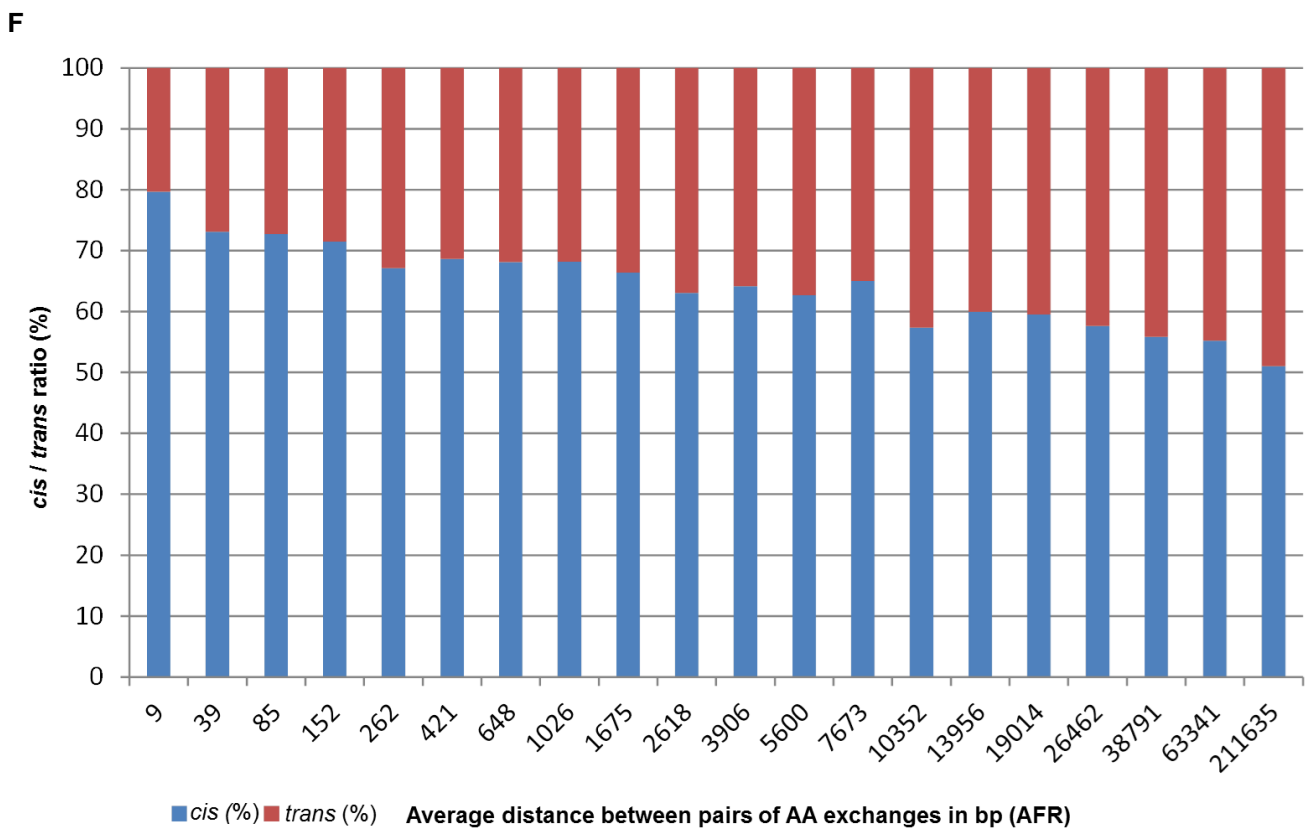
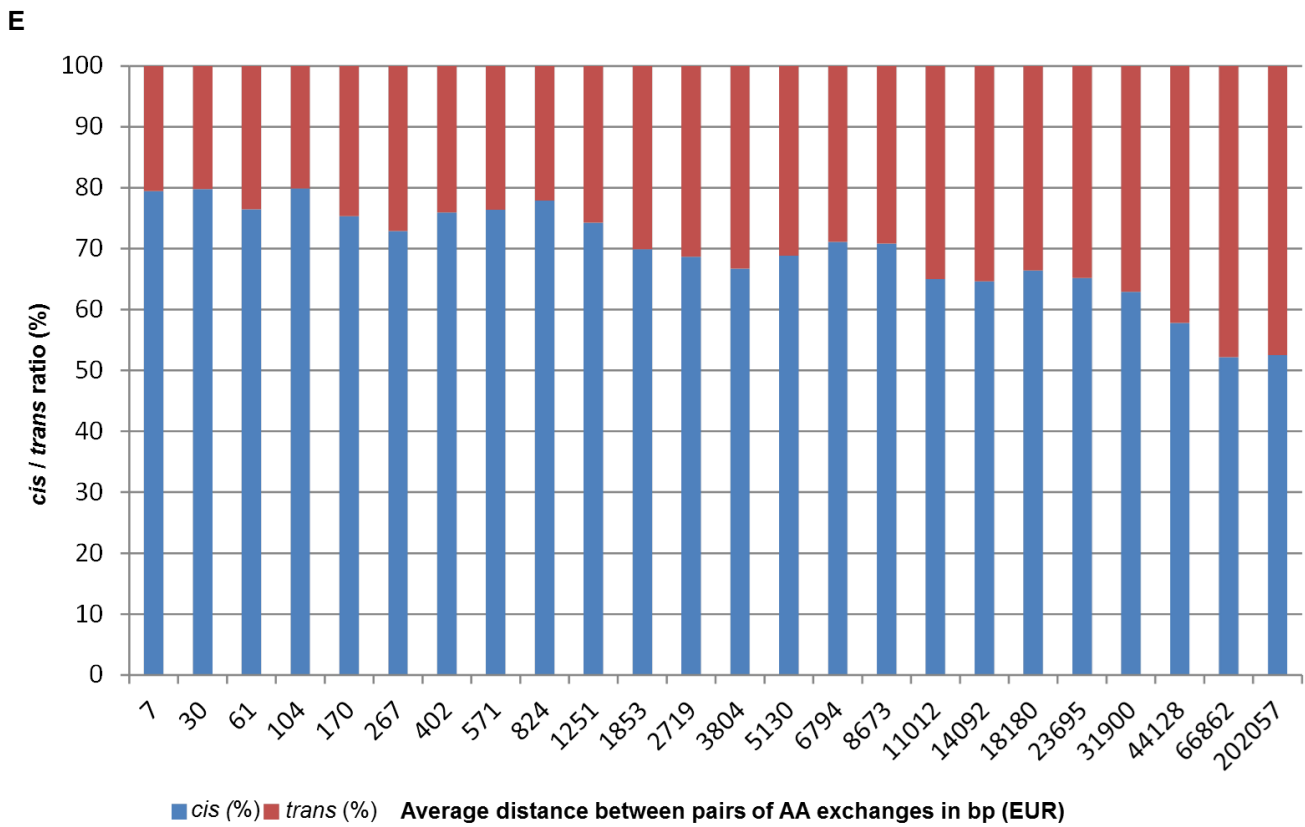
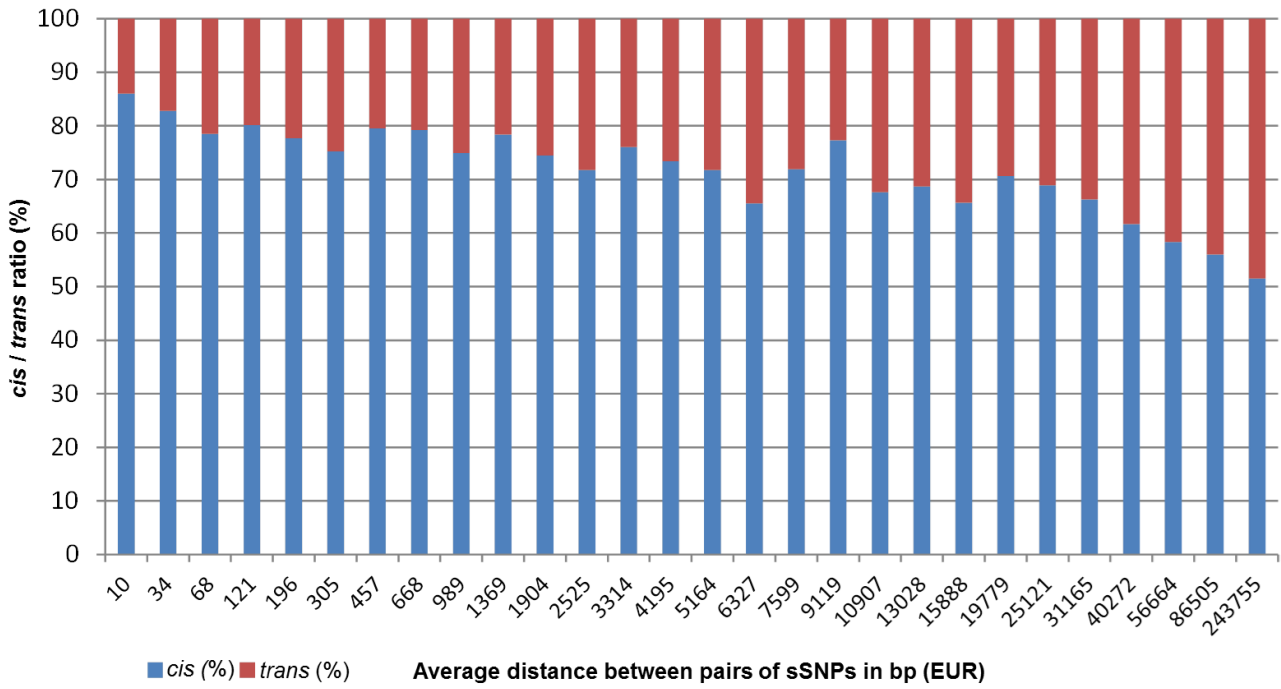
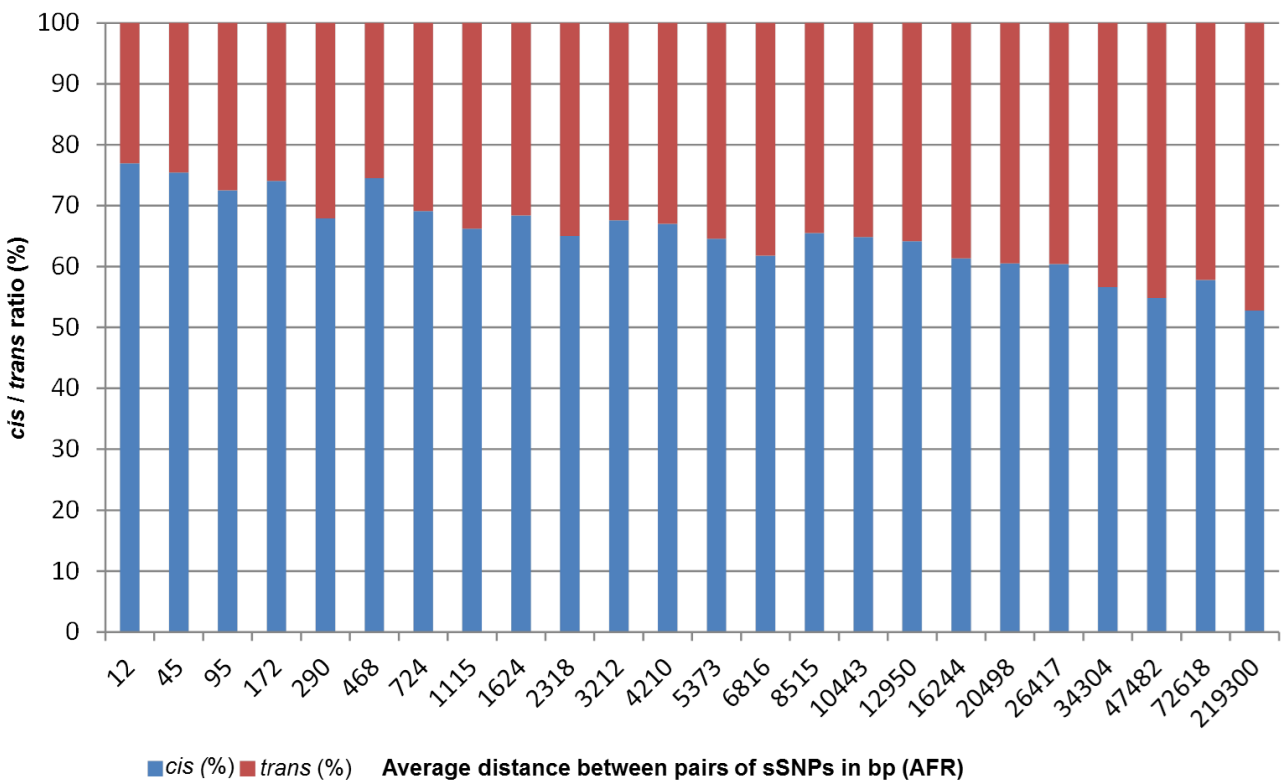


Figure S1. Relationship of inter-mutation distance with *cis/trans* ratio

(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).

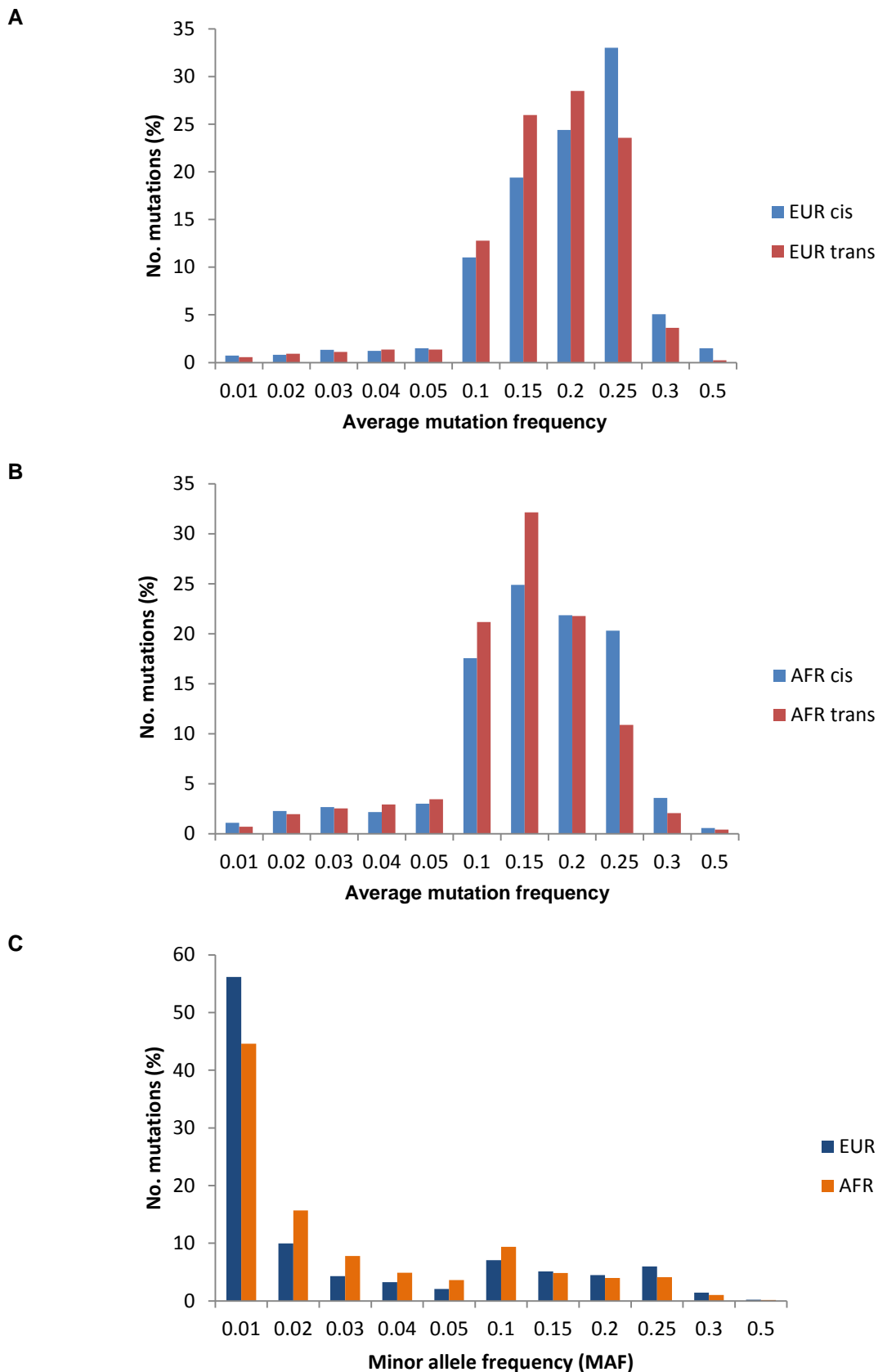


Figure S2. Average mutation frequency spectrum of pairs of mutations in *cis* and *trans*

(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.

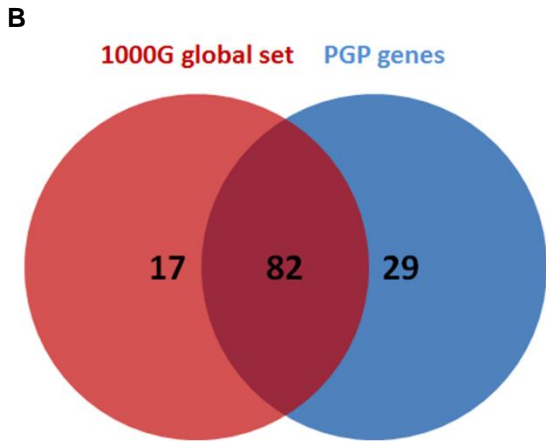
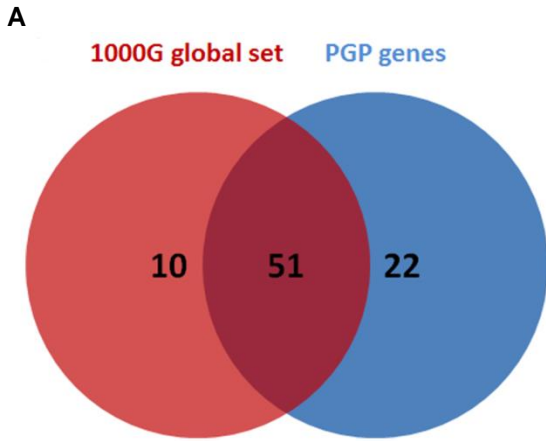


Figure S3. Pathways and GO terms shared between 1000G global set and PGP

(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%).

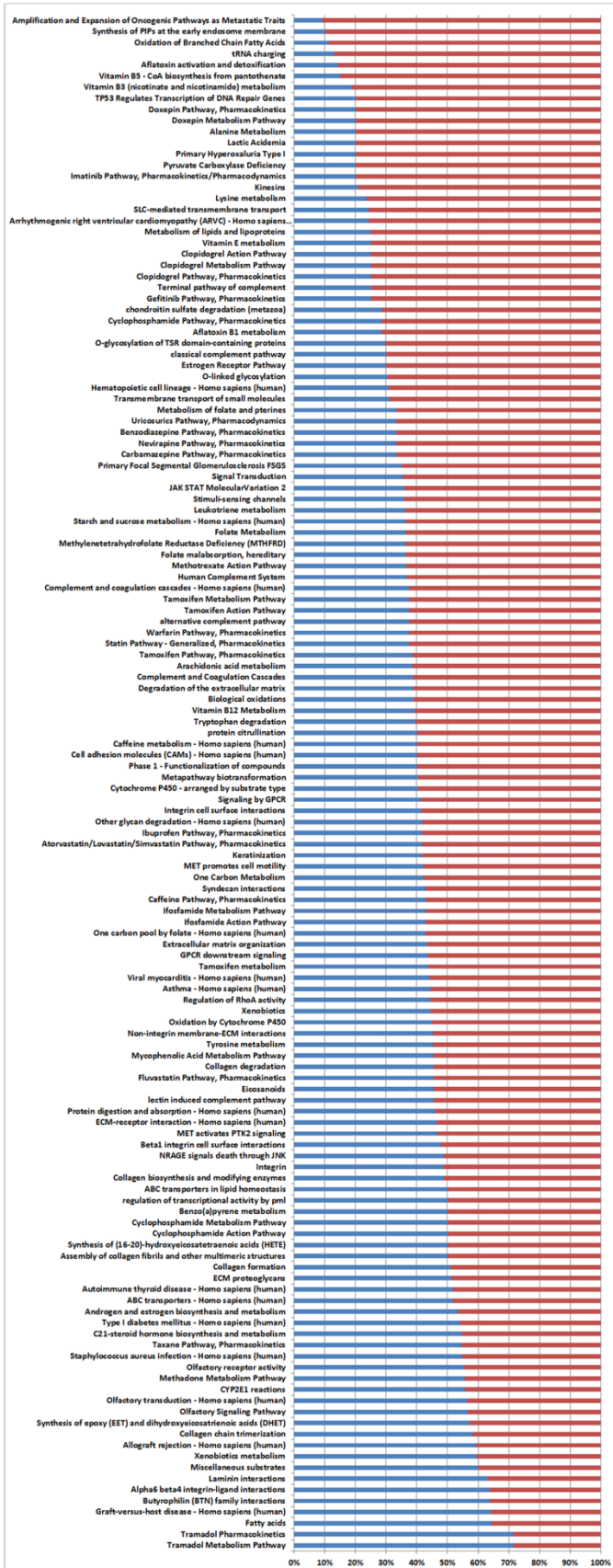
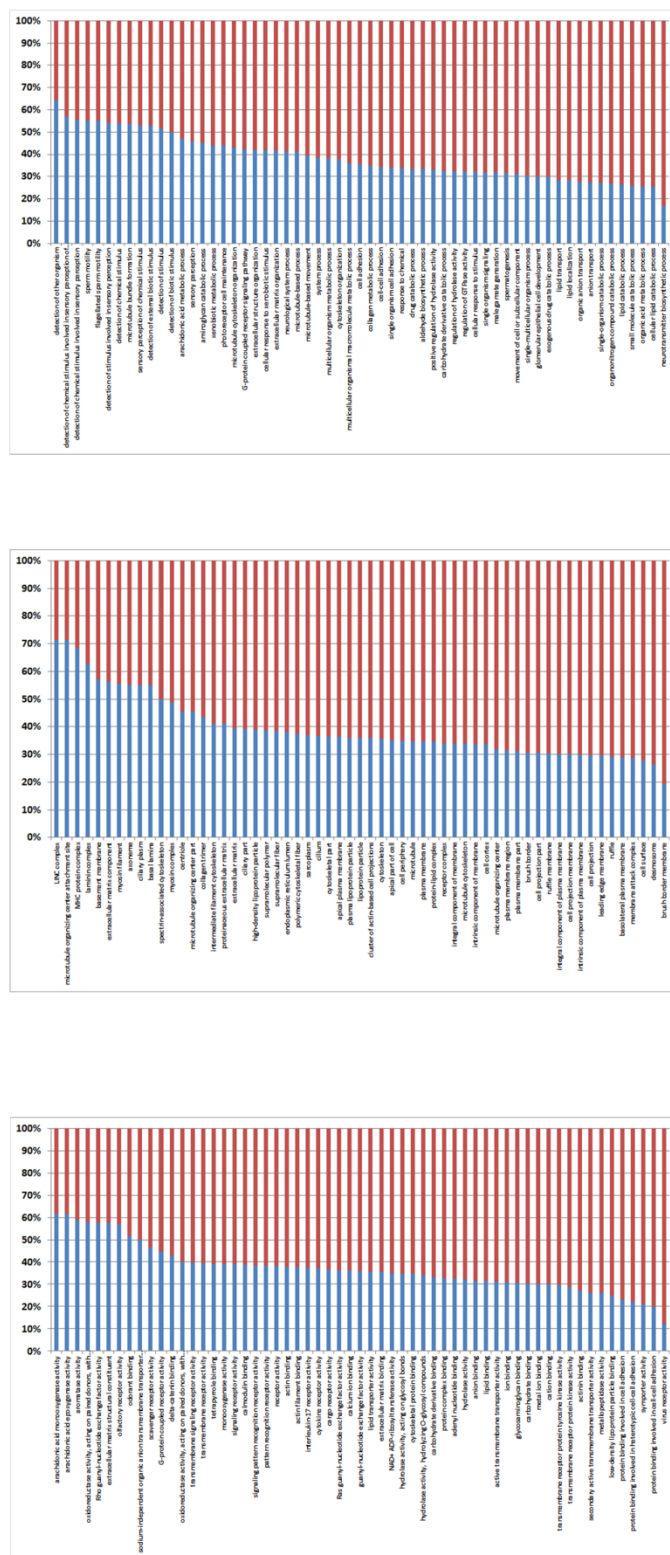
A**B**

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'.

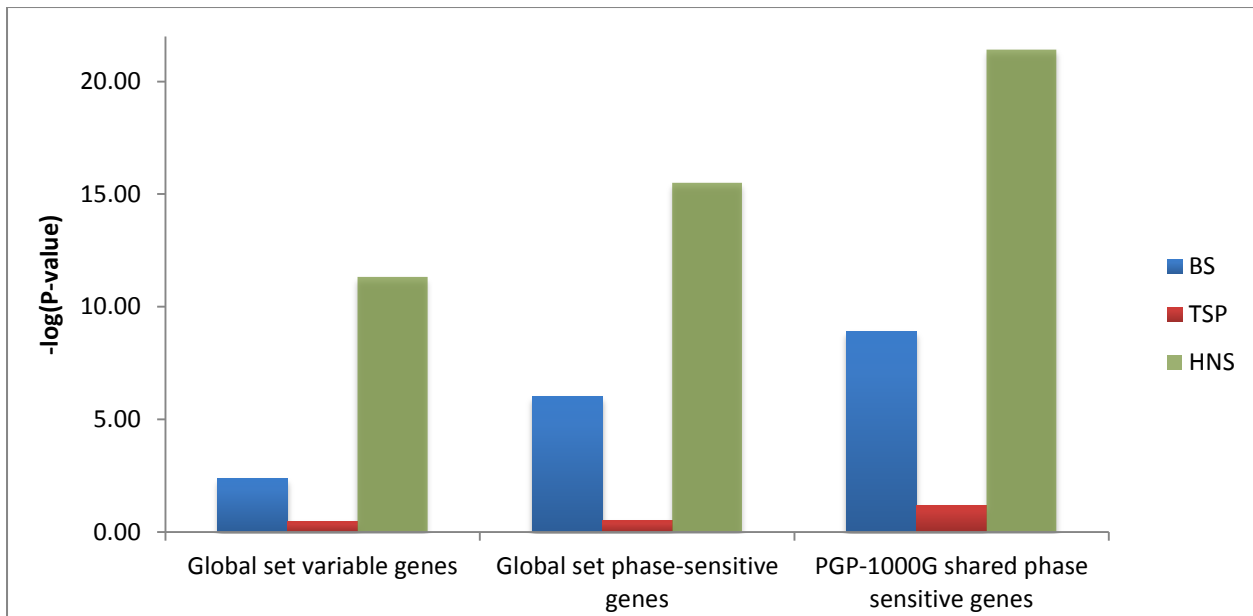
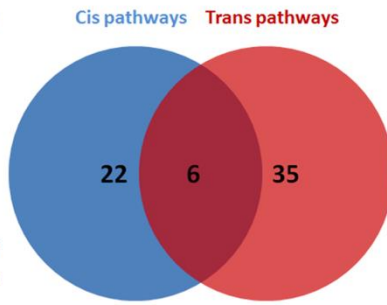


Figure S5. Enrichment of 1000G and PGP variable genes for gene sets of evolutionary significance

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).

A

Signaling by GPCR
Signal Transduction
Androgen and estrogen biosynthesis and metabolism
Xenobiotics metabolism
Fatty acids
Butyrophilin (BTN) family interactions
Miscellaneous substrates
Taxane Pathway, Pharmacokinetics
Oxidation by Cytochrome P450
C21-steroid hormone biosynthesis and metabolism
Arachidonic acid metabolism
Tyrosine metabolism
Linoleate metabolism
Other glycan degradation - Homo sapiens (human)
Fanconi anemia pathway - Homo sapiens (human)
Biological oxidations
Cytochrome P450 - arranged by substrate type
Leukotriene metabolism
Metapathway biotransformation
Tamoxifen metabolism
Apoptotic cleavage of cellular proteins
Phase 1 - Functionalization of compounds

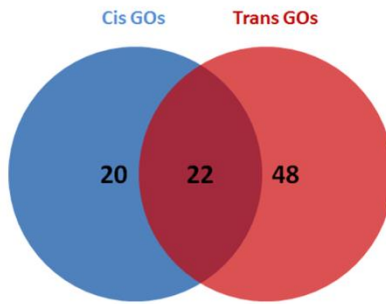


Olfactory transduction - Homo sapiens (human)
Olfactory Signaling Pathway
Olfactory receptor activity
GPCR downstream signaling
Keratinization
Extracellular matrix organization

Graft-versus-host disease - Homo sapiens (human)
Autoimmune thyroid disease - Homo sapiens (human)
Allograft rejection - Homo sapiens (human)
Beta1 integrin cell surface interactions
ECM-receptor interaction - Homo sapiens (human)
Type I diabetes mellitus - Homo sapiens (human)
Antigen processing and presentation - Homo sapiens (human)
Allograft Rejection
Collagen chain trimerization
Viral myocarditis - Homo sapiens (human)
Endosomal/Vacuolar pathway
Toxoplasmosis - Homo sapiens (human)
Staphylococcus aureus infection - Homo sapiens (human)
Collagen formation
ECM proteoglycans
Laminin interactions
Syndecan-1-mediated signaling events
Focal Adhesion
Asthma - Homo sapiens (human)
Herpes simplex infection - Homo sapiens (human)
Integrin
Cell adhesion molecules (CAMs) - Homo sapiens (human)
Collagen biosynthesis and modifying enzymes
Histidine degradation
Non-integrin membrane-ECM interactions
Antigen Presentation: Folding, assembly and peptide loading of class I MHC
Histidine metabolism - Homo sapiens (human)
Intestinal immune network for IgA production - Homo sapiens (human)
Epstein-Barr virus infection - Homo sapiens (human)
Protein digestion and absorption - Homo sapiens (human)
Antigen processing-Cross presentation
Focal Adhesion-P13K-Akt-mTOR-signaling pathway
Amine compound SLC transporters
ABC transporters - Homo sapiens (human)
NRAGE signals death through JNK

B

G-protein coupled receptor activity
neurological system process
G-protein coupled receptor signaling pathway
system process
odorant binding
sperm motility
flagellated sperm motility
response to chemical
cytoskeletal part
microtubule cytoskeleton organization
cilium movement
intermediate filament cytoskeleton
intermediate filament
aromatase activity
regulation of microtubule-based movement
microtubule bundle formation
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen
3M complex
polymeric cytoskeletal fiber
sodium-independent organic anion transmembrane transporter activity



olfactory receptor activity
detection of chemical stimulus involved in sensory perception
perception
detection of stimulus involved in sensory perception
detection of chemical stimulus
sensory perception of chemical stimulus
detection of stimulus
sensory perception
transmembrane signaling receptor activity
transmembrane receptor activity
signaling receptor activity
receptor activity
cell periphery
plasma membrane
integral component of membrane
intrinsic component of membrane
Rho guanyl-nucleotide exchange factor activity
basement membrane
supramolecular fiber
supramolecular polymer
cytoskeleton
extracellular matrix component

proteinaceous extracellular matrix
extracellular matrix
MHC protein complex
luminal side of endoplasmic reticulum membrane
integral component of luminal side of endoplasmic reticulum membrane
extracellular matrix structural constituent
peptide antigen binding
plasma membrane part
basal lamina
cell adhesion
detection of other organism
interferon-gamma-mediated signaling pathway
apical plasma membrane
MHC class I protein complex
detection of biotic stimulus
cytoskeleton organization
extracellular matrix organization
extracellular structure organization
ER to Golgi transport vesicle membrane
fibrillar collagen trimer
banded collagen fibril
LINC complex
microtubule organizing center attachment site
plasma membrane region
axonemal dynein complex
cytoskeletal anchoring at nuclear membrane
detection of external biotic stimulus
multicellular organism metabolic process
axoneme part
apical part of cell
MHC class II protein complex
ciliary part
cell-cell adhesion
homophilic cell adhesion via plasma membrane adhesion molecules
calcium ion binding
axoneme
ciliary plasm
response to interferon-gamma
notochord development
cilium
multicellular organismal catabolic process
regulation of synaptic growth at neuromuscular junction
outer dynein arm
MHC class II receptor activity
complex of collagen trimers
collagen trimer
endocytic vesicle membrane
integral component of plasma membrane

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

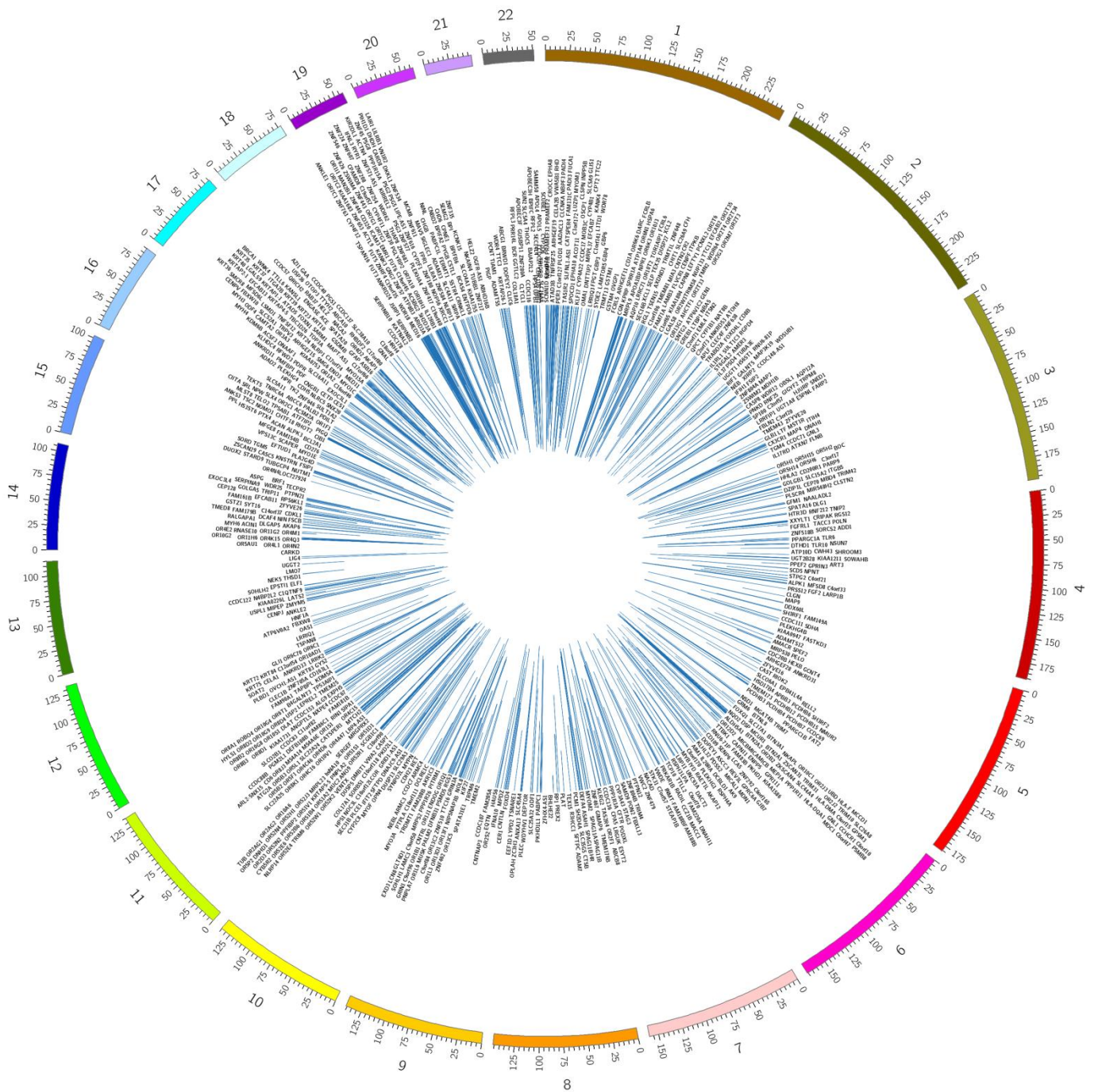


Figure S7. Distribution of *cis*- and *trans*-abundant genes across the autosomes (A) Distribution of *cis*-abundant genes.

B

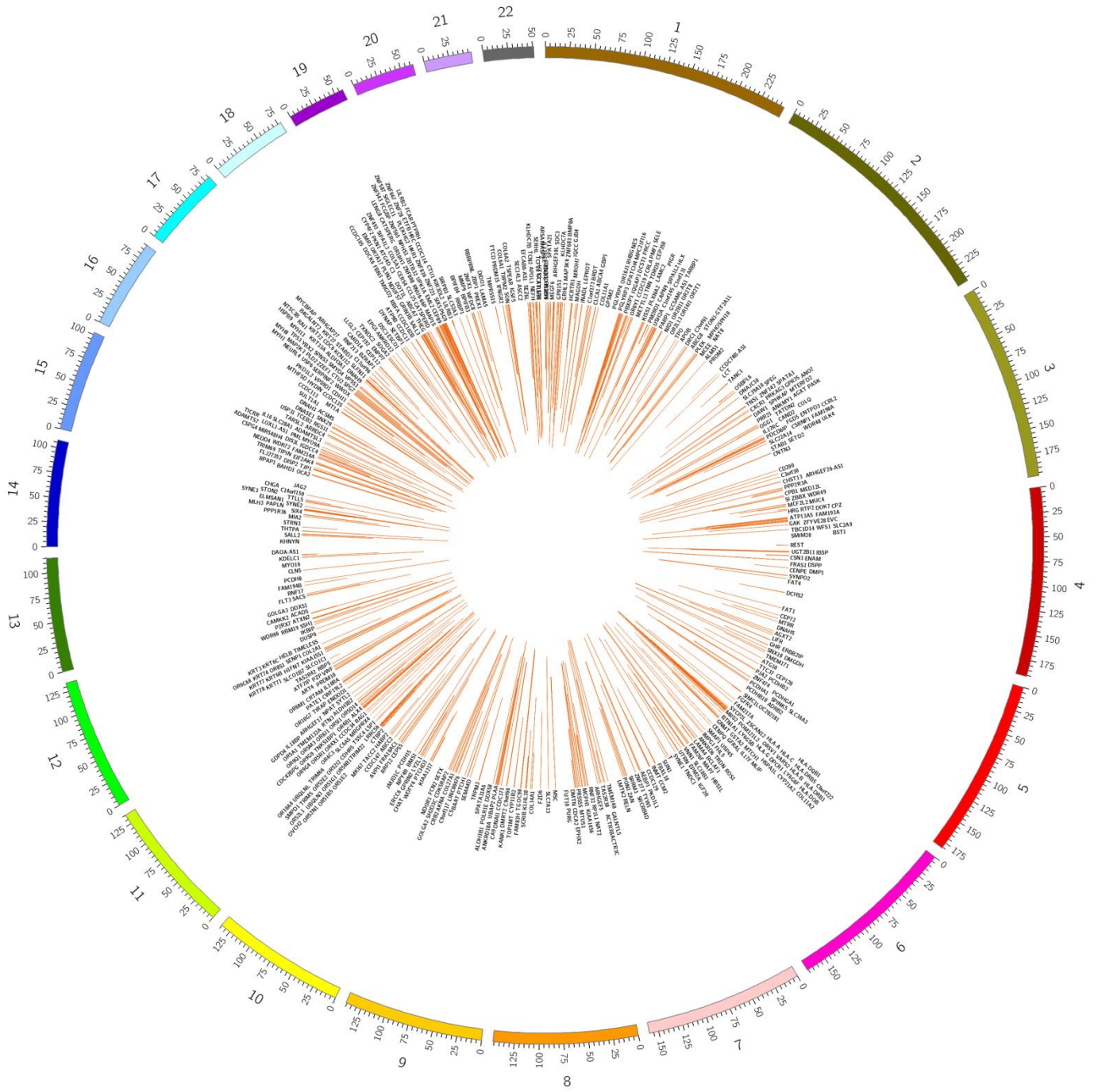


Figure S7. Distribution of *cis*- and *trans*-abundant genes across the autosomes
(B) Distribution of *trans*-abundant genes.