

Figure 1: Over-presentation of TMH-derived epitopes on most MHC-I and -II alleles (A) The percentage of epitopes for MHC-I and -II alleles that are predicted to overlap with TMHs for the proteomes of SARS-CoV-2 (top row), human (middle row) and *M. tuberculosis* (MtB; bottom row). The pair of horizontal red lines in each plot indicate the lower and upper bound of the 99% confidence interval. See supplementary Tables S5 and S7 for the exact TMH and epitope counts. **(B-C)** Correlation between the percentages of predicted TMH-derived epitopes and the hydrophobicity score of all predicted epitopes for human MHC-I **(B)** and MHC-II alleles **(C)**. Diagonal red line: linear regression analysis. Labels are shorthand for the HLA alleles, see the supplementary Table S8 for the names.

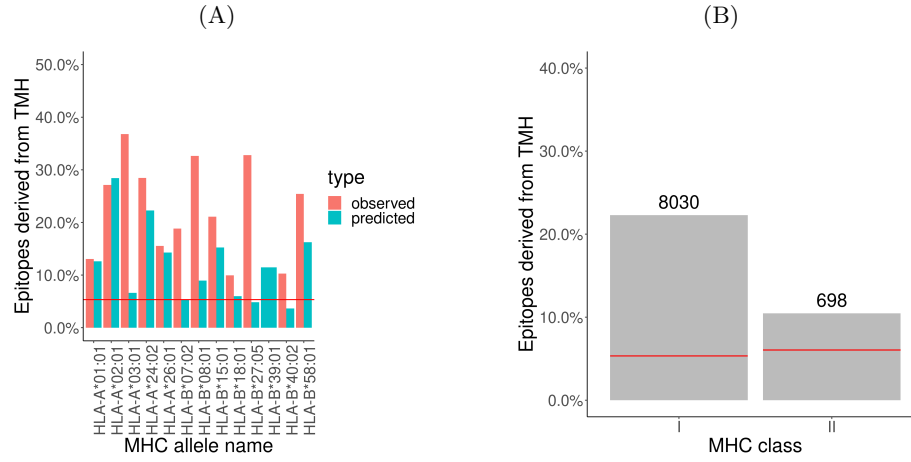


Figure 2: **Analysis of epitope database shows that TMH derived epitopes are over presented.** The percentage of epitopes for MHC-I and -II alleles that overlap with TMHs that are presented. The pair of horizontal red lines in each plot indicate the lower and upper bound of the 99% confidence interval. Note that only one line is visible as this interval is relatively narrow. Alleles are listed in Table S8). **(A)** Observed and predicted percentage of TMH-derived epitopes for MHC-I alleles. **(B)** MHC ligands from IEDB corresponding to TMH-derived epitopes. The numbers above the bars denotes the number of TMH derived epitopes obtained.

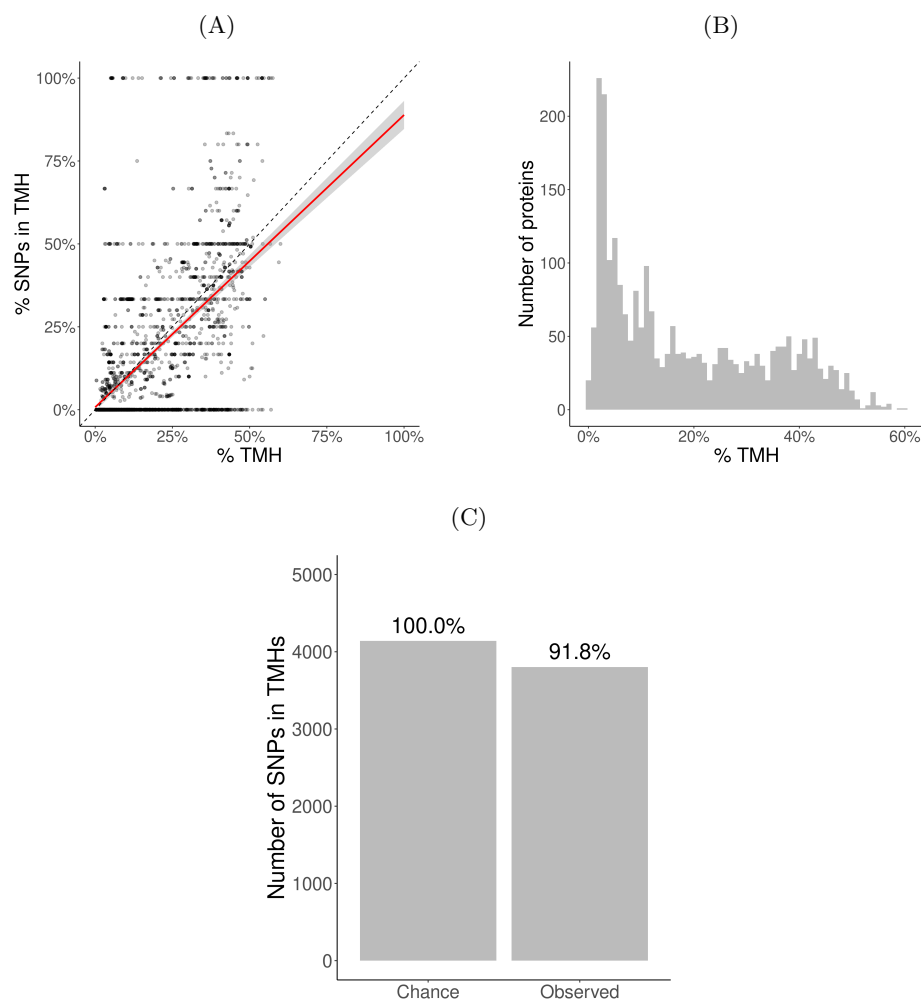


Figure 3: **Evolutionary conservation of human TMHs.** (A) Percentage of SNPs found in TMHs. Each point shows for one protein the predicted percentage of amino acids that are part of a TMH (x -axis) and the observed occurrence of SNPs being located within a TMH (y -axis). The dashed diagonal line shows the line of equality (i.e., equal conservation of TMHs and soluble protein regions). The diagonal red line indicates a linear fit, the gray area its 95% confidence interval. (B) Distribution of the percentages of TMH in the TMPs used in this study. (C) The number of SNPs in TMHs as expected by chance (left bar) and found in the dbSNP database (right bar). Percentages show the relative conservation of SNPs in TMHs found relative to stochastic chance.

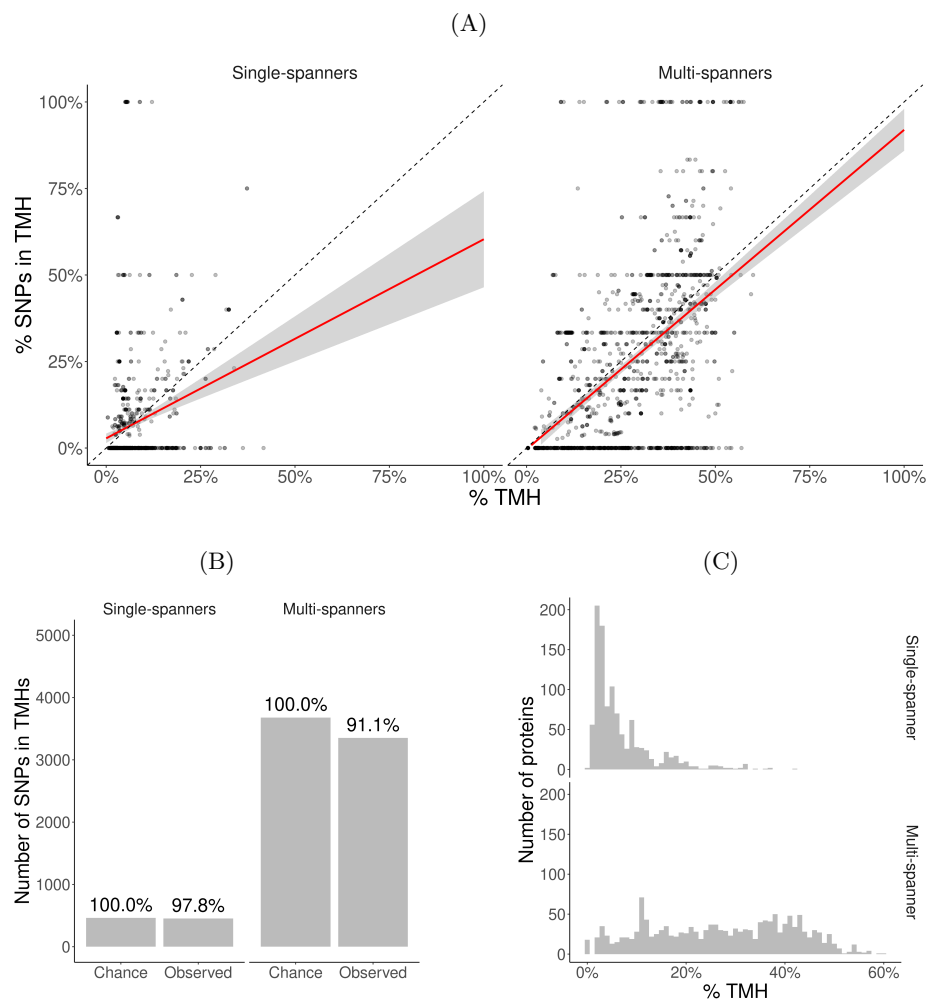


Figure 4: Membrane proteins with multiple TMHs are evolutionary more conserved than proteins with only a single TMH. (A) Percentage of SNPs found in TMPs predicted to have only a single (left) or multiple (right) TMHs. Each point shows for one protein the predicted percentage of amino acids that are part of a TMH (x -axis) and the observed occurrence of SNPs being located within a TMH (y -axis). The dashed diagonal lines show the line of equality (i.e., equal conservation of TMHs and soluble protein regions). The diagonal red lines indicate a linear fit, the gray areas their 95% confidence intervals. (B) The number of SNPs in TMHs as expected by chance and observed in the dbSNP database, for TMPs with one TMH (single-spanners) and multiple TMHs (multi-spanners). Percentages show the relative conservation of SNPs in TMHs found relative to the stochastic chances. (C) Distribution of the proportion of amino acids residing in the plasma membrane.