

**Figure 1: Over-presentation of TMH-derived epitopes on most MHC-I and -II haplotypes** (A) The percentage of epitopes for MHC-I and -II haplotypes that are predicted to overlap with TMHs for the proteomes of SARS-CoV-2 (top row), human (middle row) and *M. tuberculosis* (bottom row). The dashed lines indicate the percentages as expected by chance. See supplementary Tables 7 and 8 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 MHC-I (B) and MHC-II haplotypes (C). Diagonal line: linear regression analysis. Labels are shorthand for the HLA haplotypes, see the supplementary Table 6 for the names.

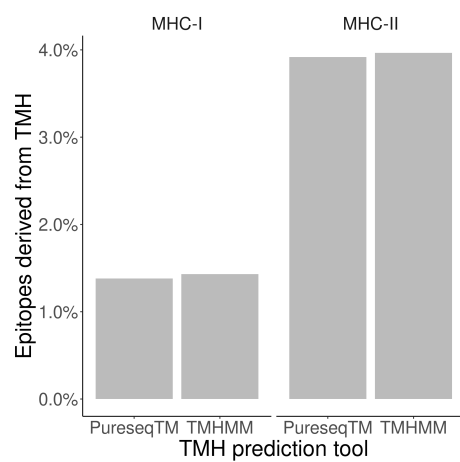


Figure 2: **Robust prediction that TMH epitopes are presented *in vivo*.** Bars show the percentage of peptides obtained from elution studies that is derived from a TMH, for MHC-I and -II, for two TMH prediction tools.

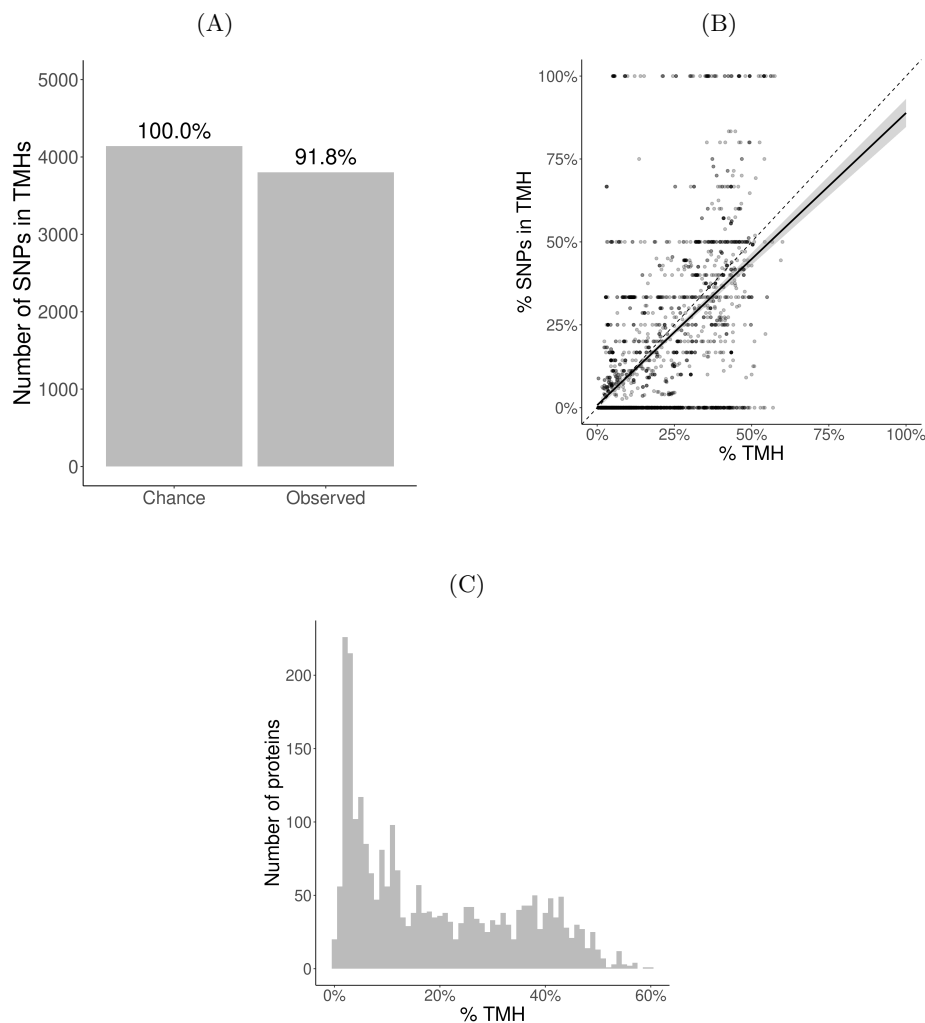
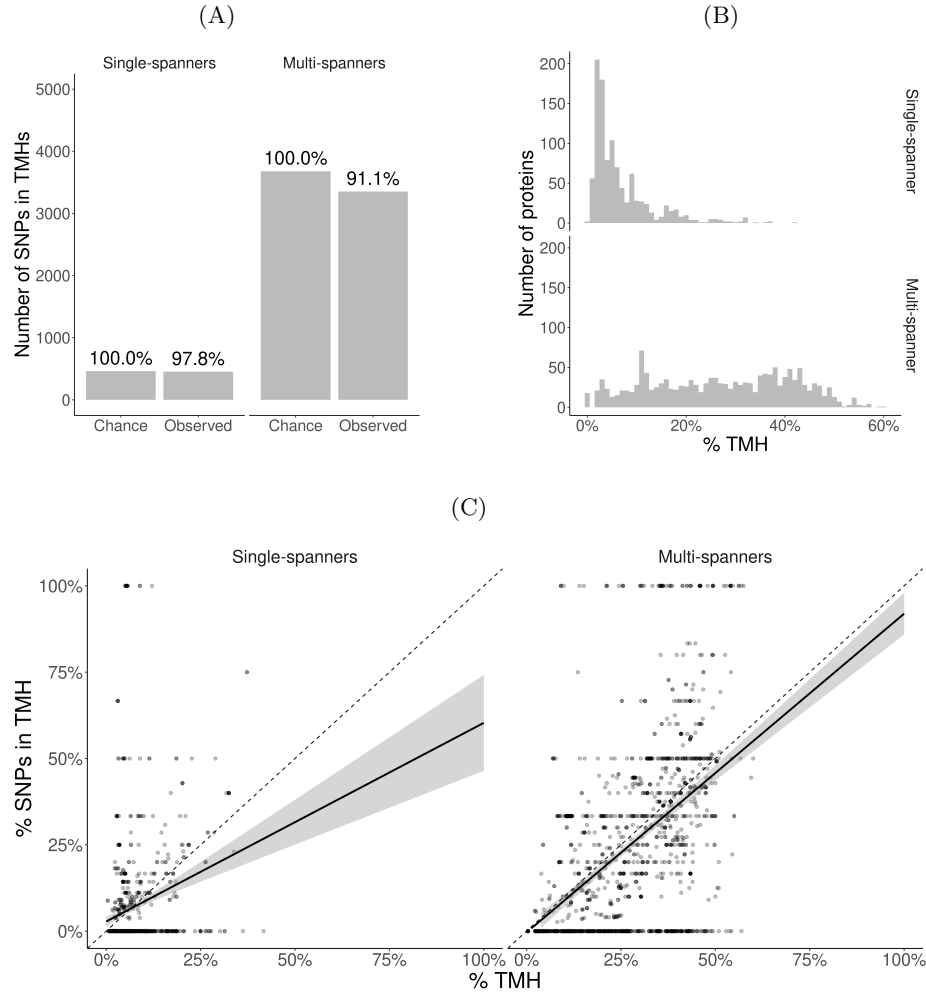


Figure 3: **Evolutionary conservation of human TMHs.** (A) The number of SNPs in TMHs as expected by chance (left bar) and found in the dbSNP database (right bar). Percentages show the relative conservation of SNPs in TMHs found. (B) Percentage of SNPs found in TMHs. Each point shows for one protein the predicted percentage of 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 line indicates a linear fit, the gray area its 95% confidence interval. (C) Distribution of the percentages of TMH in the TMPs used in this study.



**Figure 4: Membrane proteins with multiple TMHs are evolutionary more conserved than proteins with only a single TMH.** (A) 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. (B) Distribution of the proportion of amino acids residing in the plasma membrane. (C) 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 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 line indicates a linear fit, the gray area its 95% confidence interval.