Cell Reports Methods, Volume *4*

Supplemental information

SIGMA leverages protein structural information to predict the pathogenicity of missense variants

Hengqiang Zhao, Huakang Du, Sen Zhao, Zefu Chen, Yaqi Li, Kexin Xu, Bowen Liu, Xi Cheng, Wen Wen, Guozhuang Li, Guilin Chen, Zhengye Zhao, Guixing Qiu, Deciphering Disorders Involving Scoliosis & Comorbidities (DISCO) Study, Pengfei Liu, Terry Jianguo Zhang, Zhihong Wu, and Nan Wu

Supplemental figures

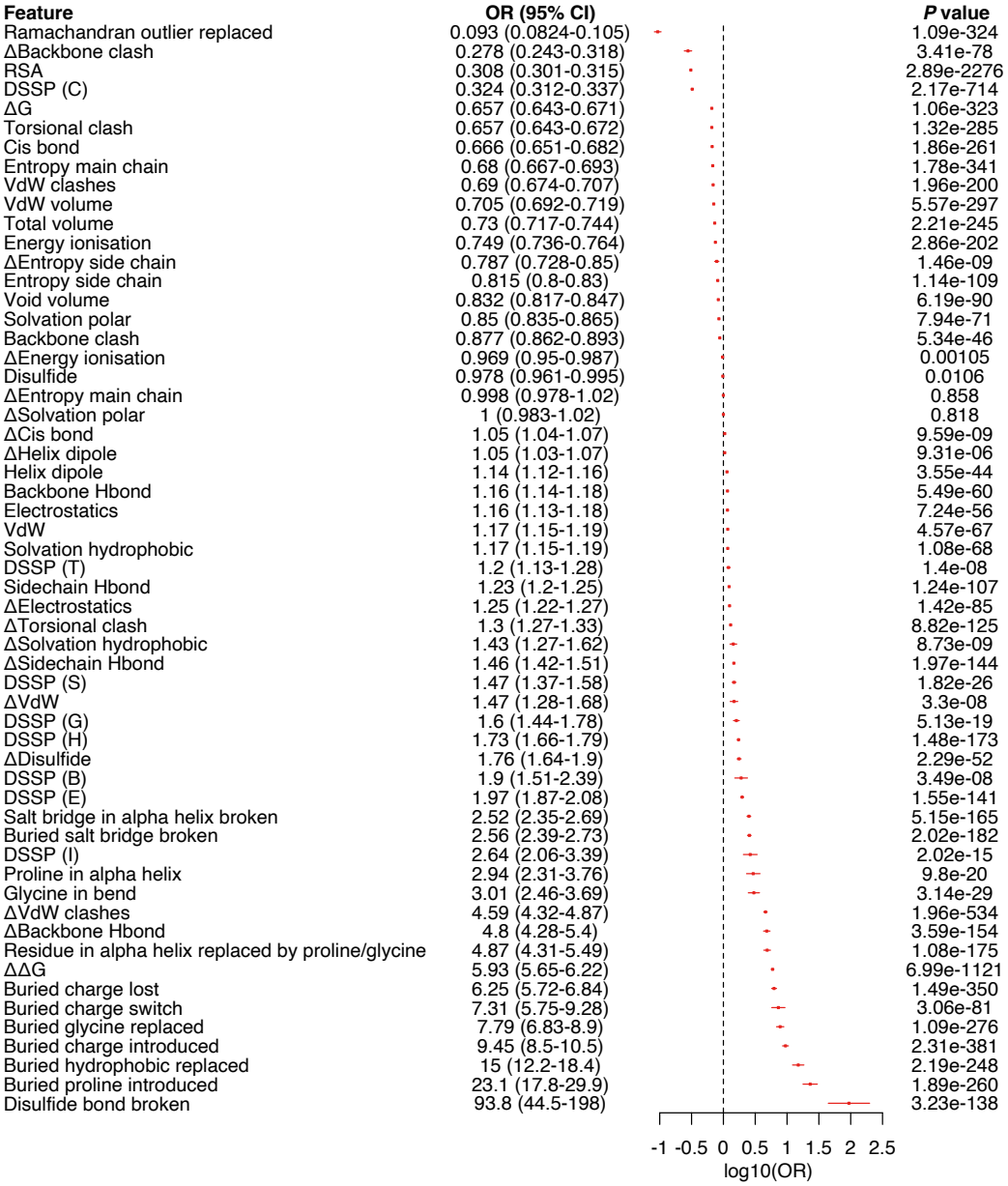

Figure S1: Forest plot for the associations of structure-informed features with variant pathogenicity. Center values represent OR, and error bars represent the 95% CI for the OR. P values were calculated by using Pearson's chi-squared test. RSA, relative solvent accessibility; ΔΔG, the unfolding free energy difference between the wild-type and mutant protein; ΔG, the unfolding free energy of the wild-type protein; OR, odds ratio; SS-bond, disulfide bond; DSSP codes: C, loop; H, alpha-helix; E, beta-sheet; S, bend; G, 3-helix; I, 5-helix; T, hydrogen-bonded turn; B, residue in isolated beta-bridge; VdW, Van der Waals. Related to Figure 2.

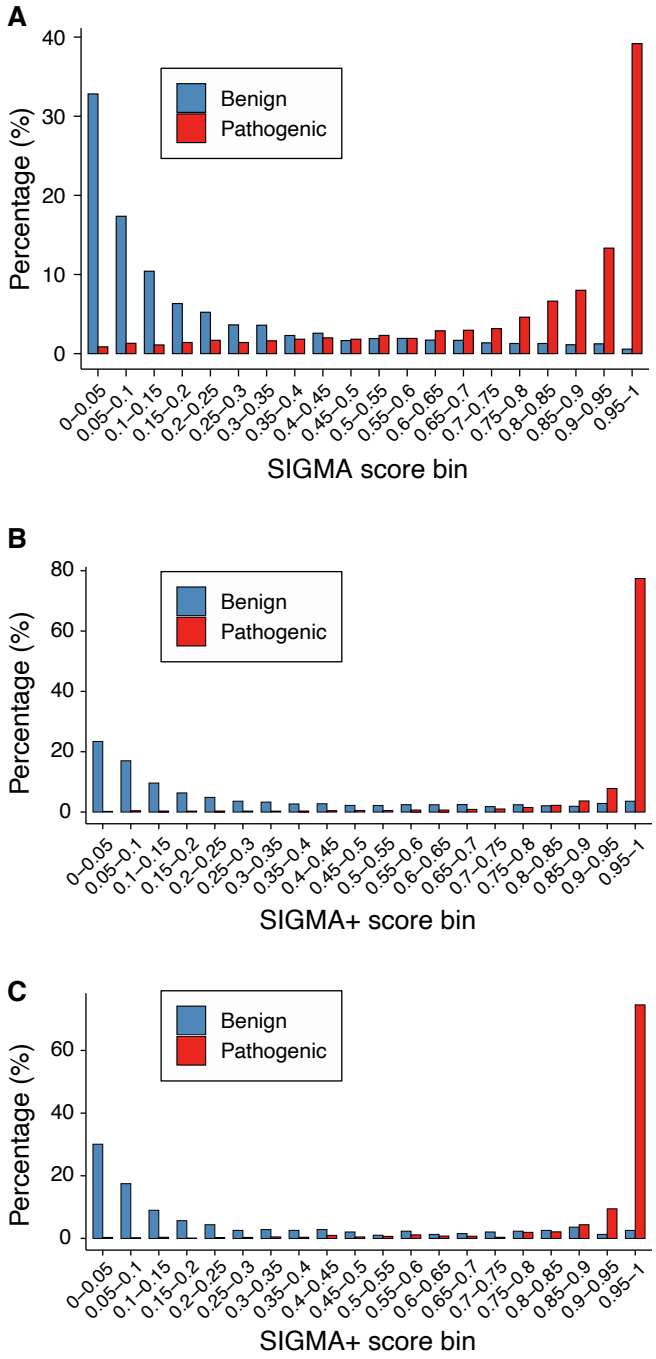

Figure S2: The distribution of SIGMA (SIGMA+) scores for pathogenic (red) and benign (blue) variants. (A) SIGMA scores in the test set. (B) SIGMA+ scores in the training set. (C) SIGMA+ scores in the test set. Related to Figure 3 and Figure 7.

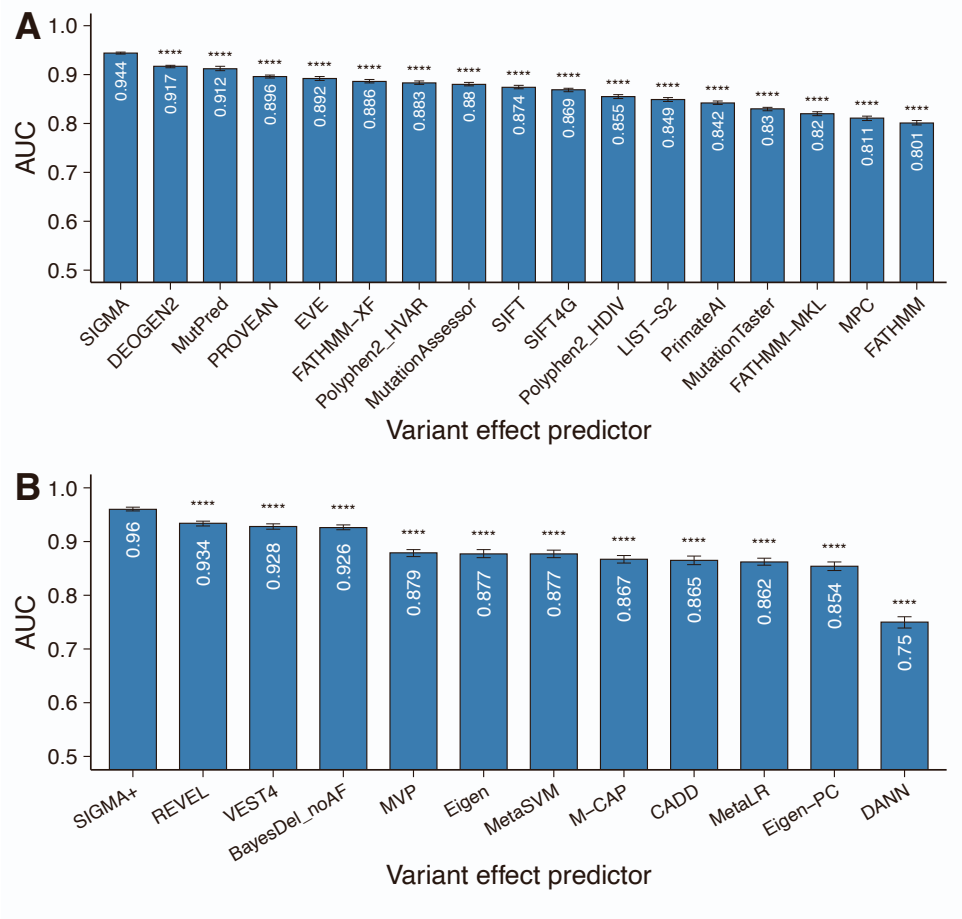

Figure S3: The performance of predictors on the training set. Related to Figure 3.

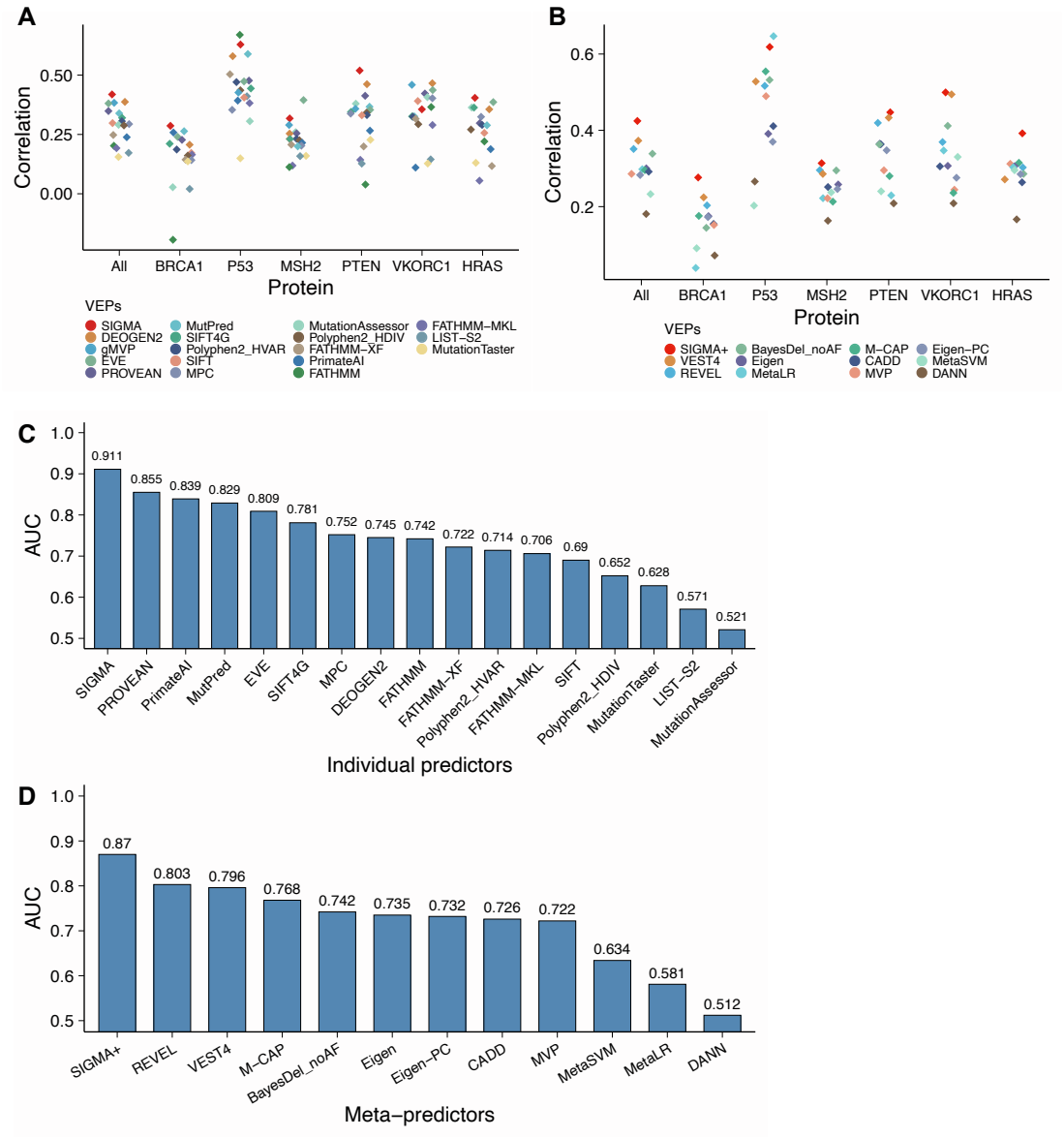

Figure S4: Performance of variant effect predictors on deep mutational scanning (DMS) datasets. (A) Correlation between individual predictors and DMS measurements. (B) Correlation between meta-predictors and DMS measurements. (C) The area under the ROC curves (AUCs) of individual predictors on the BRCA1 DMS dataset. (D) The AUCs of meta-predictors on the BRCA1 DMS dataset. Spearman's correlation was calculated between functional scores from DMS experiments and prediction scores from predictors. Related to Figure 3 and Figure 7.

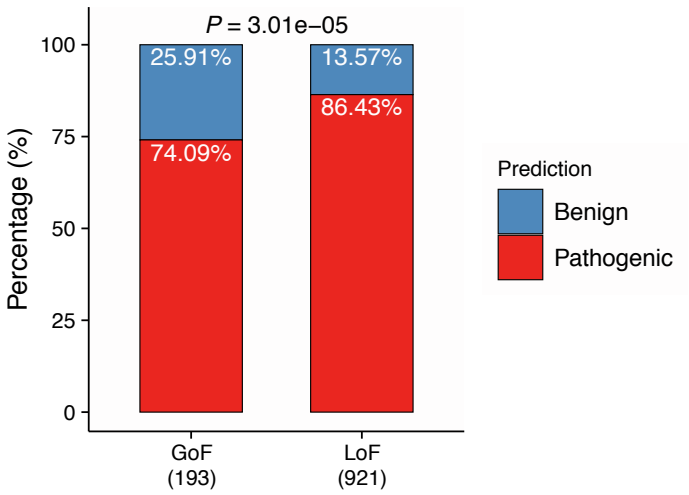

Figure S5: SIGMA predictions for 193 GoF variants and 921 LoF variants. Related to Figure 5.

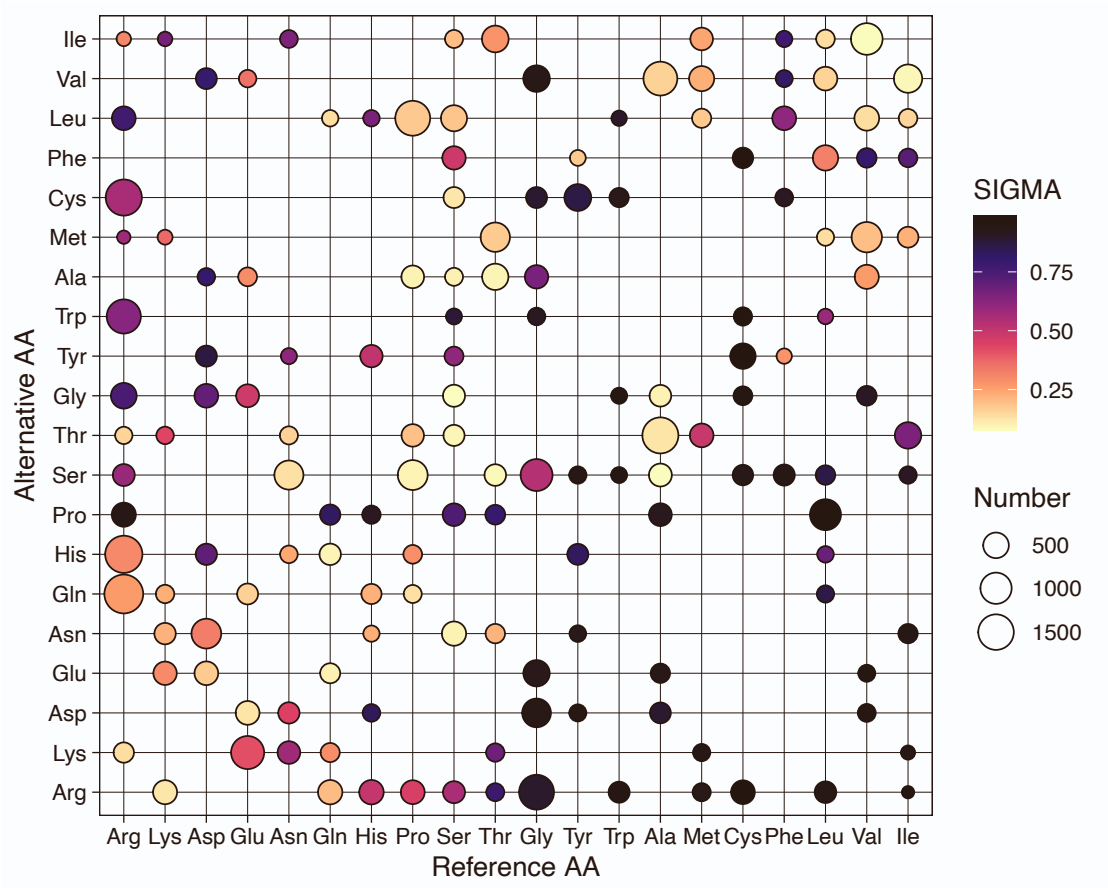

Figure S6: Pathogenicity matrix for each type of amino acid substitution in the labeled dataset. Related to Figure 6.

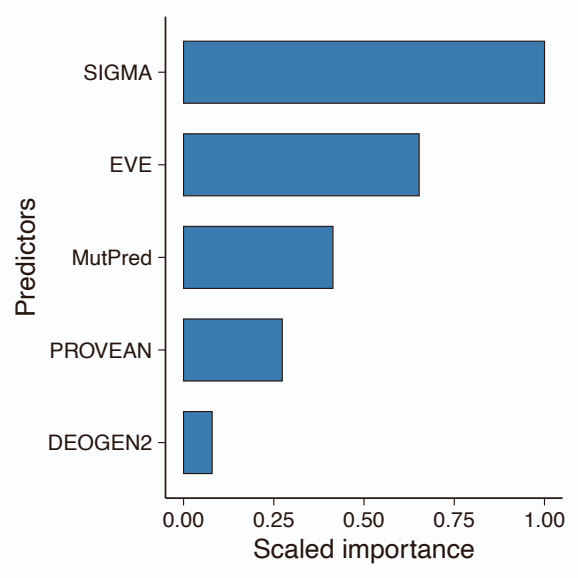

Figure S7: The importance of five predictors that contributed to SIGMA+. Related to Figure 7.

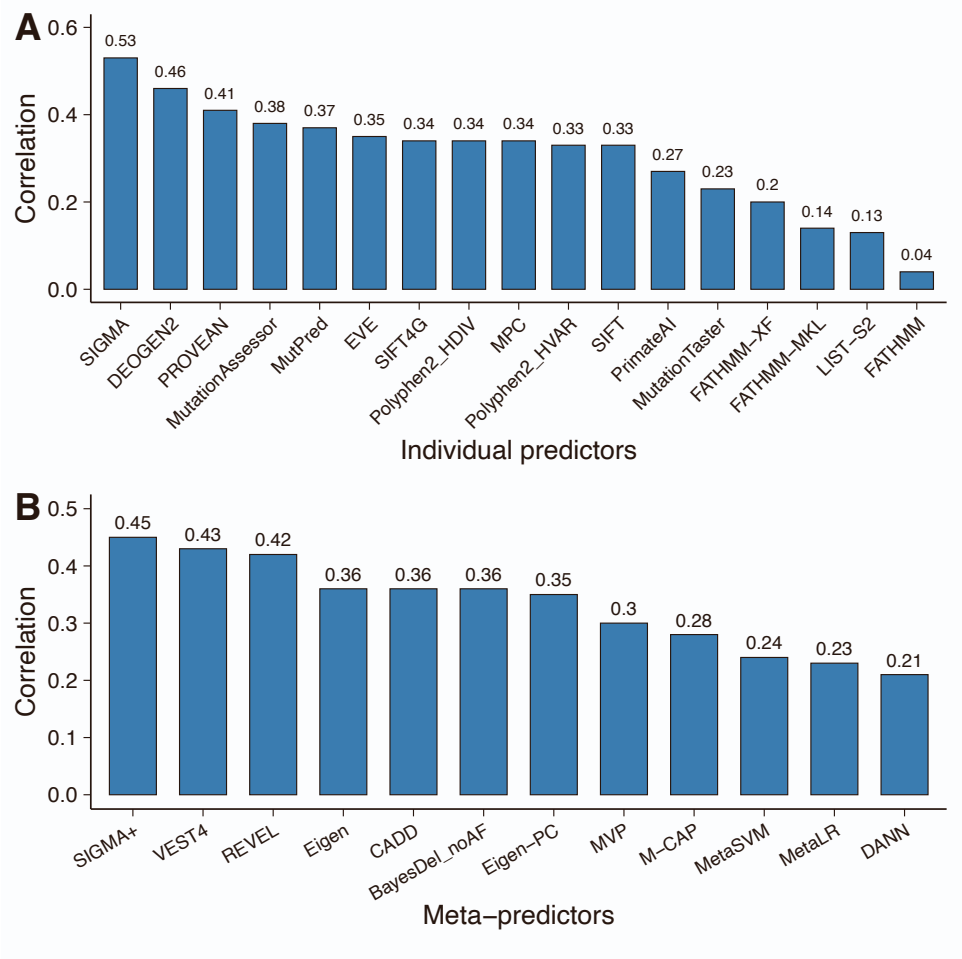

Figure S8: The correlations between the prediction scores and the DMS scores measuring the protein abundance of PTEN. Related to STAR Methods.