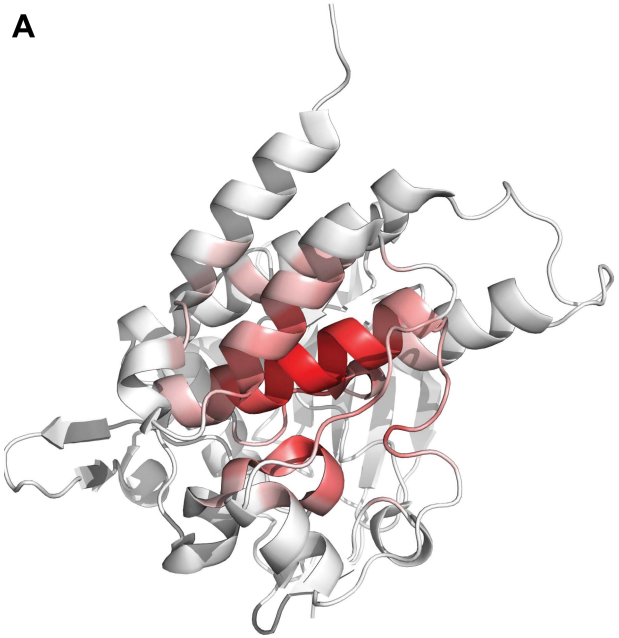
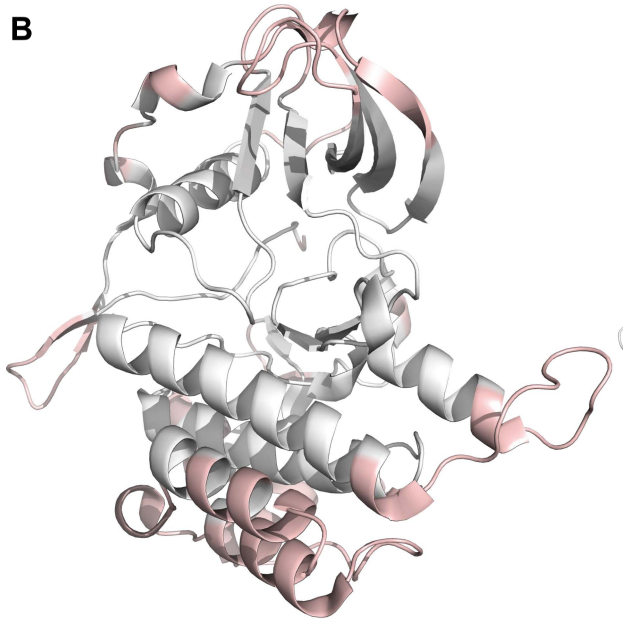
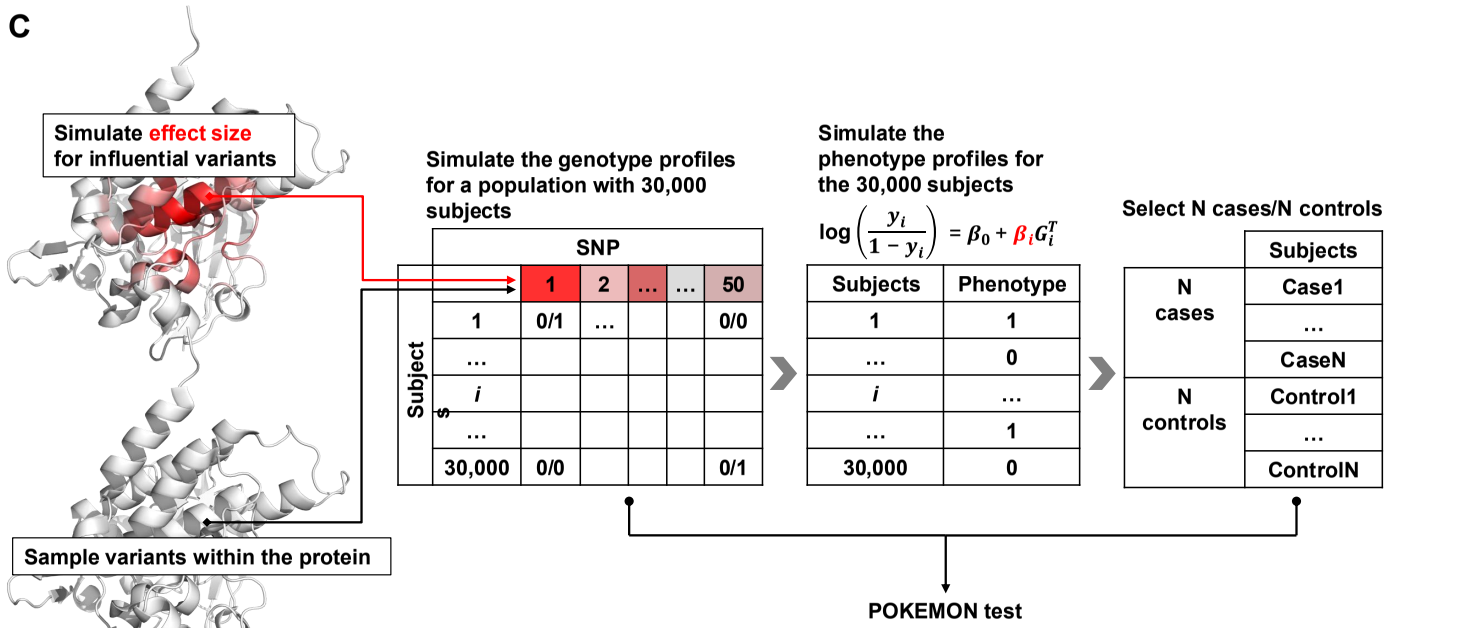
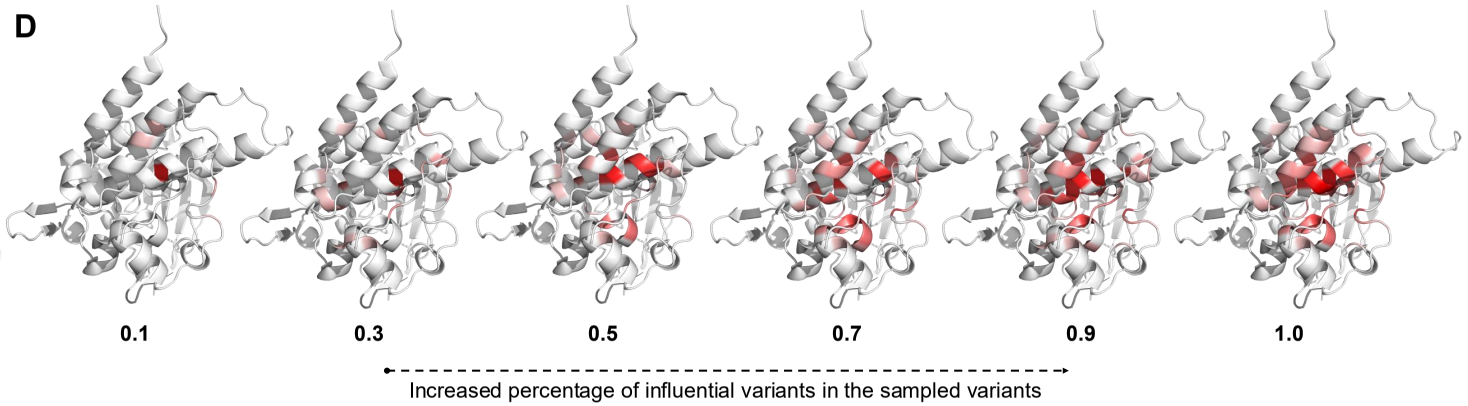
SUPPLEMENTAL FIGURES



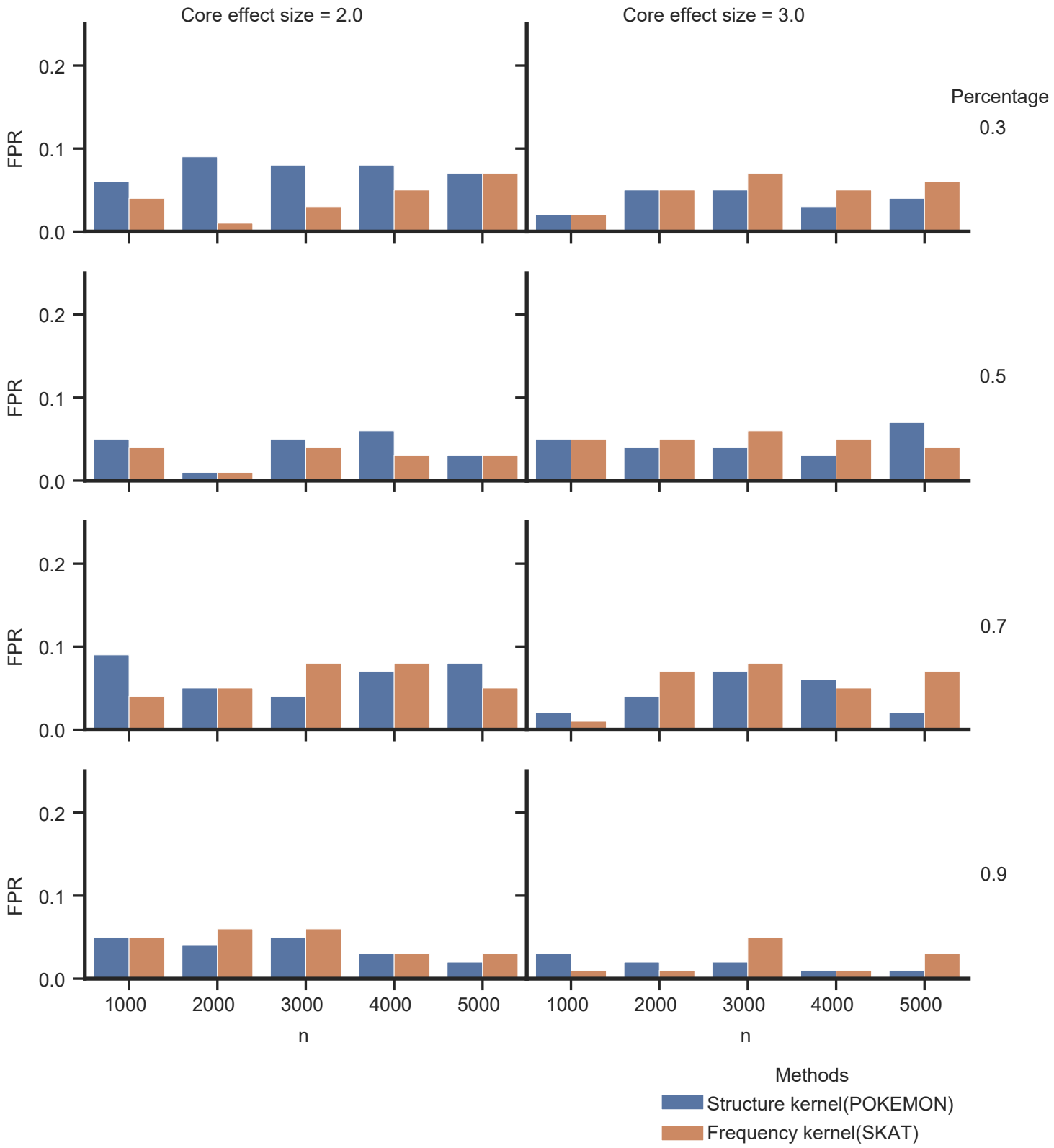






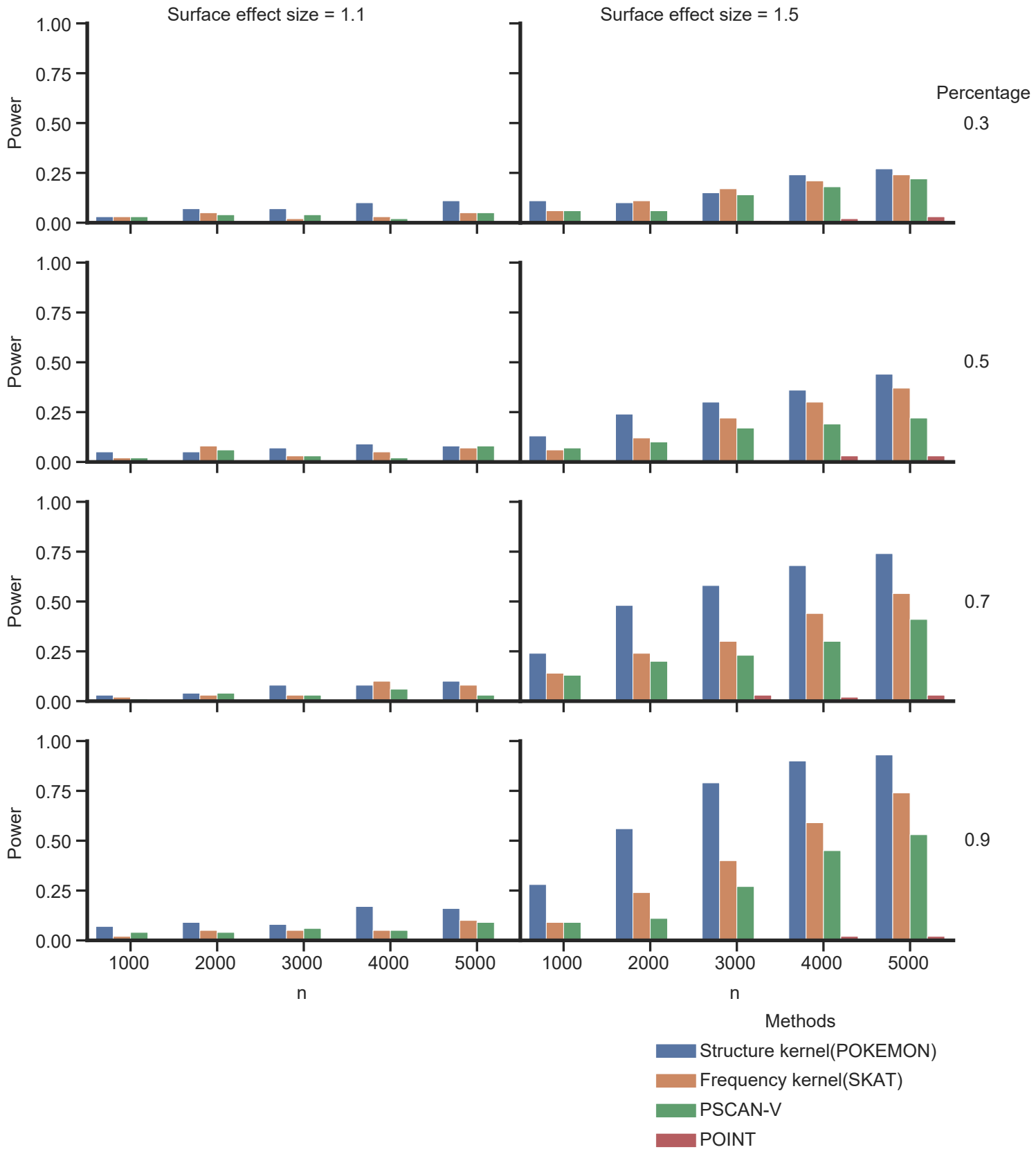
Supplementary Fig S1

Supplementary Fig S1. The model setup and simulation workflow for power simulation with POKEMON. (A) The simulated odds ratio for the clustering pattern. The odds ratio radiates from the core variant with an exponential function. (B) The simulated odds ratio for dispersion pattern. Lighter red indicates a smaller odds ratio. The selection of shell variants is based on their distance from the core. Variants that are within 21 Å from the core are considered neutral and not colored. (C) The simulation workflow. (D) The increased percentage of influential variants shown with shades formed a better-characterized core pattern.



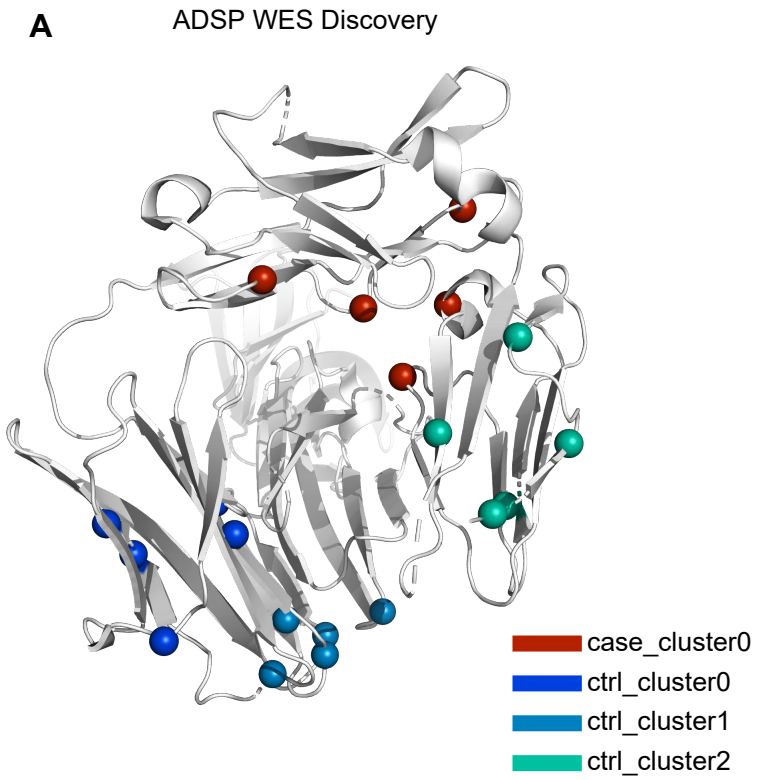
Supplemental Fig S2

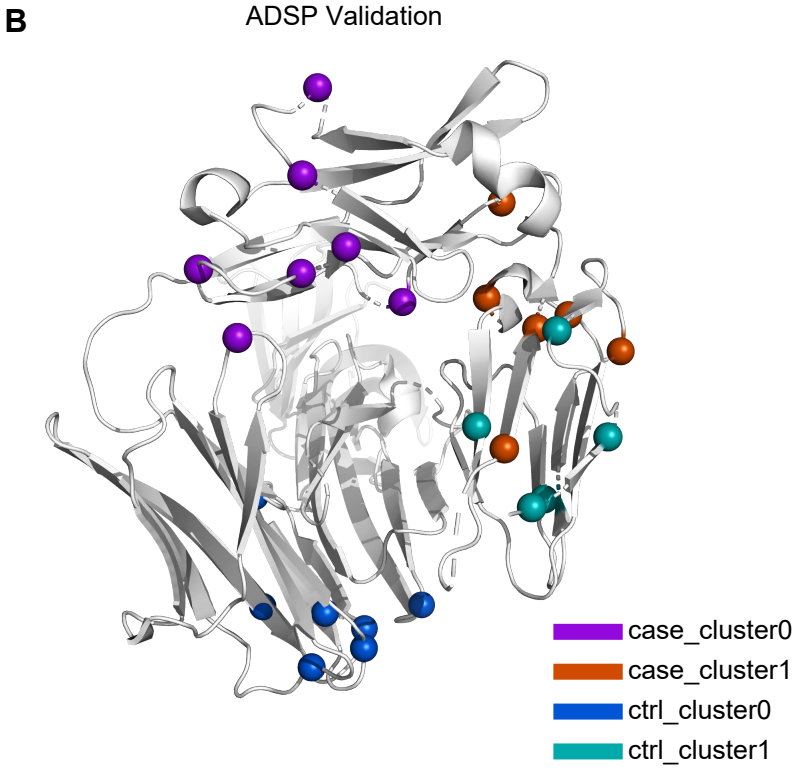
Supplementary Fig S2. (A) The empirical False Positive Rate for detecting the association between the phenotype and a core pattern on the protein between structure kernel (POKEMON) and frequency kernel (SKAT). While the phenotype file was generated according to the core pattern, we shuffled the variants' position on the protein. The empirical background false positive rate is calculated by the percentage of tests with a p-value below the significance level out of 100 replicates.



Supplemental Fig S3

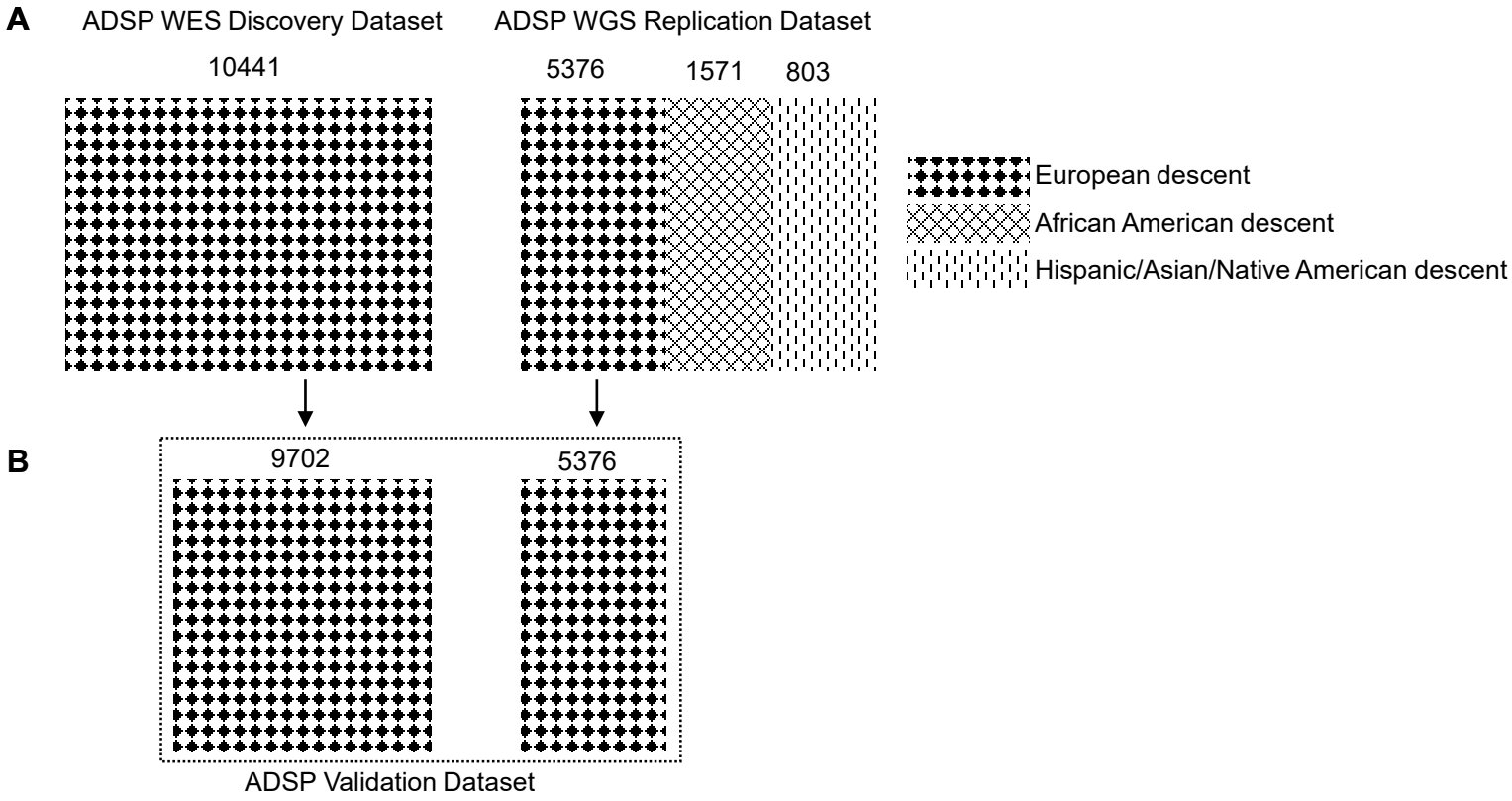
Supplementary Fig S3. (A) The empirical power for detecting the association between the phenotype and a dispersion pattern on the protein among structure kernel (POKEMON), frequency kernel (SKAT), PSCAN with variance (PSCAN-V), and POINT. The percentage of pathological variants within the selected 50 variants ranges from 0.3 to 0.9 (top to bottom). The simulated phenotype is calculated based on the surface variant odds ratio and the percentage of pathological variants. The empirical power is calculated by the percentage of tests with a p-value below the significance level out of 100 replicates.





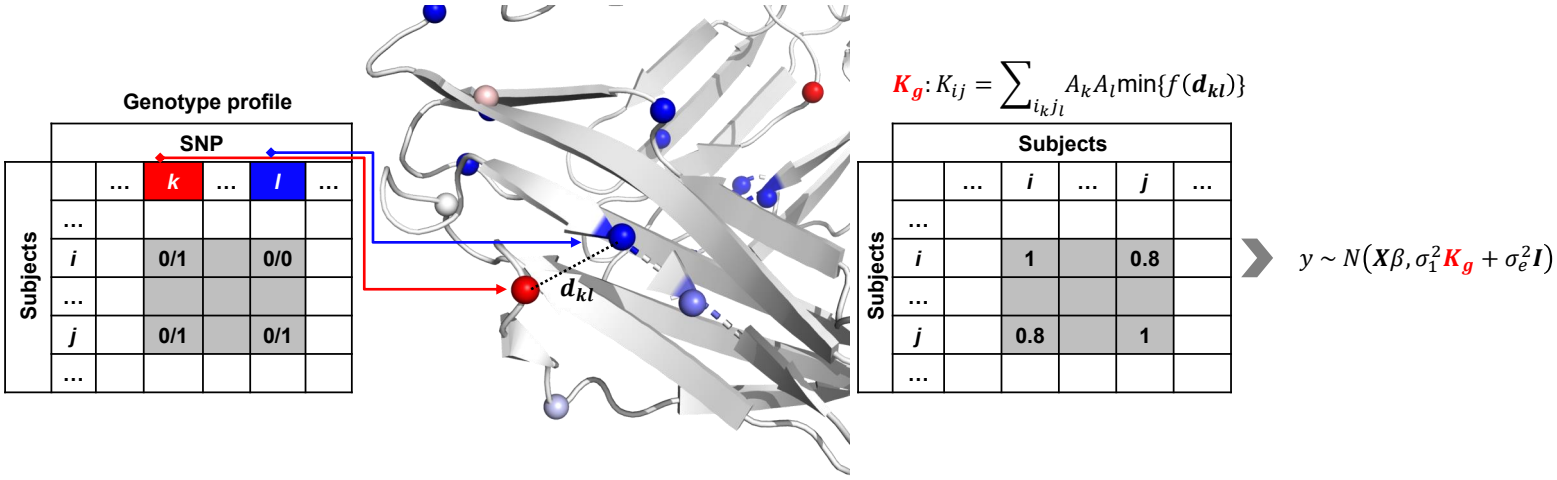
Supplemental Fig S4

Supplementary Fig S4. (A) *CSF1R* has a signal region identified in the ADSP WES discovery dataset (B) The signal region is replicated only in the ADSP validation dataset but not in the ADSP WGS replication dataset.



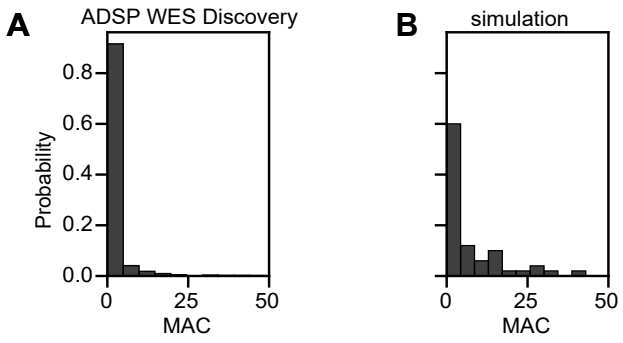
Supplemental Fig S5

Supplementary Fig S5. Graphic demonstrations of the composition for ADSP WES Discovery, ADSP WGS Replication, and ADSP validation datasets. (A) ADSP WES Discovery Dataset only has European descent subjects. ADSP WGS Replication Dataset contains multi-ancestry individuals. (B) ADSP validation dataset has 9,702 European descent subjects from the ADSP WES Discovery Dataset and 5,376 European descent subjects from the ADSP WGS Replication Dataset. The genotype for 9,702 European descent subjects is reprocessed and recalled with an updated genotype calling process implemented for the ADSP WES Discovery Dataset.



Supplemental Fig S6

Supplementary Fig S6. (A) POKEMON workflow



Supplemental Fig S7

Supplementary Fig S7. The simulated minor allele count resembles the minor allele count in the ADSP WES discovery dataset. (A) Minor allele count (MAC) is calculated with randomly selected 40 genes from the ADSP WES discovery dataset. (B) MAC is calculated with 40 independent simulations calculated in core pattern simulation.