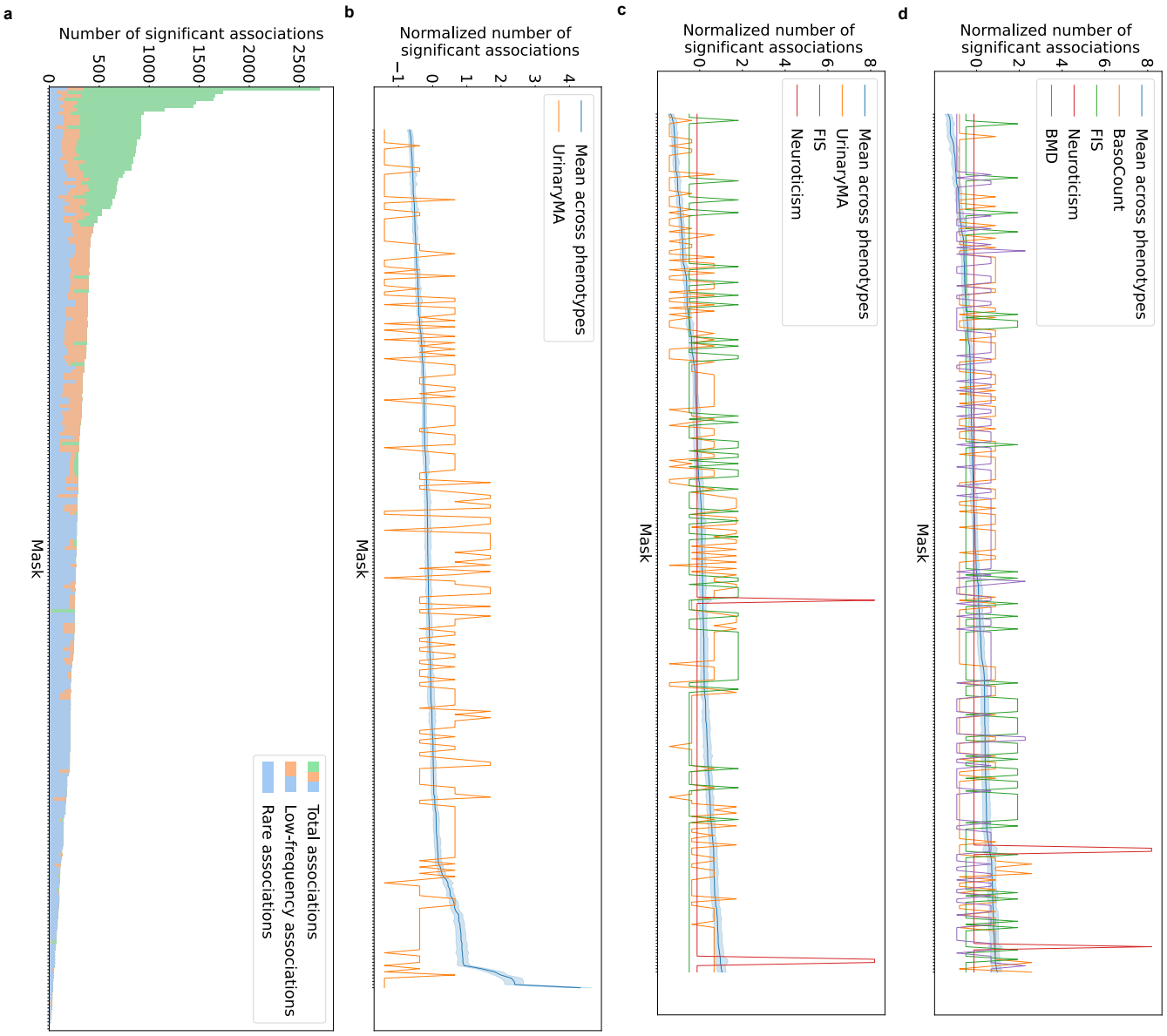
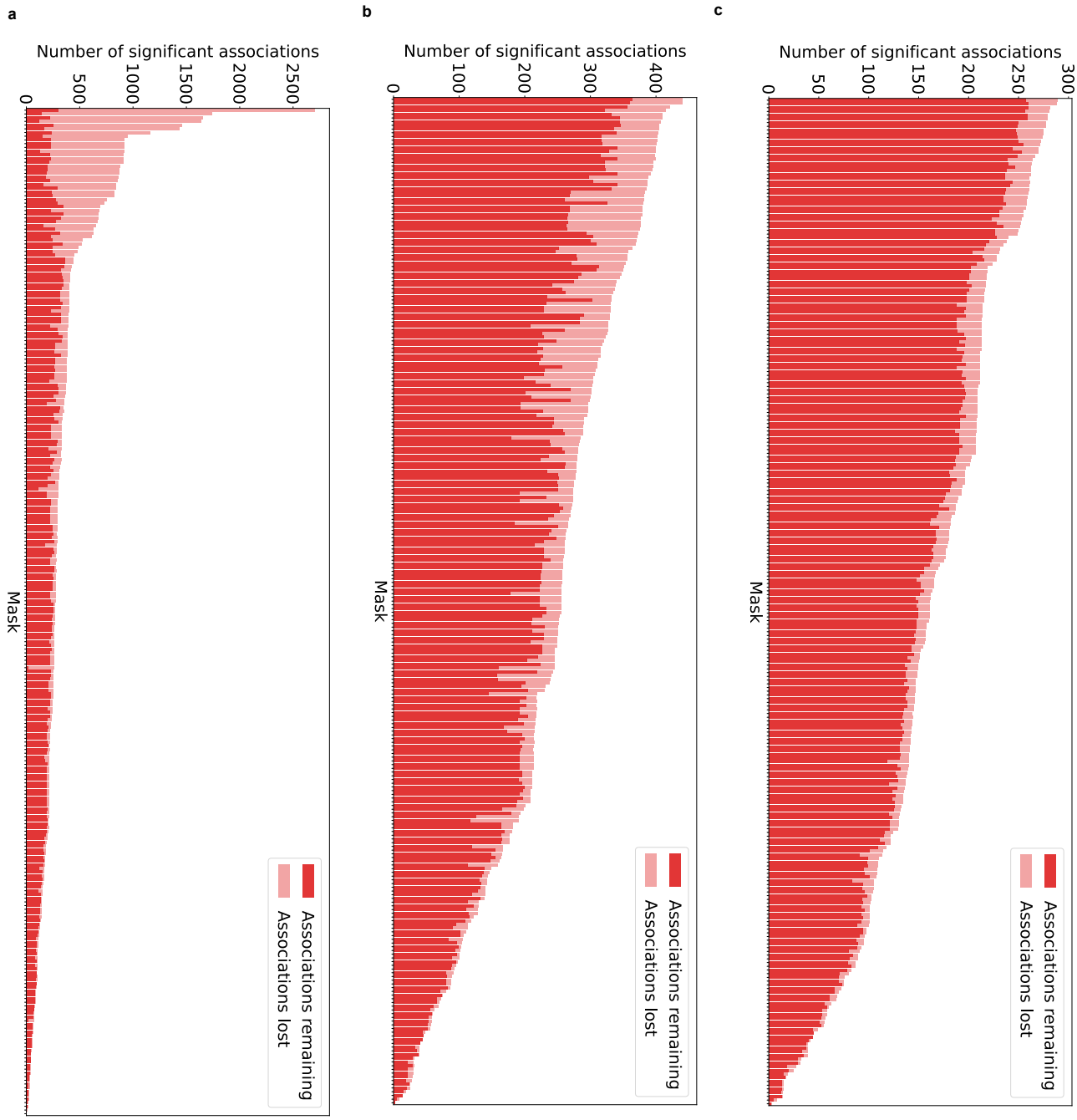
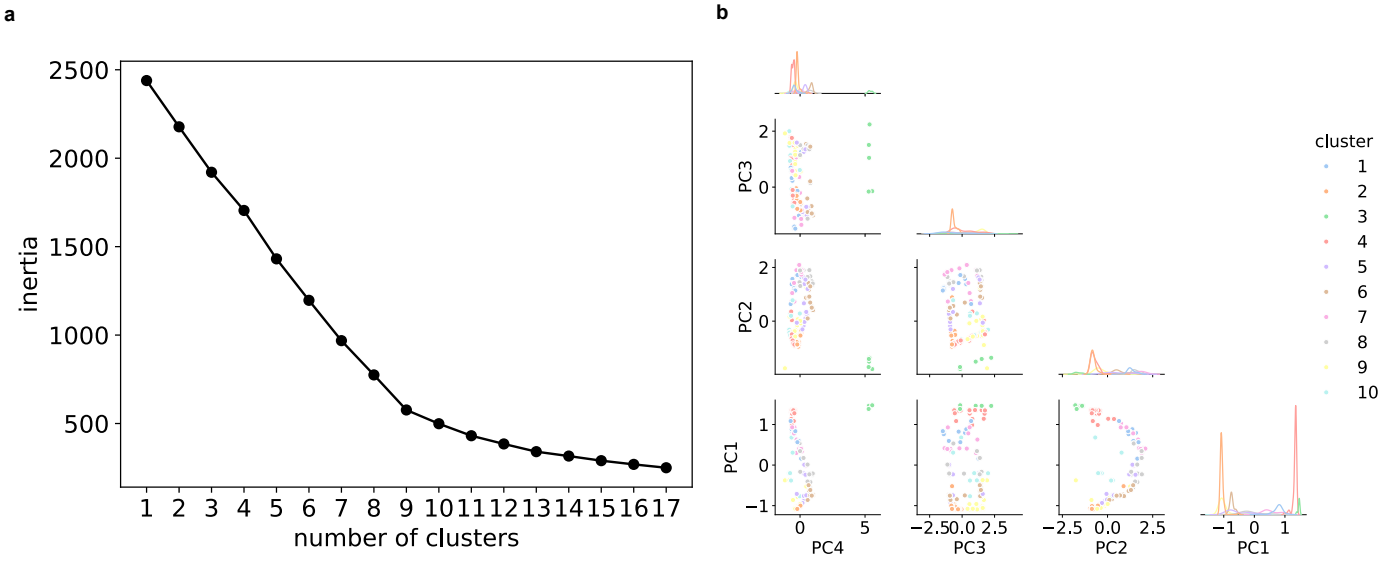
Supplementary Figure 1. Summary of the number of significant associations across 54 phenotypes. (a) Number of significant associations detected by each mask (**Supplementary Table 7a)**. Blue bars represent rare associations, blue + orange bars low-frequency associations, blue + orange + green bars total associations. **(b)** Normalized number of total significant associations in each mask for the average and outlier phenotypes. **(c)** Normalized number of low-frequency significant associations in each mask for the average (blue line) and UrinaryMA (orange line), fluid intelligence score (FIS, green line), neuroticism (red line) as outlier phenotypes. **(d)** Normalized number of rare significant associations in each mask for the average (blue line) and basophil count (BasoCount, orange line), fluid intelligence score (FIS, green line), neuroticism (red line), estimated heel bone mineral density (BMD, purple line) as outlier phenotypes.



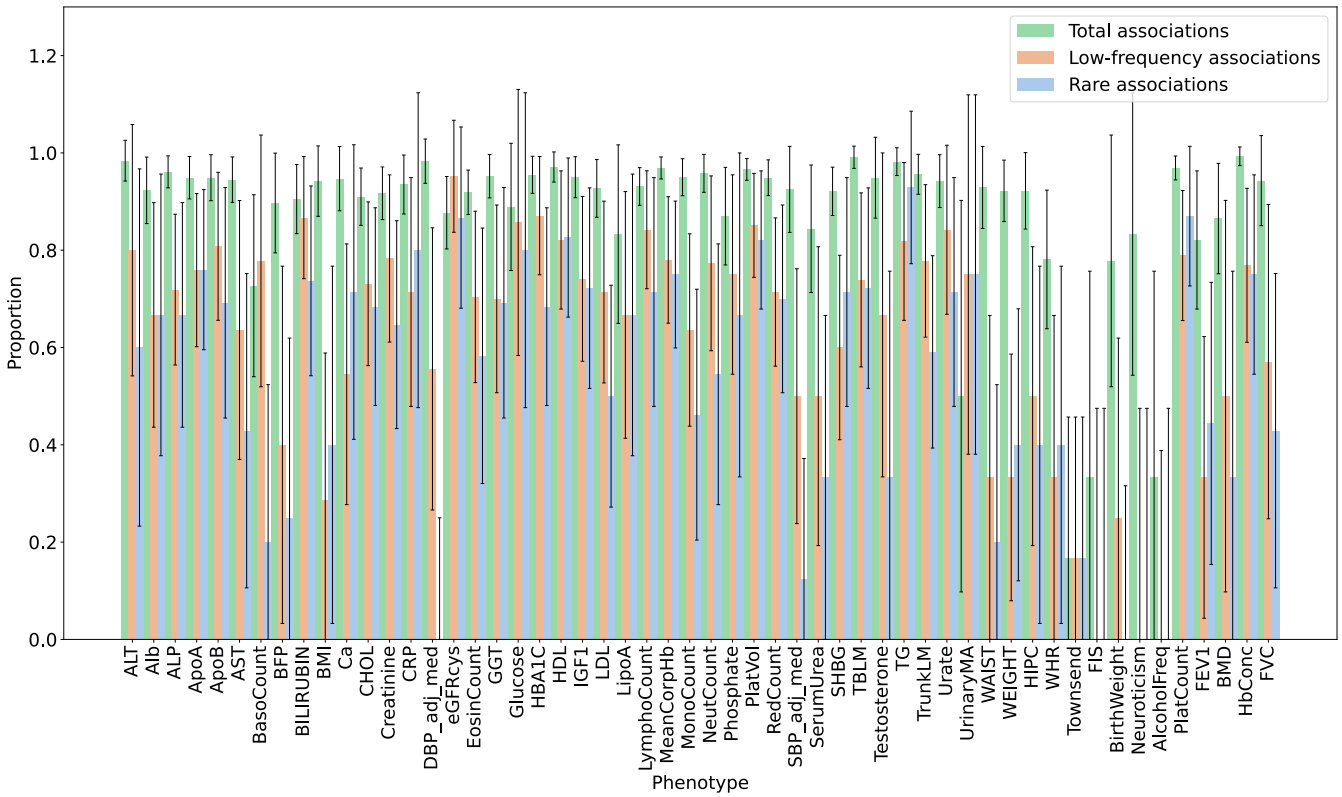
Supplementary Figure 2. Number of significant associations detected by each mask across 54 phenotypes before and after conditional analysis on common variants (Supplementary Table 7b). Dark red bars represent the number of significant associations remaining and light red bars represent the number of associations that were no longer significant after including common variants as co-variates in the burden tests. **(a)** Total associations. **(b)** Low-frequency associations. **(c)** Rare associations.



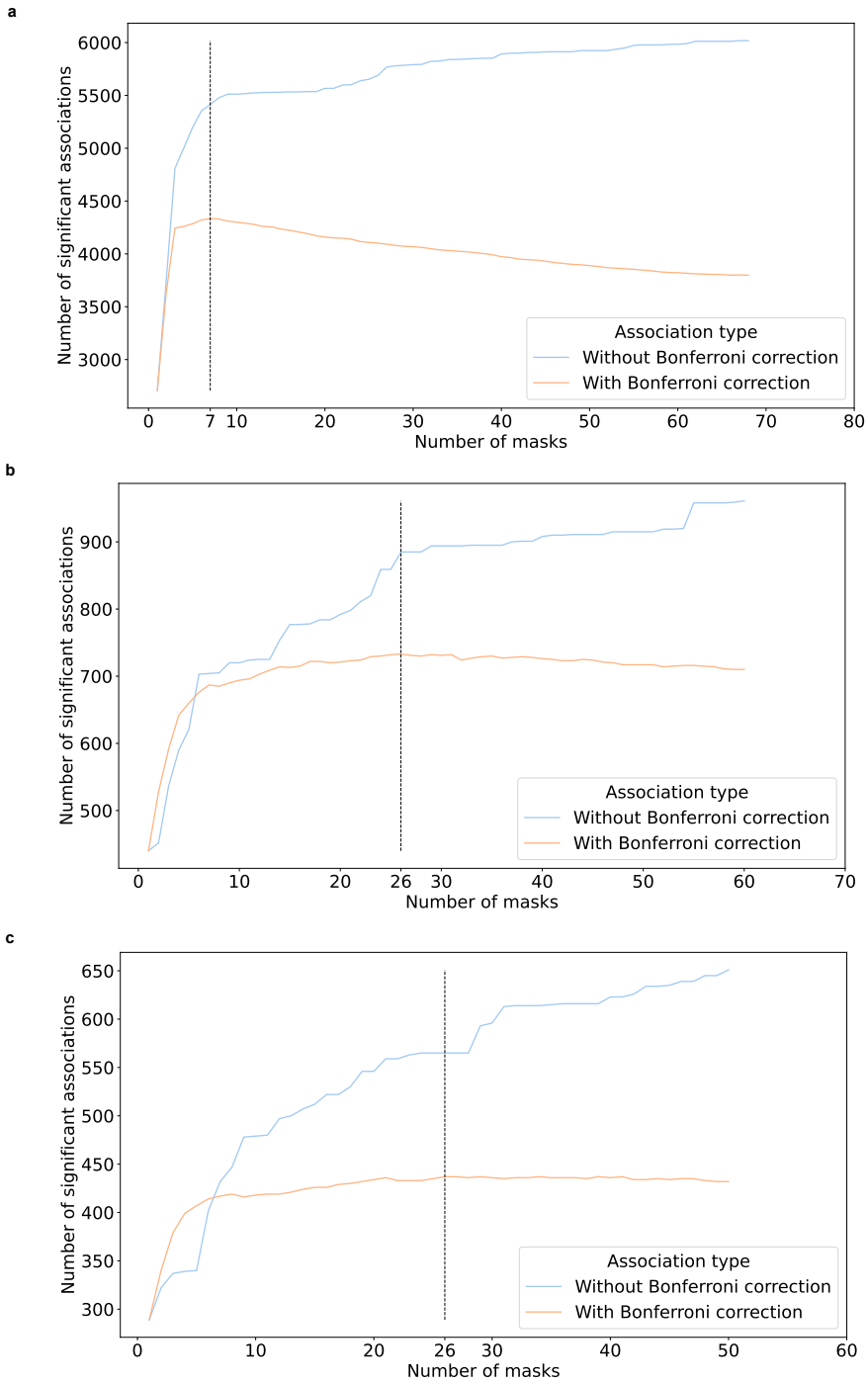
Supplementary Figure 3. Principal component (PC) analysis for 271 masks based on binary variant inclusion. (a) The plot shows the k-means clustering model's inertia decreasing less rapidly after 10 clusters, suggesting 10 as an appropriate number of clusters. **(b)** The plot shows the first 4 PCs of 271 masks grouped into 10 clusters by k-means clustering (**Supplementary Table 9b**).



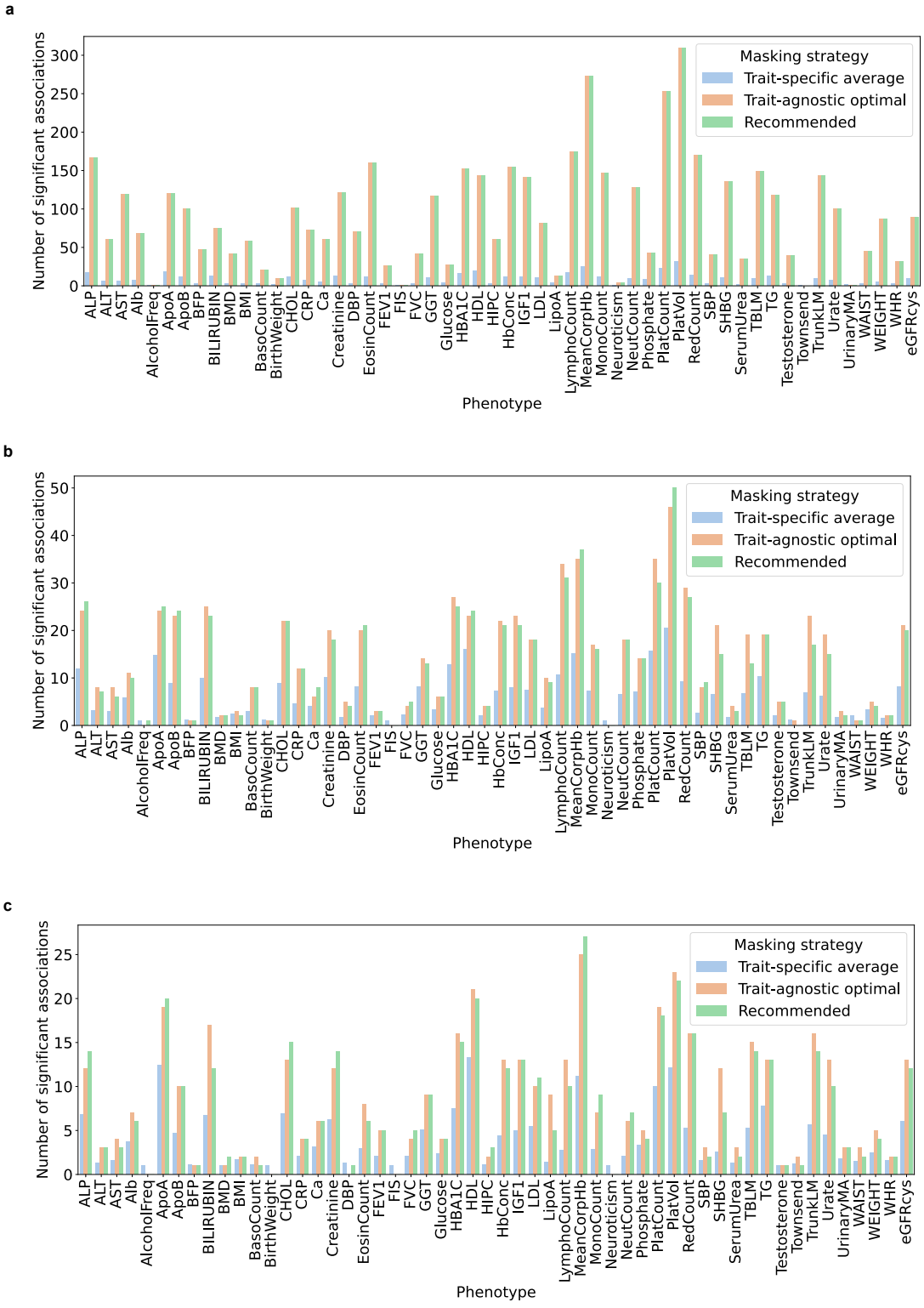
Supplementary Figure 4. Cross-validation analysis of the greedy covering method. For each phenotype ***p***, we ran the greedy covering method to obtain the trait-specific optimal masking strategies which produced the maximum number of significant associations. Next, we applied the greedy covering method to the rest of the 53 phenotypes to obtain the 53-trait optimal masking strategies, calculated the number of significant associations for phenotype ***p*** these strategies produced, and calculated the proportion compared to the maximum number of significant associations detected by the trait-specific strategies (**Supplementary Table 11)**. Green bars represent total associations, orange bars low-frequency associations, and blue bars rare associations. Error bars represent 95% confidence intervals for the proportion following a binomial distribution.



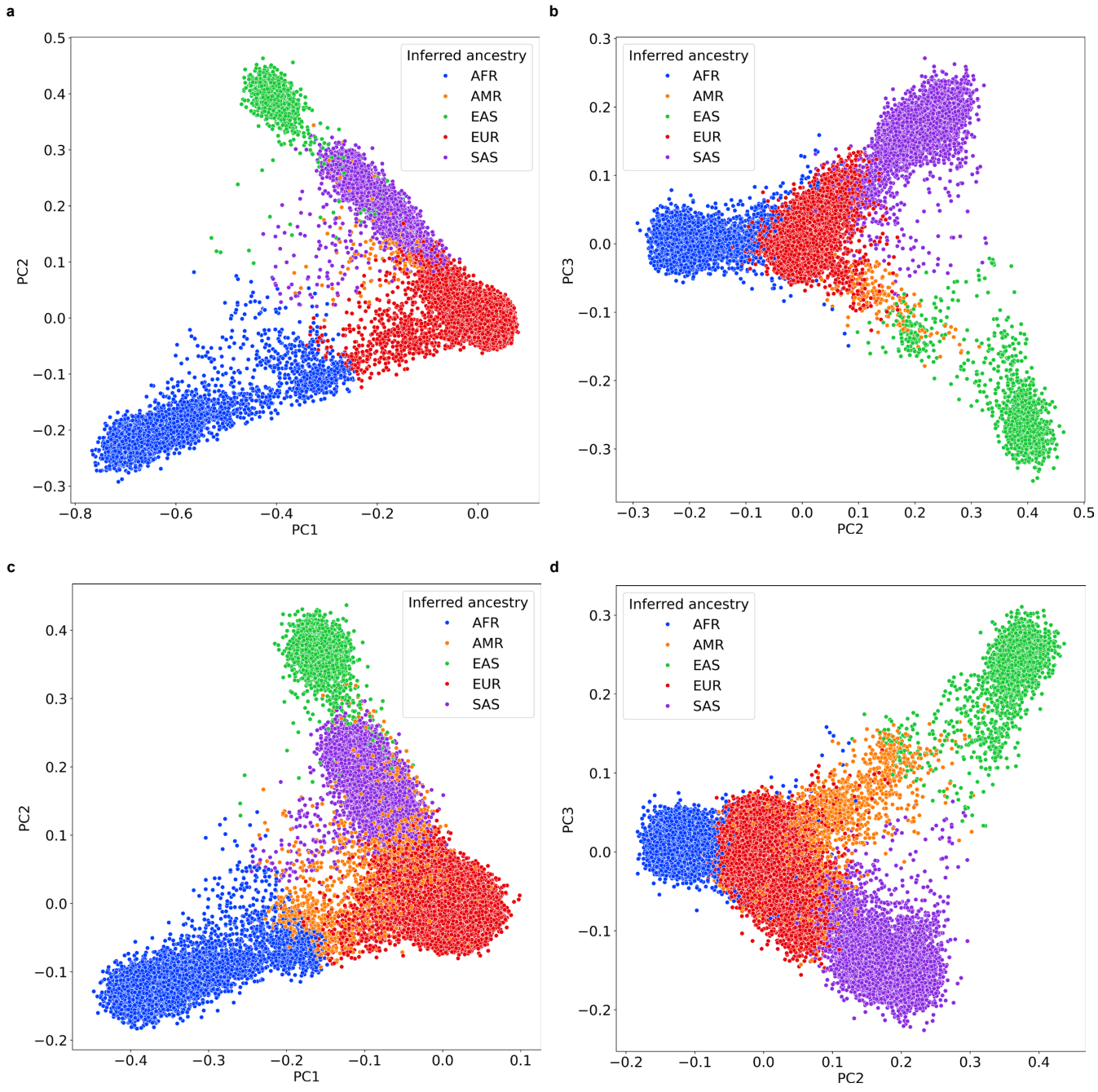
Supplementary Figure 5. Results from the greedy covering method when applied to 271 masks. The plots show the maximum number of significant associations across 54 phenotypes at each iteration (equivalent to the number of masks in the masking strategy; **Methods**) in the cases of **(a)** total associations, **(b)** low-frequency associations, and **(c)** rare associations. The blue lines represent the number of significant associations without Bonferroni correction. The orange lines represent the number of significant associations with Bonferroni correction. The black vertical lines indicate the optimal masking strategies which produced the maximum number of significant associations overall.



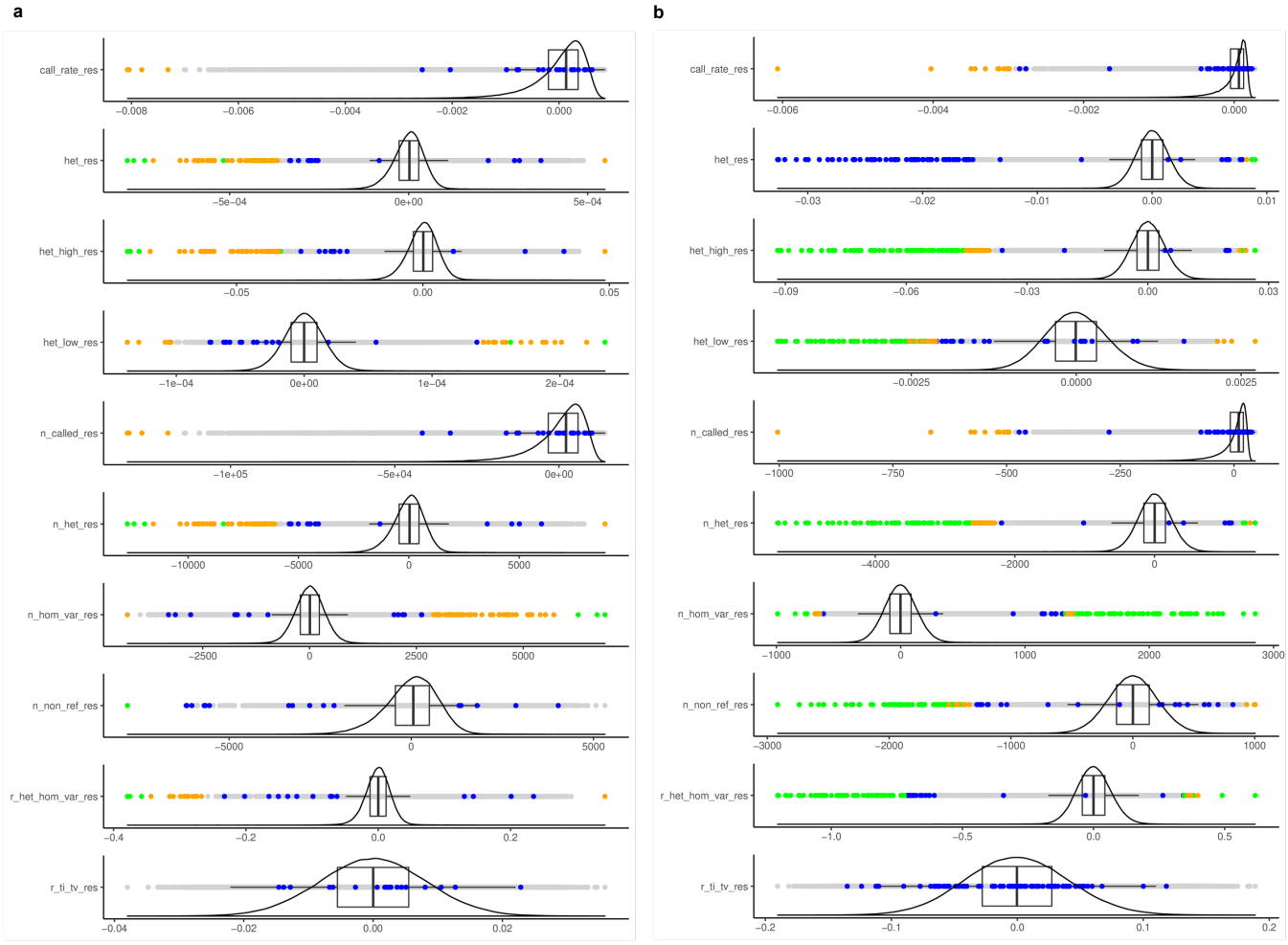
Supplementary Figure 6. Number of significant associations for each phenotype detected by three different masking strategies. The blue bars represent the number of significant associations detected by the trait-specific average masking strategies previously employed. The orange bars represent the number of significant associations detected by the trait-agnostic optimal masking strategies from 424 masks. The green bars represent the number of significant associations detected by the 6-mask or 8-mask recommended masking strategies. **(a)** Total associations. **(b)** Low-frequency associations. **(c)** Rare associations.



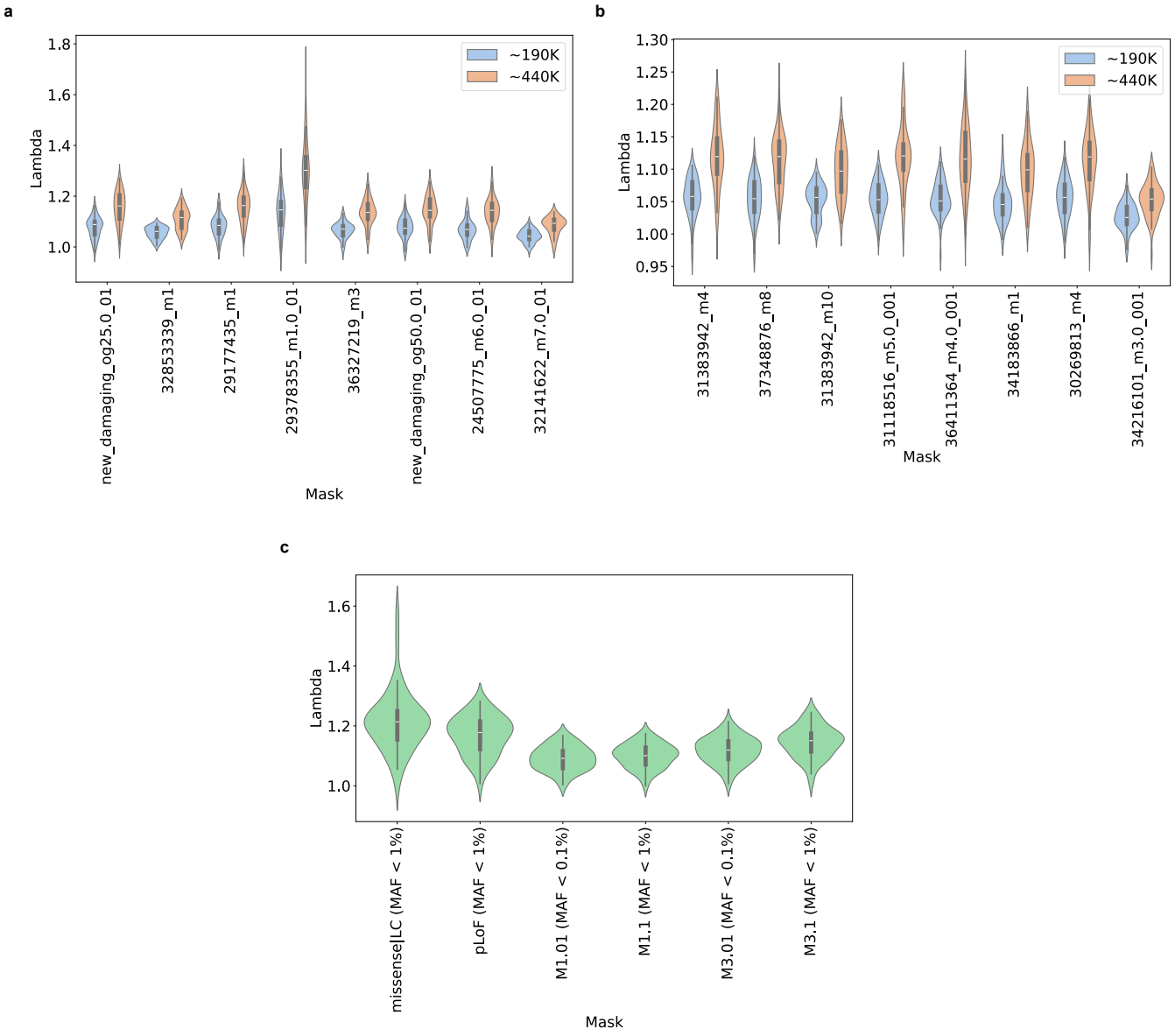
Supplementary Figure 7. Clusters of inferred ancestries using K-nearest neighbor method. (a) PC1 vs. PC2 for the samples in the ~190K analysis. **(b)** PC2 vs. PC3 for the samples in the ~190K analysis. **(c)** PC1 vs. PC2 for the samples in the ~440K analysis. **(d)** PC2 vs. PC3 for the samples in the ~440K analysis. Blue dots represent samples with inferred African (AFR) ancestry, orange dots represent samples with inferred Admixed American (AMR) ancestry, green dots represent samples with inferred East Asian (EAS) ancestry, red dots represent samples with inferred European (EUR) ancestry, purple dots represent samples with inferred South Asian (SAS) ancestry.



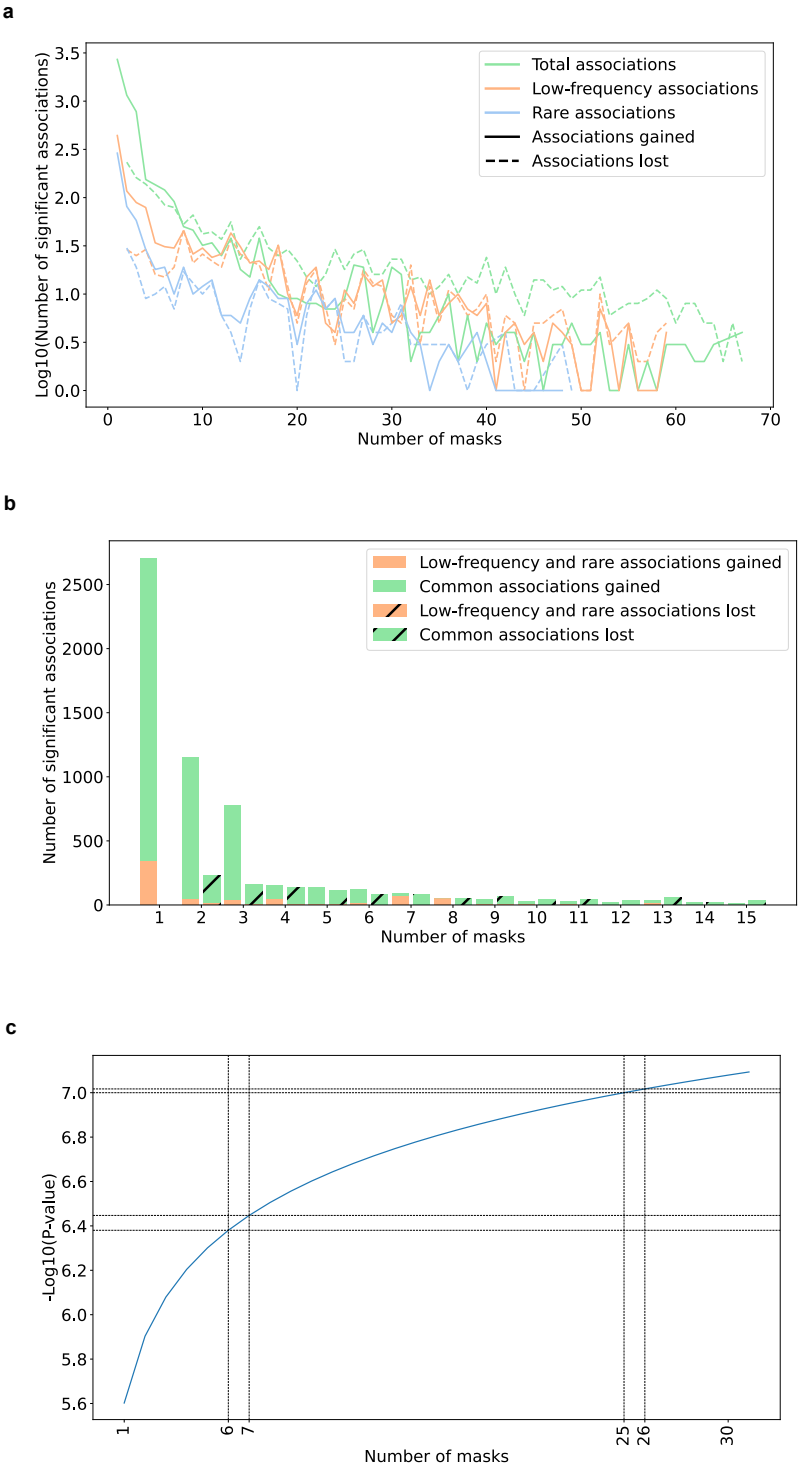
Supplementary Figure 8. Adjusted sample metric distributions. (a) For the samples in the ~190K analysis. **(b)** For the samples in the ~440K analysis. Grey dots represent samples clustered into Gaussian distributed subsets (not flagged), orange dots represent samples flagged as outliers based on individual PCARM's, blue dots represent samples flagged as outliers based on PCs of PCARM's, green dots represent samples flagged as outliers for both methods.



Supplementary Figure 9. Genomic inflation factors (lambdas) for masks in various masking strategies. (a) Lambdas of the gene-level associations for 54 phenotypes produced by the masks in the recommended low-frequency 8-mask strategy. **(b)** Lambdas of the gene-level associations for 54 phenotypes produced by the masks in the recommended rare 8-mask strategy. Blue violins represent the lambdas from the ~190K analysis. Orange violins represent the lambdas from the ~440K analysis. See **Table 1** for full definitions of masks. **(c)** Lambdas of the gene-level associations for 46 phenotypes produced by the low-frequency and rare masks in two high-profile studies of the UK Biobank (PMID: 36778668, ~500K samples and PMID: 34662886, ~455K samples). Original mask names were extracted from the studies. pLoF: high-confidence putative loss-of-function variants with MAF < 1% predicted by LOFTEE, missense|LC: missense variants and low-confidence pLoF variants with MAF < 1%, M1.01 or M1.1: pLoF variants with MAF < 0.1% and MAF < 1%, M3.01 or M3.1: pLoF variants or missense variants predicted to be damaging by 5 bioinformatic algorithms with MAF < 0.1% or MAF < 1%. Lambdas for PMID: 34375979 were not calculated because the full association results were not publicly available.



Supplementary Figure 10. Comparison between the number of masks in the optimal masking strategies for each of the total, low-frequency and rare association analyses. (a) Number of associations gained and lost as the number of mask increases (in log10). Green lines represent total associations, orange lines low-frequency associations, blue lines rare associations. Solid lines represent associations gained, dashed lines associations lost. **(b)** Number of common variant significant associations gained and lost compared to low-frequency and rare significant associations as the number of masks increases. Orange bars represent low-frequency and rare associations, green bars common variant associations. Plain bars represent associations gained, bars with forward slashed lines associations lost. **(c)** Significance threshold decreases as the Bonferroni correction penalty increases with the number of masks.



75

76

77