Supplementary Discussion

PheWAS analysis in UK Biobank and *All of Us* datasets.

In the main text, we described the PheWAS results for CC-Lassosum and GPS-Lassosum PRS, and we observed the GPS-based PRSs are frequently associated with disease-specific conditions, while CC-PRSs were more broadly associated with treatment-related consequence of the disease or comorbidities.

Here, we further performed PheWAS in UK Biobank and *All of Us* datasets using the PRS calculated from the remaining 21 methods for RA and SLE (Supplementary Data 5-8). Of the 21 remaining methods, 7 were stacking based methods. The number of significantly associated PheWAS codes (Bonferroni corrected p-value < 0.05, Supplementary Data 5-8, Supplementary Figures 10-13) with each PRS method varied (UK Biobank mean 93 PheWAS codes, range 0-335; *All of Us* mean 16 PheWAS codes, range 0-129).

In UK Biobank PheWAS for RA, 4 PRS methods were not significantly associated with any PheWAS codes (Bonferroni corrected p-value $\geqslant$ 0.05), which included MTAG-Lassosum, MTAG-PRS-CS, MV-Lassosum, and PROG-PRS-CS. These methods together with MTAG-LDpred2 was not even significantly associated with RA (Bonferroni corrected p-value $\geqslant$ 0.05). Of the 18 remaining methods, we observed 5 PRS methods having >100 PheWAS code associations among which 3 were case-control based PRS methods (PRS-CS-Stacking: 335 PheWAS codes, CC-PRS-CS: 335 PheWAS codes, CC-LDpred2: 292 PheWAS codes, TL-PRS-PRS-CS: 184 PheWAS codes, and CC-Lassosum: 142 PheWAS codes) (Supplementary Figure 10). The 255 PheWAS codes uniquely associated with these 5 PRS methods were related to either frequently observed comorbidities due to RA treatment (e.g., sepsis, renal failure, lung disease) or less related phenotypes (e.g., abdominal pain, actinic keratosis, hematemesis, varicose veins). Of the remaining 13 PRS methods, the 4 GPS-based PRS had the most disease-specific PheWAS code associations compared to other methods (GPS-Lassosum: 34 PheWAS codes, GPS-LDpred2: 29 PheWAS codes, GPS-PRS-CS: 29 PheWAS codes, GPS-stacking: 29 PheWAS codes) (Supplementary Figure 10). The 34 PheWAS codes associated with GPS-Lassosum were often closely related autoimmune diseases (e.g., ankylosing spondylitis, polymyalgia rheumatica, sarcoidosis, psoriasis, lupus, and multiple sclerosis). We observed 61 PheWAS codes uniquely associated with at least one of the remaining 9 non-GPS-based PRS methods but not associated with 4 GPS related PRS. The 61 PheWAS codes were also related to either frequently observed comorbidities due to RA treatment (e.g., thyroiditis, poisoning by antibiotics, vitamin deficiency) or less related phenotypes (e.g., type 1 diabetes, anemias, depression, systemic lupus erythematosus, angina pectoris, obstructive chronic bronchitis, GERD). These results further support GPS related PRS capture a potentially more biologically relevant set of associations specific to RA.

In UK Biobank for SLE, 5 PRS methods were not significantly associated with any PheWAS codes, including SLE (Bonferroni corrected p-value $\geqslant$ 0.05). These PRS methods are MTAG-Lassosum, MTAG-LDpred2, MTAG-PRS-CS, MV-Lassosum, and PROG-PRS-CS. Of the remaining 18 PRS methods, 3 were also not significantly associated with SLE (Bonferroni corrected p-value $\geqslant$ 0.05), which included PROG-Lassosum, TL-PRS-Lassosum, and TL-PRS-LDpred2. Of the remaining 15 methods, we observed 5 PRS methods had >50 PheWAS code associations among which 3 were case-control based PRS methods (CC-PRS-CS 79 PheWAS codes, PRS-CS-Stacking 78 PheWAS

codes, Lassosum-Stacking 65 PheWAS codes, CC-Lassosum 64 PheWAS codes, and CC-LDpred2 63 PheWAS codes) (Supplementary Figure 11). The 44 PheWAS codes uniquely associated with these 5 PRS methods were related to either frequently observed comorbidities due to severe SLE (e.g., ischemic heart disease, pleurisy) or less related phenotypes (e.g., obesity, gastritis, urinary tract infection, or lymphoid leukemia). Of the remaining 10 PRS methods, the 4 GPS-based PRS had the most specific PheWAS code associations compared to other methods (GPS-PRS-CS: 43 PheWAS codes, GPS-stacking: 43 PheWAS codes, GPS-Lassosum: 23 PheWAS codes, and GPS-LDpred2: 14 PheWAS codes) (Supplementary Figure 11). The 47 PheWAS codes associated with at least one of the 4 GPS related PRS were also associated with at least one of the non-GPS PRS methods. The 47 PheWAS codes were often closely related autoimmune diseases (e.g., Celiac disease, Multiple Sclerosis, Primary biliary cirrhosis, Systemic sclerosis, or Rheumatoid arthritis). Lastly, we observed 14 PheWAS codes uniquely associated with at least one of the remaining 6 non-GPS PRS methods but not associated with the 4 GPS related PRS. The 14 PheWAS codes were also less related phenotypes (e.g., hypovolemia, hypotension, disorders of iron metabolism, septicemia, or disorders of fluid, electrolyte, and acid-base balance). This provides further support that phenotypes associated with GPS-related PRS are more relevant for SLE when compared to those associated with other non-GPS PRS methods.

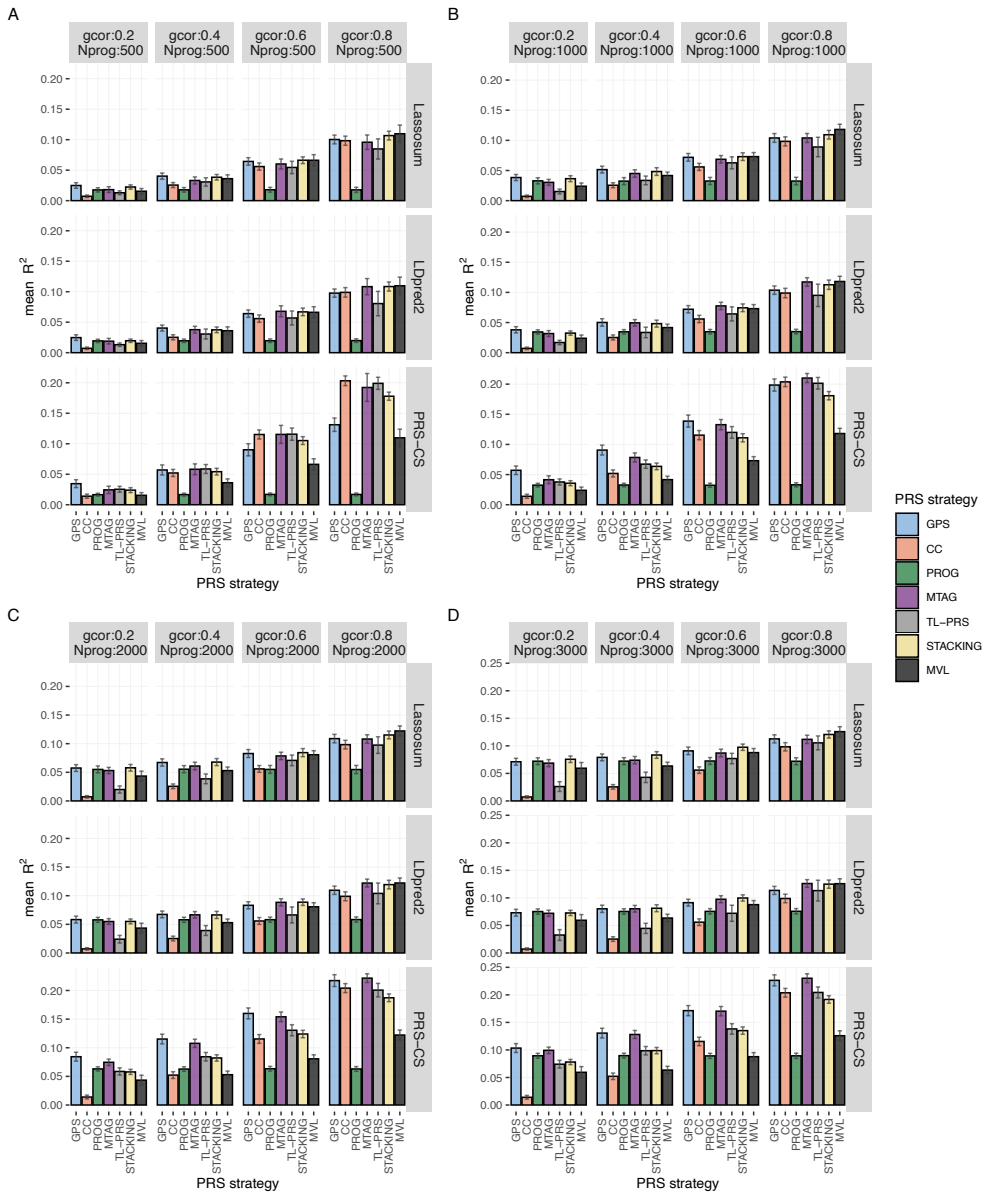
To verify if the specificity of GPS related PRS for RA and SLE is reproducible in other biobanks, we conducted PheWAS for all 23 methods in the *All of Us* dataset. We observed an even higher specificity in the *All of Us* dataset for GPS related PRS for RA and SLE.

In *All of Us* PheWAS for RA, 9 PRS methods were not significantly associated with any PheWAS codes (Bonferroni corrected p-value $\geqslant$ 0.05), which included MTAG-Lassosum, MTAG-LDpred2, MTAG-PRS-CS, MTAG-Stacking, MV-Lassosum, PROG-LDPred2, TL-PRS-LDpred2, PROG-PRS-CS, and LDPred2-Stacking. Similar to the observation in UK Biobank, of the 14 remaining methods, we observed 4 PRS methods had >40 PheWAS code associations among which 3 were case-control based PRS methods (PRS-CS-Stacking 129 PheWAS codes, CC-PRS-CS 129 PheWAS codes, CC-LDpred2 124 PheWAS codes, and CC-Lassosum 48 PheWAS codes) (Supplementary Figure 12). The 133 PheWAS codes uniquely associated with these 4 PRS methods were also either frequently observed comorbidities due to RA treatment (e.g., sepsis, pneumonia, renal failure, lung disease, and opiates and related narcotics causing adverse effects in therapeutic use) or less related phenotypes (e.g., anxiety, alcoholism, hernia, actinic keratosis, and abdominal pain). The 15 PheWAS codes associated with at least one of the 4 GPS related PRS were also associated with at least one of the non-GPS PRS methods (Supplementary Figure 12). Similar to the observation in UK Biobank, the 15 PheWAS codes were closely related autoimmune diseases (e.g., Graves' disease, multiple sclerosis, celiac disease, psoriasis, and type 1 diabetes). We observed 6 PheWAS codes uniquely associated with at least one of the remaining 6 non-GPS PRS methods but not associated with 4 GPS related PRS. The 6 PheWAS codes were all less relevant phenotypes (e.g., neoplasm of skin, secondary hypothyroidism, substance addiction and disorders, edema). These results validate the observations in UK Biobank for RA. Compared to UK Biobank, in the *All of Us* PheWAS, the GPS-based PRS were only associated with 15 PheWAS codes, most of which were closely related autoimmune diseases. This supports the specificity of GPS related PRS to RA.

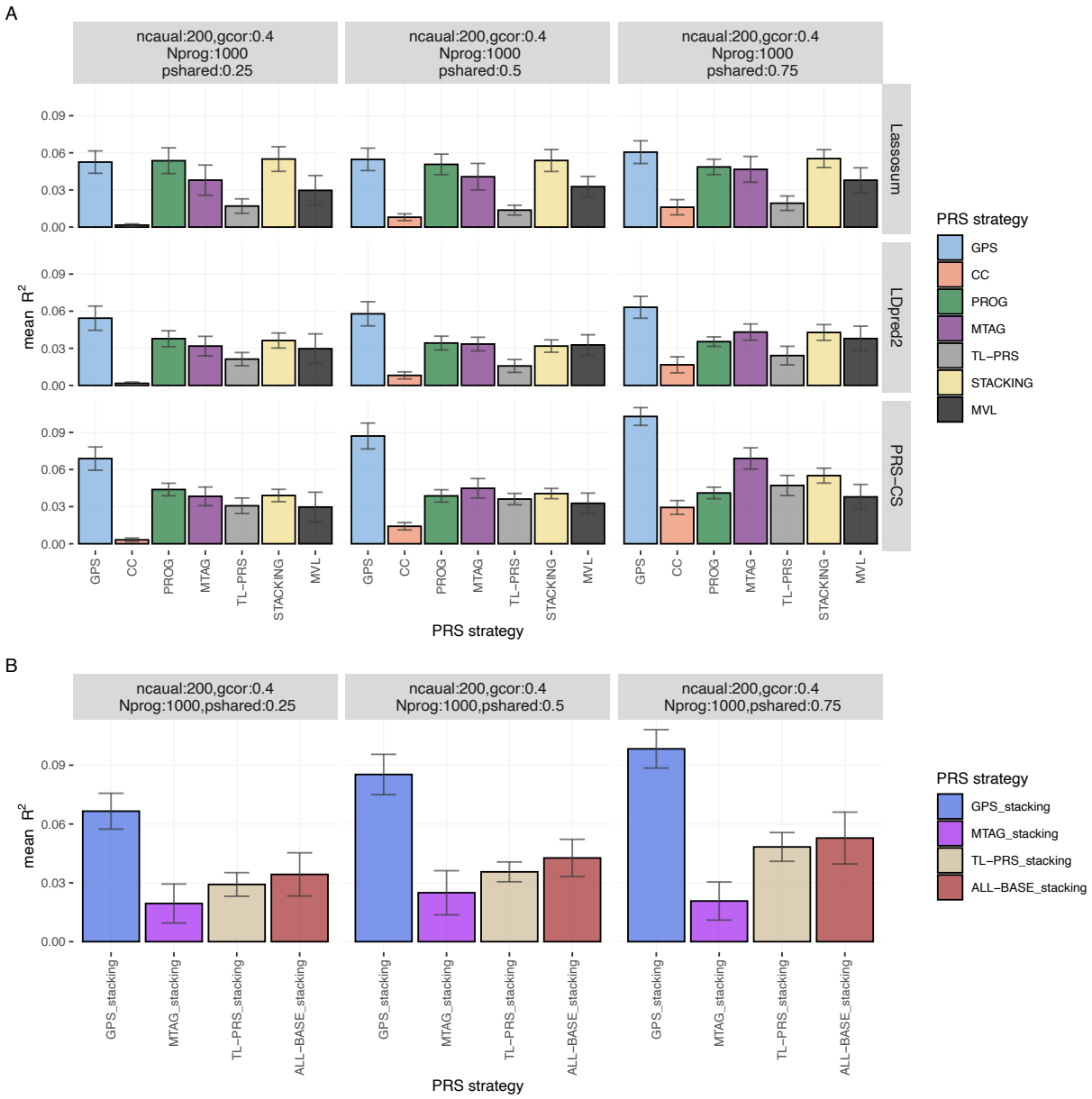
Lastly, in *All of Us* PheWAS for SLE, 9 PRS methods were not significantly associated with any PheWAS codes (Bonferroni corrected p-value $\geqslant$ 0.05), which included MTAG-Lassosum, MTAG-LDpred2, MTAG-PRS-CS, PROG-Lassosum, TL-PRS-Lassosum, PROG-PRS-CS, TL-PRS-PRS-CS, TL-PRS-Ldpred2, and TL-PRS-Stacking. Of the remaining 14 methods, 3 were not even associated with

SLE (Bonferroni corrected p-value $\geqslant$ 0.05), which included PROG-LDPred2, MVL, and MTAG-Stacking (Supplementary Figure 13). We observed 44 PheWAS codes were associated with at least one of the remaining 11 PRS methods. The GPS-based PRS had the most specific PheWAS code associations compared to other methods (GPS-stacking: 18 PheWAS codes, GPS-PRS-CS: 18 PheWAS codes, GPS-Lassosum: 15 PheWAS codes, and GPS-LDpred2: 5 PheWAS codes) (Supplementary Figure 13). The 20 PheWAS codes associated with at least one of the 4 GPS-based PRS were also associated with at least one of the 7 non-GPS PRS methods. Similar to the observation in UK Biobank, the 20 PheWAS codes were also closely related autoimmune diseases (e.g., Graves' disease, multiple sclerosis, rheumatoid arthritis, celiac disease, and systemic sclerosis). The remaining 24 PheWAS codes uniquely associated with at least one of the remaining 7 non-GPS PRS methods but not associated with 4 GPS related PRS were less related phenotypes (e.g., skin cancer, actinic keratosis, tobacco use disorder, hypertension, and hemorrhoids). These results validate the observations in the UK Biobank for SLE and supports the specificity of GPS related PRS to SLE.

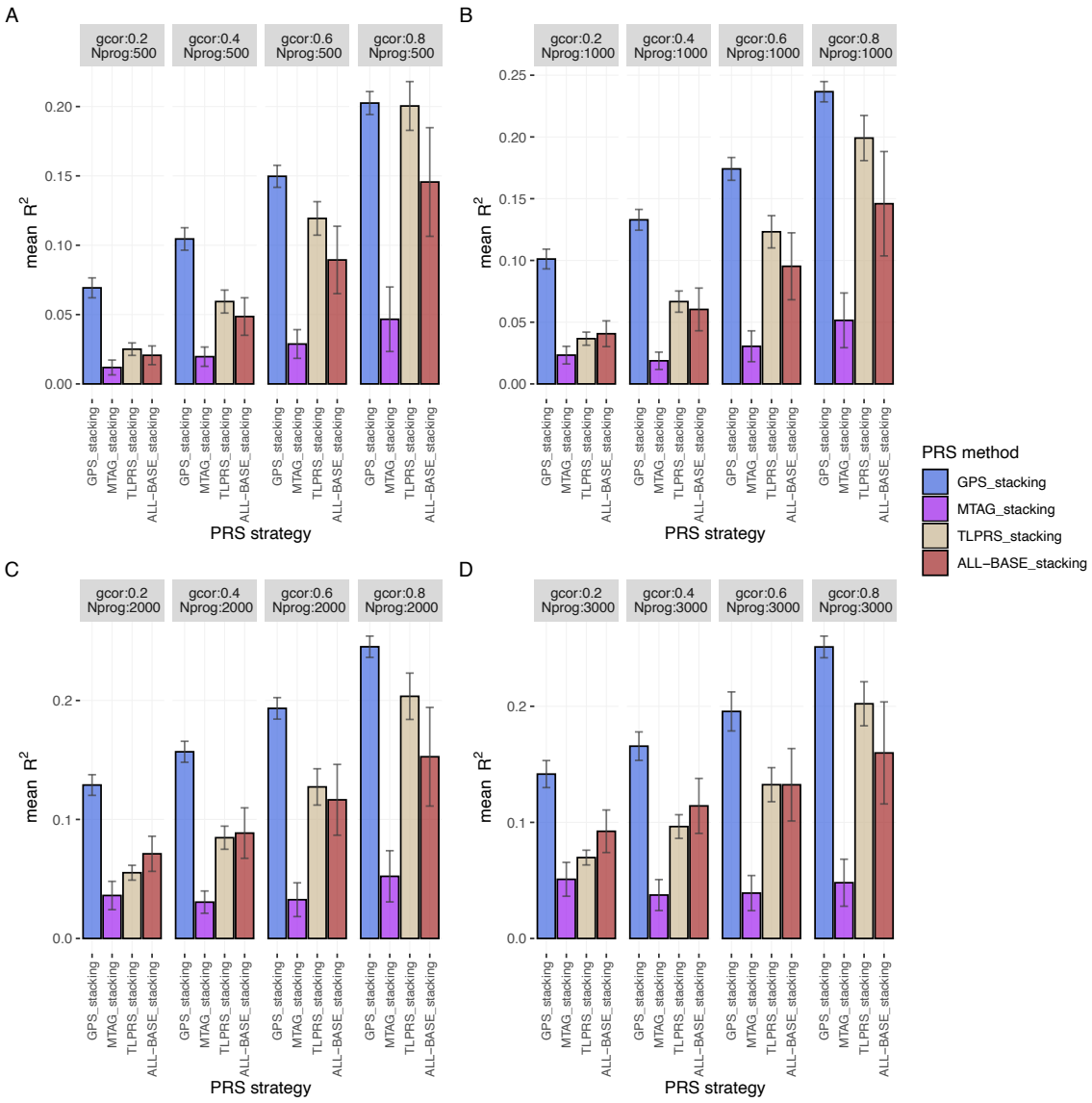
Supplementary Figures



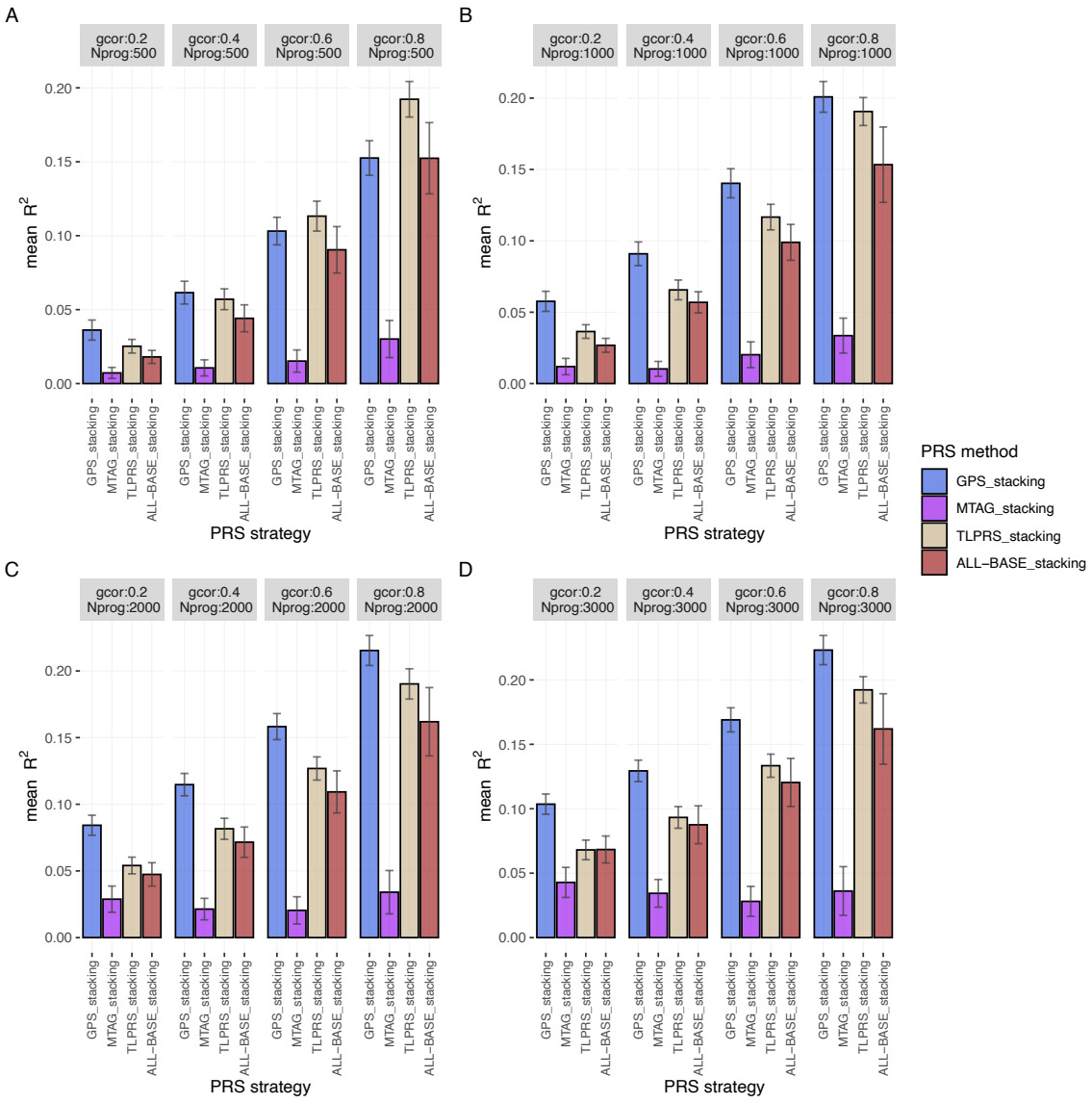
Supplementary Figure 1. Prediction accuracy of different PRS models in simulations (500 causal variants). All causal variants are shared between progression and case-control phenotype in this simulation. The prediction accuracy is evaluated by the mean prediction R^2 across 20 simulated replicates. The error bar indicates the standard deviation of prediction $\boldsymbol{R^2}$ across 20 simulation replicates. Each row represents different PRS models using the same baseline PRS method. MVL uses Lassosum as baseline framework, so it cannot accommodate alternative baseline PRS methods. To facilitate the comparisons, we estimate the prediction R^2 of MVL by repeating the scenarios in different rows (baseline methods) and taking the average. The sample size of the progression cohort is 500 in (A), 1000 in (B), 2000 in (C), and 3000 in (D). The number of causal variants is set as 500. gcor: genetic correlation, Nprog: sample size of biobank study of progression phenotype. Super-stacking models are not included.



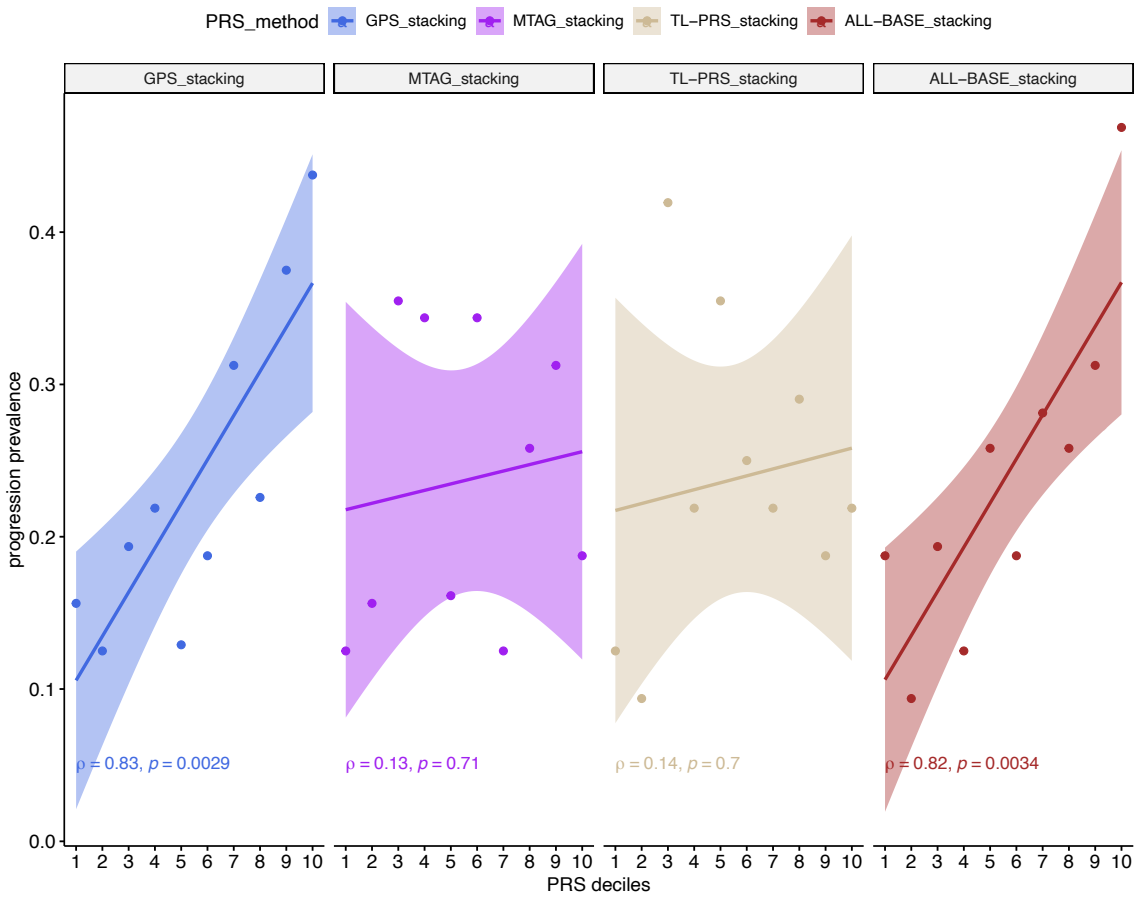
Supplementary Figure 2. The prediction accuracy of PRS models when only a proportion of causal variants are shared in simulations. The prediction accuracy is evaluated by the mean prediction $\boldsymbol{R^2}$ across 20 simulated replicates. The error bar indicates the standard deviation of prediction $\boldsymbol{R^2}$ across 20 simulated replicates. The number of causal variants, genetic correlations between case-control and progression phenotype, and the sample size of progression cohort are set as 200, 0.2, and 1000, respectively. Proportion of shared causal variants are varied between 0.25, 0.5, and 0.75. (A) Prediction accuracy of non-super-stacking PRS models (B) Prediction accuracy of super-stacking PRS models. ncausal: number of causal variants. gcor: the genetic correlation between the case-control and progression phenotypes. Nprog: the sample size of the biobank cohort measuring the progression phenotype. pshared: proportion of shared causal variants



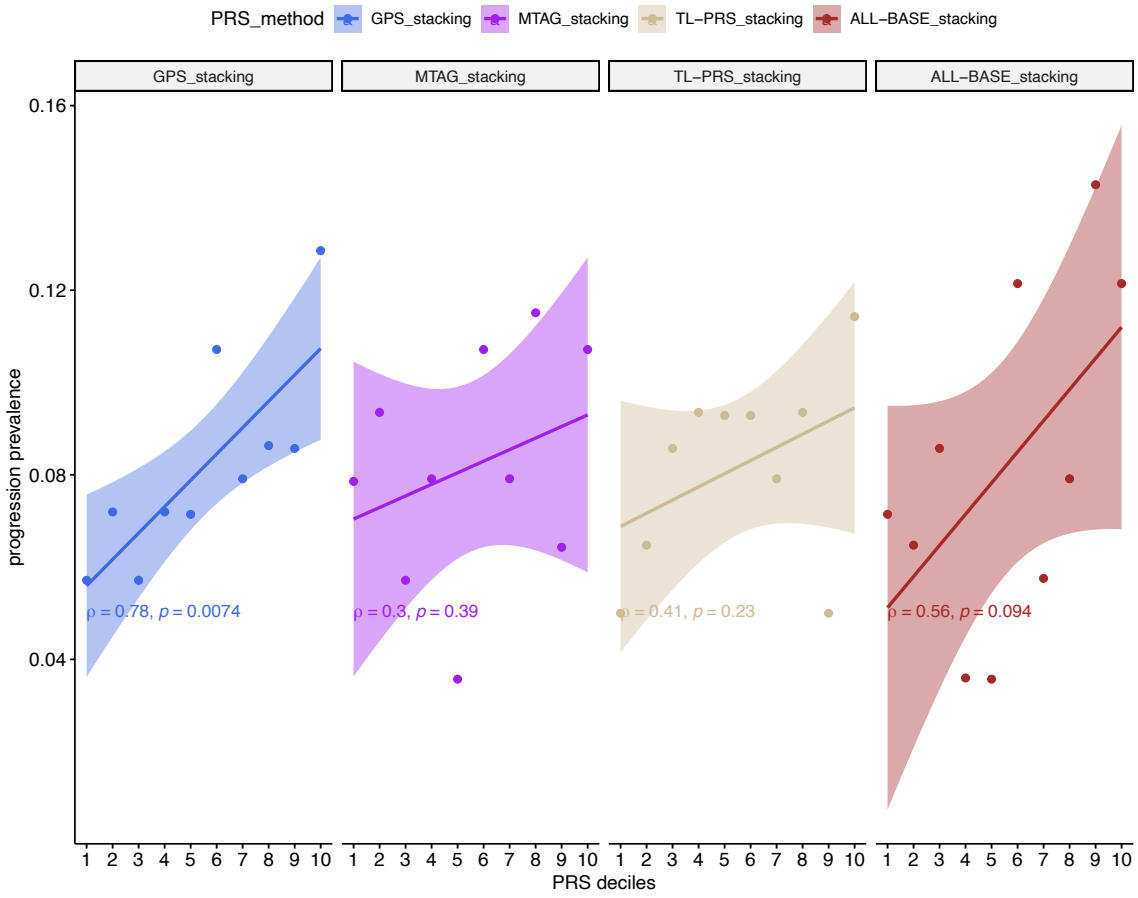
Supplementary Figure 3. The prediction accuracy of super-stacking models in simulations (200 causal variants). All causal variants are shared between progression and case-control phenotype in this simulation. The prediction accuracy is evaluated by the mean prediction $\boldsymbol{R^2}$ across 20 simulated replicates. The error bar indicates the standard deviation of prediction $\boldsymbol{R^2}$ across 20 simulation replicates. The sample size of the progression cohort is 500 in (A), 1000 in (B), 2000 in (C), and 3000 in (D). The number of causal variants is set as 200. gcor: the genetic correlation between the case-control and progression phenotypes. Nprog: the sample size of the biobank cohort measuring the progression phenotype.



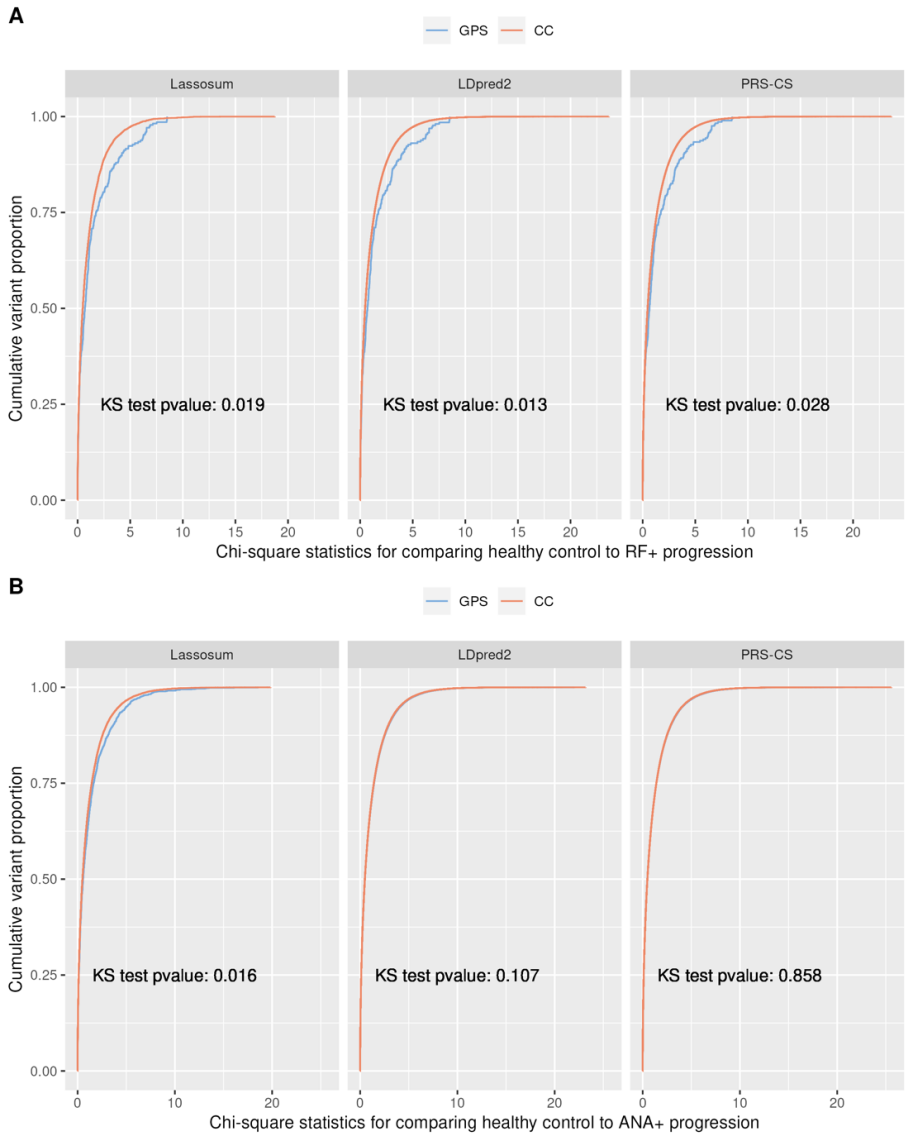
Supplementary Figure 4. The prediction accuracy of super-stacking models in simulations (500 causal variants). All causal variants are shared between progression and case-control phenotype in this simulation. The prediction accuracy is evaluated by the mean prediction $\boldsymbol{R^2}$ across 20 simulated replicates. The error bar indicates the standard deviation of prediction $\boldsymbol{R^2}$ across 20 simulation replicates. The sample size of the progression cohort is 500 in (A), 1000 in (B), 2000 in (C), and 3000 in (D). The number of causal variants is set as 500. gcor: the genetic correlation between the case-control and progression phenotypes. Nprog: the sample size of the biobank cohort measuring the progression phenotype.



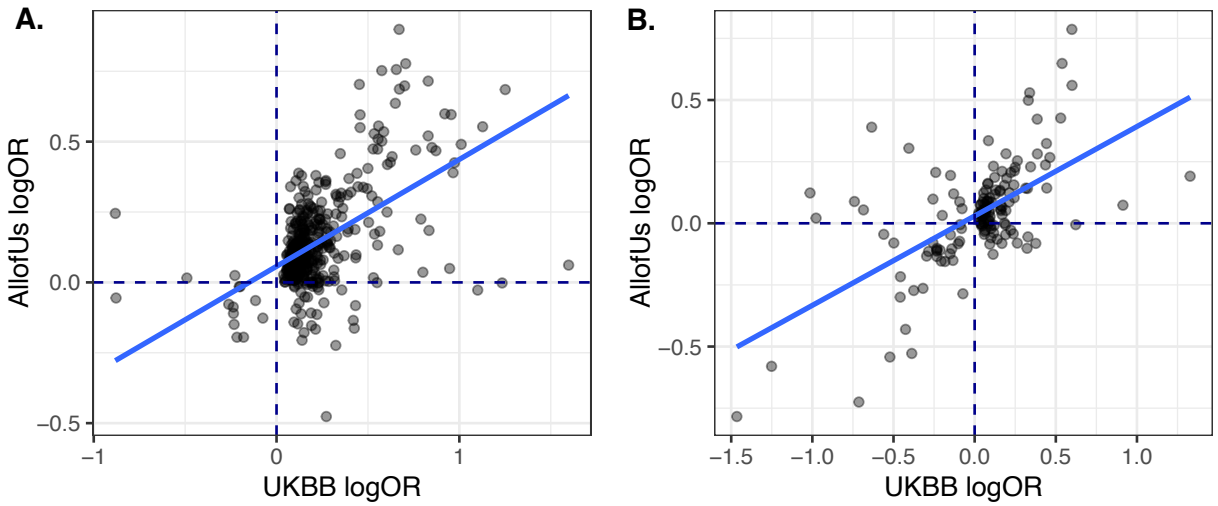
Supplementary Figure 5. The association between PRS and the prevalence of RF positive → RA progressions in the *All of Us* data for super-stacking models. The *All of Us* data is not used to train genetic risk scores. The Pearson correlation coefficient (and corresponding p-values from two-sided t-test) between PRS and the progression prevalence in the *All of Us* data are labeled on the plot. The error bands represent 95% confidence intervals of fitted linear regression lines. GPS_stacking yields the strongest correlations between predicted and observed progression in the independent test dataset among super-stacking models.



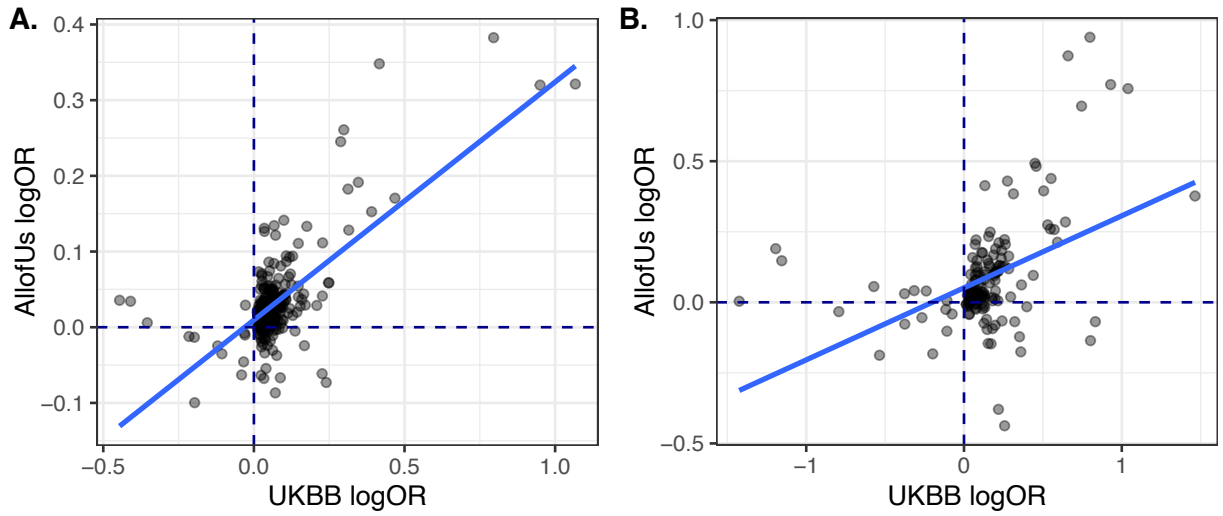
Supplementary Figure 6. The association between PRS and the prevalence of ANA positive → SLE progressions in the *All of Us* data for super-stacking models. The *All of Us* data is not used to train genetic risk scores. The Pearson correlation coefficient (and corresponding p-values from two-sided t-test) between PRS and the progression prevalence in the *All of Us* data are labeled on the plot. The error bands represent 95% confidence intervals of fitted linear regression lines. GPS_stacking yields the strongest correlations between predicted and observed progression in the independent test dataset among super-stacking models.



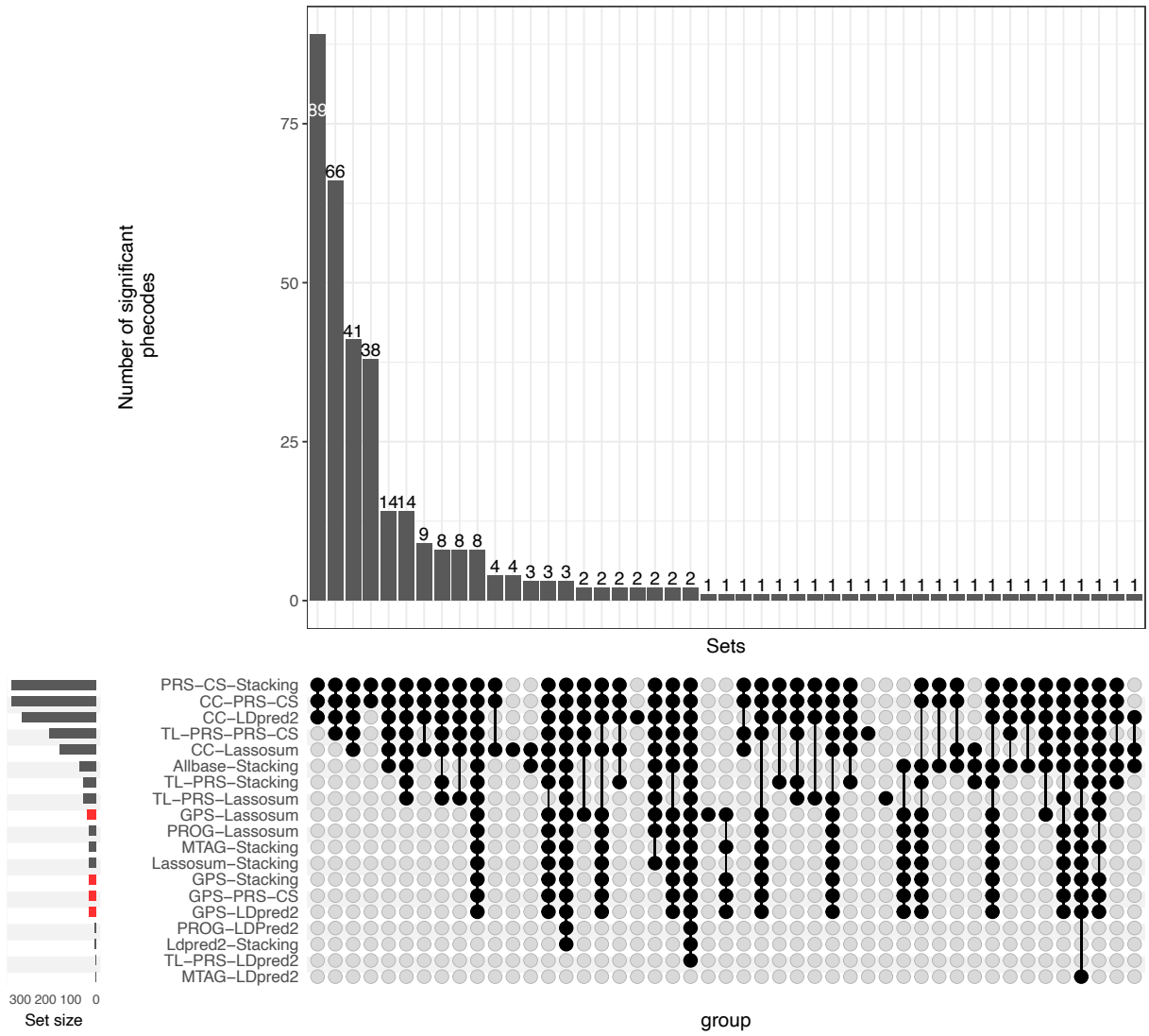
Supplementary Figure 7. Cumulative distributions of marginal association statistics testing genetic association with healthy control to preclinical disease progressions in the *All of Us* dataset. We trained the progression risk scores in the BioVU biobank. We also performed GWAS, comparing healthy controls to preclinical disease samples in the *All of Us* data, which is not used in model training. For variants selected by GPS or the risk scores using case-control samples only, we compare the distribution of marginal χ^2 statistics testing genetic associations with healthy control → preclinical disease progression. The cumulative distribution functions of the marginal χ^2 statistics are plotted for (A) healthy control to RF positive progressions and (B) healthy controls to ANA positive progressions for the variants selected by the risk scores. Two-sided Kolmogorov-Smirnov (KS) test are performed to compare the distributions and the p-values of the KS test are labeled on each subpanel. At each quantile, the variants selected by GPS are more significantly associated with control to preclinical progression phenotype, compared to variants selected by the risk scores based on case-control studies. Together with Figure 5, this comparison shows that GPS tends to select variants that can better distinguish preclinical states from controls and disease cases.



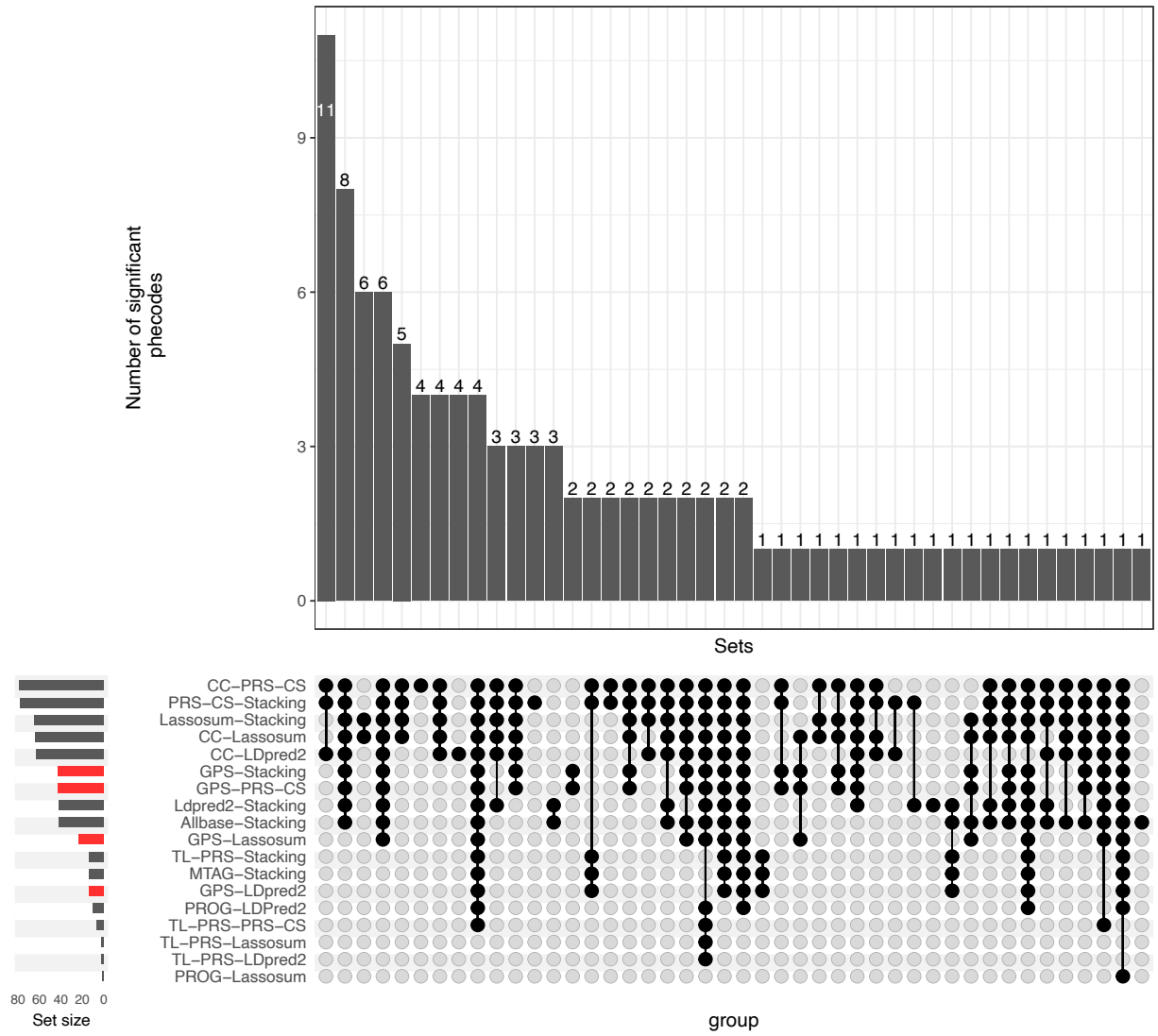
Supplementary Figure 8. Comparison of PheWAS log odds ratio (logOR) for RA CC-PRS and GPS-PRS in UK Biobank and All of Us. As UK Biobank was included in the case-control meta-analysis GWAS for RA, we conducted PheWAS for RA CC-PRS and GPS-PRS in All of Us, which consists of independent samples. We observed significant correlations in logOR between two biobanks for RA CC-PRS (A) (r2 =0.53, T-test two-sided p-value<2.2x10-16) and RA GPS-PRS (B) (r2 =0.6, T-test two-sided p- value<2.2x10-16) using PheWAS codes with two-sided p-value < 0.05 in UK Biobank.



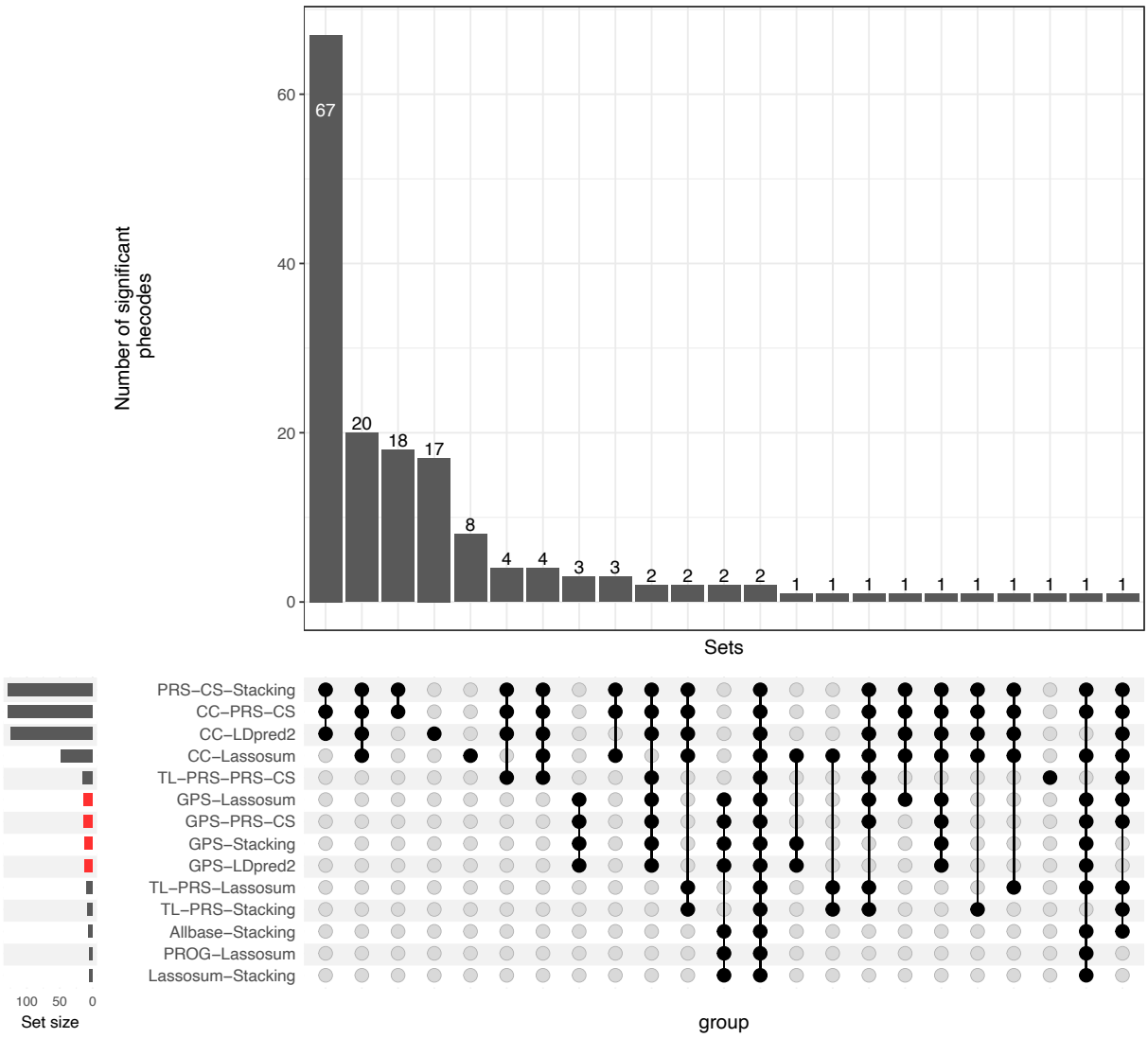
Supplementary Figure 9. Comparison of PheWAS log odds ratio (logOR) for SLE CC-PRS and GPS-PRS in UK Biobank and All of Us. As UK Biobank was included in the case-control meta-analysis GWAS for SLE, we conducted PheWAS for SLE CC-PRS and GPS-PRS in All of Us, which consists of independent samples. We observed significant correlations of logOR between the two biobanks for SLE CC-PRS (A) (r2 =0.70, T-test two-sided p-value=<$2.2x10^{-16}$) and SLE GPS-PRS (B) (r2 =0.44, T-test two-sided p- value=$5x10^{-9}$) using PheWAS codes with two-sided p-value < 0.05 in the UK Biobank.



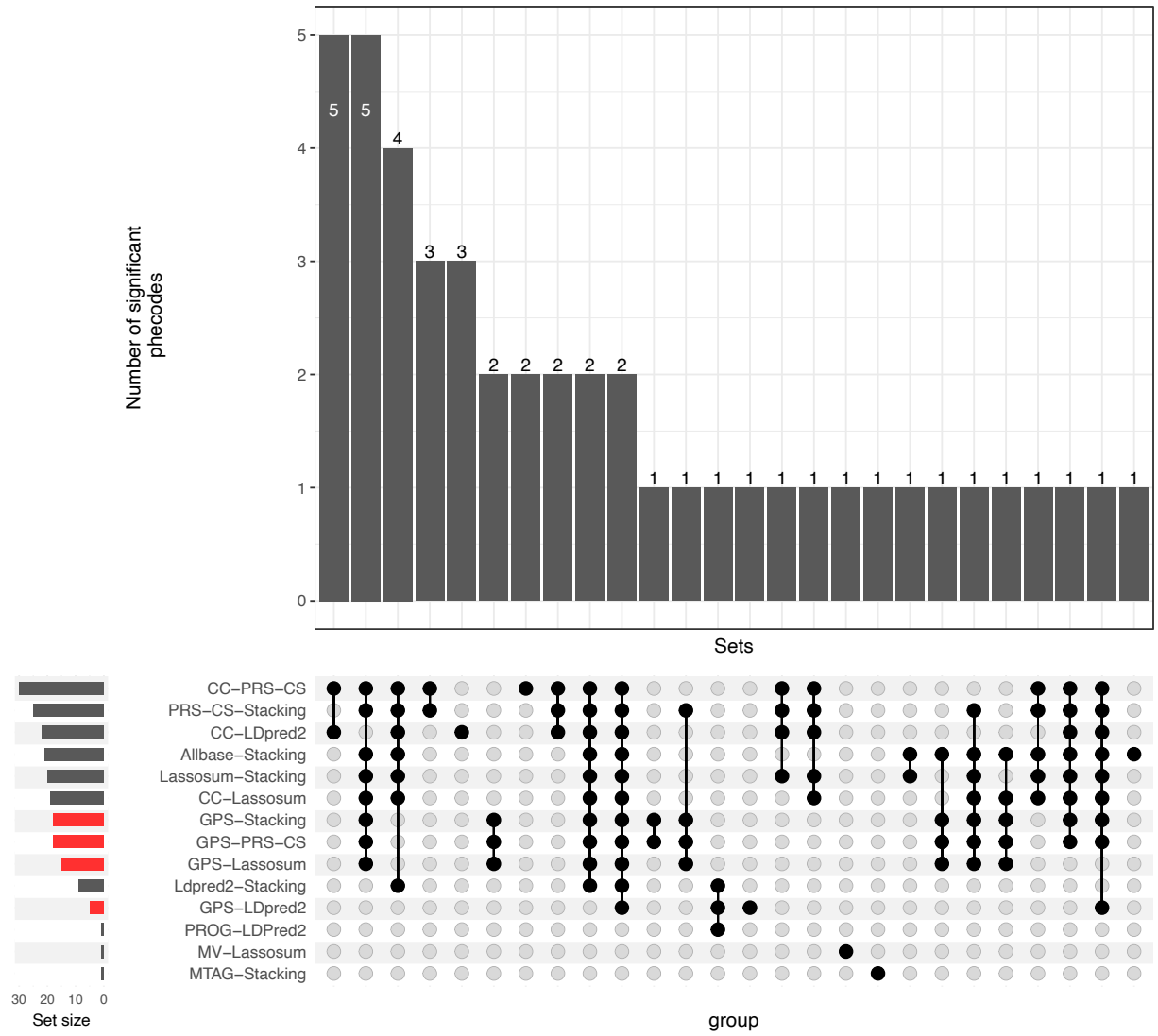
Supplementary Figure 10. Upset plot summarizing the overlaps of PheWAS codes that are significantly associated with different PRS for RA in UK Biobank. The number of PheWAS codes identified by each of the 4 GPS related PRS methods (GPS-Lassosum: 34 PheWAS codes, GPS-LDpred2: 29 PheWAS codes, GPS-PRS-CS: 29 PheWAS codes, GPS-stacking: 29 PheWAS codes) are shown with red bars. PRS methods without significant associations are omitted from the plot



Supplementary Figure 11. Upset plot summarizing overlaps of PheWAS codes that are significantly associated with different PRS for SLE in UK Biobank. The number of PheWAS codes identified by each of the 4 GPS related PRS methods (GPS-PRS-CS: 43 PheWAS codes, GPS-stacking: 43 PheWAS codes, GPS-Lassosum: 23 PheWAS codes, and GPS-LDpred2: 14 PheWAS codes) are shown with red bars. PRS methods without significant associations are omitted from the plot.



Supplementary Figure 12. Upset plot summarizing overlaps of PheWAS codes that are significantly associate with different PRS for RA in *All of Us*. The number of PheWAS codes identified by each of the 4 GPS related PRS methods (GPS-Lassosum: 14 PheWAS codes, GPS-PRS-CS: 13 PheWAS codes, GPS-LDpred2: 12 PheWAS codes, GPS-stacking 12 PheWAS codes) are shown with red bars. PRS methods without significant associations are omitted from the plot



Supplementary Figure 13. Upset plot summarizing overlaps of PheWAS codes that are significantly associated with different PRS for SLE in *All of Us*. The number of PheWAS codes identified by each of the 4 GPS related PRS methods (GPS-PRS-CS: 18 PheWAS codes, GPS-stacking: 18 PheWAS codes, GPS-Lassosum: 15 PheWAS codes, and GPS-LDpred2: 5 PheWAS codes) are shown with red bars. PRS methods without significant associations are omitted from the plot.