Description of Additional Supplementary Files

File name: Supplementary Data 1

Description: Detailed information of the 23 PRS models constructed using three baseline PRS methods and eight different combination strategies for integrating case-control risk scores.

File name: Supplementary Data 2

Description: Sample sizes and demographics for controls, preclinical individuals, and disease cases in the BioVU and All of Us biobanks. All samples are of European ancestry. The fraction of females is also reported.

File name: Supplementary Data 3

Description: The accuracy for predicting RF positive to RA progressions in the All of Us biobank for super-stacking models. GPS_stacking is the best method across all categories.

File name: Supplementary Data 4

Description: The accuracy for predicting ANA positive to SLE progressions in the All of Us biobank for super-stacking models. GPS_stacking is the best method across all categories.

File name: Supplementary Data 5

Description: PheWAS results for RA risk scores in UK Biobank for 23 PRS methods. We correlated the PRS from 23 models for RA with different PheWAS codes in UK Biobank. P-values are derived using two-sided Chi-square test. Manhattlan plots for the CC-Lassosum and GPS-Lassosum results are provided in Figure 6. Comparison of PheWAS log odds ratio (logOR) for RA CC-Lassosum and GPS-Lassosum in UK Biobank and All of Us are provided in Supplementary Figure 8. Upset plot summarizing the overlaps of significant PheWAS code across PRS methods are provided in Supplementary Figure 10.

File name: Supplementary Data 6

Description: PheWAS results for RA PRS in the All of Us data for 23 PRS methods. We correlated the PRS from 23 models for RA with different PheWAS codes in the All of Us Biobank. P-values are derived using two-sided Chi-square test. Comparison of PheWAS log odds ratio (logOR) for RA CC-Lassosum and GPS-Lassosum in UK Biobank and All of Us are provided in Supplementary Figure 8. Upset plot summarizing the overlaps of significant PheWAS code across PRS methods are provided in Supplementary Figure 12.

File name: Supplementary Data 7

Description: PheWAS results for SLE PRS in UK Biobank for 23 PRS methods. We correlated the PRS from 23 models for SLE with different PheWAS codes in the UK Biobank. P-values are derived using two-sided Chi-square test. Manhattlan plots for the CC-Lassosum and GPS-Lassosum results are provided in Figure 7. Comparison of PheWAS log odds ratio (logOR) for SLE CC-Lassosum and GPS-Lassosum in UK Biobank and All of Us are provided in Supplementary Figure 9. Upset plot summarizing the overlaps of significant PheWAS code across PRS methods are provided in Supplementary Figure 11.

File name: Supplementary Data 8

Description: PheWAS results for SLE PRS in the All of Us dataset for 23 PRS methods. We correlated the PRS from 23 models for SLE with different PheWAS codes in the All of Us Biobank. P-values are derived using two-sided Chi-square test. Upset plot summarizing the overlaps of significant PheWAS code across PRS methods are provided in Supplementary Figure 13.

File name: Supplementary Data 9

Description: Prediction accuracy of European progression PRS in non-European samples in the All of Us dataset.

File name: Supplementary Data 10

Description: Sample size of non-European progression phenotypes in the All of Us dataset.

File name: Supplementary Data 11

Description: GWAS studies used for constructing PRS models for case-control phenotypes.