The American Journal of Human Genetics, Volume *107*

Supplemental Data

Non-parametric Polygenic Risk Prediction via Partitioned GWAS Summary Statistics

Sung Chun, Maxim Imakaev, Daniel Hui, Nikolaos A. Patsopoulos, Benjamin M. Neale, Sekar Kathiresan, Nathan O. Stitziel, and Shamil R. Sunyaev

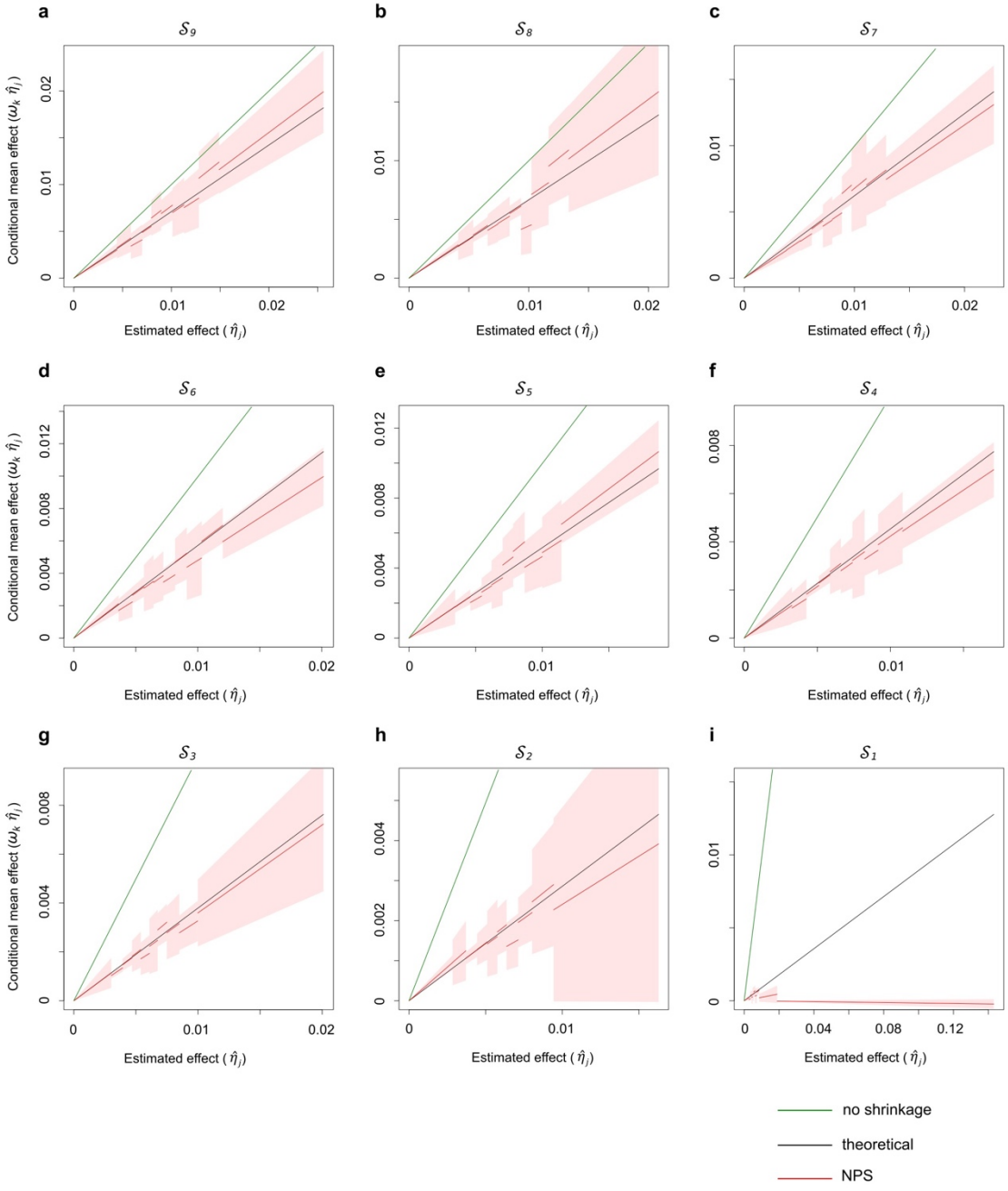

Figure S1. NPS approximates the conditional mean effects: infinitesimal genetic architecture ($\mathcal{S}_1, \ldots, \mathcal{S}_9$). NPS shrinkage weights ω_k (red line) are compared to the theoretical optimum (black line), $\lambda_j/(\lambda_j + \frac{M}{N_g h^2})$, under the infinitesimal architecture. $\mathcal{S}_1, \ldots, \mathcal{S}_{10}$ indicate the partitions of lowest to highest eigenvalues of projection. The mean NPS shrinkage weights (red line) and their 95% CIs (red shade) were estimated from 5 replicates. No shrinkage line (green) indicates $\omega_k = 1$. The number of markers *M* is 101,296. The discovery GWAS size *N* equals to *M*. The heritability h^2 is 0.5. See Figure 2B for $\mathcal{S}_{10}$.

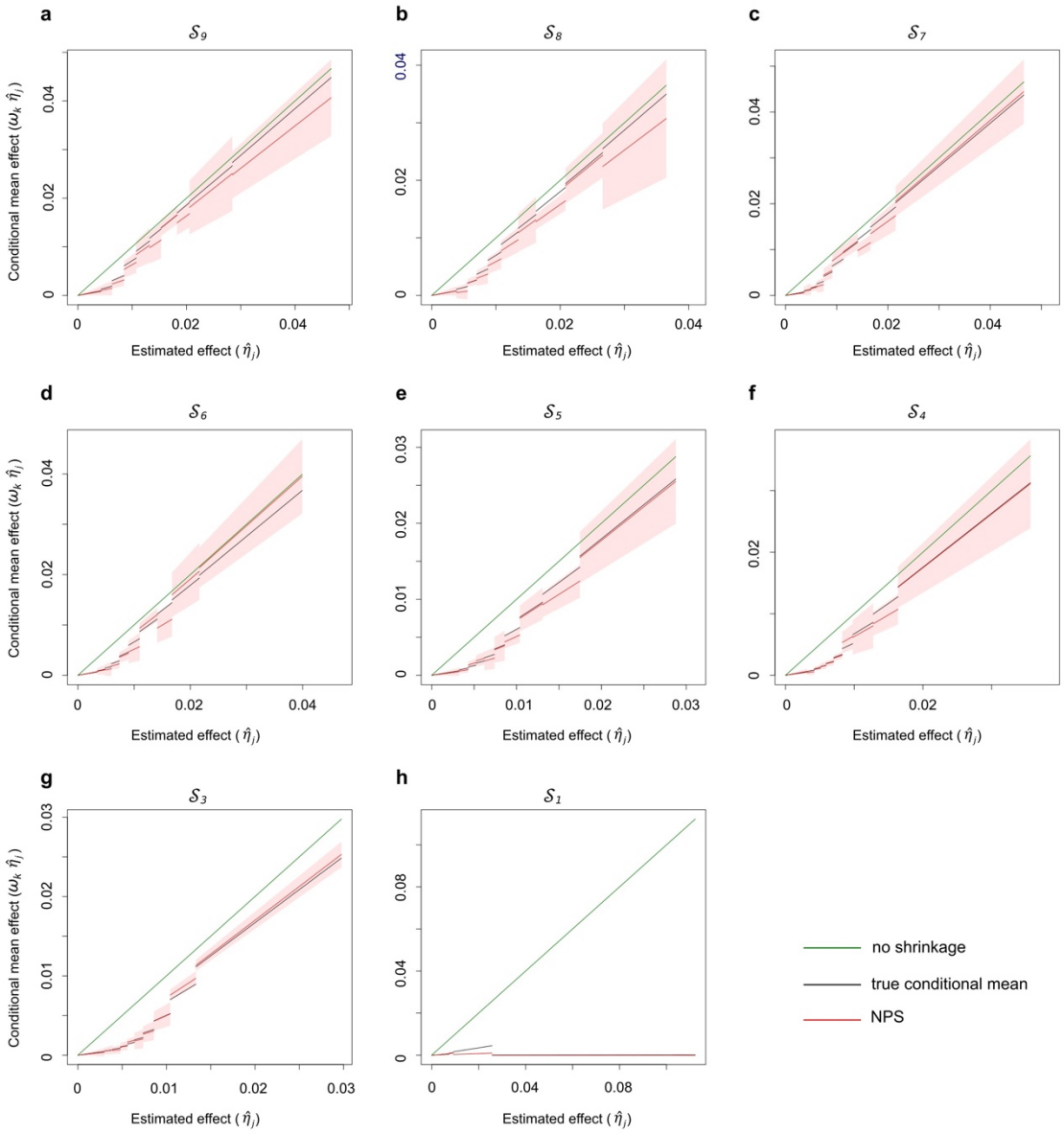

Figure S2. NPS approximates the conditional mean effects: non-infinitesimal genetic architecture ($\mathcal{S}_1, \mathcal{S}_3, \ldots, \mathcal{S}_9$). NPS shrinkage weights ω_k (red line) are compared to the true conditional means (black line), which were estimated empirically from 40 simulation runs. $\mathcal{S}_1, \ldots, \mathcal{S}_{10}$ indicate the partitions of lowest to highest eigenvalues of projection. The mean NPS shrinkage weights (red line) and their 95% CIs (red shade) were estimated from 5 replicates. No shrinkage line (green) indicates $\omega_k = 1$. The number of markers *M* is 101,296. The discovery GWAS size *N* equals to *M*. The heritability h^2 is 0.5. The fraction of causal SNPs is 1%. See Figure 2C-D for $\mathcal{S}_2$ and $\mathcal{S}_{10}$, respectively.

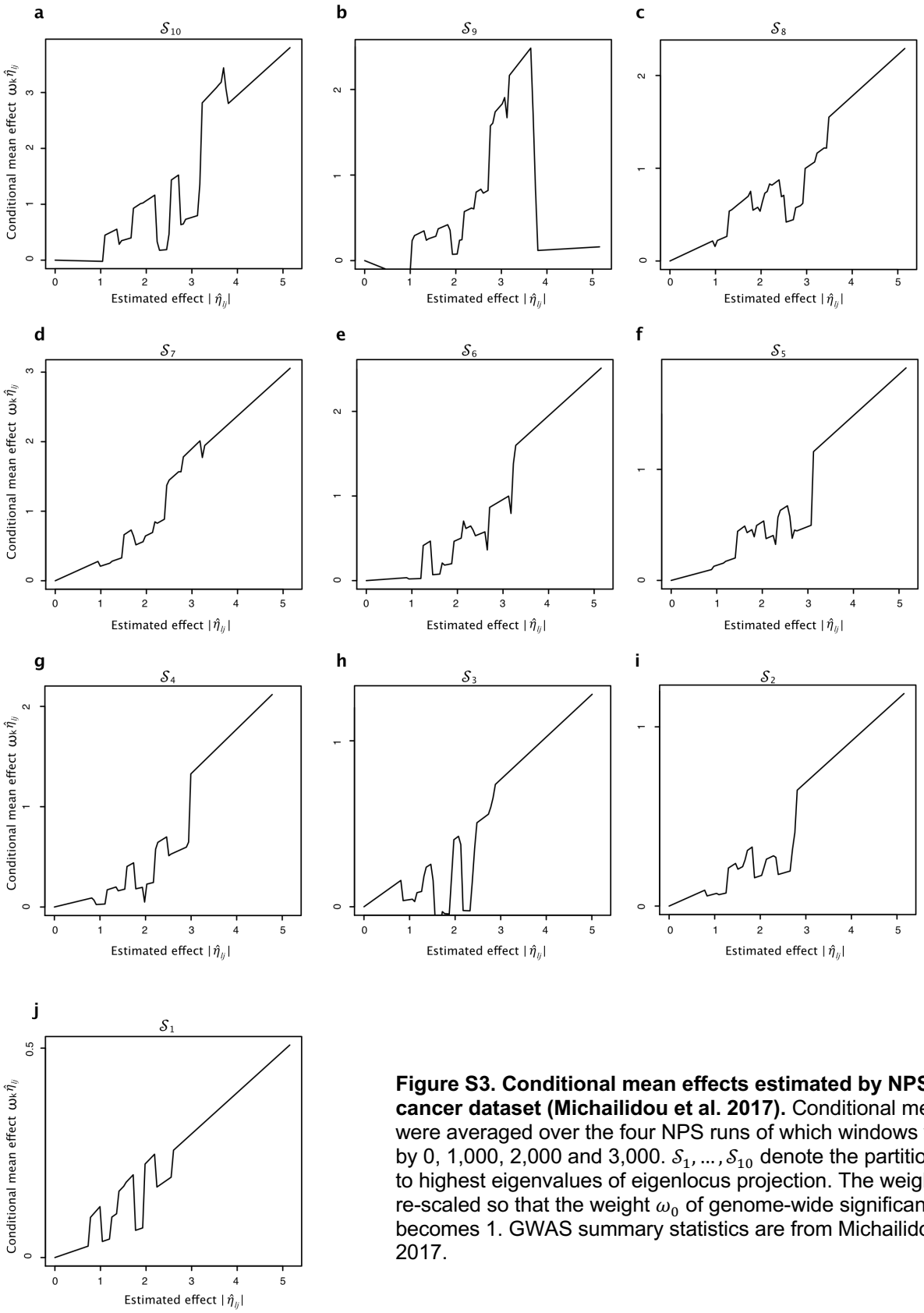

Figure S3. Conditional mean effects estimated by NPS in breast cancer dataset (Michailidou et al. 2017). Conditional mean effects were averaged over the four NPS runs of which windows were shifted by 0, 1,000, 2,000 and 3,000. $\mathcal{S}_1, \ldots, \mathcal{S}_{10}$ denote the partitions of lowest to highest eigenvalues of eigenlocus projection. The weights ω_k were re-scaled so that the weight ω_0 of genome-wide significant partition $\mathcal{S}_0$ becomes 1. GWAS summary statistics are from Michailidou et al. 2017.

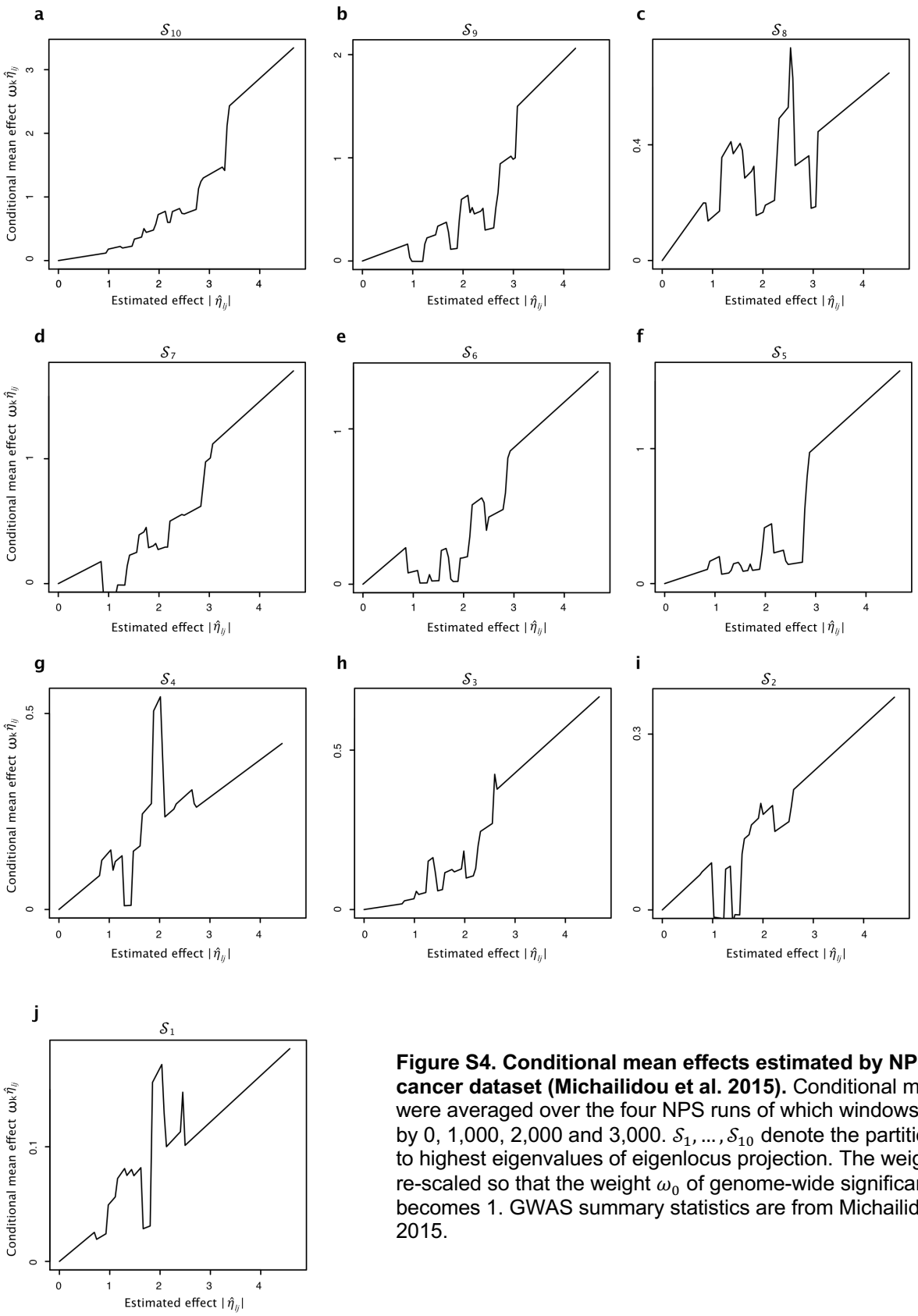

Figure S4. Conditional mean effects estimated by NPS in breast cancer dataset (Michailidou et al. 2015). Conditional mean effects were averaged over the four NPS runs of which windows were shifted by 0, 1,000, 2,000 and 3,000. $\mathcal{S}_1, \ldots, \mathcal{S}_{10}$ denote the partitions of lowest to highest eigenvalues of eigenlocus projection. The weights ω_k were re-scaled so that the weight ω_0 of genome-wide significant partition $\mathcal{S}_0$ becomes 1. GWAS summary statistics are from Michailidou et al. 2015.

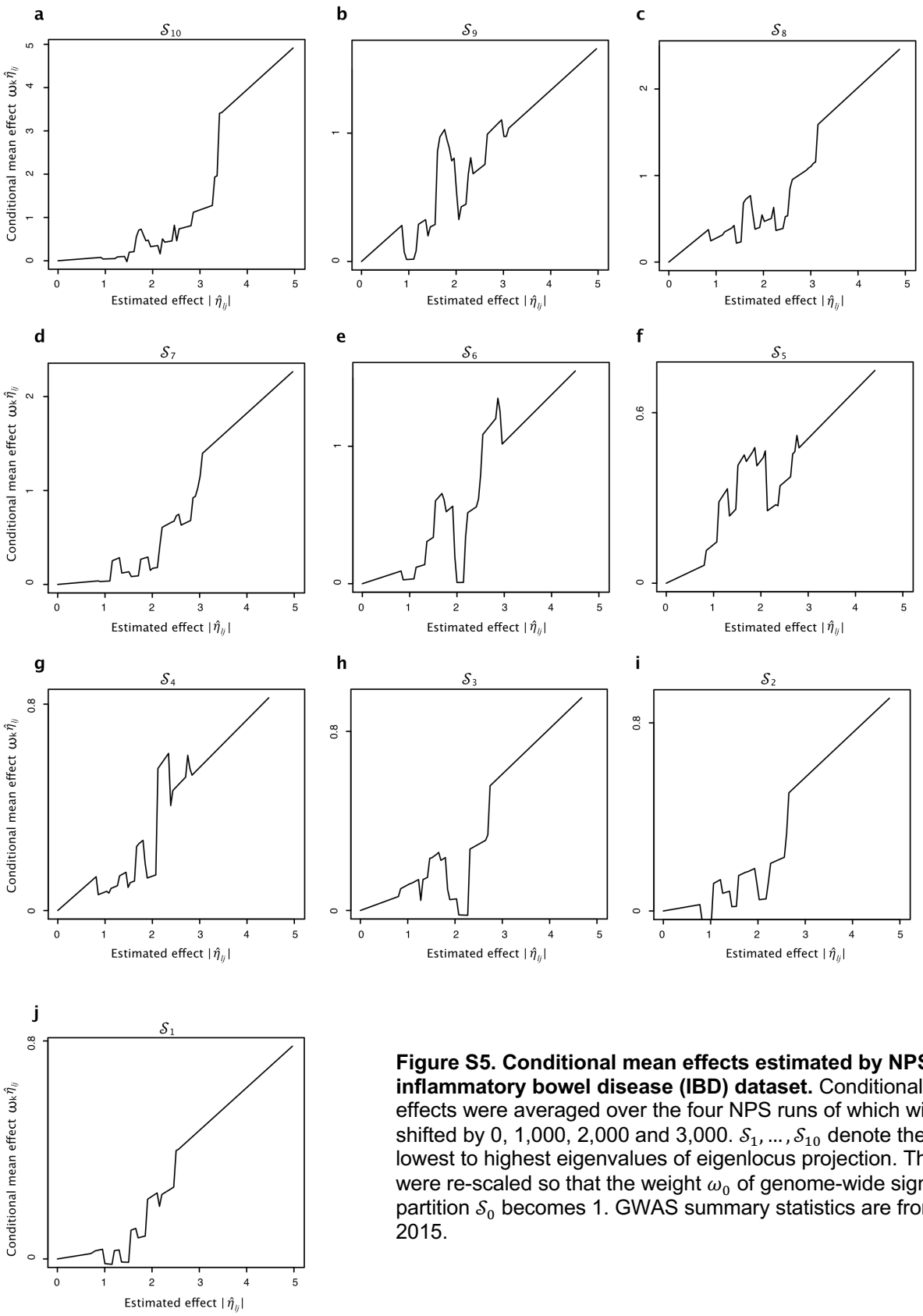

Figure S5. Conditional mean effects estimated by NPS in inflammatory bowel disease (IBD) dataset. Conditional mean effects were averaged over the four NPS runs of which windows were shifted by 0, 1,000, 2,000 and 3,000. $\mathcal{S}_1, \ldots, \mathcal{S}_{10}$ denote the partitions of lowest to highest eigenvalues of eigenlocus projection. The weights ω_k were re-scaled so that the weight ω_0 of genome-wide significant partition $\mathcal{S}_0$ becomes 1. GWAS summary statistics are from Liu et al. 2015.

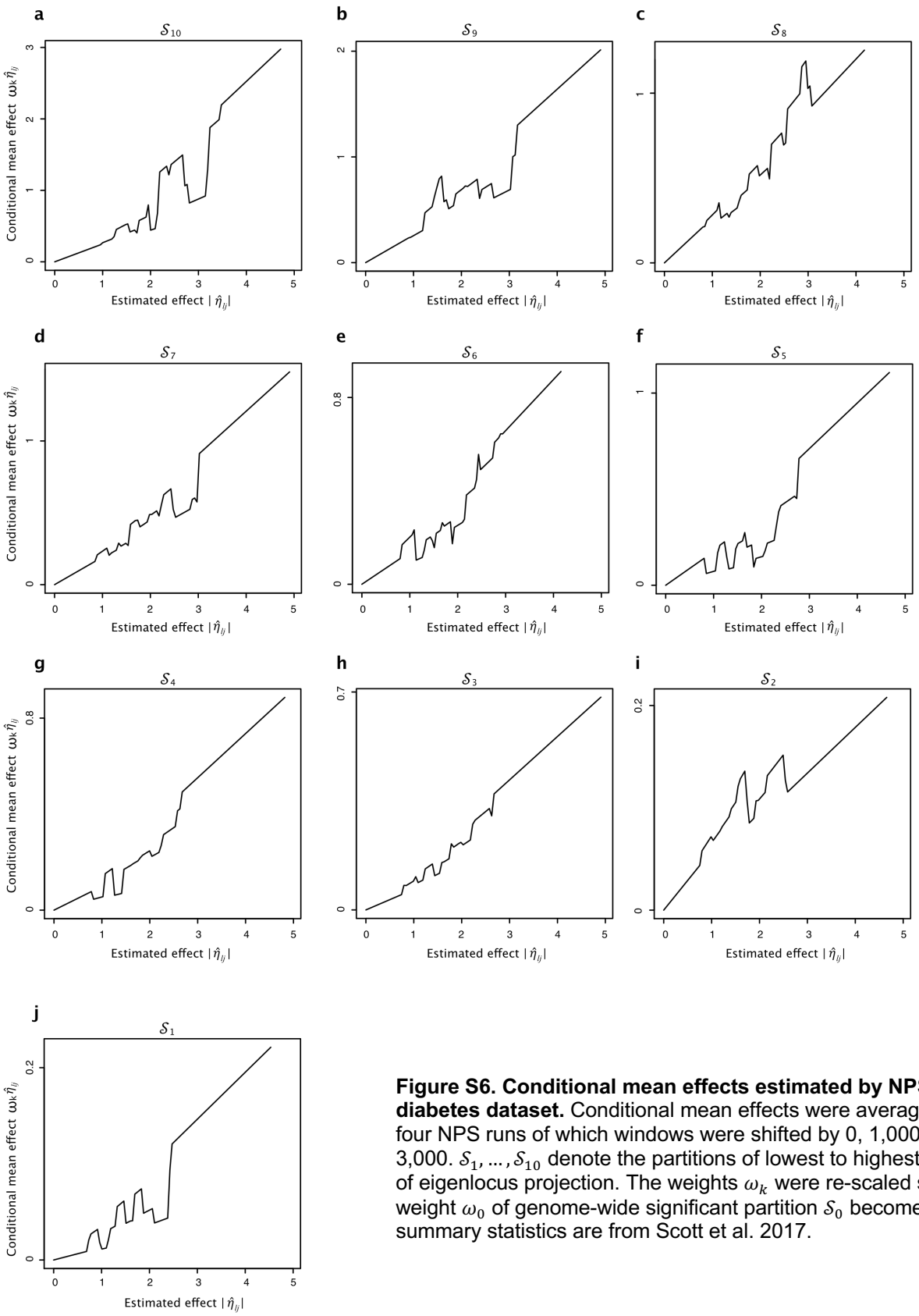

Figure S6. Conditional mean effects estimated by NPS in type 2 diabetes dataset. Conditional mean effects were averaged over the four NPS runs of which windows were shifted by 0, 1,000, 2,000 and 3,000. $\mathcal{S}_1, \ldots, \mathcal{S}_{10}$ denote the partitions of lowest to highest eigenvalues of eigenlocus projection. The weights ω_k were re-scaled so that the weight ω_0 of genome-wide significant partition $\mathcal{S}_0$ becomes 1. GWAS summary statistics are from Scott et al. 2017.

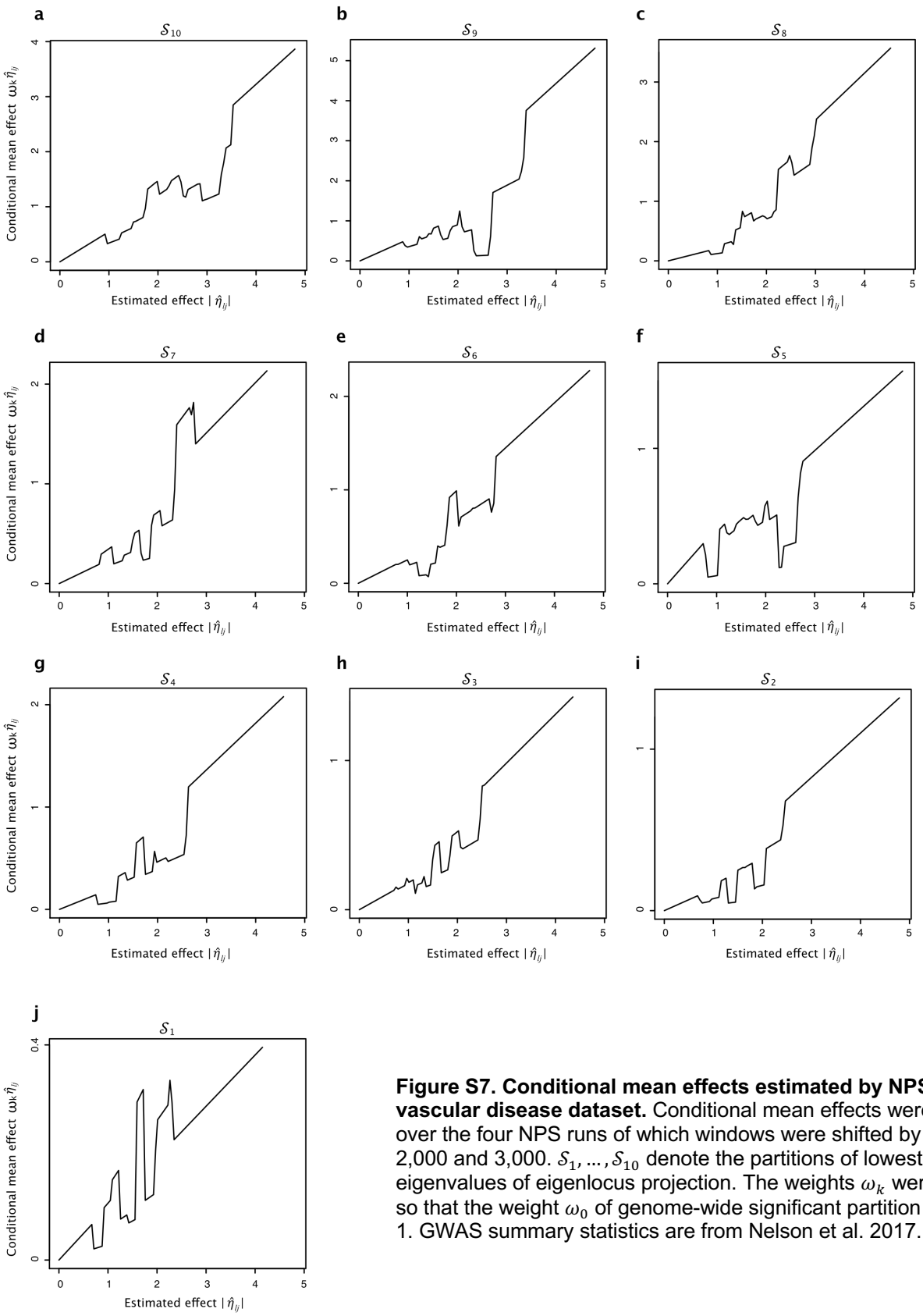

Figure S7. Conditional mean effects estimated by NPS in cardiovascular disease dataset. Conditional mean effects were averaged over the four NPS runs of which windows were shifted by 0, 1,000, 2,000 and 3,000. $\mathcal{S}_1, \ldots, \mathcal{S}_{10}$ denote the partitions of lowest to highest eigenvalues of eigenlocus projection. The weights ω_k were re-scaled so that the weight ω_0 of genome-wide significant partition $\mathcal{S}_0$ becomes 1. GWAS summary statistics are from Nelson et al. 2017.

Table S1. Comparison of prediction accuracy in genetic architectures simulating uniformly distributed causal SNPs.

Genetic Architecture	% causal SNPs	Method	Validation		NPS R^2_{Nag} compared to		
			$R^2_{Nagelkerke}$	$R^2_{Liability}$	P+T	LDPred	PRS-CS
(**a**) Point-Normal (GCTA)	1%	P+T	0.049	0.072			
		LDPred	0.071	0.103			
		PRS-CS	0.072	0.105			
		NPS	0.082	0.120	1.66 *	1.15 *	1.14 *
	0.1%	P+T	0.141	0.205			
		LDPred	0.071	0.102			
		PRS-CS	0.140	0.199			
		NPS	0.169	0.241	1.20 *	2.37 *	1.21 *
	0.01%	P+T	0.189	0.273			
		LDPred	0.076	0.110			
		PRS-CS	0.224	0.325			
		NPS	0.329	0.465	1.74 *	4.36 *	1.47 *
(**b**) Point-Normal with MAF dependency ($\alpha = -0.25$)	1%	P+T	0.050	0.071			
		LDPred	0.073	0.101			
		PRS-CS	0.081	0.115			
		NPS	0.093	0.131	1.87 *	1.27 *	1.16 *
	0.1%	P+T	0.142	0.206			
		LDPred	0.076	0.112			
		PRS-CS	0.152	0.220			
		NPS	0.175	0.253	1.24 *	2.31 *	1.15 *
	0.01%	P+T	0.199	0.293			
		LDPred	0.087	0.126			
		PRS-CS	0.230	0.330			
		NPS	0.329	0.471	1.66 *	3.78 *	1.43 *

NPS is more accurate than Pruning and Thresholding (P+T), LDPred and PRS-CS in simulated datasets. Here, two sets of Point-Normal architectures were simulated: (**a**) a spike-and-slab GCTA model which assumes the independence of heritability on minor allele frequency (MAF) and (**b**) an architecture incorporating the dependency of heritability on MAF ($\alpha = -0.25$). Under each model and for each causal fraction, three instances of genetic architecture were generated. Recent studies have found that low frequency SNPs contribute less heritability than previously expected under no dependency (Speed et al. 2017, Zeng et al. 2018). Low-frequency SNPs tend to be captured by eigenvectors of small eigenvalues and are challenging to handle with spectral decomposition. More realistic simulations (**b**) lowering the overall heritability contribution of low-frequency SNPs made NPS slightly more accurate than under (**a**) GCTA models. Binary phenotypes were simulated with the heritability of 0.5 on the liability scale and prevalence of 5%. The number of markers was 5,012,500. The GWAS sample size was 100,000. Prediction models were optimized in the training cohort of 2,500 cases and 2,500 controls. The prediction accuracies were measured in the validation cohort of 50,000 samples and averaged over three simulations. The star (*) indicates that Nagelkerke's R^2 is significantly different (paired t-test; $P < 0.05$).

Table S2. Accuracy of NPS in genetic architectures simulating the enrichment of causal SNPs within DNase I Hypersensitive Sites (DHS).

Fraction of causal SNPs	Training	Validation		
	AUC	$R^2_{Nagelkerke}$	$R^2_{Liability}$	AUC
1%	0.746	0.082	0.126	0.708
	0.737	0.083	0.125	0.708
	0.725	0.089	0.118	0.716
0.1%	0.800	0.174	0.249	0.793
	0.811	0.188	0.262	0.808
	0.802	0.179	0.261	0.798
	0.810	0.176	0.254	0.798
	0.810	0.176	0.250	0.798
	0.813	0.178	0.259	0.799
0.01%	0.891	0.325	0.463	0.887
	0.894	0.323	0.462	0.885
	0.887	0.336	0.463	0.889

Each row represents the prediction accuracy of NPS in an individual simulation run. The prediction accuracy of NPS decreased slightly compared to simulations of uniformly distributed causal SNPs (Table S1) but still remained robust. We did not train NPS prediction models using functional annotations. The causal fractions of 1% and 0.01% were replicated three times each, and the causal fraction of 0.1% was replicated six times. The simulation incorporates the dependency of heritability on minor allele frequency ($\alpha = -0.25$) and five-fold enrichment of causal SNPs in DHS elements. Binary phenotypes were simulated with the heritability of 0.5 on the liability scale and prevalence of 5%. The number of markers was 5,012,500. The GWAS sample size was 100,000. Prediction models were optimized in the training cohort of 2,500 cases and 2,500 controls. The prediction accuracies were measured in validation cohorts of 50,000 samples. AUC – Area Under the Curve.

Table S3. Accuracy of LDPred in genetic architectures simulating the enrichment of causal SNPs within DNase I Hypersensitive Sites (DHS).

Fraction of causal SNPs (p)	Input SNPs	Training		Validation		
		Estimated p	AUC	$R^2_{Nagelkerke}$	$R^2_{Liability}$	AUC
1%	All SNPs (M=5,012,500)	1.0	0.706	0.065	0.100	0.684
		1.0	0.695	0.068	0.102	0.689
		1.0	0.686	0.071	0.105	0.693
0.1%		0.3	0.695	0.080	0.108	0.705
		1.0	0.690	0.083	0.116	0.711
		1.0	0.686	0.075	0.107	0.699
		0.3	0.698	0.078	0.118	0.704
		1.0	0.693	0.069	0.103	0.694
		0.1	0.644	0.098	0.140	0.727
0.01%		0.3	0.726	0.093	0.141	0.721
		0.3	0.723	0.098	0.143	0.729
		0.01	0.840	0.268	0.373	0.854
1%	Genotyped SNPs Only (M=490,504)	1.0	0.699	0.062	0.094	0.680
		1.0	0.683	0.062	0.095	0.680
		1.0	0.674	0.066	0.095	0.687
0.1%		0.003	0.756	0.149	0.210	0.773
		1.0	0.679	0.079	0.106	0.707
		0.0001	0.729	0.116	0.165	0.715
		0.001	0.765	0.138	0.197	0.764
		0.3	0.718	0.100	0.144	0.730
		0.0003	0.753	0.123	0.183	0.753
0.01%		0.0003	0.786	0.150	0.222	0.780
		0.001	0.749	0.115	0.166	0.743
		0.001	0.816	0.222	0.317	0.827

Each row represents the prediction accuracy of LDPred in an individual simulation run. The causal fractions of 1% and 0.01% were replicated three times each, and 0.1% was replicated six times. The simulation incorporates the dependency of heritability on MAF ($\alpha = -0.25$) and five-fold enrichment of causal SNPs in DHS. Binary phenotypes were simulated with the heritability of 0.5 on the liability scale and prevalence of 5%. LDPred was run using all 5,012,500 SNPs (top) as well as a sparse set of 490,504 SNPs taken from HumanHap550v3 genotyping array (bottom). With sparse SNPs, LDPred converged to closer-to-truth simulated causal fractions and resulted a higher average but lower maximum accuracy than using all markers. The prediction model reaching the highest accuracy in a training cohort was selected for validation. The estimated causal fraction (p) represents the causal fraction of best performing prediction model in training. p=1.0 denotes the infinitesimal model in which all SNPs are causal. The GWAS sample size was 100,000. Prediction models were optimized in the training cohort of 2,500 cases and 2,500 controls. The prediction accuracies were measured in validation cohorts of 50,000 samples. AUC – Area Under the Curve.

Table S4. Accuracy of pruning and thresholding in genetic architectures simulating the enrichment of causal SNPs within DNase I Hypersensitive Sites (DHS).

Fraction of causal SNPs	Training			Validation		
	P cutoff	# SNPs	AUC	$R^2_{Nagelkerke}$	$R^2_{Liability}$	AUC
1%	0.046	57,816	0.680	0.047	0.072	0.662
	0.097	92,163	0.661	0.050	0.076	0.664
	0.153	121,820	0.664	0.054	0.075	0.670
0.1%	0.0001	2,082	0.783	0.174	0.244	0.793
	0.00015	2,562	0.751	0.133	0.186	0.761
	0.0002	2,765	0.735	0.119	0.164	0.747
	0.0001	2,147	0.795	0.160	0.247	0.787
	0.0001	2,296	0.736	0.105	0.163	0.738
	0.00015	2,529	0.759	0.128	0.190	0.757
0.01%	0.0001	1,662	0.827	0.209	0.305	0.823
	0.0001	1,631	0.807	0.176	0.263	0.797
	0.0001	1,553	0.833	0.252	0.352	0.848

Each row represents the prediction accuracy of pruning and thresholding (P+T) algorithm in an individual simulation run. The causal fractions of 1% and 0.01% were replicated three times each, and the causal fraction of 0.1% were replicated six times. The simulation incorporates the dependency of heritability on minor allele frequency ($\alpha = -0.25$) and five-fold enrichment of causal SNPs in DHS elements. Binary phenotypes were simulated with the heritability of 0.5 on the liability scale and prevalence of 5%. The prediction model reaching the highest accuracy in a training cohort was selected for validation. The P-value cutoff of best-performing model is reported here along with the number of SNPs after pruning and thresholding. The GWAS sample size was 100,000. Prediction models were optimized in the training cohort of 2,500 cases and 2,500 controls. The prediction R^2 was measured in validation cohorts of 50,000 samples. AUC – Area Under the Curve.

Table S5. Accuracy of PRS-CS in genetic architectures simulating the enrichment of causal SNPs within DNase I Hypersensitive Sites (DHS).

Fraction of causal SNPs	Training		Validation		
	$\widehat{\phi}$	AUC	R^2_{Nag}	$R^2_{Liability}$	AUC
1%	0.01	0.720	0.072	0.110	0.693
	0.0001	0.696	0.074	0.107	0.697
	0.01	0.700	0.079	0.113	0.705
0.1%	0.0001	0.771	0.157	0.221	0.780
	0.0001	0.773	0.164	0.227	0.789
	0.0001	0.769	0.155	0.224	0.781
	0.0001	0.782	0.156	0.226	0.781
	0.0001	0.768	0.148	0.217	0.777
	0.0001	0.777	0.157	0.229	0.781
0.01%	0.000001	0.835	0.230	0.332	0.835
	0.000001	0.835	0.222	0.322	0.830
	0.000001	0.819	0.232	0.326	0.833

Each row represents the prediction accuracy of PRS-CS in an individual simulation run. The causal fractions of 1% and 0.01% were replicated three times each, and the causal fraction of 0.1% were replicated six times. The simulation incorporates the dependency of heritability on minor allele frequency ($\alpha = -0.25$) and five-fold enrichment of causal SNPs in DHS elements. Binary phenotypes were simulated with the heritability of 0.5 on the liability scale and prevalence of 5%. The prediction model reaching the highest accuracy in a training cohort was selected for validation. $\hat{\phi}$ denotes the model parameter ϕ of best-performing model in training. The reference LD panel was derived from a cohort sampled under the same LD structure. The GWAS sample size was 100,000. Prediction models were optimized in the training cohort of 2,500 cases and 2,500 controls. The prediction R^2 was measured in validation cohorts of 50,000 samples. AUC – Area Under the Curve.

Table S6. Accuracy of NPS applied to real GWAS summary statistics and UK Biobank datasets.

	Training		Validation (UK Biobank)	
GWAS	**# Projections**	**AUC**	**AUC**	**Tail OR (5%)**
Breast Cancer 2015	120,886	0.656	0.627 [0.62-0.64]	2.53 [2.3-2.8]
Breast Cancer 2017	124,061	0.678	0.654 [0.65-0.66]	3.01 [2.7-3.3]
IBD	110,157	0.686	0.659 [0.65-0.67]	3.60 [3.2-4.0]
Type 2 Diabetes	139,106	0.697	0.686 [0.68-0.69]	3.81 [3.6-4.1]
CAD	105,162	0.778	0.738 [0.72-0.76]	5.21 [4.3-6.2]

GWAS summary statistics for breast cancer, inflammatory bowel disease (IBD), type 2 diabetes, coronary artery disease (CAD) were obtained from Michailidou et al. 2015, Michailidou et al. 2017, Liu et al. 2015, Scott et al. 2017, and Nelson et al. 2017, respectively. The training and validation cohorts were both assembled using UK Biobank samples (see Table 2). The number of projections represents the total number of independent projection eigenvectors used for NPS training across the genome. The 5% tail OR denotes the odds ratio at the 5% highest risk tail compared to the rest of cohort. The numbers in brackets are the 95% confidence intervals for AUCs (Area Under the Curve) and tail ORs, which were estimated by DeLong's method and bootstrapping, respectively. T2D and CAD models were trained and validated with the sex covariate.

Table S7. Accuracy of LDPred applied to real GWAS summary statistics and UK Biobank datasets.

GWAS	Training			Validation (UK Biobank)	
	# SNPs	Estimated causal fraction	AUC	AUC	Tail OR (5%)
Breast Cancer 2015	3,417,759	0.01	**0.630**	0.618 [0.61-0.63]	2.42 [2.2-2.7]
Breast Cancer 2017	3,478,993	0.1	**0.621**	0.615 [0.61-0.62]	2.33 [2.1-2.6]
IBD	3,396,783	0.03	**0.640**	0.641 [0.63-0.65]	2.77 [2.4-3.1]
Type 2 Diabetes	3,451,818	0.01	**0.680**	0.679 [0.67-0.68]	3.51 [3.3-3.8]
CAD	3,405,299	0.003	0.753	0.738 [0.72-0.76]	5.17 [4.3-6.1]
Breast Cancer 2015	351,917	0.3	0.605	0.597 [0.59-0.61]	2.25 [2.0-2.5]
Breast Cancer 2017	353,627	1.0	0.606	0.604 [0.60-0.61]	2.03 [1.8-2.3]
IBD	353,325	1.0	0.618	0.622 [0.61-0.63]	2.76 [2.4-3.1]
Type 2 Diabetes	354,110	0.1	0.679	0.680 [0.67-0.69]	3.63 [3.4-3.9]
CAD	329,644	0.03	**0.757**	0.742 [0.72-0.76]	5.65 [4.7-6.7]

GWAS summary statistics for breast cancer, inflammatory bowel disease (IBD), type 2 diabetes, coronary artery disease (CAD) were obtained from Michailidou et al. 2015, Michailidou et al. 2017, Liu et al. 2015, Scott et al. 2017, and Nelson et al. 2017, respectively. The training and validation cohorts were both assembled using UK Biobank samples (see Table 2). LDPred was ran using all hard-called common SNPs (top) as well as directly genotyped SNPs (bottom). The prediction models producing a higher AUCs in training cohorts, indicated in bold, were chosen for Table 2. LDPred runs only with hard-called genotypes and automatically excludes complementary alleles; therefore, the number of input SNPs are fewer than the number of all available imputed SNPs across the genome. The estimated causal fraction represents the causal fraction parameter of best performing prediction model in training cohort. The estimated causal fraction of 1.0 denotes the infinitesimal model in which all SNPs are causal. The tail OR denotes the odds ratio at the 5% highest risk tail compared to the rest of cohort. The numbers in brackets are the 95% confidence intervals for AUCs (Area Under the Curve) and tail ORs, which were estimated by DeLong's method and bootstrapping, respectively. T2D and CAD models were trained and validated with the sex covariate.

Table S8. Accuracy of pruning and thresholding applied to real GWAS summary statistics and UK Biobank datasets.

GWAS	Training			Validation (UK Biobank)	
	P cutoff	# SNPs	AUC	AUC	Tail OR (5%)
Breast Cancer 2015	0.0001	427	0.615	0.607 [0.60-0.62]	2.07 [1.9-2.3]
Breast Cancer 2017	0.0003	1,521	0.627	0.621 [0.61-0.63]	2.37 [2.1-2.6]
IBD	0.0002	621	0.648	0.644 [0.63-0.65]	3.00 [2.7-3.4]
Type 2 Diabetes	0.0004	691	0.661	0.659 [0.65-0.67]	3.04 [2.8-3.3]
CAD	0.025	8,915	0.739	0.719 [0.70-0.74]	5.17 [4.3-6.1]

GWAS summary statistics for breast cancer, inflammatory bowel disease (IBD), type 2 diabetes, coronary artery disease (CAD) were obtained from Michailidou et al. 2015, Michailidou et al. 2017, Liu et al. 2015, Scott et al. 2017, and Nelson et al. 2017, respectively. The training and validation cohorts were both assembled using UK Biobank samples (see Table 2). The prediction model reaching the highest accuracy in a training cohort was selected for validation. The P-value cutoff of best-performing model is reported here along with the number of SNPs after pruning and thresholding. The tail OR denotes the odds ratio at the 5% highest risk tail compared to the rest of cohort. The numbers in brackets are the 95% confidence intervals for AUC (Area Under the Curve) and tail OR, which were estimated by DeLong's method and bootstrapping, respectively. T2D and CAD models were trained and validated with the sex covariate.

Table S9. Accuracy of PRS-CS applied to real GWAS summary statistics and UK Biobank datasets.

GWAS	Training			Validation (UK Biobank)	
	# SNPs	$\hat{\phi}$	AUC	AUC	Tail OR (5%)
Breast Cancer 2015	712,303	0.0001	0.635	0.626 [0.62-0.63]	2.60 [2.3-2.9]
Breast Cancer 2017	711,549	0.0001	0.657	0.651 [0.64-0.66]	2.96 [2.7-3.3]
IBD	707,371	0.0001	0.665	0.668 [0.66-0.68]	3.67 [3.3-4.1]
Type 2 Diabetes	715,952	0.0001	0.686	0.688 [0.68-0.69]	3.99 [3.7-4.3]
CAD	708,976	0.0001	0.763	0.739 [0.72-0.76]	4.92 [4.1-5.8]

GWAS summary statistics for breast cancer, inflammatory bowel disease (IBD), type 2 diabetes, coronary artery disease (CAD) were obtained from Michailidou et al. 2015, Michailidou et al. 2017, Liu et al. 2015, Scott et al. 2017, and Nelson et al. 2017, respectively. The training and validation cohorts were both assembled using UK Biobank samples (see Table 2). The prediction model reaching the highest accuracy in a training cohort was selected for validation. $\hat{\phi}$ denotes the model parameter ϕ of best-performing model in training. The European reference LD panel provided with the software was used for training. PRS-CS uses only HapMap 3 SNPs. The tail OR denotes the odds ratio at the 5% highest risk tail compared to the rest of cohort. The numbers in brackets are the 95% confidence intervals for AUC (Area Under the Curve) and tail OR, which were estimated by DeLong's method and bootstrapping, respectively. T2D and CAD models were trained and validated with the sex covariate.

.

Table S10. Accuracy of NPS in independent validation cohorts.

	Training		Validation (Partners Biobank)	
GWAS	**# Projections**	**AUC**	**AUC**	**Tail OR (5%)**
Breast Cancer 2017	124,061	0.678	0.624 [0.60-0.64]	2.32 [1.7-3.0]
IBD	110,157	0.686	0.686 [0.67-0.70]	4.32 [3.5-5.2]
Type 2 Diabetes	139,106	0.697	0.647 [0.63-0.66]	2.97 [2.6-3.5]
CAD	105,162	0.778	0.615 [0.58-0.65]	4.10 [2.8-5.8]

The polygenic risk models trained in UK Biobank (Table 2) were validated in US white population (Table 3; Partners Biobank). The identical polygenic risk prediction models reported in Tables 2 and S6 were validated in Partners Biobank without re-training or model adjustment. The tail OR denotes the odds ratio at the 5% highest risk tail compared to the rest of cohort. The numbers in brackets are the 95% confidence intervals for AUC (Area Under the Curve) and tail OR, which were estimated by DeLong's method and bootstrapping, respectively. T2D and CAD models were trained and validated with the sex covariate.

Table S11. Accuracy of LDPred in independent validation cohorts.

GWAS	Training (UK Biobank)			Validation (Partners Biobank)	
	# SNPs	Estimated causal fraction	AUC	AUC	Tail OR (5%)
Breast Cancer 2017	1,261,292	0.1	0.600	0.580 [0.56-0.60]	1.78 [1.3-2.3]
IBD	1,238,654	0.03	0.609	0.639 [0.62-0.66]	3.07 [2.5-3.8]
Type 2 Diabetes	1,243,787	0.01	0.665	0.635 [0.62-0.65]	2.51 [2.1-2.9]
CAD	1,224,034	0.003	0.724	0.595 [0.56-0.63]	2.31 [1.4-3.5]

The polygenic risk models were trained with LDPred in UK Biobank cohorts and validated in US white population (Table 3; Partners Biobank). The training cohorts are identical to those in Tables 2 and S7, however, the prediction models were reconstructed by re-running LDPred on the SNPs found in both training and validation cohorts as recommended by the authors. LDPred runs only with hard genotypes and automatically excludes complementary alleles; therefore, the number of hard-called input SNPs are fewer than the number of all available imputed SNPs. The estimated causal fraction represents the causal fraction parameter of best performing prediction model in training cohort. The estimated causal fraction of 1.0 denotes the infinitesimal model in which all SNPs are causal. See Table 3 for case/control sample sizes of validation cohorts. The tail OR denotes the odds ratio at the 5% highest risk tail compared to the rest of cohort. The numbers in brackets are the 95% confidence intervals for AUC (Area Under the Curve) and tail OR, which were estimated by DeLong's method and bootstrapping, respectively. T2D and CAD models were trained and validated with the sex covariate.

Table S12. Accuracy of pruning and thresholding in independent validation cohorts.

GWAS	Training (UK Biobank)			Validation (Partners Biobank)	
	P cutoff	# SNPs	AUC	AUC	Tail OR (5%)
Breast Cancer 2017	0.00035	801	0.613	0.589 [0.57-0.61]	1.56 [1.2-2.1]
IBD	0.0002	331	0.629	0.659 [0.64-0.68]	3.57 [2.9-4.4]
Type 2 Diabetes	0.0001	165	0.656	0.623 [0.61-0.64]	2.10 [1.8-2.5]
CAD	0.15	15,908	0.739	0.611 [0.58-0.65]	2.72 [1.8-3.9]

The polygenic risk models were trained with pruning and thresholding algorithm in UK Biobank cohorts and validated in US white population (Table 3; Partners Biobank). The training cohorts are identical to those in Tables 2 and S8, however, the prediction models were reconstructed by re-running P+T on the SNPs found in both training and validation cohorts. However, the prediction models were reconstructed by re-running pruning and thresholding algorithm on the SNPs found in both training and validation cohorts. The prediction model reaching the highest accuracy in a training cohort was selected for validation. The P-value cutoff of best-performing model is reported here along with the number of SNPs after pruning and thresholding. See Table 3 for case/control sample sizes of validation cohorts. The tail OR denotes the odds ratio at the 5% highest risk tail compared to the rest of cohort. The numbers in brackets are the 95% confidence intervals for AUC (Area Under the Curve) and tail OR, which were estimated by DeLong's method and bootstrapping, respectively. T2D and CAD models were trained and validated with the sex covariate.

Table S13. Accuracy of PRS-CS in independent validation cohorts.

GWAS	Training			Validation (Partners Biobank)	
	# SNPs	Estimated ϕ	AUC	AUC	Tail OR (5%)
Breast Cancer 2017	512,117	0.0001	0.647	0.624 [0.60-0.64]	2.23 [1.7-2.9]
IBD	509,143	0.0001	0.663	0.682 [0.66-0.70]	4.11 [3.3-5.0]
Type 2 Diabetes	515,164	0.0001	0.685	0.649 [0.64-0.66]	2.80 [2.4-3.3]
CAD	510,103	0.0001	0.751	0.621 [0.58-0.66]	3.16 [2.1-4.4]

The polygenic risk models were trained with PRS-CS in UK Biobank cohorts and validated in US white population (Table 3; Partners Biobank). The training cohorts are identical to those in Tables 2 and S9, however, the prediction models were reconstructed by re-running PRS-CS on the SNPs found in both training and validation cohorts as recommended by the authors. PRS-CS uses only HapMap 3 SNPs. The prediction model reaching the highest accuracy in a training cohort was selected for validation. $\hat{\phi}$ denotes the model parameter ϕ of best-performing model in training. The European reference LD panel provided with the software was used for training. See Table 3 for case/control sample sizes of validation cohorts. The tail OR denotes the odds ratio at the 5% highest risk tail compared to the rest of cohort. The numbers in brackets are the 95% confidence intervals for AUC (Area Under the Curve) and tail OR, which were estimated by DeLong's method and bootstrapping, respectively. T2D and CAD models were trained and validated with the sex covariate.