The American Journal of Human Genetics

Supplemental Data

Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores

Bjarni J. Vilhjálmsson, Jian Yang, Hilary K. Finucane, Alexander Gusev, Sara Lindström, Stephan Ripke, Giulio Genovese, Po-Ru Loh, Gaurav Bhatia, Ron Do, Tristan Hayeck, Hong-Hee Won, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Discovery, Biology, and Risk of Inherited Variants in Breast Cancer (DRIVE) study, Sekar Kathiresan, Michele Pato, Carlos Pato, Rulla Tamimi, Eli Stahl, Noah Zaitlen, Bogdan Pasaniuc, Gillian Belbin, Eimear E. Kenny, Mikkel H. Schierup, Philip De Jager, Nikolaos A. Patsopoulos, Steve McCarroll, Mark Daly, Shaun Purcell, Daniel Chasman, Benjamin Neale, Michael Goddard, Peter M. Visscher, Peter Kraft, Nick Patterson, and Alkes L. Price

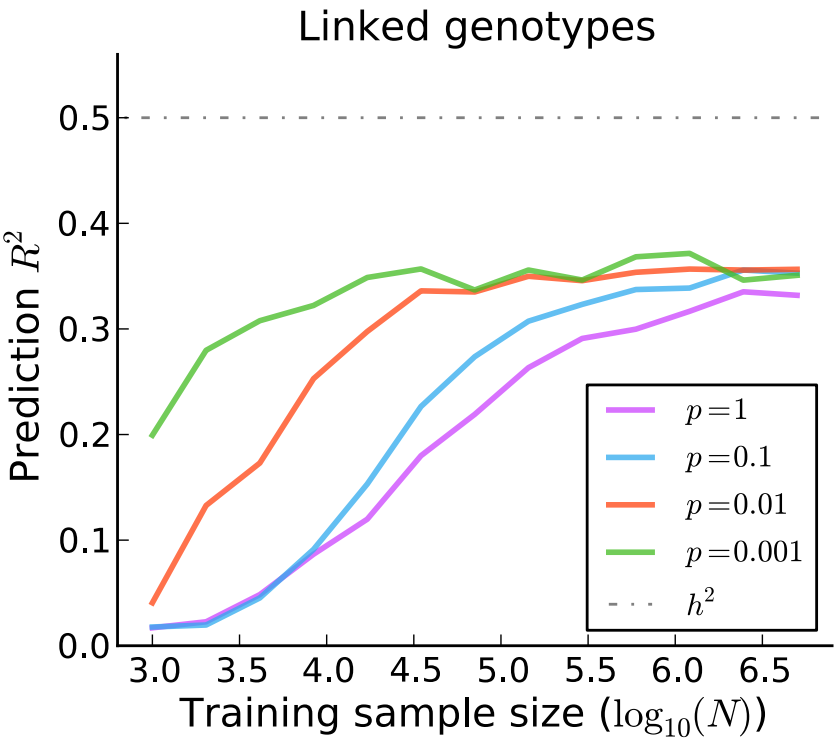

Figure S1. Performance of P+T when causal markers cluster. Performance of P+T (LD-pruning followed by thresholding) for an alternative genetic architecture where causal markers cluster. The results are averaged over 3000 simulated traits with 200K simulated genotypes where the average fraction of causal variants *p* was let vary. The simulated genotypes are linked, where we simulated independent batches of 100 markers where the squared correlation between adjacent variants in a batch was fixed to 0.9. For each simulated 100 SNP region of LD, we sampled the fraction of causal markers within a region from a Beta(*p*,*1-p*) distribution, ensuring that the expected fraction of causal markers across the genome is still *p*. This will cause causal variants to cluster in some regions of the genome. As expected, the impact of LD on the prediction accuracy of P+T is greater when causal variants cluster, and still substantial for small values of *p*.

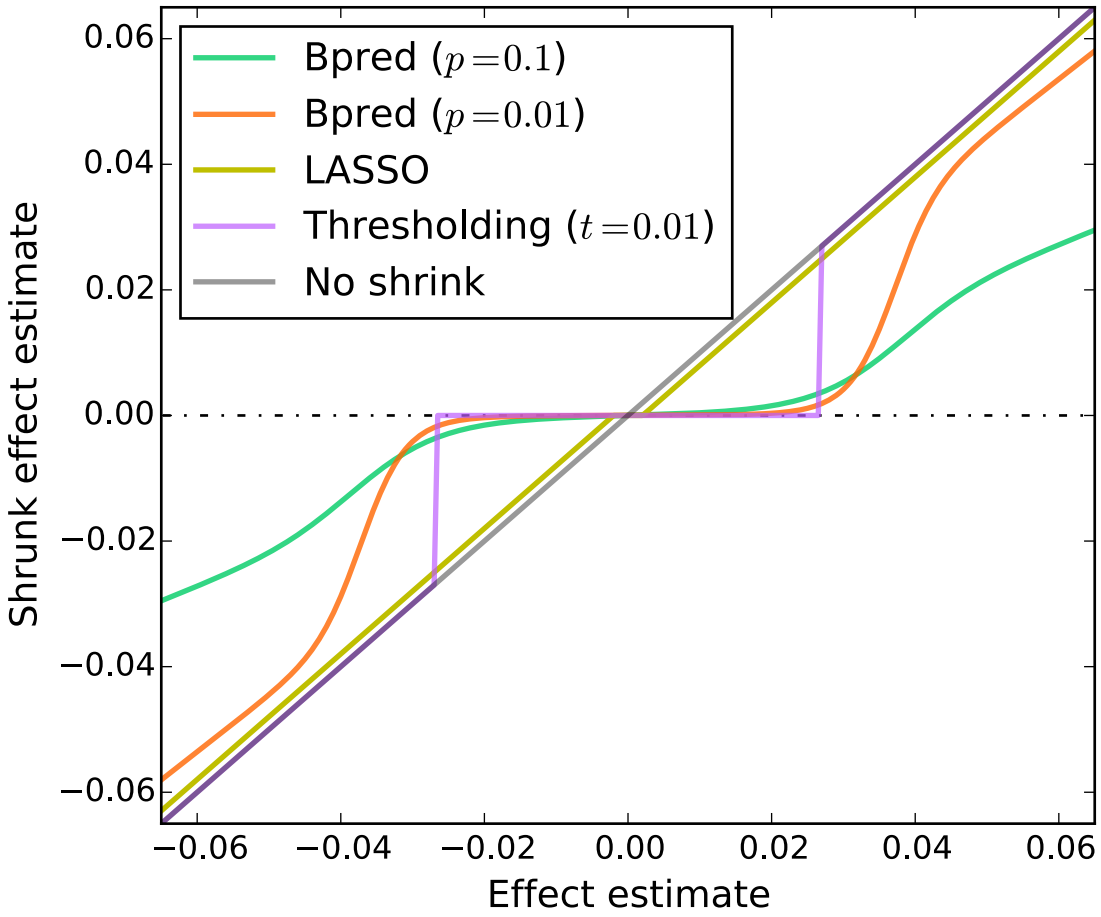

Figure S2. Comparison of five different shrinks in the absence of LD. Bpred corresponds to LDpred without LD and can be derived analytically (see Materials and Methods for details). The marginal (least square) effect estimate is plotted against the shrunk estimate for the five different shrinks. Bpred denotes the analytical solution to LDpred, which can be derived in the absence of LD (see Appendix A for details). The Bpred shrink shown here assumes that the heritability is 0.5 and the training sample size is 10,000 and the number of markers is 60,000. Similarly, the LASSO shrink shown here corresponds to the (marginal) posterior mode effect under a Laplace prior for the causal effects. Compared to *P*-value thresholding, and LASSO, Bpred can be viewed as a smoother shrink.

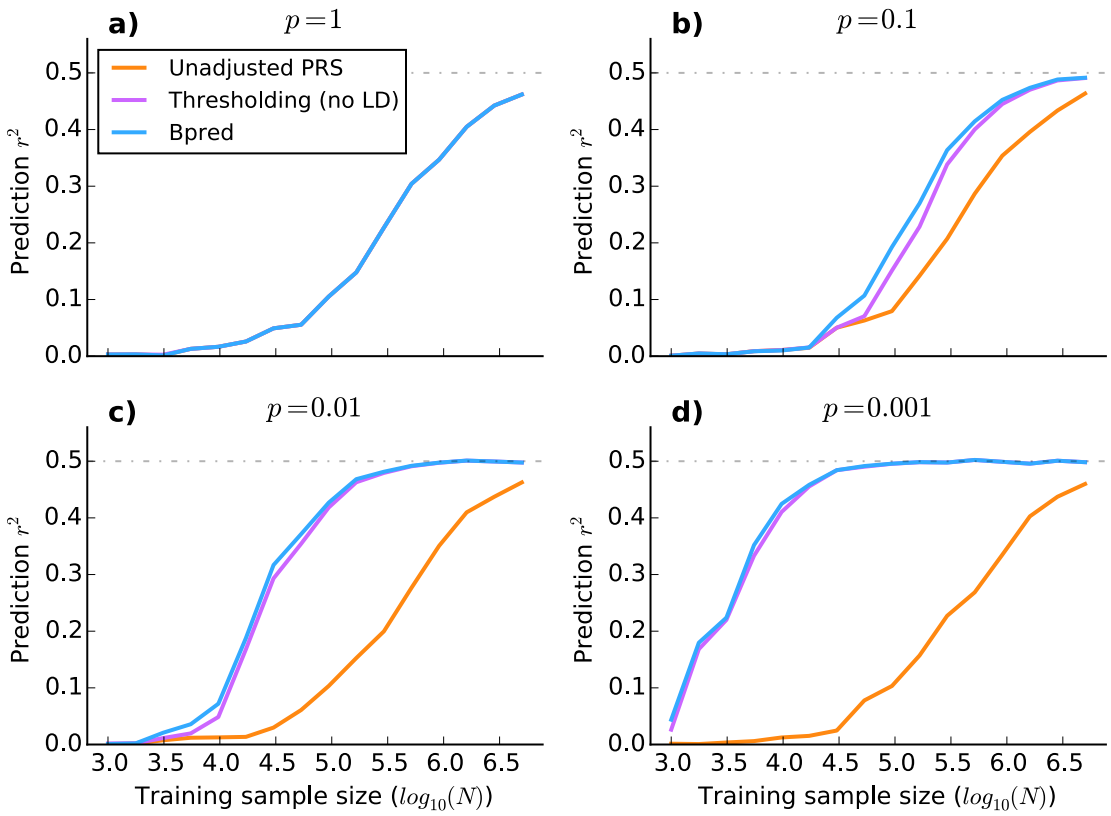

Figure S3. Comparison of methods using simulated genotypes without LD. The four subfigures **a-d** correspond to different genetic architectures where we vary *p*, the fraction of variants with (non-zero) effects drawn from a Gaussian distribution. Bpred denotes the analytical solution to LDpred, which can be derived in the absence of LD (see Appendix A for details). As expected, Bpred outperforms *P*-value thresholding in the absence of LD, although not by much.

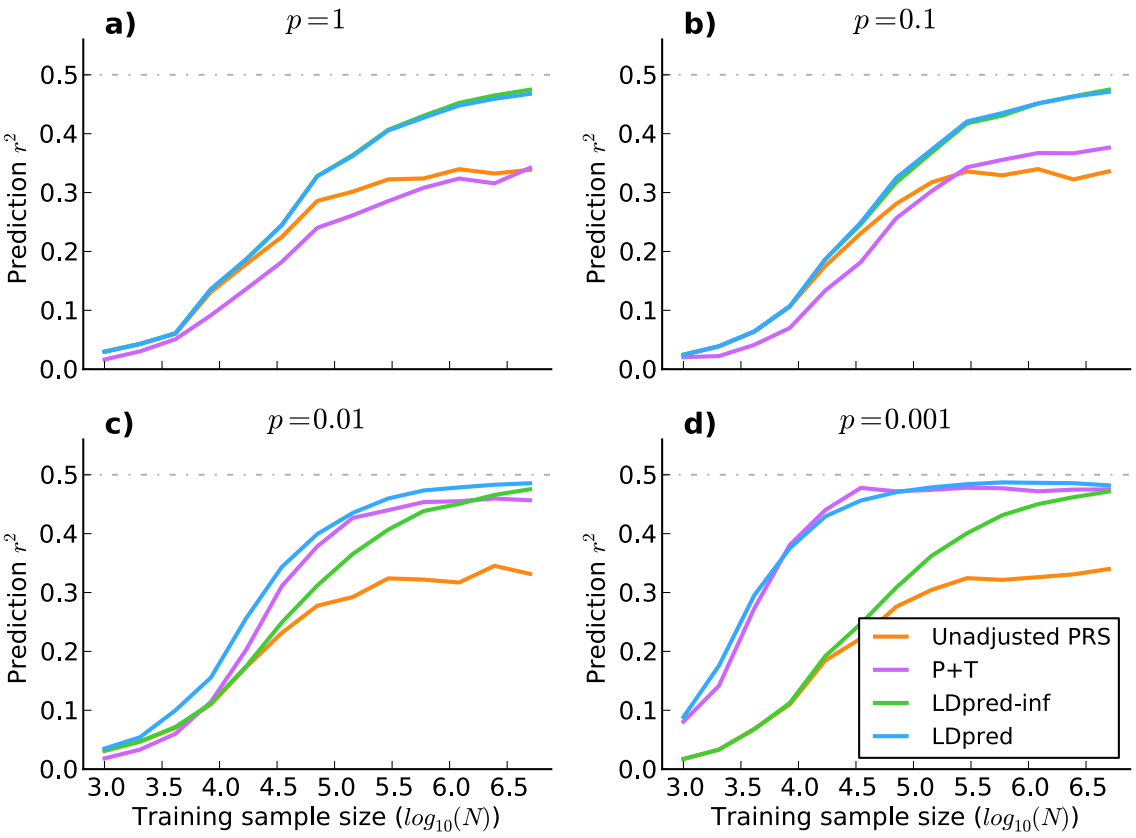

Figure S4. Comparison of methods using simulated genotypes with LD. The four subfigures **a-d** correspond to different genetic architectures where we vary *p*, the fraction of variants with (non-zero) effects drawn from a Gaussian distribution. We simulated marginal least square effect estimates with LD (see Materials and Methods for details). This enabled us to evaluate the behavior of the methods at large sample sizes. The LD structure consisted of 100 SNP regions where adjacent markers had $r^2=0.9$. For validation we simulated 200000 SNPs in 2000 individuals. For each point in the plot we averaged the results over 100 independent phenotype simulations keeping the simulated genotypes fixed (see Materials and Methods for details). Note that when p=0.001, the chance of two causal variants being in LD is very small (~1%), and thus the improvement from accounting for LD in LDpred is negligible compared to P+T.

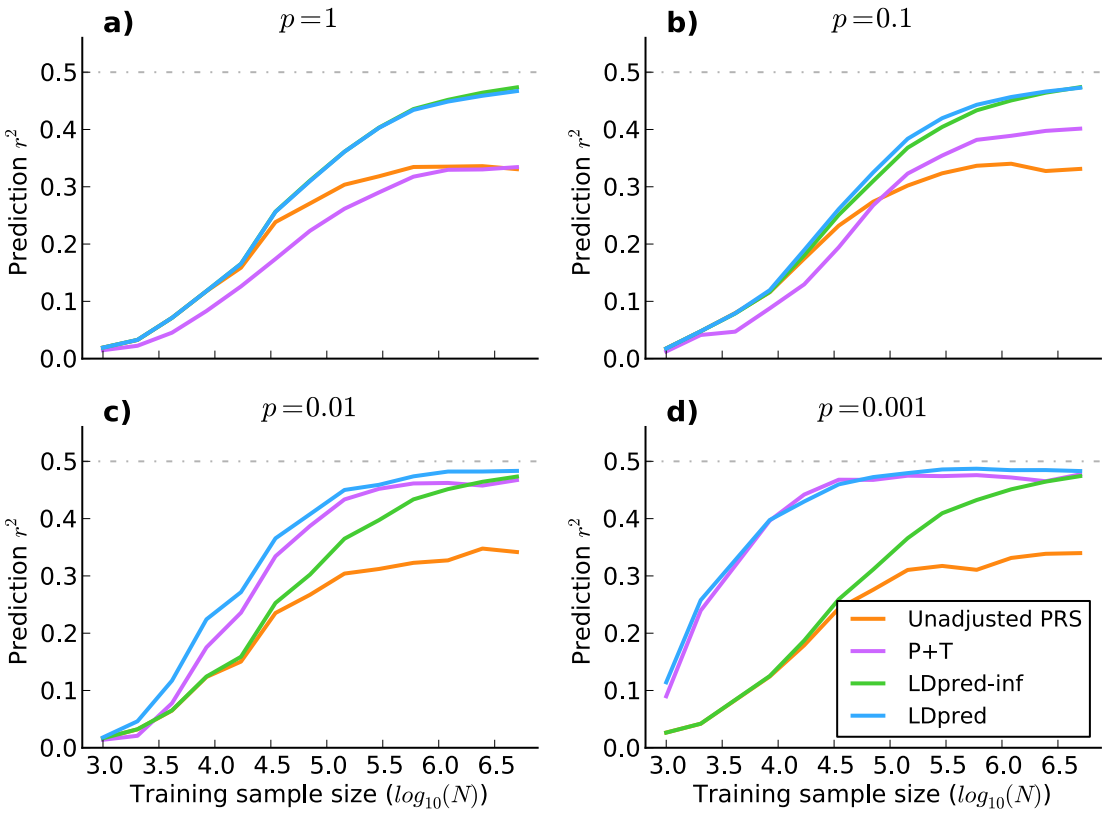

Figure S5. Comparison of methods when effects follow a Laplace distribution. Here the genotypes were simulated with LD using same simulation setup as in **Figure S4**, except the effect estimates were drawn from a Laplace mixture distribution instead of Gaussian mixture distribution. The change in prior appears to have minimal effect on the shape of the curve and the relative performance.

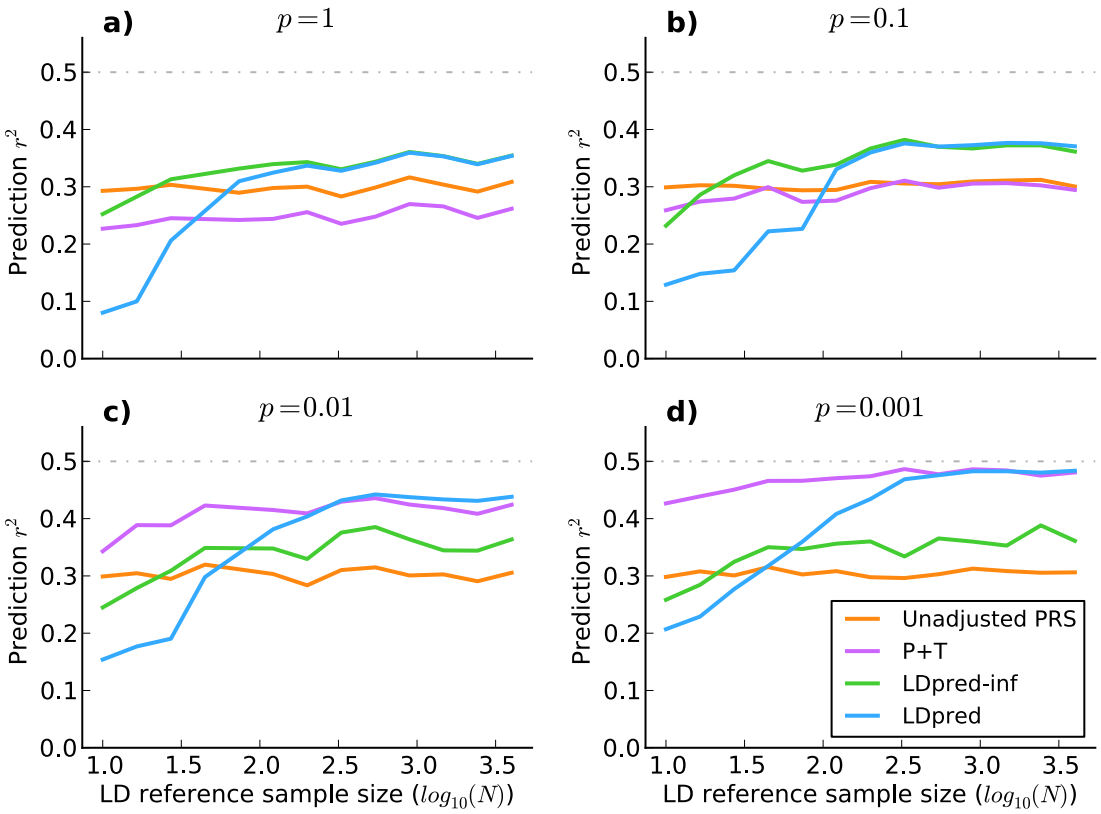

Figure S6. Prediction accuracy for methods as a function of LD reference sample size. Following the simulation setup from before (see **Figure S4.**) we simulated marginal least square effect estimates with LD (see Materials and Methods for details). We simulated segments of 100 SNPs where adjacent markers had $r^2=0.9$. We simulated in total 200000 SNPs and 2000 validation individuals. For each point in the plot we averaged the results over 100 independent phenotype simulations keeping the simulated validation genotypes fixed (see Materials and Methods for details). In addition, we simulated an LD reference panel with varying sample size along the x-axis. From these plots we see that a LD reference panel with more than 1000 individuals is necessary to ensure accurate LDpred scores. The accuracy of LDpred appears to be more sensitive to poor LD estimates than both P+T and LDpred-inf. Note that the Unadjusted PRS does not depend on LD information and is therefore expected to be a straight line, and thus providing a baseline.

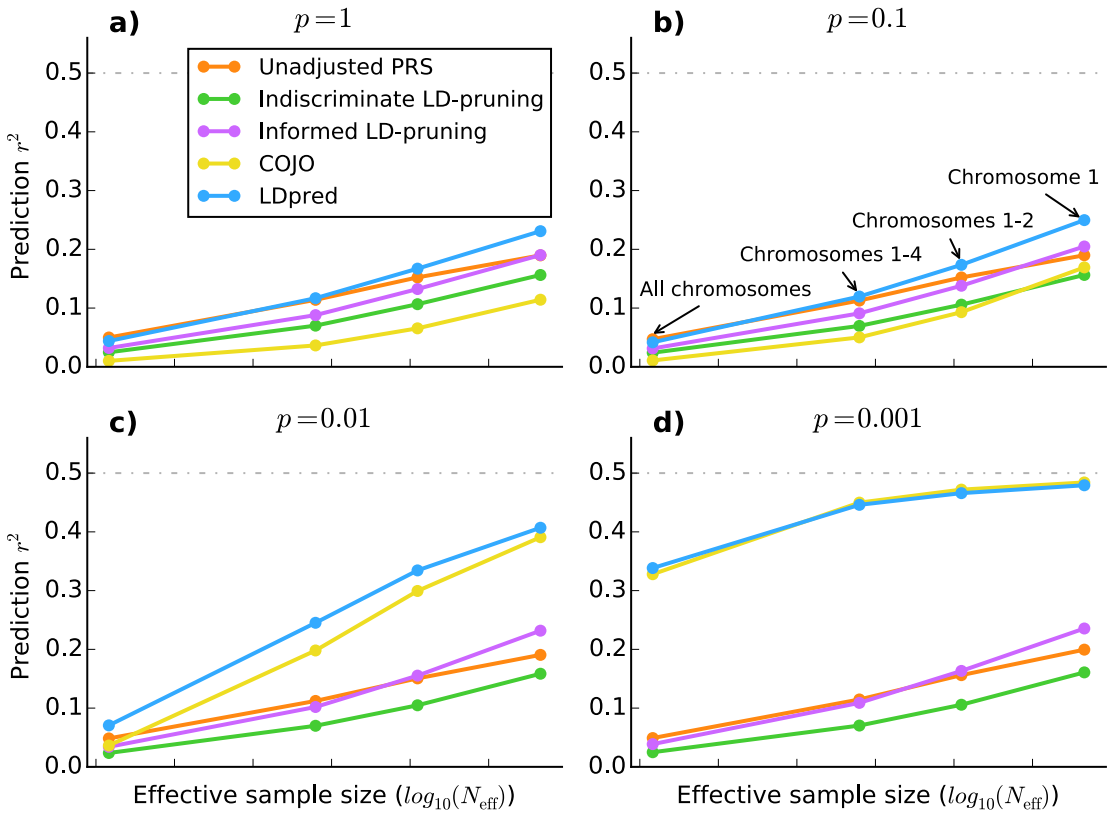

Figure S7. Comparisons to other methods using simulated traits and real WTCCC genotypes. As expected COJO[1,2] performs close to optimal with sufficient training data, or more precisely, when the ratio *(Nh²)/(Mp)* is approximately larger than 10. The comparison between the two types of LD-pruning clearly demonstrates the advantage of informed LD-pruning over indiscriminate LD-pruning, which randomly prunes either marker of a pair of markers in LD. For both LD-pruning strategies a pair of markers was considered in LD if $r^2>0.2$. When LDpred is compared to conditional joint analysis (COJO), LDpred outperforms COJO as long as the data does not overwhelm the prior, i.e. when *(Nh²)/(Mp)* is not sufficiently large (<10). For most of the diseases considered in this paper, current sample sizes are still not large enough for joint estimates to yield accurate risk scores.

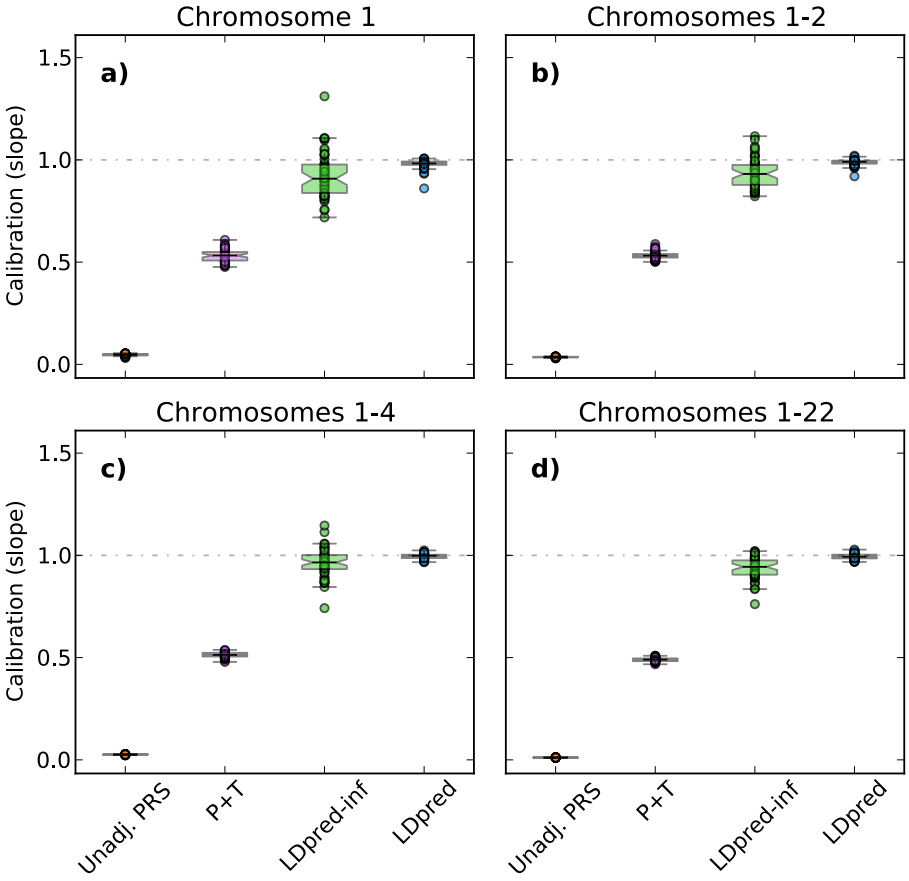

Figure S8. Boxplots of calibration slopes for simulations in Figure 2. Boxplots of calibration slopes for the four prediction methods evaluated in Figure 2 for *p=0.001* (the fraction of variants with non-zero effects). The subfigures **a-d** correspond to different number of SNPs used, ranging from 30,004 SNPs on chromosome 1 in **a)** to 376,901 SNPs or the full genome in **d)**. If the prediction conditional on the true value is unbiased then we expect a slope of one. A slope of less than one implies that the predicted value is mis-calibrated by a factor of 1/slope. Results for other values of *p (p=1; p=0.1; p=0.01)* gave similar results, and even stronger bias for P+T (LD-pruning followed by *P*-value thresholding).

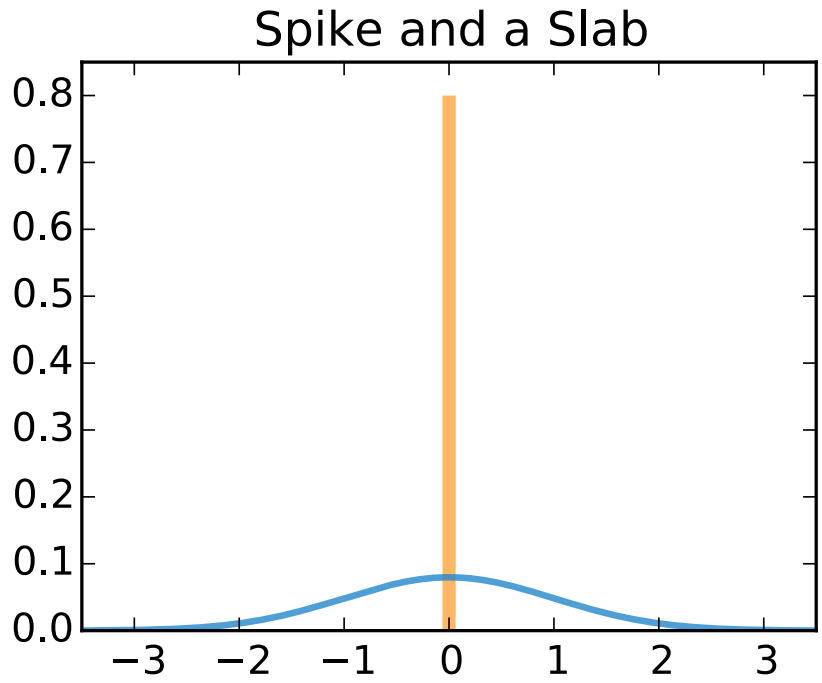

Figure S9. A spike and slab prior. An illustration of a spike and slab prior with a Gaussian slab.

Prediction Method	Accounts for LD?	Accounts for non-infinitesimal genetic architecture?	Comments
Unadjusted polygenic risk score	No.	No.	
LD-pruning followed by *P*-value thresholding (P+T)	Yes*.	Yes.	A heuristic that discards information from pruned and thresholded markers.
LDpred-inf	Yes.	No.	An analytical solution that assumes an infinitesimal prior for effects.
LDpred	Yes.	Yes.	A Gibbs sampler that assumes a point-normal mixture prior for effects.

Table S1. Overview of methods. A list of the main polygenic risk score methods (using summary association statistics as input) considered in this study. (*Although P+T prunes SNPs in high LD, it ignores bias induced by linked causal markers.)

Disease	Prediction accuracy measurement	Unadjusted PRS using all SNPs	Pruning + Thresholding	LDpred-inf	LDpred
T1D	Observed scale R^2	0.1064	0.3195	0.1062	0.3832
	Nagelkerke R^2	0.1442	0.4228	0.1438	0.5084
	Liability scale R^2	0.0426	0.0934	0.0426	0.1037
	AUC	0.6915	0.8410	0.6912	0.8738
T2D	Observed scale R^2	0.0360	0.0465	0.0404	0.0467
	Nagelkerke R^2	0.0488	0.0631	0.0547	0.0633
	Liability scale R^2	0.0257	0.0327	0.0287	0.0329
	AUC	0.6094	0.6243	0.6180	0.6275
CAD	Observed scale R^2	0.0250	0.0349	0.0290	0.0333
	Nagelkerke R^2	0.0338	0.0473	0.0393	0.0451
	Liability scale R^2	0.0191	0.0263	0.0221	0.0253
	AUC	0.5880	0.6087	0.5963	0.6043
CD	Observed scale R^2	0.0428	0.0485	0.0461	0.0824
	Nagelkerke R^2	0.0585	0.0661	0.0630	0.1122
	Liability scale R^2	0.0148	0.0167	0.0159	0.0267
	AUC	0.6212	0.6313	0.6279	0.6693
RA	Observed scale R^2	0.0483	0.1151	0.0462	0.1354
	Nagelkerke R^2	0.0656	0.1540	0.0627	0.1801
	Liability scale R^2	0.0239	0.0508	0.0229	0.0579
	AUC	0.6277	0.6994	0.6267	0.7162
BD	Observed scale R^2	0.0707	0.0876	0.0798	0.0816
	Nagelkerke R^2	0.09578	0.1185	0.1080	0.1105
	Liability scale R^2	0.0308	0.0371	0.0342	0.0349
	AUC	0.6552	0.6744	0.6662	0.6682
HT	Observed scale R^2	0.0306	0.0424	0.0348	0.0376
	Nagelkerke R^2	0.0414	0.0574	0.0471	0.0509
	Liability scale R^2	0.0258	0.0351	0.0292	0.0314
	AUC	0.6005	0.6180	0.6072	0.6109

Table S2. Numerical values of results displayed in Figure 3. The values are displayed on four different R^2 or AUC scales. To transform the prediction R^2 to liability scale we used the Lee *et al.* R^2 transformation[3] using values of disease prevalence specified in Supplementary Table 2.

Disease	Optimal fraction of causal markers used in LDpred	Optimal *P*-value threshold for Pruning + Thresholding	LDpred estimated heritability	LDpred estimated heritability on liability scale	Assumed disease prevalence
T1D	0.001	10^{-6}	1.3250	0.7258	0.005
T2D	0.03	1	0.6206	0.5125	0.03
CAD	0.03	1	0.6160	0.5181	0.035
CD	0.01	0.0001	0.7974	0.2904	0.001
RA	0.0001	10^{-6}	0.9097	0.5145	0.0075
BD	0.1	1	0.9695	0.4959	0.005
HT	0.03	1	0.6216	0.5939	0.05

Table S3. P+T and LDpred parameters for methods evaluated in Figure 3. The heritabilities are calculated as averages over 5 cross validations. The Lee *et al*. heritability transformation[4] was used to obtain the heritability on the liability scale. The LD window size used in the simulations was 400 SNPs.

Disease	Unadjusted PRS using all SNPs	Pruning + Thresholding	LDpred-inf	LDpred
T1D	0.0082	0.4301	3.2282	0.6365
T2D	0.0056	0.0278	1.2678	1.0198
CAD	0.0058	0.0231	2.1214	1.6566
CD	0.0059	0.0231	1.4159	0.8570
RA	0.0069	0.3163	2.3133	0.7755
BD	0.0076	0.0249	1.2348	1.1472
HT	0.0055	0.0301	1.7345	1.7039

Table S4. Calibration comparison for methods evaluated in Figure 3. We report the slope, where a value close to 1 represents a well-calibrated prediction. LDpred yields the most appropriately calibrated predictions.

Disease	Prediction accuracy measurement	Unadjusted PRS using all SNPs	Pruning + Thresholding	LDpred-inf	LDpred
SCZ-MGS	Observed scale R^2	0.1591	0.1510	0.1870	0.1898
	Nagelkerke R^2	0.2119	0.2014	0.2488	0.2528
	Liability scale R^2	0.0616	0.0594	0.0688	0.0694
	AUC	0.7294	0.7248	0.7499	0.7519
SCZ-ISC	Observed scale R^2	0.1169	0.0970	0.1334	0.1367
	Nagelkerke R^2	0.1574	0.1304	0.1803	0.1836
	Liability scale R^2	0.0518	0.0446	0.0578	0.0585
	AUC	0.6988	0.6784	0.7127	0.7165
MS	Observed scale R^2	0.0316	0.0674	0.0363	0.0840
	Nagelkerke R^2	0.0474	0.0978	0.0512	0.1198
	Liability scale R^2	0.0149	0.0302	0.0170	0.0368
	AUC	0.6169	0.6714	0.6187	0.6918
BC	Observed scale R^2	0.0071	0.0324	0.0092	0.0386
	Nagelkerke R^2	0.0097	0.0437	0.0119	0.0519
	Liability scale R^2	0.0040	0.0184	0.0052	0.0220
	AUC	0.5489	0.6052	0.5549	0.6156
T2D	Observed scale R^2	0.0159	0.0247	0.0214	0.0273
	Nagelkerke R^2	0.0212	0.0330	0.0309	0.0365
	Liability scale R^2	0.0112	0.0170	0.0149	0.0187
	AUC	0.5747	0.5854	0.5825	0.5953
CAD	Observed scale R^2	0.0109	0.0101	0.0124	0.0125
	Nagelkerke R^2	0.0146	0.0137	0.0168	0.0170
	Liability scale R^2	0.0085	0.0080	0.0097	0.0098
	AUC	0.5612	0.5557	0.5645	0.5647

Table S5. Numerical values of results displayed in Figure 4. The numerical results are shown on four different R^2 or AUC scales.

Trait	Prediction accuracy measurement	Unadjusted PRS using all SNPs	Pruning + Thresholding	LDpred-inf	LDpred
Height	R^2	0.0927	0.0841	0.0906	0.1014
	PC-adjusted R^2	0.0697	0.0634	0.0656	0.0853
	Risk Score + PC R^2	0.1205	0.1146	0.1166	0.1353

Table S6. Prediction accuracy for height. Height and the polygenic risk score for height is stratified by population structure. The prediction accuracy is therefore substantially reduced when we account for the first 5 principal components. Interestingly, LDpred improves the PC-adjusted prediction accuracy by 30% compared to P+T.

Diesease	Prediction accuracy measurement	Unadjusted PRS using all SNPs	Pruning + Thresholding	LDpred-inf	LDpred
T2D	Observed scale R^2	0.0022	0.0109	0.0023	0.0095
	Nagelkerke R^2	0.0029	0.0175	0.0030	0.0126
	Liability scale R^2	0.0011	0.0055	0.0012	0.0048
	AUC	0.5247	0.5633	0.5249	0.5573
CAD	Observed scale R^2	0.0029	0.0058	0.0030	0.0048
	Nagelkerke R^2	0.0039	0.0078	0.0040	0.0065
	Liability scale R^2	0.0023	0.0046	0.0024	0.0038
	AUC	0.5284	0.5418	0.5289	0.5374

Table S7. Additional validation for T2D and CAD when training on WTCCC data. Prediction accuracy for type-2 diabetes and coronary artery disease when training on WTCCC cases and controls and predicting into the WGHS data.

Disease	Optimal *P*-value threshold for Pruning + Thresholding	Optimal Gaussian mixture weight (fraction of causal markers) for LDpred	LDpred/ LD-pruning window size (# of SNPs)	GWAS sample size used in LDpred	LDpred estimated heritability	LDpred estimated heritability on liability scale	Assumed prevalence
SCZ-MGS	0.1	0.3	500	65K	0.5738	0.4231	0.01
SCZ-ISC	0.1	0.3	500	65K	0.4718	0.3479	0.01
MS	0.001	0.01	400	27K	0.3694	0.1321	0.001
BC	0.00003	0.003	400	50K	0.1934	0.1124	0.01
T2D	0.00003	0.1	300	69K	0.2061	0.1582	0.0075
CAD	0.1	1	300	86K	0.2943	0.2494	0.035

Table S8. Model parameters for results in Figure 4. Parameters inferred or assumed by P+T and LDpred for results displayed in **Figure 4**. The Lee *et al*. heritability transformation[52] was used to obtain the heritability on the liability scale.

Disease	Unadjusted PRS using all SNPs	Pruning + Thresholding	LDpred-inf	LDpred
SCZ-MGS	0.0063	0.0467	0.3845	0.3918
SCZ-ISC	0.0130	0.0407	0.4683	0.4413
MS	0.0089	0.0717	0.9092	0.2011
BC	0.0017	0.1327	1.2323	0.5650
T2D	0.0032	0.1002	0.6421	0.4057
CAD	0.0035	0.0137	0.2244	0.1868

Table S9. Calibration slopes for methods evaluated in Figure 4. We report the slope, where a value close to 1 represents a well-calibrated prediction.

Schizophrenia Cohort	Prediction accuracy measurement	Unadjusted PRS using all SNPs	Pruning + Thresholding	LDpred-inf	LDpred
MGS (European ancestry)	Observed scale R^2	0.1591	0.1510	0.1870	0.1898
	Nagelkerke R^2	0.2119	0.2014	0.2488	0.2528
	Liability scale R^2	0.0616	0.0594	0.0688	0.0694
	AUC	0.7294	0.7248	0.7499	0.7519
JPN1 (Japanese ancestry)	Observed scale R^2	0.0477	0.0702	0.0691	0.0695
	Nagelkerke R^2	0.0635	0.0944	0.0923	0.0929
	Liability scale R^2	0.0232	0.0323	0.0319	0.0320
	AUC	0.6276	0.6527	0.6523	0.6531
TCR1 (Chinese ancestry)	Observed scale R^2	0.0570	0.0616	0.0704	0.0717
	Nagelkerke R^2	0.0761	0.0821	0.0939	0.0956
	Liability scale R^2	0.0274	0.0294	0.0329	0.0336
	AUC	0.6331	0.6391	0.6483	0.6488
HOK2 (Chinese ancestry)	Observed scale R^2	0.0253	0.0306	0.0374	0.0373
	Nagelkerke R^2	0.0414	0.0511	0.0609	0.0609
	Liability scale R^2	0.0187	0.0225	0.0271	0.0271
	AUC	0.6176	0.6250	0. 6352	0.6352
AFAM (African American ancestry)	Observed scale R^2	0.0170	0.0151	0.0279	0.0280
	Nagelkerke R^2	0.0233	0.0202	0.0382	0.0383
	Liability scale R^2	0.0095	0.0084	0.0152	0.0152
	AUC	0.5745	0.5682	0.5936	0.5936

Table S10. Prediction accuracy for schizophrenia risk scores when validating in non-European populations. The accuracy is reported on four different R^2 or AUC scales.

SCZ cohort	Genetic ancestry	Optimal *P*-value threshold for Pruning + Thresholding	Optimal Gaussian mixture weight (fraction of causal markers) for LDpred	LDpred/ LD-pruning window size (# of SNPs)	GWAS sample size used in LDpred
JPN1	Japanese (Tokai)	0.1	0.3	1000	65000
TCR1	Chinese (Singapore)	0.1	0.3	1000	65000
HOK2	Chinese (Hong Kong)	1	1	1000	65000
AFAM	African American	0.3	1	400	69000

Table S11. Parameters inferred or assumed by P+T and LDpred for analysis of the non-European validation samples in **Table S10**.

References

1. Rietveld, C.A. *et al.* GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment. *Science* **340**, 1467-1471 (2013).
2. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics* **44**, 369-375 (2012).
3. Lee, S.H., Goddard, M.E., Wray, N.R. & Visscher, P.M. A Better Coefficient of Determination for Genetic Profile Analysis. *Genetic Epidemiology* **36**, 214-224 (2012).
4. Lee, S., Wray, N., Goddard, M. & Visscher, P. Estimating missing heritability for disease from genome-wide association studies. *American Journal Of Human Genetics* **88**, 294-305 (2011).