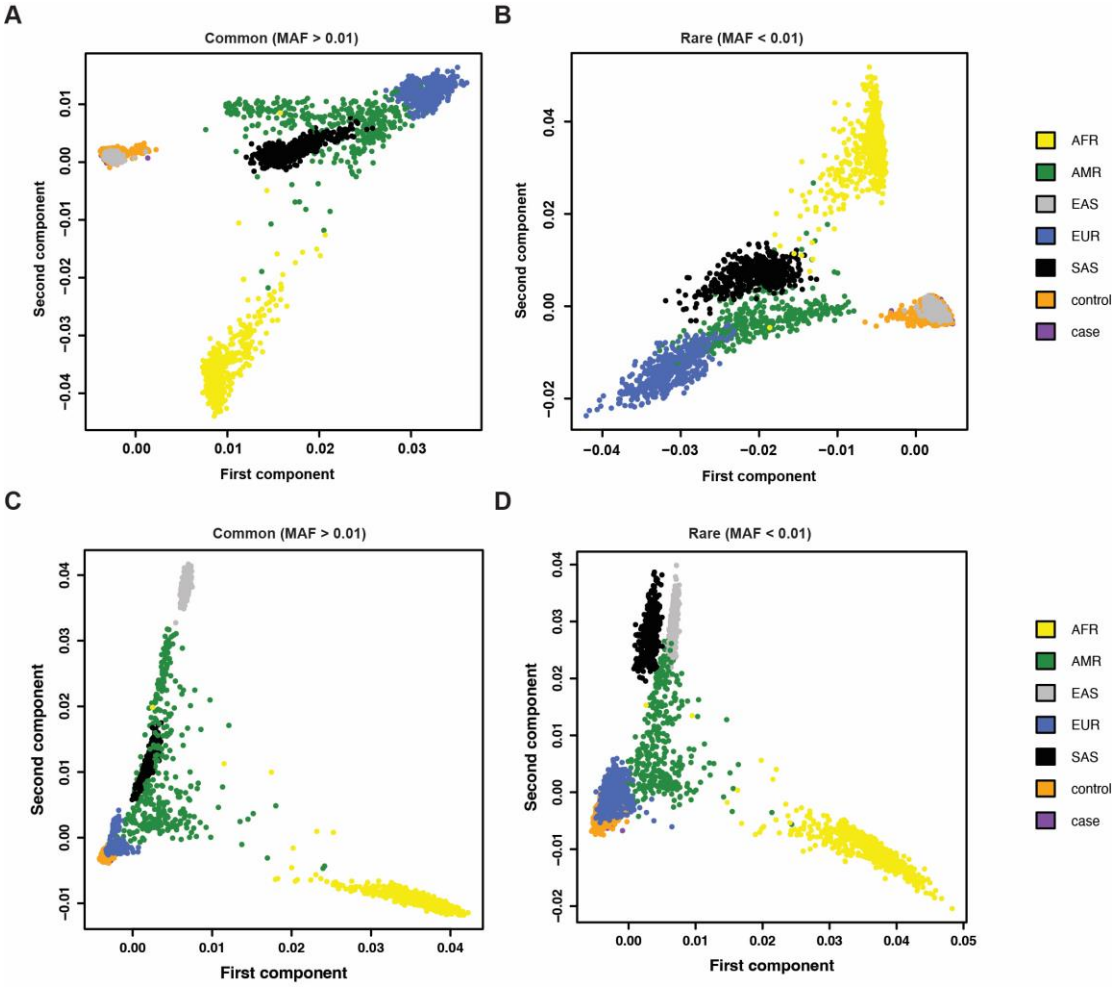
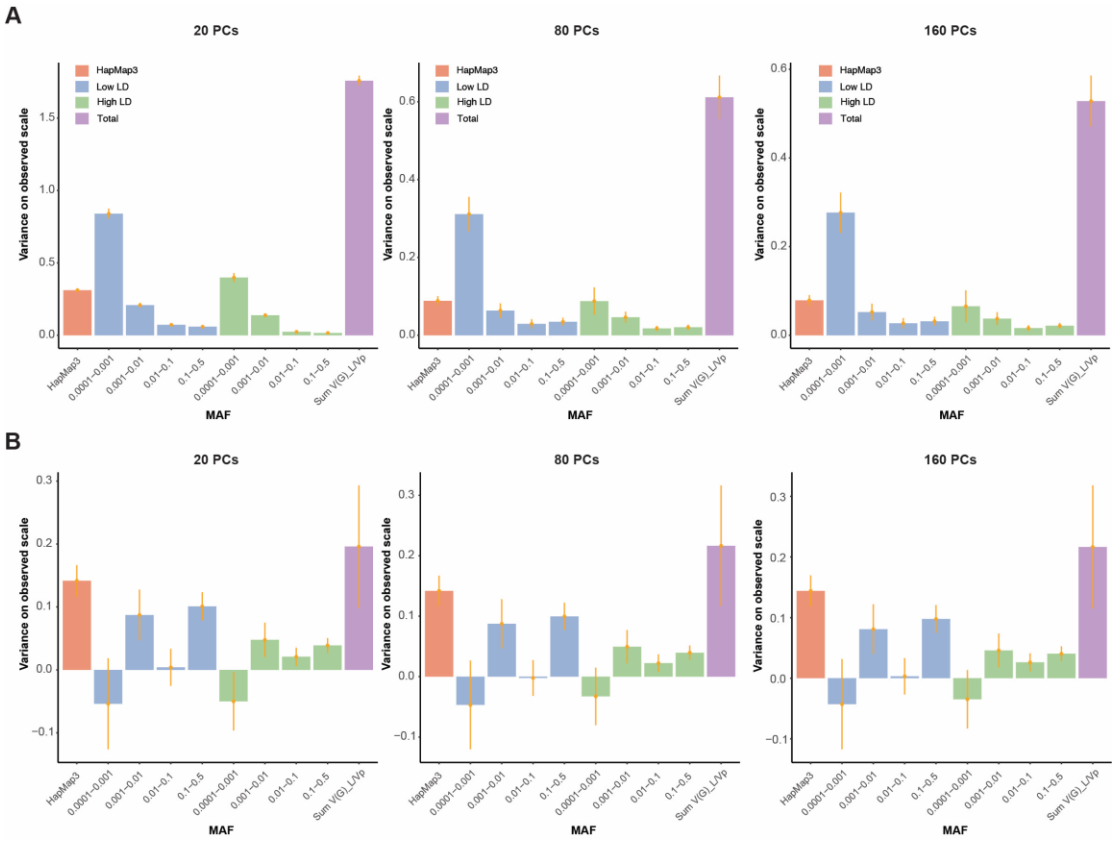
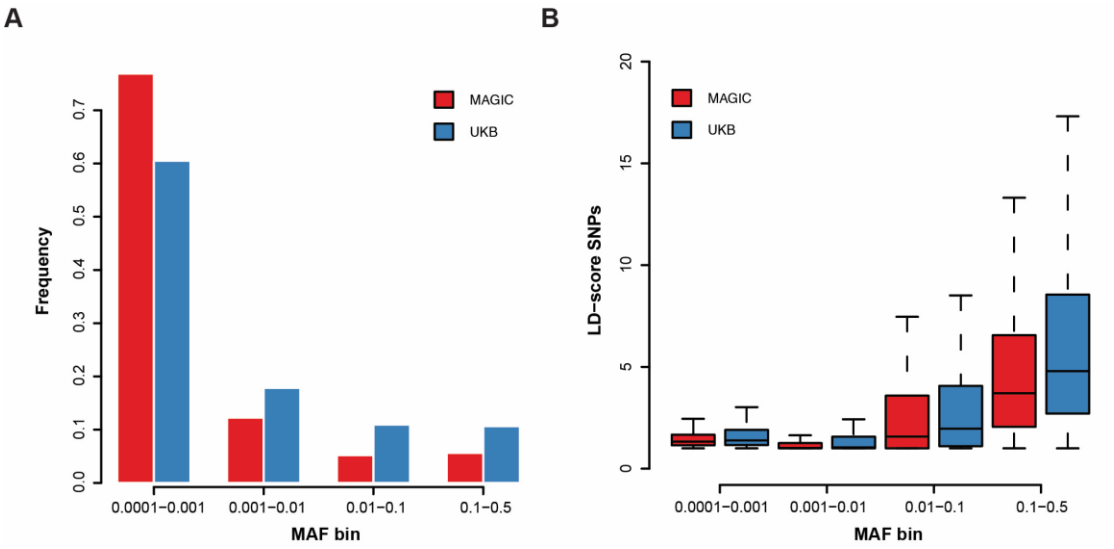


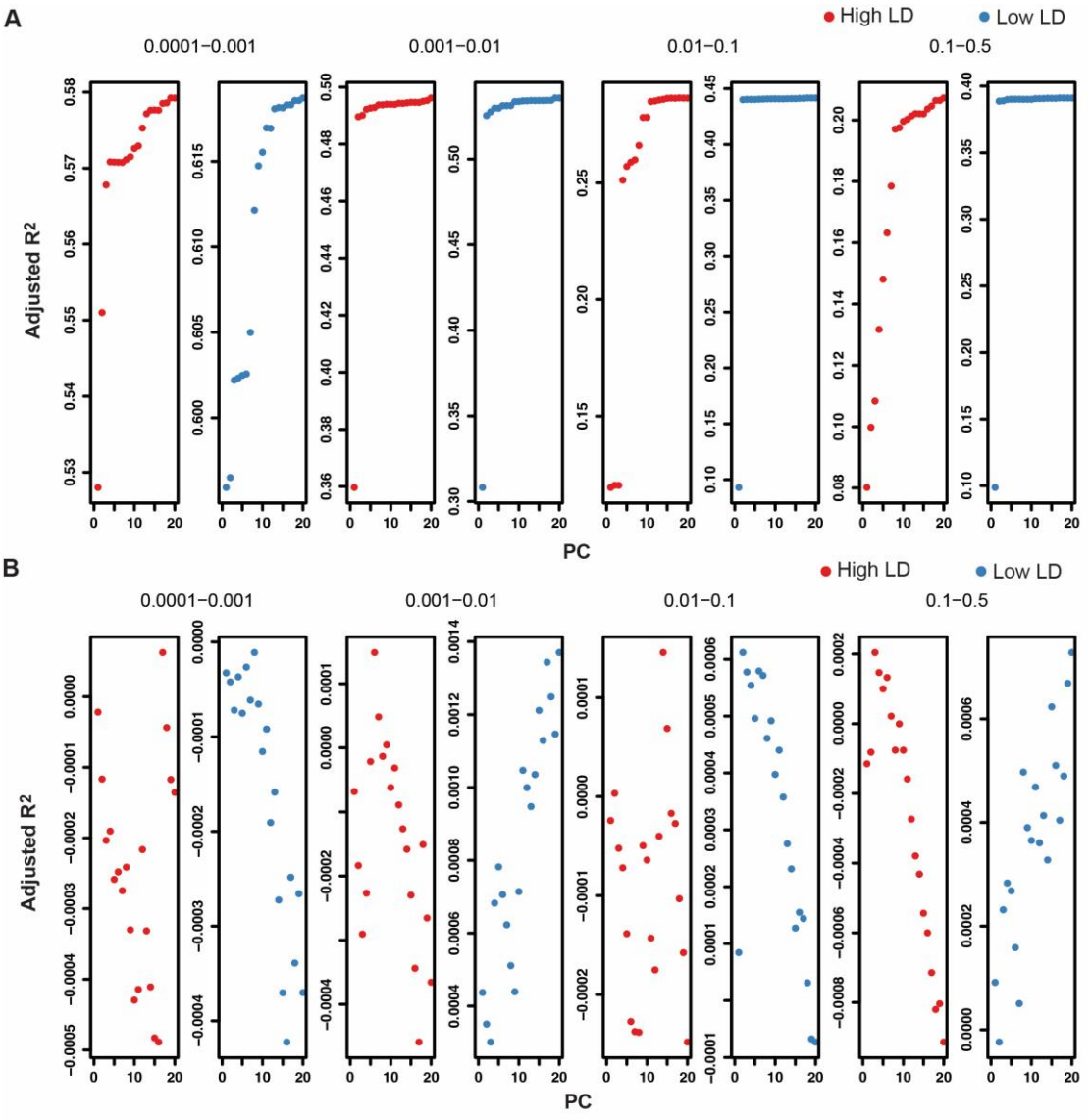
Supplementary Figure 1. Principal Component Analysis plot of individuals compared to 1000 Genomes populations. (A) PCA of MAGIC samples and 1000G populations using common SNPs. (B) PCA of MAGIC samples and 1000G populations using rare SNPs. (C) PCA of UKB samples and 1000G populations using common SNPs. (D) PCA of UKB samples and 1000G populations using rare SNPs. AFR, African; AMR, Ad Mixed American; EAS, East Asian; EUR, European; SAS, South Asian.



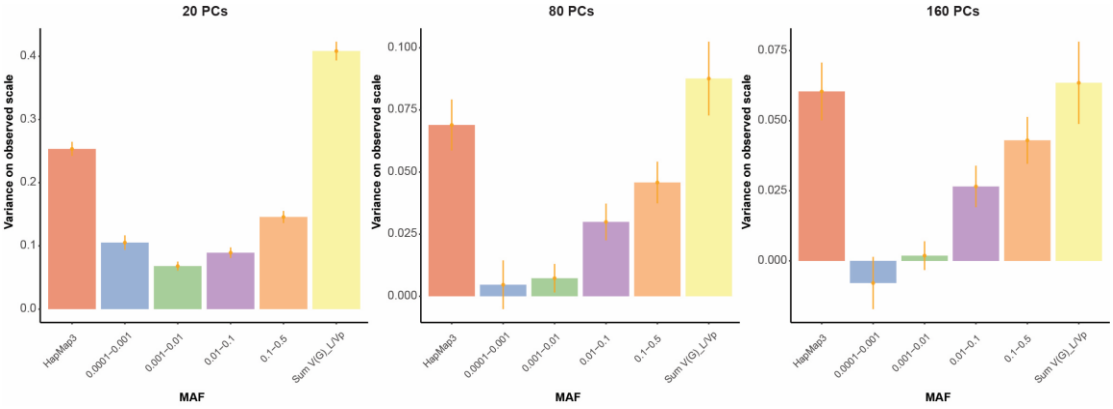
Supplementary Figure 2. GREML-LDMS estimates from WES data stratified in 8 bins (4 MAF bins in 2 LD bins) with correction for 20 PCs (on HM3 SNPs), 80 PCs (10 * 8 bins) or 160 PCs (20 * 8 bins). (A) Estimates heritability for HM in MAGIC. (B) Estimates heritability for HM in UKB. Error bars indicate standard errors (SE).



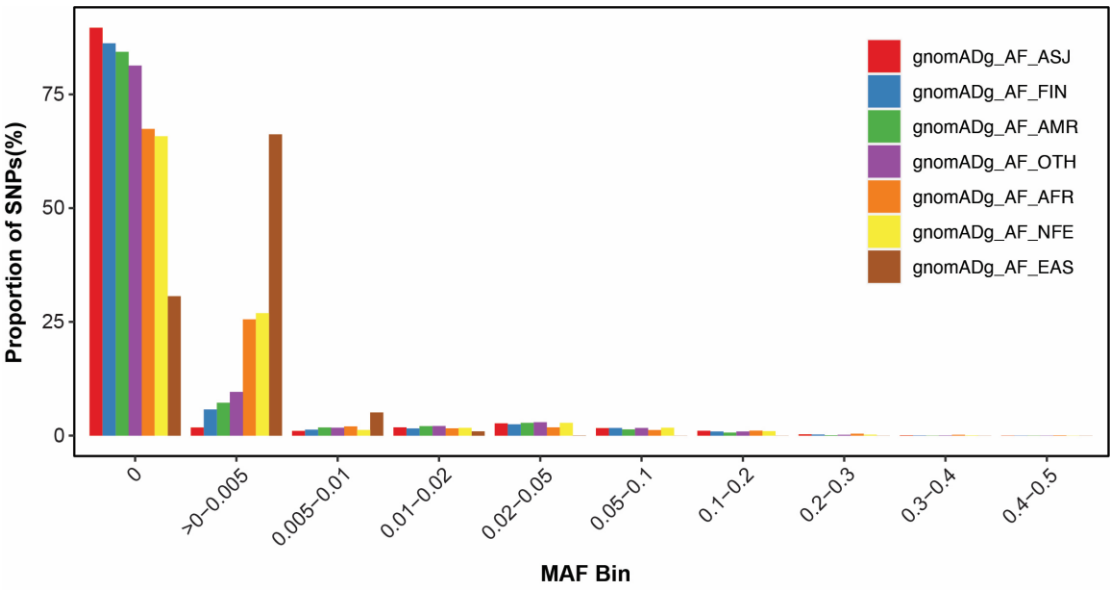
Supplementary Figure 3. (A) Fraction of the variants as a percentage of the total number of variants in the four MAF bin. (B) Boxplot of the distribution of individual SNPs LD values within each bin. Lower and upper hinges correspond to the first and third quartiles and whiskers extend to +/- 1.5 * inter-quartile range.



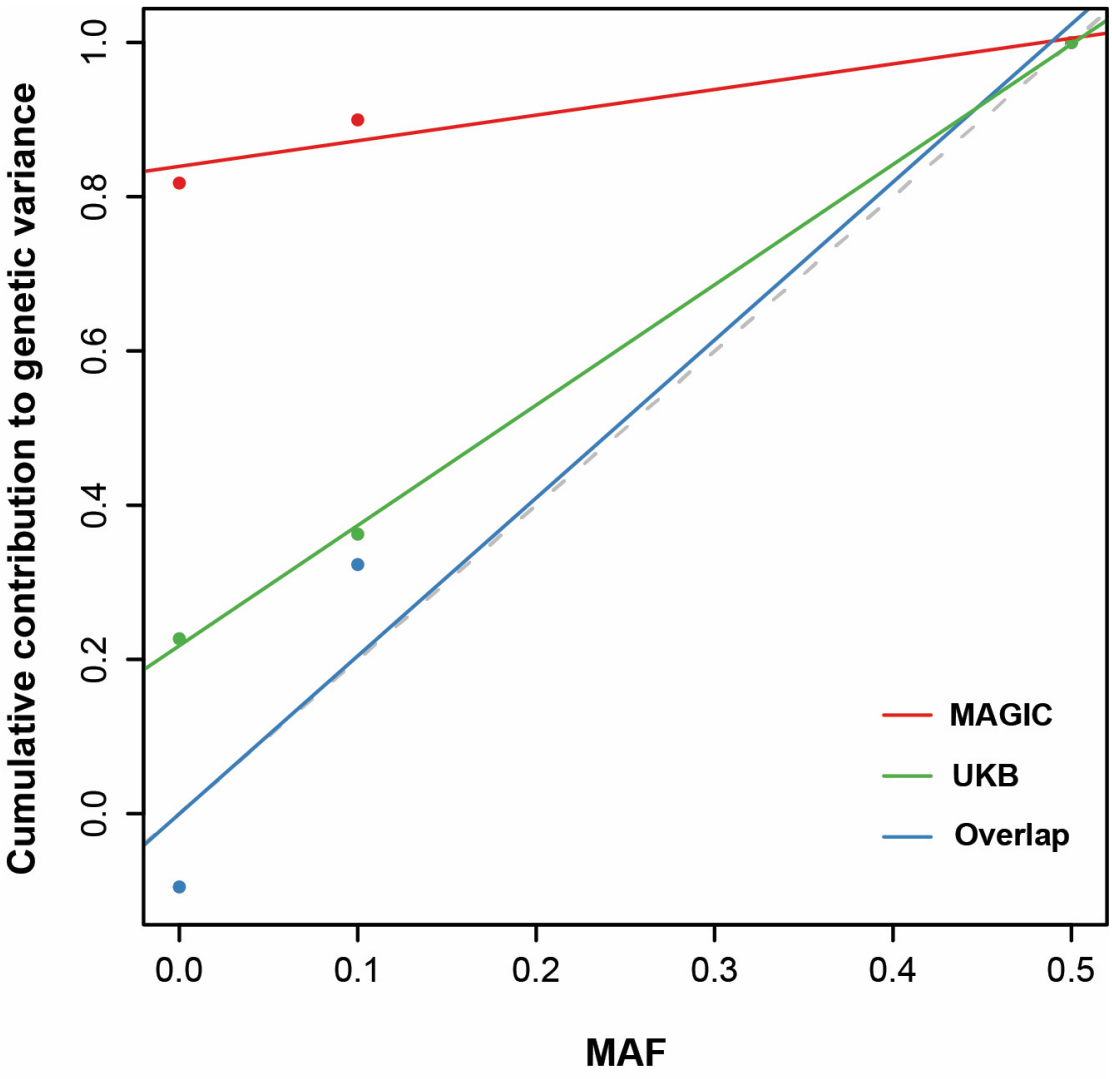
Supplementary Figure 4. Adjusted R^2 from fitting 20 PCs computemagId from LD-pruned SNPs for different MAF and LD bins. (A) Adjusted R^2 for the 12,000 MAGIC Exome samples. (B) Adjusted R^2 for the N= 8,682 UKB Exome samples.



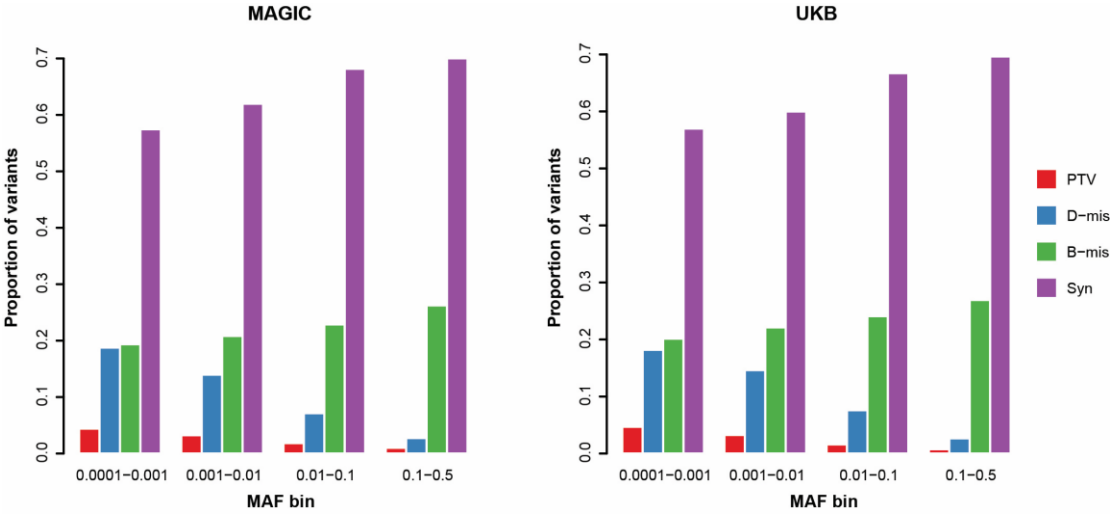
Supplementary Figure 5. GREML-LDMS estimates for overlapped variant between MAGIC and UKB with correction for 20 PCs (on HM3 SNPs), 80 PCs (10 * 8 bins) or 160 PCs (20 * 8 bins).



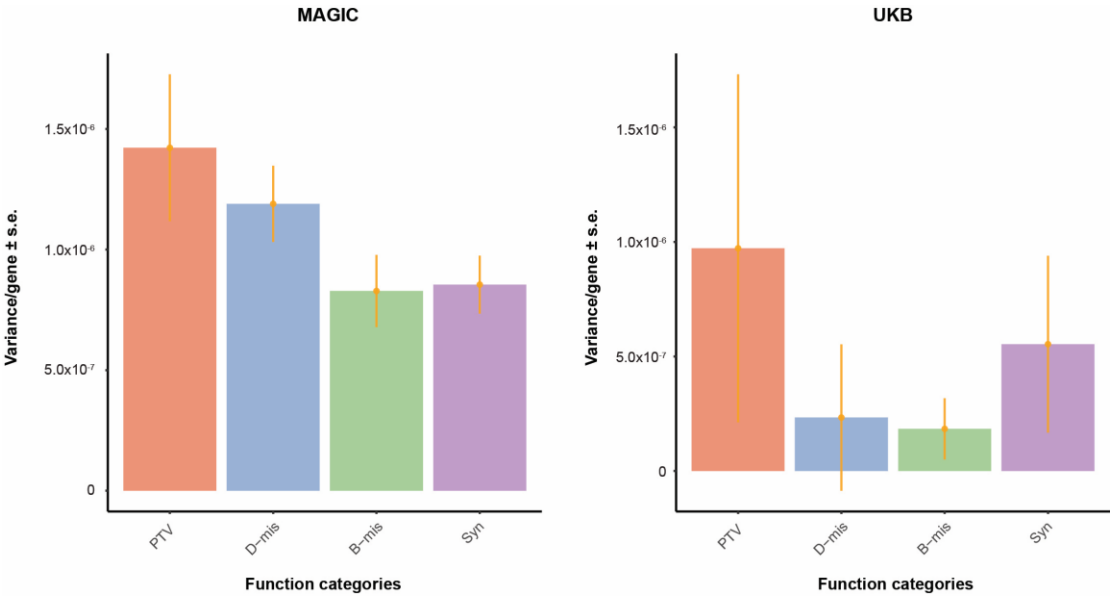
Supplementary Figure 6. MAF distribution for MAGIC specific rare variants across ancestries in gnomAD. Proportion of polymorphic variants (as a % of total) are described on the y-axis with the x-axis illustrating bins of the minor allele frequencies (MAF) for seven different ancestries.



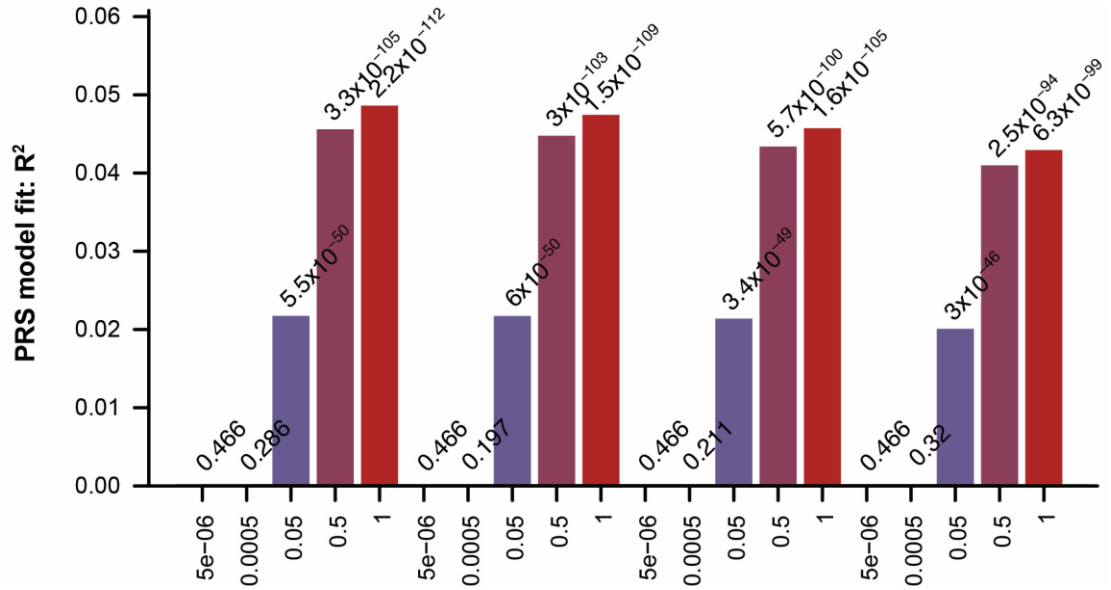
Supplementary Figure 7. Estimate of the cumulative contribution of variants, for HM, from GREML-LDMS analysis. The dash line represents the expected contribution under a neutral evolutionary model. The deviation from this dash line suggests that HM in UKB and MAGIC are under negative selection. The first bin represents variants with MAF < 0.01 (cumulative contribution to genetic variance for HM) when second bin shows all variants with MAF < 0.1. The third MAF bin includes all variants in the dataset.



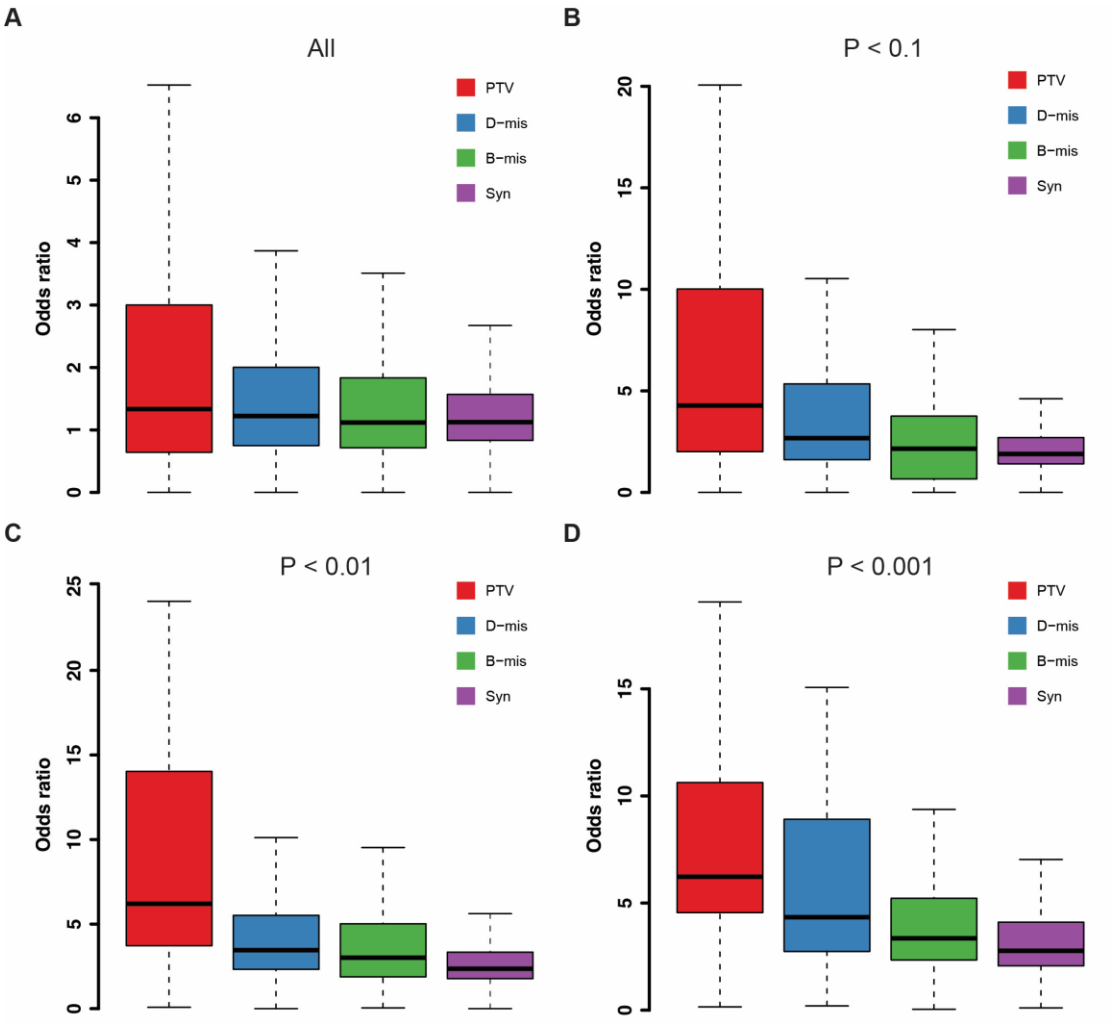
Supplementary Figure 8. Proportion of the variants categorized by annotations as a percentage of the total number of variants in the four MAF bin.



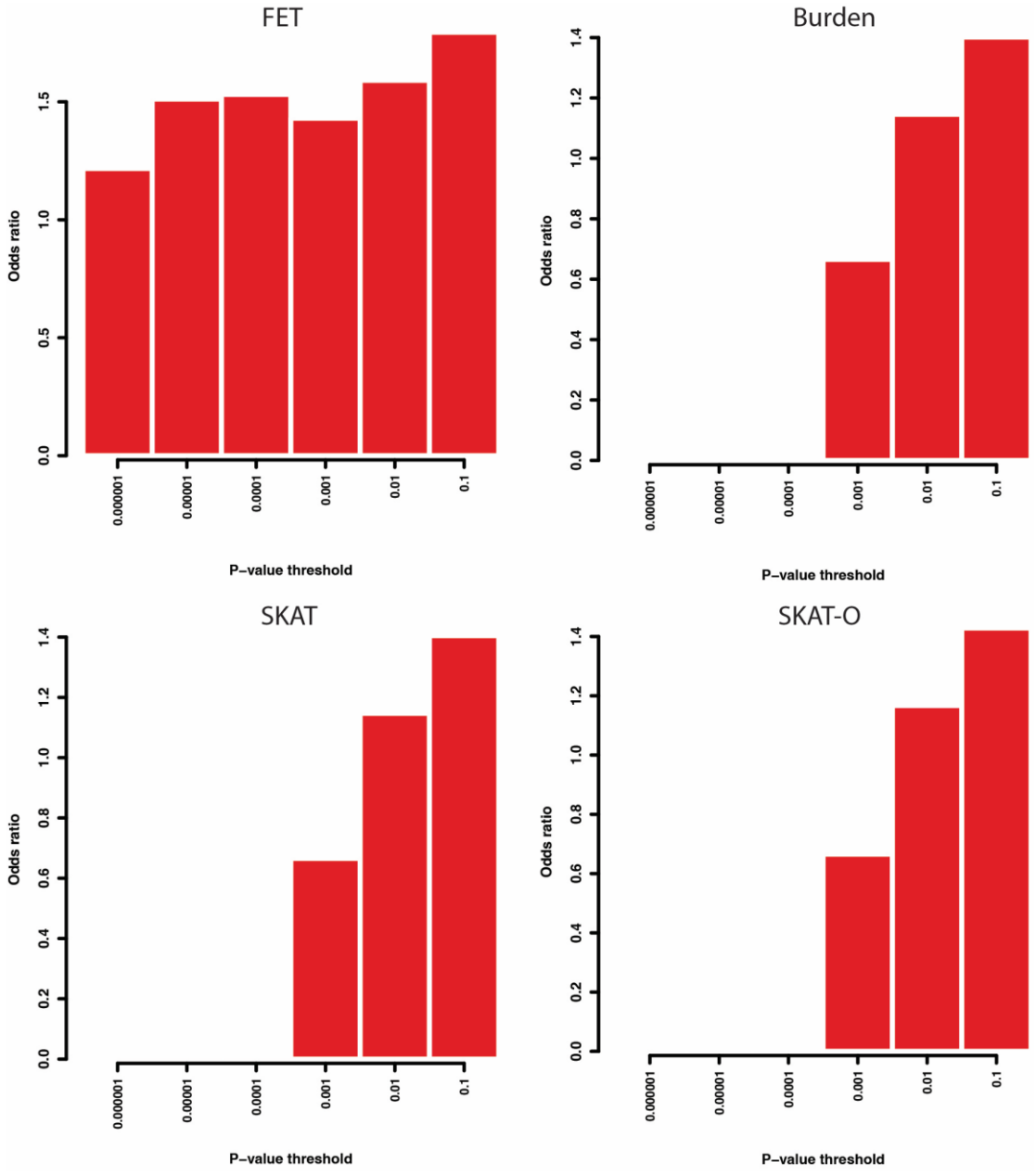
Supplementary Figure 9. Variance explained per genes (the estimate of genetic variance divided by the number of genes in each bin) from burden heritability regression (BHR) with the rare variant partitioned into four distinct categories according to the VEP putative effect of the variant. Error bars indicate s.e.



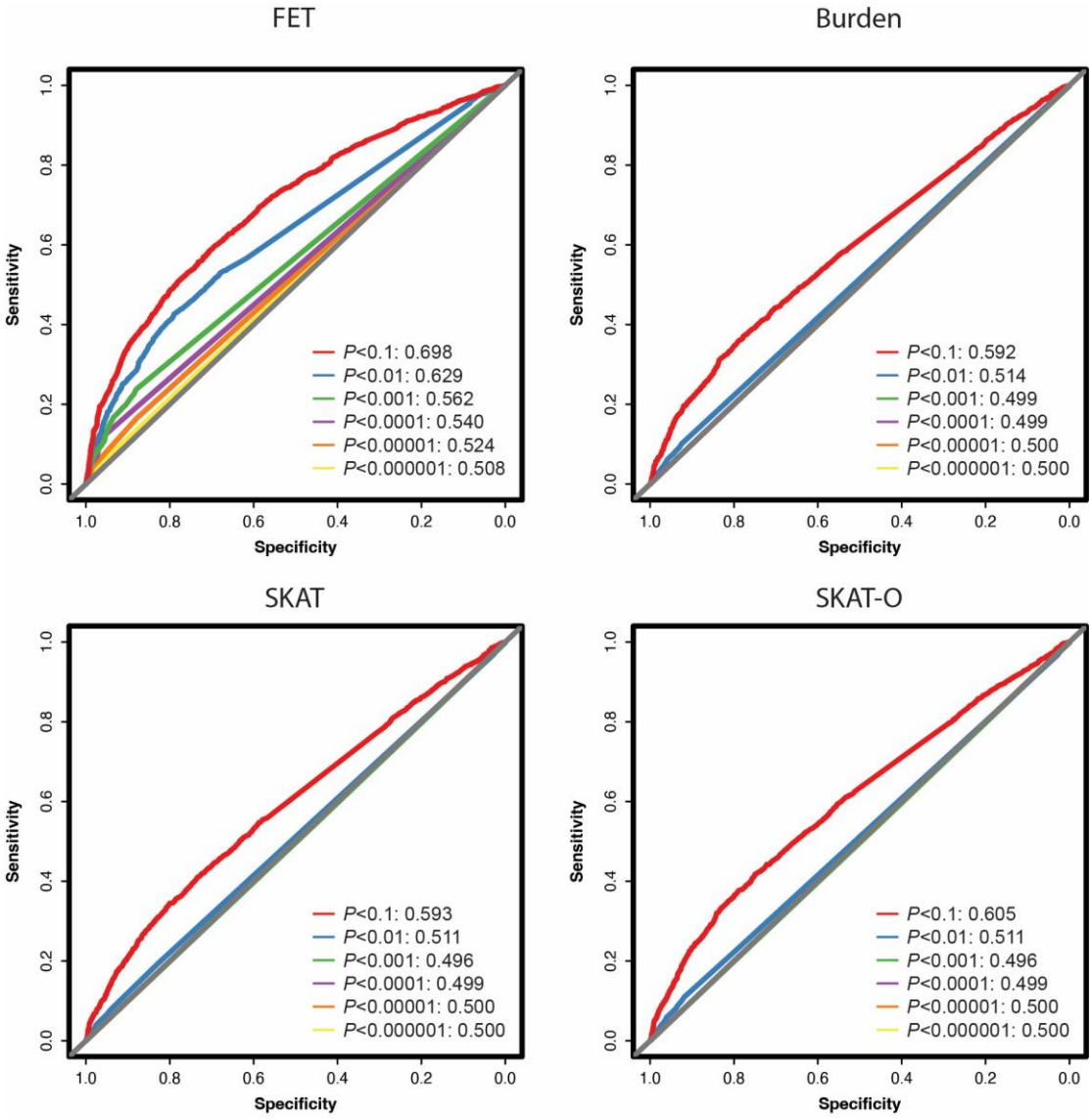
Supplementary Figure 10. Evaluation of predictive power of GRS derived from all common variants on HM. GRSs were derived from MAGIC cohort by plink.



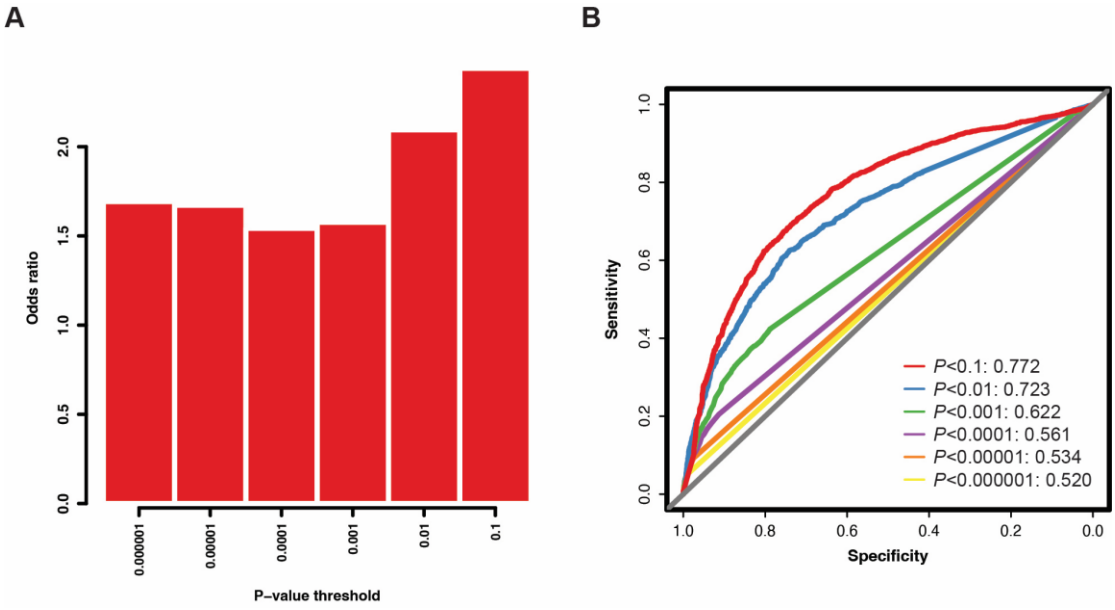
Supplementary Figure 11. Odds ratio of HM-associated gene in rare variant association study under different *P* value threshold.



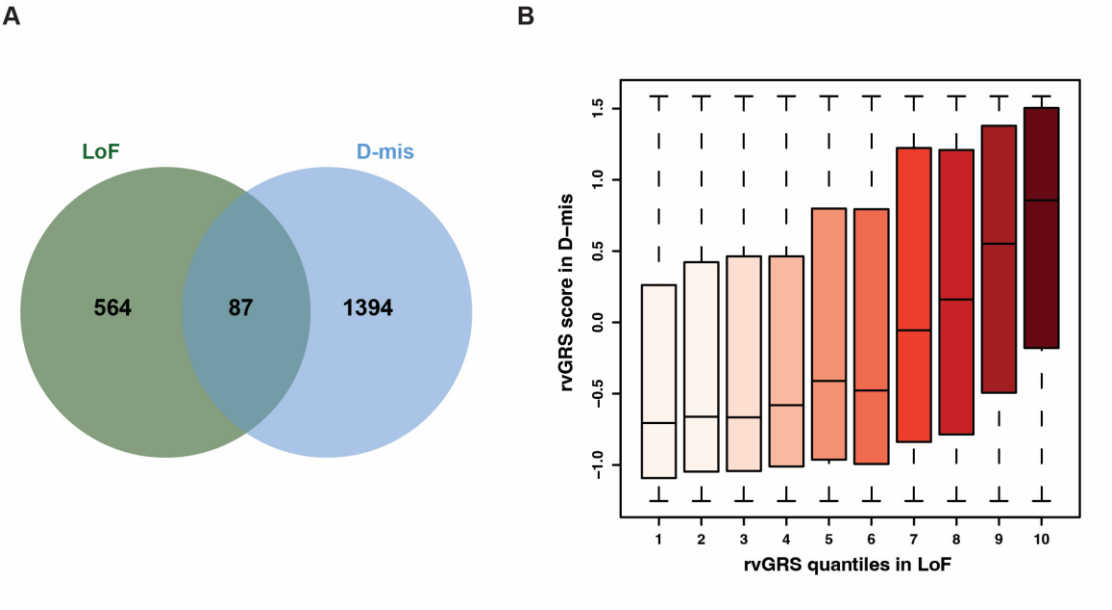
Supplementary Figure 12. Odds ratio for rvGRS model derived from HM-associated gene in rare variant association study by using different Methods.



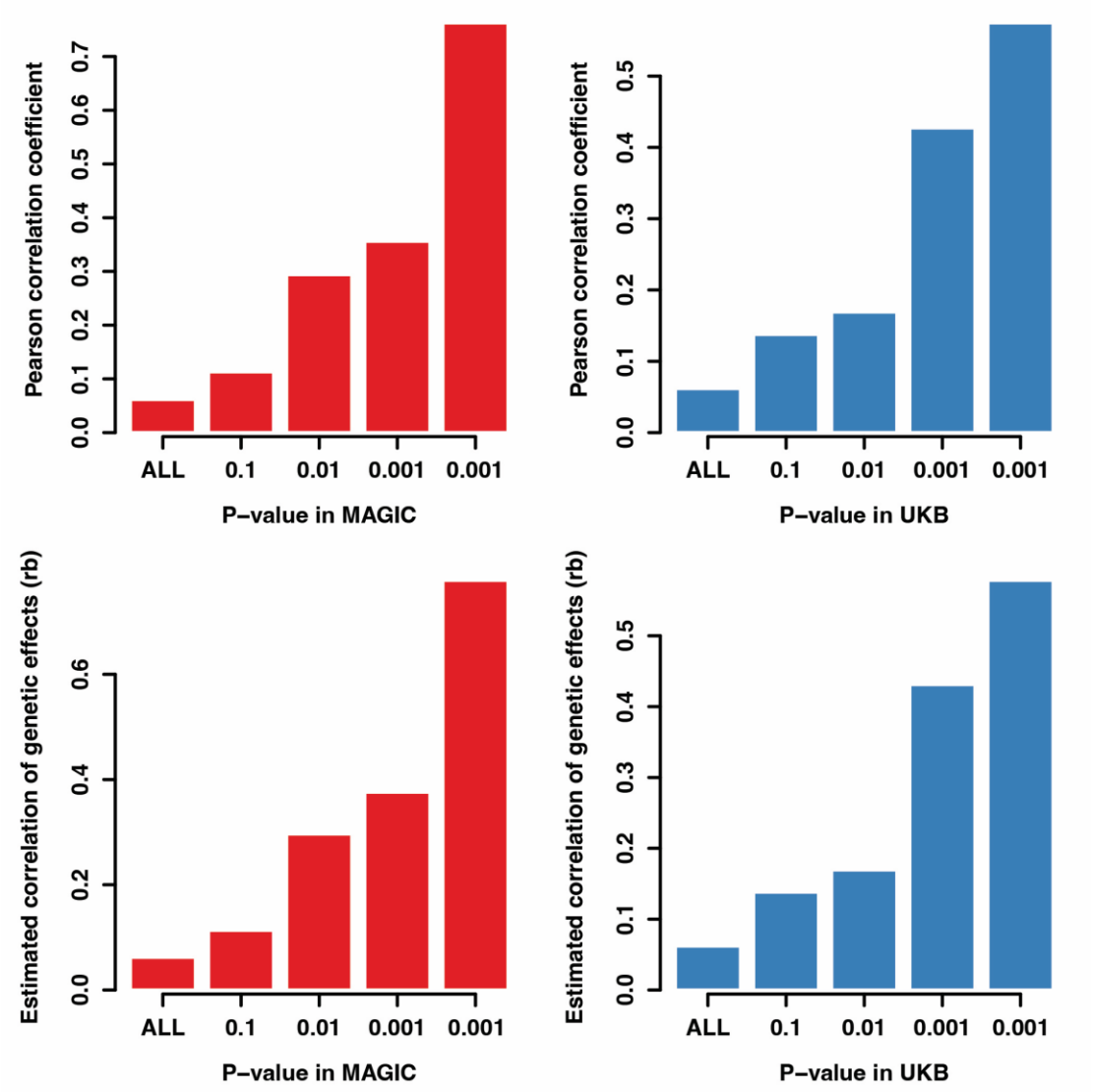
Supplementary Figure 13. ROC curve and AUC for rvGRS model derived from HM-associated genes with rare PTV.



Supplementary Figure 14. Evaluated performance of rvGRS model derived from HM-associated genes with rare D-mis selected by Fisher's exact test. (A) Odds ratio for rvGRS model; (B) ROC curve and AUC for rvGRS model.



Supplementary Figure 15. Comparison of performance of rvGRS model with PTVs or D-mis. (A) Comparison of the PTV and D-mis on the number of rare variant association study under matched significance thresholds. (B) rvGRS in D-mis for each individual according to 10 groups of the validation dataset binned according to the quantiles of the rvGRS in PTVs.



Supplementary Figure 16. Comparisons of the effect sizes (ORs) at different ExWAS P-value thresholds in the MAGIC (left) and UKB (right). Bar plot displaying the Pearson correlation coefficient (up panel) and rb (bottom pane) of log(ORs) at *P*-value threshold. The x-axis indicates the P-value thresholds used to filter the variants to be included in the correlation computation.

Supplementary Table 1. Variants identified from exome sequencing in MAGIC

Annotation	**0.1-0.5**	**0.01-0.1**	**0.001-0.01**	**0.0001-0.001**
3_prime_UTR_variant	2720	2567	5754	34517
5_prime_UTR_variant	1315	1177	2765	16393
downstream_gene_variant	1025	918	2007	12070
frameshift_variant	55	113	440	3655
inframe_deletion	35	75	243	2038
inframe_insertion	23	30	94	616
intergenic_variant	68	48	89	495
intron_variant	14199	12028	24811	143689
mature_miRNA_variant	2	6	14	78
missense_variant	7569	8572	25940	182251
non_coding_transcript_exon_variant	759	560	1282	6696
regulatory_region_variant	30	33	62	280
splice_acceptor_variant	12	5	46	329
splice_donor_variant	12	18	54	492
splice_region_variant	836	741	1669	10276
start_lost	11	8	48	302
stop_gained	45	82	345	3403
stop_lost	6	5	14	106
stop_retained_variant	4	7	10	78
synonymous_variant	9408	8460	18075	107530
TF_binding_site_variant	4	3	5	46
upstream_gene_variant	1294	1083	2230	13191
Total	39434	36541	86000	538560

Supplementary Table 2. Variants identified from exome sequencing in UKB

Annotation	**0.1-0.5**	**0.01-0.1**	**0.001-0.01**	**0.0001-0.001**
3_prime_UTR_variant	2659	2550	4079	13282
5_prime_UTR_variant	1612	1566	2657	8910
downstream_gene_variant	683	656	1012	3350
frameshift_variant	33	82	278	1477
inframe_deletion	27	29	83	370
inframe_insertion	23	34	66	285
intergenic_variant	78	111	124	378
intron_variant	35310	35795	53162	172392
missense_variant	9402	11604	24625	92030
non_coding_transcript_exon_variant	303	308	566	1560
regulatory_region_variant	12	25	41	75
splice_acceptor_variant	16	21	90	418
splice_donor_variant	14	38	112	545
splice_region_variant	1993	1876	3124	10064
start_lost	11	17	39	179
stop_gained	51	96	355	1837
stop_lost	9	12	18	70
stop_retained_variant	13	10	10	59
synonymous_variant	11467	10705	16286	54847
TF_binding_site_variant	4	2	2	13
upstream_gene_variant	1017	1034	1778	5484
Total	64738	66571	108508	367630

Supplementary Table 3. Number of variants and LD properties (based on individual variant LD score) of each of the four groupings of the MAGIC and UKB dataset according to the allele frequency of the variant.

		LD score properties in MAGIC				**LD score properties in UKB**		
MAF bin	**Number of SNPs in MAGIC**	**Mean**	**Median**	**SD**	**Number of SNPs in UKB**	**Mean**	**Median**	**SD**
0.0001-0.001	538560	1.56	1.33	1.25	367587	1.92	1.39	2.08
0.001-0.01	85955	1.61	1.01	3.33	108493	1.75	1.04	2.66
0.01-0.1	36466	3.26	1.58	4.37	66566	3.82	1.96	5.63
0.1-0.5	39425	5.22	3.7	4.82	64722	6.89	4.79	7.16

101 Supplementary Table 4. Grouping of functional consequences of the called sites

Consequence	**VEP Terms/annotation**
Protein-truncating variant (PTV)	“frameshift_variant”, “splice_acceptor_variant”, ”splice_donor_variant”, “stop_gained”, "start_lost"
Missense	"inframe_insertion", "inframe_deletion", "missense_variant", "stop_lost"
Damaging missense (D-Mis)	PolyPhen-2 "probably_damaging" & SIFT "deleterious" & CADD > 15
Benign missense (B-Mis)	PolyPhen-2 "benign" & SIFT "tolerated" & CADD < 15
Synonymous	"synonymous_variant"

102

103

104 Supplementary Table 5. Association of candidate genetic risk scores with high myopia

Derivation Strategy	Tuning Parameter(r2)	Tuning Parameter(p)	R2	N variant in score	OR	CI1	CI2	AUC
Pruning & Thresholding	0.2	0.000005	0.0001	5	1.283	1.169	1.409	0.598
Pruning & Thresholding	0.2	0.0005	0.0001	44	1.452	1.321	1.596	0.621
Pruning & Thresholding	0.2	0.05	0.0217	2464	3.77	3.375	4.214	0.851
Pruning & Thresholding	0.2	0.5	0.0456	22250	4.402	3.953	4.908	0.893
Pruning & Thresholding	0.2	1	0.0486	40491	4.991	4.45	5.606	0.895
Pruning & Thresholding	0.4	0.000005	0.0001	5	1.283	1.169	1.409	0.598
Pruning & Thresholding	0.4	0.0005	0.0002	46	1.454	1.322	1.599	0.624
Pruning & Thresholding	0.4	0.05	0.0217	2583	3.606	3.238	4.019	0.85
Pruning & Thresholding	0.4	0.5	0.0448	23992	4.593	4.108	5.14	0.892
Pruning & Thresholding	0.4	1	0.0474	45132	4.73	4.228	5.299	0.894
Pruning & Thresholding	0.6	0.000005	0.0001	5	1.283	1.169	1.409	0.598
Pruning & Thresholding	0.6	0.0005	0.0002	47	1.459	1.326	1.605	0.625
Pruning & Thresholding	0.6	0.05	0.0214	2670	3.623	3.251	4.042	0.848
Pruning & Thresholding	0.6	0.5	0.0434	25355	4.514	4.045	5.045	0.891
Pruning & Thresholding	0.6	1	0.0457	48878	4.699	4.2	5.263	0.893
Pruning & Thresholding	0.8	0.000005	0.0001	5	1.283	1.169	1.409	0.598
Pruning & Thresholding	0.8	0.0005	0.0001	48	1.422	1.295	1.562	0.623
Pruning & Thresholding	0.8	0.05	0.0201	2792	3.525	3.163	3.932	0.843
Pruning & Thresholding	0.8	0.5	0.0410	26811	4.449	3.987	4.97	0.889
Pruning & Thresholding	0.8	1	0.0429	52624	4.572	4.091	5.116	0.89
LDpred2	na	0.001	0.0055	35348	4.048	2.334	7.204	0.727
LDpred2	na	0.003	0.0138	35348	5.684	3.283	10.066	0.774
LDpred2	na	0.01	0.0238	35348	7.957	4.459	14.785	0.822
LDpred2	na	0.03	0.0296	35348	10.523	5.649	20.866	0.856
LDpred2	na	0.1	0.0330	35348	20.637	8.836	58.935	0.882
LDpred2	na	0.3	0.0310	35348	16.456	8.674	33.474	0.888
LDpred2	na	1	0.0331	35348	28.051	11.251	90.239	0.894
lassosum	na	na	0.0599	74550	1.549	1.404	1.709	0.640

105

106

Supplementary Table 6. The performance metrics of the GRS in the UKB testing cohorts

Models	OR per s.d. (95% CI), P	AUC	PRS threshold	OR (95% CI), P	Prevalence of HM
cvGRS	1.78 (1.69-1.88), 2.14e-105	0.662	Top 20% versus other 80%	2.84 (2.54-3.19), 4.70e-72	0.41
			Top 10% versus other 90%	3.49 (3.01-4.04), 1.86e-62	0.49
			Top 5% versus other 95%	3.67 (3.00-4.48), 1.96e-37	0.52
			Top 2% versus other 980%	5.64 (4.09-7.83), 2.36e-28	0.63
			Top 1% versus other 99%	8.47 (5.20-14.22), 1.92e-21	0.72
rvGRS	1.11 (1.06-1.16), 1.42e-6	0.528	Top 20% versus other 80%	1.35 (1.10-1.41), 0.0003	0.27
			Top 10% versus other 90%	1.36 (1.16-1.59), 0.0001	0.29
			Top 5% versus other 95%	1.53 (1.24-1.89), 0.00008	0.32
			Top 2% versus other 980%	1.85 (1.33-2.56), 0.00016	0.37
			Top 1% versus other 99%	1.42 (0.86-2.28), 0.132	0.31
ExGRS	1.46 (1.41-1.52), 2.35e-82	0.657	Top 20% versus other 80%	2.78 (2.48-3.11), 1.81e-68	0.41
			Top 10% versus other 90%	2.74 (2.36-3.17), 3.55e-40	0.43
			Top 5% versus other 95%	2.03 (1.65-2.49), 2.28e-11	0.38
			Top 2% versus other 980%	2.43 (1.77-3.33), 2.85e-11	0.43
			Top 1% versus other 99%	3.26 (2.09-5.10), 8.94e-9	0.51