A resource-efficient tool for mixed model association analysis of large-scale data

Jiang *et al.*

Supplementary Note 1-6

Supplementary Tables 1-4

Supplementary Figures 1-14

Supplementary References

Supplementary Notes

Supplementary Note 1. Strategies to simulate SNP genotypes from existing GWAS data

To generate a cohort with relatedness and substantial population stratification, we sampled segments of SNP genotypes from existing GWAS data based on a mosaic simulation scheme modified from Ref. [1]. Detailed procedures have been listed below (see **Supplementary Figure 3** for a schematic diagram):

1) Randomly selecting two groups of individuals with different ancestry backgrounds from the UKB participants. Based on the self-reported ethnic background, we first extracted all the UKB participants reported as "British" and "Irish" (see Data Field 21000 of UKB at http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=21000). However, the self-reported ancestry may not be accurate, as we observed inconsistency between one's belief and his/her actual genetic background estimated from genotype data (**Supplementary Figure 1)**. Thus, we selected 9,000 unrelated self-reported "British" and 9,000 unrelated "Irish" participants with relatively large differences in the first two PCs (**Supplementary Figure 2**) to ensure sufficient genetic separation between the two groups. These individuals were the "ancestors" (or founders) of our simulated individuals in step 2.
2) Generating genotypes of 90,000 unrelated individuals (i.e., 45,000 unrelated "British" and 45,000 unrelated "Irish"). To generate 45,000 unrelated "British" individuals, we first divided the genomes (536,684 SNPs in total) of all 9,000 "British" ancestors into 269 consecutive segments of approximately 2,000 SNPs. Then, to simulate the genotype of one individual, 100 ancestors were randomly sampled from the 9,000 "British" ancestors. Next, we randomly selected each segment from one of the 100 ancestors, and aligned all the 269 sampled segments back together to form a new complete genome. This would be one simulated "British" individual. By repeating these steps, 45,000 unrelated "British" individuals were generated. We used the same strategy to generate 45,000 unrelated "Irish" individuals by sampling the segments from the "Irish" ancestors.
3) Generating genotypes of 10,000 related individuals (i.e., 5,000 related "British" and 5,000 related "Irish"). A similar scheme as in step 2) was applied to generate related individuals. We define "related individuals" as those who are related with at least one other individual in the sample with genetic relatedness ≥ 0.05. Two individuals, as one related pair, were generated simultaneously each time. To mimic different degrees of relatedness, the segments for each pair of 1st degree of relatives were randomly sampled from 2 "common ancestors", and the segments for each pair of 2nd degree of relatives were randomly sampled from 4 "common ancestors". Instead of using the original 18,000 ancestors from step 1), we generated an independent set of additional 10,000 "British" individuals and 10,000 "Irish"

individuals as the ancestors for these related individuals. A total of 2,500 pairs of 1st degree of relatives (genetic relatedness = ~0.5; 1,250 "British" pairs and 1,250 "Irish" pairs) and 2,500 pairs of 2nd degree of relatives (genetic relatedness = ~0.25; 1,250 "British" pairs and 1,250 "Irish" pairs) were eventually simulated.

In summary, we generated genotype data of two groups of individuals with reasonably large difference in genetic ancestry between the two groups and substantial proportion of closely related individuals within each group. The difference in ancestry allowed us to simulate the effect of population stratification and the related individuals allowed us to simulate shared environmental effects.

Supplementary Note 2. Strategies for simulating phenotype

We used a set of different parameters to simulate phenotypes based on the simulated genotype and the following model:

$$\boldsymbol{y} = \boldsymbol{g} + \boldsymbol{z}b_p + \boldsymbol{e}_c + \boldsymbol{e}$$

1) $\boldsymbol{g} = \sum_{i=1}^{m} \boldsymbol{x}_i b_i$ where $\boldsymbol{x_i}$ is a vector of genotypes of the i^{th} causal SNP across all individuals with its effect b_i generated from a standard normal distribution $N(0, 1)$. The number of causal variants (m) was set to 10,000 to mimic a polygenic trait, all of which were randomly sampled from SNPs on the odd chromosomes (leaving the SNPs on the even chromosomes to quantify the inflation in test-statistics under the null). We also tested the methods with more causal SNPs, i.e., 20,000, 40,000, or 80,000.
2) $\boldsymbol{z}$ is an indicator vector consists of 0 (indicating "British") and 1 (indicating "Irish") with b_p being the mean difference in phenotype between the two groups. The value of b_p does not matter as $\boldsymbol{z}b_p$ is standardised in the final step, and any positive value of b yields the same result. The purpose of this step is to simulate population stratification effect by creating a phenotypic mean difference between the two ancestry groups.
3) $\boldsymbol{e}_c$ is a vector of shared environmental effects generated by: a) identifying "close relatives" (including the simulated first- and second-degree relatives as well as those pairs with estimated genetic relatedness > 0.05; **Supplementary Note 1**) and grouping them into "families" and extended "families"; b) randomly sampling an effect from a normal distribution to all the individuals in each family.
4) $\boldsymbol{e}$ is a vector of residual effects, randomly generated from a normal distribution $\boldsymbol{e} \sim N(0, 1)$.

The phenotypic value for each sample was a weighted sum of all the standardised components above. The phenotypic variance (V_p) was set to 1. The weights were the square root of the variance proportion of each component, determined by the following: a) genetic variance: $V_g = 0.4 \times V_p$; b) variance due to population stratification: $V_{pop} = 0.05 \times V_p$; c) variance due to common environmental effects: $V_{related} = 0.1 \; or \; 0.2 \times V_p$ (for related individuals); and d) residual variance: $V_{residual} = V_p - V_g - V_{pop} - V_{related}$.

Supplementary Note 3. Constructing relatedness matrix from pedigree data and sparse GRM from SNP data

If pedigree information is available, we can perform a fastGWA-Ped analysis with a relatedness matrix constructed from the expected relatedness coefficients (e.g., 0.5, 0.25, 0.125 and 0.0675 for the first-, second-, third- and fourth-degree relatives, respectively and 1 for monozygotic twins), similar to the traditional family-based MLMs [2-5]. We have provided an R-script (see **URLs**) to construct FAM based on unknown relatedness information (e.g., the inferred relatedness information provided by the UKB). Note that for fastGWA-Ped, to avoid singularity of matrix **V** for traits for which the estimate of residual variance component was negative or close to zero (e.g., < 10% of the phenotypic variance), the non-zero elements in matrix $\boldsymbol{\pi}$ (**Equation 1**) can be recomputed from SNP data with minimal computing cost.

If the pedigree information is incomplete or is not available, the relatedness matrix can be replaced by a sparse GRM. Constructing a GRM could be time-consuming with a runtime of $O(MN^2/2)$ especially when the sample size is large. We have provided a new tool in GCTA to compute a GRM in a very efficient manner via bitwise operations (**URLs**). We have also provided a shell script in the GCTA website (**URLs**) to divide the whole computation process into a number of computing jobs to be parallelized in a high-performance computing system.

In the fastGWA analysis of the UKB data, we computed the GRM for 456,422 individuals of European ancestry using 565,631 slightly LD-pruned HapMap3 [6] SNPs (LD-pruning parameters used in PLINK: window size = 1000kb, step size = 100, r^2 = 0.9 and MAF ≥ 0.01). Note that this set SNPs are sufficient to capture the relatedness among close relatives (e.g., those with genetic relatedness > 0.05). In the fastGWA-ped analysis, we inferred the pedigree relationships based on the KING [7,8] relatedness estimates provided by the UKB. The runtime to build a GRM for the full UKB cohort was around 7 hours with 100 jobs (each job was assigned with 13 GB memory and 6 CPUs). We also have an efficient tool in GCTA to convert a full-dense GRM to a sparse GRM given a relatedness threshold (default value = 0.05) and the computing cost of this conversion is low even for large data set like the UKB. Note that the same set of SNPs were used in the BOLT-REML and GCTA-GREML analyses to estimate the "genetic variance" and/or variance due to common environmental effect (see **Discussion** and **Supplementary Figure 9**).

We noticed that there were some discrepancies between the pedigree relatedness inferred from relatedness estimates provided by the UKB and our sparse GRM (**Supplementary Figure 11**). This is mainly because we used GRM to pick up more distant related pairs with relatedness coefficients between 0.05 to 0.125. In contrast, the relatedness estimates from the UKB were

evaluated according to a more sophisticated strategy (e.g. using a different set of markers with *m* = 93,511 and excluding a small proportion of individuals with higher missingness rates) [8]. The number of related pairs provided by the UKB was 107,162 (no further than third-degree relatives) involving 147,731 unique individuals, while the number of related pairs based on our sparse GRM (estimated from 565,631 common HapMap3 SNPs) was 178,075 with 213,620 unique individuals. These two relatedness estimates were used for two different primary purposes: the UKB estimate was primarily used to make explicit inference about the family relatedness among samples, while our estimate was used to capture the relatedness between all close and distant relatives with the primary aim to control for the confounding in association test.

Supplementary Note 4. Principal component analysis

We compared three different principal component analysis (PCA) methods using our simulated genotype data, namely flashPCA2 (or pruned PCA, with a recommended pruning step and a projection step, see **URL** and Ref. [9]), exact PCA (implemented in GCTA using all the SNPs without pruning, see Ref. [10]), and projection PCA (proj. PCA, implemented in GCTA). The proj. PCA method can be described as: 1) randomly extracting 10,000 individuals from the sample (n = 100,000); 2) performing exact PCA using the subset of individuals and estimating the loadings of each SNP to the top PCs; and 3) computing the top PCs of the whole sample based on the SNP loadings estimated from the subset.

The first and second PCs calculated from flashPCA2, as well as those from exact PCA, were plotted in **Supplementary Figure 4**. Apart from the scale difference, both methods managed to separate the individuals into two groups based on the first two PCs. We then examined the proportion of explained phenotypic variance (V_{pheno}) by the top PCs. There was no significant difference among all three methods (**Supplementary Figure 14**), as all of them accounted for approximately 5% of V_{pheno}, consistent with the parameter used to simulate data. However, fitting PCs calculated from all the SNPs (by exact PCA) would lead to slight deflation of test statistics (**Supplementary Figure 12**). This is because each PC is essentially a feature extracted from the SNPs, and fitting PCs as covariates while testing a target SNP is equivalent to fitting the target SNP more than once in the model, leading to deflated test-statistics under the null. If PCs are computed using a set of LD-pruned SNPs with a relatively stringent threshold, the target SNP is much less likely to be included in computing the PCs. In this case, the test statistics are less likely to be deflated under the null (**Figure 1a)**. We therefore adjusted the phenotypes by the top 10 PCs from flashPCA2 (76,103 LD-pruned SNPs with window size = 1 Mb, step size = 50 SNPs, and LD r^2 threshold = 0.05 as recommended by flashPCA2) in all the subsequent association tests in the simulation study. The number of PCs used is justified by the result that the top 10 PCs were sufficient to capture the majority of phenotype variation due to population stratification (**Supplementary Figure 14**).

In real data analysis, we adjusted each of the 3,613 UKB traits by the top 20 PCs which had been computed by fastPCA [11] using a similar strategy as flashPCA2, as described in Ref. [8].

Supplementary Note 5. LD score for simulated genotype data

LD score of a SNP is defined as the sum of LD r^2 between the target SNP and all the other SNPs in a genomic region adjusting for chance correlations [12]. LD scores are required for both BOLT-LMM and LD score regression (LDSC) analyses. In real data analysis, we used LD scores provided by BOLT-LMM and LDSC software tools (computed from SNP data of individuals of European ancestry in the 1000 Genomes Project [13]; see **URLs**). In the simulation study, we computed LD scores from the simulated genotypes using GCTA [14] (window size = 10 Mb).

Supplementary Note 6. Acknowledgements

UKB: This study has been conducted using UK Biobank resource under Application Number 12514. UK Biobank was established by the Wellcome Trust medical charity, Medical Research Council, Department of Health, Scottish Government and the Northwest Regional Development Agency. It has also had funding from the Welsh Assembly Government, British Heart Foundation and Diabetes UK.

Supplementary Tables

Supplementary Table 1. Abbreviated names of the 24 UKB quantitative traits. The first column is the abbreviation, the second column is the data-field ID of each trait, the third column is the number of records available for analysis, the fourth column is the type of phenotype data (integer or continuous variable), the fifth column contains the full names of these phenotypes, and the last column indicates which traits are female-specific.

Trait Abbr.	Data-field	Count	Type	Description	Note
WC	48	500,376	Continuous	Waist circumference	
HC	49	500,317	Continuous	Hip circumference	
HT	50	499,997	Continuous	Standing height	
WT	21002	499,762	Continuous	Weight	
BMI	21001	499,431	Continuous	Body mass index	
HGSR	47	499,193	Integer	Hand grip strength (right)	
HGSL	46	499,126	Integer	Hand grip strength (left)	
MTCIM	20023	496,709	Integer	Mean time to correctly identify matches	
BMR	23105	492,388	Continuous	Basal metabolic rate	
BFP	23099	492,127	Continuous	Body fat percentage	
DBP	4079	472,416	Integer	Diastolic blood pressure, automated reading	
SBP	4080	472,411	Integer	Systolic blood pressure, automated reading	
FVC	3062	453,724	Continuous	Forced vital capacity	
FEV	3063	453,724	Continuous	Forced expiratory volume in 1-second	
PEF	3064	453,724	Integer	Peak expiratory flow	
NTS	20127	401,596	Integer	Neuroticism score	
EA	845	336,769	Integer	Age completed full time education	
hBMD	78	279,104	Continuous	Heel bone mineral density (BMD) T-score, automated	
BW	20022	277,009	Continuous	Birth weight	
AMena	2714	272,927	Integer	Age at menarche	Female-specific factors
AFLB	2754	184,987	Integer	Age at first live birth	Female-specific factors
PR	4194	170,759	Integer	Pulse rate	
FIS	20016	165,471	Integer	Fluid intelligence score	
AMeno	3581	165,363	Integer	Age at menopause (last menstrual period)	Female-specific factors

Supplementary Table 2. Estimated attenuation ratio (SE) from LD score regression analysis for the 24 UKB traits. Phenotypes are ordered by descending sample size (*n*). The GWAS summary statistics are from fastGWA analyses in this study, the Neale Lab and GeneATLAS, respectively. "\" represents that the trait is not available in GeneATLAS. The abbreviated and full names of the traits are listed in **Supplementary Table 1**.

Trait	Attenuation ratio (fastGWA)	*n* (fastGWA)	Attenuation ratio (Neale Lab)	*n* (Neale Lab)	Attenuation ratio (GeneATLAS)
WC	0.0712 (0.0088)	455,545	0.0712 (0.0104)	360,564	0.0707 (0.008)
HC	0.0865 (0.0087)	455,495	0.0865 (0.0098)	360,521	0.0867 (0.0079)
HT	0.1382 (0.0095)	455,332	0.1206 (0.0089)	360,388	0.112 (0.0068)
WT	0.0859 (0.0081)	455,010	0.0838 (0.0091)	360,116	0.0845 (0.0075)
BMI	0.0705 (0.0076)	454,841	0.0715 (0.0086)	359,983	0.0683 (0.0072)
HGSR	0.0737 (0.0108)	454,473	0.0777 (0.0127)	359,729	0.0658 (0.0096)
HGSL	0.0672 (0.0111)	454,417	0.0672 (0.0125)	359,704	0.0602 (0.0102)
MTCIM	0.0528 (0.0138)	453,043	0.052 (0.0163)	358,695	0.0472 (0.0154)
BMR	0.103 (0.0087)	448,348	0.1001 (0.0093)	354,825	0.0971 (0.0079)
BFP	0.0807 (0.0082)	448,114	0.0791 (0.0093)	354,628	0.0836 (0.0076)
DBP	0.0884 (0.0114)	430,029	0.0836 (0.0129)	340,162	\
SBP	0.0896 (0.0102)	430,025	0.0828 (0.012)	340,159	\
FVC	0.1135 (0.0098)	415,931	0.1055 (0.0104)	329,404	\
FEV	0.0981 (0.0097)	415,931	0.0891 (0.0101)	329,404	\
PEF	0.0878 (0.0135)	415,931	0.0852 (0.0155)	329,404	\
NTS	0.0433 (0.0128)	369,407	0.0443 (0.0155)	293,006	\
EA	0.0908 (0.0156)	304,998	0.084 (0.0171)	240,547	\
hBMD	0.128 (0.0125)	262,294	0.1147 (0.0119)	206,589	\
BW	0.1084 (0.0229)	258,857	0.1113 (0.0244)	205,475	\
AMena	0.0337 (0.011)	240,378	0.038 (0.0119)	188,644	\
AFLB	0.0691 (0.0172)	168,097	0.0733 (0.0186)	131,987	\
PR	0.074 (0.0242)	149,082	0.0687 (0.0284)	118,850	\
FIS	0.052 (0.015)	146,808	0.058 (0.018)	117,131	\
AMeno	0.0376 (0.0329)	141,926	0.0321 (0.0366)	111,593	\

Supplementary Table 3. Number of exome-wide significant associations from the fastGWA analysis of the WES data for 24 traits in the UKB. Phenotypes are ordered by descending sample size (*n*). The abbreviated and full names of the traits are listed in **Supplementary Table 1**. Clumping analysis criteria: *P*-value threshold = 0.05/ the number of tested variants, window size = 5 Mb, and LD r^2 threshold = 0.01. LR: linear regression (if the estimated genetic variance component is not nominally significant). Conditional fastGWA analysis: fastGWA analysis of a WES variant conditioning on the GWAS signals (with 10 Mb of the WES variant) identified from the imputed data.

Trait	*n*	Method	fastGWA analysis	Conditional fastGWA analysis
WC	46135	MLM	0	\
HC	46133	MLM	2	0
HT	46116	MLM	0	\
WT	46067	MLM	5	0
BMI	46051	MLM	66	0
MTCIM	46025	MLM	12	0
HGSL	45973	MLM	8	0
HGSR	45949	MLM	0	\
DBP	45660	MLM	2	0
SBP	45659	MLM	4	0
BMR	45172	MLM	19	0
BFP	45143	MLM	3	0
FEV	41430	MLM	6	1
FVC	41430	MLM	8	1
PEF	41430	MLM	1	0
NTS	38071	MLM	0	\
FIS	37351	MLM	7	2
PR	37166	MLM	6	2
EA	29619	MLM	0	\
BW	27450	MLM	0	\
AMena	24450	MLM	0	\
AFLB	16703	MLM	0	\
AMeno	14088	LR	8	1
hBMD	7066	MLM	1	0
Total	\	\	**158**	**7**

Supplementary Table 4. Summary statistics of the exome-wide significant loci from the fastGWAS analysis of the UKB WES data (*n* = 46,191) for the 24 traits. PLINK clumping criteria: *P*-value threshold = 0.05/the number of variants tested for a trait, window size = 5 Mb, and LD r^2 threshold = 0.01. Each variant is named in a format "Chromosome:Position:Allele 1:Allele 2" based on the Genome Reference Consortium Human Build 38. "EA" = effect allele; "Freq." = frequency of the effect allele; *n* = sample size; "beta" = estimated of variant effect; "se" = standard error of the estimated variant effect; p = p-value. Shown are also the WES GWAS summary statistics from the fastGWA analysis conditioning on GWAS signals (within 10Mb of the WES variant in either direction) identified from the analysis of the whole UKB imputed data (*n* = 456,422).

Trait	Variant	EA	Freq.	*n*	fastGWA			Conditional fastGWA analysis		
					beta	se	p	beta	se	p
AMeno	20:5967581:G:A	A	0.061	14088	0.232	0.0238	1.47E-22	0.152	0.0244	4.42E-10
AMeno	19:55319820:A:G	G	0.391	14088	-0.084	0.0118	8.53E-13	-0.046	0.0120	0.0001
AMeno	1:38874610:C:T	T	0.464	14088	-0.063	0.0115	3.54E-08	-0.025	0.0116	0.030
AMeno	8:38030499:C:G	C	0.225	14087	0.084	0.0136	6.66E-10	0.022	0.0141	0.111
AMeno	4:83472322:C:T	C	0.482	13674	0.062	0.0113	4.29E-08	-0.012	0.0114	0.311
AMeno	16:11898087:C:T	T	0.316	13652	0.063	0.0120	1.85E-07	0.006	0.0121	0.591
AMeno	6:10887043:C:G	C	0.174	14088	0.078	0.0150	1.80E-07	0.002	0.0156	0.911
AMeno	12:66310445:A:G	G	0.031	14088	0.214	0.0331	8.82E-11	0.002	0.0333	0.960
BFP	20:63738996:T:C	T	0.330	43216	-0.028	0.0052	1.14E-07	-0.003	0.0053	0.609
BFP	2:24918669:A:G	G	0.484	45143	0.027	0.0050	5.69E-08	-0.001	0.0051	0.843
BFP	11:27700751:G:T	T	0.319	45143	0.028	0.0054	2.84E-07	-0.0006	0.0054	0.910
BMI	11:27700751:G:T	T	0.319	46051	0.038	0.0069	5.26E-08	0.024	0.0070	0.0005
BMI	1:177929986:G:C	C	0.206	46031	0.047	0.0080	6.42E-09	0.018	0.0081	0.029
BMI	2:24918669:A:G	G	0.484	46051	0.039	0.0065	2.50E-09	0.009	0.0066	0.163
BMI	3:49860567:A:G	A	0.490	46051	-0.033	0.0065	2.60E-07	0.001	0.0067	0.835
BMI	15:67824962:T:A	A	0.225	46051	-0.041	0.0077	9.29E-08	0.001	0.0080	0.914
BMI	18:60372043:C:T	T	0.021	46051	-0.121	0.0229	1.24E-07	-0.002	0.0237	0.929

BMI	4:25407216:G:A	A	0.233	46043	-0.046	0.0077	2.09E-09	0.0003	0.0078	0.969
BMI	19:45678134:G:C	C	0.194	46051	-0.044	0.0082	6.26E-08	-3.71E-05	0.0083	0.996
BMR	3:129252270:T:C	C	0.088	45026	0.041	0.0074	4.11E-08	0.036	0.0075	1.24E-06
BMR	20:35437976:G:A	G	0.407	45172	0.038	0.0043	4.91E-19	0.021	0.0045	3.03E-06
BMR	4:145159425:D:1	A	0.359	45089	0.025	0.0044	1.69E-08	0.015	0.0045	0.0009
BMR	2:23703363:C:T	T	0.130	45171	-0.037	0.0063	6.34E-09	-0.015	0.0064	0.020
BMR	16:30010081:C:T	T	0.394	45172	0.027	0.0043	9.01E-10	0.009	0.0044	0.054
BMR	4:17845658:A:C	C	0.105	43582	-0.038	0.0068	1.46E-08	-0.011	0.0068	0.115
BMR	2:36581207:C:T	T	0.349	44059	0.023	0.0044	2.90E-07	0.007	0.0045	0.145
BMR	6:130060101:T:C	T	0.312	45168	0.029	0.0046	1.31E-10	0.006	0.0047	0.197
BMR	20:33745375:G:T	T	0.252	44799	-0.027	0.0049	4.73E-08	-0.006	0.0051	0.207
BMR	7:92618019:A:G	G	0.250	45166	0.027	0.0049	3.42E-08	0.005	0.0050	0.295
BMR	9:108897159:T:C	C	0.049	45172	0.056	0.0098	1.04E-08	0.009	0.0105	0.373
BMR	1:177929986:G:C	C	0.206	45152	0.031	0.0053	3.04E-09	0.004	0.0056	0.431
BMR	1:155041950:G:A	A	0.202	45159	0.028	0.0053	7.15E-08	0.003	0.0054	0.550
BMR	12:882140:G:A	A	0.199	44544	0.030	0.0053	1.56E-08	0.002	0.0056	0.684
BMR	16:2105296:A:G	G	0.169	45172	-0.030	0.0057	1.84E-07	-0.002	0.0059	0.785
BMR	17:63930138:A:G	A	0.354	45172	-0.025	0.0044	1.61E-08	0.001	0.0047	0.829
BMR	6:7727038:G:A	A	0.467	45172	0.024	0.0043	2.29E-08	0.0004	0.0045	0.934
BMR	17:30899634:A:G	G	0.379	44776	-0.027	0.0044	6.61E-10	6.73E-05	0.0044	0.988
BMR	8:134600502:A:G	G	0.407	45172	-0.023	0.0043	9.34E-08	5.16E-05	0.0046	0.991
DBP	6:28153120:G:A	A	0.243	45660	-0.043	0.0074	5.00E-09	-0.020	0.0075	0.008
DBP	12:111446804:T:C	T	0.482	45660	0.047	0.0064	1.89E-13	0.017	0.0067	0.012
FEV	5:140671341:G:A	A	0.173	41430	-0.041	0.0072	1.39E-08	-0.042	0.0074	1.23E-08
FEV	4:145159425:D:1	A	0.359	41351	0.026	0.0050	1.89E-07	0.015	0.0052	0.004
FEV	6:32058330:C:T	C	0.293	41430	0.040	0.0053	2.82E-14	0.013	0.0054	0.019
FEV	4:105897896:G:A	A	0.257	41427	-0.035	0.0055	2.36E-10	-0.013	0.0056	0.023

FEV	17:46038946:T:C	C	0.219	40751	-0.035	0.0058	1.15E-09	-0.009	0.0059	0.118
FEV	6:35424010:C:T	C	0.228	41430	-0.031	0.0058	1.12E-07	-0.003	0.0060	0.632
FEV	19:8605262:C:T	T	0.037	41430	-0.073	0.0130	2.07E-08	-0.006	0.0144	0.677
FEV	15:83899752:A:G	A	0.475	41429	-0.026	0.0048	6.47E-08	-0.001	0.0049	0.884
FIS	1:43569801:G:A	A	0.375	37351	0.041	0.0072	8.80E-09	0.037	0.0072	3.61E-07
FIS	10:102231624:G:A	G	0.377	37351	0.037	0.0072	3.28E-07	0.025	0.0072	0.0004
FIS	6:30064745:A:C	C	0.131	37320	0.064	0.0103	5.51E-10	0.022	0.0103	0.030
FIS	14:32823916:A:G	A	0.463	37351	-0.037	0.0070	1.60E-07	-0.009	0.0070	0.205
FIS	3:49805448:I:1	TG	0.500	37351	-0.041	0.0070	6.55E-09	-0.001	0.0070	0.895
FIS	11:64242407:G:A	A	0.083	37351	-0.065	0.0126	2.67E-07	\	\	\
FIS	12:132490757:A:G	A	0.130	37348	0.066	0.0112	3.39E-09	\	\	\
FVC	5:140671322:C:T	T	0.173	41430	-0.041	0.0069	2.75E-09	-0.043	0.0070	8.79E-10
FVC	2:55922340:C:G	G	0.192	41423	-0.032	0.0059	3.78E-08	-0.021	0.0061	0.0005
FVC	17:46038946:T:C	C	0.219	40751	-0.038	0.0056	1.88E-11	-0.014	0.0058	0.017
FVC	6:32184217:A:T	A	0.144	41430	0.053	0.0066	1.92E-15	0.010	0.0070	0.167
FVC	15:100152748:G:A	A	0.110	41430	-0.039	0.0074	1.51E-07	-0.001	0.0075	0.910
FVC	19:8605262:C:T	T	0.037	41430	-0.081	0.0125	7.10E-11	-0.001	0.0136	0.965
hBMD	7:121329915:I:2	GCT	0.259	7029	0.159	0.0183	2.66E-18	0.088	0.0201	1.08E-05
HC	16:284580:G:C	C	0.284	46133	-0.038	0.0073	2.54E-07	-0.018	0.0075	0.018
HC	12:882140:G:A	A	0.199	45491	0.042	0.0082	2.79E-07	0.002	0.0083	0.842
HC	15:67824962:T:A	A	0.225	46133	-0.045	0.0079	8.28E-09	0.002	0.0079	0.844
HC	16:31110472:G:A	A	0.359	46133	-0.035	0.0068	3.10E-07	-0.001	0.0070	0.902
HC	20:35437976:G:A	G	0.407	46133	0.038	0.0067	1.08E-08	-0.001	0.0067	0.938
HGSR	17:63842363:A:G	G	0.331	45949	-0.026	0.0048	4.67E-08	-0.018	0.0048	0.0002
HGSR	6:32642624:G:A	A	0.191	45798	-0.030	0.0056	8.87E-08	-0.014	0.0056	0.015
HT	7:2762888:T:C	C	0.301	46116	-0.039	0.0049	3.71E-15	-0.025	0.0052	1.30E-06
HT	15:100152748:G:A	A	0.110	46116	-0.058	0.0072	5.66E-16	-0.040	0.0084	2.23E-06

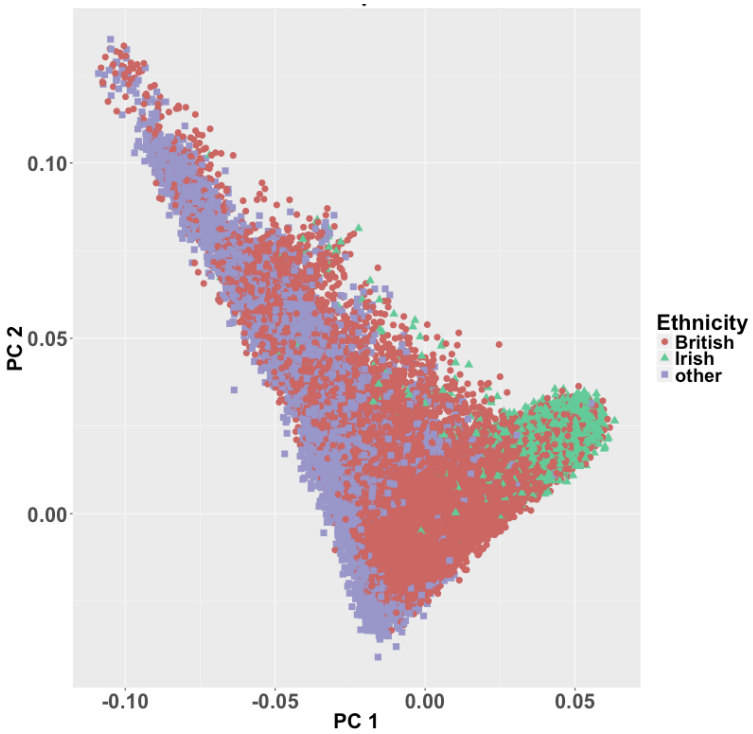
HT	6:26183874:G:A	A	0.258	46116	-0.044	0.0052	2.77E-17	-0.026	0.0058	7.78E-06
HT	5:32711527:C:A	A	0.195	46109	-0.034	0.0057	1.92E-09	-0.027	0.0063	1.27E-05
HT	20:35437976:G:A	G	0.407	46116	0.059	0.0046	9.29E-38	0.021	0.0054	9.84E-05
HT	7:92618019:A:G	G	0.250	46110	0.047	0.0052	3.20E-19	0.021	0.0056	0.0002
HT	8:134637605:G:A	A	0.259	46116	-0.029	0.0052	2.41E-08	-0.021	0.0057	0.0002
HT	10:77830146:C:T	T	0.338	46116	-0.026	0.0048	3.84E-08	-0.018	0.0051	0.0005
HT	22:45327973:C:G	C	0.381	46116	0.024	0.0047	1.83E-07	0.017	0.0051	0.0006
HT	2:219309183:C:T	T	0.117	46082	-0.035	0.0068	2.71E-07	-0.027	0.0077	0.001
HT	5:177089630:G:A	A	0.242	46116	0.042	0.0053	2.52E-15	0.020	0.0060	0.001
HT	18:23135692:G:A	A	0.495	46114	0.031	0.0045	7.07E-12	0.016	0.0047	0.001
HT	5:132336076:T:C	T	0.295	45734	-0.026	0.0049	1.47E-07	-0.017	0.0055	0.002
HT	4:144658755:G:A	A	0.484	45962	0.029	0.0045	2.40E-10	0.014	0.0048	0.003
HT	3:141608632:G:A	A	0.395	46116	0.033	0.0046	7.02E-13	0.015	0.0051	0.003
HT	1:41152770:A:T	A	0.219	46114	0.035	0.0055	2.44E-10	0.017	0.0060	0.004
HT	6:130060101:T:C	T	0.311	46112	0.035	0.0049	1.72E-12	0.015	0.0053	0.004
HT	12:28484182:G:A	A	0.316	45993	-0.035	0.0049	3.55E-13	-0.015	0.0053	0.005
HT	6:31627710:A:G	G	0.365	46101	-0.035	0.0047	1.39E-13	-0.015	0.0053	0.006
HT	9:95447312:G:A	A	0.338	46116	-0.031	0.0048	1.94E-10	-0.015	0.0056	0.009
HT	10:103076290:T:C	C	0.389	46116	0.024	0.0047	2.84E-07	0.013	0.0049	0.009
HT	15:83913152:T:C	T	0.476	46116	-0.036	0.0045	1.02E-15	-0.013	0.0051	0.009
HT	5:108820558:T:A	A	0.228	46116	0.029	0.0054	5.13E-08	0.014	0.0054	0.011
HT	17:7508451:C:T	T	0.216	46116	0.031	0.0055	1.17E-08	0.014	0.0058	0.016
HT	2:232210371:C:T	T	0.028	46116	-0.080	0.0137	5.50E-09	-0.038	0.0157	0.017
HT	1:47333967:G:C	G	0.499	46116	-0.024	0.0045	2.01E-07	-0.011	0.0050	0.027
HT	4:17883363:T:C	C	0.136	46116	-0.054	0.0066	4.66E-16	-0.015	0.0070	0.033
HT	2:25240614:G:A	A	0.411	46116	0.031	0.0046	1.81E-11	0.011	0.0052	0.034
HT	1:16986959:G:A	A	0.234	46116	-0.035	0.0053	4.65E-11	-0.013	0.0063	0.034

HT	20:33745375:G:T	T	0.252	45738	-0.036	0.0052	8.83E-12	-0.013	0.0060	0.035
HT	15:88857449:A:G	G	0.029	46116	-0.121	0.0136	6.20E-19	-0.033	0.0161	0.041
HT	19:8605262:C:T	T	0.037	46116	-0.100	0.0121	8.28E-17	-0.028	0.0144	0.055
HT	6:34246545:C:G	C	0.090	46115	0.064	0.0079	6.18E-16	0.017	0.0090	0.063
HT	6:142403781:C:T	T	0.278	46084	-0.045	0.0050	2.96E-19	-0.010	0.0055	0.064
HT	2:23958289:G:T	T	0.188	46088	-0.033	0.0058	9.73E-09	-0.012	0.0065	0.070
HT	2:55922340:C:G	G	0.192	46108	-0.039	0.0057	1.04E-11	-0.011	0.0060	0.071
HT	6:19838216:C:A	A	0.062	46116	0.076	0.0094	4.99E-16	0.017	0.0101	0.090
HT	15:61910283:C:T	C	0.457	46116	0.024	0.0046	1.10E-07	0.008	0.0048	0.105
HT	8:129748784:G:C	C	0.464	46112	-0.024	0.0045	1.75E-07	-0.007	0.0049	0.129
HT	17:63830956:G:A	A	0.284	46115	0.036	0.0050	3.84E-13	0.010	0.0064	0.133
HT	1:149934520:T:C	C	0.407	46116	0.040	0.0046	2.08E-18	0.007	0.0047	0.147
HT	6:34857885:T:C	C	0.138	46116	0.042	0.0066	2.51E-10	0.010	0.0075	0.191
HT	3:53099510:A:G	A	0.402	46031	0.028	0.0046	7.15E-10	0.006	0.0051	0.242
HT	6:151807942:T:C	C	0.477	46116	0.027	0.0045	3.31E-09	0.005	0.0050	0.286
HT	6:34871867:G:A	A	0.016	46116	-0.108	0.0182	3.25E-09	-0.021	0.0205	0.298
HT	3:172447937:C:T	T	0.313	46116	0.025	0.0049	2.36E-07	0.005	0.0051	0.357
HT	20:49158557:C:T	T	0.238	46093	0.029	0.0053	7.51E-08	0.004	0.0059	0.473
HT	11:65965994:G:A	A	0.061	46116	-0.069	0.0095	3.48E-13	-0.008	0.0106	0.474
HT	17:30784350:A:G	G	0.177	46116	-0.040	0.0060	1.40E-11	-0.004	0.0065	0.520
HT	7:66286473:T:A	A	0.190	46066	0.030	0.0058	1.54E-07	0.004	0.0062	0.530
HT	1:88770346:T:A	T	0.458	45939	-0.026	0.0045	1.58E-08	-0.002	0.0046	0.640
HT	6:7727038:G:A	A	0.466	46116	0.033	0.0045	3.07E-13	0.002	0.0049	0.663
HT	19:4954443:G:A	A	0.198	46116	-0.030	0.0057	9.09E-08	0.003	0.0063	0.667
HT	3:129301935:A:G	A	0.219	46114	-0.030	0.0055	4.05E-08	-0.002	0.0058	0.743
HT	1:184051811:G:A	A	0.348	46116	0.032	0.0048	2.46E-11	0.001	0.0050	0.778
HT	11:75566583:A:G	A	0.115	46116	0.040	0.0071	1.34E-08	0.002	0.0077	0.800

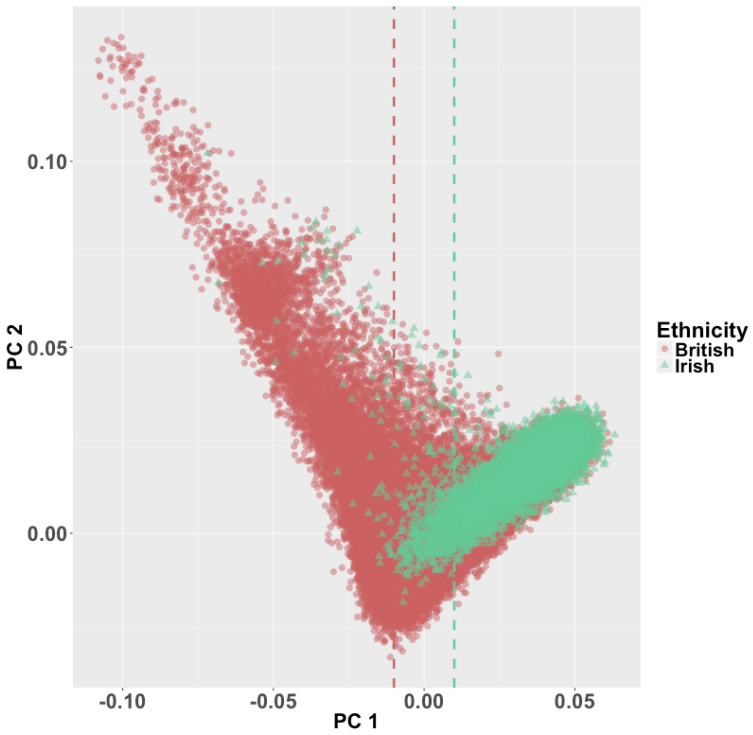
HT	12:123364549:C:G	G	0.205	46116	0.032	0.0056	1.75E-08	0.001	0.0058	0.802
HT	14:65075889:C:T	T	0.422	46116	-0.025	0.0046	3.63E-08	-0.001	0.0051	0.806
HT	10:68199588:D:6	G	0.493	46094	0.025	0.0045	2.00E-08	0.001	0.0048	0.901
HT	2:88575373:C:A	C	0.281	46116	0.030	0.0050	2.35E-09	-0.001	0.0051	0.905
HT	14:94378610:C:T	T	0.019	46116	0.117	0.0165	1.42E-12	0.001	0.0168	0.930
HT	15:67165360:A:G	G	0.056	46116	0.069	0.0098	1.76E-12	-0.001	0.0105	0.935
HT	13:49663119:G:A	A	0.022	46116	0.083	0.0153	6.54E-08	0.001	0.0162	0.936
HT	17:29562968:T:C	T	0.343	46116	-0.029	0.0048	8.29E-10	-0.0003	0.0051	0.958
HT	9:96388497:A:G	G	0.163	46110	0.036	0.0061	3.68E-09	-0.0003	0.0070	0.961
HT	19:55482069:G:T	T	0.026	46116	-0.073	0.0143	3.20E-07	0.0003	0.0153	0.986
PEF	9:133372523:G:C	C	0.137	41014	-0.042	0.0078	6.20E-08	-0.025	0.0078	0.001
PR	20:44311053:C:T	T	0.015	37166	0.163	0.0297	3.95E-08	0.161	0.0297	5.74E-08
PR	16:16056044:C:T	T	0.143	36809	-0.054	0.0104	2.15E-07	-0.055	0.0104	1.62E-07
PR	20:38213354:T:C	C	0.474	37166	-0.074	0.0073	1.82E-24	-0.032	0.0073	8.64E-06
PR	14:23396676:G:A	A	0.357	37166	0.070	0.0076	5.31E-20	0.003	0.0076	0.711
PR	7:100889133:G:C	C	0.183	37166	0.052	0.0094	3.41E-08	-0.001	0.0094	0.947
PR	2:178856319:G:A	A	0.086	37166	0.094	0.0130	4.18E-13	-0.001	0.0130	0.958
SBP	5:32713221:T:C	C	0.393	45659	-0.035	0.0061	1.20E-08	-0.013	0.0063	0.035
SBP	11:47792728:A:G	G	0.454	45643	0.031	0.0060	2.02E-07	0.007	0.0061	0.246
SBP	16:24823847:G:A	A	0.194	45659	-0.044	0.0076	6.14E-09	0.001	0.0076	0.934
SBP	1:11823674:C:T	T	0.162	45659	-0.049	0.0081	2.39E-09	-0.0004	0.0085	0.966
WT	9:108990879:G:T	T	0.075	46067	0.061	0.0109	2.13E-08	0.046	0.0111	3.18E-05
WT	16:284580:G:C	C	0.284	46067	-0.033	0.0064	1.90E-07	-0.015	0.0065	0.020
WT	1:177929986:G:C	C	0.206	46047	0.042	0.0071	2.78E-09	0.015	0.0073	0.033
WT	2:36581207:C:T	T	0.349	44945	0.030	0.0059	2.96E-07	0.011	0.0060	0.076
WT	15:67824962:T:A	A	0.225	46067	-0.035	0.0068	2.82E-07	-0.012	0.0070	0.093
WT	16:30010081:C:T	T	0.394	46067	0.032	0.0058	6.71E-08	0.009	0.0060	0.123

WT	20:35434589:C:A	C	0.346	46067	0.041	0.0060	1.61E-11	0.009	0.0062	0.154
WT	4:145159425:D:1	A	0.359	45981	0.031	0.0060	1.45E-07	0.004	0.0062	0.551
WT	17:30899634:A:G	G	0.379	45665	-0.032	0.0059	5.89E-08	-0.003	0.0059	0.587
WT	12:882140:G:A	A	0.199	45425	0.038	0.0071	7.71E-08	0.002	0.0073	0.787
WT	11:27700751:G:T	T	0.319	46067	0.033	0.0061	7.76E-08	-0.001	0.0063	0.932
WT	18:60372043:C:T	T	0.021	46067	-0.105	0.0202	2.44E-07	0.001	0.0210	0.974

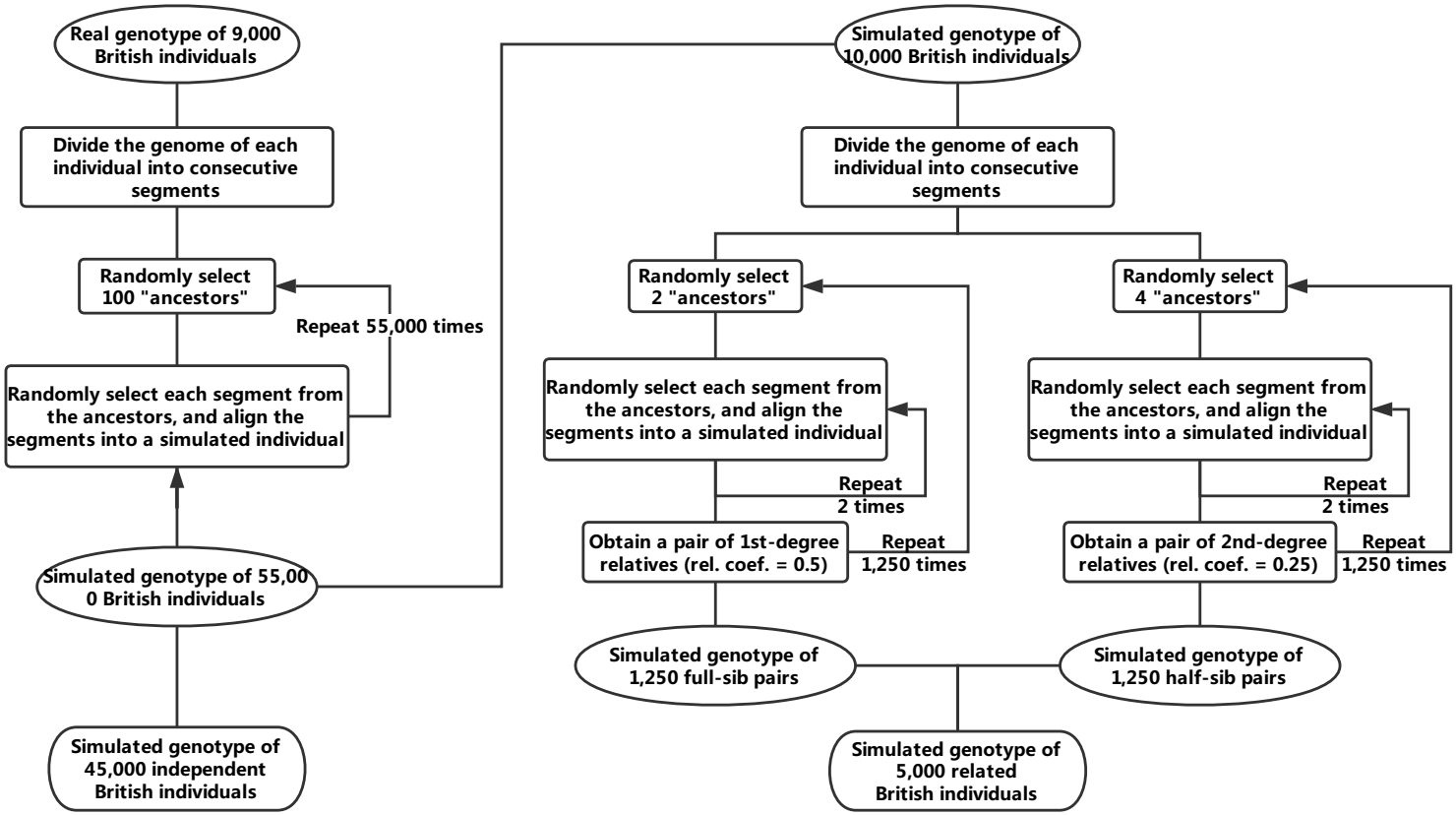
Supplementary Figures



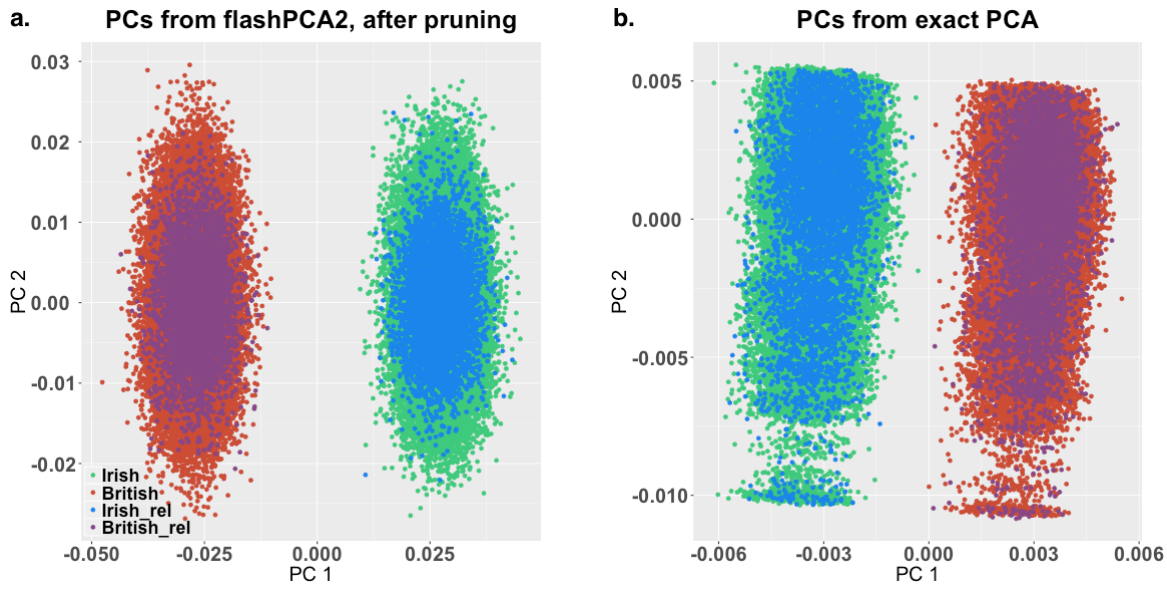
Supplementary Figure 1. The first and second principal components (PC1 and PC2) of all the UK Biobank participants of European ancestry compared to their self-reported ethnicity. The red dots represent the ones self-reported as "British", the green dots represent those self-reported as "Irish", and the purple dots represent those self-reported as "other-white background".



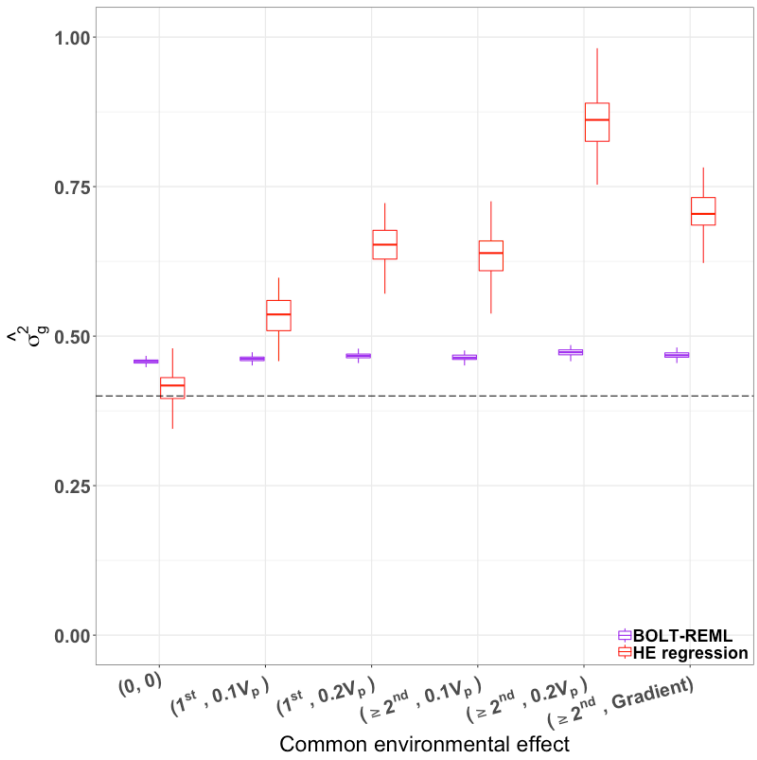
Supplementary Figure 2. The first and second principal components (PC1 and PC2) plotted against self-reported ethnicity among individuals of Irish and British ancestry from the UKB. In the simulation, we randomly selected 9,000 "Irish" individuals from the green dots on the right-hand side of the green vertical line (PC1 ≥ 0.01), and 9,000 "British" individuals from the red dots on the left-hand side of the red vertical line (PC1 ≤ -0.01) (see **Supplementary Note 1** for details of the simulation).



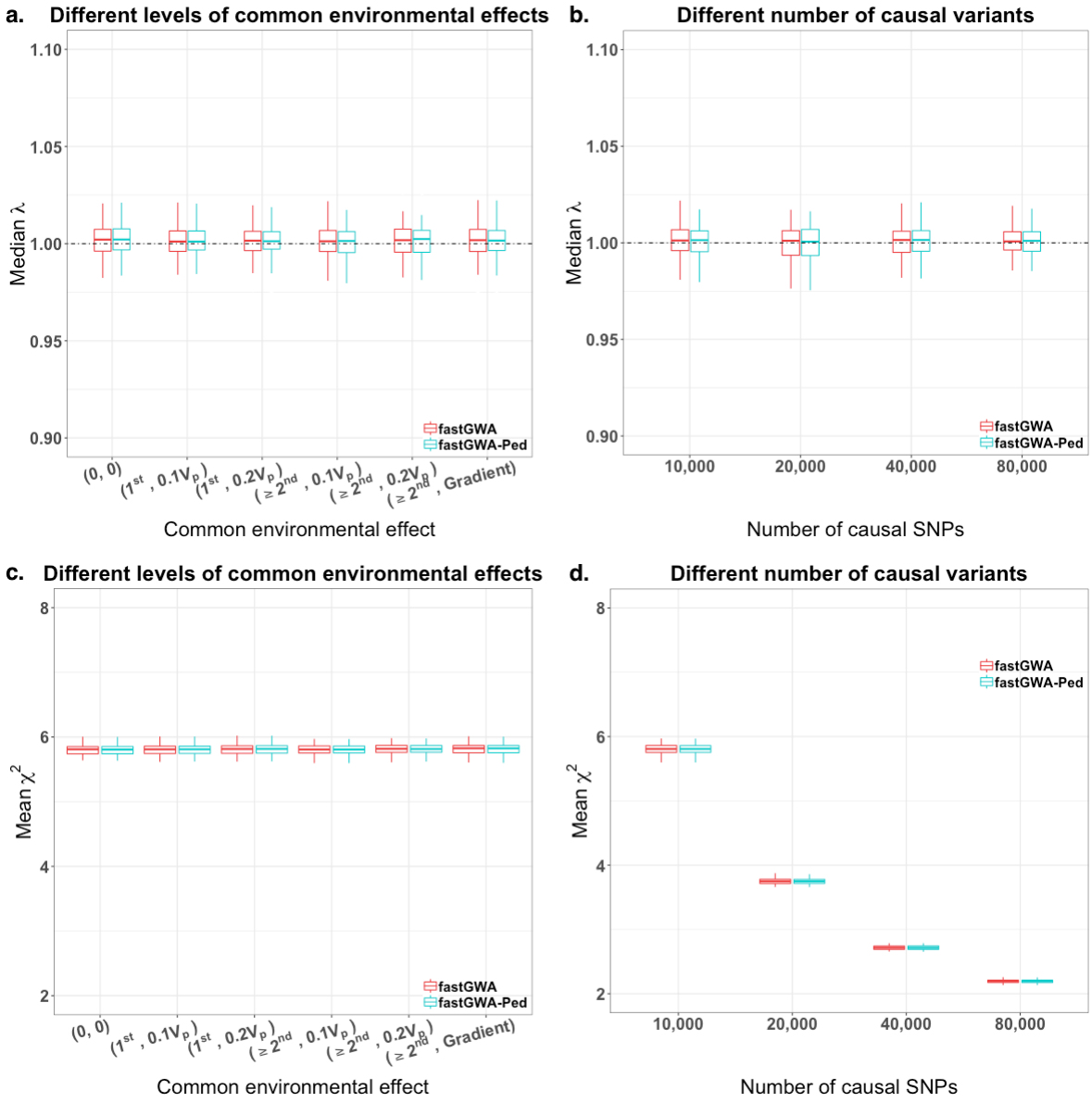
Supplementary Figure 3. Schematic diagram of simulating a GWAS data set with relatedness and population stratification from existing GWAS data. "rel. coef.": relatedness coefficient.



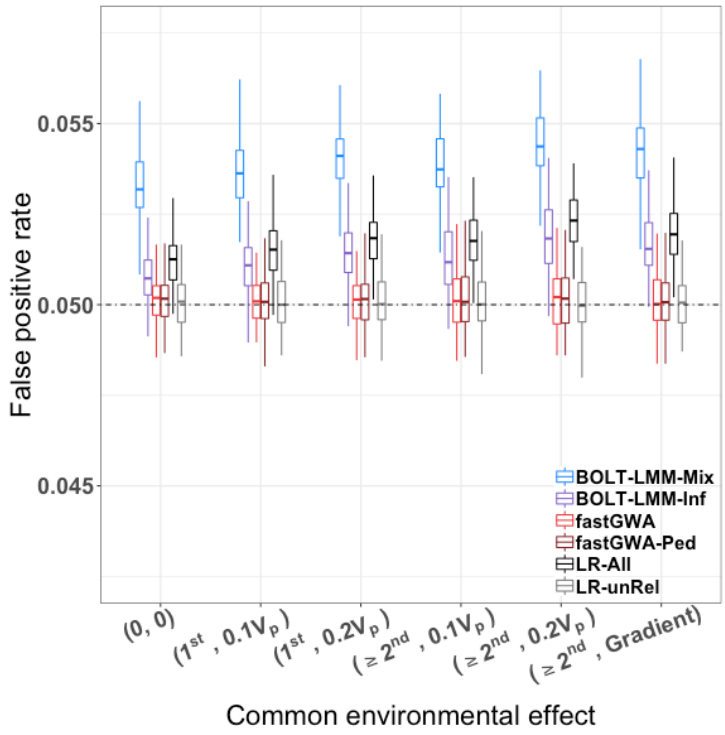
Supplementary Figure 4. Principal component analysis of all the 100,000 simulated individuals. The left panel (a) shows the first two PCs (PC1 and PC2) computed from a set of pruned SNPs (window size = 1 Mb, step size =50 and LD r^2 threshold = 0.05) by flashPCA2, while the right panel (b) shows the first two PCs from exact PCA without LD pruning implemented in GCTA. The red dots represent the simulated British individuals, while the green dots represent the simulated Irish individuals. In panels a) and b), the related individuals (relatedness coefficients ≥ 0.05) were labelled with a slightly darker colour in each group (Irish_rel and British_rel).



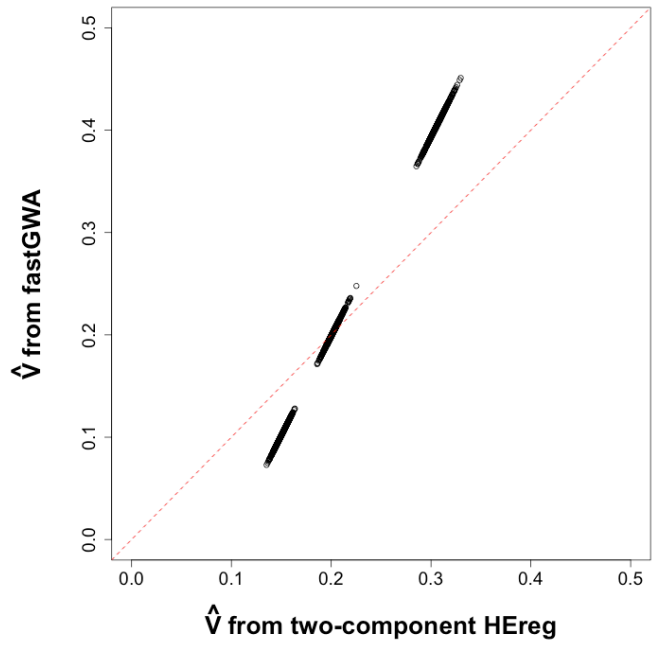
Supplementary Figure 5. Comparison of $\hat{\sigma}_g^2$ between HE regression (used in fastGWA) and BOLT-REML (used in BOLT-LMM) at different levels of relatedness in simulations. The x-axis represents different levels of relatedness, where (0, 0) represents no common environmental effect; (1^{st}, $0.1V_p$) or (1^{st}, $0.2V_p$) represents that common environmental effects explained 10% or 20% of the phenotypic variance (V_p) among 1^{st} degree relatives; ($\geq 2^{st}$, $0.1V_p$) or ($\geq 2^{st}$, $0.2V_p$) represents that common environmental effects explained 10% or 20% of V_p among all pairs of the 1^{st} and 2^{nd} degree relatives; ($\geq 2^{st}$, Gradient) represents that common environmental effects explained 20% of V_p among the 1^{st} degree relatives and 10% of V_p among the 2^{nd} degree relatives. The y-axis represents the value of $\hat{\sigma}_g^2$. The black dashed line represents the true simulation parameter (h^2 = 0.4).



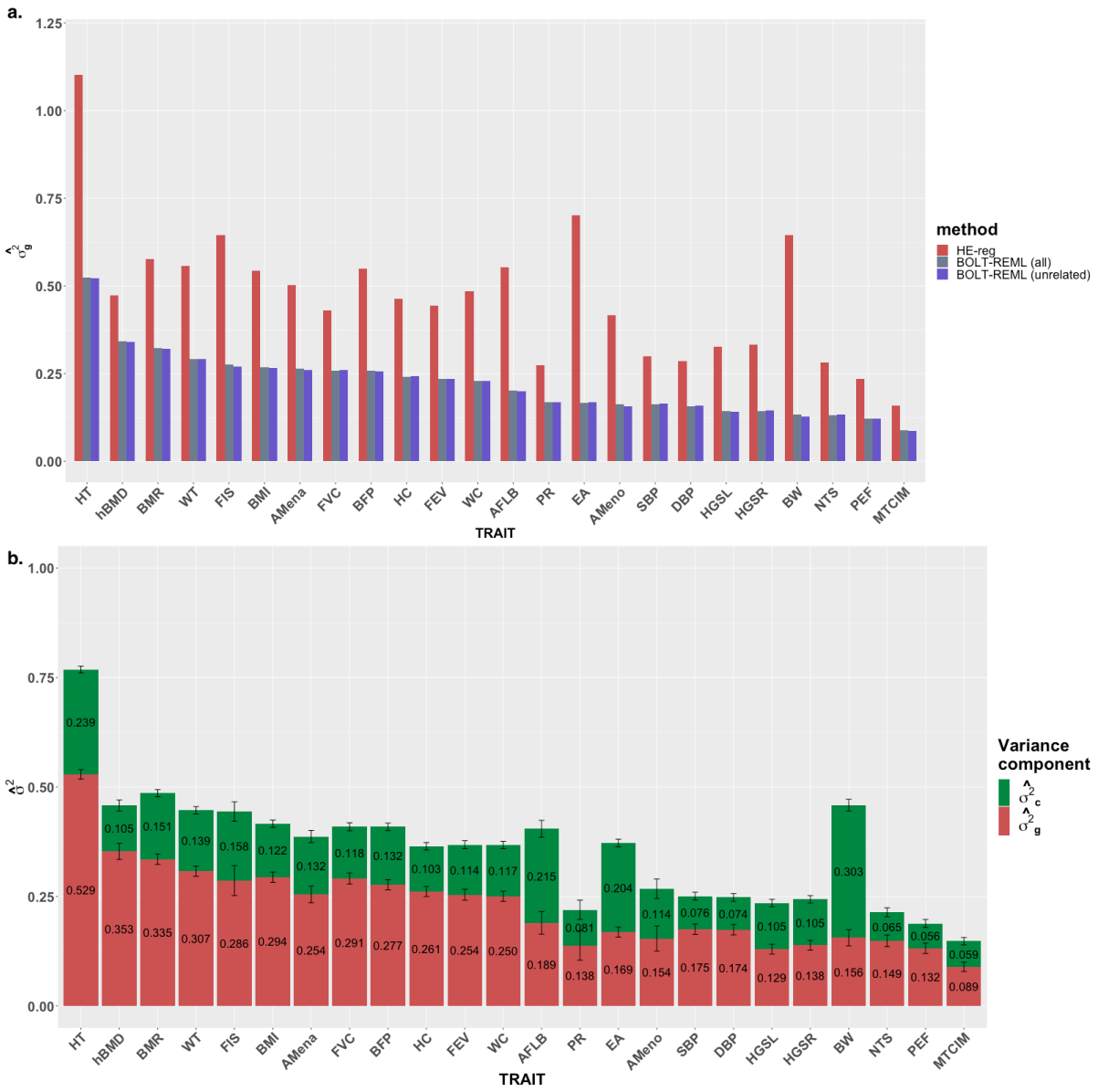
Supplementary Figure 6. Comparison between fastGWA and fastGWA-Ped. Panels a) and b) show the median λ of null SNPs for fastGWA and fastGWA-Ped, respectively. Panel a) shows the median λ with different levels of common environmental effects, and panel b) shows the median λ with different number of simulated causal variants. Panels c) and d) show the mean χ^2 vlaue of causal SNPs for fastGWA and fastGWA-Ped. Panel c) shows the mean χ^2 value with different levels of common environmental effects, and panel d) shows the mean χ^2 value with different number of simulated causal variants. In all the panels, each box plot represents the distribution of the estimates (i.e., median λ and mean χ^2) across 100 simulation replicates.



Supplementary Figure 7. Comparison of false positive rate (FPR) for different association methods. We used the simulated data as presented in Figures 1 and 2 to compute the FPR of each association method across different simulation scenarios with different levels of common environmental effects. Each boxplot represents the distribution of FPR across 100 simulation replicates.

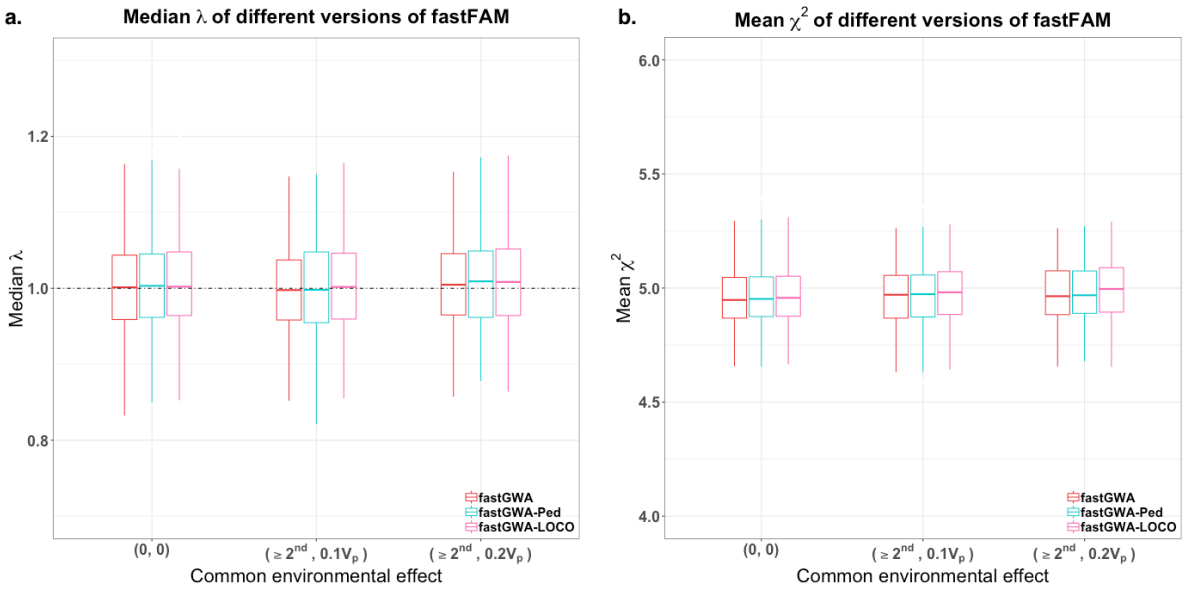


Supplementary Figure 8. Comparison of the estimated matrix $\boldsymbol{V}$ from a single-component model (i.e., the fastGWA model) with that from a the two-component model. Single-component model (fastGWA): $\boldsymbol{y} = \boldsymbol{g} + \boldsymbol{e}$ with $\boldsymbol{V} = \boldsymbol{\pi}\sigma_g^2 + \boldsymbol{I}\sigma_e^2$ (see Equation 1 in the main text for the definitions of all the parameters and variables). Two-component model: $\boldsymbol{y} = \boldsymbol{g} + \boldsymbol{e_C} + \boldsymbol{e}$ with $\boldsymbol{V} = \boldsymbol{\pi}\sigma_g^2 + \boldsymbol{C}\sigma_c^2 + \boldsymbol{I}\sigma_e^2$ where $\boldsymbol{e_C}$ is a vector of shared environmental effects (see Equation 3 in the main text) and $\boldsymbol{C}$ is a design matrix with 1 or 0 to indicate whether a pair of individuals belong to the same family. Shown are the result from the analyses of a simulated data set based on the simulation strategy described in **Supplementary Note 2**. We generated a sample of 1,000 pairs of first-degree relatives, 1,000 pairs of second-degree relatives, and 1,000 pairs of third-degree relatives. We then generated phenotype for each individual based on the two-component model with $\sigma_g^2 = 0.4$, $\sigma_c^2 = 0.2$ among the first-degree relatives, $\sigma_c^2 = 0.1$ among the second-degree relatives, and $\sigma_c^2 = 0$ among the third-degree relatives. We then estimated $\widehat{\boldsymbol{V}}$ based on both the single- and two-component models using HE regression. Plotted are the non-zero off-diagonal elements of the estimated $\widehat{\boldsymbol{V}}$ from fastGWA against those from the two-component model.

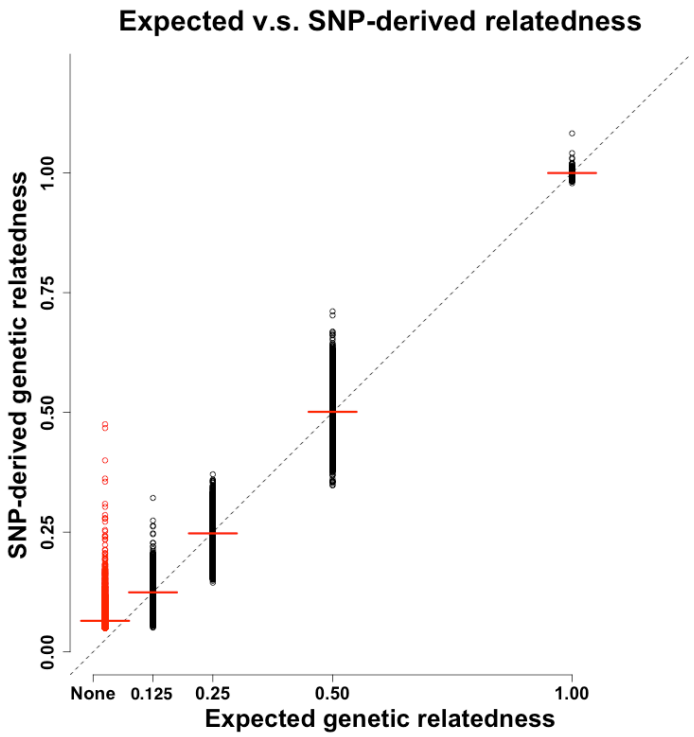


Supplementary Figure 9. Estimates of genetic variance by HE regression and BOLT-REML for 24 traits in the UKB. A full list of phenotype abbreviations can be found in **Supplementary Table 1**. Shown in panel a) are the estimates of the genetic variance (i.e., $\hat{\sigma}_g^2$) by HE regression based on the sparse GRM (used in fastGWA), by BOLT-REML [1,15] using all individuals (used in BOLT-LMM), and by a refined-version of BOLT-REML [16] using only unrelated individuals. In panel b), we analysed a subset of the UKB data (21,815 inferred full-sib pairs, consisting of 39,934 individuals from 19,386 families; see **Supplementary Note 3**) based on a two-component model: $\boldsymbol{y} = \boldsymbol{g} + \boldsymbol{e_C} + \boldsymbol{e}$ with $\boldsymbol{V} = \boldsymbol{\pi}\sigma_g^2 + \boldsymbol{C}\sigma_c^2 + \boldsymbol{I}\sigma_e^2$ where $\boldsymbol{g}$ is a vector of total genetic effects, $\boldsymbol{e_C}$ is a vector of shared environmental effects (see Equation 3 in the main text), $\boldsymbol{\pi}$ is the full dense GRM estimated from the same slightly-clumped HapMap3 SNPs used in real data analyses (m = 565,631, see **Supplementary Note 3** for details), and $\boldsymbol{C}$ is a design matrix with 1 or 0 to indicate whether a

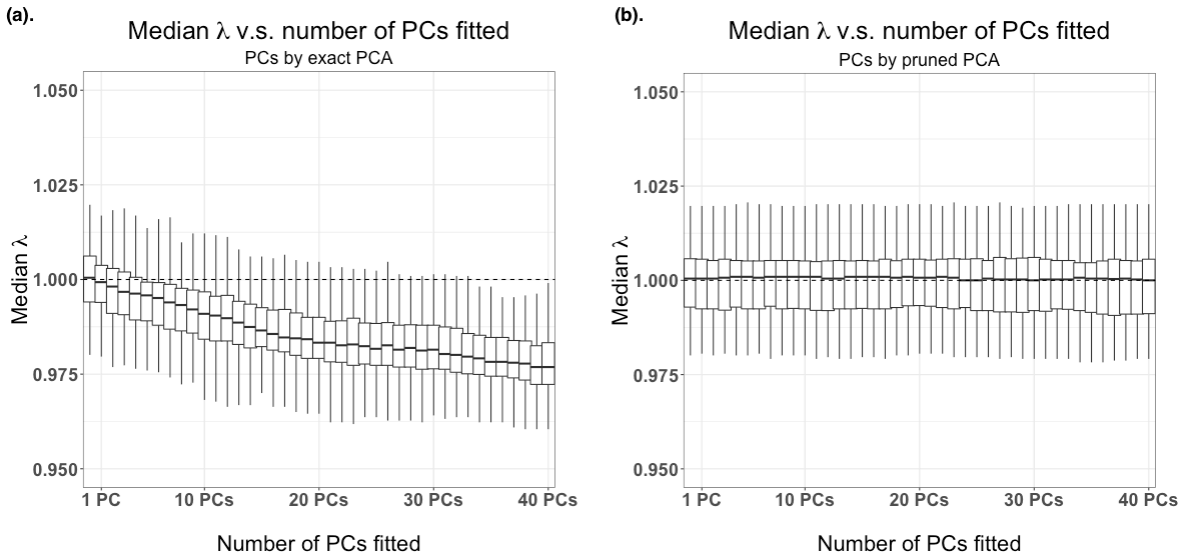
pair of individuals belong to the same family. Standard errors of the estimates are represented by the error bars.



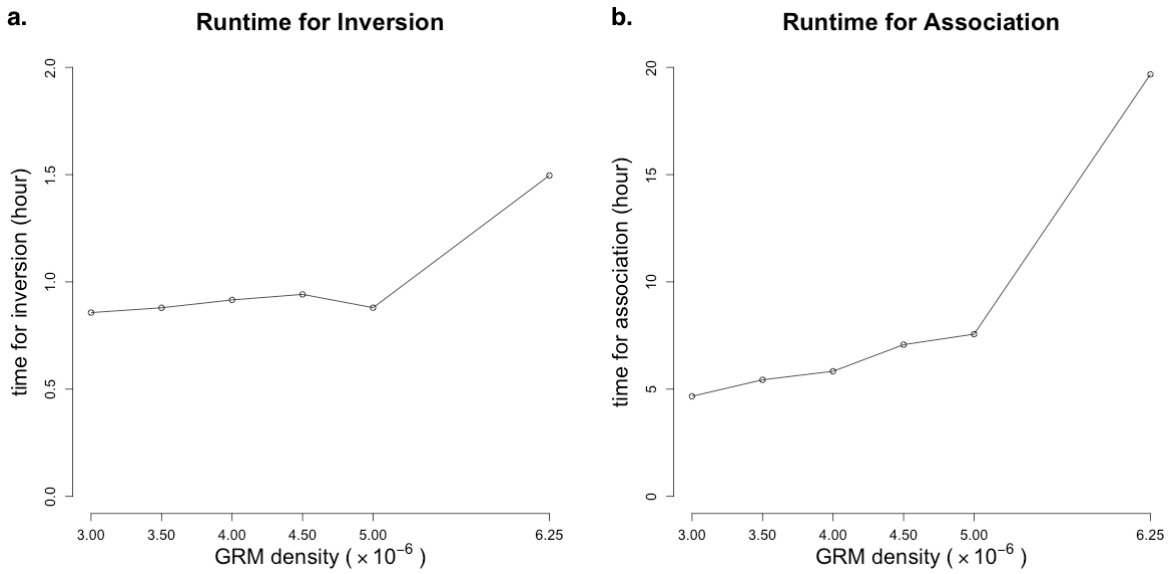
Supplementary Figure 10. Comparison of the power and inflation of fastGWA-LOCO with fastGWA and fastGWA-Ped. Shown are the results from the analyses of a simulated data set based on the simulation strategy described in **Supplementary Note 2** (with $\sigma_g^2 = 0.4V_p$, $\sigma_c^2 = 0.1V_p, or\ 0.2V_p$ for all 1st and 2nd relatives and $\sigma_c^2 = 0$ for all unrelated individuals).



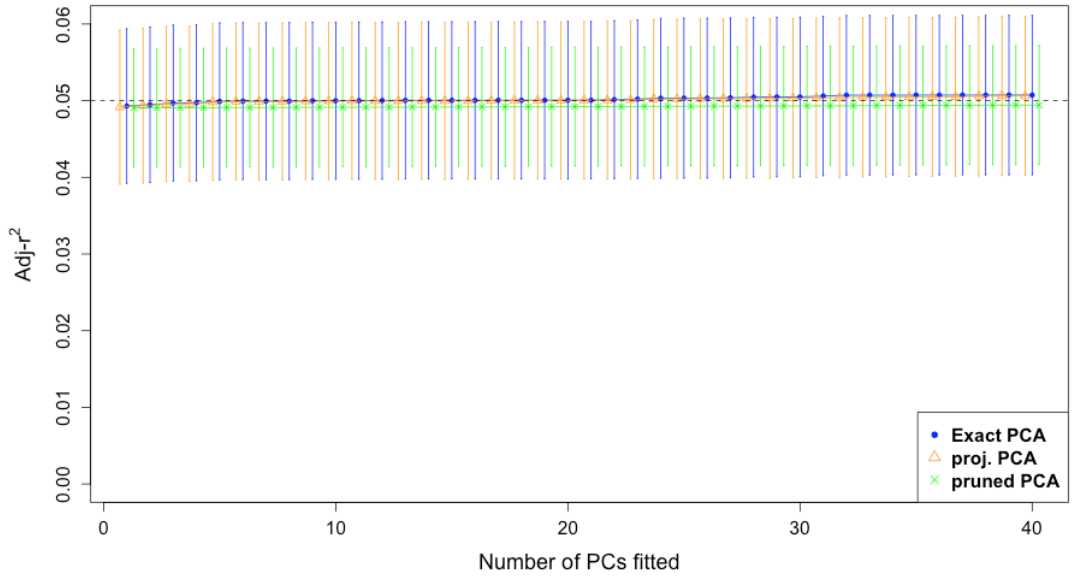
Supplementary Figure 11. A comparison between the reported genetic relatedness and the SNP-derived genetic relatedness of the UKB participants. The y-axis represents the SNP-derived genetic relatedness computed from GCTA using ~544k common SNPs on HapMap3 (178,075 individual pairs with estimated genetic relatedness ≥ 0.05). The x-axis represents the expected genetic relatedness based on the pedigree information provided by the UKB (monozygotic twin = 1, parent-offspring/full sib = 0.5, second degree relatives = 0.25, third degree relatives = 0.125, and unlabelled pair = 'none') on x-axis. Each circle represents one pair of relatives, the dashed diagonal line represents y = x, and the red horizontal lines represent the mean value of each relatedness group.



Supplementary Figure 12. The relationship between the test statistics (i.e., median λ) and the number of PCs fitted in the simulation. The left panel shows the results of fitting PCs by exact PCA (all SNPs without pruning), while the right panel shows the results of fitting PCs by PCA using LD-pruned SNPs. The phenotypes were adjusted by different number of top PCs (ranging from 1 to 40). The association analyses were performed using PLINK (linear regression). The median λ was computed from of all the null SNPs (i.e., SNPs on even chromosomes). Each boxplot represents the distribution of median λ across 100 simulation replicates.



Supplementary Figure 13. The relationship between GRM density and the runtime of fastGWA. a) runtime for inverting the variance-covariance matrix. b) runtime for association test. The x-axis represents different levels of GRM density. The total sample size was fixed to be 400,000, and the number of related pairs ranged from 40,000 (GRM density = 3×10^{-6}) to 300,000 (GRM density = 6.25×10^{-6}).



Supplementary Figure 14. Proportion of phenotypic variance explained by the top 40 PCs computed from different PCA methods in the simulation. Adjusted r^2 (adjusted coefficient of determination) was plotted against the different number of PCs fitted in the model. Three methods, the Exact PCA (Exact PCA, implemented in GCTA) using all SNPs, the PC projection approach (proj. PCA, implemented in GCTA) using all SNPs, and flashPCA2 (pruned PCA) using a set of LD-pruned SNPs, are compared. The dash line represented the parameter used to simulate the proportion of variance explained by population stratification. Each dot represents the average across 100 simulation replicates.

Supplementary References

1. Loh, P.-R., et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* (2015).
2. Abecasis, G.R., Cardon, L.R. & Cookson, W. A general test of association for quantitative traits in nuclear families. *The American Journal of Human Genetics* **66**, 279-292 (2000).
3. Lange, C., DeMeo, D.L. & Laird, N.M. Power and design considerations for a general class of family-based association tests: quantitative traits. *The American Journal of Human Genetics* **71**, 1330-1341 (2002).
4. Yu, J. *et al.* A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature genetics* **38**, 203-208 (2006).
5. Chen, W.-M. & Abecasis, G.R. Family-based association tests for genomewide association scans. *The American Journal of Human Genetics* **81**, 913-926 (2007).
6. International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52 (2010).
7. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-73 (2010).
8. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).
9. Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* **33**, 2776-2778 (2017).
10. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
11. Galinsky, K.J. *et al.* Fast principal-component analysis reveals convergent evolution of ADH1B in Europe and East Asia. *The American Journal of Human Genetics* **98**, 456-472 (2016).
12. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
13. Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56 (2012).
14. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
15. Loh, P.R., Kichaev, G., Gazal, S., Schoech, A.P. & Price, A.L. Mixed-model association for biobank-scale datasets. *Nat Genet* **50**, 906-908 (2018).
16. Loh, P.R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat Genet* **47**, 1385-92 (2015).