nature genetics

TECHNICAL REPORT

https://doi.org/10.1038/s41588-020-0613-6

Liability threshold modeling of case–control status and family history of disease increases association power

Margaux L. A. Hujoel[1]✉, Steven Gazal[2,3], Po-Ru Loh[3,4], Nick Patterson[3] and Alkes L. Price[1,2,3]✉

Family history of disease can provide valuable information in case–control association studies, but it is currently unclear how to best combine case–control status and family history of disease. We developed an association method based on posterior mean genetic liabilities under a liability threshold model, conditional on case–control status and family history (LT-FH). Analyzing 12 diseases from the UK Biobank (average *N* = 350,000) we compared LT-FH to genome-wide association without using family history (GWAS) and a previous proxy-based method incorporating family history (GWAX). LT-FH was 63% (standard error (s.e.) 6%) more powerful than GWAS and 36% (s.e. 4%) more powerful than the trait-specific maximum of GWAS and GWAX, based on the number of independent genome-wide-significant loci across all diseases (for example, 690 loci for LT-FH versus 423 for GWAS); relative improvements were similar when applying BOLT-LMM to GWAS, GWAX and LT-FH phenotypes. Thus, LT-FH greatly increases association power when family history of disease is available.

Family history of disease can provide valuable information about an individual's genetic liability for disease, potentially increasing the power of case–control association studies[1], the focus of this study. More broadly, leveraging data from ungenotyped but phenotyped relatives has a rich history in genetic risk prediction[2] and analyses of quantitative traits in humans[3] and livestock[4–6]. Standard case–control genome-wide association studies (GWAS) ignore family history information (Fig. 1a). A recent method, genome-wide association by proxy[1] (GWAX), compares disease cases and 'proxy cases' (controls with family history of disease) to controls without family history of disease (Fig. 1a). This approach has proven successful in studies of Alzheimer's disease[7,8] and has provided a valuable contribution in highlighting the value of family history information, but more accurate modeling of family history—for example, distinguishing disease cases from proxy cases—may be beneficial[1].

We propose a new association method based on posterior mean genetic liabilities under a liability threshold model, conditional on both case–control status and family history (LT-FH). The liability threshold model, in which an individual is a disease case only if an underlying continuous-valued liability lies above a threshold, has proven valuable in a wide range of settings[2,9–13]. LT-FH computes association statistics via linear regression of genotypes and posterior mean genetic liabilities; association statistics can also be computed using efficient mixed-model methods[14,15]. LT-FH allows for higher genetic liability in controls with family history as compared to controls without family history (likewise for disease cases; Fig. 1a). In contrast to GWAX, which assigns the same binary phenotype to disease cases and proxy cases, LT-FH accurately models a broad range of case–control status and family history configurations, greatly increasing power in analyses of 12 diseases of both low and high prevalence from the UK Biobank.

Results

Overview of methods. The LT-FH method relies on the liability threshold model[9] and consists of two main steps (Fig. 1b): (1) compute posterior mean genetic liabilities for each genotyped individual, conditional on available case–control status and/or family history information, and (2) compute association statistics via linear regression of genotypes and posterior mean genetic liabilities. In step 1, we first compute posterior mean genetic liabilities for every possible configuration of case–control status and family history (377 configurations, accounting for parental history, sibling history, missing data and so forth) and then perform a lookup for each genotyped individual, avoiding redundant computation. The posterior mean genetic liability for each configuration is computed via Monte Carlo integration, incorporating estimates of disease prevalence and parental disease prevalence from the target samples and estimates of narrow-sense heritability from the literature, which differs from SNP heritability[16]. We show that Monte Carlo integration outperforms an analytical approach, the Pearson–Aitken (PA) formula[2,17,18], in analyses of sibling history data.

Step 2 generalizes the Armitage trend test[19], and we show that it is equivalent to a score test; an alternative is to apply BOLT-LMM[14,15] (see Code availability) to posterior mean genetic liabilities, increasing the power in large samples. We emphasize the use of posterior mean genetic liabilities. We note that raw LT-FH effect sizes are not on the liability scale, but can nonetheless be transformed to the observed scale. Details of LT-FH are provided in the Methods; we have publicly released open-source software implementing this method (see Code availability). We consider a simple example of a disease with a prevalence of 5% and narrow-sense heritability of 50%, using parental history only. The posterior mean genetic liability is −0.10 for a control with no affected parents versus 0.39 for a control with one affected parent, but GWAS treats these individuals identically. In addition, the posterior mean genetic liability is 1.56

[1]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [2]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [3]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [4]Brigham and Women's Hospital/Harvard Medical School, Boston, MA, USA. ✉e-mail: hujoel@g.harvard.edu; aprice@hsph.harvard.edu

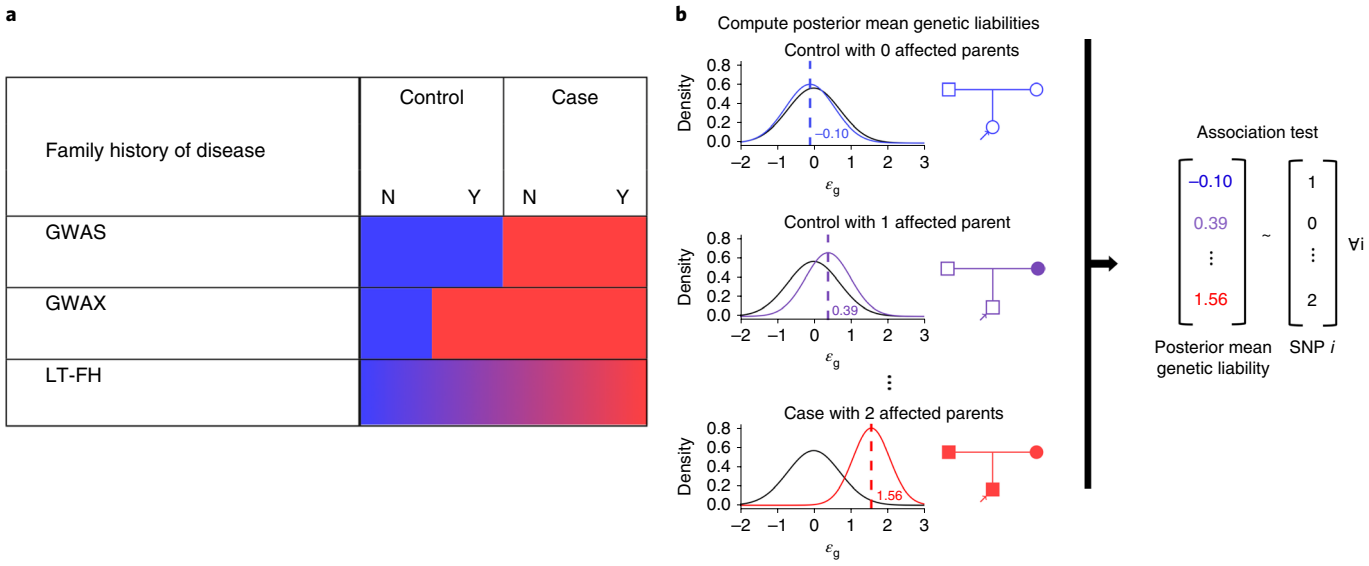


Fig. 1 | Overview of LT-FH and other methods. a, GWAS uses binary case–control status, ignoring family history; GWAX uses binary proxy case–control status, merging controls who have a family history of disease with disease cases; and LT-FH uses continuous-valued posterior mean genetic liability, appropriately differentiating all case–control and family history configurations. **b**, LT-FH computes posterior mean genetic liabilities (left) and then tests for an association between genotype and posterior mean genetic liability (right).

for a case with two affected parents versus 0.39 for a control with one affected parent, but GWAX treats these individuals identically. In contrast, LT-FH appropriately differentiates all case–control and family history configurations.

Simulations. We simulated genotypes at 100,000 unlinked SNPs and case–control status and family history (parental history for both parents) for 100,000 unrelated target samples, excluding sibling history. We simulated genotypes for both parents and used these to simulate genotypes for target samples (offspring), and we simulated case–control status for both parents and target samples using a liability threshold model; target samples were not ascertained for case–control status. Our default parameter settings involved 500 causal SNPs that explained $h^2 = 50\%$ of variance in liability, disease prevalence $K = 5\%$ (implying liability threshold $T = 1.64$ and observed-scale $h^2 = 11\%$) and accurate specification of h^2 and K to the LT-FH method. Other parameter settings were also explored. We compared three methods: GWAS, GWAX and LT-FH (see Methods for further details of the simulation framework). We note that simulations using real linkage disequilibrium (LD) patterns are essential for methods impacted by LD between SNPs; however, the LT-FH method (like GWAS and GWAX) is not impacted by LD between SNPs because genotype data are not used to compute posterior mean genetic liabilities; thus, LD between focal SNPs and other SNPs cannot impact the power of LT-FH at a focal SNP (aside from the fixed loss of information due to incomplete tagging of causal SNPs). We further note that simulations with LD using a subset of individuals from the UK Biobank would not be feasible, as simulations of family history require genotypes of both target samples and relatives (to simulate the case–control status of both target samples and relatives), but genotypes of relatives are not available for most UK Biobank samples.

We assessed the calibration of each method using average χ^2 statistics at null SNPs (Fig. 2a and Supplementary Table 1). All methods were correctly calibrated, with an average χ^2 of 1.00 for GWAS, GWAX and LT-FH. Accordingly, false-positive rates at various α levels (5×10^{-2} to 5×10^{-6}) matched the corresponding α level, as indicated by quantile–quantile plots for null SNPs (Supplementary Table 2 and Extended Data Fig. 1).

We assessed the association power of each method using both average χ^2 statistics at causal SNPs and formal power calculations (proportion of causal SNPs with $P < 5\times10^{-8}$; Fig. 2b,c and Supplementary Table 1). LT-FH was the most powerful method, with a 22% increase in average χ^2 and a 55% increase in power as compared to GWAX, which outperformed GWAS at default parameter settings.

We performed 14 secondary analyses (Supplementary Note, Supplementary Tables 1–13 and Extended Data Fig. 1). Our main findings were that (1) LT-FH outperformed both GWAX and GWAS, regardless of the underlying disease prevalence; (2) the second best method after LT-FH was GWAX for lower-prevalence diseases and GWAS for higher-prevalence diseases; (3) the impact of a mis-specified value of disease prevalence or liability-scale heritability on the association power of LT-FH was negligible; (4) shared environment led to slightly smaller improvements for LT-FH as compared to GWAS; and (5) family history reporting bias led to smaller improvements for LT-FH as compared to GWAS but no increase in type I error. We caution that the generative model in all of our simulations was the same as the liability threshold model that we used for inference, and thus these simulations should be viewed as a best-case scenario for LT-FH.

Analysis of 12 diseases in the UK Biobank. We analyzed 12 diseases in the UK Biobank[20] with genotype and case–control and/or family history data available for up to 381,493 unrelated individuals of European ancestry (20 million imputed SNPs with minor allele frequency (MAF) > 0.1%), and family history of disease available for most individuals (Methods, Table 1 and Supplementary Tables 14 and 15). Disease prevalence ranged from 0.001 for Alzheimer's disease to 0.32 for hypertension; for lower-prevalence diseases, we applied stricter MAF thresholds to avoid type I error in unbalanced case–control settings[21] (Supplementary Tables 16 and 17). Family history information, including parental (presence or absence of disease in each respective parent) and sibling (number of brothers and sisters and the presence or absence of disease in the set of all siblings) history was available for each disease. On average, across the 12 diseases, most of the target samples had complete parental and sibling history (88% and 93%, respectively; Supplementary

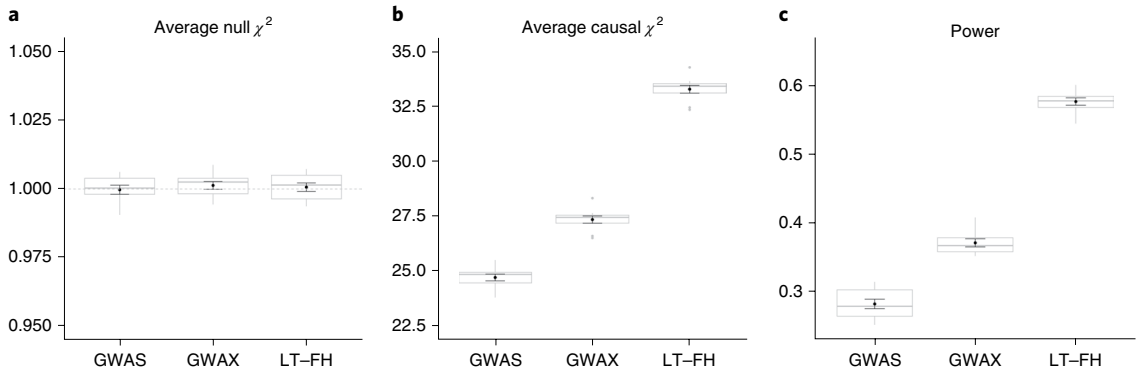


Fig. 2 | LT-FH is well calibrated and increases association power in simulations. a, Distribution of average χ^2 values for null SNPs. The dashed gray line shows the expected null value of 1. **b**, Distribution of average χ^2 values for causal SNPs. **c**, Distribution of power, defined as the proportion of causal SNPs with $P < 5 \times 10^{-8}$. Each gray boxplot represents estimates from ten simulations, with each simulation consisting of 100,000 SNPs (500 causal SNPs). The center line denotes the median; the lower and upper hinges correspond to the first and third quartiles, respectively. Whiskers extend to the minimum and maximum estimates located within 1.5 × interquartile range (IQR) from the lower and upper hinge, respectively. The black points and error bars represent the mean and ±1 s.e. Numerical results are reported in Supplementary Table 1.

Table 1 | Overview of 12 diseases in the UK Biobank

Trait	Prevalence	Parental prevalence	$\hat{h}_g^2$	s.e.($\hat{h}_g^2$)	N_{GWAS}	N_{GWAX}	$N_{LT\text{-}FH}$
AD	0.001	0.065	0.12	0.1103	381,493	324,512	381,493
PD	0.003	0.020	0.15	0.0504	381,493	318,792	381,493
Lung cancer	0.006	0.064	0.10	0.0304	381,493	323,838	381,493
Bowel cancer	0.013	0.054	0.14	0.0176	381,493	322,887	381,493
Stroke	0.023	0.144	0.08	0.0112	381,493	332,468	381,493
COPD	0.035	0.082	0.17	0.0092	381,493	329,495	381,493
Prostate cancer*	0.037	0.075	0.30	0.0189	175,450	331,458	368,940
T2D	0.042	0.092	0.36	0.0093	380,180	330,267	381,390
Breast cancer*	0.061	0.082	0.20	0.0118	206,043	348,170	371,064
Depression	0.073	0.052	0.11	0.0058	381,493	328,186	381,493
CAD	0.083	0.255	0.20	0.0058	381,493	341,810	381,493
HTN	0.318	0.260	0.31	0.0037	381,493	354,208	381,493

The UK Biobank prevalence, parental prevalence, liability-scale SNP heritability and number of samples (*N*) analyzed by GWAS, GWAX and LT-FH are reported. GWAX sample sizes are lower than GWAS and LT-FH sample sizes owing to incomplete family history data. An asterisk denotes sex-specific diseases (for which GWAX and LT-FH incorporate family history data from samples of both sexes). Estimates of liability-scale narrow-sense heritability (h^2) from the literature, which are used by the LT-FH method, are reported in Supplementary Table 15. Diseases are listed in order of disease prevalence. AD, Alzheimer's disease/dementia; PD, Parkinson's disease; COPD, chronic bronchitis/emphysema; T2D, type 2 diabetes; CAD, coronary artery disease; HTN, hypertension.

Table 18); the accuracy of family history information is assessed below. We defined GWAS, GWAX and LT-FH phenotypes accordingly (Methods, Supplementary Table 19 and Extended Data Fig. 2) and compared these three methods. We included 20 principal components (PCs), assessment center, genotype array, sex, age and age squared as covariates in association analyses for each method. In this dataset, the computational cost of computing LT-FH phenotypes (posterior mean genetic liabilities) was one to two orders of magnitude lower than the cost of computing association statistics (Supplementary Table 20).

We assessed the calibration of GWAS, GWAX and LT-FH using stratified LD score regression (S-LDSC) attenuation ratio[15,22–24], defined as (S-LDSC intercept − 1)/(mean χ^2 − 1) (Methods and Supplementary Table 21a). Attenuation ratios were similar for each method, with averages across 12 diseases of 0.18, 0.16 and 0.15, respectively (and inverse-variance-weighted averages close to 0.1, consistent with ref. [15]); we note that attenuation ratios slightly above 0 may be caused by attenuation bias[15]. Although there is no guarantee that analyses of UK Biobank data using PC covariates are completely devoid of confounding[25], these results confirm that LT-FH is well calibrated and there is no confounding specific to the LT-FH method.

We assessed the association power of each method by computing the number of independent genome-wide-significant loci identified for each disease, defined similarly to those in ref. [15] (Methods, Fig. 3a and Supplementary Table 21b); we computed the s.e. of relative improvements via block jackknife (Methods). Across all diseases, LT-FH was 63% (s.e. 6%) more powerful than GWAS and 36% (s.e. 4%) more powerful than the trait-specific maximum of GWAS and GWAX (for example, 690 independent loci for LT-FH versus 423 for GWAS). For the eight lower-prevalence diseases ($K < 5\%$), LT-FH was 39% (s.e. 7%) more powerful than the trait-specific maximum of GWAS and GWAX, and GWAX was more powerful than GWAS, consistent with simulations at lower prevalence. For the four higher-prevalence diseases ($K > 5\%$), LT-FH was 35% (s.e. 4%) more powerful than the trait-specific maximum of GWAS and GWAX, and GWAS was more powerful than GWAX, consistent with simulations at higher prevalence; GWAS versus

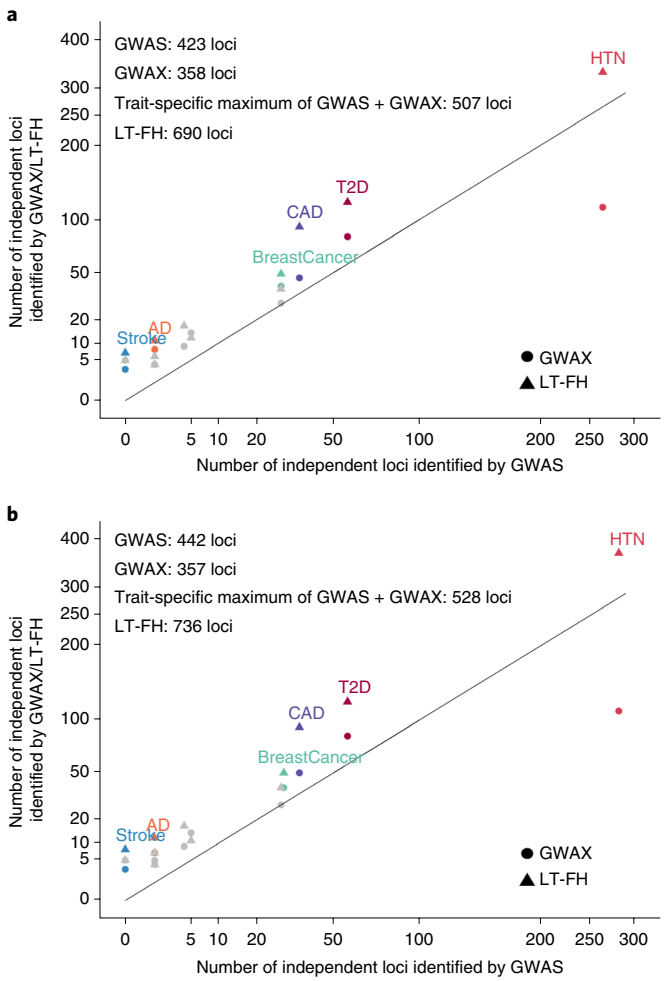


Fig. 3 | LT-FH increases association power across 12 diseases from the UK Biobank. a,**b**, We report results of GWAS, GWAX and LT-FH using either linear regression (**a**) or BOLT-LMM (**b**) on unrelated European individuals. Numerical results are reported in Supplementary Tables 21 and 35.

GWAX results were dominated by hypertension, the disease with the highest prevalence. We also evaluated the association power of each method using relative effective sample size[15]; the average relative effective sample size was 1.31 for LT-FH versus GWAS and 1.27 for LT-FH versus the trait-specific maximum of GWAS and GWAX (Supplementary Table 22). Finally, we assessed the association power of each method across all 12 diseases by computing the average χ^2 across all genome-wide-significant SNPs identified by any method and the average χ^2 across all SNPs (Supplementary Table 21c,d); we determined that these quantities were much larger for LT-FH than for the other methods.

We assessed whether GWAS, GWAX and LT-FH phenotypes reflect the same underlying genetic architectures by estimating pairwise genetic correlations between these phenotypes with BOLT-REML[26] (Table 2). Genetic correlations were much higher than the corresponding phenotypic correlations (for example, correlations between LT-FH and GWAS were 0.97 and 0.70, respectively, averaged across 12 diseases), indicating that incorporating family history preserves the same underlying genetic architecture but provides substantial independent information. Phenotypic correlations for LT-FH (or GWAX) versus GWAS were particularly low for low-prevalence diseases, for which family history is more informative than case–control status. We also computed correlations between $-\log_{10}$ P values, which mirrored the phenotypic correlations (Supplementary Table 23). LT-FH phenotypes had higher total genetic signal (sample size multiplied by observed-scale SNP heritability[23]) than GWAS phenotypes for all 12 diseases (59% higher on average; Supplementary Table 24). These findings support the use of the LT-FH method to increase association power.

We performed a replication analysis using independent data to assess whether the new associations identified by LT-FH were genuine. We analyzed four diseases—coronary artery disease (CAD), type 2 diabetes (T2D), breast cancer and prostate cancer—with publicly available summary statistics for case–control data independent from the UK Biobank (Methods). For both genome-wide-significant loci identified by GWAS and genome-wide-significant loci identified by LT-FH, we computed the replication slope (the slope of a regression of standardized effect sizes in case–control replication data versus UK Biobank discovery data[27]); association statistics for case–control replication data were always computed using GWAS. The replication slopes were similar for GWAS (slope = 0.81, s.e. = 0.02; 124 loci) and LT-FH (slope = 0.79, s.e. = 0.01; 243 loci; Fig. 4 and Supplementary Table 25). In addition, the replication slopes were similar for loci detected only by GWAS (slope = 0.67, s.e. = 0.06; 7 loci) and only by LT-FH (slope = 0.69, s.e. = 0.03; 126 loci); as expected, these slopes were lower than 0.81 and 0.79 because loci detected by only one method are more susceptible to winner's curse. These results indicate that the new loci identified by LT-FH are genuine associations, as assessed by the GWAS method in replication data.

We performed six secondary analyses (Supplementary Note, Supplementary Tables 27–34 and Extended Data Fig. 3). Our main findings showed that using only parental history was slightly less powerful than using parental and sibling history; that downweighting family history information based on its accuracy had little impact on association results; that explicitly accounting for age when computing posterior mean genetic liabilities had little impact on association results; and that our Monte Carlo LT-FH method attained higher power than the analytical PA formula when using parental and sibling history. In summary, LT-FH was well calibrated and more powerful than all other methods in primary and secondary analyses.

BOLT-LMM increases GWAS, GWAX and LT-FH association power. A method for efficient Bayesian mixed-model analysis, BOLT-LMM, was shown to increase association power (for GWAS phenotypes) as compared to linear regression[14,15]. We thus investigated the application of BOLT-LMM to GWAS, GWAX and LT-FH phenotypes, analyzing the same data for the 12 UK Biobank diseases as above. We first confirmed that GWAS, GWAX and LT-FH remained well calibrated (Supplementary Table 35a), and then we assessed association power (Fig. 3b and Supplementary Table 35b). Across all diseases, LT-FH was 67% (s.e. 6%) more powerful than GWAS and 39% (s.e. 4%) more powerful than the trait-specific maximum of GWAS and GWAX (for example, 736 independent loci for LT-FH versus 442 for GWAS), implying similar relative power among the three methods when using BOLT-LMM; the absolute increases in power attained by BOLT-LMM were modest (Fig. 3b versus Fig. 3a) and smaller than those in analyses of highly heritable UK Biobank traits[15] because disease traits have lower observed-scale SNP heritability (particularly for lower-prevalence diseases) and the increase in power attained by BOLT-LMM scales with SNP heritability[15]. Analyses of average χ^2 across all genome-wide-significant SNPs identified by any method (Supplementary Table 35c) and average χ^2 across all SNPs (Supplementary Table 35d) yielded similar conclusions.

Our above analyses using both linear regression and BOLT-LMM were restricted to unrelated target samples (up to 381,493 unrelated individuals of European ancestry), consistent with previous studies leveraging family history using GWAX[1,7,8]. We investigated the consequences of including related individuals in BOLT-LMM analyses. We first applied BOLT-LMM to GWAS, GWAX and LT-FH

Table 2 | Genetic and phenotypic correlations between GWAS, GWAX and LT-FH phenotypes

Trait	Genetic correlation (r_G)						Phenotypic correlation		
	GWAS–GWAX		GWAS–LT-FH		GWAX–LT-FH		GWAS–GWAX	GWAS–LT-FH	GWAX–LT-FH
	r_G	s.e.	r_G	s.e.	r_G	s.e.			
AD	1.00	0.38	1.00	0.34	1.00	0.002	0.09	0.30	0.96
PD	1.00	0.16	1.00	0.10	1.00	0.009	0.26	0.57	0.93
Lung cancer	1.00	0.14	1.00	0.12	1.00	0.004	0.20	0.49	0.91
Bowel cancer	0.95	0.06	0.97	0.03	0.99	0.006	0.30	0.64	0.91
Stroke	0.83	0.07	0.88	0.04	0.99	0.008	0.24	0.62	0.86
COPD	0.90	0.02	0.95	0.01	0.99	0.003	0.40	0.77	0.87
Prostate cancer	1.00	0.03	1.00	0.006	1.00	0.005	0.56	0.88	0.89
T2D	0.97	0.01	0.99	0.004	0.99	0.002	0.38	0.77	0.86
Breast cancer	1.00	0.02	1.00	0.006	1.00	0.004	0.58	0.87	0.89
Depression	0.88	0.02	0.93	0.007	0.98	0.003	0.58	0.84	0.89
CAD	0.89	0.02	0.96	0.005	0.98	0.005	0.31	0.76	0.78
HTN	0.95	0.006	0.98	0.001	0.98	0.003	0.52	0.86	0.75
Average	0.95		0.97		0.99		0.37	0.70	0.88

We report the genetic correlation (estimated using BOLT-REML) and phenotypic correlation between each pair of GWAS, GWAX and LT-FH phenotypes. BOLT-REML estimates were constrained to be ≤1.00. Estimates of genetic correlation are reported with s.e.; the s.e. of phenotypic correlation estimates was ≤0.003 in each case.

phenotypes of target samples without applying any filter for relatedness (up to 459,256 related individuals of European ancestry). We determined that GWAX and LT-FH suffered poor calibration (average S-LDSC attenuation ratios equal to 0.26 for GWAX and 0.22 for LT-FH versus 0.14 for GWAS; Supplementary Table 36a), rendering moot the increases in power due to the larger sample size (Supplementary Table 36b). A likely explanation for the poor calibration is the extreme concordance between sibling pair phenotypes (almost 100% for GWAX) arising from nearly identical family histories. Our results indicate that BOLT-LMM is unable to correct for inflation in test statistics in such extreme cases, as it does not have the flexibility to model excess phenotypic concordance for sibling pairs specifically. Thus, BOLT-LMM analysis of GWAX and LT-FH phenotypes should not be applied to target samples without any filter for relatedness.

We therefore applied BOLT-LMM to LT-FH phenotypes for all target samples (up to 459,256 related individuals of European ancestry) that were modified to incorporate only case–control status (ignoring family history information) for each individual from a sibling or parent–offspring pair within the set of target samples (see secondary analysis below). We determined that this approach was well calibrated, with an average S-LDSC attenuation ratio of 0.14 for LT-FH versus 0.14 for GWAS (inverse-variance-weighted average close to 0.1, consistent with ref. [15]; Supplementary Table 37a). We compared the power of this approach to applying BOLT-LMM to GWAS for all related Europeans and GWAX for all unrelated Europeans, as previously recommended[1,7,8] (see above). Across all diseases, LT-FH was 54% (s.e. 5%) more powerful than GWAS and 44% (s.e. 4%) more powerful than the trait-specific maximum of GWAS and GWAX (for example, 908 independent loci for LT-FH versus 590 for GWAS; Extended Data Fig. 4 and Supplementary Table 37b), similar to our previous analyses (Fig. 3). Thus, BOLT-LMM analysis of LT-FH phenotypes defined using only case–control status for all sibling pairs and parent–offspring pairs within the set of target samples is recommended, and we have publicly released LT-FH association statistics computed using this approach (see Data availability).

We performed a secondary analysis in which we instead applied BOLT-LMM to LT-FH phenotypes for all target samples (up to 459,256 related individuals of European ancestry) that were modified to incorporate family history information for one sibling within each set of siblings in the target samples (with no filter on family history information for parent–offspring pairs). We note that this approach may suffer from residual concordance between sibling pair phenotypes (and parent–offspring phenotypes) arising from redundancy between the family history of one target sample and the case–control status of the other target sample. Indeed, we determined that LT-FH phenotypes defined in this way were slightly but significantly miscalibrated (average S-LDSC attenuation ratio equal to 0.15 for LT-FH versus 0.14 for the recommended LT-FH approach; $P < 10^{-6}$ for the difference in inverse-variance-weighted means), rendering moot the slightly higher power of this approach (Supplementary Table 38a,b).

In summary, BOLT-LMM analysis increases the power of LT-FH, GWAS and GWAX, but care is required to avoid poor calibration of LT-FH and GWAX due to extreme concordance in sibling pair phenotypes. Our recommendation is for BOLT-LMM analysis of LT-FH phenotypes defined using only case–control status for all sibling pairs and parent–offspring pairs within the set of target samples.

Discussion

We have introduced a new method, LT-FH, that accurately models a broad range of case–control status and family history configurations to increase association power as compared to that of the prevailing methods[1]. In particular, LT-FH differs from other methods, including methods developed in the livestock genetics community[4–6], by allowing for complex modes of reporting family history (for example, binary response for sibling history, that is, reporting that at least one sibling has the disease), instead of requiring phenotypes for each relative. Across 12 diseases from the UK Biobank, LT-FH was 63% (s.e. 6%) more powerful than GWAS and 36% (s.e. 4%) more powerful than the trait-specific maximum of GWAS and GWAX, while maintaining correct calibration. LT-FH attained similar increases in power when using BOLT-LMM to compute association statistics for all three methods and when incorporating related individuals. These findings provide a strong motivation to apply LT-FH to genetic datasets for which family history information is available and to collect family history information in future genetic studies.

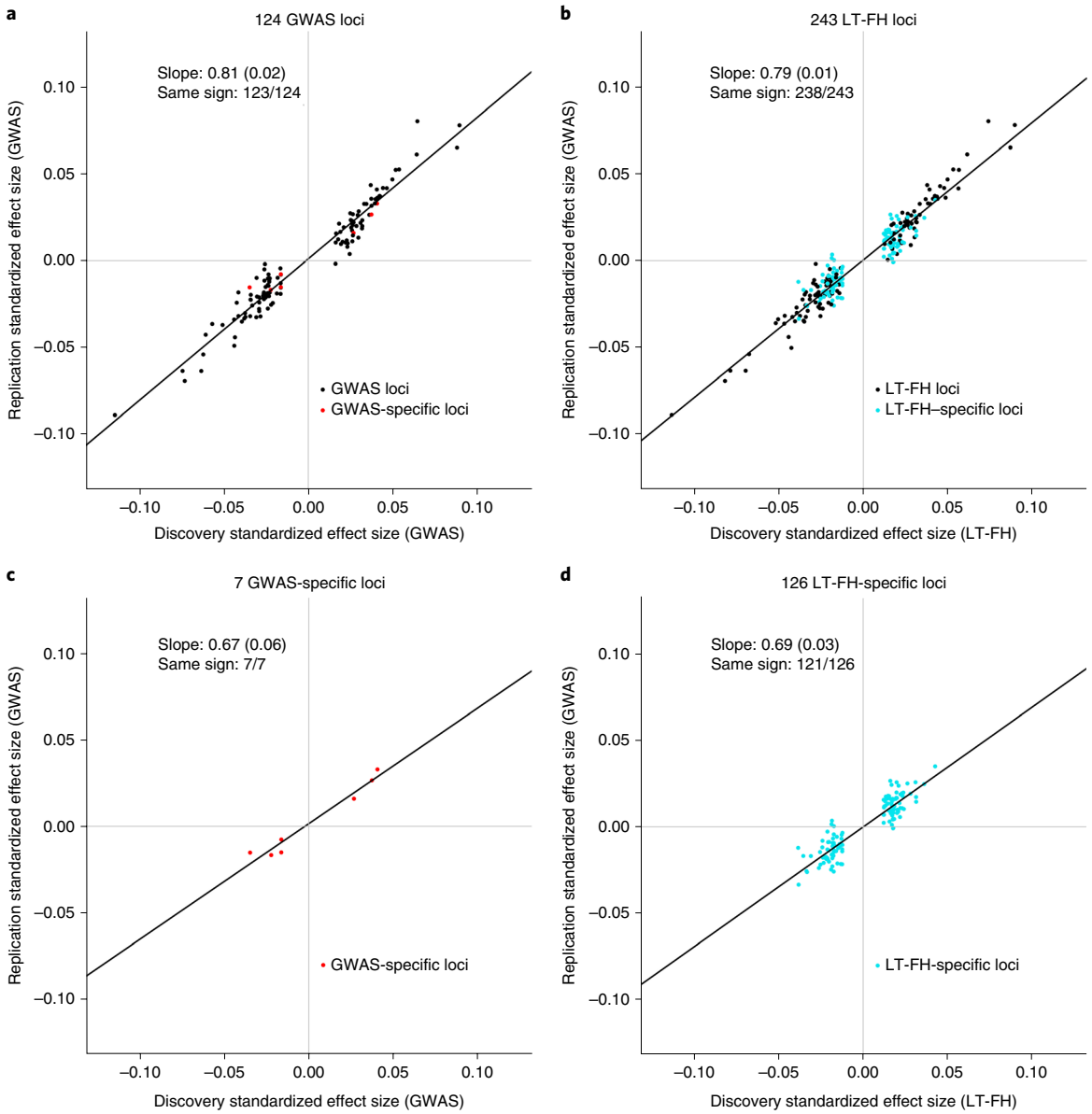


Fig. 4 | Loci identified by LT-FH replicate in independent datasets. a–d, We plotted standardized effect sizes ($Z/\sqrt{N_{eff}}$) in the non-UK Biobank replication data (average N_{eff} = 99,000 for GWAS) versus UK Biobank discovery data (average N_{eff} = 62,000 for GWAS, N_{eff} = 102,000 for LT-FH), aggregated across four diseases (CAD, T2D, breast cancer and prostate cancer) for the 124 loci identified by GWAS (**a**), the 243 loci identified by LT-FH (**b**), the 7 loci identified by GWAS but not LT-FH (**c**) and the 126 loci identified by LT-FH but not GWAS (**d**). Numerical results are reported in Supplementary Table 25.

Although LT-FH greatly increases association power, it has several limitations. First, self-reported family history information may be inaccurate. Indeed, we determined that the sibling concordance of self-reported family history was substantially lower than 100%; however, our efforts to account for this had little impact on association power (Supplementary Tables 28 and 29). Second, family history may reflect a different underlying genetic architecture than case–control status, for example, due to differences in the etiology of early-onset versus late-onset disease or differences in diagnostic criteria over time; however, we observed high genetic correlations between GWAS and LT-FH phenotypes (average of 0.97 across diseases; Table 2). Third, real diseases may not adhere to the liability threshold model used by LT-FH, which may explain why increases in power observed in analyses of real diseases were not as large as those in simulations (Figs. 2c and 3). Moreover, posterior mean genetic liabilities inferred from family history information may be biased by unmodeled effects. This might reduce the power of the LT-FH method but would not cause its test for additive association with genetic variants to produce false positives (Supplementary Tables 7 and 8). Fourth, per standard GWAS analysis, LT-FH may require stricter MAF thresholds than our default value of 0.001 to avoid type I error in unbalanced case–control settings[21]. Although the number of diseases requiring stricter MAF thresholds was reduced from eight for standard GWAS analysis to two for LT-FH (due to lower kurtosis; Supplementary Table 16), more sophisticated approaches (generalizing ref. [21]) could be explored. Fifth, just like GWAX, LT-FH should not be applied to target samples without any filter for relatedness because extreme concordance in sibling pair phenotypes leads to miscalibration—even when applying BOLT-LMM, an efficient mixed-model method that accounts for phenotypic concordance among related individuals. Specifically, GWAS analysis of related target samples can induce miscalibration because target samples have correlated genotypes and correlated phenotypes. Mixed-model methods, such as BOLT-LMM,

are designed to correct this miscalibration. However, if related target samples have extremely concordant phenotypes due to inclusion of family history information, BOLT-LMM may fail to correct this miscalibration (Supplementary Table 36). Thus, scenarios of extreme concordance between phenotypes of related target samples should be avoided. We instead recommend applying BOLT-LMM to LT-FH phenotypes defined using only case–control status for all sibling pairs and parent–offspring pairs within the set of target samples. This approach is straightforward to implement and still greatly increases association power while maintaining correct calibration. Sixth, LT-FH does not currently incorporate family history for genetically correlated diseases, which could be highly informative; however, LT-FH association results for two or more genetically correlated diseases could be jointly analyzed using existing methods[28]. Seventh, the new genome-wide-significant associations detected by LT-FH explain less variance on average, such that the increase in variance explained by LT-FH compared to other methods is lower than the increase in the number of genome-wide-significant loci. However, small genetic effects can still produce important biological insights[29,30]. Eighth, posterior mean genetic liabilities can be computed using an analytical approach, the PA formula[2,17,18], instead of LT-FH. However, the PA formula is not well suited to data that includes sibling history as a binary response (that is, at least one sibling has the disease). Indeed, we confirmed that LT-FH attains higher power than the PA formula in simulations with sibling history (Supplementary Table 11) and in analyses of UK Biobank diseases (Supplementary Table 34). Finally, we have not explored the use of family history to improve the accuracy of genetic risk prediction[2], which remains as a direction for future research. Despite these limitations, we anticipate that LT-FH will attain large increases in power in future association studies.

Online content

Any methods, additional references, reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-020-0613-6.

Received: 25 July 2019; Accepted: 12 March 2020;
Published online: 20 April 2020

References

1. Liu, J. Z., Erlich, Y. & Pickrell, J. K. Case–control association mapping by proxy using family history of disease. *Nat. Genet.* **49**, 325–331 (2017).
2. So, H.-C., Kwan, J. S. H., Cherny, S. S. & Sham, P. C. Risk prediction of complex diseases from family history and known susceptibility loci, with applications for cancer screening. *Am. J. Hum. Genet.* **88**, 548–565 (2011).
3. Visscher, P. M. & Duffy, D. L. The value of relatives with phenotypes but missing genotypes in association studies for quantitative traits. *Genet. Epidemiol.* **30**, 30–36 (2006).
4. Hayes, B. J., Bowman, P. J., Chamberlain, A. J. & Goddard, M. E. Genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci.* **92**, 433–443 (2008).
5. Misztal, I., Legarra, A. & Aguilar, I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* **92**, 4648–4655 (2009).
6. Liu, Z., Goddard, M. E., Reinhardt, F. & Reents, R. A single-step genomic model with direct estimation of marker effects. *J. Dairy Sci.* **97**, 5833–5850 (2014).
7. Marioni, R. E. et al. GWAS on family history of Alzheimer's disease. *Transl. Psychiatry* **8**, 99 (2018).
8. Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
9. Falconer, D. S. The inheritance of liability to diseases with variable age of onset, with particular reference to diabetes mellitus. *Ann. Hum. Genet.* **31**, 1–20 (1967).
10. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* **88**, 294–305 (2011).
11. Zaitlen, N. et al. Informed conditioning on clinical covariates increases power in case–control association studies. *PLoS Genet.* **8**, e1003032 (2012).
12. Weissbrod, O., Lippert, C., Geiger, D. & Heckerman, D. Accurate liability estimation improves power in ascertained case–control studies. *Nat. Methods* **12**, 332–334 (2015).
13. Hayeck, T. J. et al. Mixed model with correction for case–control ascertainment increases association power. *Am. J. Hum. Genet.* **96**, 720–730 (2015).
14. Loh, P.-R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
15. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nat. Genet.* **50**, 906–911 (2018).
16. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
17. Pearson, K. Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **200**, 1–66 (1903).
18. Aitken, A. C. Note on selection from a multivariate normal population. *Proc. Edinb. Math. Soc. B* **4**, 106–110 (1934).
19. Armitage, P. Tests for linear trends in proportions and frequencies. *Biometrics* **11**, 375–386 (1955).
20. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
21. Zhou, W. et al. Efficiently controlling for case–control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
22. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
23. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
24. Gazal, S. et al. Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. *Nat. Genet.* **49**, 1421–1427 (2017).
25. Haworth, S. et al. Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nat. Commun.* **10**, 333 (2019).
26. Loh, P.-R. et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance components analysis. *Nat. Genet.* **47**, 1385–1392 (2015).
27. Marigorta, U. & Navarro, A. High trans-ethnic replicability of GWAS results implies common causal variants. *PLoS Genet.* **9**, e1003566 (2013).
28. Turley, P. et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).
29. Price, A. L., Spencer, C. C. A. & Donnelly, P. Progress and promise in understanding the genetic basis of common diseases. *Proc. Biol. Sci.* **282**, 20151684 (2015).
30. Visscher, P. M. et al. 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).

Methods

LT-FH method: estimating posterior mean genetic liability. The liability threshold model can be written as $\epsilon = \boldsymbol{X\beta} + \epsilon_e$, where ϵ is the liability, $\boldsymbol{X}$ is the genotype at candidate SNPs (normalized to mean 0 and variance 1), $\boldsymbol{\beta}$ are the effect sizes of SNPs on the liability scale, and an individual is a case ($z=1$) if and only if $\epsilon \geq T$ and is a control otherwise ($z=0$). T determines the disease prevalence ($\Phi(T) = P(x \geq T)$ where $x \sim N(0,1)$). We further assume that $\epsilon_g = \boldsymbol{X\beta} \sim N(0, h^2)$ and, when testing a SNP g for association, the effect size is small enough that the liability is $\beta g + \epsilon = \beta g + \epsilon_g + \epsilon_e$, where $\epsilon \sim N(0,1)$, $\epsilon_g \sim N(0, h^2)$ and $\epsilon_e \sim N(0, 1-h^2)$.

We assume a multivariate normal distribution for ϵ: for two individuals $\mathrm{Cov}(\epsilon_1, \epsilon_2) = K_{12}h^2$, where K_{12} is the coefficient of the relationship for the pair of individuals, for example, 1/2 for a parent–offspring pair. We assume that we have the genotype and case status of an individual and potentially the disease status of family members. For example, to incorporate parental history and sibling history, we assume

$$\begin{pmatrix} \epsilon_{o,e} \\ \epsilon_{o,g} \\ \epsilon_{p1} \\ \epsilon_{p2} \\ \epsilon_{s} \end{pmatrix} \sim \mathrm{MVN}_5 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1-h^2 & 0 & 0 & 0 & 0 \\ 0 & h^2 & 0.5h^2 & 0.5h^2 & 0.5h^2 \\ 0 & 0.5h^2 & 1 & 0 & 0.5h^2 \\ 0 & 0.5h^2 & 0 & 1 & 0.5h^2 \\ 0 & 0.5h^2 & 0.5h^2 & 0.5h^2 & 1 \end{pmatrix} \right)$$

where $\epsilon_{o,e}$ and $\epsilon_{o,g}$ are the environmental and genetic components of the liability of the target sample (offspring), ϵ_{p1} and ϵ_{p2} are the liabilities of the parents and ϵ_s is the liability of the sibling(s) (for simplicity, we include only one sibling, but this can be extended to an arbitrary number of siblings). We estimate the posterior mean genetic liability $E\left[\epsilon_{o,g} | z_o, z_{p1}, z_{p2}, \overrightarrow{z_s}\right]$ for each individual given the case–control status of the genotyped individual (z_o), both parents (z_{p1}, z_{p2}), and sibling(s) ($\overrightarrow{z_s}$). When disease status is unknown for subsets of these individuals, we condition only on known information. We are interested in the genetic signal and therefore we estimate the posterior mean genetic liability, rather than the posterior mean liability; genetic association power is increased by integrating out the noise contributed by the environmental component of the liability.

We estimate $E\left[\epsilon_{o,g} | z_o, z_{p1}, z_{p2}, z_s\right]$ using Monte Carlo integration. Briefly, (1) we sample from a multivariate normal with the given covariance structure and compute the posterior mean genetic liability of the offspring in all possible configurations of case–control status and family history (using liabilities that fall above or below a given threshold to determine case–control status and family history). If the posterior mean genetic liabilities in any configuration has an s.e.m. above 0.01, we resample from a conditional multivariate normal to ensure that we sample in a manner representative of the tails of the distribution appropriately. (2) We sample relevant liabilities conditional on the offspring's case–control status (z_o) or conditional on the parents (that is, $z_{p1} + z_{p2} = \{0,1,2\}$). (3) We then combine all samples of $\epsilon_{o,g}$ from given configurations. (4) We compute the posterior mean genetic liabilities and assess the s.e. of each posterior mean. We then repeat 2–4 until the s.e.m. is less than 0.01 for all configurations (Supplementary Note and Supplementary Table 39). We note that family history reporting bias may reduce the power of LT-FH, but would not lead to false positives (Supplementary Table 8). In addition, distinct from family history, non-random missingness of target sample phenotypes can equally impact GWAS and LT-FH. Specifically, LT-FH is identical to GWAS in the absence of family history information, because in this setting both methods exclude individuals with missing case–control status.

We considered the PA formula as an approximate analytical approach and implemented this approximation for LT-FH$_{\text{no-sib}}$ (refs. [2,17,18]; Supplementary Note and Supplementary Table 32). However, in UK Biobank data, sibling history is provided as a binary response, that is, reporting that at least one sibling has the disease. This 'at least one' condition is complex to incorporate using analytical approaches but straightforward to incorporate using Monte Carlo integration. For a given number of siblings (S), we sampled S liabilities ($\epsilon_{s1}, \epsilon_{s2}, \ldots, \epsilon_{sS}$) and can easily assess the presence or absence of disease in the set of all siblings by comparing each sibling's liability to the relevant threshold; the absence of disease corresponds to $\epsilon_{si} < T$ for all $i \in \{1,2,\ldots,S\}$ whereas the presence of disease corresponds to at least one i with $\epsilon_{si} \geq T$. Thus, we estimate posterior mean genetic liabilities using Monte Carlo integration in our main analyses and open-source software.

LT-FH method: LT-FH association statistic. The LT-FH association statistic is a measure of association between genotype g and posterior mean genetic liability across samples. We can compute this statistic either using linear regression or using other methods such as BOLT-LMM[14,15] while incorporating additional covariates such as PCs. If we treat the posterior mean genetic liability as a continuous variable and compute the number of samples multiplied by the squared correlation between g and the posterior mean genetic liability (generalizing the Armitage trend test[19]), this is equivalent to the score test (Supplementary Note).

LT-FH effect sizes. Although LT-FH estimates the posterior mean genetic liability, the raw effect sizes computed using LT-FH are not on the liability scale. However, the raw effect sizes computed using linear regression can be transformed to the observed scale by making use of the effective sample size[15]. In detail, to obtain per-allele observed-scale effect sizes for a non-standardized phenotype (as is computed by BOLT-LMM software), we compute:

$$\hat{\beta}_{\mathrm{LT-FH,obs}} = \frac{\hat{\beta}_{\mathrm{LT-FH}}}{\mathrm{s.e.}\left(\hat{\beta}_{\mathrm{LT-FH}}\right)\sqrt{c \times N_{\mathrm{GWAS}}}} \times \sqrt{\frac{K(1-K)}{2 \times \mathrm{MAF} \times (1-\mathrm{MAF})}}$$

where $\hat{\beta}_{\mathrm{LT-FH}}$ is the raw per-allele effect size, c is the relative effective sample size for LT-FH versus GWAS, and K is the disease prevalence. We note that it is straightforward to use MAF to convert between per-allele and per-standardized-genotype effect sizes, distinct from converting between liability and observed scales. Further details are provided in the Supplementary Note.

GWAS and GWAX methods. GWAS aims to discover associations between variants and disease by comparing cases to controls of a disease. GWAX compares cases and proxy cases (controls with a family history of the disease) to controls with no family history of disease. GWAX can be particularly valuable when the case–control status of the genotyped individual is unknown; however, when the case–control status of the genotyped individual is known, incorporating this information increases power[1]. Because of this, we incorporated the case–control status of the genotyped individual (if known) when computing GWAX association statistics. For both GWAS and GWAX, association statistics can be computed using methods such as linear regression, logistic regression, BOLT-LMM[15] or SAIGE[21].

For the GWAX-2df test, we performed an F-test, regressing genotype on case status and control with family history status (two binary variables) and testing whether the coefficients associated with these two indicator variables were both 0 (Supplementary Note). When no covariates were included, the results were virtually identical to a Pearson-χ^2 test on a 2×3 table.

Simulations. We simulated genotypes at 100,000 unlinked SNPs and case–control status and family history (parental history for both parents) for 100,000 unrelated target samples; we did not include sibling history in these simulations. We simulated genotypes for both parents and used these to simulate genotypes for target samples (offspring); the underlying MAF was ~U(0.01,0.5). Using these genotypes, we simulated case–control status for both parents and target samples using a liability threshold model in which we normalized genotype to a mean of 0 and variance of 1 according to the true MAF. We set the effect sizes for the C causal SNPs on the liability scale to be $\sqrt{h^2/C}$ and added environmental noise with a variance of $1-h^2$. Target samples were not ascertained for case–control status. Our default parameter settings involved 500 causal SNPs explaining $h^2 = 50\%$ of the variance in liability, disease prevalence $K = 5\%$ (implying liability threshold $T = 1.64$ and observed-scale $h^2 = 11\%$) and accurate specification of h^2 and K to the LT-FH method. For each parameter setting, we computed ten simulation replicates in which there was a different underlying MAF distribution for each simulation replicate. Further details are provided in the Supplementary Note.

UK Biobank datasets. We analyzed genetic data from the UK Biobank[20] consisting of 381,493 unrelated individuals or 459,256 related individuals of European ancestry (based on self-reported white ethnicity; sample sizes depended on the trait analyzed; Table 1 and Supplementary Table 14) and ~20 million imputed variants with MAF > 0.1% (ref. [15]). We identified pairs of siblings (full siblings and monozygotic twins) as outlined previously[20,31]. We note that it is easy to distinguish sibling pairs from parent–child pairs using identity by descent. We identified parent–offspring pairs as individuals having a kinship coefficient in the range of (0.177, 0.354) and having a proportion of SNPs with identity by state equal to $0 < 0.0012$. We computed association statistics applying either linear regression or default BOLT-LMM analysis (both implemented in BOLT-LMM software). We note that, for this analysis, BOLT-LMM association statistics reduce to infinitesimal mixed-model association statistics (BOLT-LMM-inf) when the non-infinitesimal BOLT-LMM association statistics are not expected to increase power[14,15]. We controlled for assessment center, genotype array, sex, age, age squared and the first 20 PCs[15].

We considered 12 diseases in the UK Biobank for which family history of disease was available (Table 1). These diseases were considered when GWAX was proposed[1] and had a wide range of narrow-sense heritabilities (Supplementary Table 15). We assigned case–control status for all genotyped individuals on the basis of the presence or absence of ICD-9 codes, ICD-10 codes or other relevant codes (Supplementary Table 14). Individuals with no ICD-9 or ICD-10 codes were assigned as controls. An alternative would have been to assign missing phenotypes to individuals with no ICD-9 or ICD-10 codes, which would have reduced sample size while increasing disease prevalence (Supplementary Table 40). We decided against this possibility, because we determined that it slightly but consistently reduced total genetic signal[23] and had very little impact on GWAS power (Supplementary Table 41).

Parental history and sibling history information was available for all 12 diseases. Parental history consisted of the presence or absence of disease in each respective parent; we note that the parental prevalence was often greater than the disease prevalence (Table 1). Sibling history consisted of the number of brothers and sisters and the presence or absence of disease in the set of all siblings.

Family history is self-reported, and thus there is a risk of misreporting, recall bias or phenotype misclassification. For example, we assumed that participants' family history of 'diabetes' corresponded to that specifically of T2D, consistent with previous studies[1]. To assess the accuracy of family history information, we computed the correlation of self-reported family history between siblings. When computing the correlation of self-reported sibling history, we restricted the analysis to concordant sibling pairs (for example, both cases or both controls). For sex-specific diseases (such as breast and prostate cancer), we restricted the analysis to concordant sibling pairs of the relevant sex, sibling pairs of the non-relevant sex and sibling pairs of discordant sex where the relevant sex was a control. We observed a moderately high concordance of self-reported family history among siblings (Supplementary Table 28).

LT-FH in the UK Biobank. There are 377 possible configurations of case–control status and family history of disease. For example, there are 2 configurations when case–control status but no family history are available (case and control) and 4 configurations when case–control status and one parent's disease status is available (case with affected parent, case with unaffected parent, control with affected parent and control with unaffected parent). In detail, 252 configurations are possible when case–control status is available: case–control status with no family history (2; see above); case–control status and one parent's disease status (4; see above); case–control status and both parents' disease status (6); case–control status and sibling's disease status (40; 1–10 siblings, at least one or none affected); case–control status, one parent's disease status and sibling disease status (80); and case–control status, both parents' disease status and sibling disease status (120). There are 125 possible configurations when case–control status is unavailable: one parent's disease status (2); sibling disease status (20); both parents' disease status (3); one parent's disease status and sibling disease status (40); and both parents' disease status and sibling disease status (60). A comprehensive list of configurations is provided in Supplementary Table 42.

We conditioned on known information to compute the posterior mean genetic liability. We computed $E[\epsilon_{o,g}|\cdot]$, where • is all known disease status information, through Monte Carlo integration as described above. We allowed for genotyped individuals and their siblings to have a different disease prevalence than their parents. We selected the prevalence for genotyped individuals and their siblings using the prevalence of the disease in genotyped (unrelated European) individuals, and the prevalence for parents using the parental disease prevalence (for parents of unrelated European individuals). h^2 values were obtained from twin studies in the published literature (Supplementary Table 15), but we note the existence of h^2 estimates based on family history[32].

For individuals who reported having no siblings, more than ten siblings, or did not know or preferred not to answer, we ignored sibling history and used parental history only for LT-FH; for all other individuals, we incorporated both the number of siblings and whether at least one was affected in the computation of $E[\epsilon_{o,g}|\cdot]$. For sex-specific diseases, our exclusion criteria remained consistent; however, we next incorporated the number of siblings of the relevant sex, as well as whether at least one was affected, in the computation of $E[\epsilon_{o,g}|\cdot]$. Further details are provided in the Supplementary Note.

GWAS and GWAX in the UK Biobank. We implemented GWAS considering the case–control status of the genotyped individual only (Supplementary Table 19 and Table 1). Due to the increased risk of a type I error in unbalanced case–control settings, we selected a stringent MAF threshold when computing the number of independent loci for some diseases[15,21]. However, a method such as SAIGE that controls type I error in unbalanced case–control settings could also be applied[21].

We considered the case–control status of the genotyped individual and the disease history of parents and siblings when implementing GWAX (Supplementary Table 19). The GWAX prevalence (that is, the prevalence of proxy cases) was generally more than double the parental prevalence and many times larger than the disease prevalence (Supplementary Table 43). For individuals with at least one sibling or who did not know or preferred not to answer, we assigned using sibling and parental disease history; for individuals with no reported siblings, we assigned using parental disease history only (Supplementary Table 19). For sex-specific diseases, the conditions were replaced with siblings of the relevant sex.

Assessing calibration in the UK Biobank. We assessed the calibration of each method using S-LDSC attenuation ratio[15,22–24]. We report two estimates for the mean attenuation ratio (and mean attenuation ratio difference) across all traits. We let x_i denote the attenuation ratio (or attenuation ratio difference) for trait i and σ_i the reported jackknife s.e. (for the ratio or the difference of ratios). We computed a simple average across traits, with s.e. computed as $\frac{1}{12}\sqrt{\sum_i \sigma_i^2}$. We also computed an inverse-variance-weighted mean where the weights were determined by the variance of the GWAS attenuation ratio. The inverse-variance-weighted estimate was $\sum_i \left\{x_i/\sigma^2_{\mathrm{GWAS},i}\right\} / \sum_i \left\{1/\sigma^2_{\mathrm{GWAS},i}\right\}$, with s.e. computed as $\sqrt{\sum_i \left\{\sigma_i^2/\sigma^4_{\mathrm{GWAS},i}\right\}} / \sum_i \left\{1/\sigma^2_{\mathrm{GWAS},i}\right\}$. Further details are provided in the Supplementary Note.

Assessing power in the UK Biobank. We computed the number of independent loci as defined similarly to ref. [15]. The relative improvement for a method M versus a method m was defined as $\frac{\sum_t N_{t,M} - \sum_t N_{t,m}}{\sum_t N_{t,m}}$, where $N_{t,M}$ is the number of independent loci for trait t and method M; the s.e. was computed via block jackknife of the genome (200 blocks). Further details are provided in the Supplementary Note.

Assessing heritability and correlation in the UK Biobank. We estimated observed-scale SNP heritability $\left(\hat{h}^2_{\mathrm{g,o}}\right)$ for each method by applying BOLT-REML[26] to unrelated European individuals. Observed-scale SNP heritability was converted to liability-scale SNP heritability using the in-sample disease prevalence[10]. Genetic correlation was estimated using BOLT-REML[26]. The phenotypic correlation and correlation of $-\log_{10}P$ values (across all variants with a minimum MAF determined by case prevalence; Supplementary Table 16) were estimated in R. For both genetic and phenotypic correlation, we restricted our attention to individuals with non-missing values for both methods that were compared.

Replication analysis. We conducted a replication analysis of loci identified by GWAS and/or LT-FH in independent non-UK Biobank datasets for four diseases—CAD, T2D, breast cancer and prostate cancer—with publicly available summary statistics[33–36] (Supplementary Table 25). The replication summary statistics were from studies consisting of predominantly non-UK Europeans and were always computed using GWAS. We computed the replication slope as above for both genome-wide-significant loci identified in the UK Biobank by GWAS and genome-wide-significant loci identified in the UK Biobank by LT-FH.

In detail, we restricted association results from linear regression on unrelated Europeans for GWAS and LT-FH to SNPs that (1) appeared in the replication studies' SNP set, (2) met the MAF threshold for that trait from our main analysis and (3) were significant in GWAS or LT-FH. We then LD 'clumped' this restricted set of association statistics (stringent 5-Mb window and R^2 threshold of 0.01 for LD clumping). For GWAS (and LT-FH, respectively) for each LD clump with at least one significant SNP, we selected the most significant SNP as the lead SNP. When an LD clump contained only SNPs that were significant for LT-FH (or GWAS, respectively), this was considered an LT-FH-only locus (or GWAS only). The replication slope[27,37] was then computed as the regression slope of standardized effect sizes of lead SNPs in case–control replication data versus GWAS or LT-FH UK Biobank discovery data. Standardized effect sizes were computed as $Z/\sqrt{N_{\mathrm{eff}}}$, where Z denotes z-score. For GWAS (and GWAS replication data), N_{eff} was computed as $\frac{4}{1/N_A+1/N_U}$, and for LT-FH N_{eff} was computed as $N_{\mathrm{eff,GWAS}} \times c$, where c is the relative effective sample size of LT-FH versus GWAS (Supplementary Table 22).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

This study analyzed data from the UK Biobank, which are publicly available by application (http://www.ukbiobank.ac.uk/). We have publicly released summary association statistics computed by applying our LT-FH method to UK Biobank data; LT-FH summary association statistics for 12 diseases are available at https://data.broadinstitute.org/alkesgroup/UKBB/LT-FH/sumstats/.

Code availability

We have publicly released open-source software implementing our LT-FH method; LT-FH software (v1 and v2): https://data.broadinstitute.org/alkesgroup/UKBB/LTFH/; BOLT-LMM v2.3 software: https://data.broadinstitute.org/alkesgroup/BOLT-LMM, LTSOFT software: https://data.broadinstitute.org/alkesgroup/LTSOFT/; and PLINK software: https://www.cog-genomics.org/plink2.

References

31. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
32. Munoz, M. et al. Evaluating the contribution of genetic and familial shared environment to common disease using the UK Biobank. *Nat. Genet.* **48**, 980–983 (2016).
33. Schunkert, H. et al. Large-scale association analyses identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338 (2011).
34. Morris, A. P. et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
35. Michailidou, K. et al. Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92–94 (2017).
36. Schumacher, F. et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat. Genet.* **50**, 928–936 (2018).
37. Kichaev, G. et al. Leveraging polygenic functional enrichment to improve GWAS power. *Am. J. Hum. Genet.* **104**, 65–75 (2019).

Acknowledgements

We are grateful to L. O'Connor, O. Weissbrod, N. Zaitlen, G. Kichaev and A. Gusev for helpful discussions and E.M. Pedersen for computational suggestions. This research was funded by NIH grants R01 HG006399 (N.P. and A.L.P.), R01 MH101244 (A.L.P.), R01 MH107649 (A.L.P.), NSF CAREER award DBI-1349449 (S.G. and A.L.P.) and 5T32CA009337-32 (M.L.A.H.). P.-R.L. was supported by the Next Generation Fund at the Broad Institute of MIT and Harvard and a Sloan Research Fellowship. This research was conducted using the UK Biobank resource under application no. 10438.

Author contributions

M.L.A.H. and A.L.P. designed the experiments. M.L.A.H. performed the experiments and statistical analysis; N.P. assisted in proving the equivalence to a score test. M.L.A.H., S.G., P.-R.L. and A.L.P. analyzed the data. M.L.A.H. and A.L.P. wrote the manuscript with assistance from S.G., P.-R.L. and N.P.

Competing interests

The authors declare no competing interests.

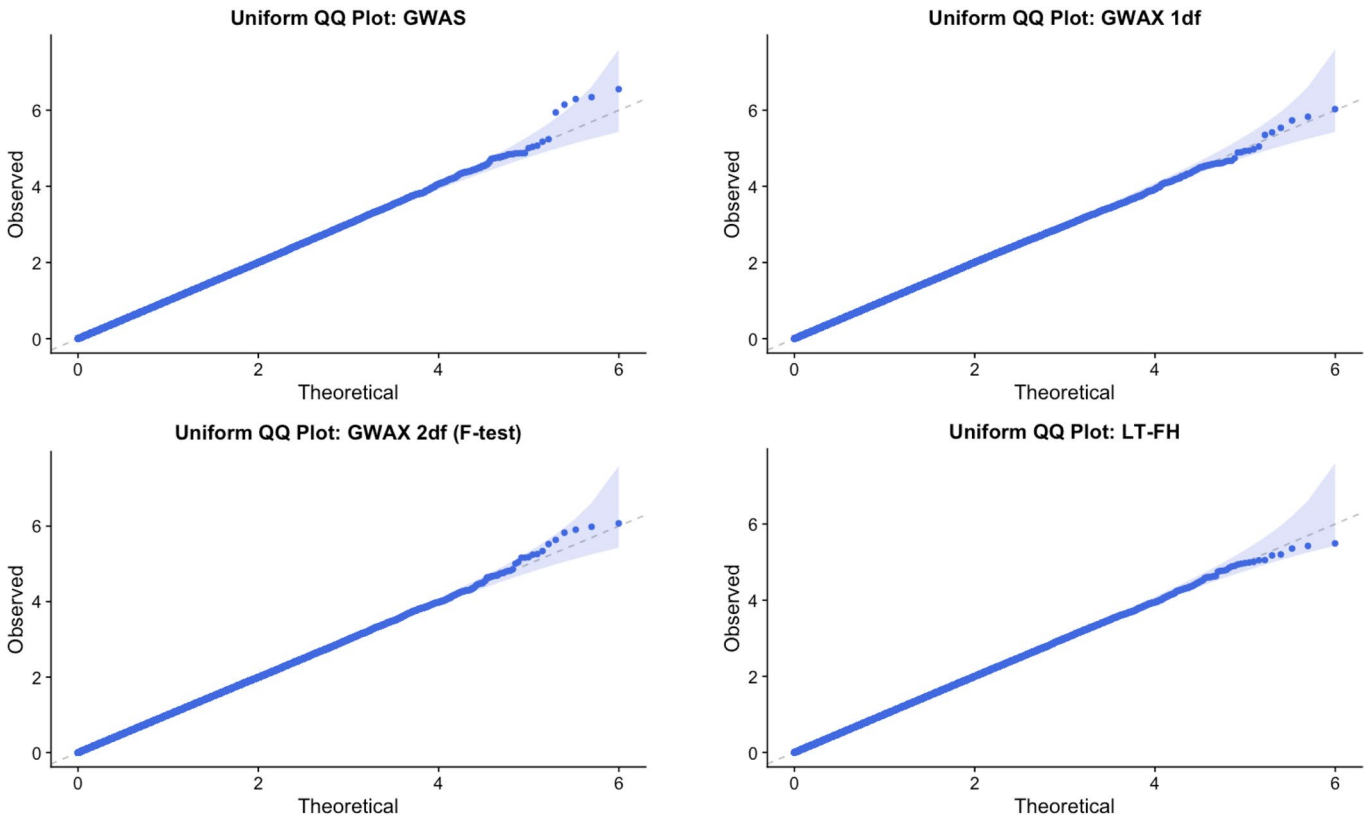
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-020-0613-6.

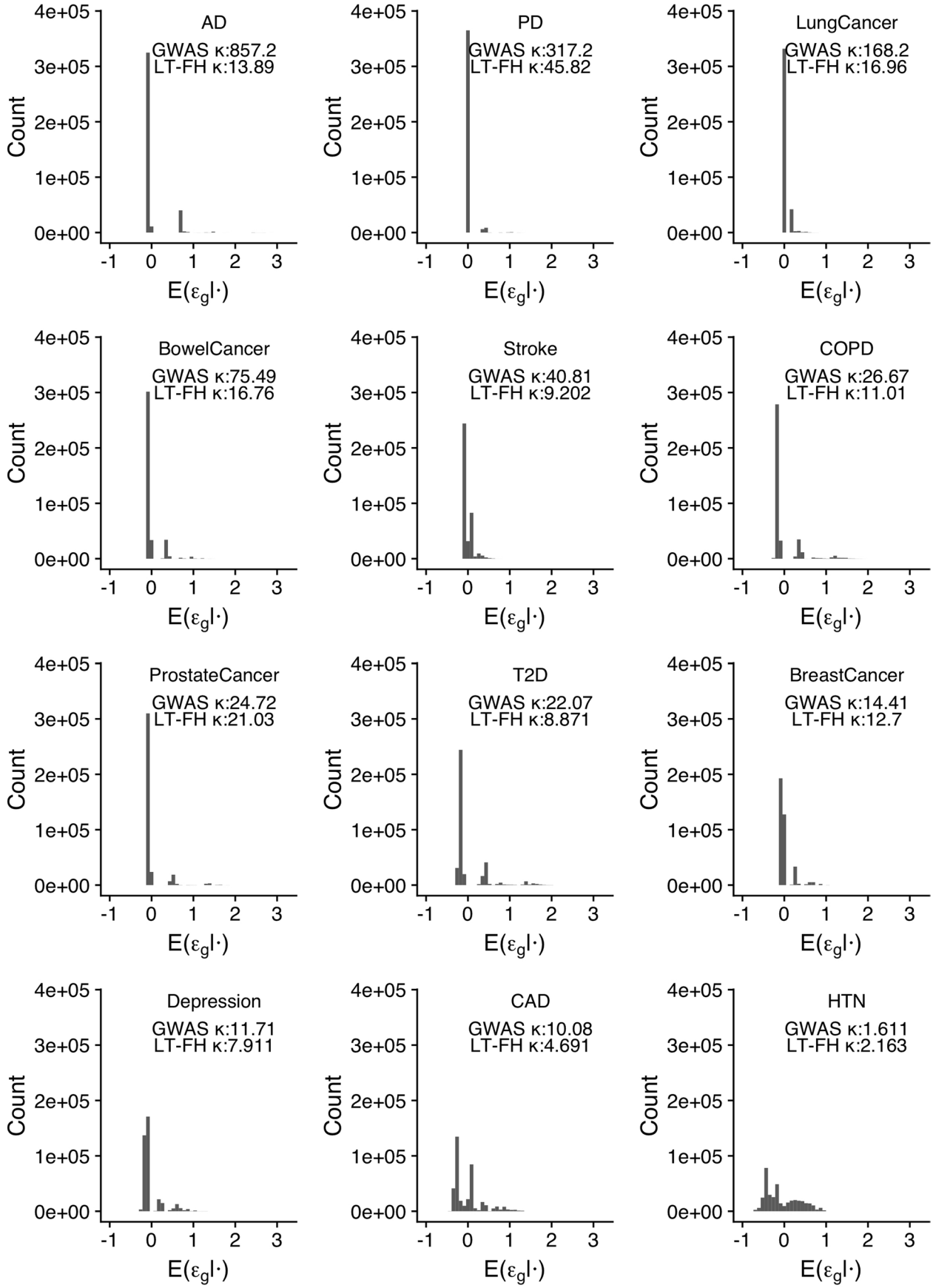
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-020-0613-6.

Correspondence and requests for materials should be addressed to M.L.A.H. or A.L.P.

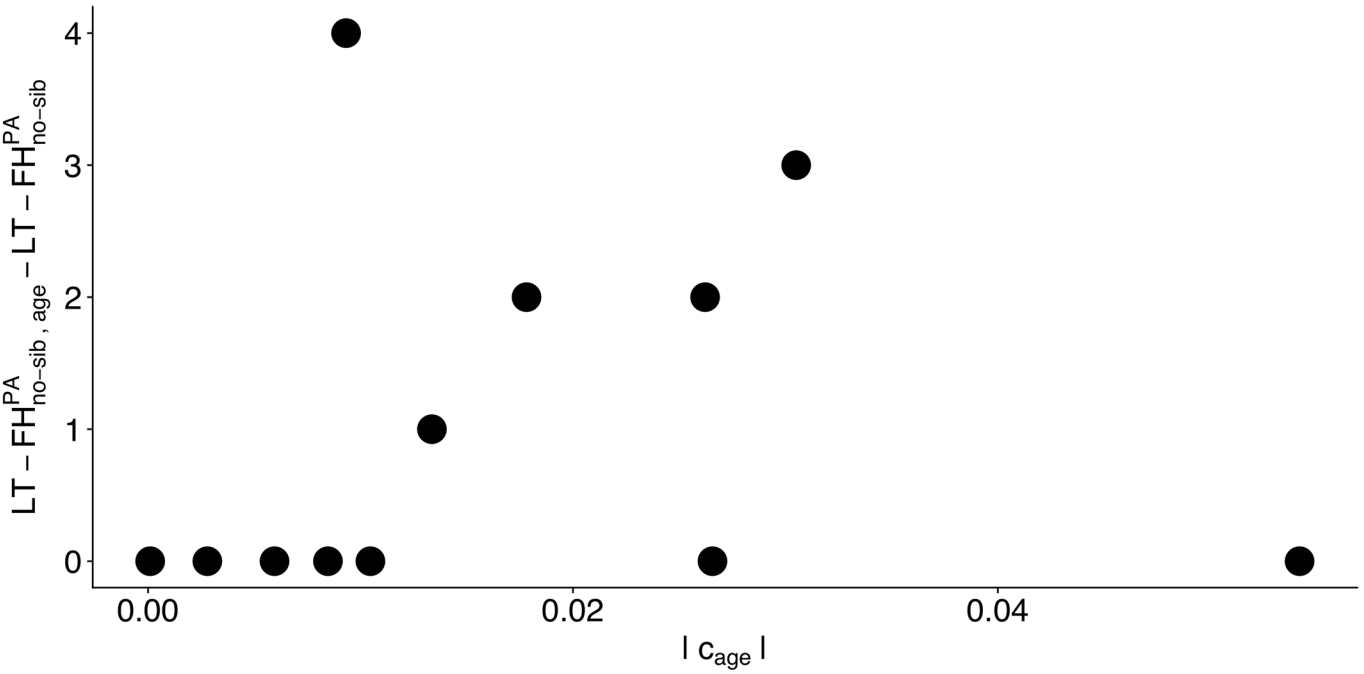
Reprints and permissions information is available at www.nature.com/reprints.



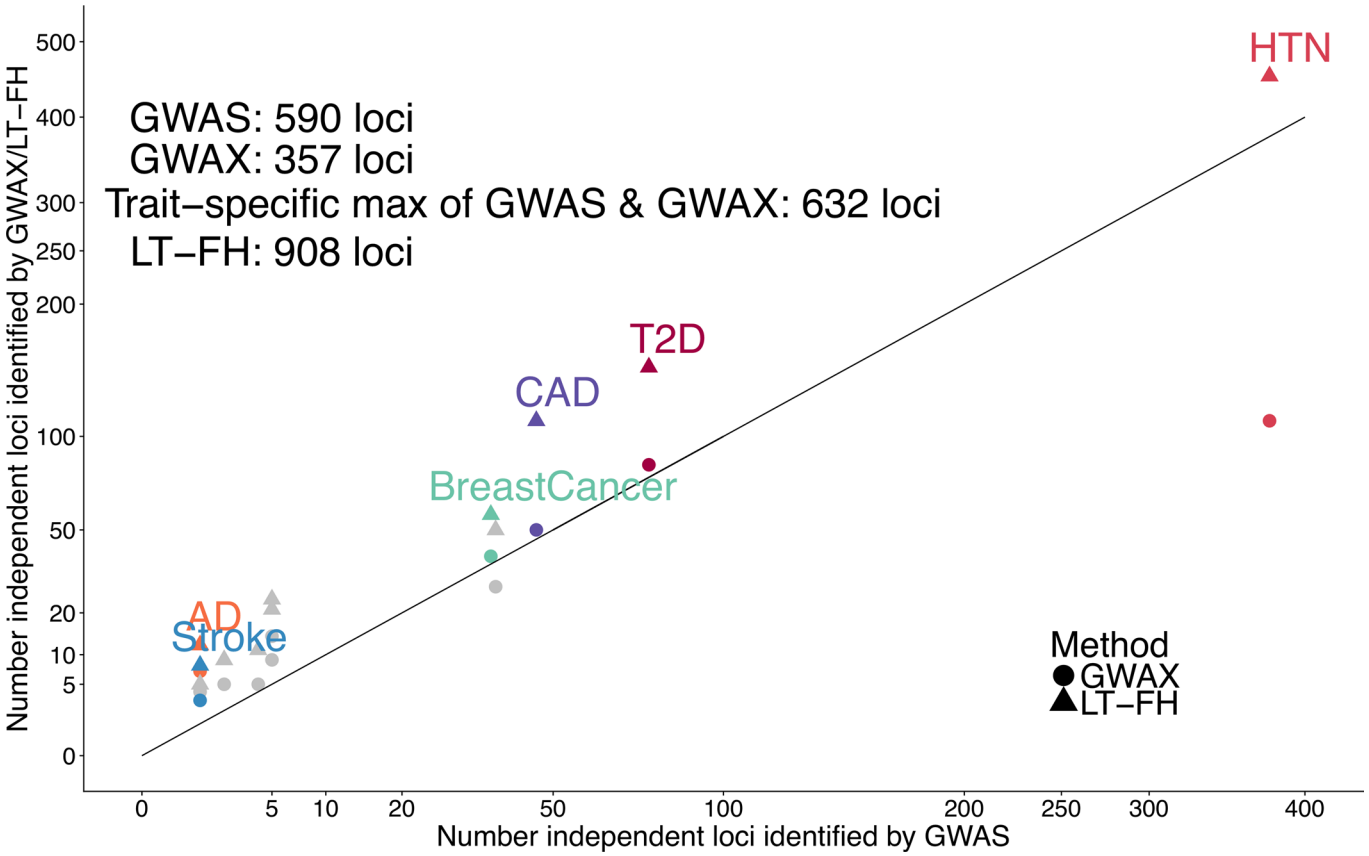
Extended Data Fig. 1 | QQ plots from simulations with default parameter settings. We report quantile-quantile (QQ) plots for null SNPs in simulations with default parameter settings. Results are based on 10 simulation replicates. These QQ plots compare the observed distribution of p-values with the standard uniform distribution. We plot the observed $-\log_{10}(p)$ as a function of $-\log_{10}\left(\frac{rank}{n+1}\right)$ and the 95% confidence bands are constructed pointwise using the beta distribution.



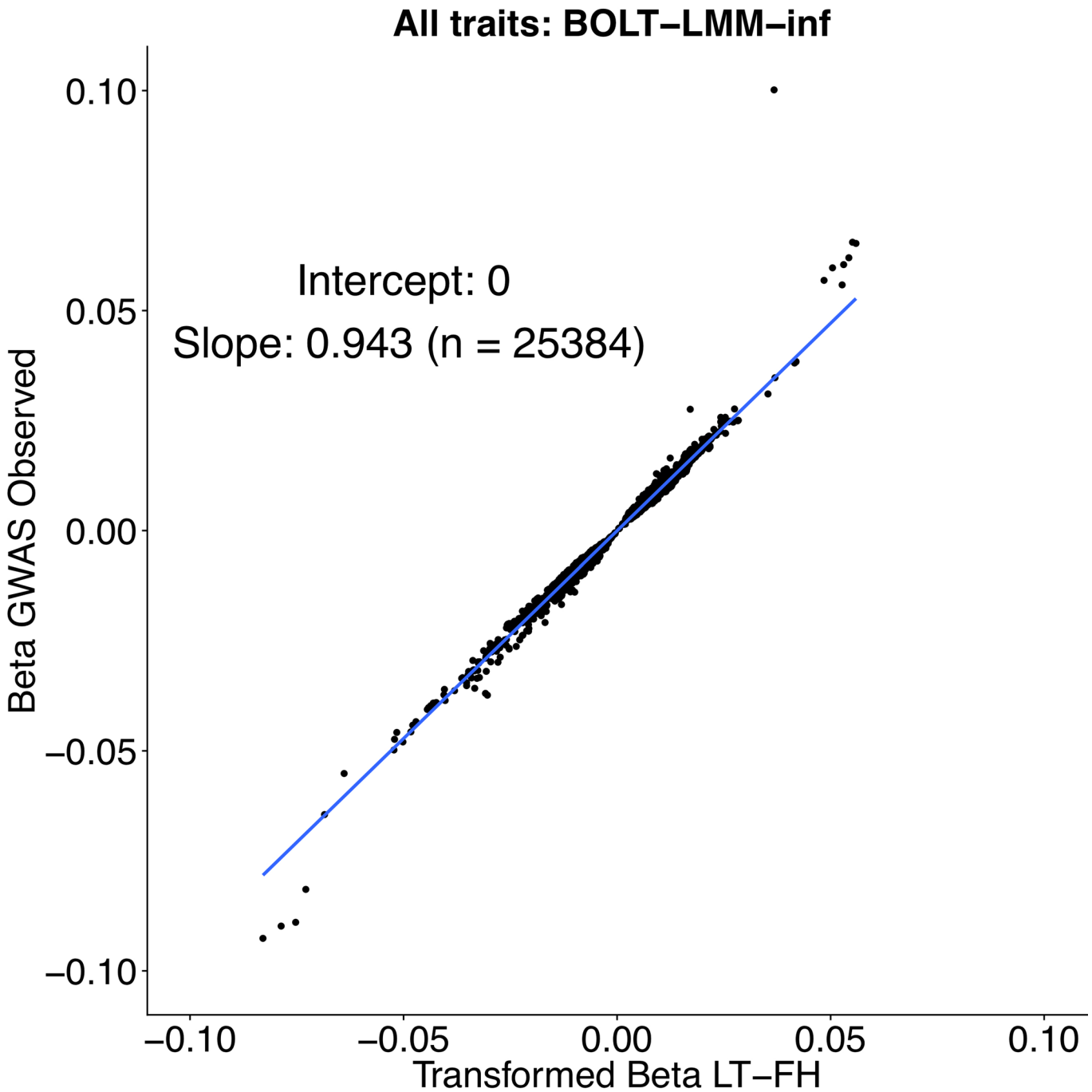
Extended Data Fig. 2 | Distribution of LT-FH phenotypes for 12 UK Biobank diseases. We plot the distribution of the LT-FH phenotype for each disease. We also report the kurtosis for both GWAS and LT-FH; Pearson's measure of kurtosis, $\kappa = \frac{E\left[(X-\mu)^4\right]}{\left(E\left[(X-\mu)^2\right]\right)^2}$, is calculated using the R package moments.



Extended Data Fig. 3 | Impact of modifying the LT-FH method to incorporate age information as a function of the liability threshold model parameter for age for 12 UK Biobank diseases. We plot the increase in number of independent loci for LT-FH$^{PA}_{no\text{-}sib,age}$ relative to for LT-FH$^{PA}_{no\text{-}sib}$ (Supplementary Table 32) against the liability threshold model parameter $|c_{age}|$(Supplementary Table 30).



Extended Data Fig. 4 | LT-FH increases association power across 12 diseases from the UK Biobank in analyses incorporating related individuals. We report results of GWAS using BOLT-LMM on related Europeans, GWAX using BOLT-LMM on unrelated Europeans, and LT-FH using BOLT-LMM on related Europeans using only case–control status for all sibling pairs and parent-offspring pairs within the set of target samples. Numerical results are reported in Supplementary Table 37.



Extended Data Fig. 5 | Strong concordance between GWAS BOLT-LMM-inf effect sizes and transformed LT-FH BOLT-LMM-inf effect sizes. We plot GWAS BOLT-LMM-inf effect sizes and transformed LT-FH BOLT-LMM-inf effect sizes for genome-wide significant effect sizes ($P \leq 5*10^{-8}$ for both GWAS and LT-FH BOLT-LMM-inf). We note that BOLT-LMM only outputs effect size estimates for BOLT-LMM-inf, the BOLT-LMM approximation to the infinitesimal mixed model. Our effect size for GWAS is the outputted $\beta_{GWAS,BOLT-LMM-inf}$ (per-allele observed scale) and for LT-FH we estimate a (per-allele observed scale) effect size as $\beta = \frac{\beta_{LT-FH,BOLT-LMM-inf}}{se(\beta_{LT-FH,BOLT-LMM-inf})\sqrt{N_{GWAS}*c}} \frac{\sqrt{K(1-K)}}{\sqrt{2(MAF)(1-MAF)}}$, where c is the boost in N_{eff} for LT-FH relative to GWAS, K is disease prevalence in GWAS and MAF is the minor allele frequency of the SNP.

nature research

Corresponding author(s): Margaux L.A. Hujoel
Alkes L. Price

Last updated by author(s): 3/11/2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	N/A
Data analysis	Data was analyzed using R, BOLT-LMM v2.3 (https://data.broadinstitute.org/alkesgroup/BOLT-LMM), LTSOFT (https://data.broadinstitute.org/alkesgroup/LTSOFT/), PLINK (https://www.cog-genomics.org/plink2), ldsc (https://github.com/bulik/ldsc), and LT-FH (https://data.broadinstitute.org/alkesgroup/UKBB/LTFH/).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

This study analyzed data from the UK Biobank, which is publicly available by application (see Data availability). We have publicly released summary association statistics computed by applying our LT-FH method to UK Biobank data (see Data availability).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes depended on the availability of individual level disease data as well as family history of disease information within UK Biobank.
Data exclusions	Our study was restricted to individuals of European ancestry within UK Biobank. We excluded variants based on quality control metrics and MAF described in our manuscript.
Replication	We perform a replication of significant loci discovered via GWAS and LT-FH for 4 traits.
Randomization	N/A
Blinding	N/A

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	UK Biobank consists of ~500,000 individuals between the ages of 40 and 69 (see Bycroft et al. 2018 Nature for more information on the population). In our study, we restricted attention to self-reported Europeans.
Recruitment	Not applicable; our study had no involvement in recruitment.
Ethics oversight	This research is not Human Subjects Research, because all data is anonymized and obtained from providers that are prohibited by established regulations or policies from releasing identifiers.

Note that full information on the approval of the study protocol must also be provided in the manuscript.