Combining 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], Alkes L. Price[1,2,3,*]

[1]) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA
[2]) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA
[3]) Broad Institute of MIT and Harvard, Cambridge, MA
[4]) Brigham and Women's Hospital/Harvard Medical School, Boston, MA

Abstract

Family history of disease can provide valuable information about an individual's genetic liability for disease in case-control association studies, but it is currently unclear how to best combine case-control status and family history of disease. We developed 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); association statistics are computed via linear regression of genotypes and posterior mean genetic liabilities, equivalent to a score test. We applied LT-FH to 12 diseases from the UK Biobank (average N=350K). We compared LT-FH to genome-wide association without using family history (GWAS) and a previous proxy-based method for incorporating family history (GWAX). 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, based on the number of independent genome-wide significant loci detected across all diseases (e.g. 690 independent loci for LT-FH vs. 423 for GWAS); the second best method was GWAX for lower-prevalence diseases and GWAS for higher-prevalence diseases, consistent with simulations. We also confirmed that LT-FH was well-calibrated (assessed via stratified LD score regression attenuation ratio), consistent with simulations. When using BOLT-LMM (instead of linear regression) to compute association statistics for all three methods (increasing the power of each method), 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. In summary, LT-FH greatly increases association power in case-control association studies when family history of disease is available.

*) Corresponding authors: MLAH (hujoel@g.harvard.edu) and ALP (aprice@hsph.harvard.edu)

Introduction

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] (and the accuracy of genetic risk prediction[2,3]). Standard 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[4–6], 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 if and only if an underlying continuous-valued liability lies above a threshold, has proven valuable in a wide range of settings[2,3,7–13]. 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 captures the fact that controls with family history have higher genetic liability than controls without family history (and 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 for diseases of both low and high prevalence in analyses of 12 diseases from the UK Biobank.

Results

Overview of methods

The LT-FH method relies on the liability threshold model[7], 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; (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, etc.) 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[16] (which differs from SNP-heritability[17]) from the literature. Step 2 generalizes the Armitage trend test[18], and we show that it is equivalent to a score test; an alternative is to apply BOLT-LMM[14,15] (see URLs) to posterior mean genetic liabilities, increasing power in large samples. We emphasize that we utilize posterior mean *genetic* liabilities. Details of the LT-FH method are provided in the Methods section; we have publicly released open-source software implementing the method (see URLs).

We consider a simple example of a disease with a prevalence of 5% and narrow-sense heritability of 50%, utilizing parental history only. The posterior mean genetic liability is -0.10 for a control with 0 affected parents vs. 0.39 for a control with 1 affected parent, but GWAS treats these individuals identically. In addition, the posterior mean genetic liability is 1.56 for a case with 2 affected parents vs. 0.39 for a control with 1 affected parent, but GWAX treats these individuals identically. On the other hand, LT-FH appropriately differentiates all case-control and family history configurations.

Simulations

We performed simulations by simulating genotypes at 100,000 unlinked SNPs and case-control status plus 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, used these to simulate genotypes for target samples (offspring), and 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 explaining $h^2 = 50\%$ 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. Further details of the simulation framework are provided in the Methods section. We note that simulations using real LD patterns are essential for methods impacted by LD between SNPs, but the LT-FH method is not impacted by LD between SNPs.

We assessed the calibration of each method using average χ^2 statistics at null SNPs (Figure 2a and Table S1). 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 (Table S2), as indicated by QQ plots for null SNPs (Figure S1).

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 $p < 5 * 10^{-8}$) (Figure 2b,c and Table S1). LT-FH was the most powerful method, with a +22% increase in average χ^2 and a +55% increase in power compared to GWAX, which outperformed GWAS at default parameter settings.

We also considered a 2 degree-of-freedom (df) extension of GWAX that treats cases, proxy cases, and controls without family history of disease as 3 distinct groups (GWAX-2df; see Methods). In contrast to previously proposed 2df extensions of GWAX[1], our GWAX-2df test allows for incorporation of covariates, enabling its application to real traits (see below). In our simulations, GWAX-2df was well-calibrated (Table S1, Table S2 and Figure S1) and more powerful than GWAX, but LT-FH still attained a +25% increase in power compared to GWAX-2df (Table S1).

We investigated how the improvement in power attained by LT-FH varied as a function of disease

prevalence K by performing additional simulations at lower prevalence ($K = 1\%$) and higher prevalence ($K = 25\%$). We varied the number of causal SNPs to approximately match the average χ^2 and power (for the GWAS method) in our main simulations, and used default settings for other parameters. Results are reported in Table S3. At lower prevalence, LT-FH attained a +28% increase in power as compared to GWAX, which far outperformed GWAS. At higher prevalence, LT-FH attained a +82% increase in power as compared GWAS, which far outperformed GWAX. These results are consistent with previous findings that the GWAX method is most useful for diseases of low prevalence[1].

We performed four secondary analyses to assess the robustness of LT-FH. First, we varied the number of causal SNPs by performing additional simulations with 250 or 750 causal SNPs (Table S4). As expected, decreasing (resp. increasing) the number of causal SNPs — which increases (resp. decreases) causal effect sizes — increased (resp. decreased) the power of all methods; average χ^2 statistics scaled inversely with the number of causal SNPs for each method, such that the relative ordering of the methods was unchanged. Second, we varied the liability-scale heritability (h^2) by performing additional simulations with h^2 equal to 0.25 or 0.75 (Table S5). Decreasing (resp. increasing) the heritability led to larger (resp. smaller) improvements for GWAX and LT-FH compared to GWAS but a smaller (resp. larger) improvement for LT-FH vs. GWAX; LT-FH still attained a $\geq$45% increase in power compared to GWAX at each value of h^2. Third, we performed simulations in which the LT-FH method utilized a misspecified value of disease prevalence ($K = 2.5\%$ or $K = 7.5\%$, vs. true $K = 5\%$) or liability-scale heritability ($h^2 = 0.25$ or $h^2 = 0.75$, vs. true $h^2 = 0.50$) (Table S6). The impact on association power was negligible in each case. Fourth, we performed simulations with shared environment, which introduces a covariance in the non-genetic component of liability between parents and target samples (offspring); this covariance was set to 0.5 times the parent-offspring covariance in the genetic component of liability, i.e. $0.5 * 0.5h^2 = 0.125$ (Table S7). This led to smaller improvements for GWAX and LT-FH compared to GWAS but a larger (+72%) increase in power for LT-FH vs. GWAX, which still slightly outperformed GWAS.

In summary, LT-FH was well-calibrated and more powerful than all other methods at all parameter settings tested. However, 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[19] with genotype data available for up to 381,493 unrelated individuals of European ancestry (20 million imputed SNPs with MAF >0.1%) and family history of disease available for most individuals (Methods, Table 1, Table S8,Table S9). 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[20] (Table S10, Table S11). Family history information was available for each disease, including parental history (presence or absence of disease in each respective parent) and sibling history (number of brothers and sisters + presence or absence of disease in the set of all siblings). Averaged across the 12 diseases, 88% of target samples had complete parental history information and 93% of target samples had complete sibling history information (Table S12); the accuracy of family history information is assessed below. We defined GWAS, GWAX and LT-FH phenotypes accordingly (Methods, Table S13; Figure S2) 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.

We assessed the calibration of GWAS, GWAX and LT-FH using stratified LD score regression (S-LDSC) attenuation ratio[15,21–23], defined as (S-LDSC intercept -1)/(mean χ^2 - 1) (Methods and Table S14a). 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 can be no guarantee that analyses of UK Biobank data using PC covariates are completely devoid of confounding[24], these results confirm that that LT-FH is well-calibrated and there is no confounding that is 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, as defined in ref.[15] (Methods, Figure 3a and Table S14b); we computed standard errors (s.e.) on 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 (e.g. 690 independent loci for LT-FH vs. 423 for GWAS). For the 8 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 4 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 vs. GWAX results were dominated by hypertension, the disease with highest prevalence. We also 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 (Table S14c) and average χ^2 across all SNPs (Table S14d); 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 using BOLT-REML[25] (Table 2). Genetic correlations were high, much higher than the corresponding phenotypic correlations (e.g. 0.97 and 0.70 respectively for correlations between LT-FH and GWAS, 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) vs. GWAS were particularly low for low-prevalence diseases, for which family history is more informative than case-control status. We also computed correlations between $-\log_{10}$ association p-values, which mirrored the phenotypic correlations (Table S15). Notably, LT-FH phenotypes had higher values of sample size times observed-scale SNP-heritability (a measure of total genetic signal[22]) than GWAS phenotypes for all 12 diseases (59% higher on average; Table S16). These findings support the use of the LT-FH method to increase association power.

We performed five secondary analyses. First, we computed association statistics for GWAX-2df, restricting this analysis to 672,292 genotyped SNPs to limit computational cost (Methods). GWAX-2df was substantially more powerful than GWAX (assessed using the number of independent loci), but this had little impact on our conclusions about LT-FH, which was +23% (s.e. 3%) more powerful than the trait-specific maximum of GWAS and GWAX-2df (vs. +35% (s.e. 4%) more powerful than the trait-specific maximum of GWAS and GWAX) in analyses of genotyped SNPs (Table S17). Second, we computed association statistics for analogues of GWAX and LT-FH that incorporate only parental history and not sibling history (GWAX_{no-sib} and LT-FH_{no-sib}). We determined that using only parental history was slightly less powerful, e.g. LT-FH_{no-sib} was +54% (s.e. 5%) more powerful than GWAS (whereas LT-FH was +63% (s.e. 6%) more powerful than GWAS) (Table S18). Third, we assessed the accuracy of family history information by computing the correlation of self-reported family history between siblings (Methods and Table S19). Averaged across traits, the correlation between siblings was 0.685 for number of affected parents and 0.583 for presence or absence of disease in the set of all siblings; it follows that the correlation between true and self-reported family history is equal to the square root of these numbers (0.827 and 0.764), if errors are uncorrelated between siblings. To investigate the importance of accounting for inaccurate family history information, we modified the LT-FH method to downweight family history information based on its accuracy for each disease (Methods), but we determined that this had little impact on association results (0% increase in power for modified LT-FH vs. LT-FH, average phenotypic correlation = 0.996 across 12 diseases; Table S20). Fourth, we investigated whether LT-FH could be improved by explicitly accounting for age when computing posterior mean genetic liabilities[10] (Table S21, Table S22). We determined that this had little impact on association results (< 2% increase in power; Table S23, Figure S3), consistent with previous findings that including age as a simple covariate is sufficient in case-control studies with random ascertainment[10]. Finally, for the 8 lower-prevalence diseases for which we applied stricter MAF thresholds to avoid type I error in unbalanced case-control settings[20], we recomputed association statistics

for GWAX and LT-FH using MAF thresholds chosen specifically for each method, based on the kurtosis of the corresponding phenotypes (Table S10, Table S11; the number of diseases requiring stricter MAF thresholds reduces to 0 for GWAX and 2 for LT-FH, due to lower kurtosis). The number of independent loci identified by GWAX and LT-FH increased for these diseases, but overall results were little changed, e.g. LT-FH was +37% (s.e. 4%) more powerful than the trait-specific maximum of GWAS and GWAX (vs. +36% (s.e. 4%) more powerful than the trait-specific maximum of GWAS and GWAX in our primary analysis) (Table S24).

In summary, LT-FH was well-calibrated and much more powerful than all other methods in all primary and secondary analyses.

BOLT-LMM increases GWAS, GWAX and LT-FH association power

BOLT-LMM, a method for efficient Bayesian mixed-model analysis, has been 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 12 UK Biobank diseases as above. We first confirmed that GWAS, GWAX and LT-FH remained well-calibrated (Table S25a). We then assessed association power (Figure 3b and Table S25b). 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 (e.g. 736 independent loci for LT-FH vs. 442 for GWAS), implying similar relative power among the three methods when using BOLT-LMM; the absolute increases in power attained by using BOLT-LMM were modest (Figure 3b vs. Figure 3a) and smaller than 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 (Table S25c) and average χ^2 across all SNPs (Table S25d) 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[1,4-6]. We investigated the consequences of including related individuals in BOLT-LMM analyses. We first applied BOLT-LMM to GWAS, GWAX and LT-FH phenotypes of target samples defined 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, vs. 0.14 for GWAS) (Table S26a), rendering moot the increases in power due to larger sample size (Table S26b). A likely explanation for the poor calibration is the extreme concordance between sibling pair phenotypes (close to 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 thus 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 pair or parent-offspring pair within the set of target samples (also 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, vs. 0.14 for GWAS (and inverse variance-weighted average close to 0.1, consistent with ref.[15]) (Table S27a). We compared the power of this approach to applying BOLT-LMM to GWAS for all related Europeans (see above) and GWAX for all unrelated Europeans (as previously recommended[1,4-6]; 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 (e.g. 908 independent loci for LT-FH vs. 590 for GWAS) (Figure S4 and Table S27b), similar to our previous analyses (Figure 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 the recommended approach; we have publicly released LT-FH association statistics computed using this approach

(see URLs).

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 exactly one sibling for each set of siblings within the set of 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, vs. 0.14 for the recommended LT-FH approach; $p < 10^{-6}$ for difference in inverse-variance weighted means) (Table S28a), rendering moot the slightly higher power of this approach (Table S28b).

In summary, BOLT-LMM analysis increases the power of LT-FH (and 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. 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 data sets for which family history information is available, and to collect family history information in future genetic studies.

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% (Table S19); however, our efforts to account for this had little impact on association power (Table S20). Second, family history may reflect a different underlying genetic architecture than case-control status, e.g. due to differences in the etiology of early-onset vs. 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 in analyses of real diseases (Figure 3) were not as large as in simulations (Figure 2c). Fourth, just like standard GWAS analysis, LT-FH may require MAF thresholds stricter than our default value of 0.001 to avoid type I error in unbalanced case-control settings[20]. Although the number of diseases requiring stricter MAF thresholds was reduced from 8 for standard GWAS analysis to 2 for LT-FH (due to lower kurtosis; Table S10), more sophisticated approaches (generalizing ref.[20]) 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 is intended to account for phenotypic concordance among related individuals. 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 information 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[26]. Seventh, we have not explored the use of family history information for genetic risk prediction[2,3]. Despite all these limitations, we anticipate that LT-FH will attain large increases in power in future association studies.

URLs

BOLT-LMM v2.3 software: `https://data.broadinstitute.org/alkesgroup/BOLT-LMM`. LT-FH software: `https://data.broadinstitute.org/alkesgroup/UKBB/LTFH/`. LT-FH summary association statistics for 12 diseases: `https://data.broadinstitute.org/alkesgroup/UKBB/LTFH/sumstats/`. LTSOFT software: `https://data.broadinstitute.org/alkesgroup/LTSOFT/`. PLINK software: `https://www.cog-genomics.org/plink2`.

Methods

LT-FH method

The liability threshold model can be written as $\epsilon = \boldsymbol{X\beta} + \epsilon_e$, where ϵ is the liability, $\boldsymbol{X}$ is the genotypes at candidate SNPs (normalized to mean 0, variance 1), β 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 assume that when testing a SNP g for association, the effect size is small enough such that the liability is $\beta g + \epsilon = \beta g + \epsilon_g + \epsilon_e$ where $\epsilon \sim N(0, 1), \epsilon_g \sim N(0, h^2), \epsilon_e \sim N(0, 1 - h^2)$.

We assume a multivariate normal distribution for $\boldsymbol{\epsilon}$: for two individuals $Cov(\epsilon_1, \epsilon_2) = K_{12}h^2$, where K_{12} is the coefficient of relationship for the pair of individuals (e.g. 1/2 for parent-offspring). 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

$$\begin{pmatrix} \epsilon_{o,e} \\ \epsilon_{o,g} \\ \epsilon_{p_1} \\ \epsilon_{p_2} \\ \epsilon_s \end{pmatrix} \sim MVN_5 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1-h^2 & 0 & 0 & 0 & 0 \\ 0 & h^2 & 0.5h^2 & 0.5h^2 & 0.5h^2 \\ 0 & 0.5h^2 & 1 & 0 & 0.5h^2 \\ 0 & 0.5h^2 & 0 & 1 & 0.5h^2 \\ 0 & 0.5h^2 & 0.5h^2 & 0.5h^2 & 1 \end{pmatrix} \right). \tag{1}$$

where $\epsilon_{o,e}, \epsilon_{o,g}$ are the environmental and genetic components of the liability of the target sample (offspring), ϵ_{p_1} and ϵ_{p_2} are the liabilities of the parents, and ϵ_s is the liability of the sibling(s) (for simplicity we include only one sibling here, but this can be extended to an arbitrary number of siblings). We estimate the posterior mean genetic liability $E[\epsilon_{o,g}|z_o, z_{p_1}, z_{p_2}, \vec{z}_s]$ for each individual given the case-control status of the genotyped individual (z_o), both parents (z_{p_1}, z_{p_2}), and sibling(s) ($\vec{z}_s$). When disease status is unknown for subsets of these individuals, we condition only on known information. We are interested in 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 liability.

We estimate $E[\epsilon_{o,g}|z_o, z_{p_1}, z_{p_2}, \boldsymbol{z}_s]$ 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 liabilites in any configuration has a standard error of the mean above 0.01, we resample from a conditional multivariate normal to ensure that we sample in a manner that represents the tails of the distribution appropriately. (2) We sample relevant liabilities conditional on the offspring's case-control status (z_o) or we sample conditional on the parents (i.e.$z_{p_1} + z_{p_2} = \{0, 1, 2\}$). (3) We then combine all samples of $\epsilon_{o,g}$ from given configurations. (4) We compute posterior mean genetic liabilities and assess the standard error of each posterior mean. We then repeat (2)-(4) until the standard error of the mean is less than 0.01 for all configurations.

We considered the Pearson-Aitken (PA) formula as an approximate analytical approach and implemented this approximation for LT-FH$_{no-sib}$[2,27,28] (Table S23). However, in UK Biobank data, sibling history is provided as a binary response, i.e., 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 sample 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.

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 principal components. If we treat the posterior mean genetic liability as a continuous variable and compute the number of samples times the squared correlation between g and posterior mean genetic liability (generalizing the Armitage trend test[18]), this is equivalent to the Score test (see 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], and as such we incorporate 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[20].

For GWAX-2df test we perform 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 are both 0 (see Supplementary Note). When no covariates are included, results are virtually identical to a Pearson-χ^2 test on a 2x3 table.

Simulations

We simulate genotypes at 100,000 unlinked SNPs and case-control status plus family history (parental history for both parents) for 100,000 unrelated target samples; we do not include sibling history in these simulations. We simulate genotypes for both parents and use these to simulate genotypes for target samples (offspring); the underlying MAF is $\sim U(0.01, 0.5)$. Using these genotypes we simulate case-control status for both parents and target samples using a liability threshold model in which we normalize genotype to a mean of 0 and variance of 1 according to the true MAF, set the effect sizes for the C causal SNPs on the liability scale to be $\sqrt{h^2/C}$, and add environmental noise with a variance of $1 - h^2$. Target samples are not ascertained for case-control status. Our default parameter settings involve 500 causal SNPs explaining $h^2 = 50\%$ 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. For each parameter setting we compute 10 simulation replicates in which there is a different underlying MAF distribution for each simulation replicate.

We apply linear regression (implemented in BOLT-LMM software) to obtain χ^2 statistics for GWAS (case vs. control), GWAX (case+proxy-case vs. control), and LT-FH ($E[\epsilon_{o,g}|z_o, z_{p_1}, z_{p_2}]$ for each individual). We compute the posterior mean genetic liability for each of the 6 configurations of case-control status and family history using 1,000,000 values of ϵ_g sampled conditional on the 3 parent history configurations (i.e.$z_{p_1} + z_{p_2} = \{0, 1, 2\}$).

We investigate 12 simulation scenarios total in which we vary C (number of causal SNPs), K (prevalence), h^2, the assumed h^2 and prevalence when calculating $E[\epsilon_{o,g}|z_o, z_{p_1}, z_{p_2}]$ for each individual, and the environmental correlation between parents and offspring (Table S1-Table S7). For simulations in which we vary parameters other than C we manipulate C such that the average χ^2 for causal SNPs for GWAS is approximately the same as in the default parameter setting for comparative reasons. Using the same underlying genotypes and phenotypes as in the default parameter setting, we investigate the effect of misspecifying h_l^2 and K on the performance of LT-FH. When introducing environmental covariance we assume the following covariance structure between the non-genetic (environmental) components of the offspring (o) and the two parents (p_1, p_2):

$$\begin{pmatrix} \epsilon_{o,e} \\ \epsilon_{p_1,e} \\ \epsilon_{p_2,e} \end{pmatrix} \sim MVN_3 \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1-h^2 & \rho & \rho \\ \rho & 1-h^2 & 0 \\ \rho & 0 & 1-h^2 \end{pmatrix} \right), \tag{2}$$

where in simulations we let $\rho = 0.5 * (0.5h^2)$ (half of the genetic covariance between offspring and parents).

Note that in this scenario we are assuming the parents still share no environmental covariance.

UK Biobank data set

We analyzed genetic data from the UK Biobank[19] 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; Table S8) and ∼20 million imputed variants with MAF >0.1%[15]. We identified pairs of siblings (full siblings and MZ twins) as outlined previously[19,29]. We note that it is easy to distinguish sibling pairs from parent-child pairs using identity-by-descent (IBD). We identified parent-offspring pairs as individuals having a kinship in $(0.177, 0.354)$ and having proportion of SNPs with identity-by-state equal to 0 (IBS0) < 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 default 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 principal components[15].

We considered 12 diseases in UK Biobank for which family history of disease was available (Table 1); these 12 diseases were considered when GWAX was proposed[1]. These 12 diseases had a wide range of narrow-sense heritabilities (Table S9). We assigned case-control status for all genotyped individuals (Table S8). We assigned individuals with no ICD9/10 codes as controls. An alternative would be to assign missing phenotypes to individuals with no ICD9/10 codes, which reduces sample size while increasing disease prevalence (Table S29). We decided against this possibility, because we determined that it slightly but consistently reduces sample size times observed-scale SNP-heritability (a measure of total genetic signal[22]) and has very little impact on GWAS power (Table S30)

Parental history and sibling history information was available for all 12 diseases. Parental history information consisted of 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 information consisted of number of brothers and sisters + presence or absence of disease in the set of all siblings. Family history information is self-reported and thus there is a risk of misreporting, recall bias, or phenotype misclassification. For example, participants were asked about their family history of "diabetes" but we assume that this family history corresponds to that specifically of type II diabetes, consistent with previous studies[1]. To assess the accuracy of family history information we compute the correlation of self-reported family history between siblings. When computing the correlation of self-reported sibling history we restricted to concordant sibling pairs (e.g. both cases or both controls) and for sex-specific diseases (breast and prostate cancer) we restricted 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 is a control. We observed a moderately high concordance of self-reported family history among siblings (Table S19).

LT-FH in UK Biobank

There are 377 possible configurations of case-control status and family history of disease. There are 252 configurations when case-control status is available: 2 configurations of case-control status alone (no family history available); 4 of case-control status and one parent's disease status; 6 of case-control status and both parents' disease status; 40 of case-control status and sibling's disease status (1-10 siblings; at least one or none affected); 80 of case-control status, one parent's disease status, and sibling's disease status; 120 of case-control status, both parents' disease status, and sibling's disease status. There are 125 configurations when case-control status is unavailable: 2 of one parent's disease status; 20 of sibling's disease status; 3 of both parents' disease status; 40 of one parent's disease status and sibling's disease status; 60 of both parents' disease status and sibling's disease status.

We condition on known information to compute the posterior mean genetic liability. We computed $E[\epsilon_{o,g}|\cdot]$ (where $\cdot$ is all known disease status information) through Monte Carlo integration as described above. We allow for genotyped individuals and their siblings to have a different disease prevalence than their parents. We select 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 the published literature (Table S9). We used h^2 estimates from twin studies, but we note the existence of h^2 estimates based on family history[30].

For individuals who report having 0 siblings, more than 10 siblings, or who do not know or prefer not to answer how many brothers and/or sisters they have, we ignore sibling history and use parental history only for LT-FH; for all other individuals we incorporate both the number of siblings as well as whether or not at least one is affected in the computation of $E[\epsilon_g|\cdot]$. For sex-specific diseases our exclusion criteria remains consistent however we now incorporate the number of siblings of the relevant sex as well as whether or not at least one is affected in the computation of $E[\epsilon_g|\cdot]$.

In a secondary analysis, we modified the LT-FH method to downweight family history information based on its accuracy for each disease such that LT-FH$^{FHweighted}_{no-sib} = E(\epsilon_g|z) + \gamma[E(\epsilon_g|z, FH) - E(\epsilon_g|z)]$ where z was case-control status, FH represented parental history information (z_{p1}, z_{p2}), and γ was the accuracy of parental history estimated through sibling disease concordance rates (Table S19). Thus, if reported history was 0% accurate we assign the LT-FH phenotype to be $E(\epsilon_g|z_o)$ (z_o is case-control status of genotyped individual), if reported history was 100% accurate we assign the LT-FH phenotype as before, $E(\epsilon_g|z_o, z_{p1}, z_{p2})$, we use linear interpolation for intermediate values of accuracy.

GWAS and GWAX in UK Biobank

We implemented GWAS considering the case-control status of the genotyped individual only (Table S13). The prevalence ranged from 0.1% to 32% (Table 1). Due to the increased risk of a type I error in unbalanced case-control settings, for some diseases we selected a stringent MAF threshold when computing the number of independent loci (see below)[15,20]. However, one could also use a method such as SAIGE that controls type I error in unbalanced case-control settings[20].

We considered the case-control status of the genotyped individual and the disease history of parents and siblings when implementing GWAX (Table S13). The GWAX prevalence (i.e. prevalence of proxy cases) was generally more than double the parental prevalence and many times larger than the disease prevalence (see Table S31). For individuals with at least 1 sibling or who do not know or prefer not to answer how many brothers and/or sisters they have we assign using sibling and parental disease history; for individuals with 0 reported siblings we assign using parental disease history only (Table S13). For sex-specific diseases (breast and prostate cancer) the conditions are replaced with siblings of the relevant sex.

We considered the case-control status of the genotyped individual and the disease history of parents and siblings when implementing GWAX-2df (Table S13). We controlled for assessment center, genotype array, sex, age, age squared, and the first 20 principal components[15]. We restricted the GWAX-2df analysis to 672,292 genotyped SNPs to limit computational cost.

Assessing calibration in UK Biobank

We assessed the calibration of each method using stratified LD score regression (S-LDSC) attenuation ratio[15,21-23]. We used stratified LD score regression with the baselineLD (v1.1) model to compute the attenuation ratio, defined as (S-LDSC intercept -1)/(mean χ^2 - 1), for each set of association statistics[15,22,23]; the standard error of attenuation ratios is computed as s.e.(S-LDSC intercept)/(mean χ^2 - 1). Regression SNPs for S-LDSC are HapMap Project Phase 3 (HapMap3) SNPs with an INFO score > 0.9, MAF > 0.01 and $0 < P \leq 1$. If the intercept is less than 1 S-LDSC reports that the attenuation ratio is less than 0; in these scenarios we set the attenuation ratio equal to 0 but still compute the standard error of the attenuation ratio as s.e.(S-LDSC intercept)/(mean χ^2 - 1). We also compute the difference in attenuation ratios between various methods; the standard error of the difference is computed via block-jackknife.

We report two estimates for the mean attenuation ratio (and mean attenuation ratio difference) across all traits. Let x_i denote the attenuation ratio (or attenuation ratio difference) for trait i and σ_i the reported jackknife standard error (for the ratio or the difference of ratios). We compute a simple average across traits, with standard error computed as $\frac{1}{12}\sqrt{\sum_i \sigma_i^2}$. We also compute an inverse-variance

weighted mean where the weights are determined by the variance of the GWAS attenuation ratio. The inverse-variance weighted estimate is $\sum_i\{x_i/\sigma^2_{gwas,i}\}/\sum_i\{1/\sigma^2_{gwas,i}\}$, with standard error computed as $\sqrt{\sum_i\{\sigma^2_i/\sigma^4_{gwas,i}\}}/\sum_i\{1/\sigma^2_{gwas,i}\}$.

Assessing power in UK Biobank

We computed the number of independent loci as defined previously[15]. In detail, we applied PLINK's LD clumping algorithm[31] (see URLs) using LD computed in N=113,851 unrelated British individuals[32] at 9.6 million imputed SNPs with MAF>0.1% and INFO>0.6, employing a conservative genome-wide significance threshold of $p<5*10^{-9}$. We used a stringent 5Mb window and R^2 threshold of 0.01 for LD clumping. We then collapsed independent signals that were within 100 kb of one another (as assessed using the top SNP from a LD-clump) into a single locus.

When computing the number of independent loci (as well as average χ^2) we restrict to a MAF above a given threshold determined by prevalence in order to avoid an increased type I error in unbalanced case-control settings[15,20]. In the case of triallelic SNPs in which one variant allele has MAF $>$ 0.1% and one variant allele has MAF $<$ 0.1%, BOLT-LMM software retains association statistics for both variant alleles; we post-processed association statistics by manually removing all association statistics for variant alleles with MAF below a given threshold, thus removing the association statistic for the variant allele below the MAF threshold in these triallelic instances. The MAF threshold selected for rare diseases is a conservative choice for both GWAX and LT-FH as both of these methods have reduced risk of increased false-positives at lower MAF, due to higher case prevalence (or lower kurtosis for LT-FH) (Table S10).

The relative improvement for a method M versus a method m is defined as $\frac{\sum_t \#_{t,M}-\sum_t \#_{t,m}}{\sum_t \#_{t,m}}$ where $\#_{t,M}$ is the number of independent loci for trait t and method M; the standard error is computed via block jackknife of the genome (200 blocks).

Assessing heritability and correlation in UK Biobank

We estimate observed-scale SNP-heritability ($\hat{h}^2_{g,o}$) for each method by applying BOLT-REML[25] to unrelated European individuals. Observed-scale SNP-heritability is converted to liability-scale SNP-heritability using the in-sample disease prevalence[33]. Genetic correlation is estimated using BOLT-REML[25]. The phenotypic correlation and correlation of $-log_{10}(p)$ (across all variants with a minimum MAF determined by case prevalence; Table S10) is estimated in `R`. For both genetic and phenotypic correlation we restrict attention to individuals with non-NA values for both methods being compared.

Acknowledgements

We are grateful to L. O'Connor, O. Weissbrod, N. Zaitlen, G. Kichaev, and A. Gusev for helpful discussions. This research was funded by NIH grants R01 HG006399, R01 MH101244, R01 MH107649 and 5T32CA009337-32. This research was conducted using the UK Biobank Resource under Application #10438.

References

[1] Jimmy Z Liu, Yaniv Erlich, and Joseph K Pickrell. Case-control association mapping by proxy using family history of disease. *Nature Genetics*, 49:325–331, 2017.

[2] H.-C. So, J. S.H. Kwan, S. S. Cherny, and P.C. Sham. Risk prediction of complex diseases from family history and known susceptibility loci, with applications for cancer screening. *American Journal of Human Genetics*, 88:548–565, 2011.

[3] C.B. Do, D.A. Hinds, U. Francke, and N. Eriksson. Comparison of family history and snps for predicting risk of complex disease. *PLOS Genetics*, 8, 2012.

[4] R.E. Marioni, S.E. Harris, Q. Zhang, et al. Gwas on family history of alzheimer's disease. *Translational Psychiatry*, 8, 2018.

[5] T. Raj, Y.I. Li, G. Wong, et al. Integrative transcriptome analyses of the aging brain implicate altered splicing in alzheimer?s disease susceptibility. *Nature Genetics*, 50:1584–1592, 2018.

[6] I.E. Jansen, J.E. Savage, K. Watanabe, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing alzheimer's disease risk. *Nature Genetics*, 51:404–413, 2019.

[7] D. S. Falconer. The inheritance of liability to diseases with variable age of onset , with particular reference to diabetes mellitus. *Annals of Human Genetics*, 31:1–20, 1967.

[8] D. D. Campbell, P. C. Sham, A. J. Knight, H. Wickham, and S. Landau. Software for generating liability distributions for pedigrees conditional on their observed disease states and covariates. *Genetic Epidemiology*, 34:159–170, 2010.

[9] S. H. Lee, N. R. Wray, M. E. Goddard, and P. M. Visscher. Estimating missing heritability for disease from genome-wide association studies. *The American Journal of Human Genetics*, 88:294–305, 2011.

[10] N. Zaitlen, S. Lindstrom, B. Pasaniuc, M. Cornelis, et al. Informed conditioning on clinical covariates increases power in case-control association studies. *PLoS Genetics*, 8, 2012.

[11] O. Weissbrod, C. Lippert, D. Geiger, and D. Heckerman. Accurate liability estimation improves power in ascertained case-control studies. *Nature Methods*, 12:332–334, 2015.

[12] T.J. Hayeck, N.A. Zaitlen, P-R Loh, B. Vilhjalmsson, et al. Mixed model with correction for case-control ascertainment increases association power. *The American Journal of Human Genetics*, 96: 720–730, 2015.

[13] T.J. Hayeck, P-R Loh, S. Pollack, et al. Mixed model association with family-biased case-control ascertainment. *The American Journal of Human Genetics*, 100:31–39, 2017.

[14] Po-Ru Loh, George Tucker, Brendan K Bulik-Sullivan, Bjarni J Vilhjalmsson, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature Genetics*, 47:284–290, 2015.

[15] Po-Ru Loh, Gleb Kichaev, Steven Gazal, Armin P. Schoech, and Alkes L. Price. Mixed-model association for biobank-scale datasets. *Nature Genetics*, 50:906–911, 2018.

[16] P.M. Visscher, W.G. Hill, and N.R. Wray. Heritability in the genomics era – concepts and misconceptions. *Nature Reviews Genetics*, 9:255–266, 2008.

[17] J. Yang et al. Common SNPs explain a large proportion of heritability for human height. *Nature Genetics*, 42:565–569, 2010.

[18] P. Armitage. Tests for Linear Trends in Proportions and Frequencies. *Biometrics*, 11:375–386, 1955.

[19] C Bycroft et al. The uk biobank resource with deep phenotyping and genomic data. *Nature*, 562: 203–209, 2018.

[20] Wei Zhou, Jonas B Nielsen, Lars G Fritsche, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature Genetics*, 50:1335–1341, 2018.

[21] B. Bulik-Sullivan et al. LD Score Regression Distinguishes Confounding from Polygenicity in Genome-Wide Association Studies. *Nature Genetics*, 47:294–305, 2015.

[22] Hilary K Finucane, Brendan Bulik-Sullivan, Alexander Gusev, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics*, 47(11): 1228–1235, 2015.

[23] Steven Gazal, Hilary K Finucane, Nicholas A Furlotte, et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nature Genetics*, 49(10):1421–1427, 2017.

[24] S. Haworth et al. Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nature Communications*, 10, 2019.

[25] Po-Ru Loh, Gaurav Bhatia, Alexander Gusev, Hilary K Finucane, et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance components analysis. *Nature Genetics*, 47:1385–1392, 2015.

[26] P. Turley, R.K. Walters, et al. Multi-trait analysis of genome-wide association summary statistics using mtag. *Nature Genetics*, 50:229 – 237, 2018.

[27] K. Pearson. Mathematical contributions to the theory of evolution. xi. on the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society A. Mathematical, Physical, and Engineering Sciences*, 200:1–66, 1903.

[28] A.C. Aitken. Note on selection from a multivariate normal population. *Proceedings of the Edinburgh Mathematical Society B.*, 4:106–110, 1934.

[29] A Manichaikul et al. Robust relationship inference in genome-wide association studies. *Bioinformatics*, 26:2867–2873, 2019.

[30] M. Munoz, R. Pong-Wong, O. Canela-Xandri, et al. Evaluating the contribution of genetic and familial shared environment to common disease using the UK Biobank. *Nature Genetics*, 48:980–983, 2016.

[31] C.C. Chang et al. Second-generation plink: rising to the challenge of larger and richer datasets. *GigaScience*, 4:1–16, 2015.

[32] K. J. Galinsky, P.-R. Loh, S. Mallick, N. J. Patterson, and A. L Price. Population structure of uk biobank and ancient eurasians reveals adaptation at genes influencing blood pressure. *American Journal of Human Genetics*, 99:1130–1139, 2016.

[33] Sang Hong Lee, Naomi R. Wray, Michael E. Goddard, and Peter M. Visscher. Estimating Missing Heritability for Disease from Genome-wide Association Studies. *The American Journal of Human Genetics*, 88:294–305, 2011.

Table 1: **Overview of 12 diseases in the UK Biobank.** We report the prevalence, parental prevalence, liability-scale SNP-heritability, and number of samples analyzed by GWAS, GWAX and LT-FH. GWAX sample sizes are lower than GWAS and LT-FH sample sizes due to incomplete family history data. * 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 Table S9. 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.

Trait	UKBB Prevalence	UKBB Parental Prevalence	$\hat{h}_g^2(s.e.)$	N_{GWAS}	N_{GWAX}	N_{LT-FH}
AD	0.001	0.065	0.12 (0.11)	381493	324512	381493
PD	0.003	0.020	0.15 (0.05)	381493	318792	381493
Lung cancer	0.006	0.064	0.10 (0.03)	381493	323838	381493
Bowel cancer	0.013	0.054	0.14 (0.018)	381493	322887	381493
Stroke	0.023	0.14	0.08 (0.011)	381493	332468	381493
COPD	0.035	0.082	0.17 (0.0092)	381493	329495	381493
Prostate cancer *	0.037	0.075	0.30 (0.019)	175450	331458	368940
T2D	0.042	0.092	0.36 (0.0093)	380180	330267	381390
Breast cancer*	0.061	0.082	0.20 (0.012)	206043	348170	371064
Depression	0.073	0.052	0.11 (0.0058)	381493	328186	381493
CAD	0.083	0.25	0.20 (0.0058)	381493	341810	381493
HTN	0.32	0.26	0.31 (0.0037)	381493	354208	381493

Table 2: **Genetic and phenotypic correlations between GWAS, GWAX and LT-FH phenotypes.** 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 are constrained to be $\leq$1.00. Standard errors of genetic correlation estimates are reported in parentheses. Standard errors of phenotypic correlation estimates are $\leq$0.003 in each case.

Trait	Genetic Correlation (r_g)			Phenotypic Correlation		
	GWAS/ GWAX	GWAS/ LT-FH	GWAX/ LT-FH	GWAS/ GWAX	GWAS/ LT-FH	GWAX/ LT-FH
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

Figure 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 with family history of disease with disease cases; 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 panel) and then tests for association between genotype and posterior mean genetic liability (right panel).

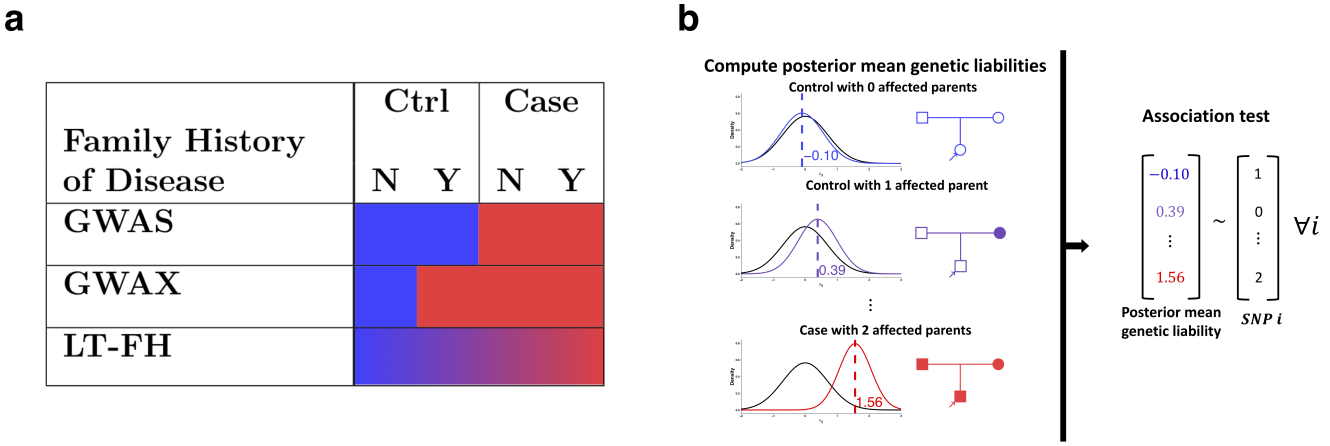

Figure 2: **LT-FH is well-calibrated and increases association power in simulations. (a)** Average χ^2 for null causal SNPs. **(b)** Average χ^2 for causal SNPs. **(c)** Power, defined as the proportion of causal SNPs with $p < 5 * 10^{-8}$. Results are averaged across 10 simulations with 100,000 SNPs (500 causal SNPs) per simulation. Error bars denote standard errors. Numerical results are reported in Table S1.

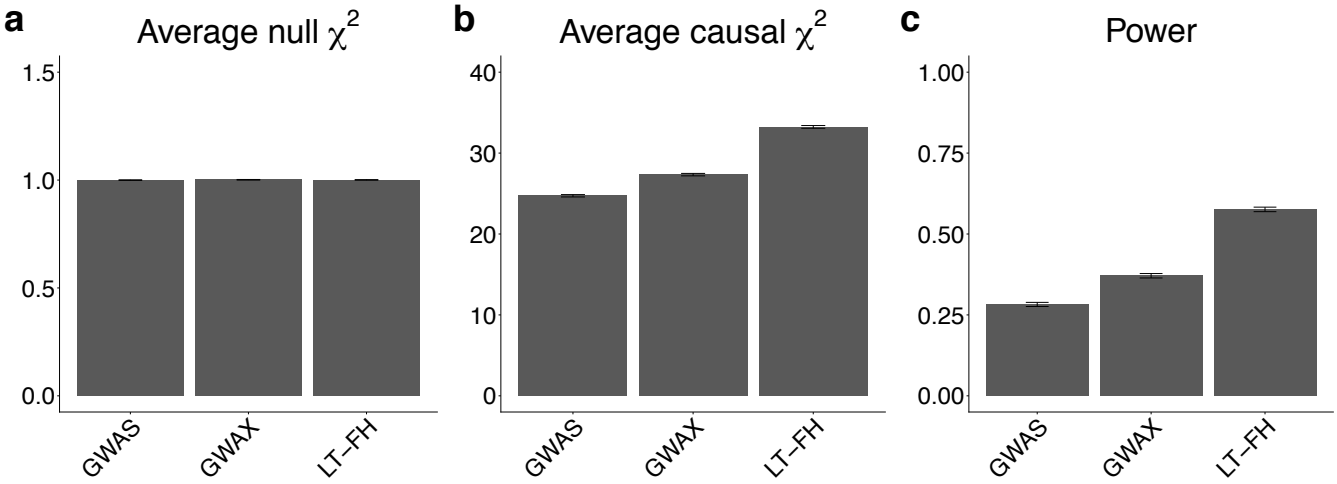
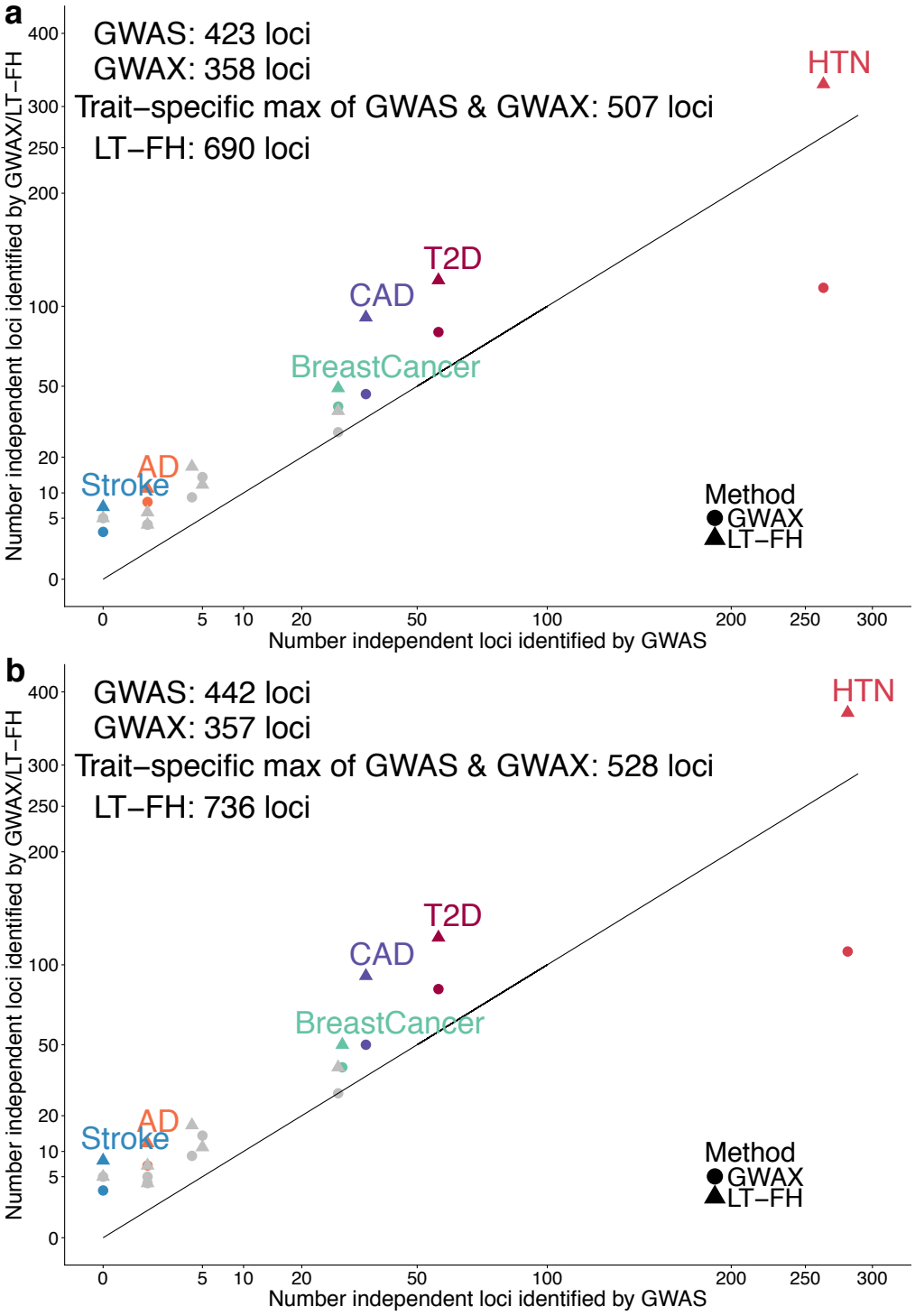
Figure 3: **LT-FH increases association power across 12 diseases from the UK Biobank.** We report results of GWAS, GWAX and LT-FH using either **(a)** linear regression or **(b)** BOLT-LMM on unrelated European individuals. Numerical results are reported in Table S14 and Table S25.