Fast variance component analysis using large-scale ancestral recombination graphs

Jiazheng Zhu[a,*], Georgios Kalantzis[a,b,*], Ali Pazokitoroudi[c], Árni Freyr Gunnarsson[a,d], Hrushikesh Loya[a,d], Han Chen[e,f], Sriram Sankararaman[g,h,i], Pier Francesco Palamara[a,d]

[a]*Department of Statistics, University of Oxford, Oxford, OX1 3LB, UK*
[b]*Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, CB10 1SA, UK*
[c]*Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, 02115, Massachusetts, USA*
[d]*Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK*
[e]*Human Genetics Center, Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, 77030, Texas, USA*
[f]*Center for Precision Health, McWilliams School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, 77030, Texas, USA*
[g]*Department of Computer Science, UCLA, Los Angeles, 90095, California, USA*
[h]*Department of Human Genetics, David Geffen School of Medicine, UCLA, Los Angeles, 90095, California, USA*
[i]*Department of Computational Medicine, David Geffen School of Medicine, UCLA, Los Angeles, 90095, California, USA*

Abstract

Recent algorithmic advancements have enabled the inference of genome-wide ancestral recombination graphs (ARGs) from genomic data in large cohorts. These inferred ARGs provide a detailed representation of genealogical relatedness along the genome and have been shown to complement genotype imputation in complex trait analyses by capturing the effects of unobserved genomic variants. An inferred ARG can be used to construct a genetic relatedness matrix, which can be leveraged within a linear mixed model for the analysis of complex traits. However, these analyses are computationally infeasible for large datasets. We introduce a computationally efficient approach, called ARG-RHE, to estimate narrow-sense heritability and perform region-based association testing using an ARG. ARG-RHE relies on scalable randomized algorithms to estimate variance components and assess their statistical significance, and can be applied to multiple quantitative traits in parallel. We conduct extensive simulations to verify the computational efficiency, statistical power, and robustness of this approach. We then apply it to detect associations between 21,374 genes and 52 blood-related traits, using an ARG inferred from genotype data of 337,464 individuals from the UK Biobank. In these analyses, combining ARG-based and imputation-based testing yields 8% more gene-trait associations than using imputation alone, suggesting that inferred genome-wide genealogies may effectively complement genotype imputation in the analysis of complex traits.

*Contributed equally

1. Introduction

Modern biobank initiatives have led to the collection of growing volumes of genomic and phenotypic information [45, 8, 67, 22, 32]. These datasets have been utilized in various applications, including the detection of phenotypic associations [73, 1], the study of complex trait architectures [71, 12], and the development of polygenic scores [25]. A key step in these analyses is to study the patterns of genetic and genealogical relationships among the analyzed samples, which provide insights into phenotypic and environmental variation. A wide range of approaches has been developed to analyze details of these relationships, ranging from methods that use genomic data to detect shared recent ancestry [18, 6, 51, 46, 60, 78] to approaches that reconstruct genome-wide genealogical relationships using an ancestral recombination graph (ARG) [57, 62, 24, 75].

Inferred genome-wide genealogies have predominantly been used to facilitate analyses in population genetics [57, 62, 24, 66], but have also recently been used for the study of complex traits [75, 58, 34]. One approach is to use an ARG inferred from a subset of common variants to detect the presence of unobserved genomic variation, which can then be tested for association with a phenotype of interest [43, 75]. The unobserved variation inferred from the ARG can also be used to construct a genetic relatedness matrix (GRM), which in turn can be used to estimate heritability, perform polygenic prediction, or detect associations within the framework of linear mixed models (LMMs). In this work, we adopt the terminology of [75], where ARG-derived GRMs were introduced for LMM-based complex trait analyses and referred to as ARG-GRMs. A contemporary approach [11] independently described a similar model for population genetic analyses, referred to as eGRM. Both models build on related work on the genealogical interpretation of principal component analysis [42].

When used to estimate heritability, ARG-GRMs have been shown to potentially lead to improved estimates by including the contribution of variants that would otherwise not be observed [75]. This strategy can also be utilized for variance component association testing [34] by constructing the GRM for a specific region of interest and assessing whether heritability significantly deviates from zero. By reducing the number of tests performed and pooling potentially heterogeneous effects within a region, as is often done in the analysis of rare variants [47, 68, 33, 53], this ARG-based variance component testing approach may lead to increased association power. However, estimating global or local heritability using the approaches of [75] and [34] requires generating and manipulating ARG-GRMs that contain $N \times N$ entries for N studied individuals, which becomes computationally infeasible in datasets comprising more than a few thousand individuals.

Here, we develop an algorithm, called ARG-RHE, that uses genealogical data to test for association within a genomic region and scales linearly with sample size, making it applicable to ARGs inferred from large biobank datasets. This approach employs fast randomized algorithms for moment-based estimation of narrow-sense heritability [69, 52] and methods for the approximation of tail probabilities to allow for significance testing [68, 38]. We extensively test the scalability, calibration, and statistical power of the ARG-RHE

algorithm using simulated data. We then apply this approach to detect associations between 52 quantitative blood-related traits and an ARG inferred using SNP array data for $337,464$ samples from the UK Biobank data set (UKBB; [8]). We find that the associations detected using ARG-RHE in these analyses are complementary to those detected by testing ∼54 million variants imputed from ∼65 thousand sequenced haplotypes from the Haplotype Reference Consortium (HRC) panel.

2. Methods

In this section, we first describe an extension to the RHE-mc algorithm [52] that allows estimating heritability for multiple traits simultaneously. We then describe the ARG-RHE approach, which enables the use of an ARG for computationally efficient variance component, region-based association testing for multiple traits in parallel. Finally, we describe the experimental setup used to benchmark these methods using simulated and real datasets.

We assume a linear relationship between the phenotype and the genetic and environmental components, $y = W\gamma + g + e$, where y is the phenotype vector for N samples, W denotes an $N \times c$ matrix of covariates, such as age and principal components (PCs) of ancestry, γ represents the corresponding fixed effects (including an intercept term), and g and e represent the random additive genetic and environmental components, respectively. We simplify the notation by assuming that the term used to represent covariates $W\gamma$ has been regressed out from the phenotype y and that y is standardized to have zero mean and unit variance. We treat g and e as random effects such that $g \sim \mathcal{N}(0, \sigma_g^2 R)$ and $e \sim \mathcal{N}(0, \sigma_e^2 I_N)$, where σ_g^2 and σ_e^2 capture the genetic and environmental variances, I_N is the $N \times N$ identity matrix, and R denotes the $N \times N$ genetic relatedness matrix (GRM). Assuming that g and e are uncorrelated, the sample variance-covariance is given by $V = \text{cov}(y) = \sigma_g^2 R + \sigma_e^2 I_N$.

2.1. Efficient ARG-based heritability estimation across multiple traits

We aim to infer the narrow-sense heritability (hereafter heritability), the fraction of phenotypic variance that may be ascribed to additive genetic effects [64], defined as $h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_e^2)$. RHE-mc [52] computes an efficient moment-based estimator of heritability, allowing to partition heritability across multiple variance components. Here, we extend this approach to allow it to simultaneously estimate heritability for multiple traits, decreasing computational costs.

Multiple genetic components can be utilized to estimate heritability, by extending the linear model for the phenotype to consider a partition of the genome into $K \geq 1$ components $g_1, g_2, ..., g_K$, of sizes $M_1, M_2, ..., M_K$ variants, respectively. Hence, we have $y = \sum_{k=1}^{K} g_k + e$, where $g_k \sim \mathcal{N}(0, \sigma_k^2 X_k X_k^\top / M_k)$, X_k is the $N \times M_k$ matrix of standardized genotypes for component k, $M_1 + ... + M_K = M$, $\sigma_1^2 + ... + \sigma_K^2 + \sigma_e^2 = 1$, and

$e \sim \mathcal{N}(0, \sigma_e^2 I_N)$. In the above, $R_k := X_k X_k^\top / M_k$ is used to estimate the GRM for the k-th component. In this case, the moment-based estimator satisfies the normal equation

$$\begin{bmatrix} H & b \\ b^\top & N \end{bmatrix} \begin{bmatrix} \tilde{\sigma}_g^2 \\ \tilde{\sigma}_e^2 \end{bmatrix} = \begin{bmatrix} c \\ y^\top y \end{bmatrix} \tag{1}$$

where $\tilde{\sigma}_g^2 = [\tilde{\sigma}_1^2, ..., \tilde{\sigma}_K^2]^\top$, H is a $K \times K$ matrix with elements $H_{k,l} = \mathrm{Tr}(R_k R_l)$, b is a vector with the K traces $\mathrm{Tr}(R_k)$, $k \in \{1, ..., K\}$, and c is a K-vector with elements $c_k = y^\top R_k y$. To avoid forming the GRMs, we calculate the required traces using Hutchinson's trace estimator [20, 69, 52] using B random vectors. Throughout this work, unless specified otherwise, the genotype matrix X contains column-centered standardized genotypes, which is equivalent to assuming an equal contribution from each SNP to the overall phenotypic variance. This commonly adopted assumption corresponds to choosing $\alpha = -1$ in the model introduced in [61], where the j-th variant with frequency f_j has effect size β_j of variance $\mathrm{Var}(\beta_j) \propto \left[f_j(1 - f_j)\right]^\alpha$. However, ARG-RHE also allows specifying arbitrary values of α, which corresponds to mean-centering the j-th variant in X and scaling by a factor proportional to $\left[f_j(1 - f_j)\right]^{\alpha/2}$ so that the GRM $R \propto XX^\top$ has $\mathrm{Tr}\, R = N$.

To perform these operations using an (implicit) ARG-GRM, we follow the approach of [75] and sample mutations from the ARG. In [75], these mutations were used to compute an approximation of the ARG-GRM, referred to as Monte Carlo ARG-GRM, which, when a sufficiently high mutation rate is used to generate mutations, yields results equivalent to using an exact ARG-GRM. Here, we use the above calculations to avoid explicitly forming the ARG-GRM. The genotype matrix X is also never stored in memory or saved to disk. Instead, multiplication operations are performed within the ARG, using efficient graph traversal algorithms. This type of traversal algorithm has been used in the past to efficiently generate mutations [50, 23] or to compute useful summaries of the ARG, such as haplotype sharing [50] or other statistics [56].

When analyzing multiple phenotypes without missing values, the left-hand side (LHS) of equation (1) is independent of the phenotype, allowing us to solve the above normal equations for any phenotype by reading the genotypes and forming the required quantities only once. In cases where several individuals have missing phenotypes, the above LHS is not the same, and we need to compute several phenotype-specific H and b from the genotyped samples. In practice, assuming that the intersection of non-missing samples across analyzed phenotypes is sufficiently large, we estimate H using all available genotype data, while reading the genotype data once. This approach has a computational cost of $\mathcal{O}\left(\frac{NMB}{\max(\log_3(N), \log_3(M))} + P(K^3 + NM) + K^2 NB\right)$, where K denotes the number of variance components, P is the number of traits, and B is the number of random vectors used for the estimation [52]. In contrast, the restricted maximum likelihood (REML) approach implemented in the GCTA software [72, 70] performs matrix operations that require $\mathcal{O}(N^2)$ memory and computational cost that scale at least quadratically with sample size.

2.2. *Efficient ARG-based variance component association testing*

To perform variance component association testing, we propose another extension so that in addition to obtaining an estimate of the variance explained by mutations that may be generated using the ARG within a region, we also efficiently assess whether this estimate is statistically significant. In the case of a single variance component, i.e., $K = 1$, and a phenotype y standardized to unit variance, solving equation (1) leads to a closed-form expression for the heritability estimator

$$\tilde{\sigma}_g^2 = \frac{y^\top (R - I_N) y}{\mathrm{Tr}(R^2) - N}. \tag{2}$$

We consider the case where R is a local GRM, constructed using variants within a region of interest. To test for association with the phenotype y, we focus on the distribution of the estimand $\tilde{\sigma}_g^2$ under the null hypothesis $H_0 : \sigma_g^2 = 0$, where $y \overset{H_0}{\sim} \mathcal{N}(0, I_N)$. The random part in the numerator of equation (2) gives the test statistic $T := y^\top R y$. After eigen-decomposing $R = \Sigma^\top \Lambda \Sigma$, where $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_N)$ holds the eigenvalues of R in decreasing order and columns of Σ store the corresponding eigenvectors, we have

$$\Pr(T < t \mid H_0) = \Pr\left(y^\top R y < t \mid H_0\right) = \Pr\left((y^\top \Sigma^\top)\Lambda(\Sigma y) < t \mid H_0\right) = \Pr\left(\sum_{i=1}^{N} \lambda_i \chi_{1,i}^2 < t\right), \tag{3}$$

where $y \sim \mathcal{N}(0, I_N)$ is invariant under the action of Σ, and $\chi_{1,i}^2$ are i.i.d. chi-squared variables with one degree of freedom. This test statistic coincides with the score-based test adopted in methods such as the sequence kernel association test (SKAT) [68], which is instead derived by considering the derivative of the restricted likelihood at $\sigma_g^2 = 0$. We note that, while it is also possible to efficiently estimate the standard error of $\tilde{\sigma}_g^2$ [69], using a Wald test on H_0 requires T to be asymptotically normally distributed. However, since T in equation (3) has significant skew and kurtosis, convergence to a normal distribution depends on the number of non-zero λ_i. This is bounded by the number of variants in a gene region and is usually not sufficient to apply large-sample asymptotics [63]. Hence, to obtain a tail probability that is precise enough for association testing but also computationally efficient, we follow the strategy recently proposed in fastSKAT [38], which relies on a mixture of moment-based and saddle-point approximations. In more detail, we first perform randomized matrix sketching [19] to estimate $\lambda_1, \dots, \lambda_L$ as the top-L leading eigenvalues of Λ. We then approximate the right-hand side of equation (3) using a truncated sum of $L + 1$ chi-squared variables, which take the form

$$\tilde{T} = \sum_{i=1}^{L} \lambda_i \chi_{1,i}^2 + a \chi_d^2, \tag{4}$$

where

$$a = \frac{\sum_{i=L+1}^{N} \lambda_i^2}{\sum_{i=L+1}^{N} \lambda_i} = \frac{\mathrm{Tr}\, R^2 - \sum_{i=1}^{L} \lambda_i^2}{\mathrm{Tr}\, R - \sum_{i=1}^{L} \lambda_i},$$

$$d = \frac{\left(\sum_{i=L+1}^{N} \lambda_i\right)^2}{\sum_{i=L+1}^{N} \lambda_i^2} = \frac{\left(\mathrm{Tr}\, R - \sum_{i=1}^{L} \lambda_i\right)^2}{\mathrm{Tr}\, R^2 - \sum_{i=1}^{L} \lambda_i^2},$$

are scalars that satisfy the moment-matching conditions $\mathbb{E}[T] = \mathbb{E}\left[\tilde{T}\right]$ and $\mathbb{E}\left[T^2\right] = \mathbb{E}\left[\tilde{T}^2\right]$. We estimate these quantities using the leading eigenvalues and Hutchinson's estimator for the trace of powers of the GRM R. Next, we estimate association p-values via a less accurate but fast moment-based estimation approach [36]. If the p-value is below a specified threshold, which we set to 0.01, we proceed to estimate the tail probability more accurately through a saddle-point approximation technique [31]. This approach is designed to avoid the numerical instability of saddle-point approximation near the mean of the distribution. As done in the computation of point estimates for heritability, we use efficient graph traversal algorithms to directly perform these computations within the ARG.

2.3. Simulation experiments

We evaluated ARG-RHE by testing its scalability, the accuracy of its h^2 and p-value estimates, and its power when applied to perform variance component association analysis. To this end, we generated synthetic ARGs spanning 10 kilobase (kb) and 10 megabase (Mb) gene windows and containing between 5,000 and 500,000 diploid individuals. Synthetic ARGs were generated using msprime v1.3.0 [23] under a constant effective population size $N_e = 50{,}000$ and recombination rate 10^{-8}. We used these simulated ARGs to generate sequencing data by sampling mutations at a rate of 10^{-8} and used them to generate heritable phenotypes, varying their polygenicity by designating 10% to 80% of variants as causal. We then sampled each effect size $\beta \sim \mathcal{N}(0, [f(1-f)]^{\alpha})$ independently and added a normal random noise term to obtain the desired total heritability h^2. Our simulations assumed the availability of an accurate ARG, focusing on quantifying any differences between the use of ARG-RHE, which relies on highly scalable approximations, to GCTA-REML, which is a more accurate but slower likelihood-based approach (see Methods). Simulation analyses that compare the use of true and inferred ARGs, as well as genotyped, imputed, and sequenced variants, may be found in [75] (Figure 3, Extended Data Figure 6) and [34] (Figure S2).

To work with Monte Carlo ARG-GRMs, we discarded the previous mutations from the ARG and resampled a new set of mutations at a rate of 10^{-6} for 10-kb windows and 10^{-8} for 10-Mb windows, obtaining a genotype matrix X, which we zero-centered and normalized by $[f(1-f)]^{\alpha/2}$ using allele frequencies f. Using this approach, the ARG-GRM can be written as $R = X^\top X/M$, where M is a normalizing factor such that $\mathrm{Tr}\, R = N$, and used to calculate p-values of the test statistic $T = y^\top R y$. In some experiments involving GCTA, we explicitly formed the ARG-GRM after writing these mutations to disk. When running the ARG-RHE algorithm, the mutations remained stored in the ARG and were used to obtain heritability estimates and assess significance, as previously described. We used 200 random vectors to compute the Hutchinson's trace estimator and the top 200 leading eigenvalues obtained from the randomized SVD.

2.3.1. Heritability estimation accuracy

To assess the accuracy of the heritability estimates, we generated ARGs containing 20,000 individuals and spanning 10 kb and 10 Mb gene windows. We then generated heritable traits with h^2 ranging from 0 to 0.8. We varied polygenicity from 10% to 80% and the MAF-dependency parameter α from -1 to 0.5 (Supplementary Table 4). We generated 100 independent simulations for each combination of parameters. When estimating h^2 using ARG-RHE, we assumed the correct α used to generate the phenotypes.

We also performed experiments in which we estimated σ_g^2 using time-stratified GRMs. For these analyses, we simulated 10-Mb windows on 20,000 individuals as above. We additionally partitioned the resampled genotypes into 6 GRMs containing an equal number of mutations based on their age quantiles, and jointly estimated the σ_g^2 across time bins using multi-component ARG-RHE. We generated heritable traits with $h^2 = 0.4$ and effect sizes α ranging from -1 to -0.25, and formed the ARG-GRMs using a fixed $\alpha = -1$.

2.3.2. Computational scalability experiments

To assess computational efficiency, we simulated up to 500,000 individuals as previously described and compared the costs of running ARG-RHE to the use of GCTA to estimate heritability with an explicitly constructed ARG-GRM, as done in [75] and [34]. When comparing the use of GCTA to the ARG-RHE algorithm, we used the same set of resampled mutations to explicitly construct the ARG-GRM using GCTA (v1.94.1) or to run the ARG-RHE algorithm, which is implemented in the arg-needle-lib library (see URLs). We measured the computational time required to run GCTA on pre-computed ARG-GRMs, excluding the time required to construct such ARG-GRMs, and compared with the time required to compute heritability and test for significance using ARG-RHE. Unlike ARG-RHE, GCTA does not support analyzing multiple phenotypes in parallel, so in these benchmarks, we tested ARG-RHE with a single phenotype, as well as with 50 random phenotypes processed in parallel. All tests were performed using Intel(R) Xeon(R) 2.40GHz CPUs.

2.3.3. Assessing p-value calibration and statistical power

We performed additional simulations to verify that variance component testing using ARG-RHE results in sufficiently calibrated p-values. To this end, we obtained exact quantiles for the test statistic T at p-values ranging from 0.05 to 10^{-8} for simulations comprising 5,000 individuals. We used Davies' method [9], obtaining a complete SVD of R to verify that the randomized approach used by ARG-RHE produces accurate p-values for a range of significance levels. For each quantile, we tested 20 independently generated ARGs each with 5 sets of random trace estimator seeds, totaling 100 estimators at each p-value threshold.

We compared the statistical power of ARG-RHE testing, which uses a moment-based approach, to that of GCTA, which relies on a more accurate but less scalable likelihood-based approach. We ran GCTA using two sets of variants for the GRM. In one case, we computed the local GRM using the actual set of sequencing variants used to simulate the phenotype. This simulation provides an upper bound to the statistical power

that can be achieved using ARG-RHE, which relies on a different set of resampled mutations. In the other case, we explicitly constructed a Monte Carlo ARG-GRM, using the same set of resampled mutations used by ARG-RHE as input GRM for GCTA to test for association. This simulation is meant to assess the extent to which the approximate randomized approach of ARG-RHE results in decreased association power compared to GCTA's likelihood-based approach. We assessed the statistical power at two significance levels: $s = 0.05$ and 2.5×10^{-6}, which is a Bonferroni-corrected threshold assuming 20,000 gene regions. We performed 100 repetitions for each combination of parameters.

2.4. Analysis of UK Biobank data

We applied ARG-RHE to perform variance component association testing using an ARG inferred for 337,464 unrelated white British individuals from the UK Biobank dataset [8]. The ARG was inferred using Threads [17] (v0.1) applied with default parameters to 798,103 genotyped variants, each with less than 10% missing values, which were computationally phased using Beagle (v5.1) [7]. We tested for association between 21,159 protein-coding and non-coding RNA regions from the HUGO Gene Nomenclature Committee (HGNC) table on the University of California, Santa Cruz (UCSC) genome browser (see URLs) and 52 quantitative blood cell traits, including 27 blood cell indices and 25 blood biochemistry marker levels. The lists of tested regions and traits are provided in Supplementary Tables 1 and 2.

We compared the results obtained using ARG-RHE to those obtained by applying the same variance component association test of equation (4) to genotype dosages imputed from the Haplotype Reference Consortium (HRC) reference panel, as performed in [8]. We included HRC-imputed variants with a minimum allele count (MAC) ≥ 5, INFO score ≥ 0.3, missingness $\leq 10\%$, and Hardy-Weinberg equilibrium $p \geq 10^{-15}$, leading to a total of 54,356,141 variants tested for association. We refer to analyses computing the test statistic of equation (4) using HRC-imputed data as HRC-RHE. We also performed standard single-variant association testing for the same set of ∼54 million HRC-imputed genotypes using Regenie. We refer to these single-variant analyses of HRC-imputed data as HRC-SV. Finally, we also performed ACAT-V gene-based testing [37] by combining HRC-SV p-values from a region using uniform weights. We set the genome-wide significance threshold to be $0.05/21159 \approx 2.36 \times 10^{-6}$ for gene-based tests, and 10^{-9} for single-variant tests.

We compared the signals from both gene-based and single-variant tests by grouping p-values from the tested regions into approximately independent LD blocks, using LDetect [4]. We then checked whether each block-trait pair contains genome-wide significant signals. We also considered a simple joint test that uses both HRC-RHE and ARG-RHE; for each gene region, we took the smaller p-value $p_{\min}$ from the two methods and treated them as independent tests by comparing $2p_{\min}$ to the same genome-wide significant threshold. When running ARG-RHE, we used $L = 200$ leading eigenvalues in estimating the tail distribution and $B = 200$ random vectors in Hutchinson's trace estimator.

To preprocess the analyzed traits, we followed the approach of [2]. We first stratified samples based on sex and menopause status and applied a rank-inverse-normal transformation (RINT) to the adjusted phenotype values. We regressed out covariates, which included age, age squared, alcohol use, smoking status, height, body mass index, assessment center, genotyping array, and 20 principal components. We then applied RINT a second time and merged all strata. Missing phenotypes were mean-imputed as 0.

We verified the associations we detected by checking whether they were previously reported on the Open Targets platform (version 22.10) [15, 44]. Specifically, for each significant gene-trait association, we used Open Targets' locus-to-gene pipeline to check for associated variants from previous studies involving the same gene. We considered an LD block-trait pair as previously detected if the LD block contained at least one such associated gene for the trait in Open Targets. For this analysis, we matched our 52 quantitative traits with 495 traits having a closely related description available on Open Targets (Supplementary Table 3).

3. Results

3.1. Scalability, accuracy, and power of ARG-RHE

We performed extensive simulations to verify the scalability, accuracy, *p*-value calibration, and statistical power of variance component analyses using ARG-RHE. We first verified the scalability of ARG-RHE under large sample sizes by performing association in up to 500,000 individuals. We compared ARG-RHE to an approach based on first computing an ARG-GRM and then providing it to GCTA to estimate heritability using restricted maximum likelihood (GCTA-REML), as done in [75] and [34] (see Methods). The results of these analyses are shown in Figure 1. As expected, GCTA-REML, which involves manipulating dense ARG-GRM matrices containing N^2 entries for N individuals, required substantially more computation and memory than ARG-RHE. As an example, testing an ARG region for association using 20,000 individuals with GCTA-REML required ∼300× more time and ∼24× more memory. The gap widened as the sample size increased, with an extrapolated difference of four orders of magnitude for an analysis involving 500,000 individuals. In addition, we verified that our extension of the RHE-mc algorithm allows for efficiently estimating heritability across multiple traits in parallel, requiring negligible additional computation to analyze 50 traits rather than a single trait, as shown in Figure 1.

We next evaluated the accuracy of using ARG-RHE to infer heritability across 96 genetic architectures simulated using different parameters (see Methods). The results of these analyses are shown in Figure 2 and Supplementary Tables 4 and 5. Applied to 10 Mb gene windows, ARG-RHE yielded accurate estimates, with an estimated relative bias ranging from −3.6% to 1.0%, which is consistent with results observed when applying the RHE algorithm to genotype data [52]. On smaller 10 kb gene windows with simulated h^2 up to 0.02, ARG-RHE yielded slightly conservative estimates of heritability, with a relative bias ranging from −2.6% to −19.4%, depending on the α parameter, which governs the relationship between allele frequency

and effect sizes [61, 59]. A small downward bias in ARG-based heritability estimates may also be observed if the mutation rate used when resampling mutations to work with Monte Carlo ARG-GRMs is not sufficiently large [75].

In addition to representing information on the genotypes of the analyzed samples, the ARG contains information on the age of the analyzed variants and can therefore be used to assess the phenotypic contribution of variants originating at different time scales. These analyses could allow estimating the strength of natural selection on variants linked to the trait [40, 55, 30, 27] or help quantify the relative phenotypic contribution of variation that is yet to be sequenced. To verify that ARG-RHE allows studying time-stratified variance components, we simulated phenotypic data using MAF-dependent architectures under different values of α, corresponding to different effect size distributions for variants of different allele frequencies (see Methods). Due to the link between the frequency and age of an allele [28, 16], these simulated architectures also imply different effect size distributions for variants of different ages, with larger negative values of α leading to larger effect sizes for rarer and more recent variants. These trends were captured by ARG-RHE when we applied it to jointly estimate multiple σ_g^2 values using implicit time-stratified GRMs (see Methods), as shown in Supplementary Figure S1.

We next assessed the calibration of p-values obtained using ARG-RHE when performing variance component association testing. We compared the p-values from ARG-RHE, which are based on a fast randomized algorithm (see Methods), to the exact p-values obtained using Davies' method. To this end, we assessed the relative error between exact and approximate log-p-values, $(-\log_{10} p_{\text{approx}} + \log_{10} p_{\text{exact}})/(-\log_{10} p_{\text{exact}})$, shown in Figure 2(b). Consistent with previous applications of this class of randomized approaches [38], we found a small inflation in the mean estimated log p-value, particularly at more extreme p-values, which corresponds to the exact critical values being estimated to be positioned around 0.93 times their nominal tail quantile.

Finally, we compared the statistical power of performing variance component association testing with ARG-RHE and GCTA-REML. Results for varying polygenicity and significance settings (see Methods) are shown in Figure 3. We first considered an analysis in which GCTA-REML is applied to the true causal variants used to simulate the trait. As the use of ARG-GRMs involves working with an estimate of the expected genetic similarity between individuals [75, 34], this setting provides an upper bound to the accuracy that can be achieved using ARG-RHE or GCTA-REML applied to an ARG-GRM. As expected, and consistent with previous observations [75, 34], having access to the true causal SNPs yielded higher power than relying on an ARG-GRM built using resampled mutations, with differences between true and resampled mutations depending on the simulation settings. We next aimed to assess the extent to which using fast randomized algorithms results in a drop in association power compared to GCTA's likelihood-based approach. To this end, we compared ARG-RHE and GCTA-REML applied to the same set of mutations resampled from the

ARG, which are distinct from the underlying causal variants (see Methods). For simulations with a MAF-dependency parameter $\alpha = 0.05$, we observed minimal differences between ARG-RHE and GCTA-REML. At $\alpha = 2.5 \times 10^{-6}$, ARG-RHE achieved a statistical power of around 60% relative to GCTA-REML for h^2 up to 0.02. Previous studies have observed a similar difference in statistical power between kernel-based score tests, such as in SKAT, compared to the likelihood ratio test used in GCTA [35].

Overall, ARG-RHE provided substantial gains in computational speed and memory requirements compared to GCTA-REML, while providing good accuracy and well-calibrated p-values. The high computational scalability of this approach comes at the cost of a slight reduction in association power compared to exact likelihood-based strategies, which is consistent with previous applications of randomized approaches to variance component association testing. However, this decrease in power is offset by the possibility of applying ARG-RHE to datasets containing orders of magnitude more individuals.

3.2. Analysis of a UK Biobank ARG

We applied ARG-RHE to an ARG comprising 337,464 unrelated white British individuals from the UK Biobank dataset, which was inferred using 798,103 genotyped variants. We leveraged this ARG to test for association between 21,159 protein-coding and non-coding RNA regions and 52 quantitative blood cell counts and serum biomarker level traits (see Methods). We compared the results of ARG-RHE to those obtained by applying the same variance component association test to ∼54 million variants imputed using the HRC reference panel (HRC-RHE). We also compared to associations detected by running Regenie to perform single-variant testing using the same set of HRC-imputed variants (HRC-SV), and to using the ACAT-V algorithm to combine single-variant association statistics from imputed data (see Methods). Results of these analyses are shown in Figure 4. The computing resources required to apply HRC-RHE and ARG-RHE to these UK Biobank datasets were lower than those required by highly optimized GWAS approaches. Applying HRC-RHE to test 52 traits across 21,159 regions spanning ∼24 million imputed variants required 299.6 CPU hours, ARG-RHE applied to the UK Biobank ARG required 601.2 CPU hours, and applying Regenie to individually test ∼54 million variants required 2,369 CPU hours, without including step 1 of the analysis (see Figure 4(a)).

We first considered the sets of approximately independent genomic regions [4] in which gene-trait associations were detected using each method, reported in Figure 4(b). Taken individually, running ARG-RHE, which relies on an ARG inferred using only genotype array markers, identified 10,374 unique locus-trait associations, while HRC-RHE and ACAT-V, which rely on imputed data, identified 12,162 and 13,720, respectively. Consistent with recent analyses [75, 34], however, we found association signals detected using the ARG to complement signals detected using imputation alone. Among the signals identified using ARG-RHE, 1,858 (18%) were not found by HRC-RHE, 1,376 (13%) were not found by ACAT-V, 1,770 (17%) were not found by HRC-SV, and 1,054 (10%) were not found by either HRC-RHE or ACAT-V (Figure 4 and Supplementary Figure S2). Performing a joint test using both imputed variants and the inferred ARG (see Methods) yielded

an increase in the number of associations. Combining ARG-RHE with HRC-RHE led to 13,190 associations, 8.5% more than only using HRC-RHE; ARG-RHE with ACAT-V led to 14,401 associations, 5% more than ACAT-V alone; while combining all three tests led to 14,831 associations, 21.9% and 8.1% more than only using HRC-RHE and ACAT-V, respectively. The full list of all genome-wide significant associations found by ARG-RHE and estimated heritability $\widehat{h}^2$ are provided in Supplementary Table 6.

We sought to validate the associations detected using ARG-RHE by looking for overlap with associations from other studies reported in the Open Targets database [49] (see URLs), which often involve larger sample sizes. Overall, we found 9,212 of the 10,374 regions associated with a trait by ARG-RHE to be reported on the Open Targets platform (see Methods). These include established signals found using ARG-RHE but not detected in our analysis of this subset of UK Biobank samples using HRC-RHE, HRC-SV, or ACAT-V, such as the association between RNF168 with red blood cell count [26], and the association of SLC4A4 with neutrophil count [2]. Of the remaining 1,162 signals that are not present in Open Targets, several have been linked to related traits in other studies. Examples include associations found on Chromosome 6 with lipoprotein A (Lp(a)), shown in Figure 5(a). Among the highly significant genes detected using ARG-RHE, SYNE1 has reported association with Lp(a) levels in the Hutterites, a founder population of South Dakota [48]. Associations with ESR1 have been reported in [29], and the RGS17 gene was significantly associated using exome sequencing data [2]. Other signals uniquely identified using ARG-RHE in our analyses and not reported in Open Targets include loci previously linked to traits that are related to those we have found to be associated. An example is the associations found on Chromosome 4 with serum urate levels, shown in Figure 5(b). The LINC00504 non-coding RNA and the CPEB2 gene were found to be associated with gout in a recent study involving ∼2.6 million individuals [39] and the CD38 gene was found to be associated with gout in another study [65].

Overall, the ARG-RHE algorithm was scalable enough to be applied to a genome-wide ARG inferred for 337,464 UK Biobank individuals using only genotyping array variants, and allowed detecting associations that complemented those that would be identified using the same variance component test applied to HRC-imputed data alone. Signals uniquely detected using ARG-RHE in these analyses were often implicated by studies that involved different cohorts or closely related traits. Further applications of ARG-RHE to non-European UK Biobank subgroups are described in [17].

4. Discussion

We developed ARG-RHE, an approach that builds on recent work on moment-based estimation of narrow-sense heritability [52] and on randomized algorithms to approximate tail probabilities [68, 38] to enable highly scalable heritability estimation and variance component association testing using an ARG. In simulations, ARG-RHE was several orders of magnitude faster than the use of explicitly computed ARG-GRMs to esti-

mate heritability [75] and to perform variance component association testing [34]. We verified ARG-RHE's accuracy and calibration, finding it to have only modestly reduced association power compared to using explicit ARG-GRMs coupled with a restricted maximum likelihood approach, which cannot be applied to large samples. Finally, we applied ARG-RHE to an ARG inferred for 337,464 UK Biobank individuals using only genotyping array variants, detecting associations that complement signals identified using the same variance component test or single-variant linear mixed model testing applied to ∼54 million HRC-imputed variants.

Our study contributes to a growing body of work leveraging inferred genome-wide genealogies for the analysis of complex traits [75, 58, 34]. One key advantage of using the ARG in this setting is that it provides a route to account for genomic variants that cannot be observed or accurately imputed. This is particularly the case in populations that are poorly represented in sequencing reference panels used for genotype imputation [75, 34], as also evidenced by further applying ARG-RHE to non-European UK Biobank individuals in [17]. Compared to testing individual edges of an inferred ARG, as done in [75], the use of a variance component test improves association power by pooling the effects of multiple variants and reducing the statistical burden by reducing the number of performed tests [34]. Our work substantially improves the computational efficiency of ARG-based variance component analyses of heritability and association, making this approach easy to apply to significantly larger samples.

Although our UK Biobank applications focused on association testing, our simulations show that ARG-RHE is scalable enough to infer genome-wide, rather than local, narrow-sense heritability. By leveraging information on allele ages contained in the ARG, it may also be used to gain insights into time-stratified phenotypic variance components. Natural selection acting on variants linked to polygenic traits often leads to larger per-allele effect sizes for variants of lower frequencies and younger ages [40, 55, 30, 27], with stronger selection leading to a larger overall phenotypic contribution from rare and recent variation [74, 59]. ARG-based, time-stratified analysis may therefore quantify the strength of selection on variants across different time strata, complementing models that link selection and allele frequencies [54, 61, 74, 59], as well as previous time-stratified approaches limited to higher frequency variation [14, 13]. These analyses may provide insights into the fraction of phenotypic variance explained by rare and recent alleles, which may not yet have been sequenced and are more likely to be population-specific. However, we caution that obtaining unbiased heritability estimates is more challenging than testing for significance, as it imposes additional requirements on the accuracy of the inferred ARG.

We highlight several avenues for future development and limitations of this work. First, ARG-RHE currently only supports quantitative traits, but it could be extended to support the association of rare variants with binary traits that have low case/control ratios [76, 41, 21]. Second, ARG-RHE implements a simple form of over-dispersion testing. Future work may involve implementing different strategies for weighting and pruning markers to improve power under various assumptions about the trait's architecture [37, 5, 79, 10, 77, 3]. Third,

the multi-trait extension of the RHE-mc algorithm, which we use to decrease computational costs, relies on a single estimate of $\boldsymbol{H}$ and b for equation (1), assuming that the intersection of individuals with non-missing phenotype data is large. This approach may be extended to obtain different $\boldsymbol{H}$ and b estimates for each trait. Despite these limitations, ARG-RHE provides a highly scalable approach for leveraging an inferred ARG to complement the use of genotype imputation in variance component analyses.

5. URLs

arg-needle-lib library, implementing ARG-RHE, `https://palamaralab.github.io/software/argneedle`
GENIE library, implementing multi-trait RHE-mc for genotype data, `https://github.com/sriramlab/GENIE`
Open Targets database, `https://genetics.opentargets.org`
UCSC genome browser, `https://genome.ucsc.edu/cgi-bin/hgTables`

6. Competing interests

A.F.G. is currently employed by deCODE genetics.

7. Acknowledgments

This research was conducted using the UK Biobank Resource (application 43206). We thank the participants of the UK Biobank project. This work was supported by the EPSRC Centre for Doctoral Training in Health Data Science (EP/S02428X/1, to J.Z.); ERC Starting Grant 850869 (to P.F.P. and G.K.); EPSRC and MRC grant EP/L016044/1 (to G.K.); the Keble College de Breyne Clarendon Scholarship (to A.F.G.); Wellcome Trust grant 222336/Z/21/Z (to A.F.G.); the Clarendon Scholarship (to H.L.); Wellcome Trust Studentship (to H.L.); NSF CAREER 1943497 and R35GM153406 (to S.S.). Computation used the Oxford Biomedical Research Computing (BMRC) facility, a joint development between the Wellcome Centre for Human Genetics and the Big Data Institute supported by Health Data Research UK and the NIHR Oxford Biomedical Research Centre. Financial support was provided by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

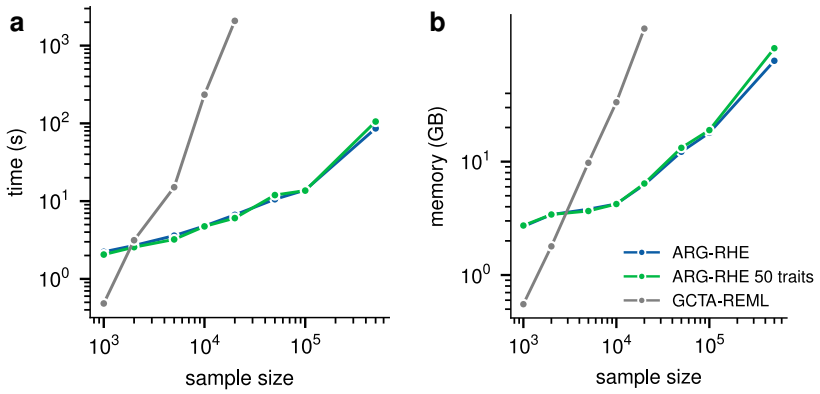

Figure 1: Scalability of ARG-RHE. (a) Computing time and (b) memory required for testing a single trait or 50 traits in parallel on a simulated 10 kb window for sample sizes up to 500,000 individuals. Both ARG-RHE and GCTA-REML used random mutations generated from the ARG at a rate of $\mu = 10^{-6}$. All methods were run on a single thread.

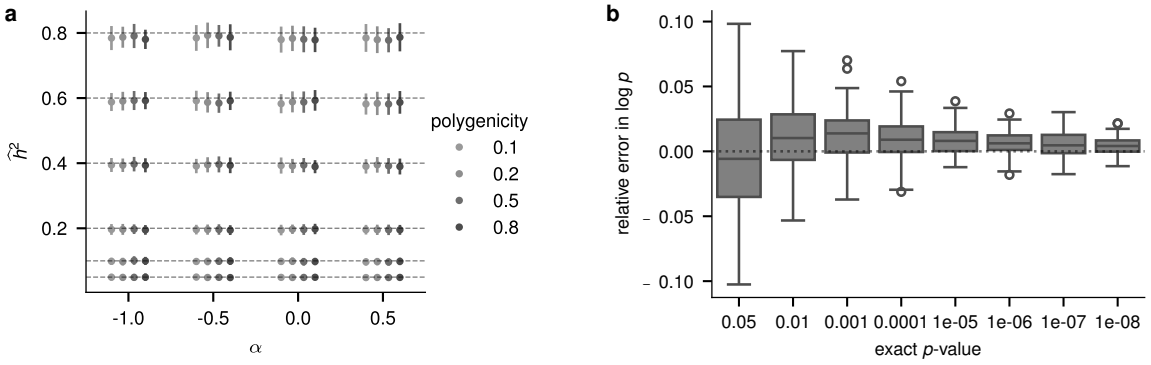

Figure 2: Accuracy of estimated heritability $\widehat{h}^2$ and p-values using ARG-RHE. (a) Estimated heritability $\widehat{h}^2$ over a range of polygenicity levels, true h^2 (shown as dashed lines), and MAF-dependency parameter α. Heritability was estimated for a 10-Mb region and 20,000 samples, resampling mutations at a rate of $\mu = 10^{-8}$. Error bars show standard deviations estimated using 100 independent repetitions. (b) Distribution of the relative error in estimated p-values $(-\log_{10} p_{\text{approx}} + \log_{10} p_{\text{exact}})/(-\log_{10} p_{\text{exact}})$ corresponding to exact p-value quantiles, using $L = 200$ leading eigenvalues across 100 random samples. p-values were estimated for a 10-kb region and 5,000 samples, resampling mutations at a rate of $\mu = 10^{-6}$.

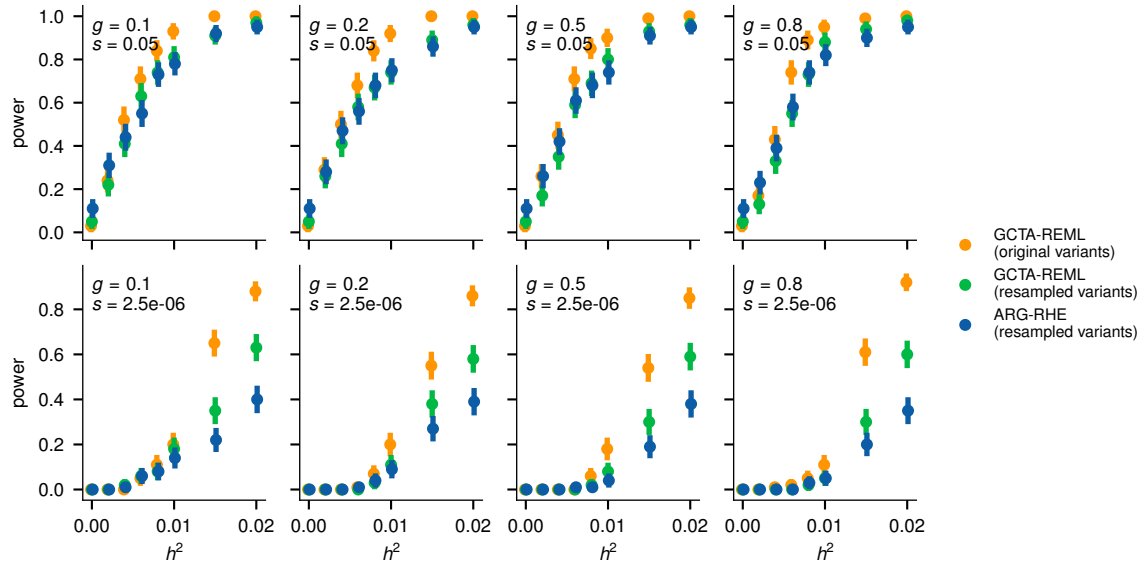

Figure 3: Statistical power of ARG-RHE. We computed power for a 10 kb gene window under different polygenicity levels g (fraction of causal variants among all sequencing variants simulated at rate 10^{-8}), local heritability h^2 (fraction of phenotypic variance explained by causal variants), and significance threshold s. We performed variance component association using either GCTA-REML applied to all sequencing variants, GCTA-REML applied to variants resampled from the ARG at a rate of 10^{-6}, and ARG-RHE applied to the same resampled variants. Error bars show standard errors estimated from 100 independent repetitions.

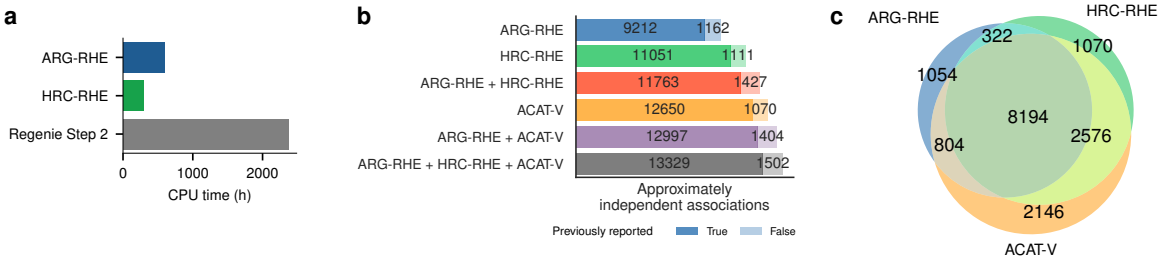

Figure 4: UK Biobank association analyses. (a) Total CPU time taken to test 52 traits for 337,464 unrelated white British individuals from the UK Biobank dataset. Regenie Step 2 processed all ∼54 million imputed variants, while HRC-RHE processed a set of ∼25 million variants overlapping any considered gene region. Both the ARG-RHE and HRC-RHE methods were run on a single thread, while Regenie was run using 12 cores; we report the total CPU time across all cores. (b) Total number of LD blocks containing genome-wide significant gene-trait associations detected using different methods, including their combination using joint tests. We annotate the counts by whether the significant LD block-trait pair contains known gene-trait associations reported on the Open Targets platform (see URLs). (c) Overlap in LD blocks containing gene-trait associations detected using ARG-RHE applied to ARGs inferred from SNP array data, or HRC-RHE, ACAT-V applied to HRC-imputed variants.

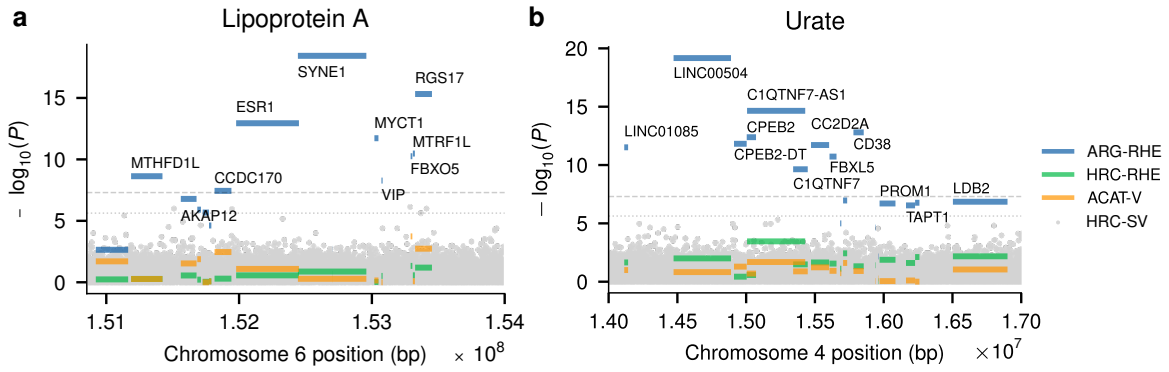

Figure 5: Manhattan plots for two regions containing ARG-RHE associations. (a) Manhattan plot for a genomic region containing genes, including SYNE1 and ESR1, associated using ARG-RHE to lipoprotein A levels. (b) Manhattan plot for a genomic region containing genes, including the LINC00504 non-coding RNA, CPEB2, and CD38 genes, associated with serum urate levels. Dotted lines represent the genome-wide significance threshold for gene-based tests. Dashed lines represent the genome-wide significance threshold for HRC-SV single-variant tests.

8. Supplementary Figures

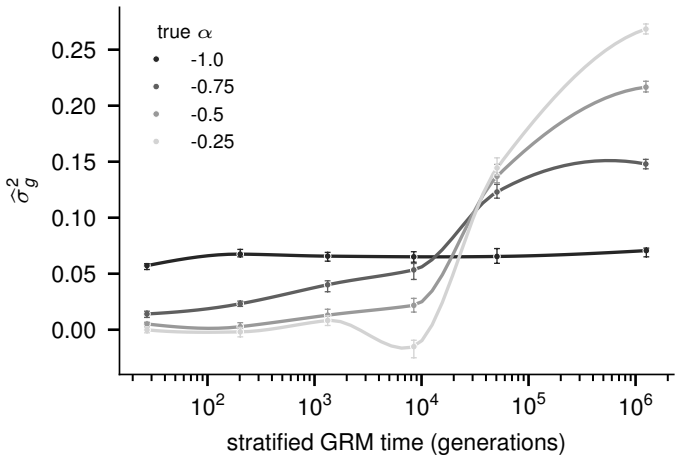

Figure S1: Joint estimates of heritability using time-stratified GRMs. Analysis of a 10 Mb gene window for 20,000 individuals and a heritable phenotype with $h^2 = 0.4$ and polygenicity $g = 0.5$. Resampled mutations are partitioned into 6 time bins containing an equal number of mutations, for which we jointly estimate σ^2_g values. The true underlying MAF-dependency parameter α, which mimics the effects of negative selection and modulates the h^2 contribution of variants of different ages, ranges from -1 to -0.25. The GRMs are built assuming $\alpha = -1$. We report the median values for estimated σ^2_g; error bars show 95% confidence intervals from 500 independent repetitions.

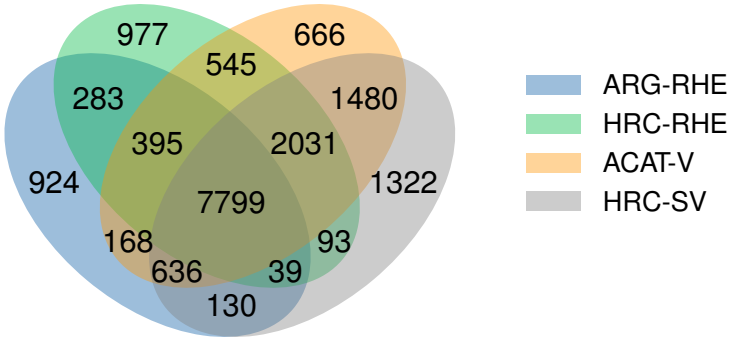

Figure S2: Overlap in LD blocks containing gene-trait associations detected using ARG-RHE and other methods. As in Figure 4b, now including HRC-SV. ARG-RHE was applied to ARGs inferred using Threads on SNP array data, while HRC-RHE, ACAT-V, and HRC-SV were applied to HRC-imputed variants.

References

[1] Abdel Abdellaoui et al. "15 Years of GWAS Discovery: Realizing the Promise". *American Journal of Human Genetics* (2023).

[2] Alison R. Barton et al. "Whole-Exome Imputation within UK Biobank Powers Rare Coding Variant Association and Fine-Mapping Analyses". *Nature Genetics* (2021).

[3] Nadezhda M. Belonogova et al. "sumSTAAR: A Flexible Framework for Gene-Based Association Studies Using GWAS Summary Statistics". *PLOS Computational Biology* (2022).

[4] Tomaz Berisa and Joseph K. Pickrell. "Approximately Independent Linkage Disequilibrium Blocks in Human Populations". *Bioinformatics* (2016).

[5] Takiy-Eddine Berrandou, David Balding, and Doug Speed. "LDAK-GBAT: Fast and Powerful Gene-Based Association Testing Using Summary Statistics". *American Journal of Human Genetics* (2023).

[6] Brian L Browning and Sharon R Browning. "Improving the Accuracy and Efficiency of Identity-by-Descent Detection in Population Data". *Genetics* (2013).

[7] Sharon R Browning and Brian L Browning. "Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies by Use of Localized Haplotype Clustering". *American Journal of Human Genetics* (2007).

[8] Clare Bycroft et al. "The UK Biobank Resource with Deep Phenotyping and Genomic Data". *Nature* (2018).

[9] Robert B. Davies. "The Distribution of a Linear Combination of X2 Random Variables". *Journal of the Royal Statistical Society. Series C (Applied Statistics)* (1980).

[10] Christiaan A. De Leeuw et al. "MAGMA: Generalized Gene-Set Analysis of GWAS Data". *PLOS Computational Biology* (2015).

[11] Caoqi Fan, Nicholas Mancuso, and Charleston W.K. Chiang. "A Genealogical Estimate of Genetic Relationships". *American Journal of Human Genetics* (2022).

[12] Hilary K Finucane et al. "Partitioning Heritability by Functional Annotation Using Genome-Wide Association Summary Statistics". *Nature Genetics* (2015).

[13] Steven Gazal et al. "Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations". *Nature Genetics* (2018).

[14] Steven Gazal et al. "Linkage Disequilibrium–Dependent Architecture of Human Complex Traits Shows Action of Negative Selection". *Nature Genetics* (2017).

[15] Maya Ghoussaini et al. “Open Targets Genetics: Systematic Identification of Trait-Associated Genes Using Large-Scale Genetics and Functional Genomics”. *Nucleic Acids Research* (2021).

[16] Robert C Griffiths and Simon Tavaré. “The age of a mutation in a general coalescent tree”. *Stochastic Models* (1998).

[17] Árni Freyr Gunnarsson et al. “A Scalable Approach for Genome-Wide Inference of Ancestral Recombination Graphs”. *bioRxiv* (2024).

[18] Alexander Gusev et al. “Whole Population, Genome-Wide Mapping of Hidden Relatedness”. *Genome research* (2009).

[19] Nathan Halko et al. “An Algorithm for the Principal Component Analysis of Large Data Sets”. *SIAM Journal on Scientific Computing* (2011).

[20] Michael F Hutchinson. “A Stochastic Estimator of the Trace of the Influence Matrix for Laplacian Smoothing Splines”. *Communications in Statistics-Simulation and Computation* (1989).

[21] Longda Jiang et al. “A Generalized Linear Mixed Model Association Tool for Biobank-Scale Data”. *Nature Genetics* (2021).

[22] Ruth Dolly Johnson et al. “Leveraging Genomic Diversity for Discovery in an HR-linked Biobank: The UCLA ATLAS Community Health Initiative”. *medRxiv* (2021).

[23] Jerome Kelleher, Alison M Etheridge, and Gilean McVean. “Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes”. *PLoS computational biology* (2016).

[24] Jerome Kelleher et al. “Inferring Whole-Genome Histories in Large Population Datasets”. *Nature Genetics* (2019).

[25] Amit V Khera et al. “Genome-Wide Polygenic Scores for Common Diseases Identify Individuals with Risk Equivalent to Monogenic Mutations”. *Nature Genetics* (2018).

[26] Gleb Kichaev et al. “Leveraging Polygenic Functional Enrichment to Improve GWAS Power”. *American Journal of Human Genetics* (2019).

[27] Adam Kiezun et al. “Deleterious alleles in the human genome are on average younger than neutral alleles of the same frequency”. *PLOS Genetics* (2013).

[28] Motoo Kimura and Tomoko Ohta. “The age of a neutral mutant persisting in a finite population”. *Genetics* (1973).

[29] Kathy L.E. Klos et al. “ESR1 Polymorphism Is Associated with Plasma Lipid and Apolipoprotein Levels in Caucasians of the Rochester Family Heart Study”. *Journal of Lipid Research* (2008).

[30] Gregory V Kryukov, Len A Pennacchio, and Shamil R Sunyaev. “Most rare missense alleles are deleterious in humans: implications for complex disease and association studies”. *American Journal of Human Genetics* (2007).

[31] D Kuonen. “Saddlepoint Approximations for Distributions of Quadratic Forms in Normal Variables”. *Biometrika* (1999).

[32] Mitja I Kurki et al. “FinnGen: Unique Genetic Insights from Combining Isolated Population and National Health Register Data”. *medRxiv* (2022).

[33] Seunggeung Lee et al. “Rare-Variant Association Analysis: Study Designs and Statistical Tests”. *American Journal of Human Genetics* (2014).

[34] Vivian Link et al. “Tree-Based QTL Mapping with Expected Local Genetic Relatedness Matrices”. *American Journal of Human Genetics* (2023).

[35] Christoph Lippert et al. “Greater Power and Computational Efficiency for Kernel-Based Association Testing of Sets of Genetic Variants”. *Bioinformatics* (2014).

[36] Huan Liu, Yongqiang Tang, and Hao Helen Zhang. “A New Chi-Square Approximation to the Distribution of Non-Negative Definite Quadratic Forms in Non-Central Normal Variables”. *Computational Statistics & Data Analysis* (2009).

[37] Yaowu Liu et al. “ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in Sequencing Studies”. *American Journal of Human Genetics* (2019).

[38] Thomas Lumley et al. “FastSKAT: Sequence Kernel Association Tests for Very Large Sets of Markers”. *Genetic Epidemiology* (2018).

[39] Tanya J Major et al. “A genome-wide association analysis of 2,622,830 individuals reveals new pathogenic pathways in gout”. *medRxiv* (2022).

[40] T Maruyama. “The age of a rare mutant gene in a large population.” *American journal of human genetics* (1974).

[41] Joelle Mbatchou et al. “Computationally Efficient Whole-Genome Regression for Quantitative and Binary Traits”. *Nature Genetics* (2021).

[42] Gil McVean. “A Genealogical Interpretation of Principal Components Analysis”. *PLOS Genetics* (2009).

[43] Mark J Minichiello and Richard Durbin. “Mapping Trait Loci by Use of Inferred Ancestral Recombination Graphs”. *American Journal of Human Genetics* (2006).

[44] Edward Mountjoy et al. “An Open Approach to Systematically Prioritize Causal Variants and Genes at All Published Human GWAS Trait-Associated Loci”. *Nature Genetics* (2021).

[45] Akiko Nagai et al. “Overview of the BioBank Japan Project: Study Design and Profile”. *Journal of Epidemiology* (2017).

[46] Juba Nait Saada et al. “Identity-by-Descent Detection across 487,409 British Samples Reveals Fine Scale Population Structure and Ultra-Rare Variant Associations”. *Nature Communications* (2020).

[47] Benjamin M Neale et al. “Testing for an Unusual Distribution of Rare Variants”. *PLOS Genetics* (2011).

[48] Carole Ober et al. “Genome-Wide Association Study of Plasma Lipoprotein(a) Levels Identifies Multiple Genes on Chromosome 6q”. *Journal of Lipid Research* (2009).

[49] David Ochoa et al. “The Next-Generation Open Targets Platform: Reimagined, Redesigned, Rebuilt”. *Nucleic acids research* (2023).

[50] Pier Francesco Palamara. “ARGON: Fast, Whole-Genome Simulation of the Discrete Time Wrightfisher Process”. *Bioinformatics* (2016).

[51] Pier Francesco Palamara et al. “High-Throughput Inference of Pairwise Coalescence Times Identifies Signals of Selection and Enriched Disease Heritability”. *Nature Genetics* (2018).

[52] Ali Pazokitoroudi et al. “Efficient Variance Components Analysis across Millions of Genomes”. *Nature Communications* (2020).

[53] Gundula Povysil et al. “Rare-Variant Collapsing Analyses for Complex Traits: Guidelines and Applications”. *Nature Reviews Genetics* (2019).

[54] Alkes L Price et al. “Pooled association tests for rare variants in exon-resequencing studies”. *American Journal of Human Genetics* (2010).

[55] Jonathan K Pritchard. “Are rare variants responsible for susceptibility to complex diseases?” *American Journal of Human Genetics* (2001).

[56] Peter Ralph, Kevin Thornton, and Jerome Kelleher. “Efficiently Summarizing Relationships in Large Samples: A General Duality Between Statistics of Genealogies and Genomes”. *Genetics* (2020).

[57] Matthew D. Rasmussen et al. “Genome-Wide Inference of Ancestral Recombination Graphs”. *PLOS Genetics* (2014).

[58] Pouria Salehi Nowbandegani et al. “Extremely Sparse Models of Linkage Disequilibrium in Ancestrally Diverse Association Studies”. *Nature Genetics* (2023).

[59] Armin P. Schoech et al. “Quantification of Frequency-Dependent Genetic Architectures in 25 UK Biobank Traits Reveals Action of Negative Selection”. *Nature Communications* (2019).

[60] Ruhollah Shemirani et al. “Rapid Detection of Identity-by-Descent Tracts for Mega-Scale Datasets”. *Nature communications* (2021).

[61] Doug Speed et al. “Improved Heritability Estimation from Genome-wide SNPs”. *American Journal of Human Genetics* (2012).

[62] Leo Speidel et al. “A Method for Genome-Wide Genealogy Estimation for Thousands of Samples”. *Nature Genetics* (2019).

[63] Walter W. Stroup. *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. CRC Press, Taylor & Francis Group, 2013.

[64] Peter M Visscher, William G Hill, and Naomi R Wray. “Heritability in the Genomics Era—Concepts and Misconceptions”. *Nature reviews genetics* (2008).

[65] Shijie Wen, Hiroshi Arakawa, and Ikumi Tamai. “CD38 Activation by Monosodium Urate Crystals Contributes to Inflammatory Responses in Human and Murine Macrophages”. *Biochemical and Biophysical Research Communications* (2021).

[66] Anthony Wilder Wohns et al. “A Unified Genealogy of Modern and Ancient Genomes”. *Science* (2022).

[67] Genevieve L Wojcik et al. “Genetic Analyses of Diverse Populations Improves Discovery for Complex Traits”. *Nature* (2019).

[68] Michael C. Wu et al. “Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test”. *American Journal of Human Genetics* (2011).

[69] Yue Wu and Sriram Sankararaman. “A Scalable Estimator of SNP Heritability for Biobank-Scale Data”. *Bioinformatics* (2018).

[70] Jian Yang et al. “Advantages and Pitfalls in the Application of Mixed-Model Association Methods”. *Nature Genetics* (2014).

[71] Jian Yang et al. “Common SNPs Explain a Large Proportion of the Heritability for Human Height”. *Nature Genetics* (2010).

[72] Jian Yang et al. “GCTA: A Tool for Genome-wide Complex Trait Analysis”. *American Journal of Human Genetics* (2011).

[73] Loïc Yengo et al. “A saturated map of common genetic variants associated with human height”. *Nature* (2022).

[74] Jian Zeng et al. “Signatures of Negative Selection in the Genetic Architecture of Human Complex Traits”. *Nature Genetics* (2018).

[75] Brian C Zhang et al. “Biobank-Scale Inference of Ancestral Recombination Graphs Enables Genealogical Analysis of Complex Traits”. *Nature Genetics* (2023).

[76] Wei Zhou et al. “Efficiently Controlling for Case-Control Imbalance and Sample Relatedness in Large-Scale Genetic Association Studies”. *Nature Genetics* (2018).

[77] Wei Zhou et al. “Scalable Generalized Linear Mixed Model for Region-Based Association Tests in Large Biobanks and Cohorts”. *Nature Genetics* (2020).

[78] Ying Zhou, Sharon R. Browning, and Brian L. Browning. “A Fast and Simple Method for Detecting Identity-by-Descent Segments in Large-Scale Data”. *American Journal of Human Genetics* (2020).

[79] Andrey Ziyatdinov et al. “Joint testing of rare variant burden scores using non-negative least squares”. *bioRxiv* (2023).