Tree-based QTL mapping with expected local genetic relatedness matrices

Authors

Vivian Link, Joshua G. Schraiber, Caoqi Fan, Bryan Dinh, Nicholas Mancuso, Charleston W.K. Chiang, Michael D. Edge

Correspondence

edgem@usc.edu

We identify quantitative trait loci by testing the eGRM, a measure of local relatedness based on the ancestral recombination graph, for association with phenotypes using a variance-components framework. This is beneficial for phenotypes with allelic heterogeneity and allows us to capture untyped variation without imputation in understudied populations.

Link et al., 2023, The American Journal of Human Genetics *110*, 2077–2091
December 7, 2023 © 2023 American Society of Human Genetics.
https://doi.org/10.1016/j.ajhg.2023.10.017

CellPress

Tree-based QTL mapping with expected local genetic relatedness matrices

Vivian Link,[1] Joshua G. Schraiber,[1] Caoqi Fan,[1,2] Bryan Dinh,[1,2] Nicholas Mancuso,[1,2] Charleston W.K. Chiang,[1,2] and Michael D. Edge[1,*]

Summary

Understanding the genetic basis of complex phenotypes is a central pursuit of genetics. Genome-wide association studies (GWASs) are a powerful way to find genetic loci associated with phenotypes. GWASs are widely and successfully used, but they face challenges related to the fact that variants are tested for association with a phenotype independently, whereas in reality variants at different sites are correlated because of their shared evolutionary history. One way to model this shared history is through the ancestral recombination graph (ARG), which encodes a series of local coalescent trees. Recent computational and methodological breakthroughs have made it feasible to estimate approximate ARGs from large-scale samples. Here, we explore the potential of an ARG-based approach to quantitative-trait locus (QTL) mapping, echoing existing variance-components approaches. We propose a framework that relies on the conditional expectation of a local genetic relatedness matrix (local eGRM) given the ARG. Simulations show that our method is especially beneficial for finding QTLs in the presence of allelic heterogeneity. By framing QTL mapping in terms of the estimated ARG, we can also facilitate the detection of QTLs in understudied populations. We use local eGRM to analyze two chromosomes containing known body size loci in a sample of Native Hawaiians. Our investigations can provide intuition about the benefits of using estimated ARGs in population- and statistical-genetic methods in general.

Introduction

Identifying trait-associated genetic loci is one of the central aims of genetics. Over the past several decades, a range of approaches—prominently including linkage mapping and genome-wide association studies (GWASs)—have been developed in order to fill this need.[1] In humans, GWASs have become a tremendous research enterprise, with millions of study participants enrolled and hundreds of thousands of trait-associated variants identified.[2]

For decades, geneticists have noted the usefulness of tree-based structures for describing genetic variation and for characterizing the genealogical and evolutionary processes that create genetic variation. At a single non-recombining locus, a tree called a gene genealogy describes the shared ancestry of individual copies of the locus.[3] For entire genomes or genomic regions in which recombination events occurred in the history of the sample, one can represent the sample's shared ancestry via an ancestral recombination graph (ARG) that encodes the sequence of "local" or "marginal" trees along the genome,[4] with recombination events as the source of differences in topology between neighboring trees. The ARG encodes all mutation, recombination, and shared ancestry events in the history of a sample of genomes.

Tree-based approaches to quantitative trait locus (QTL) mapping—in which a trait is tested for association with a tree or set of trees describing genetic variation in a region—have been proposed several times and shown to provide some advantages,[5–19] as have approaches to haplotype-based mapping that leverage awareness of tree-like relatedness patterns among sets of haplotypes.[20–23] At the same time, explicitly tree-based approaches have until recently been limited by difficulties in estimating locus-level trees at scale. Further, the dominance of meta-analysis in GWASs[24] and other methods based on summary statistics has meant that individual-level genetic data are often not available to data analysts, precluding most tree-based approaches.

In principle, tree-based approaches have the potential to address three long-standing difficulties of GWASs. First, a GWAS entails a huge number of statistical tests and requires a substantial correction for multiple testing as a result.[25] Many of these tests are correlated or redundant because the variants tested occur on the same or very similar underlying gene-genealogical trees. Testing the trees themselves may allow for fewer tests.

Second, GWASs are known to be prone to miss trait-associated genetic loci characterized by allelic heterogeneity, in which multiple nearby causal variants affect a trait of interest.[26–29] Under allelic heterogeneity, causal alleles with opposing effects on a trait might be associated with the same marker allele, diminishing the association signal at the marker. Allelic heterogeneity is not rare, appearing in many Mendelian loci identified during the linkage era[30]—linkage mapping is robust to allelic heterogeneity—and estimated recently to occur at a substantial fraction of complex trait loci[29] and expression QTLs.[31,32] Tree-based approaches, by focusing on local relatedness of haplotypes in the sample, can offer robustness to allelic heterogeneity of a sort similar to that obtained in linkage analysis.

[1]Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA, USA; [2]Center for Genetic Epidemiology, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
*Correspondence: edgem@usc.edu
https://doi.org/10.1016/j.ajhg.2023.10.017.
© 2023 American Society of Human Genetics.

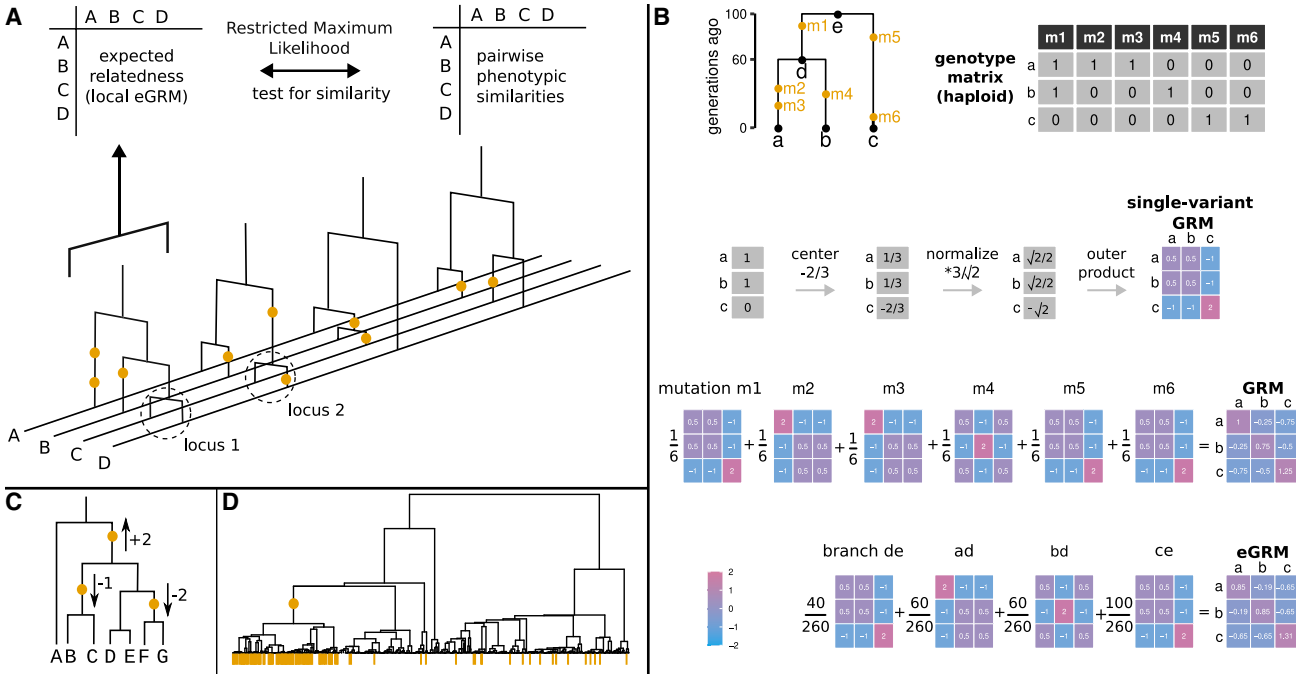

Figure 1. Main ideas of the testing framework

(A) Local eGRM framework. Schematic of the marginal trees of an ARG of 4 haploid individuals (A–D). The gold circles on the trees correspond to mutations. The clade marked with the dotted circles is identical at loci 1 and 2, and the mutation at locus 2 is informative about the branch length at locus 1. One or more marginal trees are used to calculate a local eGRM using the method described in the work of Fan and colleagues.[44] This matrix is then tested for association with the phenotypes using restricted maximum likelihood (REML).

(B) Computing the eGRM. This panel is redrawn from the paper of Fan and colleagues.[44] A genome-wide genetic relatedness matrix (GRM) can be viewed as an average of single-locus GRMs for every genotyped locus. An expected GRM (eGRM) can be obtained as a weighted average of single-locus GRMs defined by each branch in the ARG, where the weights are proportional to the expected number of mutations falling on each branch.

(C) Allelic heterogeneity. One marginal tree of an ARG with three causal mutations with opposing phenotypic effects.

(D) One marginal tree with one causal mutation. A hypothetical haploid tree in which a causal mutation (gold circle) partitions the individuals of the tree into descendants and non-descendants. The pronounced phenotype is detected in almost all descendants (gold dashes at tips) but also in other individuals.

Third, modern GWASs are fueled by imputation, in which a reference sample is fully sequenced, and then study samples that are more sparsely genotyped have their missing genotypes imputed statistically.[33,34] The imputed genotypes can then be tested for association with the trait of interest. The success of modern imputation approaches is made possible by the fact that genetic variation is structured locally in a tree-like way.[35,36] At the same time, imputation is most successful if the reference and study samples are closely related,[37–39] and closely related reference samples are not always available. Testing the tree structures that underlie imputation may offer a more direct approach to identifying QTLs that could circumvent the need for closely related reference samples.

Due to advances in ARG estimation, it may now be possible to apply tree-based methods at sufficient scale to detect QTLs. Although estimation of the ARG is extremely difficult, approximate estimation procedures that operate on single-nucleotide polymorphism (SNP) array data and scale to thousands of samples have emerged in the last few years.[40–43] Further, for researchers studying QTLs in humans, the emergence of large biobanks has meant that individual researchers or research teams have access to individual-level genetic data in sample sizes that might allow the identification of trait-associated loci.

Here, we present a tree-based approach to QTL mapping (Figure 1). We build on a recently proposed representation of tree-based relatedness, the expected genetic relatedness matrix, or eGRM[44] (see Figure 1B here), independently identified by Zhang and colleagues[42] as the ARG-GRM and in a phylogenetic context by Wang and colleagues[45] as the expected genetic similarity matrix (also see equation 10 in McVean et al.[46] for a similar computation). Genetic relatedness matrices (GRMs) are used in a wide array of statistical genetic tasks, including adjusting for population stratification and estimation of heritability.[47] Given an ARG encoding the history of a sample, the eGRM is the expectation of the GRM assuming that mutations are placed on the ARG as a Poisson process. In general, we can compute a tree-based analog of any statistic computed from genetic variation by taking its expectation given the ARG.[48]

Our procedure is to test eGRMs built from local segments of the ARG for concordance with a phenotype using a random-effects model fit by restricted maximum likelihood (REML). Loosely, the test is sensitive to cases in which

individuals who are more closely related in some local segment of the genome are likely to be more similar on the phenotype. This approach is essentially a tree-based version of previous methods to test local GRMs for concordance with a phenotype, which have been framed variously as QTL mapping approaches[49,50] or local heritability estimation.[51–54] Using tree sequences estimated by Relate,[41] we test our approach in simulations of varying degrees of allelic heterogeneity, including the case of a single causal variant. We also use our approach to analyze real data from a sample of Native Hawaiians. We analyzed chromosomes 5 and 16 with particular attention to (1) *CREBRF*, in which the lack of a population-specific reference panel has previously precluded the detection by a GWAS of a known polymorphism with a large effect on body mass index (BMI),[39,55] and (2) *FTO*, which is also strongly associated with BMI[56–58] and demonstrates allelic heterogeneity.[59]

Subjects and methods

Characterizing local relatedness

The key to our approach is a matrix **A**, called a local genetic relatedness matrix (local GRM), that characterizes the relatedness of individuals in a local region to be tested as a candidate QTL. Classically, a local GRM is calculated on the basis of the observed variants in a window (e.g., Yang and colleagues[60]). Our method is instead based on using the expectation of a local GRM (local eGRM) given an estimated ancestral recombination graph (ARG).

Figure 1 describes the main ideas of our framework: in each window, we calculate one genetic relatedness matrix using the ARG's marginal trees in that window and then use REML (restricted maximum likelihood) to test whether the local genetic relatedness explains phenotypic similarities in the sample. One advantage of using the ARG to describe genetic relatedness is that information from neighboring trees is naturally shared. To illustrate this idea, the dotted circles in Figure 1A show a clade that exists in trees 1 and 2, with a mutation at locus 2 that differentiates tips C and D. Though the mutation is at marginal tree 2, its presence is informative about the branch lengths at marginal tree 1, since the relevant subtree is identical in marginal trees 1 and 2.

Figure 1B describes the method we use to calculate the pairwise expected relatedness matrix (eGRM), developed by Fan and colleagues,[44] and also how the genetic relatedness matrix is conventionally calculated. Figure 1C shows an example of allelic heterogeneity: multiple causal alleles are in close linkage (e.g., on the same marginal tree of the ARG), so tag SNPs will be linked to several causal alleles with opposing effects. If the causal variants are themselves untyped, this can lead to association signals interfering or even canceling each other out at the typed variant. Even if the causal variants are typed, association power will be improved if they are tested for association jointly rather than separately. Figure 1D depicts a hypothetical marginal tree in which a binary phenotype's expression across all (haploid) individuals is shown at the tree's tips. A single causal mutation on a specific branch results in a majority of the descendants from that branch displaying the phenotype. If this branch is known, it can be included in tests, even if no genotyped mutations fall on it.

Local GRM

We compute local GRMs from bi-allelic variants in the local window to be tested as a QTL (as done, for example, by Yang and colleagues[60]). The entry relating individuals i and j in the GRM can be written as

$$GRM_{i,j} = \frac{1}{\ell}\sum_{k=1}^{\ell}\frac{\left(y_{i,k}-2p_k\right)\left(y_{j,k}-2p_k\right)}{\left[2p_k(1-p_k)\right]^{\alpha}}, \qquad \text{(Equation 1)}$$

where k in an index over the sites considered, ℓ is the total number of sites considered, $y_{i,k}$ is the number of focal alleles carried by individual i, and p_k is the sample frequency of the focal allele.[47] The constant α determines the relative emphasis placed on rarer variants in computing relatedness estimates, with larger values giving greater weight to rarer variants. In this paper, we use $\alpha = 1$. With $\alpha = 1$, the local GRM is a covariance matrix of mean-centered, standardized genotype counts among individuals, where the standardization is by $\sqrt{2p(1-p)}$, the standard deviation of the genotype under Hardy-Weinberg equilibrium.

Local eGRM

Fan and colleagues[44] compute a genome-wide global eGRM, which is the expectation of the genetic relatedness matrix (GRM) described by Equation 1 (with $\alpha = 1$) conditional on the ARG, assuming that infinite-sites mutations are placed on the ARG as a Poisson process. The global eGRM can be computed as a weighted sum of single-locus GRMs implied by each branch in the ARG. Specifically, each branch in the ARG defines a clade of tips that descend from the branch. A mutation on that branch would be inherited by all these tips and so would define a single-locus GRM following Equation 1. The eGRM is equal to a weighted average of all such branch-wise GRMs, with weights per branch proportional to a product $\mu(b)l(b)t(b)$, where $\mu(b)$ is the mutation rate on the branch, $l(b)$ is the length of the genomic region spanned by the branch, and $t(b)$ is the length of the branch (i.e., the time in the tree that the branch exists). Here, as in Fan et al.,[44] we assume that the mutation rates are the same on all branches.

We compute the local eGRM for genomic regions of a predefined size. The local eGRM for one tree is a weighted sum over the tree's branches. In order to calculate the local eGRM for a genomic window, we first calculate the local eGRM for all trees whose genomic intervals overlap the window and then take a weighted average of these matrices, where the weights correspond to the fraction of the window covered by each tree's genomic interval. This approach to computation is redundant (because many branches exist across multiple marginal trees) and can be ameliorated in principle via an approach that records unique branches only once.[48] We did not pursue this solution because of our decision to work with Relate trees, which do not preserve branch lengths exactly between neighboring marginal trees.

We computed local eGRMs using egrm software.[44]

The variance-components model

Let $\mathbf{y}$ be quantitative phenotypes for n individuals, $\mathbf{X}$ be an $n \times k$ design matrix of covariates, and β the covariates' regression coefficients. The k covariates may include nuisance variables and potentially confounding factors, such as age, sex, and descriptions of population structure or global relatedness. Additionally, let $\mathbf{I}_n$ be the $n \times n$ identity matrix, and σ_ϵ^2 the variance of environmental noise. Given a GRM $\mathbf{A}$ of dimensions $n \times n$ representing

relatedness among individuals in a local segment of the genome, we model the phenotypic variation in the sample by

$$\mathbf{y}|\beta,\sigma_a^2,\sigma_\epsilon^2 \sim N(\mathbf{X}\beta,\sigma_a^2\mathbf{A}+\sigma_\epsilon^2\mathbf{I}_n). \quad \text{(Equation 2)}$$

We estimate β, σ_a^2, and σ_e^2 with restricted maximum likelihood (REML). We identify a QTL if the parameter σ_a^2 is significantly different from 0.

To understand the difference between using a GRM based on observed variants compared with an expectation conditional on an estimated ARG, note that when $\mathbf{A}$ is based on observed variants, the model in Equation 2 is equivalent to one in which the typed sites in the window receive random, uncorrelated effect sizes with expectation 0 and variance proportional to $1/(2p(1-p))^{\alpha}$[61,62] that contribute additively to the trait. That is, $\gamma = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{e}$, with $\mathbf{X}$ an $n \times k$ design matrix of covariates with fixed effects β ($k \times 1$), $\mathbf{Z}$ an $n \times l$ matrix of genotypes at the sites considered with random effects $\mathbf{u}$ ($\ell \times 1$), and $\mathbf{e}$ an $n \times 1$ vector of random, uncorrelated environmental effects.

On the other hand, if $\mathbf{A}$ is computed by taking an expectation over an ARG, the model in Equation 2 is equivalent to one in which each branch of the ARG incorporated in $\mathbf{A}$ receives a random, uncorrelated effect size with expectation 0 and variance proportional to $\mu(b)l(b)t(b)/(2p(1-p))^{\alpha}$, where p is the proportion of tips that descend from a branch in the relevant span of the genome, and again $\mu(b)$ is the mutation rate on the branch, $l(b)$ is the length of the genomic region spanned by the branch, and $t(b)$ is the length of the branch.

Simulating genealogy and genotypes

Simulating ARGs for one population

We simulated ARGs for one population using stdpopsim[63,64] v.0.1.2 using the Python API. We simulated chromosome 1 for 2,000 haploid individuals of African ancestry using the "OutOfAfrica_3G09" model and msprime[65] v.1.1.1 and otherwise default parameters. We then extracted the genomic region starting at position 49,000,000 and ending at position 50,000,000. Then, we randomly assigned pairs of haplotypes to 1,000 individuals to create diploids.

Estimating ARGs with relate

To simulate genotyping array data, we filtered the simulated ARG's variants by retaining 20% of those with a minor allele frequency of at least 1%. We then used the retained variants to estimate ARGs with Relate[41] using parameters '–mode All,' '–mutation_rate 1.25e−8,' '–effectiveN 2000,' and the human recombination map (-HapMap phase II, build GRCh37, provided with the Relate software). We then converted the output to treeSequence format[66] with Relate's tool RelateFileFormats and '–mode ConvertToTreeSequence.'

Simulating phenotypes

Choosing the causal variants

We selected causal variants among those that were not retained in the downsampling scheme described above ("untyped"). In the experiments with one causal variant, the selection was uniformly at random among variants of a predefined frequency. If no branch in the local trees subtended the desired frequency, we chose the nearest possible frequency. In the experiments with allelic heterogeneity, we defined a causal window with a physical length of 5 kb in the center of the ARG and randomly selected a given proportion of untyped variants within the window to be causal.

Choosing the effect sizes

We chose the effect size for each variant on the basis of its allele frequency, sampling from a normal distribution with expectation 0 and standard deviation inversely proportional to $\sqrt{p(1-p)}$, where p is the variant's minor allele frequency. This corresponds to the LDAK model with $\alpha = -1$ (depending on how the model is parametrized, sometimes $\alpha = 1$).[67] It leads variants with lower allele frequencies to have effects with larger absolute sizes and also matches the normalization of the local eGRM/GRM that we perform. We assume an additive model, where each copy of an allele contributes to the phenotype equally, regardless of the other copy at the position, or genotypes at other positions.

In order to obtain the desired local heritability, we added random noise to the phenotypes such that $V_E = V_G(1-h^2)/h^2$, where V_E is the phenotypic variance due to environmental effects uncorrelated with genotype, V_G is the phenotypic variance due to genetic effects, and h^2 is the desired local heritability.

QTL testing

Local REML

We tested each local relatedness matrix (GRM and eGRM) for association with the phenotypes using GCTA (v.1.94.1)[68] and its implementation of restricted maximum likelihood (REML) with tag '–reml' and providing the local relatedness matrix (tag '–grm'), the phenotypes (tag '–pheno'), and running the algorithm for a maximum of 500 iterations. Note that GCTA p values for random effects in such a model are never larger than 1/2.

GWAS

We tested each typed variant for association with the phenotypes using python's statsmodels[69] (v.0.13.2) OLS function.

ACAT-V

We use function ACAT from R package ACAT[70] v.0.91 to run ACAT-V on the p values from the GWAS results in a window.

Correcting for population stratification

Simulating phenotypes affected by genetic confounding

In order to test different methods to correct for population stratification in the Hawaiian data, we simulated 50 sets of genetically confounded phenotypes based on the Native Hawaiian population in the Multiethnic Cohort MEC.[71] We randomly sampled 1,000 causal variants from the GWAS variant set (filtered for minor allele frequency and LD, see below) from chromosomes 1–3 to be causal. We assigned the causal variants with phenotypic effect sizes that depend on their allele frequencies (see simulating phenotypes). We added random noise to the phenotypes such that the simulated trait heritability was 0.45, which corresponds to the estimated heritability for BMI in Samoans.[55]

Correcting for population stratification

We assessed methods for correcting for population stratification by testing a region near *CREBRF* (chr5:172,750,000–173,000,000) for association with the simulated phenotypes affected by genetic confounding. The correction methods used were (1) 100 principal components taken from the LOCO eGRM without chromosome 5 (in the case of association testing with local eGRM) or LOCO GRM without chromosome 5 (in the case of association testing with a GWAS and local GRM) as fixed effects, (2) the LOCO GRM as a random effect in the case of a GWAS, and (3) a two-step approach where we first obtained the residuals from fitting a model with the LOCO eGRM (or GRM) as a random effect, and then fit the local eGRM (or local GRM) to these residuals. The GWAS with the fixed-effect model were run using PLINK2[72] (v.v2.00a3.7LM

AVX2 Intel [24 Oct 2022], www.cog-genomics.org/plink/2.0/), and the GWAS with the random effect model were run using GCTA (v.1.94.1).[68] The PCA for GWAS was run with EIGENSTRAT[73] and the PCA for local eGRM and GRM was run using GCTA.

Estimating statistical power

Null simulations

In order to determine a significance cutoff for each simulation configuration, we used 300 simulated ARG replicates, and we assigned each individual from each ARG a random $\mathcal{N}(0,1)$ phenotype value irrespective of genotype. We performed association tests for each association method, for each variant set/tree type, and for each testing window size, i.e., for every power simulation configuration that affects the number of association tests. We set the significance cutoff such that the family-wise error rate (i.e., the fraction of replicates containing at least one significant association) was 5%.

Power as a function of genetic architecture

For each parameter combination of variant set/tree type, causal variant proportion, causal window size, testing window size, and local heritability, we counted the number of replicates for which the p value of at least one window (for ACAT-V, local eGRM, and local GRM) or variant (for GWAS) exceeded the significance threshold defined with the null simulations.

Application to real data

Transforming the phenotypes

Phenotype data for body mass index (BMI) was available for 5,371 people from the Native Hawaiian population of the Multiethnic Cohort MEC,[71] along with sex and age. To ensure that the phenotype residuals would follow a standard normal distribution, we performed a transformation typical for BMI data. Namely, we stratified by sex, regressed out age and age squared, and removed individuals for which the residual was more than six standard deviations removed from the sex's mean. Then, we inverse rank normalized the phenotypes.[74]

Estimating the ARG with relate

In total, 5,384 self-identified Native Hawaiians from the Multiethnic Cohort (MEC) were genotyped on two separate GWAS arrays: Illumina MEGA and Illumina Global Diversity Array (GDA). After taking the intersection of SNPs found on both arrays, we removed variants that were genotyped in fewer than 95% of people in the sample, as well as variants out of Hardy-Weinberg equilibrium ($p < 10^{-6}$). We also applied a filter for people with more than 2% missing genotypes but removed no one with this filter.

With approximately 990,000 SNPs after quality control, we phased the genotypes with EAGLE[75] by using its default hg38 genetic map. We inferred ancestral alleles by using the Relate add-on module with ancestral genome homo_sapiens_ancestor_GRCh38_e86.tar.gz, downloaded from ftp://ftp.ensembl.org/pub/release-86/fasta/ancestral_alleles/. We divided the genome into segments containing 10,000 SNPs and ran Relate on these segments in parallel with all default parameters per the user manual.

Inferring the LOCO eGRM to correct for population structure

For LOCO eGRM, we first inferred the segment-wise eGRMs for all chromosomes except chromosome 5 or 16, depending on which chromosome we were analyzing. We combined the segment-wise eGRMs into a global eGRM by taking their weighted sum, where the weights were given by the expected number of mutations in each eGRM, which is a parameter that is provided in the output of egrm.[44] For LOCO GRM, we used PLINK2's "–make-rel" functionality.

Determining the significance cutoffs

The cutoffs in Table S1 were calculated for a genomic region of length 1 Mb. We computed the effective number of independent tests for each method as the number of tests for which the significance cutoff we obtain corresponds to a Bonferroni correction. For our GWAS, the effective number of tests was 281.2, and for local eGRM with 5 kb testing windows, 31. The standard genome-wide significance for a GWAS is 5×10^{-8}, which corresponds to the cutoff for one million independent test with Bonferroni correction. To approximate the genome-wide cutoff value for local eGRM, we assume that the ratio of $\frac{281}{31}$ GWAS tests to local eGRM tests for a given region holds across the genome. We thus set the genome-wide local eGRM cutoff to $\frac{0.05}{10^6 * \frac{31}{281}} \approx 4.5 \times 10^{-7}$. Following the same logic, we set the genome-wide local GRM cutoff to $\frac{0.05}{10^6 * \frac{60}{281}} \approx 2.3 \times 10^{-7}$.

Testing for QTLs

We ran our local eGRM method to test for correspondence between the transformed phenotypes and the estimated ARG around the *CREBRF* region in windows of 5 kb. We corrected for population stratification by using egrm to estimate the LOCO eGRM (leaving out chromosome 5 or 16). We further used PLINK[76] (v.1.07) to test for association between the genotypes on chromosomes 5 and 16 and the transformed phenotypes. To generate principal components as covariates for the GWAS, we held out the focal chromosome being tested for association, and generated the 100 PCs using EIGENSTRAT[73] after additionally filtering out variants with minor allele frequency <1% and filtering for LD using command using –indep-pairwise 50 5 0.8 in PLINK. Finally, we removed the windows with positions overlapping the Encode blacklist of problematic regions in the genome.[77]

Results

We compared our framework based on computing the expected genetic relatedness matrix from an inferred ARG, referred to here as local eGRM, with three other association methods: a GWAS, in which each variant is tested separately; local GRM, which for each testing window calculates a genetic relatedness matrix based on the typed variants within the window (see subjects and methods); and ACAT-V,[78] which for each window combines the variant-level p values from the GWAS and is especially powerful when a small proportion of variants within a window are causal. Although ACAT-V is designed for and most typically applied to sequence data, our focus here is on array data, and so we show results of ACAT-V applied to simulated array data in the main text, deferring comparisons with complete data to supplementary figures.

Calibrating the type I error rate by simulation

To determine the p value cutoffs for each method, we performed null simulations for each parameter combination of variant set or tree type and testing window size, i.e., for every simulation configuration that could lead to a

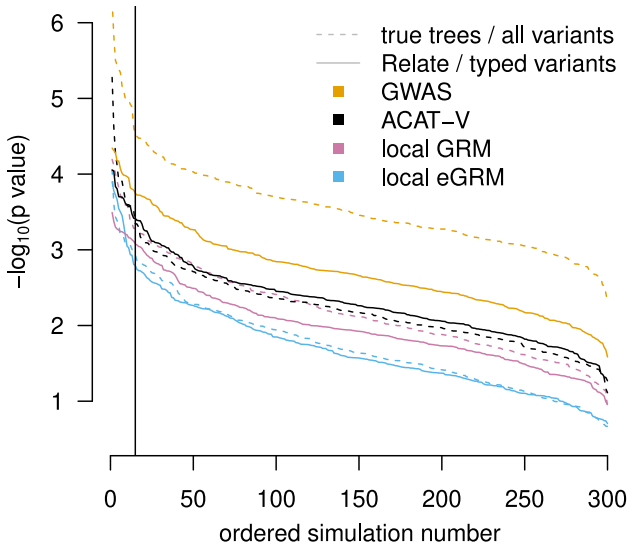

Figure 2. Setting p value cutoffs for family-wise error rate of 5%
The most significant p value for any window (local GRM, local eGRM, ACAT-V) or any SNP (GWAS) in each of the 300 replicates is shown on the y axis for each association method and ARG type/variant set, starting with the lowest minimum p values on the left. The testing window size was 5 kb. The vertical black line corresponds to 5% of the replicates.

different number of tests required per ARG. For each parameter combination, we simulated random phenotypes for all individuals in the sample, and we recorded the smallest p value resulting from the tests of the simulated chromosome against the null phenotypes. We determined the significance cutoff such that the family-wise error rate was 5% in null simulations (Table S1). Figure 2 shows the ordered p values for one of these null simulations. It shows the general pattern that can be seen for all parameter combinations, namely that the multiple testing burden is highest for a GWAS and lower for the window-based association tests ACAT-V, local GRM, and local eGRM. We can compare the results in terms of the number of "effective tests" implied by the p value cutoffs necessary to achieve a family-wise error rate (FWER) of 0.05—that is, the number of tests that would lead to the same cutoff under a Bonferroni correction. For 5 kb windows and Relate trees, the local eGRM method applied to a 1 megabase window entails $\approx$31 effective tests. In contrast, a GWAS on typed variants implies $\approx$280 effective tests, or $\approx$ 9 times as many as the local eGRM method. Further, the cutoffs are more stringent for a GWAS when all variants are used rather than only the subset of variants selected for genotyping, whereas the difference between using only typed variants and all variants is much smaller for the window-based methods.

The null simulations were also useful to determine how well the local eGRM method is calibrated with regard to the distribution of p values under the null. The quantile-quantile plots in Figure S1 confirm for multiple simulation configurations that both local eGRM and local GRM produce close-to-uniformly-distributed but slightly conservative p values. Details of the p value distribution do not influence the simulation results below, since we choose the cutoff for significance empirically based on the null simulations.

Local eGRM exhibits power advantages in cases of allelic heterogeneity

We used simulations to understand the power of our framework to find true trait-relevant genetic regions. As with the null simulations, we simulated 200 replicates of realistic human ARGs for chromosome 1 of 1,000 Africans under the out-of-Africa model using stdpopsim.[63] We simulated phenotypes for each individual in each ARG with a variety of architectures inside a trait-relevant genomic window by varying the size of the testing window, the number of causal variants in the window (a single causal variant or allelic heterogeneity with varying proportions of causal variants), and the heritability explained by variants in the window.

We compared the power of the following approaches: GWAS on both typed and all variants, local GRM with both typed and all variants, local eGRM with Relate trees estimated from typed variants, local eGRM with true trees, and ACAT-V with both typed and all variants.

First, we investigated power in the presence of allelic heterogeneity, i.e., multiple causal variants within close physical proximity and thus genetically linked with each other. Within a predefined causal window of the genome, each untyped variant has a given probability of being causal. Some summary statistics on the numbers of causal variants per causal window are given in Table S2.

Each causal variant is given a random phenotypic effect size such that loci with lower-frequency minor alleles tend to be assigned larger absolute effect sizes, as is observed in human data (see subjects and methods for details). Figure 3 shows power results for causal window of size 5 kb, with 20% of variants causal (Figures 3A and 3B, median 4 causal variants per window) or 50% of variants causal (Figures 3C and 3D, median 11 causal variants per window). We also varied the local heritability and the testing window sizes (5 kb for Figures 3A and 3C, 10 kb for Figures 3B and 3D) for the window-based tests. For the results obtained with a more extensive set of simulation parameters, including results that incorporate both typed and untyped variants, see Figure S2.

Across simulated genetic architectures with allelic heterogeneity, our local eGRM method consistently has higher power than other approaches when analyzing array data. Across the local heritability values simulated in Figure 3, the local eGRM approach with 5 kb analysis windows has on average 17% higher power than a GWAS with 20% of variants causal, and 31% higher power when 50% of the variants are causal. The other three methods (GWAS, local GRM, ACAT-V) performed similarly to each other. In contrast, when using the true ARG and all variant information, as would be captured by accurate sequencing data, a GWAS, ACAT-V, and local GRM all outperform local eGRM (Figure S2), even when local eGRM is performed on the true trees.

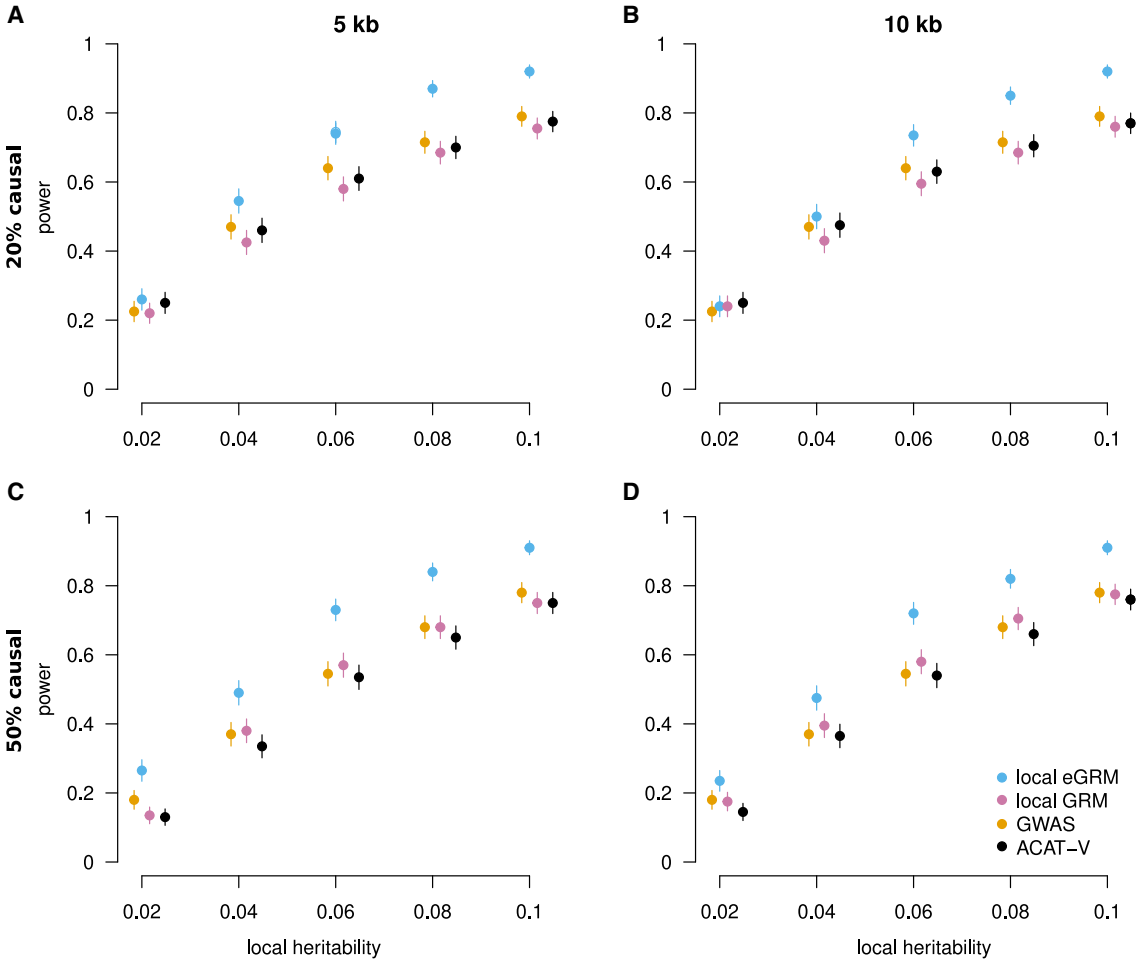

Figure 3. Power comparisons under allelic heterogeneity

Each panel shows the power to detect an association for 4 methods using array-like data when there is a 5 kb causal window. In (A) and (B), phenotypes are determined by 20% of the untyped variants in the window, while in (C) and (D), phenotypes are determined by 50% of the untyped variants in the window. In (A) and (C), a 5 kb test window is used (matching the simulated causal window size). In (B) and (D), a 10 kb test window is used. Error bars correspond to one standard error.

We also assessed the power of each method for phenotypes that have a single untyped causal variant with allele frequency 0.02 (Figures 4A and S3A) or 0.2 (Figures 4B and S3B). With a single causal variant, local eGRM is roughly comparable to the other methods, operating at a slight disadvantage when the frequency of the causal variant is low (0.02) and perhaps a slight advantage when the frequency of the causal variant is higher (0.2).

Tables S4–S7 show how many replicates were found to contain a significant peak by two methods. Generally, the concordance was higher for true ARGs than for Relate-estimated ARGs. The highest concordance is generally between a GWAS and ACAT, followed by the other pairs which have similar concordances. Tables S8 and S9 show analogous results for null simulations.

Correcting for population stratification

The simulations of the power analysis were performed on samples from a panmictic population. In real GWAS settings, however, samples are often affected by population stratification,[79–82] in which genotypes appear correlated with phenotypes because of confounding rather than because of close linkage to causal variants. In GWASs, the most popular strategies for correcting for population stratification are inclusion of a random effect for the global GRM[83] and inclusion of fixed effects for the first several principal components of a standardized genotype matrix,[73] obtained by eigendecomposition of a GRM.

We simulated phenotypes affected by genetic confounding using the Native Hawaiian genotypes in order to test several approaches for correcting population stratification. In particular, we randomly chose variants on chromosomes 1–3 in the Hawaiian individuals to be causal and simulated phenotypes on the basis of those genotypes with a heritability of 0.45. We then tested a region on chromosome 5, which contained no causal variants, for association using different methods to correct for population stratification. These methods included: adding fixed effects for 100 principal components of the LOCO (leave-one-chromosome-out) eGRM or GRM, using the LOCO GRM as a random effect (for the GWAS only) and a two-step procedure (for tests of the local eGRM and local GRM). In the two-step procedure,

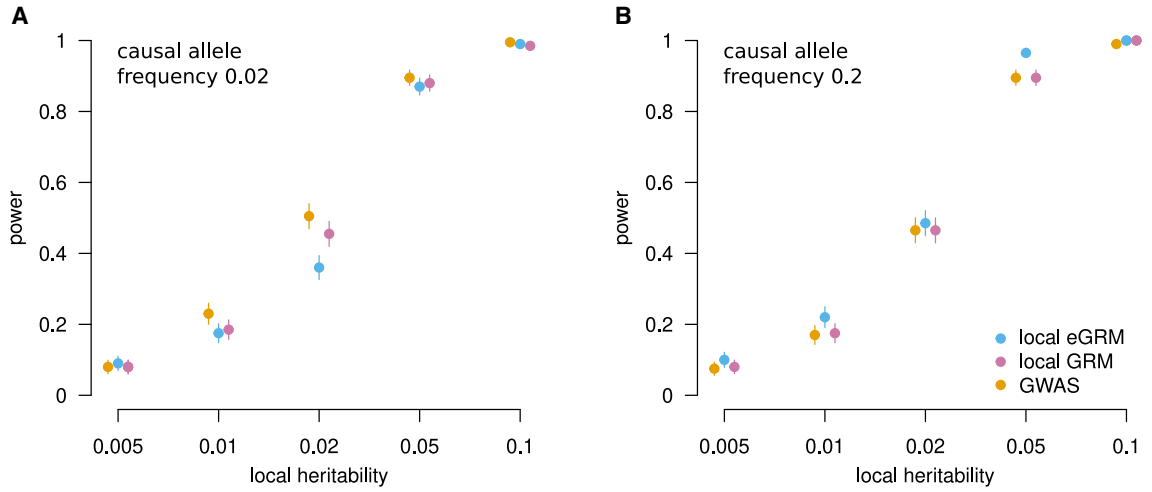

Figure 4. Power comparisons with one causal variant
Each panel shows the power to detect an association for 3 methods using array-like data when there is a causal variant at a frequency of either 0.02 (A) or 0.2 (B). Association tests with methods local eGRM and GRM were performed in genomic windows of 5 kb. The error bars correspond to one standard error.

we first fit a random effect for the LOCO eGRM or GRM and saved the residuals. In the second step, we tested the residuals for association with the local eGRMs or GRMs. We assessed how well the methods corrected for stratification through the distribution of the p values of the association tests in the test region. As shown in Figure S5, including 100 PCs did not lead to uniform p values for the GWAS or local eGRM. This is in line with the complex structure of the Hawaiian population.[84] For the GWAS, including the LOCO GRM did produce uniform p values. Correcting for stratification with a LOCO random effect, i.e., using a model with two random effects, led to convergence problems in some cases (results not shown), but the two-step procedure produced a nearly calibrated p value distribution. It is also in line with other commonly used LMM association methods.[85–89] Thus, for the local eGRM analysis of the Native Hawaiian data, we decided to use the two-step method in the case of local eGRM and the LOCO GRM method in the case of the GWAS. In the case of local GRM, all stratification correction methods that we tested led to somewhat conservative p values. We chose to use the two-step method for the analysis of the Hawaiian data in order to be consistent with local eGRM.

Applying local eGRM to known associations with BMI in the understudied Native Hawaiian cohort of the MEC

We compared the results of local eGRM, local GRM, and the GWAS using the Native Hawaiian subset within the Multiethnic Cohort (MEC),[71] for which we had genotype array data and BMI values for 5,371 individuals. We analyzed chromosomes 5 and 16, with a special focus on the loci containing *CREBRF* on chromosome 5 and *FTO* on chromosome 16.

On the basis of the phased genotypes, we used Relate to estimate the ARG for the whole genome. For the local eGRM association testing, we used egrm[44] to infer the LOCO (leave-one-chromosome-out) eGRMs from this ARG. We then fit a linear mixed model with the LOCO eGRM as a random effect to the transformed BMI phenotypes and saved the residuals. Finally, we tested the residuals for association with the ARG using local eGRM. For local GRM association testing, the procedure was analogous, except that we used the residuals that resulted from using the LOCO GRM as a random effect. In the case of the GWAS, we corrected for population stratification by including the LOCO GRM as a random effect in the association model.

The Manhattan plots of chromosome 5 (Figure 5) and chromosome 16 (Figure S6) show the results of the association tests. The gray vertical lines in all real data figures are hits for body mass index or weight identified in the GWAS catalog (https://www.ebi.ac.uk/gwas/). These GWAS catalog hits were mostly identified in samples much larger than the one we analyze here. None of the three association methods found a genome-wide significant signal on chromosome 5. However, there is a robust local eGRM peak at the end of chromosome 5, where *CREBRF* is located. On chromosome 16, all three methods have a notable peak that does not quite reach genome-wide significance centered on *FTO* (Figures S6 and S7).

CREBRF harbors a missense mutation, rs373863828 (GenBank: NM_153607.3) (c.1370G>A [p.Arg457Gln]), which has a large effect on adiposity. This variant is observed in Pacific Islanders with a frequency as high as 26% in Samoans. However, is very rare or unknown in people without recent ancestry from Polynesia.[55] The association signal was originally detected with body mass index (BMI) in a sample of ∼3k Samoans[55] at a linked tag SNP, rs12513649, which was on the Affymetrix 6.0 array. However, rs373863828 was not included in many databases typically used for imputation, not even those that include diverse populations (e.g., 1000 Genomes Project or

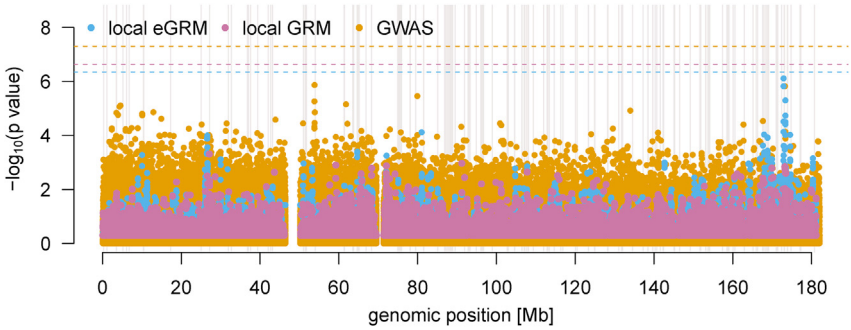

Figure 5. Association results for Hawaiian cohort of the MEC for chromosome 5

Blue dots are per-window negative $\log_{10}$ p values for the local eGRM, orange dots are per-SNP results for GWAS, and pink dots are per-window results for local GRM. The horizontal dashed lines are the genome-wide significance cutoffs (5×10^{-8} for GWAS, 2.3×10^{-7} for local GRM, and 4.5×10^{-7} for local eGRM). The gray vertical lines are SNPs found in the GWAS catalog that were found to be associated to traits "body mass index" and "weight." We corrected for population stratification with the residuals method in the case of local eGRM and GRM, and LOCO GRM in the case of GWAS. *CREBRF* is located at the end of the chromosome around 170 Mb.

Haplotype Reference Consortium), and so it was not well imputed at the time of the study. Since our method does not rely on an imputation step, testing this locus for association was of special interest. Figure 6 shows a zoomed-in view of the locus with association results for local eGRM, local GRM, and GWAS. The pink area shows the coordinates of *CREBRF*; the red vertical line is the causal variant rs373863828, and the gray vertical line immediately to the left of *CREBRF* is rs12513649, the tag SNP identified in a sample of ∼3k Samoans.[55] Local eGRM has a robust peak around the causal SNP, with the window with the lowest p value falling just short of the genome-wide cutoff (0.2 $\log_{10}$ units below, or approximately 7.2×10^{-7} compared with a cutoff of approximately 4.5×10^{-7}). The distances between the observed peaks and the causal SNP rs373863828 are in line with the distances between peaks and causal variants at frequency 0.02 observed in our simulations (Table S3; Figure S4). Local GRM, which uses the relatedness calculated based on the observed variants, does not have a peak at the causal locus. This is in line with our explanation of the simulation results, that the more accurate measurement of local relatedness provided by the local eGRM enhances the local GRM approach. In the GWAS results, a single SNP is visibly elevated to the right of *CREBRF*, which is 1.5 $\log_{10}$ units below the genome-wide cutoff. Correlations among p values produced by the three methods across the chromosome, which are generally low, are shown in Figures S8 and S9.

Chromosome 16 harbors *FTO*, which is a well-studied locus associated with human obesity known to have high allelic heterogeneity.[59] The three methods perform similarly here (Figure S7) with a robust, but not significant, peak at this locus. The most significant p value of local GRM, eGRM, and GWAS are 0.87, 1.2, and 1.4 $\log_{10}$ units below their respective cutoffs.

Discussion

We developed an approach to QTL mapping that uses estimated ARGs to characterize local relatedness and show that it provides advantages complementary to several existing approaches to QTL mapping with SNP array data. Specifically, our approach is robust to allelic heterogeneity and can assist in identifying QTLs even when the causal loci are not well tagged by any single array SNP and cannot be imputed because of a lack of a population-specific reference panel.

In cases of allelic heterogeneity, a marker variant can be linked with multiple causal variants that can have opposing effects, leading to their association signals interfering and causing difficulties for a GWAS. The local GRM and local eGRM approaches we consider here both naturally accommodate allelic heterogeneity, because even if there are multiple causal variants in a trait-relevant region, it should still be the case that individuals who are more closely related in the region tend to be more similar on the phenotype. One locus that is known for its strong association with BMI and that also displays allelic heterogeneity is *FTO* on chromosome 16. We found that all three methods were able to detect some evidence of a QTL in this region in the Native Hawaiian Cohort of the Multiethnic Cohort MEC.[71] However, our simulations show that tests of the local eGRM have an advantage in detecting loci with allelic heterogeneity when the causal loci are untyped. The local GRM and eGRM approaches differ in that the local eGRM takes into account information about local branch lengths drawn from mutations occurring in neighboring regions, since trees in neighboring regions tend to share many of the same coalescent events with the focal region. Thus, the local eGRM can sometimes capture local genetic relatedness more accurately than the local GRM, particularly for small testing windows.

We also analyzed *CREBRF*, which has a causal variant for adiposity that is at high frequency in humans with Polynesian ancestry but that has previously not been well imputed because of a lack of population-specific imputation panel. Although none of the QTL detection methods tested reached genome-wide significance, local eGRM had a robust, visually striking peak in the vicinity of *CREBRF* that nearly reached genome-wide significance, which was less marked in the results of our GWAS and even less in the results of local GRM. In cases like this, knowledge of the underlying shared ancestry in the region can stand in for imputation, in that, even if some variants are untyped, the local marginal trees may contain branches that pick out the same or nearly the same sets of haplotypes

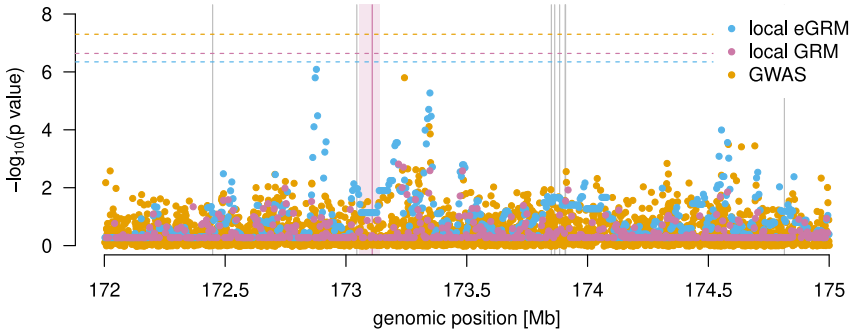

Figure 6. Association results for Hawaiian cohort of the MEC around *CREBRF*

Blue dots are per-window negative $\log_{10}$ p values for the local eGRM, orange dots are per-SNP results for GWAS, and pink dots are per-window results for local GRM. The horizontal dashed lines are the genome-wide significance cutoffs (5×10^{-8} for GWAS, 2.3×10^{-7} for local GRM, and 4.5×10^{-7} for local eGRM). The vertical shaded area delimits the coordinates of *CREBRF*, and the pink vertical line within it is the location of the causal SNP rs373863828 (GenBank: NM_153607.3) (c.1370G>A [p.Arg457Gln]). The gray vertical lines are SNPs found in the GWAS catalog that were found to be associated to traits "body mass index" and "weight." Other than rs12513649, which was found in a sample of 3,072,[55] these hits were found in samples of at least 158,284 individuals.[90–96]

(Figure 1D). In some sense, imputation followed by a GWAS on imputed markers is conceptually roundabout: an ARG-like structure is often inferred in order to perform imputation, such as by the Li & Stephens approach,[97] which is also the basis for recent approaches to ARG estimation.[40,41] In our approach, we test the structure on which imputation is performed—that is, the approximate local tree—rather than the imputed variants. Such an approach may facilitate the identification of trait-associated loci in understudied populations.

Our method adds to a long list of approaches for identifying trait-associated loci. First, and perhaps most obviously, our method is a tree-based version of methods to test local GRMs for concordance with a phenotype.[49–54] As discussed above, the advantage of our approach over such methods stems from better estimates of local relatedness achieved by estimated ARGs. Further, our method can be seen as a generalization of identity-by-descent (IBD) mapping,[53,98,99] where our method considers putative IBD over short regions as estimated by local trees in addition to the relatively large (multiple centiMorgan) segments that can be identified as recent IBD. IBD mapping, in turn, can be seen as a generalization of linkage mapping that uses IBD among pairs of people who are not closely related rather than only among close relatives. Our method is also closely related to haplotype mapping, and in particular approaches to haplotype mapping that estimate tree-like structures to describe relatedness among sets of haplotypes.[20–23] Finally, our method adds to a tradition of approaches for identifying trait-involved loci that are explicitly tree based.[5–19] Whereas most previous tree-based approaches to mapping were limited to samples in the dozens because of difficulties with ARG estimation, modern ARG estimation frameworks enable a substantial gain in power using sample sizes into the thousands.

Another recent approach that used large estimated ARGs to identify trait-associated loci came from Zhang and colleagues,[42] who developed an ARG estimation method, ARG-Needle, to identify trait-associated variants in a sample of more than 300,000 people. Our approach is complementary to theirs. Whereas Zhang and colleagues also identify and leverage the eGRM, which they term the ARG-GRM, they use it for genome-wide tasks such as heritability estimation rather than calculating the eGRM for a local region. In their searches for trait-associated variants, they sample mutations from the ARG and test them individually, which is equivalent to testing branches or clades from the ARG. A promising future direction is to combine our approach with theirs, using our method to prioritize regions and then sampling mutations within that region in an attempt to localize the signal.

Both our results and those of Zhang and colleagues[42] point to advantages of using estimated ARGs in situations in which genotype data are incomplete. In contrast, with complete data on underlying genetic variants, our simulations suggest that our tree-based approach is outperformed by other methods. This is sensible: in the scenarios we simulate, if all variants are known, then the tree provides no additional information. The local coalescent trees are helpful when data are incomplete because they provide a guide to the structure of unobserved mutations.

Local coalescent trees could in principle outperform full sequence data in other settings as well. One such setting is in combination with a model for natural selection on trait-associated variants. Selection will distort local trees, and thus signals of selection inferred from the trees might be used to prioritize trees or clades for investigation with respect to traits that could have been under selection in the history of the sample. Another relevant setting is ascertainment, in which individuals are sampled for inclusion in the study on the basis of their trait values. Such ascertainment mimics (extremely recent) natural selection in that it creates a sample of individuals selected on their phenotypes, and distortions in local trees under ascertainment could serve as evidence that the local region is trait associated.

Our work here is an initial report of some advantages of a tree-based local relatedness approach to QTL mapping. The limitations of our current approach raise promising avenues for future investigation.

Here, we included all branches in the ARG within a genomic window in the eGRM, and we weighted them as a function of their branch length, span in the genome, and the number of tips descending from them. In principle, one could alter the weighting of branches, even

choosing to leave some branches out, perhaps to form a time-specific eGRM.[44] The absolute value of GWAS effect sizes is routinely observed to be negatively correlated with minor allele frequency, a pattern that could be explained by stabilizing selection on traits keeping large-effect variants at low frequency.[100–102] The "α-model" we use to simulate effect sizes is in line with the basic observation of larger effect sizes at lower-frequency variants, as is our practice of estimating a GRM in which variants are standardized by a factor proportional to $\sqrt{p(1-p)}$, which is equivalent to assuming that the contribution to heritability of a causal variant does not depend on its frequency. However, the α-model is only a loose match to the observed distribution of effect sizes as a function of allele frequency,[102,103] and using approaches to normalization or weighting of branches informed by more refined models of selection on trait-associated variation could improve performance in real data.

We did not consider errors in estimation of the ARG, instead treating marginal tree estimates from Relate as if they represented the true marginal trees. Figure S2 shows that using estimated trees from array data decreases power compared with using the true trees. Our main focus here is hypothesis testing, but a broader consideration of local eGRMs in attempts to estimate locally explained heritability will entail consideration of the effect of errors in ARG reconstruction on heritability estimates and their standard errors.

The variance-components model underlying our approach also assumes that in QTL windows, every branch will be associated with some normally distributed effect on the phenotype. This assumption is reasonable for QTLs with high levels of allelic heterogeneity but is not designed for cases with only one causal variant, and it is likely the reason for the diminished advantage of local eGRM over a GWAS in the simulations with one causal variant compared to those with allelic heterogeneity. It is worth exploring the application of methods that allow sparse architectures to the eGRM,[104] which might present advantages even in cases with sparse genetic architectures due to reduced multiple testing. Further, although here we test an additive architecture, it may be possible to modify our approach to look for QTLs that act in a dominant, recessive, or locally epistatic manner by computing modified local eGRMs.[19,105,106]

We tested various methods of correcting for population stratification. For this, we simulated phenotypes on the basis of the real Native Hawaiian data. This allowed us to accurately capture the complicated population history. We showed that the false-positive rate of local eGRM QTL mapping can be controlled via the LOCO eGRM in a two-step procedure. However, there are many remaining avenues to explore regarding population stratification and assortative mating.

We used ARGs estimated by Relate[41] for both simulated and real data. Although tsinfer+tsdate[40,43] scales to much larger sample sizes than Relate, we used Relate because of evidence that it provides more accurate branch length estimates than tsinfer+tsdate,[107] which is reflected the observation of Fan and colleagues[44] that Relate-based eGRMs are more accurate than those formed from tsinfer+tsdate. An approach to QTL mapping based on topology rather than branch length might open up application to much larger sample sizes via tsinfer+tsdate. ARG-Needle,[42] which was recently released for general use, may also allow the procedures developed here to be used with tens or hundreds of thousands of individuals.

We tested for QTLs of size 5 kilobases or 10 kilobases. These sizes are arbitrary, but the approach of a window-based test also allows for flexibility. For example, windows could be chosen to form gene-level tests. It is likely possible to reduce the number of tests performed by adaptively choosing windows on the basis of the extent to which tree topologies change within the window. For example, in the test of *CREBRF* in Native Hawaiians, a single marginal tree spanned the entirety of *CREBRF*, likely because the genotyping array included few SNPs within *CREBRF*. Testing this marginal tree only once is more sensible than testing identical windows repeatedly, as our current approach does. Building a better approach will likely require an understanding of how estimated tree topologies change as a function of sample size, the local density of typed SNPs, the local recombination rate, and the ages of the QTL and the causal mutation.

Importantly, the method as currently implemented is computationally intense because of three time-consuming steps: estimating approximate ARGs with Relate, computing the eGRM, and fitting a linear mixed model with GCTA. Regarding the first step, although Relate is much faster than previous approaches to ARG estimation, it can still be time consuming to run on large samples. As mentioned above, tsinfer+tsdate scales to larger samples than Relate, at the cost of less accurate branch length estimates.[107] ARG-Needle is reported to run on very large samples. Improvement of tsinfer+tsdate's branch length estimates or using ARG-Needle could allow the estimation of approximate ARGs suitable for our approach on larger samples. The second step, fitting the eGRM, is slow in very large samples because the computation entails a component for every branch on the ARG. As noted above, our approach to eGRM estimation is slower than it might be because we touch redundant branches of local trees multiple times, which can be ameliorated via a branch-based approach to computing local eGRMs.[48] Further, as noted by Zhang and colleagues,[42] it is possible to take a Monte Carlo approach to eGRM estimation, placing mutations on the ARG randomly at high rate. The GRM computed from these randomly placed mutations is an approximate eGRM that retains many of the advantages of the true eGRM. Fortunately, the third step of running the mixed model has been a major target for speedups among statistical geneticists, so we will be able to adopt existing approaches when working with larger samples.[86,108]

Since before the time of Zaccheaus (Luke 19:4), people have been climbing trees to get a better view. Here, we explored a coalescent-tree-based approach to QTL mapping, showing that the expectation of the local GRM conditional on the ARG allows detection of QTLs under allelic

heterogeneity or in cases in which genotype imputation is difficult. Local eGRMs are only one case of a general framework for computing ARG-based analogues of statistics typically computed on genetic variants.[48] The advantages of this general framework for a broad range of statistical- and population-genetic tasks have yet to be explored.

Data and code availability

The datasets used for the analyses described in this manuscript were obtained from dbGaP with accession numbers phs 000220.v2.p2 and phs002183.v1.p1. The code used to perform the analyses in these paper is available on its GitHub page (https://github.com/vivilink/sycamore/), v.1.0.0 (DOI: https://doi.org/10.5281/zenodo.10011427). The repository also contains the plotting scripts and documentation.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2023.10.017.

Acknowledgments

We thank G. Coop, J. Felsenstein, G. Gorjanc, A. Harpak, H. Lee, M. Nordborg, P. Ralph, D. Runcie, and members of the Edge, Mooney, and Pennell labs for helpful conversations. We also thank the editor and anonymous peer reviewers for helpful comments that improved the manuscript.

We acknowledge support from NIH grants R35GM137758 to M.D.E.; R01HG011646, R01HG012605, and R35GM142783 to C.W.K.C.; R01HG012133 to N.M.; and F31HG012159 to B.D.

We acknowledge support from NIH grants R35GM137758 to M.D.E.; R01HG011646, R01HG012605, and R35GM142783 to C.W.K.C; R01HG012133 to N.M.; and F31HG012159 to B.D. Funding support for the PAGE Multiethnic Cohort study was provided through the National Cancer Institute (R37CA54281, R01CA6364, P01CA33619, U01CA136792, and U01CA98758) and the National Human Genome Research Institute (U01HG004802). Assistance with phenotype harmonization, SNP selection, data cleaning, meta-analyses, data management and dissemination, and general study coordination was provided by the PAGE Coordinating Center (U01HG004801-01).

Declaration of interests

The authors declare no competing interests.

Received: April 8, 2023
Accepted: October 27, 2023
Published: December 7, 2023

References

1. Balding, D.J., Moltke, I., and Marioni, J. (2019). Handbook of Statistical Genomics*1* (John Wiley & Sons).
2. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 years of gwas discovery: Biology, function, and translation. Am. J. Hum. Genet. *101*, 5–22.
3. Rosenberg, N.A., and Nordborg, M. (2002). Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nat. Rev. Genet. *3*, 380–390.
4. Griffiths, R.C., and Marjoram, P. (1996). Ancestral inference from samples of DNA sequences with recombination. J. Comput. Biol. *3*, 479–502.
5. Templeton, A.R., Boerwinkle, E., and Sing, C.F. (1987). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. l. basic theory and an analysis of alcohol dehydrogenase activity in drosophila. Genetics *117*, 343–351.
6. McPeek, M.S., and Strahs, A. (1999). Assessment of linkage disequilibrium by the decay of haplotype sharing, with application to fine-scale genetic mapping. Am. J. Hum. Genet. *65*, 858–875.
7. Larribe, F., Lessard, S., and Schork, N.J. (2002). Gene Mapping via the Ancestral Recombination Graph. Theor. Popul. Biol. *62*, 215–229.
8. Morris, A.P., Whittaker, J.C., and Balding, D.J. (2002). Fine-scale mapping of disease loci via shattered coalescent modeling of genealogies. Am. J. Hum. Genet. *70*, 686–707.
9. Zöllner, S., and Pritchard, J.K. (2005). Coalescent-Based Association Mapping and Fine Mapping of Complex Trait Loci. Genetics *169*, 1071–1092.
10. Minichiello, M.J., and Durbin, R. (2006). Mapping trait loci by use of inferred ancestral recombination graphs. Am. J. Hum. Genet. *79*, 910–922.
11. Mailund, T., Besenbacher, S., and Schierup, M.H. (2006). Whole genome association mapping by incompatibilities and local perfect phylogenies. BMC Bioinf. *7*, 454.
12. Tachmazidou, I., Verzilli, C.J., and Iorio, M.D. (2007). Genetic association mapping via evolution-based clustering of haplotypes. PLoS Genet. *3*, e111–e115.
13. Kimmel, G., Karp, R.M., Jordan, M.I., and Halperin, E. (2008). Association mapping and significance estimation via the coalescent. Am. J. Hum. Genet. *83*, 675–683.
14. Wu, Y. (2008). Association mapping of complex diseases with ancestral recombination graphs: models and efficient algorithms. J. Comput. Biol. *15*, 667–684.
15. Besenbacher, S., Mailund, T., and Schierup, M.H. (2009). Local phylogeny mapping of quantitative traits: higher accuracy and better ranking than single-marker association in genomewide scans. Genetics *181*, 747–753.
16. Zhang, Z., Zhang, X., and Wang, W. (2012). Htreeqa: Using semi-perfect phylogeny trees in quantitative trait loci study on genotype data. G3 (Bethesda). *2*, 175–189.
17. Burkett, K.M., Greenwood, C.M.T., McNeney, B., and Graham, J. (2013). Gene genealogies for genetic association mapping, with application to crohn's disease. Front. Genet. *4*, 260.
18. Thompson, K.L., and Kubatko, L.S. (2013). Using ancestral information to detect and localize quantitative trait loci in genome-wide association studies. BMC Bioinf. *14*, 200.
19. Thompson, K.L., Linnen, C.R., and Kubatko, L. (2016). Tree-based quantitative trait mapping in the presence of external covariates. Stat. Appl. Genet. Mol. Biol. *15*, 473–490.
20. Liu, J.S., Sabatti, C., Teng, J., Keats, B.J., and Risch, N. (2001). Bayesian analysis of haplotypes for linkage disequilibrium mapping. Genome Res. *11*, 1716–1724.
21. Morris, A.P. (2005). Direct analysis of unphased snp genotype data in population-based association studies via bayesian partition modelling of haplotypes. Genet. Epidemiol. *29*, 91–107.

22. Selle, M.L., Steinsland, I., Lindgren, F., Brajkovic, V., Cubric-Curik, V., and Gorjanc, G. (2020). Hierarchical modelling of haplotype effects on a phylogeny. Front. Genet. *11*, 531218.
23. Crouse, W.L., Kelada, S.N.P., and Valdar, W. (2020). Inferring the allelic series at qtl in multiparental populations. Genetics *216*, 957–983.
24. Cantor, R.M., Lange, K., and Sinsheimer, J.S. (2010). Prioritizing gwas results: A review of statistical methods and recommendations for their application. Am. J. Hum. Genet. *86*, 6–22.
25. Pe'er, I., Yelensky, R., Altshuler, D., and Daly, M.J. (2008). Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet. Epidemiol. *32*, 381–385.
26. Platt, A., Vilhjálmsson, B.J., and Nordborg, M. (2010). Conditions Under Which Genome-Wide Association Studies Will be Positively Misleading. Genetics *186*, 1045–1052.
27. Flister, M.J., Tsaih, S.W., O'Meara, C.C., Endres, B., Hoffman, M.J., Geurts, A.M., Dwinell, M.R., Lazar, J., Jacob, H.J., and Moreno, C. (2013). Identifying multiple causative genes at a single gwas locus. Genome Res. *23*, 1996–2002.
28. Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. Plant Methods *9*, 29.
29. Hormozdiari, F., Zhu, A., Kichaev, G., Ju, C.J.T., Segrè, A.V., Joo, J.W.J., Won, H., Sankararaman, S., Pasaniuc, B., Shifman, S., and Eskin, E. (2017). Widespread allelic heterogeneity in complex traits. Am. J. Hum. Genet. *100*, 789–802.
30. Terwilliger, J.D., and Weiss, K.M. (1998). Linkage disequilibrium mapping of complex disease: fantasy or reality? Curr. Opin. Biotechnol. *9*, 578–594.
31. Jansen, R., Hottenga, J.J., Nivard, M.G., Abdellaoui, A., Laport, B., de Geus, E.J., Wright, F.A., Penninx, B.W.J.H., and Boomsma, D.I. (2017). Conditional eQTL analysis reveals allelic heterogeneity of gene expression. Hum. Mol. Genet. *26*, 1444–1451.
32. Abell, N.S., DeGorter, M.K., Gloudemans, M.J., Greenwald, E., Smith, K.S., He, Z., and Montgomery, S.B. (2022). Multiple causal variants underlie genetic associations in humans. Science (New York, N.Y.) *375*, 1247–1254.
33. Marchini, J., and Howie, B. (2010). Genotype imputation for genome-wide association studies. Nat. Rev. Genet. *11*, 499–511.
34. Das, S., Abecasis, G.R., and Browning, B.L. (2018). Genotype imputation from large reference panels. Annu. Rev. Genom. Hum. Genet. *19*, 73–96.
35. Stephens, M., and Scheet, P. (2005). Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am. J. Hum. Genet. *76*, 449–462.
36. Edge, M.D., Gorroochurn, P., and Rosenberg, N.A. (2013). Windfalls and pitfalls: Applications of population genetics to the search for disease genes. Evol. Med. Public Health *2013*, 254–272.
37. Huang, L., Li, Y., Singleton, A.B., Hardy, J.A., Abecasis, G., Rosenberg, N.A., and Scheet, P. (2009). Genotype-imputation accuracy across worldwide human populations. Am. J. Hum. Genet. *84*, 235–250.
38. Jewett, E.M., Zawistowski, M., Rosenberg, N.A., and Zöllner, S. (2012). A Coalescent Model for Genotype Imputation. Genetics *191*, 1239–1255.
39. Lin, M., Caberto, C., Wan, P., Li, Y., Lum-Jones, A., Tiirikainen, M., Pooler, L., Nakamura, B., Sheng, X., Porcel, J., et al. (2020). Population-specific reference panels are crucial for genetic analyses: an example of the CREBRF locus in Native Hawaiians. Hum. Mol. Genet. *29*, 2275–2284.
40. Kelleher, J., Wong, Y., Wohns, A.W., Fadil, C., Albers, P.K., and McVean, G. (2019). Inferring whole-genome histories in large population datasets. Nat. Genet. *51*, 1330–1338.
41. Speidel, L., Forest, M., Shi, S., and Myers, S.R. (2019). A method for genome-wide genealogy estimation for thousands of samples. Nat. Genet. *51*, 1321–1329.
42. Zhang, B.C., Biddanda, A., Gunnarsson, Á.F., Cooper, F., and Palamara, P.F. (2023). Biobank-scale inference of ancestral recombination graphs enables genealogical analysis of complex traits. Nat. Genet. *55*, 768–776.
43. Wohns, A.W., Wong, Y., Jeffery, B., Akbari, A., Mallick, S., Pinhasi, R., Patterson, N., Reich, D., Kelleher, J., and McVean, G. (2022). A unified genealogy of modern and ancient genomes. Science *375*, eabi8264.
44. Fan, C., Mancuso, N., and Chiang, C.W.K. (2022). A genealogical estimate of genetic relationships. Am. J. Hum. Genet. *109*, 812–824.
45. Wang, S., Ge, S., Colijn, C., Biller, P., Wang, L., and Elliott, L.T. (2021). Estimating genetic similarity matrices using phylogenies. J. Comput. Biol. *28*, 587–600.
46. McVean, G. (2009). A genealogical interpretation of principal components analysis. PLoS Genet. *5*, 10006866–e1000710.
47. Speed, D., and Balding, D.J. (2015). Relatedness in the post-genomic era: is it still useful? Nat. Rev. Genet. *16*, 33–44.
48. Ralph, P., Thornton, K., and Kelleher, J. (2020). Efficiently Summarizing Relationships in Large Samples: A General Duality Between Statistics of Genealogies and Genomes. Genetics *215*, 779–797.
49. Wang, X., Morris, N.J., Zhu, X., and Elston, R.C. (2013). A variance component based multi-marker association test using family and unrelated data. BMC Genet. *14*, 17.
50. Sasaki, E., Zhang, P., Atwell, S., Meng, D., and Nordborg, M. (2015). missing" g x e variation controls flowering time in arabidopsis thaliana. PLoS Genet. *11*, e1005597.
51. Nagamine, Y., Pong-Wong, R., Navarro, P., Vitart, V., Hayward, C., Rudan, I., Campbell, H., Wilson, J., Wild, S., Hicks, A.A., et al. (2012). Localising loci underlying complex trait variation using regional genomic relationship mapping. PLoS One *7*, e46501.
52. Uemoto, Y., Pong-Wong, R., Navarro, P., Vitart, V., Hayward, C., Wilson, J.F., Rudan, I., Campbell, H., Hastie, N.D., Wright, A.F., and Haley, C.S. (2013). The power of regional heritability analysis for rare and common variant detection: simulations and application to eye biometrical traits. Front. Genet. *4*, 232.
53. Gusev, A., Bhatia, G., Zaitlen, N., Vilhjalmsson, B.J., Diogo, D., Stahl, E.A., Gregersen, P.K., Worthington, J., Klareskog, L., Raychaudhuri, S., et al. (2013). Quantifying missing heritability at known gwas loci. PLoS Genet. *9*, 10039933–e1004019.
54. Caballero, A., Tenesa, A., and Keightley, P.D. (2015). The nature of genetic variation for complex traits revealed by gwas and regional heritability mapping analyses. Genetics *201*, 1601–1613.
55. Minster, R.L., Hawley, N.L., Su, C.T., Sun, G., Kershaw, E.E., Cheng, H., Buhule, O.D., Lin, J., Reupena, M.S., Viali, S., et al. (2016). A thrifty variant in crebrf strongly influences body mass index in samoans. Nat. Genet. *48*, 1049–1054.

56. Dina, C., Meyre, D., Gallina, S., Durand, E., Körner, A., Jacobson, P., Carlsson, L.M.S., Kiess, W., Vatin, V., Lecoeur, C., et al. (2007). Variation in fto contributes to childhood obesity and severe adult obesity. Nat. Genet. *39*, 724–726.
57. Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R.B., Elliott, K.S., Lango, H., Rayner, N.W., et al. (2007). A common variant in the fto gene is associated with body mass index and predisposes to childhood and adult obesity. Science *316*, 889–894.
58. Scuteri, A., Sanna, S., Chen, W.M., Uda, M., Albai, G., Strait, J., Najjar, S., Nagaraja, R., Orrú, M., Usala, G., et al. (2007). Genome-wide association scan shows genetic variants in the fto gene are associated with obesity-related traits. PLoS Genet. *3*, e115–e1210.
59. Sobreira, D.R., Joslin, A.C., Zhang, Q., Williamson, I., Hansen, G.T., Farris, K.M., Sakabe, N.J., Sinnott-Armstrong, N., Bozek, G., Jensen-Cody, S.O., et al. (2021). Extensive pleiotropism and allelic heterogeneity mediate metabolic effects of irx3 and irx5. Science *372*, 1085–1091.
60. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common snps explain a large proportion of the heritability for human height. Nat. Genet. *42*, 565–569.
61. Lynch, M., and Walsh, B. (1998). Genetics and Analysis of Quantitative Traits (Sinauer Associates, Inc).
62. Goddard, M., Meuwissen, T., and Daetwyler, H. (2019). Prediction of Phenotype from DNA Variants (John Wiley and Sons, Ltd), pp. 799–820. chapter 28.
63. Adrion, J.R., Cole, C.B., Dukler, N., Galloway, J.G., Gladstein, A.L., Gower, G., Kyriazis, C.C., Ragsdale, A.P., Tsambos, G., Baumdicker, F., et al. (2020). A community-maintained standard library of population genetic models. Elife *9*, 549677–e55039.
64. Lauterbur, M.E., Cavassim, M.I.A., Gladstein, A.L., Gower, G., Pope, N.S., Tsambos, G., Adrion, J., Belsare, S., Biddanda, A., Caudill, V., et al. (2023). Expanding the stdpopsim species catalog, and lessons learned for realistic genome simulations. Elife *12*.
65. Kelleher, J., Etheridge, A.M., and McVean, G. (2016). Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. PLoS Comput. Biol. *12*, 1004842.
66. Baumdicker, F., Bisschop, G., Goldstein, D., Gower, G., Ragsdale, A.P., Tsambos, G., Zhu, S., Eldon, B., Ellerman, E.C., Galloway, J.G., et al. (2022). Efficient ancestry and mutation simulation with msprime 1.0. Genetics *220*, iyab229.
67. Speed, D., Cai, N., UCLEB Consortium, Johnson, M.R., Nejentsev, S., and Balding, D.J. (2017). Re-evaluation of snp heritability in complex human traits. Nat. Genet. *49*, 986–992.
68. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). Gcta: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.
69. Seabold, S., and Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. In 9th Python in Science Conference. https://conference.scipy.org/proceedings/scipy2010/pdfs/seabold.pdf.
70. Liu, Y., and Xie, J. (2020). Cauchy combination test: A powerful test with analytic p-value calculation under arbitrary dependency structures. J. Am. Stat. Assoc. *115*, 393–402.
71. Kolonel, L.N., Henderson, B.E., Hankin, J.H., Nomura, A.M., Wilkens, L.R., Pike, M.C., Stram, D.O., Monroe, K.R., Earle, M.E., and Nagamine, F.S. (2000). A multiethnic cohort in Hawaii and Los Angeles: Baseline characteristics. Am. J. Epidemiol. *151*, 346–357.
72. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation plink: Rising to the challenge of larger and richer datasets. GigaScience *4*, 7.
73. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909.
74. McCaw, Z.R., Lane, J.M., Saxena, R., Redline, S., and Lin, X. (2020). Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. Biometrics *76*, 1262–1272.
75. Loh, P.R., Palamara, P.F., and Price, A.L. (2016). Fast and accurate long-range phasing in a uk biobank cohort. Nat. Genet. *48*, 811–816.
76. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet. *81*, 559–575.
77. Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The encode blacklist: Identification of problematic regions of the genome. Sci. Rep. *9*, 9354–9355.
78. Liu, Y., Chen, S., Li, Z., Morrison, A.C., Boerwinkle, E., and Lin, X. (2019). ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in Sequencing Studies. Am. J. Hum. Genet. *104*, 410–421.
79. Pritchard, J.K., and Rosenberg, N.A. (1999). Use of unlinked genetic markers to detect population stratification in association studies. Am. J. Hum. Genet. *65*, 220–228.
80. Rosenberg, N.A., and Nordborg, M. (2006). A General Population-Genetic Model for the Production by Population Structure of Spurious Genotype–Phenotype Associations in Discrete, Admixed or Spatially Distributed Populations. Genetics *173*, 1665–1678.
81. Vilhjálmsson, B.J., and Nordborg, M. (2013). The nature of confounding in genome-wide association studies. Nat. Rev. Genet. *14*, 1–2.
82. Veller, C., and Coop, G. (2023). Interpreting population and family-based genome-wide association studies in the presence of confounding. Preprint at bioRxiv, 2023.02.26.530052.
83. Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. *38*, 203–208.
84. Kim, S.K., Gignoux, C.R., Wall, J.D., Lum-Jones, A., Wang, H., Haiman, C.A., Chen, G.K., Henderson, B.E., Kolonel, L.N., Le Marchand, L., et al. (2012). Population genetic structure and origins of native hawaiians in the multiethnic cohort study. PLoS One *7*, e47881.
85. Jiang, L., Zheng, Z., Qi, T., Kemper, K.E., Wray, N.R., Visscher, P.M., and Yang, J. (2019). A resource-efficient tool for mixed model association analysis of large-scale data. Nat. Genet. *51*, 1749–1755.
86. Loh, P.R., Tucker, G., Bulik-Sullivan, B.K., Vilhjálmsson, B.J., Finucane, H.K., Salem, R.M., Chasman, D.I., Ridker, P.M., Neale, B.M., Berger, B., et al. (2015). Efficient bayesian

mixed-model analysis increases association power in large cohorts. Nat. Genet. *47*, 284–290.
87. Loh, P.R., Kichaev, G., Gazal, S., Schoech, A.P., and Price, A.L. (2018). Mixed-model association for biobank-scale datasets. Nat. Genet. *50*, 906–908.
88. Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J.A., Ziyatdinov, A., Benner, C., O'Dushlaine, C., Barber, M., Boutkov, B., et al. (2021). Computationally efficient whole-genome regression for quantitative and binary traits. Nat. Genet. *53*, 1097–1103.
89. Zhou, W., Nielsen, J.B., Fritsche, L.G., Dey, R., Gabrielsen, M.E., Wolford, B.N., LeFaive, J., VandeHaar, P., Gagliano, S.A., Gifford, A., et al. (2018). Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat. Genet. *50*, 1335–1341.
90. Tachmazidou, I., Süveges, D., Min, J.L., Ritchie, G.R.S., Steinberg, J., Walter, K., Iotchkova, V., Schwartzentruber, J., Huang, J., Memari, Y., et al. (2017). Whole-genome sequencing coupled to imputation discovers genetic signals for anthropometric traits. Am. J. Hum. Genet. *100*, 865–884.
91. Zhu, Z., Guo, Y., Shi, H., Liu, C.L., Panganiban, R.A., Chung, W., O'Connor, L.J., Himes, B.E., Gazal, S., Hasegawa, K., et al. (2020). Shared genetic and experimental links between obesity-related traits and asthma subtypes in uk biobank. J. Allergy Clin. Immunol. *145*, 537–549.
92. Kichaev, G., Bhatia, G., Loh, P.R., Gazal, S., Burch, K., Freund, M.K., Schoech, A., Pasaniuc, B., and Price, A.L. (2019). Leveraging polygenic functional enrichment to improve gwas power. Am. J. Hum. Genet. *104*, 65–75.
93. Pulit, S.L., Stoneman, C., Morris, A.P., Wood, A.R., Glastonbury, C.A., Tyrrell, J., Yengo, L., Ferreira, T., Marouli, E., Ji, Y., et al. (2019). Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of european ancestry. Hum. Mol. Genet. *28*, 166–174.
94. Akiyama, M., Okada, Y., Kanai, M., Takahashi, A., Momozawa, Y., Ikeda, M., Iwata, N., Ikegawa, S., Hirata, M., Matsuda, K., et al. (2017). Genome-wide association study identifies 112 new loci for body mass index in the japanese population. Nat. Genet. *49*, 1458–1467.
95. Hoffmann, T.J., Choquet, H., Yin, J., Banda, Y., Kvale, M.N., Glymour, M., Schaefer, C., Risch, N., and Jorgenson, E. (2018). A large multiethnic genome-wide association study of adult body mass index identifies novel loci. Genetics *210*, 499–515.
96. Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshiba, S., Narita, A., Konuma, T., Yamamoto, K., Akiyama, M., et al. (2021). A cross-population atlas of genetic associations for 220 human phenotypes. Nat. Genet. *53*, 1415–1424.
97. Li, N., and Stephens, M. (2003). Modeling Linkage Disequilibrium and Identifying Recombination Hotspots Using Single-Nucleotide Polymorphism Data. Genetics *165*, 2213–2233.
98. Albrechtsen, A., Sand Korneliussen, T., Moltke, I., van Overseem Hansen, T., Nielsen, F.C., and Nielsen, R. (2009). Relatedness mapping and tracts of relatedness for genome-wide data in the presence of linkage disequilibrium. Genet. Epidemiol. *33*, 266–274.
99. Browning, S.R., and Thompson, E.A. (2012). Detecting Rare Variant Associations by Identity-by-Descent Mapping in Case-Control Studies. Genetics *190*, 1521–1531.
100. Simons, Y.B., Bullaughey, K., Hudson, R.R., and Sella, G. (2018). A population genetic interpretation of GWAS findings for human quantitative traits. PLoS Biol. *16*, e2002985.
101. Zeng, J., de Vlaming, R., Wu, Y., Robinson, M.R., Lloyd-Jones, L.R., Yengo, L., Yap, C.X., Xue, A., Sidorenko, J., McRae, A.F., et al. (2018). Signatures of negative selection in the genetic architecture of human complex traits. Nat. Genet. *50*, 746–753.
102. Simons, Y.B., Mostafavi, H., Smith, C.J., Pritchard, J.K., and Sella, G. (2022). Simple scaling laws control the genetic architectures of human complex traits. Preprint at bioRxiv.
103. Spence, J.P., Sinnott-Armstrong, N., Assimes, T.L., and Pritchard, J.K. (2022). A flexible modeling and inference framework for estimating variant effect sizes from gwas summary statistics. Preprint at bioRxiv.
104. Zhou, X., Carbonetto, P., and Stephens, M. (2013). Polygenic modeling with bayesian sparse linear mixed models. PLoS Genet. *9*, 10032644–e1003314.
105. Weissbrod, O., Geiger, D., and Rosset, S. (2016). Multikernel linear mixed models for complex phenotype prediction. Genome Res. *26*, 969–979.
106. Hivert, V., Sidorenko, J., Rohart, F., Goddard, M.E., Yang, J., Wray, N.R., Yengo, L., and Visscher, P.M. (2021). Estimation of non-additive genetic variance in human complex traits from a large sample of unrelated individuals. Am. J. Hum. Genet. *108*. 962–798.
107. Brandt, D.Y., Wei, X., Deng, Y., Vaughn, A.H., and Nielsen, R. (2022). Evaluation of Methods for Estimating Coalescence Times Using Ancestral Recombination Graphs. Genetics *221*.
108. Runcie, D.E., and Crawford, L. (2019). Fast and flexible linear mixed models for genome-wide genetics. PLoS Genet. *15*, 10079788–e1008024.