Author/s:
Balding, DJ;Speed, D

Title:
Recent Statistical Innovations in Human Genetics

Date:
2025-09-01

Citation:
Balding, D. J. & Speed, D. (2025). Recent Statistical Innovations in Human Genetics. Annals of Human Genetics, 89 (5), pp.241-254. https://doi.org/10.1111/ahg.12606.

Persistent Link:
https://hdl.handle.net/11343/362718

Annals of Human Genetics

WILEY

Annals of human genetics

REVIEW OPEN ACCESS

Recent Statistical Innovations in Human Genetics

David J. Balding[1,2] | Doug Speed[3]

[1]Melbourne Integrative Genomics, School of Mathematics & Statistics, University of Melbourne, Parkville, Victoria, Australia | [2]UCL Genetics Institute, University College London, London, UK | [3]Quantitative Genetics and Genomics, Aarhus University, Aarhus, Denmark

Correspondence: David J. Balding (dbalding@unimelb.edu.au)

Received: 11 February 2025 | **Revised:** 13 May 2025 | **Accepted:** 19 May 2025

Keywords: ancestral recombination graph | coalescent models | complex traits | forensic DNA profiles | genetic epidemiology | genome-wide association studies | heritability

ABSTRACT

We review three areas of human genetics that have been developed in the past few decades, in which statistical innovation has made a crucial contribution with recent important advances and the potential for further rapid progress. The first topic is the development of mathematical models for the genealogy underlying samples of genome-wide genetic data. Coalescent theory emerged in the 1980s, leaped ahead in the past decade and is now burgeoning into new application areas in population, evolutionary and medical genetics. The second is the development of statistical methods for genome-wide association studies which has made great strides over two decades, including exciting recent developments for association testing based on coalescent theory and improved methods for trait prediction. Finally, we review the statistical ideas that helped resolve the controversies surrounding the introduction of forensic DNA profiling in the early 1990s. Big advances in interpretation of the predominant autosomal DNA profiles have set a benchmark for other areas of forensic science, but the statistical assessment of uniparentally inherited profiles (derived from the mitochondrial DNA or the Y chromosome) remains unsatisfactory.

1 | Introduction

This review covers aspects of three areas of recent statistical advance in human genetics: modelling shared inheritance of DNA segments and what they can tell us about demographic history and selection, modelling relationships between DNA variants and phenotypes and what they can tell us about genetic causes and predicted phenotypes, and identification inferences from DNA profiles. While these topics span much of human genetics research, there is much that is missing, perhaps most notably we do not attempt to cover the vast array of multivariate statistical methods and dynamic models of processes that mediate the genotype–phenotype relationship, including transcriptomics, metabolomics, proteomics and epigenetics. Some of these processes are now studied at the level of individual cells (Sarkar and Stephens 2021; Argelaguet et al. 2021). We also exclude microbiome studies of external and internal human environments (Nearing et al. 2022). Another omission is descriptive analyses of worldwide human genetic variation, such as the People of the British Isles project (Leslie et al. 2015) and the study of indigenous Australian genomes (Silcocks et al. 2023).

2 | Coalescent Models for Genome-Wide Data

The coalescent is a mathematical model for the genealogy underlying a sample of DNA sequences. Coalescent theory was first developed in the 1980s (Kingman 1982; Hudson 1983; Tajima 1983) and inference based on the single-locus coalescent emerged in the 1990s (Griffiths and Tavaré 1994). In the past decade, genome-wide inference based on the coalescent with recombination has become feasible. Initially, this was limited to modest sample sizes but advances in graphical structures for storing genetic datasets, based on coalescent ideas, led to dramatic progress in simulation of large, genome-wide datasets and in inference of

genealogical history and evolutionary parameters. Now a new phase of advances is making inroads into many areas of genetics research, including the analysis of genome-wide association studies (GWAS), and inferences of population demographic history and mechanisms of selection. Joint modelling of the genealogies of entire genomes for large samples has long been a dream of genetics researchers and is fast becoming reality.

2.1 | Introduction to the Coalescent Model

In the absence of recombination, the standard coalescent (Hein et al. 2005; Nordborg 2019) specifies a probability distribution over binary trees, with leaf nodes, internal nodes and branches representing, respectively, observed sequences, shared ancestral sequences and genetic inheritance lineages. Coalescent time is measured backward from the most recent sampling time (at generation 0), and the branches progressively merge (or 'coalesce') backward in time until a single lineage remains, at the most recent common ancestor (MRCA) for the locus.

This backward-in-time approach was revolutionary, being much more computationally efficient than forward-in-time approaches because only ancestors of the sampled genomes need to be tracked, and due to coalescences these rapidly diminish in number going backward in time. However, forward-in-time models such as the Wright–Fisher model retain advantages of flexibility, for example, to incorporate complex models for demography and selection.

A simple version of the coalescent model is derived under the assumption of neutrality in a random-mating population of $N(g)$ diploid individuals, g generations in the past. Typically sampling is at $g = 0$, but samples can also contain ancient DNA sequences sampled at $g > 0$. Assuming that $N(0)$ is large, the coalescence times of different pairs are independent with probability $1/(2N(g))$ in generation g. In the case of the standard coalescent ($N(g)$ constant), this leads to a simple probability model with independent and geometrically distributed times between coalescence events (exponentially distributed in the continuous-time approximation). For more general $N(g)$, the simple probability structure of the standard coalescent can be recovered by a time rescaling proportional to $1/N(g)$.

A simplifying feature of neutral coalescent models is that the genealogical tree is independent of the mutation process. This paves the way for Bayesian statistical inference in which the coalescent model (without mutation) plays the role of a prior distribution that is updated based on the sequence data using a likelihood function incorporating a mutation model. To enhance computational speed, mutation models within coalescent analyses often rely on simplifying assumptions, such as that the ancestral and derived alleles are known and the 'infinite sites' assumption (Kimura 1969) that at most one mutation occurs in the genealogy underlying any site. Bayesian approaches have been implemented within Markov-chain Monte Carlo (MCMC) (Kuhner et al. 1995; Wilson et al. 2003), and within importance sampling, in which genealogical trees are simulated under an approximating model, and then reweighted according to the ratio of likelihoods (Griffiths and Tavaré 1994; Stephens and Donnelly 2000).

One of the first applications of coalescent-based inference to real (Y chromosome) data (Tavaré et al. 1997) used a rejection-sampling approach that developed into a major new branch of statistical inference, approximate Bayesian computation (ABC) (Sisson et al. 2018). ABC was initially formulated to tackle demographic inference from genetic data (Beaumont et al. 2002) by simulating datasets under a prior coalescent model which are accepted/rejected (or more generally reweighted) according to their similarity with the observed data. Thus, multiple data simulations, which are often straightforward to program though can be computationally costly, replace the often-intractable likelihood calculation. Measuring similarity between datasets usually involves the comparison of summary statistics, which should ideally be easy to compute and close to sufficient for the parameters of interest (Nunes and Balding 2010).

2.2 | The Ancestral Recombination Graph

The coalescent with recombination (CwR) (Hudson 1983) reduces to the coalescent at each genome site but incorporates recombinations which occur independently at rates specified by an arbitrary recombination map. Under the CwR, the trees at distinct sites on the same chromosome are often identical if the intersite genetic distance is small, but because of recombinations they tend to diverge as this distance increases. The CwR is remarkably robust to many of its simplifying assumptions (Wakeley et al. 2012), but it performs poorly when both the sample size n and the recombination length are large, in which case forward-in-time models may be preferred (Baumdicker et al. 2022).

Griffiths and Marjoram (1997) developed a convenient formulation of the CwR as an ancestral recombination graph (ARG). They defined the ARG as a random graph with, in addition to coalescence nodes, recombination nodes at which an ancestral lineage branches, with each branch recording the ancestry further back in time on one side of the recombination site. The term ARG is often used for one realisation of the random graph, which may represent the true genealogical history of the sample or an estimate of it, and it can be represented by different data structures (Wong et al. 2024).

In addition to being a stochastic process over time, the CwR can be viewed as a 'spatial' process along the genome, with the state at each genome site being a coalescent tree (Wiuf and Hein 1999). While this spatial sequence of coalescent trees can be derived from an ARG, its data structures are much more efficient. The spatial process is non-Markov, meaning that given the tree at one site, the trees to the left of that site can be informative about trees to its right. The dependence is due to 'trapped non-ancestral' genome segments, which are not ancestral to any of the sampled sequences but lie between segments that are ancestral. The sequentially Markov coalescent (SMC′) (Marjoram and Wall 2006), which improves the earlier SMC (McVean and Cardin 2005), prohibits coalescences that would create trapped non-ancestral sites, providing a process that closely approximates the CwR but is Markov along the genome and hence more tractable.

The development of the succinct tree sequence (TS) graphical structure for representing the ARG of a sample of genome sequences (Kelleher et al. 2016) has proved to be a breakthrough

that has stimulated many advances in ARG-based inference methods. Recombination events are not explicitly stored in the TS, but recombinations that change the tree topology can be extracted as the end points of shared genome segments. The TS achieved dramatic reductions in data storage requirements and in computation time to simulate genome-scale datasets, both under the CwR, for example, using MSPRIME (Kelleher et al. 2016; Baumdicker et al. 2022), and forward-in-time under complex demographic and selection scenarios using SLiM (Haller et al. 2019).

The key to these efficiencies is that genome segments that are very similar across observed sequences are stored as putative ancestral sequences that could have generated the similarity through shared inheritance. This goes well beyond simply recording only once a tree that is conserved over multiple genome sites: each tree node is stored only once, along with the genome segment over which it is conserved. The storage efficiency gains can be achieved even if the inference of shared inheritance is inaccurate, but in analogy with the parsimony principle, the most efficient storage is usually achieved by a good approximation to the actual genealogical history of the observed sequences since their MRCA. Thus, in addition to the data compression gains, the TS data structure can provide an efficient basis for inference of evolutionary and demographic parameters (Fan et al. 2025). It follows that, unlike most other forms of data compression, the compressed dataset is more useful than the original (Kelleher et al. 2019) and decompression is rarely needed.

2.3 | Software for Inferring ARGs

The Bayesian MCMC software ARGweaver (Rasmussen et al. 2014) provided a major advance in genome-wide ARG inference. It remains one of the most accurate methods for inferring ARGs but is limited to $n \sim 10$. ARGweaver assumes the SMC $'$ model, whereas ARGinfer (Mahmoudi et al. 2022) further improves inference by using an augmented TS under the full CwR. An extension ARGweaver-D incorporates population structure and was used to infer ancient gene flow among hominins (Hubisz et al. 2020). SINGER (Deng et al. 2024), currently under development, is claimed to go beyond the capacity and robustness of ARGweaver in Bayesian ARG inference, increasing feasible n to $\sim 10^2$.

RELATE (Speidel et al. 2019) models branch lengths probabilistically, but a single tree topology is fitted in genome regions. It uses a model for haplotype ancestry similar to that of N. Li and Stephens (2003), which underpins many developments in statistical genetics over the past two decades. RELATE can handle $n \sim 10^4$ and has been applied to infer human genealogical history (Speidel et al. 2025) and mechanisms of selection (Speidel et al. 2019).

TSinfer (Kelleher et al. 2019) uses the TS data structure to achieve big gains in computational efficiency for ARG inference. It outputs a single ARG with no measure of uncertainty of inferences, whereas ARGweaver and ARGinfer both approximate a Bayesian posterior distribution allowing flexibility in reporting uncertainty. TSinfer records only the ordering of ARG nodes and not branch lengths, but it is complemented by TSdate which provides Bayesian inference of the lengths of branches in an ARG output by TSinfer (Wohns et al. 2022). TSinfer was the first ARG inference software to handle biobank-scale datasets ($n > 10^5$) and remains the fastest. It has attracted a community of developers that have contributed to making it user-friendly and complemented by a range of ancillary software in TSkit, the TS toolkit.

Brandt et al. (2022) give quantitative comparisons finding broadly that, when computationally feasible, ARGweaver is more accurate than RELATE which is more accurate than TSinfer, the reverse ordering of the feasible n.

2.4 | Using ARGs to Advance Genetics Research

Efficient ARG inference is revolutionising genetics research because the summary statistics that have traditionally been used to extract information from genome sequence data are generally more statistically efficient when replaced by ARG-derived analogues, and these are often also more computationally efficient (Ralph 2019; Ralph et al. 2020). Among many other examples, some outlined below, Fan et al. (2022) find that ARG-derived measures of relatedness are superior to widely used correlation-based coefficients, and Salehi Nowbandegani et al. (2023) present ARG-based measures of linkage disequilibrium (LD) that are more efficient than traditional pairwise LD statistics.

In particular, an ARG estimating the genealogy underlying a genetic sample can capture all the information in the sample about the demographic history of the population from which the sequences were drawn. It can therefore be used in place of the original sequence data for inference of population size and structure, and migration rates. Further data reduction via summary statistics, typically with some loss of information, is often required to make inferences tractable. Common summary statistics include allele frequencies (Moorjani and Hellenthal 2023) and ARG-derived estimates of the TMRCA (time since the MRCA) along the genome. H. Li and Durbin (2011) used TMRCA estimates to draw inferences about historic population sizes from a single diploid individual, noting that due to recombinations the genome of an individual represents a sample of partial genomes from many individuals in past generations. The method has been developed to handle larger n, by combining pairwise comparisons (Schiffels and Wang 2020). Demographic inference remains challenging, in particular, it is difficult to distinguish population-size changes from changes in population structure (Mazet et al. 2016).

A powerful illustration of ARG-based inferences was provided by Wohns et al. (2022), who used TSinfer to estimate the genealogical history of 3601 modern and eight ancient genomes from around the world. The reconstructed ARG was used to confirm features of human evolutionary history and to describe the current population structure using pairwise TMRCA estimates extracted from the ARG. The authors also estimated the locations of our human ancestors at different time depths, assuming that an ancestor occupied a central position among the geographic locations of its descendants.

ARG inference has been applied to improve the power of GWAS (B. Zhang et al. 2023), using software ARG-Needle that, like

TSinfer, can handle biobank-scale datasets. The authors find that ARG-Needle is more accurate than TSinfer, but it is slower and uses more memory. By using simulation in inferred genealogies to investigate the effects of unobserved causal variants, it is more efficient than imputation, the prevailing approach which was itself a major advance 15 years earlier (Marchini et al. 2007). Since the ARG is constructed from only the observed sample there is no requirement for a population-specific reference panel as required for imputation. ARG-Needle can assess both rare and common variants in a single analysis, whereas previously different methods were required in separate analyses. It also performs fine-mapping by including the effects of local variants and incorporates effects of relatedness and population structure in a way that is potentially superior to current approaches based on pairwise relatedness coefficients, discussed below.

Genome segments that are similar across two sequences may share a common ancestor and are called identical-by-descent (IBD), although mutations may have occurred since the MRCA. The lengths of inferred IBD segments are widely used as summary statistics for inferences of demographic or evolutionary parameters, or population association (Tian et al. 2022; Chen et al. 2023; Moorjani and Hellenthal 2023) and hence methods to infer IBD segments have been a focus of research (Zhou et al. 2020). Short IBDs generally correspond to more remote shared ancestry, and because these are often poorly inferred, a length threshold, typically around 2 cM, is imposed on the IBDs used for inference.

Given a TS inferred from a genome dataset, for each pair of sequences it is straightforward to partition chromosomes into segments with the same MRCA. This ARG-derived IBD is used within an ABC framework to infer demographic and evolutionary parameters (Huang et al. 2025). Imposing even a very low-length threshold was found to decrease the quality of inferences, which may be counter-intuitive given the inaccuracy of inference for short IBDs. In fact, inferences based on inferred IBD were found to be only moderately less powerful than when true IBDs (available in simulation studies) were used for inference.

CLUES (Stern et al. 2019) and SIA (Hejase et al. 2022), based on MCMC and deep learning algorithms, respectively, both exploit ARG inference to improve on previous methods for inferring selection. For reviews of recent developments in ARG-based research, see Brandt et al. (2024) and Nielsen et al. (2025), while Lewanski et al. (2024) provide a gentle introduction to the ARG with applications in evolutionary genomics.

3 | Models for Association, Prediction and Heritability

GWAS have had a revolutionary impact on human genetics over the past two decades, thanks in large part to advances in statistical methods. Much has improved since one of us reviewed early statistical methods (Balding 2006), just ahead of the groundbreaking Wellcome Trust Case Control Consortium (2007) report of seven GWAS that set a standard for statistical rigor while also introducing innovative methods. Another landmark is the UK Biobank (Bycroft et al. 2018) with sequence data and rich phenotyping for almost 5×10^5 individuals.

For each of n individuals, a GWAS typically consists of the genotypes at m SNPs, and one or more phenotypes (or traits) which can be at the organism level, such as blood pressure or a disease state, or at the molecular level, such as abundances of transcripts or proteins. Different motivations for a GWAS include

1. *Association analysis*: Which SNPs can be confirmed as contributing causally to the trait, either directly or by tagging (being in LD with) an ungenotyped causal variant?
2. *Prediction*: What is the best phenotype estimate for a new individual whose genome-wide genotype is available?
3. *Heritability analysis*: How are causal effects underlying a trait distributed over the genome?

The (narrow-sense) heritability of a trait is the proportion of variance explained by genetic predictors in an additive model. The heritability of a SNP increases with both its effect size and its frequency, providing a useful currency for measuring its genetic impact on a trait. By combining over SNPs, we can estimate the heritability of genome regions or quantify the impact of quantitative predictors such as measures of LD.

3.1 | Statistical Models for GWAS

We write x_j for the length-n vector of genotypes at the jth SNP. While genotypes for diploid individuals may initially be coded as 0, 1 or 2 copies of a reference allele, we assume that the x_j have been standardised so that $\mathbb{E}[x_j] = 0$ and $\mathbb{V}\mathrm{ar}[x_j] = 1$ for each j. The workhorse of genome-wide genetic inference is the multiple linear regression model:

$$y = \epsilon + \sum_{j=1}^{m} \beta_j x_j, \tag{1}$$

where y is a vector of phenotype values, assumed here to be standardised, and ϵ is a vector including both environmental contributions to the trait and genetic effects not captured by the x_j. The effect size β_j is either the causal effect of SNP j itself or a component of the effect of an unobserved variant that is represented only by SNP j. Note that LD can inflate single-SNP estimates $\hat{\beta}_j$, for example, $\hat{\beta}_j$ can differ significantly from zero even if $\beta_j = 0$.

Only additive effects are included in (1), ignoring both dominance and epistasis which play a role in many traits but usually make only small contributions to complex traits (Wei et al. 2014; Z. Zhu et al. 2015). The summation term in (1) is referred to as a polygenic score (PGS) or, for binary traits, a polygenic risk score (PRS). A fitted PGS (with the β_j replaced by $\hat{\beta}_j$) can be used for prediction, see below.

While (1) has the form of a common statistical model, it is ill-posed in practice for genome-wide inference because typically $m \gg n$ ('too many predictors'). Even if $m < n$, the β_j are not identifiable in the presence of LD (dependence among the x_j). Both these problems are insoluble, but there are various ways to 'regularise' (1) to obtain useful solutions. It is important to note that no one regularisation approach is correct, different

approaches can lead to noticeable differences in inferences and individual β_j can still be poorly estimated.

Early regularisation approaches emphasised subset selection, in which a large fraction of SNPs is excluded as being unlikely to contribute causally to the trait. However, many complex traits have turned out to be highly complex, with a large number of weak causal effects, dispersed widely across the genome. In recognition of this complexity, the term 'omnigenic' has been coined (Boyle et al. 2017), extending the traditional hierarchy of monogenic, oligogenic and polygenic. Weak effects can be cumulatively important, yet individually indistinguishable from no effect, which makes subset selection unsatisfactory.

Since J. Yang et al. (2010), a penalisation approach to regularisation, which 'shrinks' estimates $\hat{\beta}_j$ towards zero, has been favoured. While the penalty function need not be a probability density, it often is and it is convenient to refer to it as a prior for the β_j even though the analysis may not be Bayesian. Ideally, the prior should represent the actual distribution of effect sizes, about which previous GWAS can be informative but ascertainment effects limit this approach.

3.2 | Choice of Effect-Size Prior

Because the sign of β_j can be swapped by switching allele labels, it is common to assume a prior density that is symmetric about zero and hence $\mathbb{E}[\beta_j] = 0$. J. Yang et al. (2010) imposed the same Gaussian prior on each β_j, which has its origins in the polygenic model of Fisher (1919), also called the infinitesimal model with the prediction method based on it, similar to ridge regression, called BLUP. However, the 'thin' tails of the Gaussian provide a poor representation of actual effect sizes. The Laplace prior, corresponding to Lasso regression, also attracted interest due to its simple form and thicker tails. Early applications focussed on its sparsity property due to the posterior mode often arising at $\beta_j = 0$.

Many other effect-size priors have since been proposed (H. Zhu and Zhou 2020). Most attention has focused on the shape of the prior density, particularly in the tail thickness and whether there is a singularity or point probability mass at zero. The use of mixture priors allowed further benefits from the additional flexibility. A common form of the two-component mixture is a 'slab and spike' prior, which includes a spike (probability mass or high/infinite density) around $\beta_j = 0$, while the slab spans the plausible range for β_j given $\beta_j \neq 0$. An alternative choice is the elastic net prior, a mixture of Gaussian and Laplace.

A wide range of mixture priors can be constructed as a base distribution, usually Gaussian or sometimes Laplace, conditional on a parameter which follows a 'mixing distribution' that can be discrete or continuous, often a gamma or half-Cauchy distribution. The hierarchy can be extended by allowing the mixing distribution to itself depend on a parameter that has a higher level mixing distribution. Hoggart et al. (2008) made early use of this approach with a Laplace-gamma mixture prior, emphasising sparse solutions. PRS-CS (Ge et al. 2019) is a more recent example implementing a Gauss-gamma-gamma prior with three parameters that are chosen to obtain finite density at the origin and a desired tail thickness.

3.3 | Modelling the Prior Variance (Expected Heritability)

While there has been much focus on choice of prior density, typically the same prior is adopted for every SNP, despite overwhelming evidence that effect sizes vary with SNP properties known a priori, such as minor allele fraction (MAF), measures of LD and functional annotations. Due to standardisation, $\sigma_j^2 = \mathbb{V}\mathrm{ar}[\beta_j] = \mathbb{E}[\beta_j^2]$, the expected heritability of SNP j. Assuming that σ_j^2 is constant over SNPs is called the *uniform heritability model* and is frequently adopted by GWAS software.

While some attention had previously been paid to leveraging MAF, LD and genome function to improve GWAS analyses, Speed et al. (2022) set out a systematic way to incorporate them as predictors of σ_j^2 in a heritability model. A different approach is to model for each SNP the probability that it is causal (Hu et al. 2017). The most appropriate heritability model can vary according to the analysis goals, but for each of the three goals set out above, improving on the uniform heritability model can lead to substantial inferential gains as we discuss below.

Treating all SNPs the same a priori may seem natural, but assuming that σ_j^2 is constant over j when the x_j are standardised implies a very different relationship between MAF and effect size than the corresponding assumption for unscaled genotypes (taking values 0, 1 or 2). Which of these two assumptions best captures actual relationships between effect size and MAF? They are the cases $\alpha = -1$ and $\alpha = 0$ in the MAF-power heritability model

$$\sigma_j^2 \propto [p_j(1-p_j)]^{1+\alpha}, \tag{2}$$

where p_j is the MAF of SNP j, usually replaced by its sample value. The value of α can be interpreted as a measure of the effects of negative (or purifying) selection. Negative values of α correspond to the unscaled effect sizes tending to increase rapidly as p_j decreases to zero, which could result from selection removing deleterious alleles before they can become common.

When optimised over $\{-1, -0.75, \ldots, 0.25, 0.5\}$, the value $\alpha = -0.25$ gives overall the best fit for human traits (Speed et al. 2017). The value $\alpha = 0$ (uniform unscaled effect sizes), widely used in animal and plant breeding and corresponding to selective neutrality, fits some human traits well. The uniform heritability value $\alpha = -1$, corresponding to strong purifying selection, fits human traits poorly despite being the most widely assumed value.

The poor performance of the uniform heritability model on real data has often been masked because it is uncritically adopted in simulation studies. Any simulation that treats (standardised) SNPs the same irrespective of MAF, LD or genome context is implicitly assuming uniform heritability, and, because that model is unrealistic, results from the simulation study may be misleading.

Recognition of the importance of MAF in modelling effect sizes dates back at least to J. Yang et al. (2015), who assumed that σ_j^2 is constant over SNPs within each of 10 MAF classes. This approach avoids the non-linearity of (2) at the cost of nine

additional parameters and with no requirement for smoothness over neighbouring classes.

In addition to MAF, LD impacts heritability. For example, if two SNPs are in perfect LD they have the same genetic variation as a single SNP and so the heritability of the pair may be similar to that of a single SNP. Empirically, per-SNP heritability decreases as LD increases (Speed et al. 2012). One simple step to adjust for its effect is to remove all but one of a set of SNPs in high LD (e.g., $r^2 > 0.98$) before the use of (2).

3.4 | Association Analysis

Showing causality of a SNP requires establishing population association, and also that the association is not explained by LD with other SNPs, either the high LD that is common among SNPs located close together on the genome (disentangling this LD is referred to as fine mapping), or population structure effects mediated by long-range LD. Replication in different studies can help with both these tasks, but a key strategy when assessing the association of a target SNP is to include as covariates independent causal effects, both environmental and genetic.

Devlin and Roeder (1999) introduced a widely used adjustment for population structure effects in association analyses, called the genomic inflation factor (GIF) and defined as the ratio of the median GWAS test statistic to its null value assuming a χ_1^2 null distribution. GIF >1 is interpreted as genome-wide inflation due to effects of population structure, and dividing all test statistics by the GIF was intended to reverse this inflation. This approach relies on an assumption, now understood to often be unrealistic, that causal effects are sparse so that non-null SNPs barely affect the empirical median. An improved approach was to add as covariates to the regression model the leading principal components of the genotype correlation matrix XX'/m, where $'$ denotes transpose and X is the $n \times m$ matrix with jth column equal to x_j (Price et al. 2006).

Linear mixed model (LMM) association analysis (Yu et al. 2006; Kang et al. 2008; Astle and Balding 2009) provides a better approach to adjusting for the effects of other SNPs and population structure. It emerged in the context of non-human organisms and was initially computationally prohibitive for human GWAS, particularly for case-control studies, but computational advances have led to its widespread adoption. A simple form of the LMM for association analysis is

$$y = \epsilon + \beta_0 x_0 + \gamma, \tag{3}$$

where 0 indexes the SNP being tested. The term 'mixed' refers to the inclusion of fixed effects, here $\beta_0 x_0$ and often also covariates such as sex and age, in addition to the genetic random effect γ. To avoid diluting the signal from SNP 0, the effects of SNPs in high LD with it should be excluded from γ. For computational reasons, in practice all SNPs on the same chromosome are excluded, a leave-one-chromosome-out approach. While γ is unobserved, its $n \times n$ variance-covariance matrix K is assumed to be known and is interpreted as a matrix of pairwise kinship coefficients. Although (3) appears to be different from (1), they are in practice similar models: the common choice $K = XX'/m$ is in effect equivalent to replacing γ in (3) with a PGS similar to that in (1). In practice, the difference between (3) and (1) is that the penalisation is not applied to SNP 0, and other SNPs on its chromosome are excluded.

Fitting (3) usually proceeds in two steps, the first using penalisation to fit the PGS, and the second testing $\beta_0 = 0$ using the fitted PGS as an offset. Two of the leading software that take this approach are BOLT-LMM (Loh et al. 2015) and REGENIE (Mbatchou et al. 2021). BOLT-LMM is usually more powerful than REGENIE, because it uses more realistic penalisation when constructing the PGS, whereas REGENIE is faster because it constructs the PGS using an efficient block-based algorithm. LDAK-KVIK (Hof and Speed 2025) replaces the uniform heritability model that is implicit in BOLT-LMM and REGENIE with the MAF-power model (2). This modelling improvement, together with computational innovations, led to LDAK-KVIK having slightly higher statistical power than BOLT-LMM and being slightly more computationally efficient than REGENIE. Even the largest GWAS can now be analysed using the state-of-art LMM approach with only modest computing resources: an LDAK-KVIK analysis with $n > 3.5 \times 10^5$ required around eight CPU hours and five Gb of memory (Hof and Speed 2025).

3.5 | Analyses Based on Summary Statistics

Individual-level genotype and phenotype data are necessary for a primary association analysis, but a major breakthrough in GWAS has been the development of methods to perform subsequent analyses using results from single-SNP association tests (effect-size estimates and p-values), without individual-level data. Most of these summary-statistic methods also compute measures of LD using a reference panel, consisting of the genotypes of individuals with similar ancestry to those in the GWAS.

Summary statistic analyses are typically much faster than analyses of individual-level data, and they avoid confidentiality concerns arising from sharing individual-level data. Although some summary-statistic methods claim to adjust for effects of population structure, it remains important that the primary association analysis generating the summary statistics makes appropriate adjustments (Holmes et al. 2019), such as through the use of an LMM. Summary-statistic methods usually lose some precision relative to analyses based on the full GWAS dataset, but often the loss is small. Summary statistics are now widely used in place of individual-level data for many analyses beyond the initial association analysis, even when individual-level data are available. One immediate benefit comes from facilitating meta-analyses across GWAS. Here, we briefly discuss the use of summary statistics for prediction and heritability analyses.

3.6 | Prediction

The simplest prediction tools are 'classical' PGS with $\hat{\beta}_j$ from single-SNP association analysis. Initially only highly significant SNPs were included, but performance was improved by relaxing the p-value threshold to include more SNPs in the PGS (International Schizophrenia Consortium et al. 2009). Further improvement came from reducing double counting of causal signals by clumping: setting $\hat{\beta}_j = 0$ if SNP j is highly correlated

with a nearby SNP that has a lower association p-value (Choi and O'Reilly 2019). More advanced prediction methods typically obtain $\hat{\beta}_j$ as the posterior mean from a joint analysis, using model (1) regularised by a prior density, as discussed above. These methods vary according to the algorithm (often MCMC or variational Bayes) and the choice of prior distribution for the β_j.

If an individual-level predictor depends on y and the x_j only through vector products of the form $y'x_j$ and $x_j'x_k$, then a summary-statistic version is available as these products can be estimated from the test statistics and reference panel, respectively. LDpred (Vilhjálmsson et al. 2015) was one of the first advanced, summary-statistic PGS. It uses a slab-and-spike mixture prior with a point-mass spike and a Gaussian slab. DBSLMM (S. Yang and Zhou 2020) is another advanced method that uses a deterministic approximation to the posterior mean and a two-Gaussians prior.

Most prediction methods adopt the uniform heritability model, but Q. Zhang et al. (2021) obtained gains in accuracy by modifying existing prediction tools, both using individual-level data and summary statistics, to improve the heritability model. They also quantified the gains from each of the major advances in prediction methods of the past decade, reporting that moving from single-SNP to joint estimation using a single Gaussian prior increased average prediction accuracy by 17% while moving from this prior to a mixture prior led to a 16% increase (with little difference among different mixtures). Finally, replacing the uniform heritability model with the BLD-LDAK model (see below) led to a 14% improvement. Each of these advances is roughly equivalent to increasing n by a quarter. These results provide a guide to efficiency gains from statistical innovations, but actual gains will depend on factors such as n and the trait h^2. In particular, the relative gain from improving the heritability model tends to decrease as the heritability of the trait increases.

In practice, genomic predictions are usually a single phenotype value, without measures of uncertainty. Although many methods work with a posterior distribution for each β_j, the implied posterior distribution for a prediction is computationally challenging and may not be well calibrated due to approximations, for example, in modelling the dependence among SNPs. Many users adopt pre-fitted PGS available from public repositories, for which validation studies are typically available quantifying the average performance of the method in samples from one or more populations.

The performance of a predictor of a quantitative trait is often measured by the squared correlation between predicted and observed phenotypes (R^2). For binary traits, the area under curve (AUC) is often reported or the relative risks at different prediction thresholds. For example, disease prevalence among individuals above the top percentile of PRS values may be reported as a multiple of the population prevalence. Whereas R^2 and AUC are global measures of accuracy and can be approximated from summary statistics, relative risks are more valuable as they can be tailored to the PRS value for an individual, but they require a validation study using individual-level data.

The accuracy of complex trait prediction from SNP data is often low in the general population, although it can be high within high-risk subgroups. PGS have uses other than predicting an individual's future trait value. For example, they can be used to study the relationship of disease risk with other risk factors or disease severity or response to treatment. Khera et al. (2018) seems to have been a turning point in moving from scepticism to a positive attitude towards the value of PGS for complex traits, but there remains scepticism, for example, around incorporating the predictive value of age in PGS (Curtis 2019).

A major challenge for prediction accuracy is the role of the ancestry of the target of prediction (Scutari et al. 2016; Wang et al. 2020). The greater the genetic distance between that individual and the GWAS used to construct the predictor, the less accurate the prediction. However, the impact of this genetic distance can be hard to assess without good data resources. The range of ancestry groups used to construct prediction tools is often limited, so that currently genomic prediction is worse for much of the global human population than for members of a few well-studied populations. However, the problem of prediction across ancestries appears to be largely due to differences in LD: if the primary GWAS is powerful enough to localise causal variants, their predictive value tends to be similar across ancestries.

Jayasinghe et al. (2024) catalogue a wide range of prediction software which vary, for example, in the requirement for users to specify tuning parameters, which may require a training dataset, in the non-SNP predictors that can be included, such as genome annotations, and in claimed suitability for cross-ancestry prediction.

3.7 | Heritability Analysis

The SNP heritability (h^2_{SNP}) of a trait is the heritability tagged by all m GWAS SNPs. Usually, $h^2_{\text{SNP}} < h^2$, the classical heritability estimated by comparing pedigree-relatedness with phenotype similarity. The difference between $\widehat{h^2}$ and $\hat{h}^2_{\text{SNP}}$ is sometimes referred to as 'missing' heritability. Most of this gap is due to rare causal variants that are poorly tagged by GWAS SNPs (Young 2022), and will be reduced as sequencing replaces SNP genotyping, or by expanded imputation reference panels or by ARG-based methods to model rare variants.

Assuming that the β_j are uncorrelated, the heritability tagged by SNPs within genome region c is $h^2_c = \sum_{j\in c}\beta^2_j$. Although individual β^2_j may be poorly estimated, if there is low LD between SNP pairs within and outside c then h^2_c can be well estimated by $\sum_{j\in c}\hat{\beta}^2_j$ even if c is small, allowing us to investigate how heritability is distributed across genome regions.

Heritability analysis based on summary statistics was launched by Bulik-Sullivan et al. (2015) who introduced LD score regression, a simple linear regression of the form

$$\mathbb{E}[S_j] = 1 + A + n\xi^2_j \qquad \text{where} \qquad \xi^2_j = \sum_{k\sim j} r^2_{jk}\sigma^2_k\,. \tag{4}$$

Here, S_j is a test statistic with a χ^2_1 null distribution, and r_{jk} is a sample correlation coefficient for SNPs j and k so that $r^2_{jj} = 1$. The sum is over SNP j itself and SNPs close to it (typically, within

1 Mb). It follows that ξ_j^2 is the expected heritability tagged by SNP j, including multiple tagging, whereas σ_j^2 is the expected heritability attributed uniquely to SNP j.

The motivating intuition for (4) is that the more SNPs tagged by SNP j, the greater the causal signal contributing to S_j. This intuition appears to conflict with the empirical fact that per-SNP heritability tends to decrease as LD increases; however, as the LD tagged by SNP j increases, ξ_j^2 can increase even though σ_j^2 tends to decrease.

Bulik-Sullivan et al. (2015) considered only the uniform heritability model, the special case of (4) with σ_j^2 constant over j. However, the computational efficiency of the summary statistic approach opened up a flexible class of heritability models of the form (Finucane et al. 2015)

$$\sigma_j^2 = \sum_c \tau_c a_{jc}, \tag{5}$$

where a_{jc} indicates whether or not SNP j is in category c, while τ_c is a parameter measuring the effect on heritability of membership of category c.

Many tens of functional categories can be fitted simultaneously using (5), which was subsequently extended to allow quantitative predictors of heritability related to MAF and LD (Gazal et al. 2017). These authors proposed the BLD model with 75 predictors, which was later improved by the BLD-LDAK model with either 66 or 67 predictors, according to whether α in (2) is estimated or pre-specified (Speed et al. 2020).

Estimates $\hat{\tau}_c$ can be interpreted as the average increase or decrease in β_j^2 for a unit increase in the predictor a_{jc}. An alternative summary of heritability is the enrichment λ_c of heritability in region c which is h_c^2 expressed as a multiple of a baseline such that when $\lambda_c = 1$ the average heritability of a SNP within c is the same as that of a SNP outside c. A key difference between $\hat{\tau}_c$ and $\hat{\lambda}_c$ is that only the former is estimated jointly, adjusting for other predictors in the model. This can be important because many functional categories overlap.

Speed et al. (2022) reported average $\hat{\lambda}_c$ and $\hat{\tau}_c$ over 14 UK Biobank traits, for 24 genome annotations which included a 'conserved' category that included about 2.5% of SNPs that were the most conserved over 19 mammalian species. Both $\hat{\lambda}_c$ and $\hat{\tau}_c$ were largest for this category, indicating that being conserved over mammals was the best predictor of heritability for these traits, even after adjusting for the effects of the other 23 categories. The next most important category based on both $\hat{\lambda}$ and $\hat{\tau}$ was 'coding' with $\hat{\lambda} > 4$ (compared with $\hat{\lambda} > 7$ for 'conserved'). 5'UTR, transcription start site and H3K9ac also showed convincing evidence for both $\tau > 0$ and $\lambda > 1$. Few other categories had $\hat{\tau}$ significantly different from zero, one exception being DNAse hypersensitivity sites (DHS) for which $\hat{\tau}$ was significantly negative even though $\hat{\lambda} \approx 1.7$. Thus, although SNPs within DHS house more heritability for these traits than other SNPs, this must be due to overlapping categories because a DHS SNP has on average lower heritability than a similar non-DHS SNP.

4 | Interpretation of Forensic DNA Profiles

4.1 | A General Framework for Assessing DNA Profile Evidence

In the first decade of forensic DNA profiling, from the late 1980s, there was much controversy over both the reliability of the laboratory techniques and the interpretation of the results (Lander 1989). Over time, laboratory techniques improved and a framework for evidence interpretation emerged that not only helped DNA profiling evidence to become widely used in criminal justice systems but also allowed it to become a benchmark to highlight deficiencies in the presentation in court of other forms of evidence (PCAST 2016).

The interpretation framework is based on a probability updating formula, a version of Bayes theorem that in this context can be called the weight of evidence formula (Steele and Balding 2015). In a simple forensic identification scenario, suppose that the possible sources of a DNA sample include a known person of interest that we will call person 0 (who may be a defendant or a crime victim) and N other individuals. We write H_i for the hypothesis that the true source of the DNA is person i, $i = 0, 1, \ldots, N$, and write $\mathcal{E}_D$ for the DNA evidence, which typically consists of a crime scene profile (CSP), a reference profile from person 0 that at least partly matches the CSP, and at least one 'frequency database' of DNA profiles used to estimate population allele frequencies. We also write π_i for the probability of H_i prior to $\mathcal{E}_D$ being assessed. Then the probability that person 0 is the source of the CSP given $\mathcal{E}_D$ is

$$\mathbb{P}(H_0|\mathcal{E}_D) = \frac{\mathbb{P}(\mathcal{E}_D|H_0)\pi_0}{\sum_{i=0}^N \mathbb{P}(\mathcal{E}_D|H_i)\pi_i} = \left(\sum_{i=0}^N \frac{\mathbb{P}(\mathcal{E}_D|H_i)}{\mathbb{P}(\mathcal{E}_D|H_0)}\frac{\pi_i}{\pi_0}\right)^{-1}. \tag{6}$$

The final form of (6) is convenient because the likelihoods $\mathbb{P}(\mathcal{E}_D|H_i)$ are typically minuscule and hard to interpret, whereas a relative likelihood (or likelihood ratio, LR) can be more interpretable. We discuss computing $\mathbb{P}(\mathcal{E}_D|H_i)$ in Section 4.3. Here we define the LR such that a lower value corresponds to stronger evidence for H_0, often the inverse value is reported instead.

While (6) cannot be simplistically applied to actual cases, for example, because of the judgments required to assess values for the π_i, it is nevertheless an invaluable guide to thinking about evidential strength. For example, to have a high probability for H_0 it is not sufficient for the LR to be small for a 'random man' alternative DNA source, the *sum* in the denominator of (6), over all alternative sources, must be negligible. The widely used 'random man' alternative hypothesis is not only unnecessary it can be misleading.

Simplifying further, let $\mathbb{P}(\mathcal{E}_D|H_0) = 1$ while for $i = 1, 2, \ldots, N$, both $\pi_i = 1/(N+1)$ and $\mathbb{P}(\mathcal{E}_D|H_i) = p$, where p is the population proportion with the same DNA profile as person 0. Then (6) becomes

$$\mathbb{P}(H_0|\mathcal{E}_D) = \frac{1}{1+Np}, \tag{7}$$

providing a simple formula summarising the insight described above: here, to have $\mathbb{P}(H_0|\mathcal{E}_D) \approx 1$ it is not enough for $p \ll 1$, we require $Np \ll 1$.

Further insights come from varying $\mathbb{P}(\mathcal{E}_D|H_i)$ according to the relatedness of 0 with i, considering both direct relatedness and coancestry shared within subpopulations. For example, if the alternative DNA sources include k siblings of person 0, but no other relatives, then

$$\mathbb{P}(H_0|\mathcal{E}_D) = \frac{1}{1 + ks + (N-k)p}, \tag{8}$$

where s is the LR comparing person 0 with a sibling. Typically $s \gg p$ and $k \ll N$, so both terms in the denominator of (8) can be important.

Formula (8) extends in the obvious way for other relatives of person 0, including the important case, discussed further below, of alternative sources sharing coancestry with person 0 that is greater than that shared with members of the frequency database.

4.2 | Suspects Identified From a Database Search

There has been controversy over the effect on evidential strength of identifying person 0 via a search in an 'intelligence database' of the DNA profiles of known individuals, who may, for example, include previous offenders. A failure to grasp the principles of evidence evaluation in a forensic setting has led even distinguished scientists to error, including National Research Council USA (1992, 1996). The erroneous thinking arises because of an analogy with multiple testing or 'hypothesis trawling', which is generally understood by scientists to weaken the significance of any resulting match. This intuition is misleading in the setting of forensic DNA profiles because we know that there exists a true source of the DNA and so the more extensive the search, the higher the probability that the resulting match has identified the correct individual.

Instead, under simplifying assumptions similar to those underlying (7), if the alternative sources include M individuals in the intelligence database, and only person 0 matches the CSP, then

$$\mathbb{P}(H_0|\mathcal{E}_D) = \frac{1}{1 + (N-M)p}, \tag{9}$$

which increases with M. In the extreme case that $N = M$, all alternative sources of the DNA are members of the intelligence database and are found not to match, (9) correctly gives $\mathbb{P}(H_0|\mathcal{E}_D) = 1$. The simple cases described here exclude complexities such as the possibility of sample contamination or labelling error, and the role of the non-DNA evidence, but (6) can also guide reasoning about these and other issues (Steele and Balding 2015; Buckleton et al. 2016).

4.3 | Computing Likelihoods

Computing the likelihoods $\mathbb{P}(\mathcal{E}_D|H_i)$ required in (6) has been an evolving challenge over recent decades as DNA profiling technology has improved, and the sophistication of mathematical models and algorithms has increased (Buckleton et al. 2016). The original DNA fingerprints (Jeffreys et al. 1985) were restriction fragment length polymorphisms, referred to as multi-locus probes with the relevant loci generally being unknown. While it seemed plausible that such fingerprints were unique, statistical interpretation of the evidential value was not feasible.

This changed in the late 1980s with the introduction of single-locus probes, which were superseded a few years later by short tandem repeat (STR) profiles. For these systems, simple statistical models were available, based on the presence/absence of alleles and their database frequencies, which together with population genetics theory justifying assumptions of independence both within loci (Hardy–Weinberg equilibrium, HWE) and across loci (linkage equilibrium, LE) could be used to estimate the population proportion of a profile. The independence assumptions were not strictly valid, due to effects of inbreeding, selection, relatedness or population structure, and the estimates were generally too small for empirical verification. Experimental artefacts such as undetectable (null) alleles which could be confused with homozygotes presented further challenges.

Resolving these problems was held back by a lack of understanding of how a profile population frequency estimate was connected with evidential weight. In fact, (6) reveals that the relevant quantity is the LR which is a ratio of probabilities. When person 0 matches a single-contributor CSP, the ith LR is the probability that persons i and 0 both have a given profile, divided by the probability that person 0 has the profile. Equivalently, it is the conditional probability that i has the profile given that 0 has it (Weir 1996). It is in the conditioning that relatedness and population structure effects arise and can be modelled.

At a single locus, the conditional probability can be formulated as a probability of drawing an AA or AB (homozygote or heterozygote) allele pair given an observation of the same allele pair in an initial draw. While testing for HWE and LE in forensic databases has attracted much attention, the more important impact on weight of evidence comes from coancestry due to relatedness/population structure creating genetic correlations between pairs of individuals. Once these are dealt with using relatedness or coancestry coefficients, any remaining deviations from HWE and LE (within-individual correlations) are usually negligible.

At a single locus, if the coancestry of i and 0 is θ then a simple formula for the LR is (Balding and Nichols 1994)

$$\text{Person 0 homozygous AA:} \quad LR = \frac{(2\theta+(1-\theta)p_A)(3\theta+(1-\theta)p_A)}{(1+\theta)(1+2\theta)}, \tag{10}$$

$$\text{Person 0 heterozygous AB:} \quad LR = 2\frac{(\theta+(1-\theta)p_A)(\theta+(1-\theta)p_B)}{(1+\theta)(1+2\theta)}, \tag{11}$$

where p_A and p_B are population allele fractions, usually replaced by estimates from a frequency database, possibly with an adjustment that biases estimates upwards and avoids zero estimates. When $\theta = 0$, the LRs in (10) and (11) simplify to p_A^2 and $2p_Ap_B$, the population proportions under HWE, and they increase (reducing

$\mathbb{P}(H_0|\mathcal{E}_D)$) as θ increases from zero. The value of θ is assessed relative to the average ancestry shared between person 0 and individuals in the frequency database, so θ and hence the LRs increase (reducing $\mathbb{P}(H_0|\mathcal{E}_D)$) as the database becomes less representative of the ancestry of person 0.

In principle, there can be different θ values for different i, according to their level of coancestry with 0, but in practice it is convenient to use a single relatively large value. Often $\theta = 0.03$ is recommended for STR markers, which represents a relatively large genetic distance between, for example, a database representing a continent-scale population and a subpopulation of that population (Steele et al. 2014), which tends to err in the direction of understating $\mathbb{P}(H_0|\mathcal{E}_D)$. Such understatement typically favours defendants, but when the DNA profile evidence is powerful this causes little detriment to the prosecution case.

LRs (10) and (11) apply to alternative sources i who have no known relatedness with person 0. The LRs for relatives rely on classical relatedness coefficients measuring IBD probabilities generated by the relatedness. These formulas should also allow for coancestry in the probabilities for non-IBD alleles.

Whole-profile LRs are obtained by multiplication over loci, which implies an assumption of independence. Although (10) and (11) cannot be regarded as exact, by a generous adjustment for shared ancestry, which is the principle source of dependence, multiplication across loci is reasonable as other concerns, such as deviation from HWE or sampling error in estimates of the p_A, are negligible relative to the impact of the θ adjustment.

4.4 | Artefacts, Dropout and Mixtures

A major advantage of STR profiles is that they can be used in conjunction with DNA amplification using PCR to analyse small quantities of DNA (called low template DNA, or LTDNA). Further, a single capillary electrophoresis (CE) run can process multiple loci ('multiplexing') using both dye separation due to fluorescence labelling and length separation due to variation in the flanking regions analysed with the STR polymorphism.

These advantages generate additional interpretation challenges, including the 'stutter' peaks introduced by imperfect PCR copying of the tandem repeats, as well as high variability in heights of peaks in the electropherogram (epg) reporting the fluorescence signal as a function of DNA fragment length. The extra variability makes it difficult to visually discriminate single-allele from multiple-allele peaks, and the contributions of different DNA sources in a mixed CSP. With LTDNA, alleles present in the sample can fail to amplify and so 'drop out' from the epg, while minuscule amounts of environmental DNA may be amplified and 'drop in' to the epg. Another epg artefact is pull-up peaks due to errors in reading the fluorescence labels.

Initially, expert judgement was relied on to 'clean up' the epg data prior to statistical analysis, with the expert manually designating stutter and other artefacts, and sometimes manually deconvolving the profiles of one or more contributors. Over time, improvements in statistical analysis led to more automated analysis of the epg. An important step in this advance was the move away from a binary interpretation of the presence/absence of alleles to 'probabilistic genotyping' in which the continuous epg curve is the basis of an analysis that computes probability distributions for the genotypes of different contributors to a DNA sample, along with estimates of their relative DNA contributions. Such 'continuous' interpretation models are able to deal with multiple contributors, LTDNA, and degraded DNA, incorporating models for stutter, peak-height variability and baseline noise (Buckleton et al. 2016).

Sophisticated software for probabilistic DNA profile interpretation is now widely used, including commercial software STRmix (Bright et al. 2016) and TrueAllele (Perlin et al. 2011); for a comparison, see Greenspoon et al. (2024). EuroForMix (Bleka et al. 2016) is a free, open-source alternative. While they incorporate sophisticated models and have by now been well validated, no software can account for all sources of uncertainty or error, in particular complexities such as secondary transfer of DNA and cross-contamination of samples, either accidental or malicious. There is also an unavoidable problem that DNA profile interpretation using (6) requires judgements that lie in the domain of the finder of fact (judge or jury) whereas the software can only be safely used by trained experts.

Bayesian networks are graphical models that represent probabilistic relationships among variables such as the potential source(s) of a DNA sample, which can be updated when new evidence is introduced. Bayesian networks have been used to incorporate DNA profiles with other forms of evidence, and to assess relatedness with a sample contributor (Taroni et al. 2014; Schaapveld et al. 2019). They are most often used in an investigative setting but they have also been used to help assess weight of evidence in court (Fenton et al. 2020).

Recently, DNA sequencing has begun to replace the CE process described above that has been established since the 1990s. Sequencing can generate more informative DNA profiles that include SNPs and other markers in addition to STRs. It also avoids CE artefacts, although it introduces new interpretation challenges and requires new interpretation software. Currently, CE methods are cheaper and remain widely used for routine casework. Sequencing is widely used for mitochondrial DNA (mtDNA), initially only in some regions but now almost all of the mtDNA.

4.5 | Uniparental Profiles: Y and mtDNA

The general framework for evidence interpretation outlined above has had much success when applied to autosomal DNA profiles, which are the most widely used form of DNA evidence. However, application of (6) to DNA profiles based on the Y chromosome or mtDNA is more challenging due to their uniparental inheritance. The lack of recombination for these lineage markers limits their variation and hence their discriminatory power, but they have specific uses for example to identify DNA from a male or a female contributor to a sample that has both male and female sources and to help establish relatedness over multiple generations (King et al. 2014). Uniparental profiles can also be more stable and more likely to be recovered from degraded samples.

Lineage markers are inherited intact from parent to child unless a mutation occurs. It follows that there is a set $\mathcal{R}_0$ of close patrilineal/matrilineal relatives of person 0 sharing their lineage marker profile. The size of $\mathcal{R}_0$, denoted $|\mathcal{R}_0|$, decreases as the whole-profile mutation rate (μ) increases. The mitochondrial genome has $\mu \approx 0.015$ per generation, corresponding to about 1 mutation every 70 generations. It follows that $|\mathcal{R}_0|$ is typically large, $\sim 10^3$ or greater (Andersen and Balding 2018). Older Y-STR profiling kits generated profiles with $0.025 < \mu < 0.1$, but for more modern kits μ can be almost 0.15, or one mutation every seven generations. Then the distribution of $|\mathcal{R}_0|$ has mode 1 (corresponding to a unique profile), and mean and 99th percentile typically ~ 10 (Andersen and Balding 2017).

In assessing the weight of evidence for Y-profile evidence in court, it remains common for forensic scientists to report a measure of evidential strength that compares Male 0 with a male not paternally related to him (Bright et al. 2025). Such an approach is unsatisfactory because all males are paternally related, the relevant quantity is not a binary indicator (i is related/unrelated to 0) but l_i, the number of father–son steps in the ancestral path linking i with 0. To a close approximation

$$\mathbb{P}(\mathcal{E}_D|H_i) = (1-\mu)^{l_i}\ . \tag{12}$$

Among other limitations, (12) does not allow for two or more mutations that cancel each other and leave the profile unchanged, but these are typically very rare.

For most alternative sources of the DNA, l_i is unknown but it will typically be $\sim 10^3$ for pairs of individuals in large human populations, in which case $\mathbb{P}(\mathcal{E}_D|H_i)$ in (12) will be negligible, even cumulatively over many such individuals. However, there will be some individuals for whom l_i is small and hence $\mathbb{P}(\mathcal{E}_D|H_i)$ is large. Some members of $\mathcal{R}_0$ will be known close relatives such as cousins, but there will also be many males for whom l_i is small enough that $\mathbb{P}(\mathcal{E}_D|H_i)$ is non-negligible but large enough that the relatedness will typically not be recognised (few of us recognise as relatives individuals who share a common ancestor five or more generations in the past).

The strong dependence of $\mathbb{P}(\mathcal{E}_D|H_i)$ on the unknown l_i makes (6) difficult to use for Y-profiles. Its value is determined by the number of males with low l_i, for whom the probability of a match is non-negligible. Summing over a probability distribution for l_i is a reasonable approach but would face practical difficulties in court because the choice of distribution is a matter for the finder of fact who is unlikely to be able to perform the summation.

Instead, Andersen and Balding (2017) proposed to replace evaluation of (6) with a simulation-based upper estimate of $|\mathcal{R}_0|$. An estimate of the number of males in the population with the same Y-profile is conceptually straightforward for a finder of fact, who can use their own judgement to decide the plausibility of one of these males, and not Male 0, being the true source of the DNA. A similar approach is available for mtDNA profiles, but their lower mutation rate makes $|\mathcal{R}_0|$ much larger than for Y profiles.

Other approaches to the interpretation of uniparentally inherited profiles in court have been proposed (Andersen and Balding 2021), none of them completely satisfactory in dealing with the problem presented by (12). Until a better alternative approach becomes widely adopted, interpretation in court will continue to be unsatisfactory.

5 | Concluding Remarks

While the three areas reviewed here are distinct and important, the concept of the ARG describing the genealogy underlying a genetic sample is perhaps of central importance because it is relevant to all genetics research. We described how the ARG data structure is beginning to make inroads into GWAS analysis, and conceptually there is also a role for the ARG in the forensic interpretation of DNA profiles. There, a key quantity is the probability that an unprofiled individual has a given profile, and an ARG linking all potential sources of the queried DNA sample would provide the best structure for computing the required probabilities. Implementing this idea would be challenging and it seems likely that simple approximations will continue to prevail, except perhaps for uniparentally inherited profiles for which the ARG simplifies to a coalescent tree. Coalescent-based modelling of mtDNA match probabilities has been attempted (Wilson et al. 2003), using a query leaf node with no sequence data. This approach is not widely used and may be an area for future development.

The advances in ARG-based genetics research over the past decade have been astonishing, particularly in comparison with relatively slow advances in the previous three decades, but there is still a long way to go. The challenge of assessing uncertainty in ARGs inferred genome-wide and for large samples has only begun to be addressed. Another major challenge is to use demographic and selection inferences to improve ARG inference through joint modelling of the ARG and evolutionary processes. When these are realised, we will be able to obtain as full a picture as possible from DNA data about the evolutionary history of human populations.

Author Contributions

Sections 2 and 4, on coalescent modelling and forensic DNA profiling, were mainly written by DB with contributions from DS. The middle section on GWAS was jointly written by both authors.

Acknowledgements

This review was written in part while DJB was a visiting scholar at the National University of Singapore. Many thanks to Mark Beaumont, Yaoban Chan, Zhendong Huang, Leo Speidel and Jerome Kelleher for their helpful comments on drafts.

Open access publishing facilitated by The University of Melbourne, as part of the Wiley - The University of Melbourne agreement via the Council of Australian University Librarians.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The authors have nothing to report.

References

Andersen, M., and D. Balding. 2017. "How Convincing is a Matching Y-Chromosome Profile?" *PLoS Genetics* 13: e1007028.

Andersen, M., and D. Balding. 2018. "How Many Individuals Share A Mitochondrial Genome?" *PLoS Genetics* 14: e1007774.

Andersen, M., and D. Balding. 2021. "Assessing the Forensic Value of DNA Evidence From Y Chromosomes and Mitogenomes." *Genes* 5: 1209.

Argelaguet, R., A. Cuomo, and O. A. Stegle. 2021. "Computational Principles and Challenges in Single-Cell Data Integration." *Nature Biotechnology* 39: 1202–1215.

Astle, W., and D. Balding. 2009. "Population Structure and Cryptic Relatedness in Genetic Association Studies." *Statistical Science* 24: 451–471.

Balding, D. 2006. "A Tutorial on Statistical Methods for Population Association Studies." *Nature Reviews Genetics* 7: 781–791.

Balding, D., and R. Nichols. 1994. "DNA Profile Match Probability Calculation: How to Allow for Population Stratification, Relatedness, Database Selection and Single Bands." *Forensic Science International* 64: 125–140.

Baumdicker, F., G. Bisschop, D. Goldstein, et al. 2022. "Efficient Ancestry and Mutation Simulation With Msprime 1.0." *Genetics* 220: iyab229.

Beaumont, M., W. Zhang, and D. Balding. 2002. "Approximate Bayesian Computation in Population Genetics." *Genetics* 162: 2025–2035.

Bleka, O., G. Storvik, and P. Gill. 2016. "EuroForMix: An Open Source Software Based on a Continuous Model To Evaluate STR DNA Profiles From a Mixture of Contributors With Artefacts." *Forensic Science International: Genetics* 21: 35–44.

Boyle, E., Y. Li, and J. Pritchard. 2017. "An Expanded View of Complex Traits: From Polygenic to Omnigenic." *Cell* 169: 1177–1186.

Brandt, D., C. Huber, C. Chiang, and D. Ortega-Del Vecchyo. 2024. "The Promise of Inferring the Past Using the Ancestral Recombination Graph." *Genome Biology and Evolution* 16: evae005.

Brandt, D., X. Wei, Y. Deng, A. Vaughn, and R. Nielsen. 2022. "Evaluation of Methods for Estimating Coalescence Times Using Ancestral Recombination Graphs." *Genetics* 221: iyac044.

Bright, J., M. Andersen, D. Taylor, H. Kelly, M. Kruijver, and J. Buckleton. 2025. "Relevant Propositions for Y Chromosome Interpretation." *Journal of Forensic Sciences* 70: 271–275.

Bright, J., D. Taylor, C. McGovern, et al. 2016. "Developmental Validation of STRmix Expert Software for the Interpretation of Forensic DNA Profiles." *Forensic Science International: Genetics* 23: 226–239.

Buckleton, J., J. Bright, and D. Taylor, eds. 2016. *Forensic DNA Evidence Interpretation*, 2nd ed., CRC Press.

Bulik-Sullivan, B., P. Loh, H. Finucane, et al. 2015. "LD Score Regression Distinguishes Confounding From Polygenicity in Genome-Wide Association Studies." *Nature Genetics* 47: 291–295.

Bycroft, C., C. Freeman, D. Petkova, et al. 2018. "The UK Biobank Resource With Deep Phenotyping and Genomic Data." *Nature* 562: 203–209.

Cavalli-Sforza, L., and A. Edwards. 1967. "Phylogenetic Analysis: Models and Estimation Procedures." *American Journal of Human Genetics* 19: 233–257.

Chen, H., A. Naseri, and D. Zhi. 2023. "FiMAP: A Fast Identity-By-Descent Mapping Test for Biobank-Scale Cohorts." *PLoS Genetics* 19: e1011057.

Choi, S., and P. O'Reilly. 2019. "PRSice-2: Polygenic Risk Score Software for Biobank-Scale Data." *GigaScience* 8: giz082.

Curtis, D. 2019. "Clinical Relevance of Genome-Wide Polygenic Score May Be Less Than Claimed." *Annals of Human Genetics* 83: 274–277.

Deng, Y., Y. Nielsen, and Y. Song. 2024. "Robust and Accurate Bayesian Inference of Genome-Wide Genealogies for Large Samples." Preprint, bioRxiv, March 16. https://doi.org/10.1101/2024.03.16.585351.

Devlin, B., and K. Roeder. 1999. "Genomic Control for Association Studies." *Biometrics* 55: 997–1004.

Fan, C., J. Cahoon, B. Dinh, et al. 2025. "A Likelihood-Based Framework for Demographic Inference from Genealogical Trees." *Nature Genetics* 57: 865–874.

Fan, C., N. Mancuso, and C. Chiang. 2022. "A Genealogical Estimate of Genetic Relationships." *American Journal of Human Genetics* 109: 812–824.

Fenton, N., M. Neil, B. Yet, and D. Lagnado. 2020. "Analyzing the Simonshaven Case Using Bayesian Networks." *Topics in Cognitive Science* 12: 1092–1114.

Finucane, H., B. Bulik-Sullivan, A. Gusev, et al. 2015. "Partitioning Heritability by Functional Annotation Using Genome-Wide Association Summary Statistics." *Nature Genetics* 47: 1228–1235.

Fisher, R. 1919. "XV.-The Correlation Between Relatives on the Supposition of Mendelian Inheritance." *Transactions of the Royal Society of Edinburgh* 52: 399–433.

Fisher, R. 1925. *Statistical Methods for Research Workers*. Oliver and Boyd.

Galton, F. 1886. "Regression Towards Mediocrity in Hereditary Stature." *Journal of the Anthropological Institute* 15: 246–263.

Gazal, S., H. Finucane, N. Furlotte, et al. 2017. "Linkage Disequilibrium-Dependent Architecture of Human Complex Traits Shows Action of Negative Selection." *Nature Genetics* 49: 1421–1427.

Ge, T., C. Chen, Y. Ni, Y. Feng, and J. Smoller. 2019. "Polygenic Prediction via Bayesian Regression and Continuous Shrinkage Priors." *Nature Communication* 10: 1776.

Greenspoon, S., L. Schiermeier-Wood, and B. Jenkins. 2024. "A Tale of Two PG Systems: A Comparison of the Two Most Widely Used Continuous Probabilistic Genotyping Systems in the United States." *Journal of Forensic Sciences* 69: 1840–1860.

Griffiths, R., and P. Marjoram. 1997. "An Ancestral Recombination Graph." In *Progress in Population Genetics and Human Evolution*, 257–270. Springer.

Griffiths, R., and S. Tavaré. 1994. "Ancestral Inference in Population Genetics." *Statistical Science* 9: 307–319.

Haller, B., J. Galloway, J. Kelleher, P. Messer, and P. Ralph. 2019. "Tree-Sequence Recording in SLiM Opens New Horizons for Forward-Time Simulation of Whole Genomes." *Molecular Ecology Resources* 19: 552–566.

Hein, J., M. Schierup, and C. Wiuf. 2005. *Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory*. Oxford University Press.

Hejase, H., Z. Mo, L. Campagna, and A. Siepel. 2022. "A Deep-Learning Approach for Inference of Selective Sweeps From the Ancestral Recombination Graph." *Molecular Biology and Evolution* 39: msab332.

Hof, J., and D. Speed. 2025. "LDAK-KVIK Performs Fast and Powerful Mixed-Model Association Analysis of Quantitative and Binary Phenotypes." Preprint, medRxiv, July 26. https://doi.org/10.1101/2024.07.25.24311005.

Hoggart, C., J. Whittaker, M. De Iorio, and D. Balding. 2008. "Simultaneous Analysis of all SNPs in Genome-Wide and Re-Sequencing Association Studies." *PLoS Genetics* 4: e10000130.

Holmes, J., D. Speed, and D. Balding. 2019. "Summary Statistic Analyses Can Mistake Confounding Bias for Heritability." *Genetic Epidemiology* 43: 930–940.

Hu, Y., Q. Lu, R. Powles, et al. 2017. "Leveraging Functional Annotations in Genetic Risk Prediction for Human Complex Diseases." *PLoS Computational Biology* 13: e1005589.

Huang, Z., J. Kelleher, Y. Chan, and D. Balding. 2025. "Estimating Evolutionary and Demographic Parameters via ARG-derived IBD." *PLoS Genetics* 21: e1011537.

Hubisz, M., A. Williams, and A. Siepel. 2020. "Mapping Gene Flow Between Ancient Hominins Through Demography-Aware Inference of the Ancestral Recombination Graph." *PLoS Genetics* 16: e1008895.

Hudson, R. 1983. "Properties of a Neutral Allele Model With Intragenic Recombination." *Theoretical Population Biology* 23: 183–201.

International Schizophrenia Consortium. Purcell, S., N. Wray, J. Stone, et al. 2009. "Common Polygenic Variation Contributes to Risk of Schizophrenia and Bipolar Disorder." *Nature* 460: 748–752.

Jayasinghe, D., S. Eshetie, K. Beckmann, B. Benyamin, and S. Hong Lee. 2024. "Advances and Limitations in Polygenic Risk Score Methods for Genomic Prediction: A Scoping Review." *Human Genetics* 143: 1401–1431.

Jeffreys, A., V. Wilson, and S. Thein. 1985. "Individual-Specific "Fingerprints" of Human DNA." *Nature* 316: 76–79.

Kang, H., N. Zaitlen, C. Wade, et al. 2008. "Efficient Control of Population Structure in Model Organism Association Mapping." *Genetics* 178: 1709–1723.

Kelleher, J., A. Etheridge, and G. McVean. 2016. "Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes." *PLoS Computational Biology* 12: e1004842.

Kelleher, J., Y. Wong, A. Wohns, C. Fadil, P. Albers, and G. McVean. 2019. "Inferring Whole-Genome Histories in Large Population Datasets." *Nature Genetics* 51: 1330–1338.

Khera, A., M. Chaffin, K. Aragam, et al. 2018. "Genome-Wide Polygenic Scores for Common Diseases Identify Individuals With Risk Equivalent to Monogenic Mutations." *Nature Genetics* 50: 1219–1224.

Kimura, M. 1969. "The Number of Heterozygous Nucleotide Sites Maintained in a Finite Population due to Steady Flux of Mutations." *Genetics* 61: 893–903.

King, T., G. Fortes, P. Balaresque, et al. 2014. "Identification of the Remains of King Richard III." *Nature Communications* 5: 5631.

Kingman, J. 1982. "The coalescent." *Stochastic Processes and Their Applications* 13: 235–248.

Kuhner, M., J. Yamato, and J. Felsenstein. 1995. "Estimating Effective Population Size and Mutation Rate From Sequence Data Using Metropolis-Hastings Sampling." *Genetics* 140: 1421–1430.

Lander, E. 1989. "DNA Fingerprinting on Trial." *Nature* 339: 501–505.

Leslie, S., B. Winney, G. Hellenthal, et al. 2015. "The Fine-Scale Genetic Structure of the British Population." *Nature* 519: 309–314.

Lewanski, A., M. Grundler, and G. Bradburd. 2024. "The Era of the ARG: An Introduction to Ancestral Recombination Graphs and Their Significance in Empirical Evolutionary Genomics." *PLoS Genetics* 20: e1011110.

Li, H., and R. Durbin. 2011. "Inference of Human Population History From Individual Whole-Genome Sequences." *Nature* 475: 493–496.

Li, N., and M. Stephens. 2003. "Modeling Linkage Disequilibrium and Identifying Recombination Hotspots Using Single-Nucleotide Polymorphism Data." *Genetics* 165: 2213–2233.

Loh, P., G. Tucker, B. Bulik-Sullivan, et al. 2015. "Efficient Bayesian Mixed-Model Analysis Increases Association Power in Large Cohorts." *Nature Genetics* 47: 284–290.

Mahmoudi, A., J. Koskela, J. Kelleher, Y. Chan, and D. Balding. 2022. "Bayesian Inference of Ancestral Recombination Graphs." *PLoS Computational Biology* 18: e1009960.

Marchini, J., B. Howie, S. Myers, G. McVean, and P. Donnelly. 2007. "A New Multipoint Method for Genome-Wide Association Studies by Imputation of Genotypes." *Nature Genetics* 39: 906–913.

Marjoram, P., and J. Wall. 2006. "Fast 'Coalescent' Simulation." *BMC Genetics* 7: 16.

Mazet, O., W. Rodríguez, S. Grusea, S. Boitard, and L. Chikhi. 2016. "On the Importance of Being Structured: Instantaneous Coalescence Rates and Human Evolution-Lessons for Ancestral Population Size Inference?" *Heredity* 116: 362–371.

Mbatchou, J., L. Barnard, J. Backman, et al. 2021. "Computationally Efficient Whole-Genome Regression for Quantitative and Binary Traits." *Nature Genetics* 53: 1097–1103.

McVean, G., and N. Cardin. 2005. "Approximating the Coalescent With Recombination." *Philosophical Transactions of the Royal Society of London B* 360: 1387–1393.

Moorjani, P., and G. Hellenthal. 2023. "Methods for Assessing Population Relationships and History Using Genomic Data." *Annual Review of Genomics and Human Genetics* 24: 305–332.

National Research Council USA. 1992. *DNA Technology in Forensic Science*. National Academies Press.

National Research Council USA. 1996. *The Evaluation of Forensic DNA Evidence*. National Academies Press.

Nearing, J., G. Douglas, M. Hayes, et al. 2022. "Microbiome Differential Abundance Methods Produce Different Results Across 38 Datasets." *Nature Communications* 13: 342.

Nielsen, R., A. Vaughn, and Y. Deng. 2025. "Inference and Applications of Ancestral Recombination Graphs." *Nature Reviews Genetics* 26: 47–58.

Nordborg, M. 2019. "Chapter 25: Coalescent Theory." In *Handbook of Statistical Genomics*, edited by D. Balding, I. Moltke, and J. Marioni, 145–175. Wiley.

Nunes, M., and D. Balding. 2010. "On Optimal Selection of Summary Statistics for Approximate Bayesian Computation." *Statistical Applications in Genetics and Molecular Biology* 9: 34.

PCAST. 2016. *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*. President's Council of Advisors on Science and Technology, USA.

Perlin, M., M. Legler, C. Spencer, et al. 2011. "Validating True Allele DNA Mixture Interpretation." *Journal of Forensic Sciences* 56: 1430–1447.

Price, A., N. Patterson, R. Plenge, M. Weinblatt, N. Shadick, and D. Reich. 2006. "Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies." *Nature Genetics* 38: 904–909.

Ralph, P. 2019. "An Empirical Approach to Demographic Inference With Genomic Data." *Theoretical Population Biology* 127: 91–101.

Ralph, P., K. Thornton, and J. Kelleher. 2020. "Efficiently Summarizing Relationships in Large Samples: A General Duality Between Statistics of Genealogies and Genomes." *Genetics* 215: 779–797.

Rasmussen, M., M. Hubisz, I. Gronau, and A. Siepel. 2014. "Genome-Wide Inference of Ancestral Recombination Graphs." *PLoS Genetics* 10: e1004342.

Salehi Nowbandegani, P., A. Wohns, J. Ballard, et al. 2023. "Extremely Sparse Models of Linkage Disequilibrium in Ancestrally Diverse Association Studies." *Nature Genetics* 55: 1494–1502.

Sarkar, A., and M. Stephens. 2021. "Separating Measurement and Expression Models Clarifies Confusion in Single-Cell RNA Sequencing Analysis." *Nature Genetics* 53: 770–777.

Schaapveld, T., S. Opperman, and S. Harbison. 2019. "Bayesian Networks for the Interpretation of Biological Evidence." *WIREs Forensic Science* 1: e1325.

Schiffels, S., and K. Wang. 2020. MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. In *Statistical Population Genomics*, edited by J. Y. Dutheil, 147–166. Methods in Molecular Biology, vol. 2090. Humana.

Scutari, M., I. Mackay, and D. Balding. 2016. "Using Genetic Distance to Infer the Accuracy of Genomic Prediction." *PLoS Genetics* 12: e1006288.

Silcocks, M., A. Farlow, A. Hermes, et al. 2023. "Indigenous Australian Genomes Show Deep Structure and Rich Novel Variation." *Nature* 624: 593–601.

Sisson, S., Y. Fan, and M. Beaumont eds. 2018. *Handbook of Approximate Bayesian Computation*. Chapman and Hall/CRC.

Speed, D., N. Cai, The UCLEB Consortium. M. Johnson, S. Nejentsev, and D. Balding. 2017. "Reevaluation of SNP Heritability in Complex Human Traits." *Nature Genetics* 49: 986–992.

Speed, D., G. Hemani, M. Johnson, and D. Balding. 2012. "Improved Heritability Estimation From Genome-Wide SNP Data." *American Journal of Human Genetics* 91: 1011–1021.

Speed, D., J. Holmes, and D. Balding. 2020. "Evaluating and Improving Heritability Models Using Summary Statistics." *Nature Genetics* 52: 458–462.

Speed, D., A. Kaphle, and D. Balding. 2022. "SNP-Based Heritability and Selection Analyses: Improved Models and New Results." *BioEssays* 44: e2100170.

Speidel, L., M. Forest, S. Shi, and S. Myers. 2019. "A Method for Genome-Wide Genealogy Estimation for Thousands of Samples." *Nature Genetics* 51: 1321–1329.

Speidel, L., M. Silva, T. Booth, et al. 2025. "High-Resolution Genomic History of Early Medieval Europe." *Nature* 637: 118–126.

Steele, C., and D. Balding. 2015. *Weight of Evidence for Forensic DNA Profiles*, 2nd ed. Wiley.

Steele, C., D. Syndercombe Court, and D. Balding. 2014. "Worldwide F(ST) Estimates Relative to Five Continental-Scale Populations." *Annals of Human Genetics* 78: 468–477.

Stephens, M., and P. Donnelly. 2000. "Inference in Molecular Population Genetics." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 62: 605–635.

Stern, A., P. Wilton, and R. Nielsen. 2019. "An Approximate Full-Likelihood Method for Inferring Selection and Allele Frequency Trajectories From DNA Sequence Data." *PLoS Genetics* 15: e1008384.

Tajima, F. 1983. "Evolutionary Relationship of DNA Sequences in Finite Populations." *Genetics* 105: 437–460.

Taroni, F., A. Biedermann, S. Bozza, P. Garbolino, and C. Aitken. 2014. *Bayesian Networks for Probabilistic Inference and Decision Analysis in Forensic Science*. Wiley.

Tavaré, S., D. Balding, R. Griffiths, and P. Donnelly. 1997. "Inferring Coalescence Times From DNA Sequence Data." *Genetics* 145: 505–518.

Tian, X., R. Cai, and S. Browning. 2022. "Estimating the Genome-Wide Mutation Rate From Thousands of Unrelated Individuals." *American Journal of Human Genetics* 109: 2178–2184.

Vilhjálmsson, B., J. Yang, H. Finucane, et al. 2015. "Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores." *American Journal of Human Genetics* 97: 576–592.

Wakeley, J., L. King, B. Low, and S. Ramachandran. 2012. "Gene Genealogies Within a Fixed Pedigree, and the Robustness of Kingman's Coalescent." *Genetics* 190: 1433–1445.

Wang, Y., J. Guo, G. Ni, J. Yang, P. Visscher, and L. Yengo. 2020. "Theoretical and Empirical Quantification of the Accuracy of Polygenic Scores in Ancestry Divergent Populations." *Nature Communications* 11: 3865.

Wei, W., G. Hemani, and C. Haley. 2014. "Detecting Epistasis in Human Complex Traits." *Nature Reviews Genetics* 15: 722–733.

Weir, B. 1996. *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. Sinauer.

Wellcome Trust Case Control Consortium, T. 2007. "Genome-Wide Association Study of 14,000 Cases of Seven Common Diseases and 3,000 Shared Controls." *Nature* 447: 661–678.

Wilson, I., M. Weale, and D. Balding. 2003. "Inferences From DNA Data: Population Histories, Evolutionary Processes and Forensic Match Probabilities." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 166: 155–188.

Wiuf, C., and J. Hein. 1999. "Recombination as a Point Process Along Sequences." *Theoretical Population Biology* 55: 248–259.

Wohns, A., Y. Wong, B. Jeffery, et al. 2022. "A Unified Genealogy of Modern and Ancient Genomes." *Science* 375: eabi8264.

Wong, Y., A. Ignatieva, J. Koskela, G. Gorjanc, A. Wohns, and J. Kelleher. 2024. "A General and Efficient Representation of Ancestral Recombination Graphs." *Genetics* 228: iyae100.

Yang, J., A. Bakshi, Z. Zhu, et al. 2015. "Genetic Variance Estimation With Imputed Variants Finds Negligible Missing Heritability for Human Height and Body Mass Index." *Nature Genetics* 47: 1114–1120.

Yang, J., B. Benjamin, B. McEvoy, et al. 2010. "Common SNPs Explain a Large Proportion of the Heritability for Human Height." *Nature Genetics* 42: 565–569.

Yang, S., and X. Zhou. 2020. "Accurate and Scalable Construction of Polygenic Scores in Large Biobank Data Sets." *American Journal of Human Genetics* 106: 679–693.

Young, A. 2022. "Discovering Missing Heritability in Whole-Genome Sequencing Data." *Nature Genetics* 54: 224–226.

Yu, J., G. Pressoir, W. Briggs, et al. 2006. "A Unified Mixed-Model Method for Association Mapping That Accounts for Multiple Levels of Relatedness." *Nature Genetics* 38: 203–208.

Zhang, B., A. Biddanda, A. Gunnarsson, F. Cooper, and P. Palamara. 2023. "Biobank-Scale Inference of Ancestral Recombination Graphs Enables Genealogical Analysis of Complex Traits." *Nature Genetics* 55: 768–776.

Zhang, Q., F. Privé, B. Vilhjálmsson, and D. Speed. 2021. "Improved Genetic Prediction of Complex Traits From Individual-Level Data or Summary Statistics." *Nature Communications* 12: 4192.

Zhou, Y., S. Browning, and B. Browning. 2020. "A Fast and Simple Method for Detecting Identity-by-Descent Segments in Large-Scale Data." *American Journal of Human Genetics* 106: 426–437.

Zhu, H., and X. Zhou. 2020. "Statistical Methods for SNP Heritability Estimation and Partition: A Review." *Computational and Structural Biotechnology Journal* 18: 1557–1568.

Zhu, Z., A. Bakshi, A. Vinkhuyzen, et al. 2015. "Dominance Genetic Variation Contributes Little to the Missing Heritability for Human Complex Traits." *American Journal of Human Genetics* 96: 377–385.