REVIEW

Is population structure in the genetic biobank era irrelevant, a challenge, or an opportunity?

Daniel John Lawson[1] · Neil Martin Davies[1] · Simon Haworth[1] · Bilal Ashraf[1] · Laurence Howe[2] · Andrew Crawford[1] · Gibran Hemani[1] · George Davey Smith[1] · Nicholas John Timpson[1]

Abstract

Replicable genetic association signals have consistently been found through genome-wide association studies in recent years. The recent dramatic expansion of study sizes improves power of estimation of effect sizes, genomic prediction, causal inference, and polygenic selection, but it simultaneously increases susceptibility of these methods to bias due to subtle population structure. Standard methods using genetic principal components to correct for structure might not always be appropriate and we use a simulation study to illustrate when correction might be ineffective for avoiding biases. New methods such as trans-ethnic modeling and chromosome painting allow for a richer understanding of the relationship between traits and population structure. We illustrate the arguments using real examples (stroke and educational attainment) and provide a more nuanced understanding of population structure, which is set to be revisited as a critical aspect of future analyses in genetic epidemiology. We also make simple recommendations for how problems can be avoided in the future. Our results have particular importance for the implementation of GWAS meta-analysis, for prediction of traits, and for causal inference.

Introduction

Is population structure relevant in genetic epidemiology?

It could be taken for granted that the problem of population structure (see "Box 1"), in genetic epidemiology, is "solved". Despite early concerns that phenotypes may be stratified by population (Cardon and Palmer 2003; Freedman et al. 2004; Klein et al. 2005; Marchini et al. 2004) replication rates have been high since the arrival of the genome-wide association study (GWAS) (Pe'er et al. 2008) and consequent adoption of stringent genome-wide significance levels. Phenotype stratification is routinely corrected for using principal components analysis (PCA) (Price et al. 2006) and a range of simple methods (Bouaziz et al. 2011) all appear effective at

controlling false positives. State-of-the-art methods use linear mixed models (LMMs, Hoffman 2013; Loh et al. 2015; Zhang et al. 2010) which also control for kinship ("Box 1"). Furthermore, large-scale collaborations in genetic consortia, such as GIANT which examined over 300K individuals in over 100 studies (Locke et al. 2015), enable both replication and the pooling of effect estimates from independent populations. Heterogeneity analyses (Kulminski et al. 2016) are often used to quantify and understand variation.

Indeed, any residual relatedness or familial structure in its broadest sense can now be recruited to help analyses and potentially gain information. The restricted maximum likelihood (REML) method underlying inference in LMMs can be exploited to estimate the "heritability", or proportion of variation in a phenotype explained by genotyped single nucleotide polymorphisms (SNPs) (Yang et al. 2011). These methods exploit population structure using the genetic relatedness matrix—a particular choice for the measurement of kinship based on SNP similarity—to assess if more genetically similar individuals are more phenotypically similar.

Despite the success of GWAS and heritability analysis, we are entering a new biobank era of massive scale single data collection exercises. Examples of these include 500K participants in the UK Biobank (Sudlow et al. 2015), 500K enrolled into the China Kadoorie Biobank (Chen et al. 2011)

✉ Daniel John Lawson
dan.lawson@bristol.ac.uk

1 MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK

2 Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, Gower Street, London WC1E 6BT, UK

and the million veterans program in the US (Gaziano et al. 2016). Importantly, they are of a scale sufficient to both capture signatures of historic demographic events but also be sufficiently influenced by their sampling structure to generate properties in data that can bias association results or their interpretation. This article sets out some reasons to characterize population structure, and specifically:

1. A bias may remain in either direction of the estimated causal effect of a SNP on a trait, after correction for population structure.
2. The effect of correction for structure is a function of the dataset, especially when there isdifferent detection power.

As a consequence:

3. Prediction and heritability analyses require a thoughtful investigation regarding the types of causal pathway that are useful to retain, depending on the intended use of the analysis.
4. Applied analyses such as two-sample Mendelian randomization estimation of the effect of an exposure on an outcome may be biased by population structure when the two samples differ in composition or when they differ in size.

But there are some upsides:

5. Comparison of datasets against a standard reference population structure will resolve many of these issues.
6. Population structure can be very informative about pleiotropy or other biases in causal estimates.

Having argued that population structure is not simply "solved", this article continues with the following structure. We next address the "challenges" being faced in routine analyses. This begins by defining a goal of correction for population structure and show that it has worked in GWAS, but that there are still open problems in the understanding of selection, Mendelian randomization, and prediction. In "opportunities" we describe ways that population structure can be exploited to learn more about the causal link between genetics and biology, as well as describing methodology that might solve the problems. To validate the high-level claims being made, we will consider simple simulations as well as re-examining examples from the literature. Finally, in the discussion we consider what the problems might imply biologically and give some first steps towards solving them.

Challenges in population structure and phenotype stratification

In this section, we will demonstrate that in theory and in practice, most methods that use genetic associations are vulnerable to subtle, but important problems that derive from population structure. A key claim in this paper is that *associations* between genetic loci and traits have been reliably established, but *estimates of effect sizes* are less robust. Many uses of population structure depend crucially on unbiased effect size estimates.

The claim that population structure may have been underexplored is not new. It is now understood that structure may have led to different signatures of selection between UK Biobank and the GIANT consortium in height (Berg et al. 2018; Sohail et al. 2018). The problems may not be specific to the study of selection: Berg et al. note that "population structure corrections in GWAS may not always work exactly as expected" whilst Sohail et al. conclude that "polygenic adaptation signals based on large numbers of SNPs below genome-wide significance are extremely sensitive". Population structure has been recently confirmed as a key part of the problem (Barton et al. 2019; Berg et al. 2019; Sohail et al. 2019), and other authors report residual associations between PCs, geography and traits in the UK Biobank (Haworth et al. 2018; Liu et al. 2018).

Population structure is correlated with phenotypes

To understand why effect estimates may be biased, it is helpful to revisit ideas in population genetics. Populations do differ genetically by genetic drift and/or selection, and as a consequence these populations will also have different genetic phenotypes. For example, ancient populations had different "genetic heights" (Mathieson et al. 2015), with some potentially being taller than any modern population. Height, and other traits, appear to be "omnigenic" (Boyle et al. 2017); that is, there is no region of the genome not in linkage disequilibrium (LD) with SNPs causal for these traits. Since modern populations are a mixture of older populations, SNPs causal for the trait are themselves correlated with ancestry. It follows that the estimate of the effect of a SNP on a trait can be an underestimate when correcting for population structure.

The justification for PCA correction for population structure (Price et al. 2006) is to correct for *non-causal linear associations* between ancestry and phenotype (Fig. 1a). Causality is hard to define because we rarely measure the exact cause, but proxy it; here we are interested in proxies that are
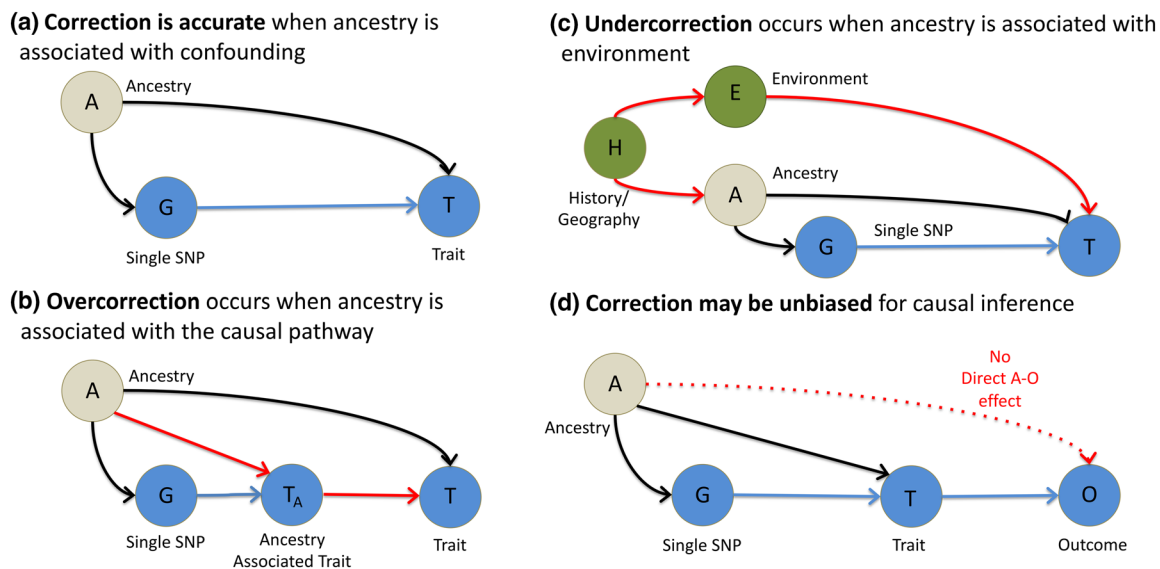
(a) Correction is accurate when ancestry is associated with confounding

(b) Overcorrection occurs when ancestry is associated with the causal pathway

(c) Undercorrection occurs when ancestry is associated with environment

(d) Correction may be unbiased for causal inference

Fig. 1 Causal models including ancestry for the effect of a SNP (*G*) on a trait (*T*). **a** Correction for structure will be accurate when ancestry (*A*) is confounding *T*. **b** Correction for structure may give biased inference when ancestry is associated with the causal pathway (*T$_A$*, which may not be measured) by which the SNP acts. For example, *T* = skin cancer is associated with *T$_A$* = skin tone. **c** Correction for structure will be incomplete when ancestry is associated with the environment (*E*) due to shared history and geography (*H*), for example *T* = BMI with *E* = diet choice. **d** Correction for structure when using causal inference is robust to complexity, provided the assumptions of Mendelian randomization (see text) are met; particularly all remaining effects of ancestry go through the trait (*T*) so there is no direct effect of ancestry (*A*) on the outcome (*O*)

genetic and act through biological pathways. A non-causal association can be generated when population structure is associated with both allele frequency and the phenotype (Fig. 1a). For example, genetic drift simultaneously changes phenotype and SNP frequencies by chance. Weak genetic drift as experienced by larger populations over short time-scales is additive, which corresponds to an additive effect on PCs (McVean 2009). Larger genetic drift, as produced by extreme bottlenecks or consanguinity, is not additive as the SNP frequency distribution becomes skewed and SNPs may become fixed or lost from a population. PCA correction and related methods are less useful when such drifted populations are included (Lawson et al. 2018).

Admixture can change SNP frequencies genome-wide, and small admixture variation is ubiquitous. Even large modern human populations not homogeneous—each individual has a slightly different ancestry proportion from earlier populations. The most ancient detectible human admixture event—Neanderthal introgression into Eurasians—has a mean of around 2% (Sankararaman et al. 2014), but varies substantially between populations and individuals (Wall et al. 2013). Many features of Neanderthal ancestry can be correctly understood using GWAS, which is associated causally with some phenotypes including increasing the risk of depression (Simonti et al. 2016), and non-causally with others, for example skin color (because Neanderthal genes entered the modern human gene pool outside of Africa).

Admixture has the potential to interact with family studies. Siblings have the same expected value of ancestry, with them both receiving a random realized amount. Realized, rather than expected, ancestry is a better predictor of phenotypes (Speed and Balding 2015). Such admixture variation can tag an environmental covariate, for example alcohol consumption influenced by ALDH2 (Price et al. 2002). It could also tag another phenotype that has a confounding relationship, for example, when mixed-race siblings vary in skin tone they may experience different societal pressures (Song 2010), which would be plausibly associated with educational attainment (Light and Strayer 2002) and other phenotypes. A causal analysis would include this pathway—i.e., in the-examples, ALDH2 is causal for alcohol consumption and skin tone for education. However, in the second example the inference does not fit our definition of being biologically caused since it is mediated solely through modifiable societal norms.

Whilst genetic drift can create phenotypic variation between populations, selection does so much more rapidly (Nielsen 2005). If a phenotype is under selection in a particular population, all SNPs that causally affect that phenotype (and also those in linkage disequilibrium) will change in frequency, inducing an association between ancestry and phenotype. Further, where some of the variants affecting a selected phenotype are pleiotropic or in LD with SNPs for another phenotype, selection can generate

genetic associations between the phenotype under selection and other phenotypes. In extreme cases selection can lead to allele frequencies being almost perfectly correlated with population structure. The *LCT* gene (Bersaglieri et al. 2004) which is associated with lactase persistence, and similarly a variant in *ADH1B* (Galinsky et al. 2016) which influences alcohol metabolism, both stratify by population.

Impact on Mendelian randomization

Population structure bias has also been discussed in relation to Mendelian randomization (Davey Smith and Ebrahim 2003; Davey Smith and Hemani 2014; Didelez and Sheehan 2007; Lawlor et al. 2008), an approach which uses a SNP or groups of SNPs as an instrument or "proxy" to test whether an exposure causes an outcome. Mendelian randomization estimates the causal effect under the assumptions (Davies et al. 2018) that; (a) SNPs are associated with the exposure; (b) SNPs do not influence the outcome through a pathway independent of the exposure; (c) that there are no confounders of the SNPs–outcome relationship. Population substructure differences can in theory affect both the strength of genetic instruments and induce confounding, for example in the study of lactase persistence (Campbell et al. 2005; Davey Smith et al. 2009), but there is little evidence the problem is widespread.

The loci that are particularly useful for Mendelian randomization may be particularly susceptible to bias from population structure. This is because strong associations are generated through strong selection, which as discussed above is typically structured. For example, Mendelian randomization studies for alcohol consumption in Europeans typically use the variant in *ADH1B* as a genetic proxy (Holmes et al. 2014; Howe et al. 2019; Lawlor et al. 2013, 2014; Zuccolo et al. 2013). The *ADH1B* variant is associated with ancestry at the country and continental level (Li et al. 2011).

Understanding ancestry correction

For detection in GWAS, a sensible aim is to have the most stringent control of any potential bias, including for phenotype stratification. In addition to PCA correction for stratification, GWAS has also been controlled using genomic control (Devlin et al. 2001) which accounts for confounding by scaling test statistics using an inflation factor to ensure that "null" SNPs (as represented by the median) behave as expected under the null model. However, if all SNPs have a true effect this approach is under-powered. Linkage disequilibrium can be exploited to separate real from confounding signals, implemented in the popular tools LDAK (Speed et al. 2012; Speed and Balding 2019) and LDSC (Bulik-Sullivan et al. 2015). The premise is that if every SNP has an effect then SNPs that are in regions of higher LD will

have larger measured associations because they are composites of their own effects and those around it. These methods confirmed that large-scale GWAS results detect real associations, but what about the size of the effect?

A central goal of genetic association studies is to estimate the "true" causal effect of a SNP (G) on a trait (T). The "true" effect is defined as the effect of G on T when all other traits *that are not in the pathway between G and T* (i.e., confounders) are accounted for (Fig. 1). Correction of GWAS for ancestry (A) is designed to remove *non-causal* associations when observable ancestry (C_A, which might be PCs) not in the causal pathway (Fig. 1a). However, it also removes *causal* associations when ancestry is associated with traits in the pathway (Fig. 1b); a phenomenon often called vertical pleiotropy. Corrected estimates of the G–T associations exclude the G–A–T association. However, they also exclude the G–T_A–T association and hence may underestimate the effect size. For example, if G increases the risk of skin cancer by changing skin tone, its effect size will be underestimated if skin tone is predicted by ancestry. In general, because modern populations are mixtures of ancient populations, many SNPs with a biological effect (including ADH1B and Lactase) may associate with ancestry PCs due to having been common in only one ancestral population.

Ancestry can also associate with the environment (E) and hence also environmental confounders (C_E) (Fig. 1c). There is no causal relationship between A and T via C_E and so *if G is associated with E then correction is desirable* to obtain a less biased estimate of the causal effect of G on T. However, the measured ancestry A is unlikely to account for all association between E and T, so observing an environmental effect indicates the need for additional phenotyping of that environment. For example, if a culture has a diet that reduces BMI then controlling for ancestry only partially corrects for diet. The same problem occurs if observable ancestry (e.g., PCs) do not completely capture the true ancestry.

Genome-wide genetic measures are strongly affected by population structure

A SNP–trait association estimate may be biased after ancestry correction when there is a correlation between the (true) SNP–trait effect and the contribution to an ancestry observable from the SNP (e.g., PC loading). There are many SNPs contributing to ancestry measures, so the bias for each SNP–trait estimate is likely to be small, but genome-wide estimates sum this bias. For example, heritability estimates can in theory be biased by population structure through the prediction of non-genetic covariates (Browning and Browning 2011; Dandine-Roulland et al. 2016), though the scale of the problem is not well quantified for most phenotypes. The robustness of heritability estimates to the existence of internal population structure can at least be tested (Speed et al. 2012, 2014).

Another genome-wide task is to use genetic data to directly predict phenotypes (called genomic prediction, Meuwissen et al. 2001). A predictor is learned using one dataset, then applied to genetic data from others which may be more or less similar in terms of the populations than make it up. This "out of sample" use case makes prediction particularly vulnerable to bias. As demonstrated in Fig. 2, conservative estimates of effect sizes are less useful than a bias–variance tradeoff accounting for the intended populations to be predicted. Adjustment for the PCs is likely to create a higher mean-square error, and it systematically reduces the variance explained in a heritability analysis. The model correcting for ancestry would be preferred for prediction only if (a) it contained enough predictive power to capture real phenotypic differences, and (b) the use case involved generalization into populations for which ancestry may have different effects; for example, predicting skin cancer would

be concerning if the predicted population's skin tone fell outside the range of study population or was caused by different underlying SNPs.

Genetic "prediction … is generally not robust to minor changes in the population" (Goddard et al. 2016). LD in Africans is lower than in Europeans, which makes prediction harder (de los Campos et al. 2010, 2015). A recent study claims that "effect sizes for European ancestry-derived polygenic scores are only 36% as large in African ancestry samples" (Duncan et al. 2018). Yet in consumer genomics (Multhaup and Lehman 2017) and many applications in medicine (Bloss et al. 2011) including drug response (Roden and George 2002), prediction is the primary goal, and ancestry is known to be important (Foll et al. 2014). Prediction is also important for ancient genomics, for example the recent reconstruction of the facial features and dark skin tone of "Cheddar Man" in Neolithic Britain (Brace et al. 2018).
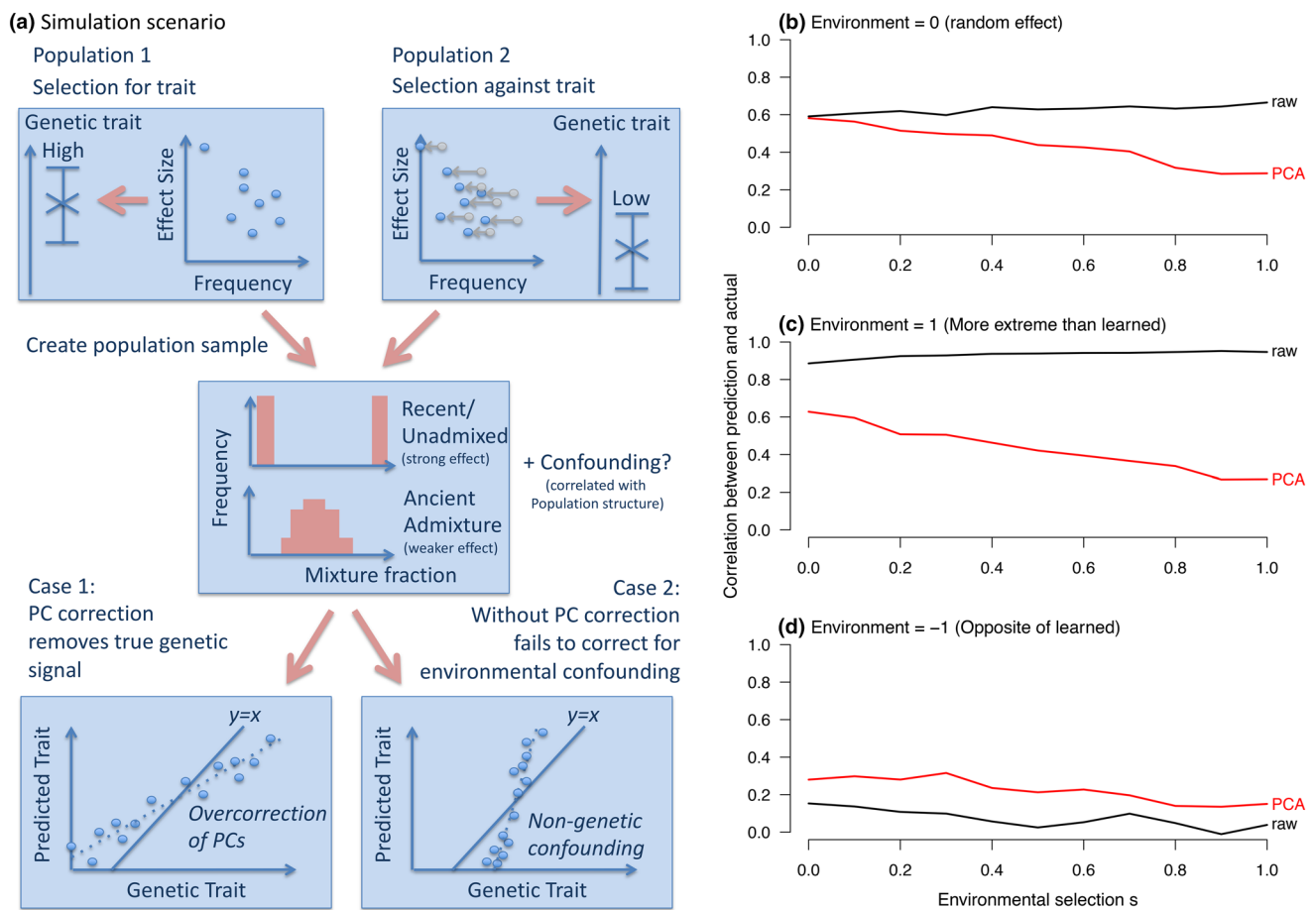

Fig. 2 When should we use PCA correction? **a** In simulation settings (see "Methods") it is straightforward to construct scenarios where correction helps or hinders prediction of traits. Top: two populations are produced with different genetic phenotype, either by drift or selection. Middle: these are mixed to make modern populations. Bottom: in Case 1 the phenotype is associated with true population structure, which can be overcorrected. In Case 2 confounding non-genetic

association is included in the prediction. **b–d** Show results for this simulation. **b** Correcting for confounding using PCA reduces prediction accuracy when traits are genetically associated with population structure. **c** Genetic structure can predict non-genetic confounding leading to apparently good performance on similarly biased populations. **d** PC correction can protect against this confounding at the cost of reduced performance

Prediction protocols typically involve summing the effect of all SNPs that are reliably associated with phenotype to make a "genetic score"—two early examples include coronary heart disease (Ripatti et al. 2010) and gout (Dehghan et al. 2008). However, other prediction models may be necessary when the genetic architecture of the trait does not follow the infinitesimal and additive assumptions (Morgante et al. 2018).

With the availability of much larger datasets, there is increasing discussion about whether polygenic risk prediction should be included in clinical care. For example, Khera et al. (2018) created a polygenic risk score consisting of millions of variants. The top 8% of the population by this score had comparable risk of coronary heart disease to carriers of rare monogenic mutations. While key coronary heart disease loci such as 9P21 have been replicated worldwide (Battram et al. 2018; Dong et al. 2013; Kral et al. 2011; Schunkert et al. 2011), the generalizability of polygenic risk scores of millions of SNPs across different populations requires further study.

Interaction between Mendelian randomization and population structure

Causal inference via Mendelian randomization exploits the effect of G on O that goes via the trait T. If the assumptions are met, Mendelian randomization estimates are robust to bias in G–T estimates, as long as there is no (uncorrected) direct A–O effect. One important mechanism by which G–O (gene–outcome) associations might go via A is linkage disequilibrium (LD). If the instrument SNP G is only a proxy for the true causal SNP, and LD differs between populations, then in theory G can be a strong genetic instrument in one population but weak in another. There is an absence of evidence for this phenomenon, likely due to a lack of large African datasets for whom LD is very different than Europeans.

In recent years, Mendelian randomization studies have increasingly used a two-sample design (Hartwig et al. 2016); in which estimates of the SNP–exposure and SNP–outcome relationships are taken from separate GWAS using non-overlapping samples. Here, an implicit assumption is that the two different samples used to estimate these relationships are drawn from the same underlying population. Typically, the two-sample design will use individuals of similar ancestry (e.g., restricting to individuals of recent European ancestry), however the effect of using two similar but ancestrally distinct samples within a broad definition such as Europeans is currently unclear. We will show below that even if the two samples are from the same population, there may be less power to detect the PCs in a smaller population. This can theoretically result in differential correction and, as a consequence, may bias causal estimates.

Simulating phenotypes with population stratification

For genome-wide questions including heritability analyses and prediction, it is easy to construct scenarios in which either correction or non-correction for structure can be misleading. Figure 2 describes two simulation scenarios: case 1 in which true genetic signal for a trait is associated with population structure (e.g., height), and case 2 in which population structure associates non-causally with the trait through the environment. PC correction is conservative when phenotypes are truly associated with ancestry (Fig. 2b). When ancestry is predictive of the environment (Fig. 2c) it can even increase genetic associations through non-causal pathways. However, when genes have moved into new environments, PC correction reduces bias (Fig. 2d).

Detecting population structure is essential for correcting for it

If not appropriately modeled, phenotype stratification can bias GWAS, heritability estimation, prediction, and Mendelian randomization. However, no single bias-correction approach is necessarily the correct choice for all scenarios. Even if the correct strategy is known, measurement of population structure is critical. As with any parameters estimated from a dataset, increasing sample size increases the ability to detect population structure (Patterson et al. 2006). Within the UK, there was no detectible structure in a subset of around 1000 people from the UK10K project (The UK10K Consortium 2015). However, with over 100,000 people (Galinsky et al. 2016) from the UK Biobank project (Sudlow et al. 2015) several axes of variation are visible in the PCs. Importantly, the latent structure proxied by these axes of variation were still in the data before they could be detected, and so correction on smaller datasets will systematically under-correct for stratification. This may explain why estimates from a single large study are different from a meta-analysis of smaller ones, though to our knowledge this has not been studied.

To the extent that detection of population structure is a problem, better methodology offers a solution. Methods based on "chromosome painting" (Lawson et al. 2012) exploit linkage disequilibrium to better detect population structure. Specifically, the approach counts recent sharing of segments of DNA that are identical by descent, rather than SNP frequencies, to detect recent structure. From the 2039 individuals in the People of the British Isles (PoBI) dataset (Leslie et al. 2015) there was only 1 geographically meaningful PC but over 50 populations detectable with chromosome painting. Studies sampled from a single location such as many cohort studies (for example ALSPAC, Boyd et al. 2013) are typically PCA corrected but the PCs are too

weak to capture the real variation and thus the population is assumed to be "homogeneous". There is a "detection threshold" in ancestry for which we can calculate the sample size required to detect ancestry variation of a given size (McVean 2009). To our knowledge, there has been no systematic study of the importance of residual population structure in small samples.

Exploiting the PoBI dataset, chromosome painting in ALSPAC (Haworth et al. 2018) (Fig. 3) reveals dramatic genetic heterogeneity which is associated with phenotype, here shown for educational attainment. In this case, the bias is predominantly associated with migration: people who

move are more likely to be educated. In ALSPAC, genetic ancestry can predict 8% of the variation in education; for comparison, the most recent published whole-genome genetic score explains 3.2% (Okbay et al. 2016), and a mega-scale analysis is expected to generate a genetic score explaining 10% of the variance (Martin 2018). These results are based on meta-analyses of many studies, in which PC correction may not have sufficiently controlled for population structure.

It is unclear how many of these GWAS hits are in fact hits for ancestry and hidden population structure or migration. The problem exists in many other phenotypes (Martin
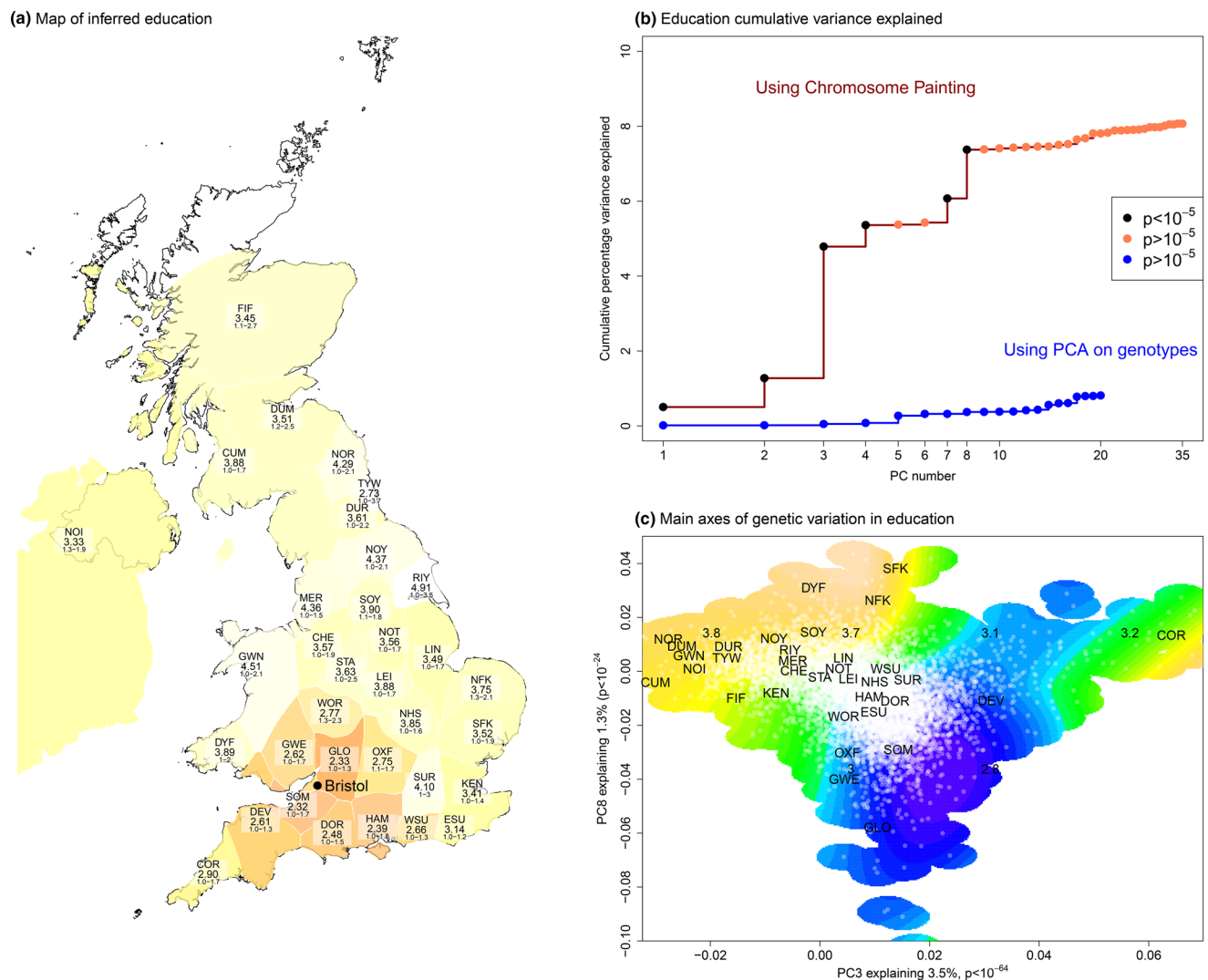

Fig. 3 Population structure can be detected in ALSPAC using the external UK reference dataset PoBI and chromosome painting (see "Methods"). This structure is associated with phenotype, and is not found using regular PCA. **a** Inferred (see "Methods") education level of people migrating from different regions of the UK into the ALSPAC cohort based in Bristol; scale is 1 = no education, 2 = vocational, 3 = GSCEs (age 16), 4 = further education (age 18), 5 = degree (reproduced from Haworth et al. 2018). Participants with ancestry

further from Bristol have considerably higher education, suggesting differential migration by education. **b** Variance explained in education by chromosome painting PCs (8%) and regular PCA (0.8%). **c** The chromosome painting PC locations of individuals and populations for chromosome painting PC 3 and 5, which have the largest associations with education. PoBI mean label locations are shown, along with ALSPAC individuals (white dots) and a kernel smoothing of education

et al. 2017): for example "height is predicted to decrease with genetic distance from Europeans" which is not empirically observed. Interpretation of these results is left to the discussion.

Opportunities from structured populations

Opportunities from natural genetic drift experiments

There are many inferential opportunities offered by population structure, of which two are widely exploited. The first is natural experiments caused by genetic drift. Phenotypic variation decreases on average with distance from Africa (Manica et al. 2007), but any given phenotype may experience extreme variation in a small population. A classic example is the Kosrae islanders in the middle of the Pacific (Lowe et al. 2009) who are at high risk of type 2 diabetes. Similarly, Greek islanders vary dramatically in longevity (Panoutsopoulou et al. 2014) and Ashkenazi Jews (Levy-Lahad et al. 1997) are at high risk of breast cancer; in all three cases, examining drifted populations has led to better understanding of disease for a wider population. In a more extreme example, only 85 individuals were required to identify the gene responsible for blonde hair in Melanesians (Kenny et al. 2012).

Exploiting genetic drift in genetic epidemiology is not limited to extreme founder events such as Finland (Cannon et al. 1998) but is actually routinely (if incidentally) used.

The well-studied European populations experienced the out-of-Africa bottleneck as well as further founder events (Pagani et al. 2016) and are (due to availability) oversampled. Whilst much variation is missing in Europe (1000 Genomes Consortium 2015), the benefit is that some variants are at higher frequency than selection would allow in the larger effective population size within Africa. Figure 4a describes the GWAS results of a meta-analysis of stroke, for individuals with European, African and Asian ancestry sampled worldwide (Pulit et al. 2016). Seven SNPs were significant in the meta-analysis, which we consider in the context of European or African ancestry. Genetic drift has changed the genomic architecture of the disease; two significant SNPs have increased massively in frequency in the (more drifted) Europeans. Under the assumption that this is drift and not European-specific selection, there may not have been power to detect them in the same sized sample of Africans.

Opportunities from replication in varied populations

The second widely exploited opportunity from population structure is replication. Replication—or more generally, joint analyses of independent datasets (McCarthy et al. 2008)—is justifiably required for a GWAS result to be accepted. However, replication should be seen as a function of the properties of the populations being studied; Li and Keating (2014) list 11 examples where trans-ethnic replication has contributed to better understanding of GWAS results. When either
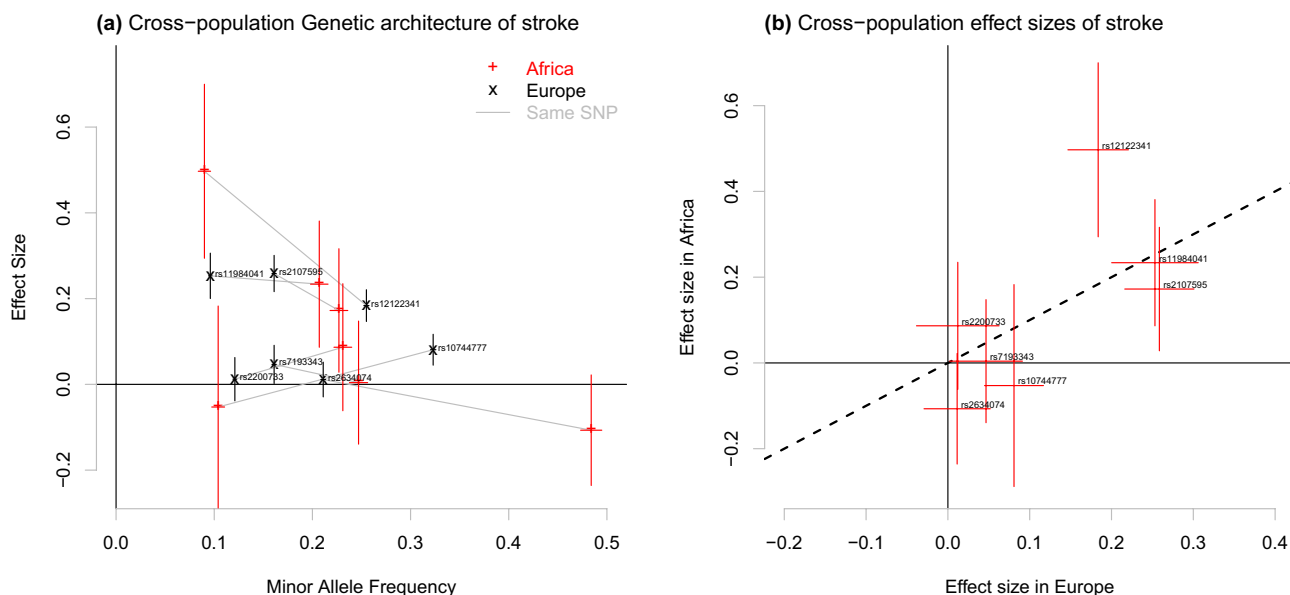

Fig. 4 Genetic architecture of significant stroke SNPs, from the GWAS meta-analysis of data from Pulit et al. (2016). **a** Compares minor allele frequency against inferred effect size for Africans and Europeans (larger sample size). **b** Compares the effect sizes only. Effect SNPs are chosen to ensure that the effect directions in the meta-analysis are positive

the environment (Kaufman et al. 2013; Zhang et al. 2015; Logue et al. 2011) or the genetics (as in European meta-analyses) do not differ substantially then confounding can still lead to an observed association by the same mechanism as that which produced the association in the discovery dataset. Examining these in detail can prove insightful (Zhang et al. 2015). In the stroke example, Fig. 4a shows that 3 of the SNPs replicate in Africans, despite lower power in that population, which increases confidence in their association. Less reassuring are the 2 SNPs with different direction of effect, however, these results are imprecise and these differences could be due to estimation error. The current focus on European datasets appears to primarily probe variation and mutations from after the out-of-Africa bottleneck (Simons et al. 2018), meaning that it misses much biology from the critical period of the evolution of humans, as well as having negative implications for genetic applications in medical science (Oh et al. 2015; Martin et al. 2018).

Population structure can be exploited to reach additional biological insights. The true effect sizes may differ between populations, either due to gene–gene or gene–environment interactions. For example, decanalization (Gibson 2009) is a relatively common form of genetic interaction involving the removal of mechanisms that regulate the genome, allowing larger effects. Correct estimation or overcorrection (Fig. 1a, b) are both possible scenarios when genetic variation is associated with ancestry via a gene–gene interaction. Undercorrection (Fig. 1c) may occur when genetics is associated with the environment via a gene–environment interaction. However, the scenario can be detected and hence properly modeled by observing genotypes in multiple genetic and physical environments; see for example Vrieze et al. (2012) who examine gene–environment interactions in psychiatric disorders.

In the stroke example in Fig. 4b, 1 SNP (rs12122341) has different size of effect in Europeans and Africans. This is interesting and important as it must be associated with a difference in the genetic or environmental background of the two populations. This means the SNP will violate the Mendelian randomization assumptions within one or both populations, but examining multiple populations allows this to be detected. Hypothetically, the difference in effect seen in stroke may contribute via some other phenotype such as smoking behavior or diet.

Opportunities from population-aware methodology

So far, we have considered what current methodology and/or simple data analysis can show. Comparisons are therefore SNP-wise, focused on strongly associated SNPs, and usually result in falsifying hypotheses formed from single population analyses. Completely new information can be extracted by instead comparing sets of SNPs or whole

genetic architectures (Timpson et al. 2018) across population structures and traits (Simons et al. 2018).

Variation between populations can be exploited as part of the statistical methodology, to further learn about the genetic structure of a phenotype. Within a population, admixture mapping was an early tool (Winkler et al. 2010) to exploit variation in ancestry, though there are relatively few recent novel discoveries using this method, one being Adhikari et al. (2016). Across populations, standard GWAS methodology has been successfully applied and extended in "transethnic" approaches (Li and Keating 2014) which start by treating ancestry as a fixed or random effect in regression. More sophisticated approaches such as MANTRA (Morris 2011) and MR-MEGA (Mägi et al. 2017) model heterogeneity in ancestry-specific effects, allowing the agreement between different populations to be measured. Popcorn (Brown et al. 2016) allows this to be done using only GWAS summary statistics. The consistent story across all phenotypes studied in these papers—rheumatoid arthritis, type 2 diabetes, gene expression, and kidney function—is that that both environment and ancestry play an important a role in explaining differences in populations. The total contribution of both is usually on the same order of magnitude. Although not conclusive in human studies, gene–environment interactions have also been explicitly measured and can contribute substantially, e.g., adding 11% to accuracy in a plant study (Desta and Ortiz 2014).

Current methods for trans-ethnic analysis perform the association and heritability stage, but they can also be used for causal inference. The Mendelian randomization framework can be extended to consider the graph of how all traits may be causally related to all other traits (Hemani et al. 2017). We have seen that access to estimates from multiple populations provides insights into the effect sizes of individual SNPs for individual traits. New methodology should be able to exploit differences across populations to automatically screen SNPs and create causal graphs unique to each population.

Discussion

Population structure is relevant for epidemiology

Population structure has always been a feature of genetic studies of phenotypic variation. The impact it has had on inference has varied considerably as the data and questions have changed. Structure confounded the early efforts of genetic discovery, but was then sidestepped by larger datasets and a focus on discovery and replication. Population structure is transitioning from a theoretical problem to a practical issue for questions that require an accurate estimate of effect sizes. This is especially important for prediction

where out-of-sample target populations must be considered and Mendelian randomization for which sample size provides a potential source of bias.

Population structure is still a challenge

The very nature of population structure is challenging, requiring approaches that are specific to the analytical context and trait. Figure 2 demonstrates that the use of stratification correction in larger datasets may overcorrect, whilst Fig. 3 implies that meta-analysis of smaller datasets will systematically under-correct for population structure. The development of new tools to address and exploit structure is an important challenge for epidemiology.

The educational attainment results highlighted by structure within the ALSPAC study reveal important complexity. The effect sizes in trio studies (Okbay et al. 2016) are theoretically not confounded by population structure, and are consistently 30–40% smaller than the inferred effects for the larger, unrelated sample. Our results show that population structure alone can predict educational outcome better than was previously thought. It is still unclear how this predictive power arises—this study implicates migration whilst other explanations include assortative mating and dynastic effects (Kong et al. 2018; Young et al. 2018), as well as sampling biases, though these are not mutually exclusive. We have discussed reasons to adjust GWAS results—or not—using higher quality ancestry estimates for the consortium datasets. There are two opposing hypotheses, which are both consistent with the available data:

(a) Educational attainment is associated with ancestry because of causal pathways that should be included in our definition of the phenotype. For example, historical biased migration could create "brain drain", or selection on ancestral populations leading to a difference in ability (Clark and Cummins 2018). Alternatively, phenotypic differences between populations might exert influences over life-choices.

(b) Educational attainment is associated with ancestry because of non-causal phenotypic pathways. Examples include access to education, cultural norms, the relationship between education and GDP (Nelson and Phelps 1966), and discrimination within the educational system (Light and Strayer 2002; Song 2010).

It is likely that a combination of the above is true. Non-causal pathways are certainly plausible (Fig. 5): average education levels and GDP per capita are correlated within countries and between countries in Europe (Mankiw et al. 1992). GDP is in turn correlated in Northern Europe and the UK with high Germanic and Scandinavian ancestry, such as England, Germany, Denmark, Netherlands, Belgium and Luxembourg.

The difference in selection signal for height between UK Biobank and GIANT (Figure 1 of Berg et al. 2018) cannot be explained by PC correction, as the difference in effect exists even in uncorrected UK Biobank estimates. Stratification may play a role, but it is more complex than a simple correction issue.

We saw in the stroke example that having access to multiple populations is transformative for how SNP effects can be interpreted. This genetic and cultural diversity is helpful in the study of all phenotypes, though we also need new methodology to further exploit the rich information available from these datasets. It may be the case that some phenotypes—including educational attainment and psychiatric disorders—are associated with traits that are actually creating the observed population structure. The arrow of causality then becomes unclear, and we may need dynamical models for historical data to complete the picture.

Epidemiology has much to gain from recognizing population structure

Stratified, admixed and otherwise heterogenous populations are an opportunity to test and validate the statistical models built on inferred genetic contributions to traits.

We also need to revise the data-sharing practices that we use. There are at least three different ways to run GWAS, which might all inform an understanding of how a SNP is associated with a phenotype:

1. *Unadjusted estimates:* These are likely to individually be overestimates of the causal effect and contain false positives.
2. *PCA- or LMM-adjusted estimates:* In many cases, these can individually be underestimates of the causal effect.
3. *Externally adjusted estimates:* Many of the problems discussed would disappear by standardizing the correction: for example, if small studies standardize against genetic variation in the UK Biobank then under-correction will be reduced. These PCs can be included alongside the standard PCs which would still be required to correct for batch effects, residual family structure, etc.

None of these approaches is "correct" for estimating the "true" causal effect sizes for individual SNPs, but having varied estimates from varied populations allows for triangulation (Lawlor et al. 2016) and hence brings us towards a better understanding of the underlying relationship between genotype and phenotype. The third approach may be particularly important for the standardization of varied sizes of study included in meta-analysis.

This article has discussed the part that population structure may play in the future of genetic epidemiology. Observational relationships between genetic variants

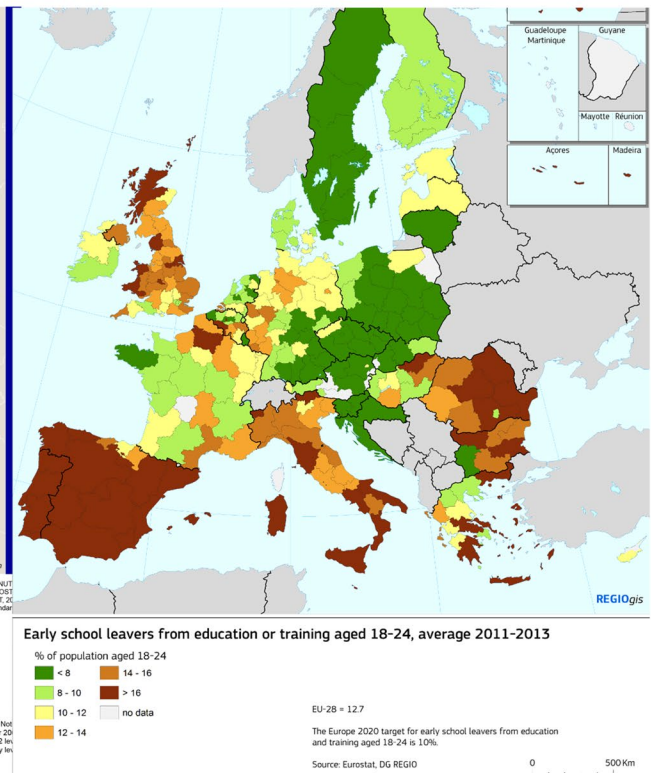**(a) GDP per capita, 2009**
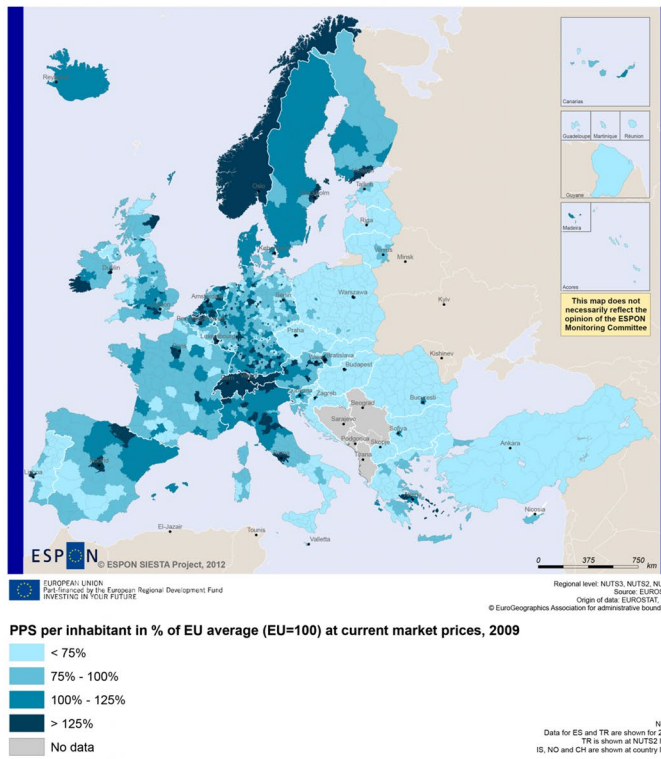
(b) Average education

Fig. 5 Maps of measures of educational attainment correlate with GDP, both within and across countries in Europe. There are large differences between North and South Europe, and this is plausibly associated with genetic ancestry. This may confound inference by generating genetic associations with education that are not biologically causal but are instead driven by access to education. Data source: Eurostat http://ec.europa.eu

and phenotypes are validated through external measurements—structured populations and structured environments. Examining robustness of association signals in different populations is just one form of triangulation, and external validation in non-human models, in vitro experiments and clinical trials all will continue to play a key role.

Box 1: Structure, stratification, ancestry, demography, kinship?

Individuals are not randomly sampled from a large homogeneous population. *Population structure* is the existence of correlated variation in allele frequencies between (sub) populations, meaning that SNPs on different chromosomes are predictive of each other. This can lead to *phenotypic population stratification* or "allele frequency differences < associated with phenotype > due to systematic ancestry differences" (Price et al. 2006). *Ancestry* refers to the proportion of the genome that individuals received from historical abstracted populations, which change over time and are related through their *demography*. Populations and their history are a modeling construct that makes sense of

the family tree (or pedigree) relating all individuals that left descendants in the sampled individuals. This is often measured through a *kinship* matrix or genetic similarity between all individuals in a dataset. Principal component analysis (PC analysis) is a dimension-reducing method to focus on ancestry by measuring the largest-scale variations in kinship, but the full kinship matrix also measures recent relatedness.

Correctly accounting for structure is important. If phenotypic stratification is insufficiently accounted for, then variants associated with population structure become associated with stratified phenotypes: for example, LCT variation is correlated not just with Lactase persistence but also with height (Campbell et al. 2005).

Methods

Model for simulating genotypes in varying environments

This simulation is designed to describe prediction quality in a range of situations where admixture has led to a single,

relatively homogeneous population from two source populations, which are different in SNP frequency for some reason (either selection or drift) that is not explicitly modeled. It treats environment as separate from ancestry, so that a "test population" can be constructed in a different environment to the training population.

We simulated a sample of N individuals at L SNPs. To construct a model including correlations between PCs and traits, we allow individuals to be admixed between two populations ($j = 1, 2$). We then increase the frequency of the SNPs that are associated with the trait within Population 1. Finally, we add an environmental confounder associated with the admixture proportion from Population 1. This leads to a situation in which genes are associated with phenotype via two pathways, a "causal" genetic pathway and a "non-causal" environmental confounded pathway.

Under these conditions, prediction accuracy is reduced by PC correction (because PCs are associated with the genetics of the trait). However, the raw predictor is fitting the environmental component as well as the genetic. The environmental effect is removed by correcting for PCs. This lowers the prediction accuracy.

Specifically, the admixture fraction for individual i from Population 1 is:

$$a_i \sim TruncNorm\big(a_0, \sigma_a^2\big),$$

where $\sigma_a^2 = a_0\big(1 - a_0\big)/L_{\mathrm{eff}}$ represents the variance expected were the admixture fraction to be sampled under a binomial with L_{eff}. This simulates recent (small L_{eff}) or ancient (large L_{eff}) admixture. a_i are truncated to lie within (0,1), allowing admixture proportions of exactly 0 or 1 to be simulated.

We then simulate ancestral the allele frequency for SNP l as $p_1 \sim Uniform(0.05, 0.5)$. Population SNP frequencies are $p_{\mathrm{lm}} = TruncNorm\big(p_1\big(1 + c_i s\big), p_1\big(1 - p_1\big)\sigma_p^2\big)$, where σ_p^2 describes genetic drift from the ancestral frequency, s approximates a "selection" or extreme drift effect for SNPs associated with the trait, and $c_1 = 1$ for SNPs that are causal for the trait and 0 otherwise. Truncation allows frequencies of exactly 0 or 1 to be simulated.

The effect sizes are $\beta_1 \sim Uniform\big(0, c_1 \beta_0 p_1\big(1 - p_1\big)\big)$. The SNP data X_{il} are then sampled $X_{\mathrm{il}} \sim bern\big(a_i p_{11} + \big(1 - a_i\big)p_{12}\big)$. The genetic contribution to the phenotype is $Y_i^{\mathrm{G}} = \sum_{l=1}^{L} X_{il}\beta_l$ and standardize Y_i^{G} to have mean 0 and variance 1. We then generate an environmental exposure $E_i = (a_i - a_0)/\sigma_a$ with expected variance 1.

Finally, we construct a final phenotype $Y_i = h Y_i^{\mathrm{G}} + (1 - h)e E_i + (1 - h)(1 - e) \in_i$, where $\in_i \sim Norm(0, 1)$, h describes the "heritability" of the trait and e describes the "environmental contribution". Y_i is therefore a mixture of components with mean 0 and variance 1 by construction.

To generate the plots, we simulate data under this model, and then examine new phenotypes in new data generated under the same model with estimated $\hat{\beta}_1$ from the function "mixed.solve" in the R package rrBLUP. The top 20 PCs are calculated using the function "irlba" in the package irlba.

Simulations use $h = 0.5$, $e = 0.5$, $N = 2000$, $L = 4000$, $L_{\mathrm{eff}} = 20$, $a_0 = 0.4$, $\beta_0 = 1$ and $s = 0.2$ by default. Sensitivity analysis shows that no conclusions are dependent on the details of these choices (not shown).

Modeling the ALSPAC data

We created a combined dataset of PoBI and ALSPAC mothers, which is described fully in Haworth et al. (2018). Briefly, these were jointly phased, and used the imputed genotypes of ALSPAC participants at a set of 508,223 SNPs chosen by Leslie et al. (2015) for the PoBI dataset. Chromosome painting (Lawson et al. 2012) was used to find the haplotypes that each individual shared with each of the 35 labeled populations from the PoBI dataset (Fig. 2). The PoBI data forms a reference dataset for which a mixture model is fit for ALSPAC mothers. The map in Fig. 2a shows what the genetic score for that region would be, should the observed education in ALSPAC mothers be generated by mixing their regional education values by the inferred admixture weights.

A singular value decomposition (SVD) was then applied to the $N = 2039$ by $K = 35$ matrix of the results of chromosome painting for the PoBI participants (after centering and standardizing variance). Each of the $M = 7739$ mothers for whom genetic and education data were available were mapped into the SVD. Variance explained (Fig. 2b) uses a linear model predicting education either from chromosome painting or raw genotype PCs.

The heatmap (Fig. 2c) is a 2D smoothing of education values using chromosome painting PC3 and PC8, using "predict.gam" from the R package "mgcv" (restricted to where there is an observation within a distance 0.03).

Compliance with ethical standards

References

1000 Genomes Consortium (2015) A global reference for human genetic variation. Nature 526:68–74. https://doi.org/10.1038/nature15393

Adhikari K, Fontanil T, Cal S, Mendoza-Revilla J, Fuentes-Guajardo M, Chacón-Duque J-C, Al-Saadi F, Johansson JA, Quinto-Sanchez M, Acuña-Alonzo V, Jaramillo C, Arias W, Lozano RB, Pérez GM, Gómez-Valdés J, Villamil-Ramírez H, Hunemeier T, Ramallo V, de Cerqueira CCS, Hurtado M, Villegas V, Granja V, Gallo C, Poletti G, Schuler-Faccini L, Salzano FM, Bortolini M-C, Canizales-Quinteros S, Rothhammer F, Bedoya G, Gonzalez-José R, Headon D, López-Otín C, Tobin DJ, Balding D, Ruiz-Linares A (2016) A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features. Nat Commun 7:10815. https://doi.org/10.1038/ncomms10815

Barton N, Hermisson J, Nordborg M (2019) Why structure matters. eLife 8:e45380. https://doi.org/10.7554/eLife.45380

Battram T, Hoskins L, Hughes DA, Kettunen J, Ring SM, Davey Smith G, Timpson NJ (2018) Coronary artery disease, genetic risk and the metabolome in young individuals. Wellcome Open Res 3:114. https://doi.org/10.12688/wellcomeopenres.14788.1

Berg JJ, Harpak A, Sinnott-Armstrong N, Joergensen AM, Mostafavi H, Field Y, Boyle EA, Zhang X, Racimo F, Pritchard JK, Coop G (2018) Reduced signal for polygenic adaptation of height in UK Biobank. bioRxiv. https://doi.org/10.1101/354951

Berg JJ, Harpak A, Sinnott-Armstrong N, Joergensen AM, Mostafavi H, Field Y, Boyle EA, Zhang X, Racimo F, Pritchard JK, Coop G (2019) Reduced signal for polygenic adaptation of height in UK Biobank. eLife 8:e39725. https://doi.org/10.7554/eLife.39725

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN (2004) Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet 74:1111–1120. https://doi.org/10.1086/421051

Bloss CS, Schork NJ, Topol EJ (2011) Effect of direct-to-consumer genomewide profiling to assess disease risk. N Engl J Med 364:524–534. https://doi.org/10.1056/NEJMoa1011893

Bouaziz M, Ambroise C, Guedj M (2011) Accounting for population stratification in practice: a comparison of the main strategies dedicated to genome-wide association studies. PLoS One 6:e28845. https://doi.org/10.1371/journal.pone.0028845

Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, Molloy L, Ness A, Ring S, Davey Smith G (2013) Cohort Profile: the 'Children of the90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. Int J Epidemiol 42:111–127. https://doi.org/10.1093/ije/dys064

Boyle EA, Li YI, Pritchard JK (2017) An expanded view of complex traits: from polygenic to omnigenic. Cell 169:1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Brace S, Diekmann Y, Booth TJ, Faltyskova Z, Rohland N, Mallick S, Ferry M, Michel M, Oppenheimer J, Broomandkhoshbacht N, Stewardson K, Walsh S, Kayser M, Schulting R, Craig OE, Sheridan A, Pearson MP, Stringer C, Reich D, Thomas MG, Barnes I (2018) Population replacement in Early Neolithic Britain. bioRxiv. https://doi.org/10.1101/267443

Brown BC, Ye CJ, Price AL, Zaitlen N (2016) Transethnic genetic-correlation estimates from summary statistics. Am J Hum Genet 99:76–88. https://doi.org/10.1016/j.ajhg.2016.05.001

Browning SR, Browning BL (2011) Population structure can inflate SNP-based heritability estimates. Am J Hum Genet 89:191–193. https://doi.org/10.1016/j.ajhg.2011.05.025

Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson N, Daly MJ, Price AL, Neale BM (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47:291–295. https://doi.org/10.1038/ng.3211

Campbell CD, Ogburn EL, Lunetta KL, Lyon HN, Freedman ML, Groop LC, Altshuler D, Ardlie KG, Hirschhorn JN (2005) Demonstrating stratification in a European American population. Nat Genet 37:868–872. https://doi.org/10.1038/ng1607

Cannon TD, Kaprio J, Lönnqvist J, Huttunen M, Koskenvuo M (1998) The genetic epidemiology of schizophrenia in a finnish twin cohort: a population-based modeling study. Arch Gen Psychiatry 55:67–74. https://doi.org/10.1001/archpsyc.55.1.67

Cardon LR, Palmer LJ (2003) Population stratification and spurious allelic association. Lancet 361:598–604. https://doi.org/10.1016/S0140-6736(03)12520-2

Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F, Li L (2011) China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. Int J Epidemiol 40:1652–1666. https://doi.org/10.1093/ije/dyr120

Clark G, Cummins N (2018) The big sort: selective migration and the decline of Northern England, 1780–2018 (SSRN Scholarly Paper No. ID 3206788). Social Science Research Network, Rochester, NY

Dandine-Roulland C, Bellenguez C, Debette S, Amouyel P, Génin E, Perdry H (2016) Accuracy of heritability estimations in presence of hidden population stratification. Sci Rep 6:26471. https://doi.org/10.1038/srep26471

Davey Smith G, Ebrahim S (2003) "Mendelian randomization": can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 32:1–22. https://doi.org/10.1093/ije/dyg070

Davey Smith G, Hemani G (2014) Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet 23:R89–R98. https://doi.org/10.1093/hmg/ddu328

Davey Smith G, Lawlor DA, Timpson NJ, Baban J, Kiessling M, Day INM, Ebrahim S (2009) Lactase persistence-related genetic variant: population substructure and health outcomes. Eur J Hum Genet 17:357–367. https://doi.org/10.1038/ejhg.2008.156

Davies NM, Holmes MV, Davey Smith G (2018) Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ 362:k601. https://doi.org/10.1136/bmj.k601

de los Campos G, Gianola D, Allison DB (2010) Predicting genetic predisposition in humans: the promise of whole-genome markers. Nat Rev Genet 11:880–886. https://doi.org/10.1038/nrg2898

de los Campos G, Sorensen D, Gianola D (2015) Genomic heritability: what is it? PLoS Genet 11:e1005048. https://doi.org/10.1371/journal.pgen.1005048

Dehghan A, Köttgen A, Yang Q, Hwang S-J, Kao WL, Rivadeneira F, Boerwinkle E, Levy D, Hofman A, Astor BC, Benjamin EJ, van Duijn CM, Witteman JC, Coresh J, Fox CS (2008) Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study. Lancet 372:1953–1961. https://doi.org/10.1016/S0140-6736(08)61343-4

Desta ZA, Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. Trends Plant Sci 19:592–601. https://doi.org/10.1016/j.tplants.2014.05.006

Devlin B, Roeder K, Wasserman L (2001) Genomic control, a new approach to genetic-based association studies. Theor Popul Biol 60:155–166. https://doi.org/10.1006/tpbi.2001.1542

Didelez V, Sheehan N (2007) Mendelian randomization as an instrumental variable approach to causal inference. Stat Methods Med Res 16:309–330. https://doi.org/10.1177/0962280206077743

Dong L, Wang H, Wang DW, Ding H (2013) Association of chromosome 9p21 genetic variants with risk of coronary heart disease in the East Asian population: a meta-analysis. Ann Hum Genet 77:183–190. https://doi.org/10.1111/ahg.12010

Duncan L, Shen H, Gelaye B, Ressler K, Feldman M, Peterson R, Domingue B (2018) Analysis of polygenic score usage and performance across diverse human populations. bioRxiv. https://doi.org/10.1101/398396

Foll M, Poh Y-P, Renzette N, Ferrer-Admetlla A, Bank C, Shim H, Malaspinas A-S, Ewing G, Liu P, Wegmann D, Caffrey DR, Zeldovich KB, Bolon DN, Wang JP, Kowalik TF, Schiffer CA, Finberg RW, Jensen JD (2014) Influenza virus drug resistance: a time-sampled population genetics perspective. PLoS Genet 10:e1004185. https://doi.org/10.1371/journal.pgen.1004185

Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN, Pato MT, Petryshen TL, Kolonel LN, Lander ES, Sklar P, Henderson B, Hirschhorn JN, Altshuler D (2004) Assessing the impact of population stratification on genetic association studies. Nat Genet 36:388–393. https://doi.org/10.1038/ng1333

Galinsky KJ, Bhatia G, Loh P-R, Georgiev S, Mukherjee S, Patterson NJ, Price AL (2016) Fast principal-component analysis reveals convergent evolution of ADH1B in Europe and East Asia. Am J Hum Genet 98:456–472. https://doi.org/10.1016/j.ajhg.2015.12.022

Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, Whitbourne S, Deen J, Shannon C, Humphries D, Guarino P, Aslan M, Anderson D, LaFleur R, Hammond T, Schaa K, Moser J, Huang G, Muralidhar S, Przygodzki R, O'Leary TJ (2016) Million veteran program: a mega-biobank to study genetic influences on health and disease. J Clin Epidemiol 70:214–223. https://doi.org/10.1016/j.jclinepi.2015.09.016

Gibson G (2009) Decanalization and the origin of complex disease. Nat Rev Genet 10:134–140. https://doi.org/10.1038/nrg2502

Goddard ME, Kemper KE, MacLeod IM, Chamberlain AJ, Hayes BJ (2016) Genetics of complex traits. Proc R Soc B 1835:20160569

Hartwig FP, Davies NM, Hemani G, Davey Smith G (2016) Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. Int J Epidemiol 45:1717–1726. https://doi.org/10.1093/ije/dyx028

Haworth S, Mitchell R, Corbin L, Wade KH, Dudding T, Budu-Aggrey A, Carslake D, Hemani G, Paternoster L, Davey Smith G, Davies N, Lawson D, Timpson N (2018) Common genetic variants and health outcomes appear geographically structured in the UK Biobank sample: old concerns returning and their implications. bioRxiv. https://doi.org/10.1101/294876

Hemani G, Bowden J, Haycock PC, Zheng J, Davis O, Flach P, Gaunt TR, Davey Smith G (2017) Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. bioRxiv. https://doi.org/10.1101/173682

Hoffman GE (2013) Correcting for population structure and kinship using the linear mixed model: theory and extensions. PLoS One 8:e75707. https://doi.org/10.1371/journal.pone.0075707

Holmes MV, Dale CE, Zuccolo L, Silverwood RJ, Guo Y, Ye Z, Prieto-Merino D, Dehghan A, Trompet S, Wong A, Cavadino A, Drogan D, Padmanabhan S, Li S, Yesupriya A, Leusink M, Sundstrom J, Hubacek JA, Pikhart H, Swerdlow DI, Panayiotou AG, Borinskaya SA, Finan C, Shah S, Kuchenbaecker KB, Shah T, Engmann J, Folkersen L, Eriksson P, Ricceri F, Melander O, Sacerdote C, Gamble DM, Rayaprolu S, Ross OA, McLachlan S, Vikhireva

O, Sluijs I, Scott RA, Adamkova V, Flicker L, van Bockxmeer FM, Power C, Marques-Vidal P, Meade T, Marmot MG, Ferro JM, Paulos-Pinheiro S, Humphries SE, Talmud PJ, Leach IM, Verweij N, Linneberg A, Skaaby T, Doevendans PA, Cramer MJ, van der Harst P, Klungel OH, Dowling NF, Dominiczak AF, Kumari M, Nicolaides AN, Weikert C, Boeing H, Ebrahim S, Gaunt TR, Price JF, Lannfelt L, Peasey A, Kubinova R, Pajak A, Malyutina S, Voevoda MI, Tamosiunas A, der Zee AHM, Norman PE, Hankey GJ, Bergmann MM, Hofman A, Franco OH, Cooper J, Palmen J, Spiering W, de Jong PA, Kuh D, Hardy R, Uitterlinden AG, Ikram MA, Ford I, Hyppönen E, Almeida OP, Wareham NJ, Khaw K-T, Hamsten A, Husemoen LLN, Tjønneland A, Tolstrup JS, Rimm E, Beulens JWJ, Verschuren WMM, Onland-Moret NC, Hofker MH, Wannamethee SG, Whincup PH, Morris R, Vicente AM, Watkins H, Farrall M, Jukema JW, Meschia J, Cupples LA, Sharp SJ, Fornage M, Kooperberg C, LaCroix AZ, Dai JY, Lanktree MB, Siscovick DS, JorgensonE, Spring B, Coresh J, Li YR, Buxbaum SG, Schreiner PJ, Ellison RC, Tsai MY, Patel SR, Redline S, Johnson AD, Hoogeveen RC, Hakonarson H, Rotter JI, Boerwinkle E, de Bakker PIW, Kivimaki M, Asselbergs FW, Sattar N, Lawlor DA, Whittaker J, Davey Smith G, Mukamal K, Psaty BM, Wilson JG, Lange LA, Hamidovic A, Hingorani AD, Nordestgaard BG, Bobak M, Leon DA, Langenberg C, Palmer TM, Reiner AP, Keating BJ, Dudbridge F, Casas JP (2014) Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. BMJ 349:g4164. https://doi.org/10.1136/bmj.g4164

Howe LJ, Sharp GC, Hemani G, Zuccolo L, Richmond S, Lewis SJ (2019) Prenatal alcohol exposure and facial morphology in a UK cohort. Drug Alcohol Depend 197:42–47. https://doi.org/10.1016/j.drugalcdep.2018.11.031

Kaufman KM, Zhao J, Kelly JA, Hughes T, Adler A, Sanchez E, Ojwang JO, Langefeld CD, Ziegler JT, Williams AH, Comeau ME, Marion MC, Glenn SB, Cantor RM, Grossman JM, Hahn BH, Song YW, Yu C-Y, James JA, Guthridge JM, Brown EE, Alarcón GS, Kimberly RP, Edberg JC, Ramsey-Goldman R, Petri MA, Reveille JD, Vilá LM, Anaya J-M, Boackle SA, Stevens AM, Freedman BI, Criswell LA, Pons-Estel BA, Lee J-H, Lee J-S, Chang D-M, Scofield RH, Gilkeson GS, Merrill JT, Niewold TB, Vyse TJ, Bae S-C, Alarcón-Riquelme ME, Jacob CO, Sivils KM, Gaffney PM, Harley JB, Sawalha AH, Tsao BP (2013) Fine mapping of Xq28: both MECP2 and IRAK1 contribute to risk for systemic lupus erythematosus in multiple ancestral groups. Ann Rheum Dis 72:437–444. https://doi.org/10.1136/annrheumdis-2012-201851

Kenny EE, Timpson NJ, Sikora M, Yee M-C, Estrada AM, Eng C, Huntsman S, Burchard EG, Stoneking M, Bustamante CD, Myles SM (2012) Melanesians blond hair is caused by an amino acid change in TYRP1. Science 336:554. https://doi.org/10.1126/science.1217849

Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, Kathiresan S (2018) Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 50:1219–1224. https://doi.org/10.1038/s41588-018-0183-z

Klein RJ, Zeiss C, Chew EY, Tsai J-Y, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308:385–389. https://doi.org/10.1126/science.1109557

Kong A, Thorleifsson G, Frigge ML, Vilhjalmsson BJ, Young AI, Thorgeirsson TE, Benonisdottir S, Oddsson A, Halldorsson BV, Masson G, Gudbjartsson DF, Helgason A, Bjornsdottir G, Thorsteinsdottir U, Stefansson K (2018) The nature of nurture:

effects of parental genotypes. Science 359:424–428. https://doi.org/10.1126/science.aan6877

Kral BG, Mathias RA, Suktitipat B, Ruczinski I, Vaidya D, Yanek LR, Quyyumi AA, Patel RS, Zafari AM, Vaccarino V, Hauser ER, Kraus WE, Becker LC, Becker DM (2011) A common variant in the *CDKN2B* gene on chromosome 9p21 protects against coronary artery disease in Americans of African ancestry. J Hum Genet 56:224–229. https://doi.org/10.1038/jhg.2010.171

Kulminski AM, Loika Y, Culminskaya I, Arbeev KG, Ukraintseva SV, Stallard E, Yashin AI (2016) Explicating heterogeneity of complex traits has strong potential for improving GWAS efficiency. Sci Rep 6:35390. https://doi.org/10.1038/srep35390

Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G (2008) Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med 27:1133–1163. https://doi.org/10.1002/sim.3034

Lawlor DA, Nordestgaard BG, Benn M, Zuccolo L, Tybjaerg-Hansen A, Davey Smith G (2013) Exploring causal associations between alcohol and coronary heart disease risk factors: findings from a Mendelian randomization study in the Copenhagen General Population Study. Eur Heart J 34:2519–2528. https://doi.org/10.1093/eurheartj/eht081

Lawlor DA, Benn M, Zuccolo L, Silva NMGD, Tybjaerg-Hansen A, Davey Smith G, Nordestgaard BG (2014) ADH1B and ADH1C genotype, alcohol consumption and biomarkers of liver function: findings from a Mendelian Randomization Study in 58,313 European Origin Danes. PLoS One 9:e114294. https://doi.org/10.1371/journal.pone.0114294

Lawlor DA, Tilling K, Davey Smith G (2016) Triangulation in aetiological epidemiology. Int J Epidemiol 45:1866–1886. https://doi.org/10.1093/ije/dyw314

Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of population structure using dense haplotype data. PLoS Genet 8:e1002453. https://doi.org/10.1371/journal.pgen.1002453

Lawson DJ, Van Dorp L, Falush D (2018) A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nat Commun 9:3258

Leslie S, Winney B, Hellenthal G, Davison D, Boumertit A, Day T, Hutnik K, Royrvik EC, Cunliffe B, Wellcome Trust Case Control Consortium 2, International Multiple Sclerosis Genetics Consortium, Lawson DJ, Falush D, Freeman C, Pirinen M, Myers S, Robinson M, Donnelly P, Bodmer W (2015) The fine-scale genetic structure of the British population. Nature 519:309–314. https://doi.org/10.1038/nature14230

Levy-Lahad E, Catane R, Eisenberg S, Kaufman B, Hornreich G, Lishinsky E, Shohat M, Weber BL, Beller U, Lahad A, Halle D (1997) Founder BRCA1 and BRCA2 mutations in Ashkenazi Jews in Israel: frequency and differential penetrance in ovarian cancer and in breast-ovarian cancer families. Am J Hum Genet60:1059–1067

Li YR, Keating BJ (2014) Trans-ethnic genome-wide association studies: advantages and challenges of mapping in diverse populations. Genome Med. https://doi.org/10.1186/s13073-014-0091-5

Li D, Zhao H, Gelernter J (2011) Strong Association of the Alcohol Dehydrogenase 1B Gene (ADH1B) with alcohol dependence and alcohol-induced medical diseases. Biol Psychiatry Common Rare Gene Variants Addict 70:504–512. https://doi.org/10.1016/j.biopsych.2011.02.024

Light A, Strayer W (2002) From Bakke to Hopwood: does race affect college attendance and completion? Rev Econ Stat 84:34–44. https://doi.org/10.1162/003465302317331900

Liu X, Price AL, Loh P-R, O'Connor LJ, Gazal S, Schoech A, Maier RM, Patterson N (2018) Quantification of genetic components of population differentiation in UK Biobank traits reveals signals of polygenic selection. bioRxiv. https://doi.org/10.1101/357483

Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, Croteau-Chonka DC, Esko T, Fall T, Ferreira T, Gustafsson S, Kutalik Z, Luan J, Mägi R, Randall JC, Winkler TW, Wood AR, Workalemahu T, Faul JD, Smith JA, Hua Zhao J, Zhao W, Chen J, Fehrmann R, Hedman ÅK, Karjalainen J, Schmidt EM, Absher D, Amin N, Anderson D, Beekman M, Bolton JL, Bragg-Gresham JL, Buyske S, Demirkan A, Deng G, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Goel A, Gong J, Jackson AU, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Mangino M, Mateo Leach I, Medina-Gomez C, Medland SE, Nalls MA, Palmer CD, Pasko D, Pechlivanis S, Peters MJ, Prokopenko I, Shungin D, Stančáková A, Strawbridge RJ, Ju Sung Y, Tanaka T, Teumer A, Trompet S, van der Laan SW, van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Isaacs A, Albrecht E, Ärnlöv J, Arscott GM, Attwood AP, Bandinelli S, Barrett A, Bas IN, Bellis C, Bennett AJ, Berne C, Blagieva R, Blüher M, Böhringer S, Bonnycastle LL, Böttcher Y, Boyd HA, Bruinenberg M, Caspersen IH, Ida Chen Y-D, Clarke R, Warwick Daw E, de Craen AJM, Delgado G, Dimitriou M, Doney ASF, Eklund N, Estrada K, Eury E, Folkersen L, Fraser RM, Garcia ME, Geller F, Giedraitis V, Gigante B, Go AS, Golay A, Goodall AH, Gordon SD, Gorski M, Grabe H-J, Grallert H, Grammer TB, Gräßler J, Grönberg H, Groves CJ, Gusto G, Haessler J, Hall P, Haller T, Hallmans G, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q, Hengstenberg C, Holmen O, Hottenga J-J, James AL, Jeff JM, Johansson Å, Jolley J, Juliusdottir T, Kinnunen L, Koenig W, Koskenvuo M, Kratzer W, Laitinen J, Lamina C, Leander K, Lee NR, Lichtner P, Lind L, Lindström J, Sin Lo K, Lobbens S, Lorbeer R, LuY, Mach F, Magnusson PKE, Mahajan A, McArdle WL, McLachlan S, Menni C, Merger S, Mihailov E, Milani L, Moayyeri A, Monda KL, Morken MA, Mulas A, Müller G, Müller-Nurasyid M, Musk AW, Nagaraja R, Nöthen MM, Nolte IM, Pilz S, Rayner NW, Renstrom F, Rettig R, Ried JS, Ripke S, Robertson NR, Rose LM, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Scott WR, Seufferlein T, Shi J, Vernon Smith A, Smolonska J, Stanton AV, Steinthorsdottir V, Stirrups K, Stringham HM, Sundström J, Swertz MA, Swift AJ, Syvänen A-C, Tan S-T, Tayo BO, Thorand B, Thorleifsson G, Tyrer JP, Uh H-W, Vandenput L, Verhulst FC, Vermeulen SH, Verweij N, Vonk JM, Waite LL, Warren HR, Waterworth D, Weedon MN, Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF, Zhang Q, The LifeLines Cohort Study, Brennan EP, Choi M, Dastani Z, Drong AW, Eriksson P, Franco-Cereceda A, Gådin JR, Gharavi AG, Goddard ME, Handsaker RE, Huang J, Karpe F, Kathiresan S, Keildson S, Kiryluk K, Kubo M, Lee J-Y, Liang L, Lifton RP, Ma B, McCarroll SA, McKnight AJ, Min JL, Moffatt MF, Montgomery GW, Murabito JM, Nicholson G, Nyholt DR, Okada Y, Perry JRB, Dorajoo R, Reinmaa E, Salem RM, Sandholm N, Scott RA, Stolk L, Takahashi A, Tanaka Toshihiro, van Hooft FM, Vinkhuyzen AAE, Westra H-J, Zheng W, Zondervan KT, The ADIPOGen Consortium, The AGEN-BMI Working Group, The CARDIOGRAMplusC4D Consortium, The CKDGen Consortium, The Glgc, The Icbp, The MAGIC Investigators, The MuTHER Consortium, The MIGen Consortium, The PAGE Consortium, The ReproGen Consortium, The GENIE Consortium, The International Endogene Consortium, Heath AC, Arveiler D, Bakker SJL, Beilby J, Bergman RN, Blangero J, Bovet P, Campbell H, Caulfield MJ, Cesana G, Chakravarti A, Chasman DI, Chines PS, Collins FS, Crawford DC, Adrienne Cupples L, Cusi D, Danesh J, de Faire U, den Ruijter HM, Dominiczak AF, Erbel R, Erdmann J, Eriksson JG, Farrall M, Felix SB, Ferrannini E, Ferrières J, Ford I, Forouhi NG, Forrester T, Franco OH, Gansevoort RT, Gejman PV, Gieger C, Gottesman O, Gudnason V, Gyllensten U, Hall AS, Harris TB, Hattersley AT, Hicks AA, Hindorff LA, Hingorani AD, Hofman A, Homuth

G, Kees Hovingh G, Humphries SE, Hunt SC, Hyppönen E, Illig T, Jacobs KB, Jarvelin M-R, Jöckel K-H, Johansen B, Jousilahti P, Wouter Jukema J, Jula AM, Kaprio J, Kastelein JJP, Keinanen-Kiukaanniemi SM, Kiemeney LA, Knekt P, Kooner JS, Kooperberg C, Kovacs P, Kraja AT, Kumari M, Kuusisto J, Lakka TA, Langenberg C, Le Marchand L, Lehtimäki T, Lyssenko V, Männistö S, Marette A, Matise TC, McKenzie CA, McKnight B, Moll FL, Morris AD, Morris AP, Murray JC, Nelis M, Ohlsson C, Oldehinkel AJ, Ong KK, Madden PAF, Pasterkamp G, Peden JF, Peters A, Postma DS, Pramstaller PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ridker PM, Rioux JD, Ritchie MD, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schunkert H, Schwarz PEH, Sever P, Shuldiner AR, Sinisalo J, Stolk RP, Strauch K, Tönjes A, Trégouët D-A, Tremblay A, Tremoli E, Virtamo J, Vohl M-C, Völker U, Waeber G, Willemsen G, Witteman JC, Carola Zillikens M, Adair LS, Amouyel P, Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle E, Bornstein SR, Bottinger EP, Bouchard C, Cauchi S, Chambers JC, Chanock SJ, Cooper RS, de Bakker PIW, Dedoussis G, Ferrucci L, Franks PW, Froguel P, Groop LC, Haiman CA, Hamsten A, Hui J, Hunter DJ, Hveem K, Kaplan RC, Kivimaki M, Kuh D, Laakso M, Liu Y, Martin NG, März W, Melbye M, Metspalu A, Moebus S, Munroe PB, Njølstad I, Oostra BA, Palmer CNA, Pedersen NL, Perola M, Pérusse L, Peters U, Power C, Quertermous T, Rauramaa R, Rivadeneira F, Saaristo TE, Saleheen D, Sattar N, Schadt EE, Schlessinger D, Eline Slagboom P, Snieder H, Spector TD, Thorsteinsdottir U, Stumvoll M, Tuomilehto J, Uitterlinden AG, Uusitupa M, van der Harst P, Walker M, Wallaschofski H, Wareham NJ, Watkins H, Weir DR, Wichmann H-E, Wilson JF, Zanen P, Borecki IB, Deloukas P, Fox CS, Heid IM, O'Connell JR, Strachan DP, Stefansson K, van Duijn CM, Abecasis GR, Franke L, Frayling TM, McCarthy MI, Visscher PM, Scherag A, Willer CJ, Boehnke M, Mohlke KL, Lindgren CM, Beckmann JS, Barroso I, North KE, Ingelsson E, Hirschhorn JN, Loos RJF, Speliotes EK (2015) Genetic studies of body mass index yield new insights for obesity biology. Nature 518:197–206. https://doi.org/10.1038/nature14177

Logue MW, Schu M, Vardarajan BN, Buros J, Green RC, Go RCP, Griffith P, Obisesan TO, Shatz R, Borenstein A, Cupples LA, Lunetta KL, Fallin MD, Baldwin CT, Farrer LA, Group, for the M.-I.R. on A.G.E. (MIRAGE) S (2011) A comprehensive genetic association study of alzheimer disease in African Americans. Arch Neurol 68:1569–1579. https://doi.org/10.1001/archneurol.2011.646

Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, Chasman DI, Ridker PM, Neale BM, Berger B, Patterson N, Price AL (2015) Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet 47:284–290. https://doi.org/10.1038/ng.3190

Lowe JK, Maller JB, Pe'er I, Neale BM, Salit J, Kenny EE, Shea JL, Burkhardt R, Smith JG, Ji W, Noel M, Foo JN, Blundell ML, Skilling V, Garcia L, Sullivan ML, Lee HE, Labek A, Ferdowsian H, Auerbach SB, Lifton RP, Newton-Cheh C, Breslow JL, Stoffel M, Daly MJ, Altshuler DM, Friedman JM (2009) Genome-Wide Association studies in an isolated founder population from the Pacific Island of Kosrae. PLoS Genet 5:e1000365. https://doi.org/10.1371/journal.pgen.1000365

Mägi R, Horikoshi M, Sofer T, Mahajan A, Kitajima H, Franceschini N, McCarthy MI, COGENT-Kidney Consortium, T2D-GENES Consortium, Morris AP (2017) Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. Hum Mol Genet 26:3639–3650. https://doi.org/10.1093/hmg/ddx280

Manica A, Amos W, Balloux F, Hanihara T (2007) The effect of ancient population bottlenecks on human phenotypic variation. Nature 448:346–348. https://doi.org/10.1038/nature05951

Mankiw NG, Romer D, Weil DN (1992) A contribution to the empirics of economic growth. Q J Econ 107:407–437. https://doi.org/10.2307/2118477

Marchini J, Cardon LR, Phillips MS, Donnelly P (2004) The effects of human population structure on large genetic association studies. Nat Genet 36:512–517. https://doi.org/10.1038/ng1337

Martin N (2018) Getting to the genetic and environmental roots of educational inequality. Npj Sci Learn 3:4. https://doi.org/10.1038/s41539-018-0021-1

Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, Daly MJ, Bustamante CD, Kenny EE (2017) Human demographic history impacts genetic risk prediction across diverse populations. Am J Hum Genet 100:635–649. https://doi.org/10.1016/j.ajhg.2017.03.004

Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ (2018) Hidden "risk" in polygenic scores: clinical use today could exacerbate health disparities. bioRxiv. https://doi.org/10.1101/441261

Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M, Sirak K, Gamba C, Jones ER, Llamas B, Dryomov S, Pickrell J, Arsuaga JL, de Castro JMB, Carbonell E, Gerritsen F, Khokhlov A, Kuznetsov P, Lozano M, Meller H, Mochalov O, Moiseyev V, Guerra MAR, Roodenberg J, Vergès JM, Krause J, Cooper A, Alt KW, Brown D, Anthony D, Lalueza-Fox C, Haak W, Pinhasi R, Reich D (2015) Genome-wide patterns of selection in 230 ancient Eurasians. Nature 528:499–503. https://doi.org/10.1038/nature16152

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, Hirschhorn JN (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet 9:356–369. https://doi.org/10.1038/nrg2344

McVean G (2009) A genealogical interpretation of principal components analysis. PLoS Genet 5:e1000686. https://doi.org/10.1371/journal.pgen.1000686

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4):1819–1829

Morgante F, Huang W, Maltecca C, Mackay TFC (2018) Effect of genetic architecture on the prediction accuracy of quantitative traits in samples of unrelated individuals. Heredity 120:500–514. https://doi.org/10.1038/s41437-017-0043-0

Morris AP (2011) Transethnic meta-analysis of Genomewide Association studies. Genet Epidemiol 35:809–822. https://doi.org/10.1002/gepi.20630

Multhaup M, Lehman A (2017) The science behind 23andMe's Genetic Weight report; White Paper 23–17. https://permalinks.23andme.com/pdf/23_17-GeneticWeight_Feb2017.pdf. Accessed 15 Aug 2018

Nelson RR, Phelps ES (1966) Investment in humans, technological diffusion, and economic growth. Am Econ Rev 56:69–75

Nielsen R (2005) Molecular signatures of natural selection. Annu Rev Genet 39:197–218. https://doi.org/10.1146/annurev.genet.39.073003.112420

Oh SS, Galanter J, Thakur N, Pino-Yanes M, Barcelo NE, White MJ, de Bruin DM, Greenblatt RM, Bibbins-Domingo K, Wu AHB, Borrell LN, Gunter C, Powe NR, Burchard EG (2015) Diversity in clinical and biomedical research: a promise yet to be fulfilled. PLOS Med 12:e1001918. https://doi.org/10.1371/journal.pmed.1001918

Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, Turley P, Chen G-B, Emilsson V, Meddens SFW, Oskarsson S, Pickrell JK, Thom K, Timshel P, de Vlaming R, Abdellaoui A,

Ahluwalia TS, Bacelis J, Baumbach C, Bjornsdottir G, Brandsma JH, Pina Concas M, Derringer J, Furlotte NA, Galesloot TE, Girotto G, Gupta R, Hall LM, Harris SE, Hofer E, Horikoshi M, Huffman JE, Kaasik K, Kalafati IP, Karlsson R, Kong A, Lahti J, van der Lee SJ, de Leeuw C, Lind PA, Lindgren K-O, Liu T, Mangino M, Marten J, Mihailov E, Miller MB, van der Most PJ, Oldmeadow C, Payton A, Pervjakova N, Peyrot WJ, Qian Y, Raitakari O, Rueedi R, Salvi E, Schmidt B, Schraut KE, Shi J, Smith AV, Poot RA, St Pourcain B, Teumer A, Thorleifsson G, Verweij N, Vuckovic D, Wellmann J, Westra H-J, Yang J, Zhao W, Zhu Z, Alizadeh BZ, Amin N, Bakshi A, Baumeister SE, Biino G, Bønnelykke K, Boyle PA, Campbell H, Cappuccio FP, Davies G, De Neve J-E, Deloukas P, Demuth I, Ding J, Eibich P, Eisele L, Eklund N, Evans DM, Faul JD, Feitosa MF, Forstner AJ, Gandin I, Gunnarsson B, Halldórsson BV, Harris TB, Heath AC, Hocking LJ, Holliday EG, Homuth G, Horan MA, Hottenga J-J, de Jager PL, Joshi PK, Jugessur A, Kaakinen MA, Kähönen M, Kanoni S, Keltigangas-Järvinen L, Kiemeney LALM, Kolcic I, Koskinen S, Kraja AT, Kroh M, Kutalik Z, Latvala A, Launer LJ, Lebreton MP, Levinson DF, Lichtenstein P, Lichtner P, Liewald DCM, Cohort Study L, Loukola A, Madden PA, Mägi R, Mäki-Opas T, Marioni RE, Marques-Vidal P, Meddens GA, McMahon G, Meisinger C, Meitinger T, Milaneschi Y, Milani L, Montgomery GW, Myhre R, Nelson CP, Nyholt DR, Ollier WER, Palotie A, Paternoster L, Pedersen NL, Petrovic KE, Porteous DJ, Räikkönen K, Ring SM, Robino A, Rostapshova O, Rudan I, Rustichini A, Salomaa V, Sanders AR, Sarin A-P, Schmidt H, Scott RJ, Smith BH, Smith JA, Staessen JA, Steinhagen-Thiessen E, Strauch K, Terracciano A, Tobin MD, Ulivi S, Vaccargiu S, Quaye L, van Rooij FJA, Venturini C, Vinkhuyzen AAE, Völker U, Völzke H, VonkJM, Vozzi D, Waage J, Ware EB, Willemsen G, Attia JR, Bennett DA, Berger K, Bertram L, Bisgaard H, Boomsma DI, Borecki IB, Bültmann U, Chabris CF, Cucca F, Cusi D, Deary IJ, Dedoussis GV, van Duijn CM, Eriksson JG, Franke B, Franke L, Gasparini P, Gejman PV, Gieger C, Grabe H-J, Gratten J, Groenen PJF, Gudnason V, van der Harst P, Hayward C, Hinds DA, Hoffmann W, Hyppönen E, Iacono WG, Jacobsson B, Järvelin M-R, Jöckel K-H, Kaprio J, Kardia SLR, Lehtimäki T, Lehrer SF, Magnusson PKE, Martin NG, McGue M, Metspalu A, Pendleton N, Penninx BWJH, Perola M, Pirastu N, Pirastu M, Polasek O, Posthuma D, Power C, Province MA, Samani NJ, Schlessinger D, Schmidt R, Sørensen TIA, Spector TD, Stefansson K, Thorsteinsdottir U, Thurik AR, Timpson NJ, Tiemeier H, Tung JY, Uitterlinden AG, Vitart V, Vollenweider P, Weir DR, Wilson JF, Wright AF, Conley DC, Krueger RF, Davey Smith G, Hofman A, Laibson DI, Medland SE, Meyer MN, Yang J, Johannesson M, Visscher PM, Esko T, Koellinger PD, Cesarini D, Benjamin DJ (2016) Genome-wide association study identifies 74 loci associated with educational attainment. Nature 533:539–542. https://doi.org/10.1038/nature17671

Pagani L, Lawson DJ, Jagoda E, Mörseburg A, Eriksson A, Mitt M, Clemente F, Hudjashov G, DeGiorgio M, Saag Lauri, Wall JD, Cardona A, Mägi R, Sayres MAW, Kaewert S, Inchley C, Scheib CL, Järve M, Karmin M, Jacobs GS, Antao T,Iliescu FM, Kushniarevich A, Ayub Q, Tyler-Smith C, Xue Y, Yunusbayev B, Tambets K, Mallick CB, Saag Lehti, Pocheshkhova E, Andriadze G, Muller C, Westaway MC, Lambert DM, Zoraqi G, Turdikulova S, Dalimova D, Sabitov Z, Sultana GNN, Lachance J, Tishkoff S, Momynaliev K, Isakova J, Damba LD, Gubina M, Nymadawa P, Evseeva I, Atramentova L, Utevska O, Ricaut F-X, Brucato N, Sudoyo H, Letellier T, Cox MP, Barashkov NA, Škaro V, Mulahasanovic L, Primorac D, Sahakyan H, Mormina M, Eichstaedt CA, Lichman DV, Abdullah S, Chaubey G, Wee JTS, Mihailov E, Karunas A, Litvinov S, Khusainova R, Ekomasova N, Akhmetova V, Khidiyatova I, Marjanović D, Yepiskoposyan L, Behar DM, Balanovska E, Metspalu A,

Derenko M, Malyarchuk B, Voevoda M, Fedorova SA, Osipova LP, Lahr MM, Gerbault P, Leavesley M, Migliano AB, Petraglia M, Balanovsky O, Khusnutdinova EK, Metspalu E, Thomas MG, Manica A, Nielsen R, Villems R, Willerslev E, Kivisild T, Metspalu M (2016) Genomic analyses inform on migration events during the peopling of Eurasia. Nat Adv Online Publ. https://doi.org/10.1038/nature19792

Panoutsopoulou K, Hatzikotoulas K, Xifara DK, Colonna V, Farmaki A-E, Ritchie GRS, Southam L, Gilly A, Tachmazidou I, Fatumo S, Matchan A, Rayner NW, Ntalla I, Mezzavilla M, Chen Y, Kiagiadaki C, Zengini E, Mamakou V, Athanasiadis A, Giannakopoulou M, Kariakli V-E, Nsubuga RN, Karabarinde A, Sandhu M, McVean G, Tyler-Smith C, Tsafantakis E, Karaleftheri M, Xue Y, Dedoussis G, Zeggini E (2014) Genetic characterization of Greek population isolates reveals strong genetic drift at missense and trait-associated variants. Nat. Commun. 5:5345. https://doi.org/10.1038/ncomms6345

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2:e190. https://doi.org/10.1371/journal.pgen.0020190

Pe'er I, Yelensky R, Altshuler D, Daly MJ (2008) Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol 32:381–385. https://doi.org/10.1002/gepi.20303

Price RK, Risk NK, Wong MM, Klingle RS (2002) Substance use and abuse by Asian Americans and Pacific Islanders: preliminary results from four national epidemiologic studies. Public Health Rep 117:S39–S50

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909. https://doi.org/10.1038/ng1847

Pulit SL, McArdle PF, Wong Q, Malik R, Gwinn K, Achterberg S, Algra A, Amouyel P, Anderson CD, Arnett DK, Arsava EM, Attia J, Ay H, Bartz TM, Battey T, Benavente OR, Bevan S, Biffi A, Bis JC, Blanton SH, Boncoraglio GB, Brown RD, Burgess AI, Carrera C, Chapman Smith SN, Chasman DI, Chauhan G, Chen W-M, Cheng Y-C, Chong M, Cloonan LK, Cole JW, Cotlarciuc I, Cruchaga C, Cuadrado-Godia E, Dave T, Dawson J, Debette S, Delavaran H, Dell CA, Dichgans M, Doheny KF, Dong C, Duggan DJ, Engström G, Evans MK, Pallejà XE, Faul JD, Fernández-Cadenas I, Fornage M, Frossard PM, Furie K, Gamble DM, Gieger C, Giese A-K, Giralt-Steinhauer E, González HM, Goris A, Gretarsdottir S, Grewal RP, Grittner U, Gustafsson S, Han B, Hankey GJ, Heitsch L, Higgins P, Hochberg MC, Holliday E, Hopewell JC, Horenstein RB, Howard G, Ikram MA, Ilinca A, Ingelsson E, Irvin MR, Jackson RD, Jern C, Conde JJ, Johnson JA, Jood K, Kahn MS, Kaplan R, Kappelle LJ, Kardia SLR, Keene KL, Kissela BM, Kleindorfer DO, Koblar S, Labovitz D, Launer LJ, Laurie CC, Laurie CA, Lee CH, Lee J-M, Lehm M, Lemmens R, Levi C, Leys D, Lindgren A, Longstreth WT, Maguire J, Manichaikul A, Markus HS, McClure LA, McDonough CW, Meisinger C, Melander O, Meschia JF, Mola-Caminal A, Montaner J, Mosley TH, Müller-Nurasyid M, Nalls MA, O'Connell JR, O'Donnell M, Ois Á, Papanicolaou GJ, Paré G, Peddareddygari LR, Pedersén A, Pera J, Peters A, Poole D, Psaty BM, Rabionet R, Raffeld MR, Rannikmäe K, Rasheed A, Redfors P, Reiner AP, Rexrode K, Ribasés M, Rich SS, Robberecht W, Rodriguez-Campello A, Rolfs A, Roquer J, Rose LM, Rosenbaum D, Rost NS, Rothwell PM, Rundek T, Ryan KA, Sacco RL, Sale MM, Saleheen D, Salomaa V, Sánchez-Mora C, Schmidt CO, Schmidt H, Schmidt R, Schürks M, Scott R, Segal HC, Seiler S, Seshadri S, Sharma P, Shuldiner AR, Silver B, Slowik A, Smith JA, Söderholm M, Soriano C, Sparks MJ, Stanne T, Stefansson K, Stine OC, Strauch K, Sturm J, Sudlow CL, Tajuddin SM, Talbert RL, Tatlisumak T, Thijs V,

Thorleifsson G, Thorsteindottir U, Tiedt S, Traylor M, Trompet S, Valant V, Waldenberger M, Walters M, Wang L, Wassertheil-Smoller S, Weir DR, Wiggins KL, Williams SR, Wloch-Kopec D, Woo D, Woodfield R, Wu O, Xu H, Zonderman AB, Worrall BB, de Bakker PIW, Kittner SJ, Mitchell BD, Rosand J (2016) The NINDS Stroke Genetics Network: a genome-wide association study of ischemic stroke and its subtypes. Lancet Neurol. 15:174–184. https://doi.org/10.1016/S1474-4422(15)00338-5

Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, Guiducci C, Perola M, Jula A, Sinisalo J, Lokki M-L, Nieminen MS, Melander O, Salomaa V, Peltonen L, Kathiresan S (2010) A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. The Lancet 376:1393–1400. https://doi.org/10.1016/S0140-6736(10)61267-6

Roden DM Jr, George AL Jr (2002) The genetic basis of variability in drug responses. Nat Rev Drug Discov 1:37–44. https://doi.org/10.1038/nrd705

Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, Patterson N, Reich D (2014) The genomic landscape of Neanderthal ancestry in present-day humans. Nature 507:354–357. https://doi.org/10.1038/nature12961

Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, Preuss M, Stewart AFR, Barbalic M, Gieger C, Absher D, Aherrahrou Z, Allayee H, Altshuler D, Anand SS, Andersen K, Anderson JL, Ardissino D, Ball SG, Balmforth AJ, Barnes TA, Becker DM, Becker LC, Berger K, Bis JC, Boekholdt SM, Boerwinkle E, Braund PS, Brown MJ, Burnett MS, Buysschaert I, Carlquist C, John F, Chen L, Cichon S, Codd V, Davies RW, Dedoussis G, Dehghan A, Demissie S, Devaney JM, Do R, Doering A, Eifert S, El Mokhtari NE, Ellis SG, Elosua R, Engert JC, Epstein SE, de Faire U, Fischer M, Folsom AR, Freyer J, Gigante B, Girelli D, Gretarsdottir S, Gudnason V, Gulcher JR, Halperin E, Hammond N, Hazen SL, Hofman A, Horne BD, Illig T, Iribarren C, Jones GT, Jukema JW, Kaiser MA, Kaplan LM, Kastelein JJP, Khaw K-T, Knowles JW, Kolovou G, Kong A, Laaksonen R, Lambrechts D, Leander K, Lettre G, Li M, Lieb W, Linsel-Nitschke P, Loley C, Lotery AJ, Mannucci PM, Maouche S, Martinelli N, McKeown PP, Meisinger C, Meitinger T, Melander O, Merlini PA, Mooser V, Morgan T, Mühleisen TW, Muhlestein JB, Münzel T, Musunuru K, Nahrstaedt J, Nelson CP, Nöthen MM, Olivieri O, Patel RS, Patterson CC, Peters A, Peyvandi F, Qu L, Quyyumi AA, Rader DJ, Rallidis LS, Rice C, Rosendaal FR, Rubin D, Salomaa V, Sampietro ML, Sandhu MS, Schadt E, Schäfer A, Schillert A, Schreiber S, Schrezenmeir J, Schwartz SM, Siscovick DS, Sivananthan M, Sivapalaratnam S, Smith A, Smith TB, Snoep JD, Soranzo N, Spertus JA, Stark K, Stirrups K, Stoll M, Tang WHW, Tennstedt S, Thorgeirsson G, Thorleifsson G, Tomaszewski M, Uitterlinden AG, van Rij AM, Voight BF, Wareham NJ, Wells GA, Wichmann H-E, Wild PS, Willenborg C, Witteman JCM, Wright BJ, Ye S, Zeller T, Ziegler A, Cambien F, Goodall AH, Cupples LA, Quertermous T, März W, Hengstenberg C, Blankenberg S, Ouwehand WH, Hall AS, Deloukas P, Thompson JR, Stefansson K, Roberts R, Thorsteinsdottir U, O'Donnell CJ, McPherson R, Erdmann J, Samani NJ (2011) Large-scale association analyses identifies 13 new susceptibility loci for coronary artery disease. Nat Genet 43:333–338. https://doi.org/10.1038/ng.784

Simons YB, Bullaughey K, Hudson RR, Sella G (2018) A population genetic interpretation of GWAS findings for human quantitative traits. PLoS Biol 16:e2002985. https://doi.org/10.1371/journal.pbio.2002985

Simonti CN, Vernot B, Bastarache L, Bottinger E, Carrell DS, Chisholm RL, Crosslin DR, Hebbring SJ, Jarvik GP, Kullo IJ, Li R, Pathak J, Ritchie MD, Roden DM, Verma SS, Tromp G,

Prato JD, Bush WS, Akey JM, Denny JC, Capra JA (2016) The phenotypic legacy of admixture between modern humans and Neandertals. Science 351:737–741. https://doi.org/10.1126/science.aad2149

Sohail M, Maier RM, Ganna A, Bloemendal A, Martin AR, Turchin MC, Chiang CWK, Hirschhorn JN, Daly M, Patterson N, Neale B, Mathieson I, Reich D, Sunyaev SR (2018) Signals of polygenic adaptation on height have been overestimated due to uncorrected population structure in genome-wide association studies. bioRxiv. https://doi.org/10.1101/355057

Sohail M, Maier RM, Ganna A, Bloemendal A, Martin AR, Turchin MC, Chiang CW, Hirschhorn J, Daly MJ, Patterson N, Neale B, Mathieson I, Reich D, Sunyaev SR (2019) Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. eLife 8:e39702. https://doi.org/10.7554/eLife.39702

Song M (2010) Does 'race' matter? A study of 'mixed race' siblings' identifications. Sociol Rev 58:265–285. https://doi.org/10.1111/j.1467-954X.2010.01903.x

Speed D, Balding DJ (2015) Relatedness in the post-genomic era: is it still useful? Nat Rev Genet 16:33–44. https://doi.org/10.1038/nrg3821

Speed D, Balding DJ (2019) SumHer better estimates the SNP heritability of complex traits from summary statistics. Nat Genet 51:277. https://doi.org/10.1038/s41588-018-0279-5

Speed D, Hemani G, Johnson MR, Balding DJ (2012) Improved heritability estimation from genome-wide SNPs. Am J Hum Genet 91:1011–1021. https://doi.org/10.1016/j.ajhg.2012.10.010

Speed D, O'Brien TJ, Palotie A, Shkura K, Marson AG, Balding DJ, Johnson MR (2014) Describing the genetic architecture of epilepsy through heritability analysis. Brain 137:2680–2689. https://doi.org/10.1093/brain/awu206

Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R (2015) UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLOS Med 12:e1001779. https://doi.org/10.1371/journal.pmed.1001779

The UK10 K Consortium (2015) The UK10 K project identifies rare variants in health and disease. Nature 526:82–90. https://doi.org/10.1038/nature14962

Timpson NJ, Greenwood CMT, Soranzo N, Lawson DJ, Richards JB (2018) Genetic architecture: the shape of the genetic contribution to human traits and disease. Nat Rev Genet 19:110–124. https://doi.org/10.1038/nrg.2017.101

Vrieze SI, Iacono WG, McGue M (2012) Confluence of genes, environment, development, and behavior in a Post-GWAS World. Dev Psychopathol 24:1195–1214. https://doi.org/10.1017/S0954579412000648

Wall JD, Yang MA, Jay F, Kim SK, Durand EY, Stevison LS, Gignoux C, Woerner A, Hammer MF, Slatkin M (2013) Higher levels of Neanderthal Ancestry in East Asians than in Europeans. Genetics 194:199–209. https://doi.org/10.1534/genetics.112.148213

Winkler CA, Nelson GW, Smith MW (2010) Admixture mapping comes of age. Annu Rev Genomics Hum Genet 11:65–89. https://doi.org/10.1146/annurev-genom-082509-141523

Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88:76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Young AI, Frigge ML, Gudbjartsson DF, Thorleifsson G, Bjornsdottir G, Sulem P, Masson G, Thorsteinsdottir U, Stefansson K, Kong A (2018) Relatedness disequilibrium regression estimates

heritability without environmental bias. Nat Genet 50:1304–1310. https://doi.org/10.1038/s41588-018-0178-9

Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–360. https://doi.org/10.1038/ng.546

Zhang Y, Zhang J, Yang J, Wang Y, Zhang L, Zuo X, Sun L, Pan H-F, Hirankarn N, Wang T, Chen R, Ying D, Zeng S, Shen JJ, Lee TL, Lau CS, Chan TM, Leung AMH, Mok CC, Wong SN, Lee KW, Ho MHK, Lee PPW, Chung BH-Y, Chong CY, Wong RWS, Mok MY, Wong WHS, Tong KL, Tse NKC, Li X-P, Avihingsanon Y, Rianthavorn P, Deekajorndej T, Suphapeetiporn K, Shotelersuk V, Ying SKY, Fung SKS, Lai WM, Wong C-M, Ng IOL, Garcia-Barcelo M-M, Cherny SS, Tam PK-H, Sham PC, Yang S, Ye DQ, Cui Y, Zhang X-J, Lau YL, Yang W (2015) Meta-analysis of GWAS on two Chinese populations followed by replication identifies novel genetic variants on the X chromosome associated with systemic lupus erythematosus. Hum Mol Genet 24:274–284. https://doi.org/10.1093/hmg/ddu429

Zuccolo L, Lewis SJ, Davey Smith G, Sayal K, Draper ES, Fraser R, Barrow M, Alati R, Ring S, Macleod J, Golding J, Heron J, Gray R (2013) Prenatal alcohol exposure and offspring cognition and school performance. A 'Mendelian randomization' natural experiment. Int J Epidemiol 42:1358–1370. https://doi.org/10.1093/ije/dyt172