Trends in Cognitive Sciences

CellPress

Review

Twin studies to GWAS: there and back again

Naomi P. Friedman,[1,2,*] Marie T. Banich,[1,3] and Matthew C. Keller[1,2]

The field of human behavioral genetics has come full circle. It began by using twin/family studies to estimate the relative importance of genetic and environmental influences. As large-scale genotyping became cost-effective, genome-wide association studies (GWASs) yielded insights about the nature of genetic influences and new methods that use GWAS data to estimate heritability and genetic correlations invigorated the field. Yet these newer GWAS methods have not replaced twin/family studies. In this review, we discuss the strengths and weaknesses of the two approaches with respect to characterizing genetic and environmental influences, measurement of behavioral phenotypes, and evaluation of causal models, with a particular focus on cognitive neuroscience. This discussion highlights how twin/family studies and GWAS complement and mutually reinforce one another.

Highlights

Traditional twin/family studies and genome-wide association studies (GWAS) are complementary methods for behavioral genetic research.

Twin/family studies estimate aggregate genetic and environmental influences on a trait, while GWAS suggest more specific plausible biological mechanisms.

Estimates of genetic (co)variance from these two approaches have different assumptions, require different sample sizes for adequate power, and have distinct interpretations.

Twin/family studies typically examine deep phenotypes, such as executive functioning or memory; GWAS tend to focus on general phenotypes (e.g., intelligence), often minimally assessed, that can be administered in large samples or harmonized across multiple studies.

Data from both approaches can be used to test causal models. Family data can be used to control for confounds in GWAS and probe gene–environment interplay.

From twin and family studies to GWAS and back again: complementary approaches

Behavioral genetics is the study of genetic and environmental influences on individual differences in behavioral **phenotypes** (see Glossary). Historically, researchers used twin/family studies to estimate the relative influences of genes and environment. The logic of these studies is simple: to the degree a trait is genetically influenced, individuals who share more genetic material should be more similar on that trait. This logic is especially sound when environments are as similar as possible between different types of relatives, which is why twin studies have taken a central place in behavioral genetics for the past half century. Both monozygotic (MZ; identical) and dizygotic (DZ; fraternal) twins share familial and other background environmental influences (e.g., age, neighborhood), so there should be no major differences in the degree of environmental similarity between the two twin types. Genetic influences can therefore be estimated by the degree to which the correlation between phenotypes of MZ twins, who share 100% of their segregating genes, is higher than the correlation between DZ twins, who share 50% on average. From such comparisons researchers estimate **heritability (h^2)**, the proportion of population variability due to genetic influences, and **genetic correlations (*rG*)**, the degree to which two traits have similar genetic influences (Box 1). Results from thousands of twin studies show that most behavioral traits are partially heritable, but also substantially influenced by the environment [1,2].

Beginning in the late 1980s, the next major phase of research in behavioral genetics emphasized the molecular basis of heritability by examining whether specific measured **genotypes** were associated with behavior. Particularly in the past decade, **genome-wide association studies (GWASs)**, which examine the effects of **single nucleotide polymorphisms (SNPs)** measured across the genome (Box 2), have ushered in a new era of scientific discovery and methodological developments [3]. Yet twin/family studies continue to have a central role in human behavioral genetic research. In this review, we highlight the complementary strengths and limitations of GWAS and twin/family designs with respect to: (i) characterization of genetic influences and genetic correlations, (ii) measurement of phenotypes, (iii) characterization of environmental influences, and (iv) evaluation of causal models.

[1]Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, CO 80309, USA
[2]Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO 80309, USA
[3]Institute of Cognitive Science, University of Colorado Boulder, Boulder, CO 80309, USA

*Correspondence: naomi.friedman@colorado.edu (N.P. Friedman).

How GWAS and twin/family data are used to understand genetic influences

Traditional behavioral genetics methods compare the covariances of close relatives, usually twins, to estimate the proportion of trait variation due to genetic versus environmental differences between individuals (Figure 1A). While these methods do not identify the specific genetic variants that influence traits, they suggest that such variants exist. To identify those variants, researchers

Box 1. Estimating heritability and genetic correlations in twin studies

Correlations among relatives can arise from a combination of shared genes and shared environments. Twin studies tease apart these influences by evaluating the extent to which genetic similarity is associated with phenotypic similarity. Structural equation twin models (see Figure 1A in main text) separate phenotypic variance into three latent (unobserved) factors: (i) additive genetic influences (A), the sum of a large number of genetic variants' effects on the trait; (ii) common or shared environmental influences (C), environmental influences that lead to within-family resemblance); and (iii) nonshared environmental influences (E), environmental influences, including measurement error, that do not lead to within-family resemblance. Heritability (h^2) is the proportion of total variance due to genetic influences: A/(A + C + E).

Multivariate twin models use cross-twin associations between different traits to estimate genetic (*r*A or *r*G) and environmental (*r*C and *r*E) correlations between those traits. The *r*G estimate increases to the degree that cross-twin, cross-trait correlations (e.g., twin 1's cognitive ability with twin 2's substance use) are greater for MZ compared with DZ twins. Evidence for *r*C occurs to the degree that these correlations in DZ twins approach those in MZ twins. Evidence for *r*E occurs when these cross-twin correlations are lower than the phenotypic (within-twin, cross-trait) correlations.

High heritability is sometimes misinterpreted as indicating that the behavior is biologically determined and immutable (genetic determinism). Heritability is the proportion of variance explained by genetic differences in a population in a given environment and at a given time. Hence, heritability can vary across populations such as countries, ages, cohorts, and other grouping criteria that reflect different environmental contexts. For example, in the United States, the heritabilities of IQ and academic achievement appear to be higher in the context of high SES [101,102], suggesting that low SES inhibits expression of genetic potential. Some argue that since heritability estimates are population-specific, they are useless theoretically [103]. However, understanding how patterns change across generation, age, sex, and environmental contexts provides a richer view of individual differences that can lead to insights about gene–environment interplay [23,104]. High heritability is also not inconsistent with change, as the mean trait level in a population can increase or decrease while rank orders remain similar. Moreover, because heritability is about differences within the population in which it is estimated, heritability has no necessary relevance to the causes of between-group differences (e.g., race, sex, age cohorts), which can reflect different environmental contexts [104].

Box 2. GWAS methods

A GWAS consists of a series of independent regressions of a phenotype on genetic variants (specifically, SNPs) across the genome, one regression per SNP (see Figure 2A in main text). A modern SNP array includes fewer than 1 million 'common' SNPs [those whose **minor allele frequency (MAF)** > 1%], even though there are ~15 million common SNPs in the human genome [105]. Despite this discrepancy, the arrays capture the influence of almost all common SNPs as well as most other types of common polymorphisms (e.g., deletions/insertions), because genetic variants that are nearby on the genome tend to be correlated (in LD). Thus, each measured SNP 'tags' other nearby variants. Most modern GWASs impute nearly all common SNPs and use these imputed scores directly in the analysis.

Due to LD, the influence of a single causal variant will typically manifest itself in nearby SNPs. Thus, GWAS hits implicate a region where a causal variant is likely to exist rather than pinpointing the actual causal variant, making biological interpretation of individual variants challenging. Nevertheless, in aggregate, GWAS hits are not random. They occur more often in or near genes and regulatory regions and they often cluster in biologically meaningful genetic pathways (e.g., SNPs associated with bipolar disorder cluster in synaptic signaling pathways [106]), giving insight into the pathophysiology of disease [107].

The field has agreed upon standards and practices designed to minimize false positives. Because systematic differences in allele frequencies exist between subpopulations, and because environmental influences can differ across these subpopulations, it is routine to control for such **population stratification** by using mixed linear models and/or by including genomic principal components as covariates. Regardless of the number of SNPs examined, the maximum effective number of independent tests conducted on common SNPs in the genome is ~1 million [108], so the standard in the field is to require a Bonferroni-corrected threshold of $0.05/10^6 = 5 \times 10^{-8}$ for a SNP association to be considered 'genome-wide significant'. Modern GWASs are performed on tens of thousands to millions of individuals, with sample size being the most important factor in determining the number of significant associations [19]. To achieve large samples, the field has adopted a highly collaborative approach, including forming large GWAS consortia, open access to large biobank datasets, and sharing summary statistics for all SNP tests following a GWAS. These practices reduce barriers of entry to the field, maximize data usage, increase transparency, and encourage crossfertilization of ideas.

Glossary

Allele: one of two or more forms of a DNA segment.
Gene–environment correlation (*r*GE): when genetic influences for a particular trait are more likely to co-occur with particular environmental contexts. *r*GE can be passive (the variants that offspring inherit also influence the parent's behavior or environment (e.g., a child who inherits risk alleles for smoking has parents who smoke), active (variants influence one's choice of environment), or evocative (an individual's genetically influenced trait evokes responses from others).
Genetic correlation (*r*G): correlation between the genetic influences of two traits. *r*G is the genetic covariance scaled by the product of the square root of the genetic variances of the two traits; a high *r*G can occur even when the traits have low heritabilities.
Genome-wide association study (GWAS): study in which a phenotype is regressed on each of a large number of SNPs measured across the genome (Figure 2A and Box 2).
Genotype: an individual's combination of two alleles at a particular location across both chromosomes.
Heritability (h^2): the proportion of population variability in a phenotype due to genetic differences. h^2 estimates based on twin studies include all sources of genetic variance, whereas estimates based on GWAS data do not: h^2_{SNP} only includes variance explained by common SNPs.
Linkage disequilibrium (LD): correlations between nearby SNPs on the genome. LD patterns differ across populations.
Minor allele frequency (MAF): the population frequency of the least common allele for a genetic variant. SNPs with MAF > 1% are typically considered common.
Missing heritability: the discrepancy between h^2 estimated in twin studies and the summed variation explained by genome-wide significant variants. h^2_{SNP}, which includes the influence of variants not reaching genome-wide significance, accounts for some missing heritability. While the cause of 'still missing' heritability (h^2–h^2_{SNP}) remains unclear, it is likely attributable to (mostly rare) variants that are poorly tagged by common SNPs on arrays.
Phenotype: observed characteristic, such as a test score.

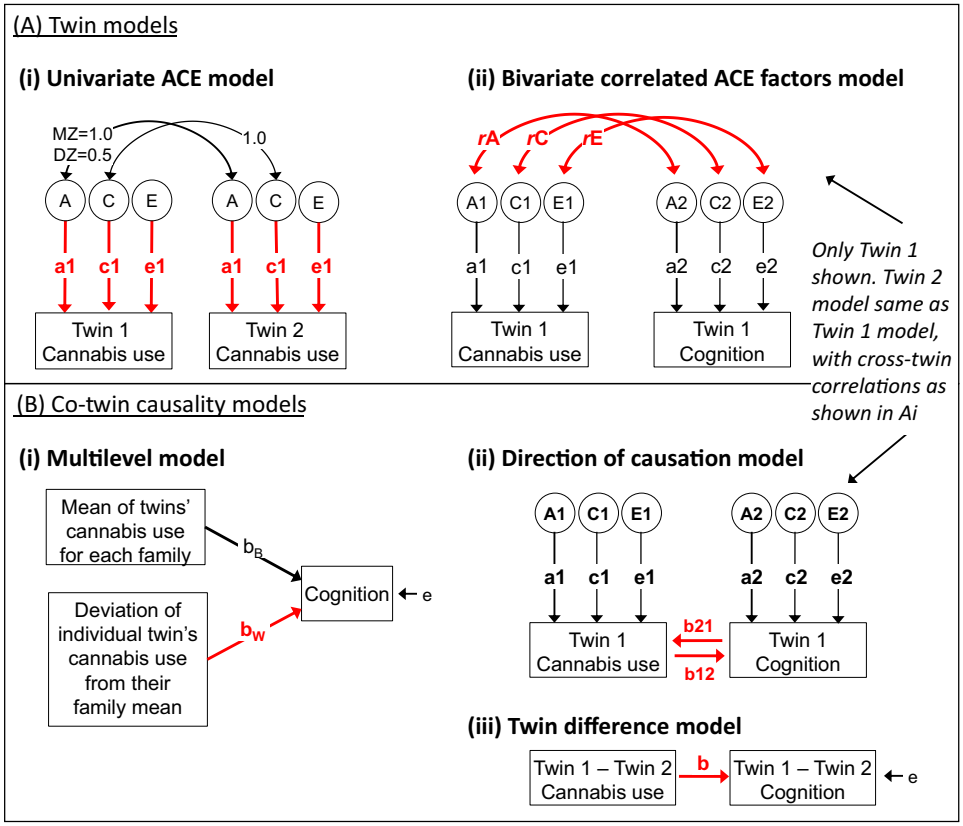

Figure 1. Illustration of common models used with family data. Red indicates parameters of key interest in each method. (A) Univariate twin models (i) leverage correlations between monozygotic (MZ) and dizygotic (DZ) twins to partition total variance into additive genetic (A), shared environment (C), and nonshared environment (E) components. Bivariate models (ii) leverage cross-twin correlations to estimate genetic (*r*A) and environmental (*r*C and *r*E) correlations between these variance components. Only one twin is shown for simplicity. (B) Causal models using twin data examine how twin differences in a purported cause relate to twin differences in a purported outcome, either with multilevel models (i) or twin difference models (iii). Direction of causation models (ii) are bivariate twin models that are constrained such that all genetic and environmental correlations must be explained with unidirectional or bidirectional paths between the observed variables; finding that this model fits significantly worse than the less restrictive bivariate twin model in panel Aii is inconsistent with causality.

typically turn to associations studies, which correlate **alleles** (alternative forms of DNA segments) at one or more locations across the genome with a trait (Figure 2A). The initial 'candidate gene' approach focused on specific alleles within genes hypothesized to have large effects on biological systems. However, for reasons discussed in Box 3, within the past 10 years, the field of human genetics has become disillusioned with this hypothesis-driven approach and has instead largely moved toward the hypothesis-free GWAS approach conducted in large samples. Within the cognitive domain, large GWASs examining general cognitive ability or intelligence quotient (IQ) [4–7], brain imaging phenotypes [8–16], and educational attainment [17] have yielded hundreds of associations. None of these associations were predicted beforehand; nonetheless, many are biologically plausible, for example, genes associated with neurogenesis and myelination influence variation in IQ [7].

Thousands of GWASs on millions of individuals have now been conducted, yielding tens of thousands of replicable associations [18,19]. From these results, a clear picture of the genetics underlying

Polygenic: when a phenotype is influenced by many genetic variants.
Polygenic score (PGS): score that captures an individual's overall genetic propensity for a phenotype. A PGS is usually constructed as the weighted sum of 'risk' alleles each person possesses, with the weights determined by effect sizes from an independent GWAS. PGS are also sometimes called polygenic risk scores or genomic profile risk scores.
Population stratification: systematic differences in allele frequencies between subpopulations. When environmental differences lead to mean trait differences across subpopulations, population stratification can confound GWAS estimates.
SNP: pronounced 'snip'; type of genetic variant in which the base letter at a single nucleotide varies across the population. SNPs are numerous across the genome and easy to genotype at large scale, making them the main type of variation examined in GWAS.

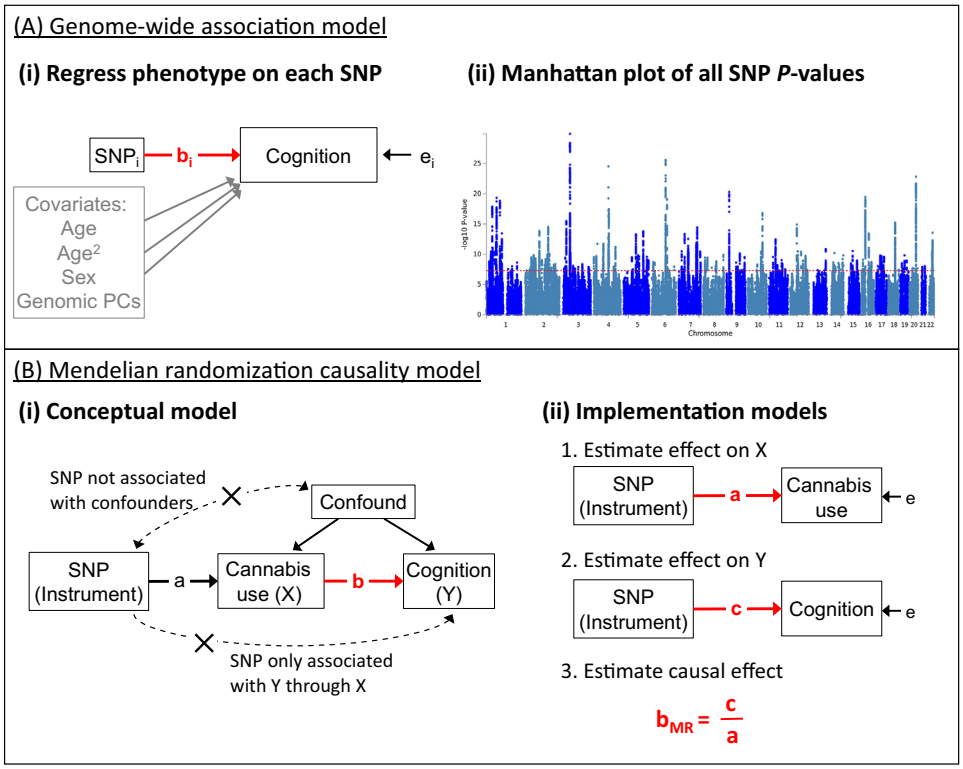

Figure 2. Illustration of genome-wide association studies (GWAS) and Mendelian randomization models. Red indicates parameters of key interest in each method. (A) A GWAS consists of a series of regressions of the phenotype on each single nucleotide polymorphism (SNP), controlling for covariates like age, sex, and ancestry as measured with genomic principal components (PCs). The Manhattan plot organizes the log *P* value (y-axis) of each SNP tested by its position on the chromosomes (x-axis); the red line indicates genome-wide significance α = 5e–8. (B) If the Mendelian randomization assumptions (i) are met, the analysis models shown in (ii) allow for estimation of causal effects with GWAS data.

complex traits has emerged. Most crucially, the effect sizes of individual genetic variants are almost always extremely small [20]. The median odds ratio for significantly associated variants across all common diseases is ~1.09, corresponding to $r^2 < 0.05\%$ (five one-hundredths of one percent) [18]. Such small effect sizes imply that most complex traits are extremely **polygenic** [20], such that their heritability is due to the sum of thousands of causal variants [21].

While identifying specific variants associated with a trait is a major focus of GWAS, methodological innovations have expanded the utility of GWAS data to estimate heritability, genetic correlations, and causality – traditional domains of twin/family studies [3,22,23]. A popular method for estimating heritability and genetic correlations using GWAS data is genomic restricted maximum likelihood (GREML). Similar to twin/family studies, GREML evaluates the extent to which individuals who are more genetically similar are also more phenotypically similar [24]. Unlike twin/family studies, however, GREML uses SNPs to calculate genetic similarity (sharing of alleles across the genome) among 'unrelated' individuals who nevertheless vary in their genetic relatedness due to sharing of distant ancestors. GREML focuses on the very narrow range of genetic similarity that is typical among unrelated individuals (e.g., more distantly related than second cousins), as opposed to the large range of genetic similarity observed in twin and family studies [e.g., twin pairs who share 50% (DZ) versus 100% (MZ) of segregating genetic variants]. Because GREML uses

Box 3. Candidate gene approach

Candidate gene approaches investigate specific genetic variants (out of ~15 million common variants in the genome) hypothesized to be relevant to a particular trait. Often, the same ~30 'usual suspect' polymorphisms are studied repeatedly across a wide range of traits [109] (e.g., the COMT val/met polymorphism alone has been associated with cognitive abilities, personality dimensions, psychiatric disorders, substance abuse, dementia, Parkinson's disease, cancer, obesity, and height). Candidate gene sample sizes are typically in the hundreds, orders of magnitude smaller than typical GWAS sample sizes, yet most published candidate gene studies reported significant associations [32,34]. However, findings over the past decade have increased skepticism about the approach. For one, effect sizes detected in nearly all GWASs on complex traits are far smaller than those detected in candidate gene studies. More damningly, when the same traits and variants have been investigated using GWAS data in samples that have ~100% power to detect typical effects reported in the literature, most candidate gene findings have failed to replicate [34,76,110–112].

What could explain the discrepancy between candidate gene findings and those reported in GWAS? Higher error in the measurement of traits in large GWAS samples is unlikely to explain this discrepancy because the power gained due to sample size more than compensates for the decrease in power caused by any realistic amount of measurement error [34,113]. Rather, it is likely that positive findings for candidate gene studies are mostly false positives [32–34]. The candidate gene literature shows evidence for a number of factors that increase the false positive rate in a field: publication bias, multiple approaches that researchers can take in performing analyses to find one yielding a significant result, poor methodological practices (e.g., lack of control for population stratification), low prior probability of a given single candidate polymorphism having a large behavioral effect, and low power due to small sample sizes [32]. In response to concerns about the utility of the approach, geneticists interested in complex traits have largely abandoned candidate gene studies and it has become difficult to get them reviewed in genetics journals [114], although behavioral science journals continue to publish them [34]. With increasing access to large GWAS datasets, genome-wide array data that is as cheap (~$40/sample) as candidate gene data, and the ability to use PGSs in relatively small samples, the era of candidate gene studies is drawing to a close.

genomic similarity at measured SNPs, its estimate of heritability ('SNP-heritability', or h^2_{SNP}) captures only the genetic effects explained by the SNPs used in the analysis. Thus, h^2_{SNP} changes as a function of how many SNPs are interrogated across the genome in a GWAS study (Box 2) and is almost always lower than h^2 estimated from twin/family studies, as explained later. **Linkage disequilibrium (LD)**-score regression [25], a related technique for estimating h^2_{SNP} and genetic correlations, has become widely adopted because it is less computationally demanding than GREML and only requires GWAS summary statistics, which are typically publicly available.

GWAS results can also be used to construct **polygenic scores (PGSs)**, which quantify an individual's overall genetic propensity or 'risk' for a disorder or trait. Similar to how an individual with a family history of schizophrenia is at higher risk to develop the disorder, an individual with a high PGS for schizophrenia carries a high burden of schizophrenia risk alleles and therefore has a higher risk of developing schizophrenia compared to someone with a low PGS. Although many methods for computing PGSs exist [3,26], in their simplest form, PGSs are estimated by multiplying an individual's allele count at a SNP by a weight for that SNP's predicted regression effect ($\hat{\beta}$) on the trait, which was estimated in an independent GWAS. These weighted allele counts are then summed at all relevant measured SNPs across the genome [3,22,27,28]. By aggregating the influences of tens of thousands of variants across the genome, PGSs explain more variation than individual SNPs. For example, in the Twins Early Development Study, PGSs for IQ and educational attainment respectively explained 6.7% of the variance in IQ and 14.8% of the variance in educational achievement at age 16 years, far more than the effect of any single SNP [29].

Once estimated, a PGS can be used like other measured variables to test a range of interesting questions. For example, PGS studies have shown that individuals with higher polygenic risk for major depressive disorder are more likely to develop depression when they have also experienced stressful life events [30,31]. This interaction effect supports a prominent model of gene–environment interplay, but focuses on polygenic risk rather than the single alleles that were the

focus of prior (inconsistent) candidate gene studies [32–34]. In addition to such tests of theoretical models, PGSs can be used to understand developmental and/or neural mechanisms for disease and normal variation [35]. For instance, PGSs for schizophrenia are associated with childhood cognitive, social, and emotional impairments [36] and PGSs for attention deficit/hyperactivity disorder (ADHD), IQ, and educational attainment are associated with smaller caudate or total brain volumes in children [37], suggesting genetic individual differences in these traits may involve early behavioral and neural differences.

All of these findings come from studies of modest sample sizes (n = 1139–12 490) compared with those required to detect individual SNP effects in GWAS, demonstrating the increased effect size and statistical power afforded by using PGSs. Because of their larger effects, PGSs can be used with smaller sample sizes (e.g., in the thousands) than those needed to detect individual SNP effects in GWASs, which require sample sizes tens to hundreds of times larger to be adequately powered [22,38]. As GWAS sample sizes increase and individual SNP effects are estimated with greater precision, the predictive ability of PGSs will also increase, better enabling identification of individuals at risk for disease to facilitate early intervention and prevention [39].

How methods based on GWAS and twin/family data complement one another

It may be tempting to view the progression from twin/family studies to approaches that estimate genetic (co)variation using GWAS data as one of outdated approaches giving way to modern ones that achieve the same goals in better ways. We believe this view is misguided. As summarized in Table 1 (Key table), the goals and advantages of twin/family and GWAS approaches are complementary and both are needed for optimal scientific discovery.

Perhaps most importantly, the complementarity of the two approaches is due to their largely nonoverlapping strengths and weaknesses. Though both approaches can be used to quantify heritability and genetic correlations, twin/family studies can often explore more nuanced and detailed hypotheses than is possible with GWAS data. Twin/family studies can probe associations by estimating multiple influences simultaneously; for example, the association between height and IQ is due to both shared genes and the tendency for individuals to select mates that are similar to themselves on these traits [40]. They also can test specific psychological models, for example, testing the 'developmental propensity model', which posits that early, heritable measures of temperament and cognition constitute distinct propensities to antisocial behavior that jointly influence later conduct problems [41].

By contrast, techniques that use GWAS data, like GREML and LD-score regression, can incorporate biological information about SNPs (e.g., whether they are in genes, regulatory regions, etc.) to investigate whether specific types of SNPs are particularly important ('enriched') compared with other SNPs/genes. For example, Feng *et al*. [42] recently found that genes that were preferentially expressed in the central nervous system and genes associated with IQ explained variation in resting-state fMRI brain connectivity in almost all the brain networks that showed significant heritability. By contrast, genes related to schizophrenia and ADHD were both related to the frontoparietal network but also showed specific relationships to the visual (schizophrenia) and somatomotor (ADHD) networks, respectively. Another major advantage of techniques like GREML and LD-score regression is that they can estimate genetic correlations between traits measured in different individuals, an impossibility in twin/family studies. For example, these methods have been used to estimate the genetic correlations between the same disorder across different ethnicities [43] and between disorders such as schizophrenia and bipolar disorder [44], which do not co-occur within individuals because their diagnoses are mutually exclusive and rarely co-occur within families due to low base rates.

Key table

Table 1. Comparison of twin/family and genome-wide association study (GWAS) approaches[a]

Attribute	Twin/family studies	GWASs of unrelated individuals
Original/primary goals	• Decompose (co)variances into genetic and environmental influences at aggregate level	• Identify individual genetic associations across the genome
Additional uses for data	• Multivariate structural equation models simultaneously modeling both genetic and environmental structure	• Bio-informatic follow-up to evaluate biological pathways and types of SNPs associated with a phenotype
	• Causal models (co-twin control and direction of causation models)	• Causal models (Mendelian randomization)
		• Polygenic (risk) scores
	• Estimate heritability (h^2)	
	• Calculate genetic correlations (*r*G)	
	• Genetic structural equation modeling	
Strengths (+) and weaknesses (–)	+ Small samples (e.g., 400+ twin pairs) can still be adequately powered to estimate h^2 and *r*G	– Requires large samples (10s to 100s of thousands of individuals) for adequate power to estimate SNP effects, h^2, and *r*G
	+ Potential for deep phenotyping, including longitudinal assessments	– Often has minimal phenotyping
	+ h^2 includes the effects of all genetic variants (common and rare)	– h^2 includes only effects of variants tagged by array (misses rare variant effects)
	+ Can estimate multiple influences (e.g., with extended twin designs: additive and nonadditive genetic influences, parental influences, *r*GE)	– Typically estimate just additive genetic influences
	+ Within-family tests control for population stratification, environmental confounds	– Sensitive to population stratification; require careful control for ancestry
	– Blind to biological information regarding the types of variants that give rise to h^2	+ Allows estimates of the relative importance of different types of SNPs (enrichment analyses)
	– Assumes environments for monozygotic and dizygotic pairs are equivalent; violations can result in biased estimates	+ In unrelated individuals, genetic similarities are unlikely to be confounded with environmental similarities after controlling for ancestry
	– Estimating *r*G requires that both traits are measured on the same individuals	+ Can estimate *r*G across different samples or across mutually exclusive traits
	– Cannot separate direct from indirect genetic effects in samples of unrelated individuals (GWAS) or using only twins (twin/family studies)	
	– Historic lack of diverse samples (with particular concern about exacerbating health disparities for GWAS discoveries)	
	Combined approaches (e.g., GWAS using within-family design)	
	+ Excellent control of population stratification and *r*GE	
	+ Differentiate direct from indirect genetic effects	
	+ Test more complex models, including gene–environment interplay	

[a]Text aligned under each method applies to that approach, whereas centered text applies to both approaches.

Importantly, however, the heritability estimated with GWAS data (h^2_{SNP}) is not the same as that estimated with twin/family data (h^2). The former is the portion of variation tagged by the specific SNPs on the array (the set of SNPs genotyped) used in the analysis. While modern arrays do a good job of capturing the influences of nearly all common SNPs, they miss the influences of certain types of variants, such as rare variants that are not well tagged by the genotyped or imputed SNPs but may nonetheless have a large influence [19]. For this reason, h^2_{SNP} is lower (typically by 40–80%) than the total h^2 of the same trait estimated in twin/family studies. The

remaining h^2 not explained by SNPs (h^2– h^2_{SNP}) is referred to as 'still-**missing heritability**' [22]. Moreover, because twin/family studies use a larger range of genetic similarity (e.g., 100% similarity in MZ versus 50% in DZ) to estimate heritability than do methods like GREML, they have greater statistical power for a given sample size and can therefore be used to test hypotheses about genetic (co-)variation, causation, or gene–environment interplay in smaller samples; twin studies are generally well powered with 400 pairs [45], whereas GREML requires $n > 4000$ to detect heritability greater than 20% [46].

Similarly, the assumptions required, and biases when these assumptions are unmet, are different across the two approaches [47,48]. For example, the twin design has been criticized for relying on the 'equal environments assumption': the assumption that MZ and DZ twins share equally similar environments [23]. Violations of this assumption can lead to overestimates of heritability, causing some to argue that twin studies are seriously flawed [49] (but see [50] for counterpoints and [51] for evaluations of the biases caused by violating this assumption). By contrast, methods that use GWAS data typically do not include close relatives, thus lessening any confounding of genetic and environmental similarity. However, estimates from GWAS can be biased by other nongenetic factors that cause differences between families, such as population stratification. Thus, GWASs require rigorous control for ancestry and to date have been largely restricted to white Europeans, the most commonly assessed ancestry. This lack of diversity means that GWAS discoveries may not translate to other ethnic groups, which could increase health disparities [52–54].

Because the assumptions and biases of twin/family studies and GWAS are different, confidence about their results are strengthened when both approaches reach consistent conclusions. Moreover, combining these approaches enables many biases to be identified and addressed, as discussed more in the following sections.

GWAS and twin/family studies differ in the depth of phenotype measurement

The observation that the genetic variance in complex traits is attributable to the sum of thousands of individually tiny effects means that very large samples are needed to detect individual genetic effects, particularly after correction for the 1 million independent tests in a typical GWAS ($\alpha = 5 \times 10^{-8}$) [19]. Large GWAS sample sizes are enabled by biobanks such as the UK Biobank [55], databases compiled by companies like 23andMe, and consortia that harmonize samples across the globe with relevant phenotypes. Two examples of such consortia are COGENT (Cognitive Genomics Network), which has identified genes associated with general cognitive function [56] and ENIGMA (Enhancing NeuroImaging Genetics through Meta-Analysis), which has identified genes associated with brain structure and a variety of psychiatric and neurological disorders [57].

The large-scale data collection and harmonization across diverse studies required for modern GWASs means that the phenotypes examined with GWASs are often 'minimal' compared with those examined in twin/family studies. For example, they may be as simple and crude as self-reported depression. But simulations show that even with gross measurement error, large sample sizes result in high power to detect genetic effects [34]. For example, incorporating 23andMe's self-reported depression diagnosis enabled the first GWAS hits for major depressive disorder [58].

Although this 'brute force' approach has proven effective for SNP discovery in GWAS, there are drawbacks. GWASs on 'minimal' phenotypes yield lower h^2_{SNP} and identify SNPs that have lower specificity compared with those discovered in GWASs on more strictly defined phenotypes [59].

Minimal phenotyping also means that these discoveries are restricted to traits measured in common across multiple GWAS samples or in large biobanks. In the cognitive domain, GWAS discoveries have accelerated for IQ [4–6,60], but there are few, if any, genome-wide significant associations for specific cognitive abilities like aspects of memory and executive functioning [61,62]. Similarly, GWAS of MRI phenotypes have focused on structural measures, which can be more easily standardized across sites than functional measures [8–10,12,14–16,63–65]. While such information about structure can provide important insights, the genetic (co)variation in brain function (that can be measured reliably [66–69]; see Outstanding questions) may provide insights into behavior different than those based on structure.

By contrast, twin/family studies allow for deeper phenotypes, which can be used to characterize genetic heterogeneity (e.g., within a domain or across age), providing insights into potential mechanisms. For instance, twin/family studies of structural MRI indicate distinct genetic contributions to cortical myelination, cortical thickness, and surface area [70,71]. In addition, the genetic relationship between these cortical features and cognition varies: general cognitive ability is genetically related to cortical myelination in the temporal lobe and insula, but shows weaker or nonsignificant genetic correlations with cortical thickness and surface area [70]. Moreover, twin/family studies have found that the heritabilities of these measures vary across regions. For example, genetic influences for the same region differ across the two hemispheres [71] and patterns of heritability vary markedly across and within classic resting state networks [72].

Finally, because many twin/family studies incorporate longitudinal assessments, they can provide information on how genetic influences change across age. For instance, twin/family studies have revealed that the heritability of, and genetic correlations between, structural brain measures vary across age [70,73,74]. Heritability can vary across the lifespan because genetic influences are expressed differently at different biological stages of life (e.g., with pubertal changes or in old age) or in different developmental contexts (e.g., genes related to cognitive ability may have different effects as children progress through formal schooling). Heritability can also vary across the lifespan because existing genetic influences are amplified by correlated environmental influences (e.g., individuals high in intellectual ability seek out or are chosen for additional education). A recent longitudinal twin study found evidence for a general genetic factor on cortical thickness and thinning that decreases in influence from childhood and adolescence compared with adulthood, but also evidence for new genetic influences on cortical thickness that arise particularly in the frontal cortex in adolescence [75], a developmental period in which this brain region undergoes extensive changes.

While increasing GWAS sample sizes for phenotypes derived from brain imaging [65] will greatly expand the range of questions that can be addressed with GWAS data, resources are limited. For this reason, twin studies will continue to enable analyses of deeper phenotypes due to their smaller sample size requirements for adequately powered heritability and genetic correlation analysis, greater diversity of existing phenotypes assessed, and relative ease with which new ones can be measured.

Speaking to the mutual compatibility between twin and GWAS studies, genetic distinctions revealed by twin/family studies can inform the design and interpretation of GWASs. Behavioral measures with very high genetic correlations in twin/family studies, such as IQ and educational attainment, can be ‘lumped’ to increase power in GWASs [7], whereas measures with lower genetic correlations can be ‘split’, enabling greater specificity of GWAS findings [76]. Twin/family studies are also useful for developing proxy phenotypes for use in GWAS samples. For example, before GWAS sample sizes increased for cognitive measures, educational attainment was used

as a proxy for IQ, because it was known from twin/family studies that the two shared genetic covariance [77]. Similarly, genetic correlations from family studies could suggest coarse proxies for other phenotypes of interest, such as brain activity during a particular task paradigm, that are difficult to obtain in large GWAS samples.

Twin/family data facilitate the separation of genetic from environmental influences

Perhaps ironically, genetically informative samples provide some of the strongest evidence for the importance of environmental influences [23]. Twin/family studies suggest that roughly half the variability in human traits is attributed to environmental differences [2]. Environmental influences can be shared (those that increase similarity among family members) or nonshared (those that lead family members' traits to be uncorrelated, including measurement error). For most traits, twin/family studies suggest that nonshared environmental influences are large whereas shared ones are small [78], but there are notable exceptions [23,78,79], such as shared environmental influences on childhood general cognitive ability [80], academic achievement, and educational attainment [81].

Although twin/family and GWAS designs typically assume genetic and environmental influences are independent, this assumption may sometimes be violated, especially for behavioral traits. For example, genetic variants associated with higher educational attainment are passed on to offspring where they have 'direct' genetic effects (influencing the offspring's own behavior), but these same genetic variants in the parents can influence the rearing environment in ways that facilitate educational attainment ('indirect' genetic effects through nurturing from relatives) [82]. This phenomenon is termed 'passive **gene–environment correlation**' **(*r*GE)** in the twin literature [83,84] and 'genetic nurture' in the GWAS literature [85].

Recently, the combination of twin/family-based designs with GWASs have allowed researchers to tease apart direct and indirect genetic effects. Typical GWAS estimates can be influenced by both types of effects, whereas estimates from within-family designs, in which differences in sibling phenotypes are regressed on differences in sibling genotypes, are only influenced by direct genetic effects [82,86]. Thus, indirect genetic effects, and by extension passive *r*GE, can be estimated by the degree to which between-family (typical) GWAS estimates are greater than within-family estimates. For instance, a recent study [87] used data from fraternal twins to separate PGS prediction of eight traits into between- and within-family components. Between-family r^2 estimates were significantly larger than within-family r^2 estimates for cognitive traits (IQ and educational achievement), suggesting passive *r*GE. This pattern was not observed for noncognitive traits, such as height, personality, and psychopathology. Controlling for socioeconomic status (SES) reduced most of the discrepancies in the between-family versus within-family estimates, suggesting that the between-family PGS prediction reflected indirect effects related to SES. Moreover, there is consistency across studies using GWAS and twin/family approaches: Traits showing evidence for genetic nurture in GWAS studies also tend to show evidence for shared environmental influences in twin studies [80,81], indicating that estimates from twin/family studies can be used to develop hypotheses about genetic nurture that can be directly tested with GWAS datasets that include close relatives.

Additionally, PGSs derived from GWAS can be leveraged in smaller twin/family studies to better interrogate deep behavioral and environmental phenotypes, sometimes longitudinally assessed. When considered in conjunction with PGSs, these deep phenotypes enable more nuanced tests of gene–environment interplay than is possible in GWAS studies with minimal phenotyping. For example, one study [88] found that mothers' PGSs for educational attainment predicted their children's educational attainment over and above the children's PGSs and that these effects

were mediated by environmental measures of increased cognitive stimulation and decreased household chaos. Thus, not only do family designs provide a means for estimation of direct and indirect genetic effects using PGSs, but family studies can also utilize results from GWAS to gain new insights into mechanisms of environmental mediation. These studies and others [84,89] illustrate the synergistic back-and-forth between twin/family studies and GWAS that is increasingly becoming the landscape of modern genetics.

GWAS and twin/family data facilitate causal inference in different ways

The presence of *r*GE and genetic correlations between traits clouds the interpretation of associations from observational studies. For example, substance use is associated with cognitive impairment, but this association might occur because drugs are neurotoxic, because cognitive impairment influences decisions about substance use, or because third (confounding) variables influence both cognition and the likelihood of substance use [90]. In some cases, genetically informative samples (both GWAS and twin/family studies) provide additional information that can be used to tease apart these alternatives.

Mendelian randomization (Box 4 and Figure 2B) is a popular method that uses GWAS data to estimate whether the relationship between two variables is consistent with causality [91]. However, the variables that can be examined are limited to the phenotypes that have been assessed in well-powered GWAS, which may not include variables relevant to the causal hypotheses of interest or may include only minimally assessed phenotypes.

Deeper phenotypes that have been assessed in twin studies, including longitudinal and neuroimaging studies, can also be used to evaluate causal hypotheses [90,92]. As described earlier, PGS associations within families, such as those within DZ twin pairs, are useful for evaluating whether genetic differences within families cause phenotypic differences, controlling for confounding factors that may vary between families [82,87]. Conversely, co-twin control or twin-difference analyses can also be used to test whether environmental differences within families (e.g., different exposure to trauma or differences in substance use) are associated with differences in behavioral outcomes. Because they focus on within-family effects, they control

Box 4. Causal models with genetically informative samples

Both GWAS and twin/family data enable tests of causality, though the methods leverage different information (see Figures 1B and 2B in main text). Mendelian randomization uses genetic variants identified in a GWAS as 'instrumental' variables to test hypotheses about direction of causation [91,115]. An instrumental variable directly affects the exposure or causal variable (X) but does not directly affect the outcome (Y). Because parental alleles are randomly distributed to offspring during meiosis, inheriting an allele that affects X is a form of random exposure; this SNP (or set of SNPs) is expected to have an effect on Y through X and the ratio of the SNP's regression coefficient predicting Y to its coefficient predicting X provides an estimate of the causal effect of X on Y. The validity of this approach depends on strong assumptions that the instrument is not associated with confounders and is not associated with Y except through X [116]. These assumptions are often violated, though new techniques have been developed to reduce resulting bias [117,118].

In contrast to Mendelian randomization, which focuses on specific genetic effects of the proposed causal variable, co-twin control or twin difference models (and sibling difference models) [90,92] control for shared genetic and familial environment effects to focus on nonfamilial environmental effects of the proposed causal variable. Because they share genes and familial environments, siblings, DZ twins, and particularly MZ twins who differ in X can be used to test for environmental effects of that exposure on Y, controlling for unmeasured genetic and environmental confounds. If the exposed twins are more likely to show the outcome than their nonexposed co-twins, then there is evidence for environmental influences of that exposure. By contrast, if the nonexposed twins are equally likely to show the outcome, the association may be due to shared familial risk. Comparing co-twin differences across MZ and DZ groups can also indicate whether a phenotypic association is likely to be genetically mediated: because MZ co-twins completely control for genetic influences but DZ co-twins only partially control for shared genetic influences, observing stronger associations of twin differences in X with twin differences in Y in DZ twins compared with MZ twins suggests genetic mediation. Twin data can also be used to estimate direction of causation models [90], which test whether the genetic and environmental associations between X and Y are proportionally consistent with causal paths between X and Y.

for factors that vary systematically between families but not within families, such as rearing effects, SES, and in the case of MZ twins, genetics. For example, compared with their adult MZ co-twins who did not experience stressors during development, MZ twins who had stressful experiences showed increased activation in limbic regions during a cognitive control task requiring one to ignore distracting emotional material [93]. Although this pattern does not prove a causal link between such early stressful life events and neural changes, by controlling for most confounds, it is consistent with a causal explanation.

Perhaps more importantly, these models allow hypotheses about causality to be rejected when there is evidence that confounding variables account for the link (Box 4 and Figure 1B). For instance, in contrast to evidence from longitudinal studies suggesting that cannabis use impairs cognition [94], co-twin control studies show that twins who use more or use earlier than their co-twins do not have lower cognitive ability or brain volume, suggesting no causal effect [95–98]. The inclusion of a large twin sample in the National Institutes of Health's multisite Adolescent Brain and Cognitive Development study (https://abcdstudy.org) [90] will enable similar co-twin analyses to evaluate effects of a number of environmental risk factors on brain and cognitive development.

Concluding remarks

The explosion of research enabled by inexpensive genotyping and the development of new methods using genome-wide data has ushered in a 'golden age of genetic research' [23]. While twin/family studies are an older method, they remain important, providing different and often complementary information to GWASs. Both approaches will continue to borrow methods, data, and ideas from one another to optimize scientific discovery. GWAS data is now being used to estimate heritability and genetic correlations and has borrowed structural equation model techniques from the twin/family literature [99,100] to elucidate the genetic architecture of complex traits. GWAS researchers have also begun to appreciate the importance of collecting twin/family data to better control for population stratification and to estimate passive *r*GE. By the same token, twin/family studies have begun incorporating measured genetic data, such as PGSs, to test causal hypotheses, to better interrogate gene–environment interplay, and to identify biological systems and neurocognitive mechanisms underlying observed genetic and environmental effects. The lines between traditional twin/family approaches and those that use GWAS data have blurred and will continue to do so, leading to answers to longstanding questions (see Outstanding questions) in complex trait genetics as well as cognitive neuroscience.

Outstanding questions

What explains the still-missing heritability, the gap between heritability estimated from twin/family studies and that estimated from SNPs?

Given the evidence that many genes of small effect underlie behavior, how can genetic discovery (e.g., GWAS hits clustering within particular biological pathways) be translated into new theoretical models and/or be used to guide new therapies?

Can GWAS and twin studies identify developmental periods during which specific traits (cognitive ability, symptoms of mental health disorders) show changes in genetic and environmental influences, which might have important implications for the timing of interventions?

How different are the genes that influence behavior across factors such as sex, ancestry, SES, etc.?

How can we increase diversity in genetic studies to ensure that resulting discoveries benefit all individuals in society?

Why are cognitive traits such as IQ and educational achievement more subject to shared environmental influences and indirect genetic effects than other traits like psychiatric disorders?

Some measures, such as task performance and fMRI activation or connectivity, tend to show lower reliability than questionnaire or structural MRI measures. What approaches (e.g., task optimization for individual differences, longer assessments/scans, multiband scanning, analysis of multiple measures with mixed models or structural equation models to remove noise) may allow both twin/family and GWAS to best investigate individual differences in these important phenotypes?

Acknowledgments

N.P.F. was supported by National Institutes of Health (NIH) grants MH063207, DA046064, DA046413, DA042742, and DA051018. M.T.B. was supported by NIH grants DA041120 and DA051018. M.C.K. was supported by NIH grant MH100141.

Declaration of interests

We have no conflicts of interest to disclose.

References

1. Turkheimer, E. (2000) Three laws of behavior genetics and what they mean. *Curr. Dir. Psychol. Sci.* 9, 160–164
2. Polderman, T.J.C. *et al.* (2015) Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* 47, 702–709
3. Maier, R.M. *et al.* (2018) Embracing polygenicity: a review of methods and tools for psychiatric genetics research. *Psychol. Med.* 48, 1055–1067
4. Savage, J.E. *et al.* (2018) Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* 50, 912–919
5. Sniekers, S. *et al.* (2017) Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat. Genet.* 49, 1107–1112
6. Davies, G. *et al.* (2018) Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat. Commun.* 9, 2098
7. Hill, W.D. *et al.* (2019) A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Mol. Psychiatry* 24, 169–181
8. Satizabal, C.L. *et al.* (2019) Genetic architecture of subcortical brain structures in 38,851 individuals. *Nat. Genet.* 51, 1624–1636

9. Adams, H.H.H. *et al.* (2016) Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat. Neurosci.* 19, 1569–1582
10. Elliott, L.T. *et al.* (2018) Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* 562, 210–216
11. Hofer, E. *et al.* (2020) Genetic correlations and genome-wide associations of cortical structure in general population samples of 22,824 adults. *Nat. Commun.* 11, 4796
12. Hibar, D.P. *et al.* (2017) Novel genetic loci associated with hippocampal volume. *Nat. Commun.* 8, 13624
13. Jawinski, P. *et al.* (2019) Human brain arousal in the resting state: a genome-wide association study. *Mol. Psychiatry* 24, 1599–1609
14. Zhao, B. *et al.* (2019) Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nat. Genet.* 51, 1637–1644
15. Jansen, P.R. *et al.* (2020) Genome-wide meta-analysis of brain volume identifies genomic loci and genes shared with intelligence. *Nat. Commun.* 11, 5606
16. Grasby, K.L. *et al.* (2020) The genetic architecture of the human cerebral cortex. *Science* 367, 1340
17. Lee, J.J. *et al.* (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* 50, 1112–1121
18. Buniello, A. *et al.* (2019) The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 47, D1005–D1012
19. Visscher, P.M. *et al.* (2017) 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* 101, 5–22
20. Chabris, C.F. *et al.* (2015) The fourth law of behavior genetics. *Curr. Dir. Psychol. Sci.* 24, 304–312
21. Zhu, Z. *et al.* (2015) Dominance genetic variation contributes little to the missing heritability for human complex traits. *Am. J. Hum. Genet.* 96, 377–385
22. Wray, N.R. *et al.* (2014) Research review: polygenic methods and their application to psychiatric traits. *J. Child Psychol. Psychiatry* 55, 1068–1087
23. Harden, K.P. (2021) "Reports of my death were greatly exaggerated": behavior genetics in the postgenomic era. *Annu. Rev. Psychol.* 72, 37–60
24. Yang, J. *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42, 565–569
25. Bulik-Sullivan, B.K. *et al.* (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* 47, 291–295
26. Lloyd-Jones, L.R. *et al.* (2019) Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat. Commun.* 10, 5086
27. Wray, N.R. *et al.* (2020) From basic science to clinical application of polygenic risk scores: a primer. *JAMA Psychiatry* 78, 101–109
28. Martin, A.R. *et al.* (2019) Predicting polygenic risk of psychiatric disorders. *Biol. Psychiatry* 86, 97–109
29. Allegrini, A.G. *et al.* (2019) Genomic prediction of cognitive traits in childhood and adolescence. *Mol. Psychiatry* 24, 819–827
30. Peyrot, W.J. *et al.* (2014) Effect of polygenic risk scores on depression in childhood trauma. *Br. J. Psychiatry* 205, 113–119
31. Colodro-Conde, L. *et al.* (2018) A direct test of the diathesis–stress model for depression. *Mol. Psychiatry* 23, 1590–1596
32. Duncan, L.E. and Keller, M.C. (2011) A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *Am. J. Psychiatry* 168, 1041–1049
33. Duncan, L.E. *et al.* (2019) How genome-wide association studies (GWAS) made traditional candidate gene studies obsolete. *Neuropsychopharmacology* 44, 1518–1523
34. Border, R. *et al.* (2019) No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. *Am. J. Psychiatry* 176, 376–387
35. Bogdan, R. *et al.* (2018) Polygenic risk scores in clinical psychology: bridging genomic risk to individual differences. *Annu. Rev. Clin. Psychol.* 14, 119–157
36. Riglin, L. *et al.* (2017) Schizophrenia risk alleles and neurodevelopmental outcomes in childhood: a population-based cohort study. *Lancet Psychiatry* 4, 57–62
37. Alemany, S. *et al.* (2019) Common polygenic variations for psychiatric disorders and cognition in relation to brain morphology in the general pediatric population. *J. Am. Acad. Child Adolesc. Psychiatry* 58, 600–607
38. Dudbridge, F. (2013) Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* 9, e1003348
39. Torkamani, A. *et al.* (2018) The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* 19, 581–590
40. Keller, M.C. *et al.* (2013) The genetic correlation between height and IQ: shared genes or assortative mating? *PLoS Genet.* 9, e1003451
41. Rhee, S.H. *et al.* (2016) An examination of the developmental propensity model of conduct problems. *J. Abnorm. Psychol.* 125, 550–564
42. Feng, J. *et al.* (2020) Partitioning heritability analyses unveil the genetic architecture of human brain multidimensional functional connectivity patterns. *Hum. Brain Mapp.* 41, 3305–3317
43. Brown, B.C. *et al.* (2016) Transethnic genetic-correlation estimates from summary statistics. *Am. J. Hum. Genet.* 99, 76–88
44. Cross-Disorder Group of the Psychiatric Genomics Consortium (2019) Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell* 179, 1469–1482
45. Verhulst, B. (2017) A power calculator for the classical twin design. *Behav. Genet.* 47, 255–261
46. Visscher, P.M. *et al.* (2014) Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet.* 10, e1004269
47. Keller, M.C. *et al.* (2010) Are extended twin family designs worth the trouble? A comparison of the bias, precision, and accuracy of parameters estimated in four twin family models. *Behav. Genet.* 40, 377–393
48. Evans, L.M. *et al.* (2018) Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nat. Genet.* 50, 737–745
49. Charney, E. (2012) Behavior genetics and postgenomics. *Behav. Brain Sci.* 35, 331–358
50. Miller, M.B. *et al.* (2012) Assumptions in studies of heritability and genotype–phenotype association. *Behav. Brain Sci.* 35, 372–373
51. Felson, J. (2014) What can we learn from twin studies? A comprehensive evaluation of the equal environments assumption. *Soc. Sci. Res.* 43, 184–199
52. Duncan, L. *et al.* (2019) Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* 10, 3328
53. Martin, A.R. *et al.* (2019) Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* 51, 584–591
54. Martin, A.R. *et al.* (2017) Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* 100, 635–649
55. Bycroft, C. *et al.* (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–209
56. Trampush, J.W. *et al.* (2017) GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium. *Mol. Psychiatry* 22, 336–345
57. Thompson, P.M. *et al.* (2020) ENIGMA and global neuroscience: a decade of large-scale studies of the brain in health and disease across more than 40 countries. *Transl. Psychiatry* 10, 100
58. Hyde, C.L. *et al.* (2016) Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* 48, 1031–1036
59. Cai, N. *et al.* (2020) Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nat. Genet.* 52, 437–447

60. Lam, M. *et al.* (2017) Large-scale cognitive GWAS meta-analysis reveals tissue-specific neural expression and potential nootropic drug targets. *Cell Rep.* 21, 2597–2613
61. Davies, G. *et al.* (2016) Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N =112 151). *Mol. Psychiatry* 21, 758–767
62. Ibrahim-Verbaas, C.A. *et al.* (2016) GWAS for executive function and processing speed suggests involvement of the CADM2 gene. *Mol. Psychiatry* 21, 189–197
63. Zhao, B. *et al.* (2019) Heritability of regional brain volumes in large-scale neuroimaging and genetic studies. *Cereb. Cortex* 29, 2904–2914
64. Hibar, D.P. *et al.* (2015) Common genetic variants influence human subcortical brain structures. *Nature* 520, 224–229
65. Smith, S.M. *et al.* (2021) An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nat. Neurosci.* 24, 737–745
66. Elliott, M.L. *et al.* (2020) What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. *Psychol. Sci.* 31, 792–806
67. Herting, M.M. *et al.* (2018) Test-retest reliability of longitudinal task-based fMRI: Implications for developmental studies. *Dev. Cogn. Neurosci.* 33, 17–26
68. Ge, T. *et al.* (2017) Heritability analysis with repeat measurements and its application to resting-state functional connectivity. *Proc. Natl. Acad. Sci. U. S. A.* 114, 5521–5526
69. Elliott, M.L. *et al.* (2019) General functional connectivity: shared features of resting-state and task fMRI drive reliable and heritable individual differences in functional brain networks. *NeuroImage* 189, 516–532
70. Schmitt, J.E. *et al.* (2020) The genetics of cortical myelination in young adults and its relationships to cerebral surface area, cortical thickness, and intelligence: a magnetic resonance imaging study of twins and families. *NeuroImage* 206, 116319
71. Strike, L.T. *et al.* (2019) Genetic complexity of cortical structure: differences in genetic and environmental factors influencing cortical surface area and thickness. *Cereb. Cortex* 29, 952–962
72. Reineberg, A.E. *et al.* (2020) Genetic and environmental influence on the human functional connectome. *Cereb. Cortex* 30, 2099–2113
73. Schmitt, J.E. *et al.* (2019) A comprehensive quantitative genetic analysis of cerebral surface area in youth. *J. Neurosci.* 39, 3028–3040
74. Schmitt, J.E. *et al.* (2018) The genetic contributions to maturational coupling in the human cerebrum: a longitudinal pediatric twin imaging study. *Cereb. Cortex* 28, 3184–3191
75. Teeuw, J. *et al.* (2019) Genetic influences on the development of cerebral cortical thickness during childhood and adolescence in a Dutch longitudinal twin sample: the Brainscale Study. *Cereb. Cortex* 29, 978–993
76. Hatoum, A.S. *et al.* (2020) Genome-wide association study of over 427,000 individuals establishes executive functioning as a neurocognitive basis of psychiatric disorders influenced by GABAergic processes. *bioRxiv* Published online September 2, 2020. http://doi/org/10.1101/674515
77. Rietveld, C.A. *et al.* (2014) Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc. Natl. Acad. Sci. U. S. A.* 111, 13790–13794
78. Plomin, R. *et al.* (2016) Top 10 replicated findings from behavioral genetics. *Perspect. Psychol. Sci.* 11, 3–23
79. Engelhardt, L.E. *et al.* (2019) Accounting for the shared environment in cognitive abilities and academic achievement with measured socioecological contexts. *Dev. Sci.* 22, e12699
80. Briley, D.A. and Tucker-Drob, E.M. (2013) Explaining the increasing heritability of cognitive ability across development: a meta-analysis of longitudinal twin and adoption studies. *Psychol. Sci.* 24, 1704–1713
81. Branigan, A.R. *et al.* (2013) Variation in the heritability of educational attainment: an international meta-analysis. *Soc. Forces* 92, 109–140
82. Young, A.I. *et al.* (2019) Deconstructing the sources of genotype-phenotype associations in humans. *Science* 365, 1396–1400
83. Plomin, R. (2014) Genotype-environment correlation in the era of DNA. *Behav. Genet.* 44, 629–638
84. Cheesman, R. *et al.* (2020) Comparison of adopted and nonadopted individuals reveals gene–environment interplay for education in the UK biobank. *Psychol. Sci.* 31, 582–591
85. Kong, A. *et al.* (2018) The nature of nurture: effects of parental genotypes. *Science* 359, 424–428
86. Young, A.I. *et al.* (2018) Relatedness disequilibrium regression estimates heritability without environmental bias. *Nat. Genet.* 50, 1304–1310
87. Selzam, S. *et al.* (2019) Comparing within- and between-family polygenic score prediction. *Am. J. Hum. Genet.* 105, 351–363
88. Wertz, J. *et al.* (2020) Using DNA from mothers and children to study parental investment in children's educational attainment. *Child Dev.* 91, 1745–1761
89. Bates, T.C. *et al.* (2018) The nature of nurture: using a virtual-parent design to test parenting effects on children's educational attainment in genotyped families. *Twin Res. Hum. Genet.* 21, 73–83
90. Iacono, W.G. *et al.* (2018) The utility of twins in developmental cognitive neuroscience research: how twins strengthen the ABCD research design. *Dev. Cogn. Neurosci.* 32, 30–42
91. Hemani, G. *et al.* (2018) The MR-Base platform supports systematic causal inference across the human phenome. *eLife* 7, e34408
92. McGue, M. *et al.* (2010) Causal inference and observational research: the utility of twins. *Perspect. Psychol. Sci.* 5, 546–556
93. Godinez, D.A. *et al.* (2016) Differences in frontal and limbic brain activation in a small sample of monozygotic twin pairs discordant for severe stressful life events. *Neurobiol. Stress* 5, 26–36
94. Meier, M.H. *et al.* (2012) Persistent cannabis users show neuropsychological decline from childhood to midlife. *Proc. Natl. Acad. Sci. U. S. A.* 109, E2657–E2664
95. Jackson, N.J. *et al.* (2016) Impact of adolescent marijuana use on intelligence: results from two longitudinal twin studies. *Proc. Natl. Acad. Sci. U. S. A.* 113, E500–E508
96. Ross, J.M. *et al.* (2020) Investigating the causal effect of cannabis use on cognitive function with a quasi-experimental co-twin design. *Drug Alcohol Depend.* 206, 107712
97. Meier, M.H. *et al.* (2018) Associations between adolescent cannabis use and neuropsychological decline: a longitudinal co-twin control study. *Addiction* 113, 257–265
98. Pagliaccio, D. *et al.* (2015) Shared predisposition in the association between cannabis use and subcortical brain structure. *JAMA Psychiatry* 72, 994
99. Grotzinger, A.D. *et al.* (2019) Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* 3, 513–525
100. Balbona, J.V. *et al.* (2021) Estimation of parental effects using polygenic scores. *Behav. Genet.* 51, 264–278
101. Tucker-Drob, E.M. and Bates, T.C. (2016) Large cross-national differences in gene × socioeconomic status interaction on intelligence. *Psychol. Sci.* 27, 138–149
102. Papageorge, N.W. and Thom, K. (2020) Genes, education, and labor market outcomes: evidence from the Health and Retirement Study. *J. Eur. Econ. Assoc.* 18, 1351–1399
103. Crusio, W.E. (2012) Heritability estimates in behavior genetics: wasn't that station passed long ago? *Behav. Brain Sci.* 35, 361–362
104. Visscher, P.M. *et al.* (2008) Heritability in the genomics era — concepts and misconceptions. *Nat. Rev. Genet.* 9, 255–266
105. Durbin, R.M. *et al.* (2010) A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073
106. Mullins, N. *et al.* (2021) Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* 53, 817–829
107. Tam, V. *et al.* (2019) Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* 20, 467–484
108. Pe'er, I. *et al.* (2008) Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* 32, 381–385
109. Dick, D.M. *et al.* (2015) Candidate gene–environment interaction research: reflections and recommendations. *Perspect. Psychol. Sci.* 10, 37–59
110. Johnson, E.C. *et al.* (2017) No evidence that schizophrenia candidate genes are more associated with schizophrenia than noncandidate genes. *Biol. Psychiatry* 82, 702–708

111. Farrell, M.S. *et al.* (2015) Evaluating historical candidate genes for schizophrenia. *Mol. Psychiatry* 20, 555–562
112. Culverhouse, R.C. *et al.* (2018) Collaborative meta-analysis finds no evidence of a strong interaction between stress and 5-HTTLPR genotype contributing to the development of depression. *Mol. Psychiatry* 23, 133–142
113. Border, R. *et al.* (2019) Measurement error cannot account for failed replications of historic candidate gene-by-environment hypotheses: response to Vrshek-Schallhorn *et al*. *Am. J. Psychiatry* 176, 668–669
114. Hewitt, J.K. (2012) Editorial policy on candidate gene association and candidate gene-by-environment interaction studies of complex traits. *Behav. Genet.* 42, 1–2
115. Zhu, Z. *et al.* (2018) Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* 9, 224
116. Evans, D.M. and Davey Smith, G. (2015) Mendelian randomization: new applications in the coming age of hypothesis-free causality. *Annu. Rev. Genomics Hum. Genet.* 16, 327–350
117. Brumpton, B. *et al.* (2020) Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nat. Commun.* 11, 3519
118. Verbanck, M. *et al.* (2018) Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* 50, 693–698