Author for correspondence:
Eleftheria Zeggini
e-mail: eleftheria@sanger.ac.uk

Statistical methods to detect pleiotropy in human complex traits

Sophie Hackinger and Eleftheria Zeggini

Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK

In recent years pleiotropy, the phenomenon of one genetic locus influencing several traits, has become a widely researched field in human genetics. With the increasing availability of genome-wide association study summary statistics, as well as the establishment of deeply phenotyped sample collections, it is now possible to systematically assess the genetic overlap between multiple traits and diseases. In addition to increasing power to detect associated variants, multi-trait methods can also aid our understanding of how different disorders are aetiologically linked by highlighting relevant biological pathways. A plethora of available tools to perform such analyses exists, each with their own advantages and limitations. In this review, we outline some of the currently available methods to conduct multi-trait analyses. First, we briefly introduce the concept of pleiotropy and outline the current landscape of pleiotropy research in human genetics; second, we describe analytical considerations and analysis methods; finally, we discuss future directions for the field.

1. Introduction

The field of human complex trait genetics aims to elucidate how genetic variation affects differences in phenotypes. Most complex phenotypes are highly polygenic (i.e. they are influenced by a large number of genetic variants with moderate effects, rather than a handful of variants with large effects [1]). Since their inception in the early 2000s [2–4], genome-wide association studies (GWAS) have become the tool of choice for complex trait analysis. In the classical GWAS approach, the association of genetic variants across the entire genome with a single phenotype of interest is tested in a group of individuals. Recent years have seen a shift towards the joint analysis of related phenotypes. As a consequence of active method development in this field there are now a number of statistical tools available to detect cross-phenotype genetic associations. This review is intended to provide an overview of available methods and their relative strengths and limitations.

1.1. Types of pleiotropy

The term 'pleiotropy' was coined over 100 years ago by German scientist Ludwig Plate to describe the phenomenon of a hereditary unit affecting more than one trait of an organism [5]. Since then, pleiotropy has been a topic of extensive research and debate. Before human genetics began to gain traction, pleiotropy was mainly studied in model organisms and, on a more theoretical level, in evolutionary biology [5,6]. Over the course of the past decades there have been several proposals on how to classify different types of pleiotropy [5,7–9]. With regards to GWAS, it is important to note that cross-phenotype associations can arise due to several reasons, not all of which are biologically meaningful [7,8]. Solovieff and colleagues [8] described three broad categories of pleiotropy in the context of complex traits:

In the case of biological pleiotropy, causal variants of different traits fall into the same gene or regulatory unit (e.g. transcription factor binding sites) [8]. In GWAS this could manifest itself in the form of two different variants in the

same region tagging the same or two separate causal variants, or as one variant tagging the causal one (figure 1a,b). In practice, fine-mapping and molecular studies are required to confidently distinguish between these different scenarios [8].

Mediated pleiotropy refers to the case where a variant directly affects one trait, which in turn affects another (figure 1c). GWAS will still pick up an association of the variant with the second trait, but this association will disappear when conditioned on the first. Causal inference can be achieved through Mendelian randomization studies, which have been widely used in genetic epidemiology [8,10–12]. An example is the association of the *FTO* gene with osteoarthritis (OA) [13], which was shown to exert its effect on OA through body mass index (BMI) [14].

Finally, cross-phenotype associations can also arise due to spurious pleiotropy. At the planning stage of a study, design artefacts may lead to inaccurate results. For example, ascertainment bias or misclassification of cases can both inflate genetic overlap estimates. At the analysis stage, causal variants in different genes may be tagged by the same GWAS variant (figure 1d). A classic example of this is the human leucocyte antigen (HLA) region on chromosome 6. Due to its high gene density and extensive linkage disequilibrium (LD), GWAS signals within the HLA region are difficult to map finely. While the HLA locus has been associated with a range of diseases [13,15–20], most prominently immune-mediated ones, it remains unclear to what extent these disorders share the same causal risk variants or genes.

1.2. Pleiotropy in human complex traits

In 2011, a systematic evaluation of associations reported in the NIHGR GWAS catalogue found that 4.6% of variants were associated with more than one trait [21]. This number is likely to have grown, as GWAS signals have been continuously added to the database.

Many cross-phenotype effects are not surprising. For example, variants in the *DSP* gene are associated with chronic obstructive pulmonary disease, as well as pulmonary fibrosis and lung function traits [22]. Others are perhaps less intuitive and can shed light into hitherto unknown connections between traits. For example, variants in the *ASTN2* gene have been shown to affect both risk to osteoarthritis [13] and migraine [23,24]. These seemingly unrelated diseases might share pathways involved in pain perception.

Until a few years ago, the focus of many consortial efforts was to combine datasets of one phenotype for large-scale GWAS and meta-analyses [15,25,26]. For many traits, results from these studies are now publicly available, providing an excellent resource for cross-phenotype analyses using summary statistics. As the appreciation of pleiotropic effects has gained traction in the scientific community, cross-disorder analyses of several related traits have been increasingly carried out to disentangle shared and disease-specific genetic determinants [27–30].

The establishment of genome-wide genotyped biobanks [31] and cohorts with in-depth phenotype information [32] has also made it possible to perform multi-trait analyses on the same sample set [12,33], for example through phenome-wide association studies (PheWAS), where the association of each genetic variant with all phenotypes in a dataset is tested [34–37].

One challenge of the PheWAS approach is the high multiple-testing burden that grows as the number of traits and variants tested increases [9]. Although this can be partly circumvented by performing targeted PheWAS at a selected number of variants hypothesized to exert pleiotropic effects [9,38], other challenges such as consistent phenotyping and selection of appropriate covariates remain [34].

Investigating pleiotropy in human traits not only holds the potential to uncover additional associations, but could additionally help to redefine disease classifications. This is of particular interest in disorders for which the aetiopathology is unclear, and for which current diagnostic tools might be inadequate. For example, psychiatric conditions are highly comorbid, and until recently [15,39] have been mostly refractory to GWAS [40]. Comparisons of different psychiatric disorders have shown that the genetic overlap among them is extensive [27,41], and that certain pairs of diseases are genetically more similar than others [41]. Together these findings suggest that shared biological mechanisms cross diagnostic boundaries, and might aid the development of more accurate disease classification systems.

2. Analytical approaches

2.1. Study design considerations

There are some practical considerations to be taken into account when selecting an appropriate method for multi-trait analysis.

First, the type of data available will determine which statistical approach is applicable. Due to limitations of data sharing policies it might not be possible to obtain individual-level genotype data for all traits analysed.

Second, the type and number of traits to include must be considered: some approaches require all traits to be continuous, while others also allow for dichotomous traits or a combination of both. Several methods, such as colocalization tests [42,43] or genetic correlation analyses [44], can currently only accommodate two traits at a time, while others lose power with an increasing number of traits [45].

Finally, if each trait is measured on a different set of individuals, sample overlap between each dataset will need to be accounted for. This has been implemented in several methods [43,44,46]. Ideally, the exact number of overlapping individuals will need to be accounted for. However, this is often not possible when using data from publicly available GWAS. One way to estimate the extent of overlap is to calculate the Pearson's correlation of the Z-scores of all independent, non-associated variants from two studies [43], although other methods have also been proposed [47–49].

2.2. Overview of methods

Pleiotropy analyses can be broadly classified into three categories according to the level at which they assess genetic overlap: genome-wide, regional and single variant.

Genome-wide methods are currently only available for pairwise trait comparisons, and can be used as an initial assessment of the global genetic overlap between two traits.

The latter two approaches aim to detect cross-phenotype effects at distinct genomic regions and at a single variant, respectively.
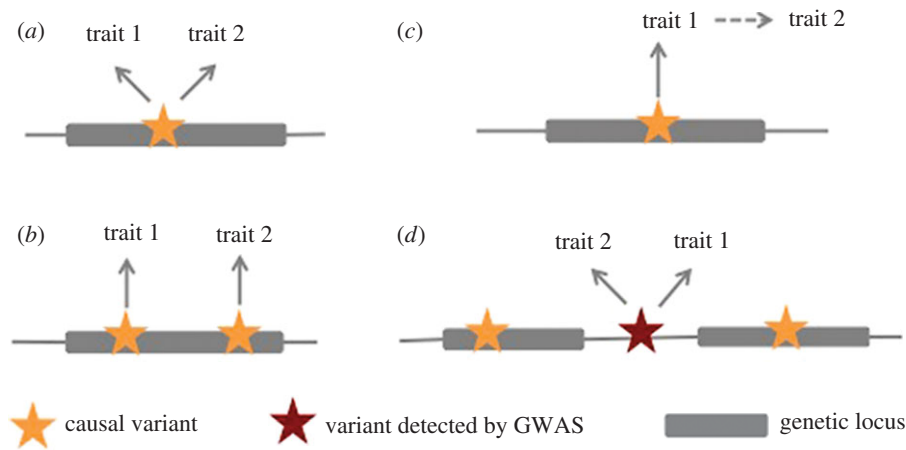
Figure 1. Schematic representation of different scenarios for cross-phenotype associations. Such effects might arise due to biological pleiotropy, whereby causal variants for two traits colocalize in the same locus (*a*,*b*), due to mediated pleiotropy, whereby a variant exerts an effect on one trait through another one (*c*), or due to spurious pleiotropy, whereby causal variants for two traits fall into distinct loci but are in LD with a variant associated with both traits (*d*).

Region-based methods bin variants into groups based on pre-defined criteria, such as LD-blocks or gene boundaries, and then test for cross-phenotype effects within each group. An advantage of such approaches is that they alleviate the multiple testing penalty incurred by single-point analyses; furthermore, they can increase power by combining information across biologically meaningful units.

Since variant-level methods test each variant separately, they provide the highest resolution. On the other hand, they are less powerful in situations where each trait is associated with a different variant in the same functional unit, and might fail to identify these cross-phenotype effects unless all relevant variants are in at least moderate LD.

The above analysis approaches can be further sub-divided based on their underlying statistical framework into univariate and multivariate. Univariate methods combine summary statistics of single-trait GWAS to search for cross-phenotype effects. This means that analyses can be carried out with each trait measured on a distinct set of individuals. Multivariate methods, on the other hand, jointly model all traits in a statistical framework, which requires that all individuals included in the study have phenotype information for all traits analysed. The statistical difference between uni- and multivariate methods is best illustrated by the example of linear regression analysis: for univariate regression, the response variable (i.e. the phenotype) will be a vector, with one data point for each individual in the study; for multivariate regression, the response variable will be a matrix, where each row represents an individual and each column represents one phenotype. Although there are exceptions, these categories are often analogous to distinguishing between methods requiring only summary data and individual-level information, respectively.

2.3. Genome-wide methods

Polygenic risk scores (PRS; or genetic risk scores) were initially used in genetic epidemiology to test how well a set of variables could predict, or distinguish between, case-control status in a study sample [50–53]. In the context of GWAS, the risk variables comprise variants known to be associated with a given trait. Odds ratios (ORs) for these variants from a 'base' GWAS are then used to construct scores for each individual in an independent 'target' dataset.

Using logistic (binary trait) or linear (continuous trait) regression to relate phenotype and score, the proportion of phenotypic variance explained in the target data by the base risk variants can be directly estimated.

This framework can also be applied to two different traits [27,36,54,55]. Using this approach, Purcell and colleagues [54] showed that risk scores for bipolar disorder are significantly associated with schizophrenia, and that the variance in phenotype captured by the risk variants could be increased by relaxing the *p*-value threshold for variant inclusion (rather than using only genome-wide significant variants). One reason for this could be that many variants with a true effect on the phenotype did not reach genome-wide significance in the base study. This is especially likely for highly polygenic traits, for which only a fraction of the heritability can be explained by currently known risk variants.

Genetic correlation (rg) captures the extent to which genetic factors influence the covariance of two traits. Multivariate methods for genetic correlation analysis include GCTA [28,56], BOLT-REML [57] and mvLMM [58]. GCTA and BOLT-REML use restricted maximum-likelihood estimation to compute rg between two traits of any type (i.e. two binary, two continuous, or one binary and one continuous), while the mvLMM algorithm is more similar to GEMMA and can only accommodate normally distributed traits. Individual-level genotype data are required as input, and samples need to have phenotype values for both traits analysed. For disease traits, overlap between the cases and controls for each trait should be negligible. While all three methods use similar algorithms, BOLT-REML and mvLMM are more efficient than GCTA in terms of run time and memory usage [57,58].

More recently, a univariate method for genetic correlation analysis, cross-trait LD score regression (LDSC), requiring only summary statistics, has been developed [44,59]. Like its multivariate counterparts, LDSC can handle any combination of traits and can adjust for sample overlap. The method requires the use of a reference panel for LD estimation. This is of particular importance when analysing distinct GWAS performed on populations of different ancestries. The LD Hub database, which acts as both a central aggregation of public summary statistics and an online interface for LDSC, enables systematic comparisons between a range of traits [60]. As the authors of LDSC point out, it is important to distinguish genetic correlation from pleiotropy [44]. A near-zero

estimate of genetic correlation between two traits does not necessarily mean that they share no common risk loci. For example, there could be no directionality to their genetic relationship (i.e. at some shared loci the risk allele is the same for both traits, while at others the risk allele for one trait is protective of the other). An example of the latter scenario is the rs7501939 variant in *TCF2*, for which the C allele confers increased risk for prostate cancer and decreased risk for type 2 diabetes [61]. As for PRS, if either or both of the input datasets are underpowered, this could also lead to a falsely low estimate of r_g. Conversely, in the case of disease traits, genetic correlation could be inflated due to ascertainment bias or misclassification of cases [8].

2.4. Regional methods

The pleiotropic region identification method (PRIMe) [62] defines regions by designating the most strongly associated variant among the included traits as a 'driver', and all variants in LD with it as 'passengers'. The next most strongly associated variant is then considered: if it has been already assigned as a passenger, it is skipped; otherwise it is designated as a driver, and so on. This process is repeated iteratively until the genome is divided into non-overlapping blocks consisting of one driver and zero or several passengers. The pleiotropy index of each region is defined as the number of traits with association *p*-values below some threshold in that region. The statistical significance of the index can be calculated using two separate approaches for traits measured on overlapping or non-overlapping datasets, respectively.

In 2013 Giambartolomei and colleagues described a Bayesian colocalization model to identify genomic regions of colocalizing expression quantitative trait loci (eQTL) and GWAS signals [42]. This method was then extended to account for sample overlap, and implemented in a software package (GWAS-PW) to enable simplicity of use for the pairwise comparison of GWAS summary statistics [43]. The model integrates the effects of all variants in a pre-defined region, such as approximately independent LD blocks [63]. It generates posterior probabilities for each of five hypotheses, the two most relevant being that in a given region the traits share one causal variant, and that they each have a separate causal variant. An advantage of this approach over many variant-level methods is that it evaluates the evidence for *both* traits being associated with a given region, thus making it possible to distinguish from the scenario of one trait driving an observed signal.

Multivariate methods for locus-based analysis also exist: Tang & Ferreira [64] extended a previously developed method using canonical correlation analysis (CCA; see §2.6) to perform association analysis between two traits and sets of genetic markers, for example within genes. Further development of the method also allowed for multiple gene-multiple phenotype tests [65]. A recently published method, metaCCA, performs CCA on summary data from single-trait GWAS at both a variant- and locus-level (see §2.5) [66]. The use of functional linear models for the analysis of multiple variants and multiple traits has also been proposed [67]. A limitation of this and of CCA-based approaches is that they are not applicable to disease or non-normally distributed traits. This was overcome by Lutz and colleagues [68] by using a permutation-based approach that accommodates any combination of traits (see §2.6). Finally, mtSET uses a multivariate mixed

model with two variance components to account both for inter-individual variation (e.g. due to population structure or relatedness) and for variation among the variants being tested together [69].

2.4.1. Rare variant tests

The substantial drop in sequencing costs over the past decade together with the establishment of better reference panels for imputation have made association studies of low frequency and rare variants feasible [70,71]. Methods for rare-variant studies usually group several variants together and perform an association test with this composite genotype. They are generally more powerful than testing individual rare variants [72], and have been the tool of choice for single-trait studies [73]. Two of the most popular burden test methods are kernel-based tests (such as SKAT [74]) and collapsing tests [75]. These and other approaches have been reviewed in detail elsewhere [73].

While some of the multi-trait methods described above are applicable to both common and low-frequency markers [67–69], approaches have also been specifically designed for rare variants. The methods described below all rely on individual-level data with phenotypes measured in the same set of individuals.

Wu and Pankow extended univariate SKAT for the application to multiple continuous traits [76]. Another method, MAAUSS, also builds on the SKAT algorithm, including a variance-covariance matrix that allows for the joint modelling of multiple phenotypes [77]. Multiple binary or a mixture of binary and continuous traits can be analysed by MAAUSS through integration of the generalized estimating equation framework (see also §2.6).

In adaptive weighting reverse regression (AWRR) [78], the genotypes in a set of variants are first combined, weighted by the strength of association and direction of effect of each variant; the resulting variable is then regressed on multiple traits and a score test used to assess significance. This reverse regression approach is similar to other methods discussed here (see §2.6), and can incorporate large numbers of traits of any kind.

2.5. Single-point univariate methods

With the increasing availability of summary data from large-scale GWAS, an important question has been how to harness these data to perform pleiotropy analyses. Perhaps the simplest way to search for cross-phenotype effects is to decide on a *p*-value threshold and declare all variants that fall below this threshold for a group of traits as cross-phenotype associations [8]. However, this approach can be underpowered, as even with large sample sizes truly associated variants with sub-threshold *p*-values will be missed. Consequently, a number of methods to statistically combine summary data for multiple traits have been developed (table 1).

2.5.1. Extensions to meta-analysis

In the classical meta-analysis approach, *p*-values or effect sizes are combined across multiple studies of the same trait [89]. For the latter, effects are typically either assumed to be consistent across studies (fixed effects meta-analysis) or allowed to vary (random effects meta-analysis). However, a

Table 1. Univariate methods for single-point association analysis and variant prioritization. impl., implementation.

method	ref.	PMID	year	data	n traits	trait type	impl.
CPMA	[79]	21852963	2011	p-values	>2	any	R
ASSET	[45]	22560090	2012	betas, SEs	≥2	any	R
CPASSOC	[80]	25500260	2015	Z-scores	≥2	any	R
MultiMeta	[54]	25908790	2015	betas, SEs	≥2	any	R
MTAG	[81]	NA	2017	betas, SEs	≥2	any	Python
cFDR	[82]	25658688	2015	p-values	2	any	R
Bayesian overlap	[83]	26411566	2015	p-values	2	any	NA
metaCCA	[66]	27153689	2016	betas, SEs	≥2	any	R
GPA	[84]	25393678	2014	p-values	2	any	R
GPA-MDS	[85]	27868058	2016	p-values	≥2	any	R
fastPAINTOR	[86]	27663501	2017	Z-scores	≥2	any	C++
EPS	[87]	27153687	2016	p-values	2	any	Matlab
RiVIERA-MT	[88]	NA	2016	p-values, betas, SEs	≥2	any	R

genetic variant might have the opposite effect on two traits. While this can be circumvented by applying a directionality-agnostic p-value-based meta-analysis, there are some limitations, such as the inability to obtain an overall effect estimate [89]. Therefore, these standard approaches are best-suited to groups of traits/disorders assumed to have similar underlying biological mechanisms [27]. The meta-analysis framework has been adapted to accommodate this and other issues that arise when combining several different traits.

Cotsapas and colleagues developed a cross-phenotype meta-analysis (CPMA) method that tests for the presence of two or more trait associations at a variant [79]. This has the advantage of protecting against the scenario of one trait driving the association. CPMA only requires p-values as input and is thus robust to heterogeneous effect directions. Since CPMA compares the distribution of p-values for all traits at a variant to the null hypothesis of uniformity, it is well suited for moderate to large numbers of phenotypes, but less so for pairs of traits.

In a generalization of fixed-effects meta-analysis, all possible subsets of traits are evaluated to identify the one with the maximum absolute Z-statistic at a variant [45]. The approach, termed ASSET, takes effect estimates as input, and can also be used to identify disease subtypes within case-control data. Extensions were also proposed to account for sample overlap and effect heterogeneity between traits [45]. The method expects all traits to be of the same type, and the number of tests performed grows exponentially with the number of traits, decreasing power. Using ASSET, investigators have identified three loci associated with five autoimmune disorders, as well as risk loci associated with different cancers [90].

Zhu and colleagues developed two meta-analysis test statistics to detect cross-phenotype associations assuming homogeneous and heterogeneous effects across studies, respectively [80]. The tests are implemented in the R package CPASSOC, and work with both univariate (i.e. one trait per cohort) and multivariate summary statistics (i.e. several traits measured in each cohort). CPASSOC requires the specification of an inter-cohort correlation matrix. Since the true phenotypic correlation is unknown in the absence of raw data, this can be derived from summary statistics and—similarly to approaches outlined above—accounts for overlapping samples. Applying CPASSOC to anthropometric trait summary data from the GIANT consortium identified one novel genome-wide significant locus within the *TOX* gene missed by conventional meta-analysis [91].

A recently published method, MultiMeta, enables the joint analysis of summary statistics obtained from multivariate GWAS, such as multivariate LMM analysis (see §2.6) [92]. MultiMeta generalizes single-trait inverse variance weighted meta-analysis to allow each variant to have a vector of effect estimates (one for each trait included).

The Multi-Trait Analysis of GWAS (MTAG) tool takes a slightly different approach to the above-mentioned methods: in a generalization of inverse variance weighted meta-analysis, it incorporates effect estimates from multiple traits and outputs adjusted effect estimates for *each* trait separately [81]. Sample overlap is accounted for by LD score regression. If there is no genetic correlation between any of the traits, and each trait has been measured on an independent sample, the MTAG effect estimates are equivalent to the single-trait estimates.

2.5.2. Bayesian methods

Conditional false discovery rate (cFDR) can detect variants associated with one 'principal' trait, given its p-values of association with both the 'principal' and the 'conditional' trait fall below a certain p-value threshold [82]. An extension to allow for shared controls has also been developed [46]. Conditional FDR can be used as a variant prioritization tool, to detect additional associated variants with one trait by leveraging information on their association with a second trait; furthermore, the method can serve to detect cross-phenotype effects explicitly by taking the maximum of the cFDRs computed for each of two traits [93].

Asimit and colleagues developed a Bayesian approach that adjusts for sample size (and, consequently, power) differences between studies [83]. Based on p-values from two input studies, the method tests for an excess of shared signals and can be used to identify a list of variants with evidence of association for both traits.

2.5.3. Other approaches

In metaCCA, a method based on canonical correlation analysis (CCA), single-trait summary statistics together with genotype data from a reference panel are used to reconstruct the genotype–phenotype covariance matrix for each study [66]. Weighted averages of these matrices are then combined across studies, and can be used to perform a multi-trait meta-analysis at a single variant or a genetic locus (based on LD).

2.5.4. Variant prioritization and fine-mapping

In order to aid variant prioritization and fine-mapping, the integration of GWAS data and functional annotations (such as chromatin marks) as well as transcriptomic data has become popular [94–96]. Some methods now allow for the inclusion of multiple GWAS of one or several traits.

The first tool allowing for the simultaneous modelling of multiple GWAS and annotation data was GPA [84]. The model tests for enrichment of annotations of (multi-)trait associated variants in functional datasets (e.g. eQTL or histone modifications). Although it can be easily extended to more than three traits/GWAS datasets, the degrees of freedom of the resulting chi-squared test grow with every trait tested, leading to a loss of power. As only binary annotations can be included in this model, the investigators built a revised approach, GPA-MDS, which combines multidimensional scaling with the GPA algorithm to draw visual representations of the relationships between multiple phenotypes [85].

FastPAINTOR, an extension to the single-trait fine-mapping algorithm PAINTOR [97], additionally models the LD structure at a locus which can improve accuracy [86]. Following a similar model to GPA, a Bayesian method entitled EPS also accounts for LD, but infers association at the gene- rather than the variant-level, and can incorporate gene expression data from a large number of tissues [87]. RiVIERA-MT further allows for multiple causal variants within one locus [88].

2.6. Single-point multivariate methods

As the availability of large-scale genetic datasets with multiple phenotype measurements increases, the focus of method development for multi-trait analyses has shifted towards multivariate methods that use individual-level data rather than summary statistics (table 2).

These approaches are generally more powerful than combining test statistics from univariate GWAS, as the inter-trait covariance can be accounted for [117,118]. In this section we will outline some of the methods available to conduct cross-phenotype analyses with individual-level data. Unless otherwise stated, all methods described in this section are based on multivariate models. For a more detailed discussion of the statistical properties of different multivariate approaches we refer the reader to two excellent reviews [117,119].

2.6.1. Dimension reduction

One efficient way to deal with multivariate phenotypes is to first apply a dimension reduction technique that collapses the individual trait values, and then perform an association between genotype and this new set of variables. Principal component analysis (PCA) derives linear combinations of the phenotypes that explain the greatest possible covariance

between them [117]. This approach was first used in linkage studies with multiple trait measurements, where principal components (PCs) were calculated to maximize the heritability in the data [120]. Extensions of this method were developed for family [98] and population-based studies [99]. The latter method, called PCHAT, requires splitting of the study sample into a 'training set' to derive the PCs and an 'analysis set' to perform the association, which can lead to a loss of power.

Recently, it has been shown that PCs can also be combined across individual multi-trait studies for meta-analysis. Average PCs are derived from the weighted means of the loadings (i.e. the linear combination of traits) in each study [100].

In contrast to PCA, CCA derives linear combinations of the traits that explain the greatest amount of covariance between a given variant and the traits. The first implementation of CCA for multivariate phenotypes was developed by Ferreira & Purcell [101], and later extended for gene-based analyses [64,65] (see §2.4).

2.6.2. Multivariate mixed models

Linear mixed models (LMMs) are an extension of standard regression analysis incorporating both fixed and random effects, and have gained popularity in GWAS due to their ability to handle relatedness among individuals [121,122]. Multivariate LMMs can be used for association testing with multiple phenotypes. They model association between a genetic marker and the traits as the fixed effect, and the inter-trait covariance as the random effect [117]. While multivariate mixed models are generally more powerful than standard univariate association tests, they perform les well when the traits under consideration are only weakly correlated [102], and assume phenotypes to be normally distributed, which does not allow for the inclusion of disease traits.

Korte and colleagues first applied multivariate LMMs to pairwise quantitative trait measurements in a human cohort [102]. The method, MTMM, showed increased power to detect loci compared with single-trait LMMs, and can also be used to decompose overall trait covariance into genetic and environmental factors.

Fitting multivariate LMMs requires a computationally intensive parameter estimation step, which until recently impeded their application to more than two traits [58,102,103]. A multivariate extension of the GEMMA algorithm [103] can accommodate a moderate number of phenotypes (between 2 and 10) and also shows substantially faster computation times compared with MTMM. A further improvement was achieved by Furlotte & Eskin with their matrix-variate linear mixed model (mvLMM), with runtime scaling linearly, rather than cubically, with the number of samples included [58]. The same research group also developed GAMMA, a mixed model that accurately adjusts for population structure, with computation time scaling linearly with the number of phenotypes, enabling it to jointly analyse large numbers of traits (more than 100) [104].

2.6.3. Generalized estimating equations

Generalized estimating equations (GEEs) are a multivariate method to jointly analyse non-normally distributed phenotypes [119,123]. As the standard GEE approach requires

Table 2. Multivariate methods for single-point association analysis. impl., implementation; ND, normally distributed.

method	ref.	PMID	year	data	*n* traits	trait type	impl.
FBAT-PC	[98]	16646795	2004	raw	≥2	any	C
PCHAT	[99]	17922480	2008	raw	≥2	any	Fortran
AvPC	[100]	27876822	2016	raw	≥2	any	NA
mvPlink	[101]	19019849	2009	raw	≥2	any	C++
MTMM	[102]	22902788	2012	raw	2	ND	R
GEMMA	[103]	24531419	2014	raw	≥2	ND	C/C++
mvLMM	[58]	25724382	2015	raw	≥2	ND	Python
GAMMA	[104]	27770036	2016	raw	≥2	ND	R
B_EGEE	[105]	18924135	2009	raw	2	any	Fortran
PleioGRiP	[106]	23419378	2013	raw	2	binary	Java
mvBIMBAM	[107]	23861737	2013	raw	≥2	ND	C/C++
Kendall's tau	[108]	20711441	2010	raw	≥2	any	NA
MultiPhen	[109]	22567092	2012	raw	≥2	any	R
ATeMP	[110]	26479245	2015	raw	≥2	any	NA
BAMP	[111]	26493781	2015	raw	≥2	any	NA
TATES	[112]	23359524	2013	*p*-values	≥2	any	R/Fortran
extension to O'Briens	[113]	20583287	2010	raw	≥2	any	upon request
Trinculo	[114]	26873930	2016	raw	≥2	categorical	C
log-linear model	[115]	21849790	2011	raw	≥2	binary	NA
PET	[116]	25044106	2014	raw	2	ND	R
PLeiotropySNP	[68]	27900789	2016	raw	≥2	any	R

traits to have the same underlying distribution, combining binary and continuous traits is not straightforward. B_EGEE was developed to overcome this constraint, using a regression procedure to incorporate two models with different underlying link functions into a unified equation system [105]. This method is implemented for two traits, although the authors suggest that the incorporation of multiple traits be possible [105]. While not as powerful as LMMs, GEEs have been used in family-based tests for both single- and multi-trait analysis due to their ability to model the random effects of relatedness [124,125].

2.6.4. Bayesian methods

Bayesian statistics allow for a model comparison between several alternative hypotheses, making them an attractive tool for pleiotropy analysis [42,43,107,126]. PleioGRiP is based on naive Bayesian classifiers that can be used to both test for association of a variant with two or more phenotypes and to perform genetic risk prediction [106,126].

A model-selection framework proposed by Stephens returns Bayes factors for each possible partitioning of phenotypes into one of three categories: unassociated, directly associated, or indirectly associated with a genetic marker [107]. The weighted average of these Bayes factors gives the overall evidence that any phenotype be associated. At markers where the evidence against this global null is strong, the individual Bayes factors can be used to determine which traits are likely to drive the association. The framework is implemented in the software MVBIMBAM and has been used to identify variants associated with low- and

intermediate-density lipoprotein subfractions [127]. While it has the advantage of accepting either individual-level or summary data as input, it is not applicable to non-normally distributed traits.

A Bayesian multivariate regression framework was also implemented in the widely used GWAS software SNPTEST [128].

2.6.5. Other approaches

A non-parametric score test based on Kendall's tau makes no assumption about the distribution of traits being analysed, and can thus incorporate any type of trait [108]. A limitation is the inability to adjust for covariates.

Another way to allow for the inclusion of traits with mixed distribution is to reverse the regression of standard GWAS. MultiPhen performs ordinal regression of the genotype (number of minor alleles at a marker) on multiple phenotypes and tests for association using a likelihood ratio test [109]. Alternatively, a score statistic can be used to test for association, which is computationally more efficient and equivalent to Kendall's tau. To prevent loss of power for non-normally distributed phenotypes, ATeMP first standardizes traits using a normalized rank or ordinal residual transformation [110]. Alternatively, allele-specific information can be related to trait values in binomial regression [111]. The latter approach, termed BAMP, is slightly more powerful in the case of non-normally distributed phenotypes [111,129].

Trait-based Association Test that uses Extended Simes procedure (TATES) is a model-free approach that combines univariate *p*-values of different traits while taking into

account the correlation between the traits [112]. Even though the test is based on summary statistics, individual-level data need to be available to calculate the correlation matrix. All samples are expected to be phenotyped across all traits.

Based on a univariate method developed by O'Brien [113], which derives a weighted sum of univariate test statistics, Yang and colleagues proposed two approaches that allow for effect heterogeneity among studies [130]: a sample-splitting approach, similar to that used in PCHAT [99], and a cross-validation approach. An advantage of this framework is that it can deal with missing phenotype data well (i.e. not all traits being measured on the same set of samples).

Multinomial models are a generalization of logistic regression, where the outcome variable (in this context the phenotype) can take more than two possible outcomes [131]. Such models are applicable to disease traits where cases can be classified into two or more sub-classes, for which risk variants might exhibit heterogeneous effects. While not technically multivariate, this type of analysis requires individual-level data. Bayesian and frequentist multinomial regression has been implemented in the software package Trinculo [114]. The Bayesian approach was recently used to determine which of a group of six immune disorders was associated with previously identified risk loci [132].

2.6.6. Explicit tests of pleiotropy

Model selection procedures are effective tools to discern associated from non-associated traits. A log-linear framework proposed by Lee and colleagues [115] distinguishes between different causal models of association, and can identify effects specific to a subgroup of individuals using information criteria. The method is only applicable to disease traits, and has been used to identify subtype-specific effects in migraine [133] and schizophrenia [134], as well as cross-disorder associations in psychiatric diseases [28,135]. Information criteria can also be used in an ordinal regression framework, such as MultiPhen [109], to identify the group of traits driving an association [129].

The pleiotropy estimation and test (PET) tool determines the pleiotropic effect of a variant on two continuous traits as the proportion of inter-trait correlation explained [116]. This quantity is assessed against the null hypothesis that neither trait is influenced by the variant.

Lutz and colleagues extended the above method for noncontinuous traits and for more than two phenotypes, as well as for rare variant or region-based tests (see also §2.4) [68]. The approach is based on performing standard univariate association analyses for each phenotype, followed by permutations of the phenotypes; evidence for pleiotropy is assessed by two alternative ways each of which compare the univariate p-values to the permutation-derived ones. The method tests for evidence against the null hypothesis of no association with *all* included phenotypes, and thus excludes the possibility of a signal being driven by one trait only. Therefore, if all but one of the analysed traits are associated with a variant or gene, the method will not reject the null.

2.7. Single-point methods summary

Several comparisons of different multi-trait methods have been conducted to date, testing power and type I error rates, as well as computational performance under different scenarios [118,129,136–138]. Since each report focused on a different combination of methods, the emerging picture seems to be that the most suitable method depends on individual study set-up.

Multivariate approaches are generally more powerful than univariate methods [138], unless only one trait is associated with a genetic marker or all traits are very highly correlated [102]. It is therefore advisable to perform both multivariate and univariate association tests in a complementary way [103]. This will not only enable the detection of additional signals, but also aid the interpretation of a multivariate association (i.e. which trait(s) is/are driving the signal). Since only a handful of currently available methods explicitly test for cross-phenotype effects, considering univariate association statistics also guards against false positive multi-trait associations.

When combining summary statistics across multiple traits in a univariate fashion, an important consideration is the power of individual studies. As for regional or genome-wide methods, single-point methods will fail to detect cross-phenotype associations if the input datasets are underpowered. Another important aspect is the ancestry of input study samples, especially for methods requiring the specification of reference panels [66], for which combining studies from different populations might lead to spurious results.

The implementation and computational efficiency of different methods are additional important considerations. This is especially relevant to multivariate methods using individual-level genotype data, which generally require more memory and processing power. Most of the multivariate approaches outlined above have been tested on small to moderate sample sizes ($n < 5000$), making it hard to predict their performance on large datasets. For example, methods implemented in the programming language R will generally be less memory-efficient than programs implemented in C or C++.

2.8. Detecting mediated pleiotropy

Determining whether the correlation between two traits is due to a causal link (i.e. trait 1 is a causal risk factor for trait 2) or due to confounding factors such as environmental exposures can be achieved through Mendelian randomization (MR). Notably, while most methods outlined in this review aim to detect biological pleiotropy and are confounded by mediated pleiotropy, the opposite is true for MR. MR uses information on the association of one or several genetic markers—instrumental variables (IVs)—with each trait to infer whether or not trait 1 causally influences trait 2 [11,139–142] (figure 2). An early example of MR is a study published in 2005 which concluded that, contrary to prior belief, C-reactive protein levels were not causal for metabolic syndrome [11]. If both traits were measured on the same samples an MR can be performed via two-sided least-squares analysis, where trait 1 is first regressed onto the IVs, and trait 2 is then regressed on the predicted values of trait 1 from the first regression. The effect size derived from the second regression is the MR estimate. If only summary data are available, or the associations for each trait were derived from different studies (referred to as two-sample MR), the Wald estimator can be used instead: this is the ratio of the effect of the variant on trait 1 over its effect on trait 2 [143].
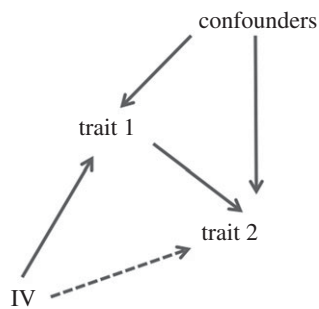
Figure 2. Directed acyclic graph of the Mendelian randomization model. IV, instrumental variable.

In order to be a valid IV, three key assumptions about the genetic marker must be met: first, the marker is associated with trait 1; second, the marker is not associated with any confounding variables, such as environmental exposures; and third, the marker is not associated with trait 2 when conditioning on trait 1. The first two assumptions are usually easy to fulfil in a GWAS context. The first assumption also implies that the function of the gene or marker used as an IV is known *a priori*. Consequently, MR is not a method to detect new genotype–phenotype associations [141]. Some consideration should be given to assumption 2, which can be violated in the case of population stratification [141,144]. Arguably the biggest uncertainty is the third assumption, which will not hold if the variant or variants used independently affect both trait 1 and trait 2 (i.e. if the variants are truly pleiotropic).

To overcome this limitation, several approaches have been proposed to both detect and subsequently adjust for pleiotropy in MR settings [145–147]. The inclusion of multiple independent variants is a prerequisite for such analyses: if all of the variants satisfy the IV assumptions there should be no heterogeneity between their individual MR estimates [141,145]. In other words, in the case of no pleiotropy, MR estimates of each variant will only vary by chance.

If individual-level data are available and both traits have been measured on the same sample, the Sargan test can be used to assess evidence against the null of all MR estimates being the same [144]. For two-sample MR, an adaptation of inverse-variance weighted meta-analysis can be used to combine Wald estimators across several variants [148]. This approach assumes that the three IV assumptions hold. The Cochran Q-statistic and the related I^2 index can be used to test for heterogeneity between individual IV estimates [145]. In Egger regression, variant-trait 2 effect sizes are regressed on the variant-trait 1 effect sizes with an unconstrained intercept [147]. An intercept term significantly different from zero is indicative of pleiotropy. If the distributions of the variant(s)-trait 1 effects and variant(s)-trait 2 effects are independent, the effect estimate obtained from Egger regression is equivalent to the MR estimate obtained from inverse variance weighted analysis and can thus be used to infer causality between trait 1 and 2. Recently, Bowden and colleagues proposed a summary data-based step-wise analysis framework which applies all three of the above methods to differentiate between the scenarios of no pleiotropy, pleiotropy without heterogeneity and pleiotropy with heterogeneity [146]. By applying this framework to summary data from two GWAS the authors showed that the observed association between plasma urate levels and cardiovascular disease was likely to be due to pleiotropy rather than a causal link, as evident from heterogeneity in the MR estimates from the 31 variants analysed.

2.9. Detecting spurious pleiotropy

In addition to methods for detecting cross-phenotype effects, a recently published tool (BUHMBOX) aims to distinguish heterogeneity in disease cases from true pleiotropy between two diseases [149]. It compares risk allele frequency of variants associated with one disease (D1) in the cases of a second disease (D2). By deriving a test statistic from the correlation matrix between all risk loci, BUHMBOX tests whether the D1 risk alleles are enriched in a subgroup of D2 cases (high correlation) or whether they are evenly distributed across all D2 cases (low correlation). The former yields a significant test statistic and is indicative of heterogeneity. On the other hand, a non-significant statistic could be the result of either true pleiotropy or insufficient power.

To our knowledge, BUHMBOX is currently the only method aimed at identifying spurious pleiotropy. BUHMBOX is not agnostic (i.e. D1 and D2 need to be specified by the user) and it is only applicable to disease traits.

3. Future directions

Pleiotropy will continue to grow as a key research area of human genetics as data availability increases.

A key question that has been the subject of debate in evolutionary biology [6] and is likely to receive more attention in human genetics is how to define a trait. For example, if two different anthropometric traits always covary, is it accurate to treat them as separate traits or are they actually two measurements of the same underlying biological phenotype? Such questions will be especially important in multivariate models where the inclusion of highly correlated traits can be costly in terms of power.

A further consideration is whether functional measurements, such as gene or protein expression, constitute traits in themselves or simply intermediate steps between genes and phenotype. This is especially important when attempting to draw biological conclusions from multi-trait analyses. MR approaches will be necessary to correctly infer causality, and indeed some work has already been done with regard to this question [150].

It is increasingly common for researchers to be confronted with sample numbers in the hundreds of thousands, and an important next step will be to develop fast and efficient algorithms with good scalability. In contrast to single-trait GWAS, where analysis approaches are now fairly standardized, multi-trait methods vary considerably in the statistical approaches they employ. Ensuring comparability and replicability of results will be a key challenge in pleiotropy research moving forward.

As data on cross-phenotype associations accumulate and our understanding of molecular links between diseases grows, these insights will be invaluable for drug development and repurposing, and for personalized medicine.

References

1. Price AL, Spencer CCA, Donnelly P. 2015 Progress and promise in understanding the genetic basis of common diseases. *Proc. R. Soc. B* **282**, 20151684. (doi:10.1098/rspb.2015.1684)

2. Ozaki K *et al.* 2002 Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. *Nat. Genet.* **32**, 650–654. (doi:10.1038/ng1047)

3. Klein RJ. 2005 Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385–389. (doi:10.1126/science.1109557)

4. Hirschhorn JN, Daly MJ. 2005 Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **6**, 95–108. (doi:10.1038/nrg1521)

5. Stearns FW. 2010 One hundred years of pleiotropy: a retrospective. *Genetics* **186**, 767–773. (doi:10.1534/genetics.110.122549)

6. Wagner GP, Zhang J. 2011 Fundamental concepts in genetics: the pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nat. Rev. Genet.* **12**, 204–213. (doi:10.1038/nrg2949)

7. Hodgkin J. 1998 Seven types of pleiotropy. *Int. J. Dev. Biol.* **505**, 501–505.

8. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. 2013 Pleiotropy in complex traits: challenges and strategies. *Nat. Rev. Genet.* **14**, 483–495. (doi:10.1038/nrg3461)

9. Paaby AB, Rockman MV. 2013 The many faces of pleiotropy. *Trends Genet.* **29**, 66–73. (doi:10.1016/j.tig.2012.10.010)

10. Gage SH, Davey Smith G, Ware JJ, Flint J, Munafò MR. 2016 G = E: what GWAS can tell us about the environment. *PLoS Genet.* **12**, e1005765. (doi:10.1371/journal.pgen.1005765)

11. Timpson NJ *et al.* 2005 C-reactive protein and its role in metabolic syndrome: Mendelian randomisation study. *Lancet* **366**, 1954–1959. (doi:10.1016/S0140-6736(05)67786-0)

12. Astle WJ *et al.* 2016 The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell.* **167**, 1415–1429. (doi:10.1016/j.cell.2016.10.042)

13. Zeggini E *et al.* 2012 Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet* **380**, 815–823. (doi:10.1016/S0140-6736(12)60681-3)

14. Panoutsopoulou K *et al.* 2014 The effect of FTO variation on increased osteoarthritis risk is mediated through body mass index: a Mendelian randomisation study. *Ann. Rheum. Dis.* **73**, 2082–2086. (doi:10.1136/annrheumdis-2013-203772)

15. Ripke S *et al.* 2014 Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427. (doi:10.1038/nature13595)

16. Liu JZ *et al.* 2012 Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. *Nat. Genet.* **44**, 1137–1141. (doi:10.1038/ng.2395)

17. Liu JZ *et al.* 2013 Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nat. Genet.* **45**, 670–675. (doi:10.1038/ng.2616)

18. Clayton DG *et al.* 2005 Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat. Genet.* **37**, 1243–1246. (doi:10.1038/ng1653)

19. Zhao W *et al.* 2017 Identification of new susceptibility loci for type 2 diabetes and shared etiological pathways with coronary heart disease. *Nat. Genet.* **49**, 1450–1457. (doi:10.1038/ng.3943)

20. Lie BA *et al.* 1999 The predisposition to type 1 diabetes linked to the human leukocyte antigen complex includes at least one non-class II gene. *Am. J. Hum. Genet.* **64**, 793–800. (doi:10.1086/302283)

21. Sivakumaran S *et al.* 2011 Abundant pleiotropy in human complex diseases and traits. *Am. J. Hum. Genet.* **89**, 607–618. (doi:10.1016/j.ajhg.2011.10.004)

22. Hobbs BD *et al.* 2017 Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat. Genet.* **49**, 426–432. (doi:10.1038/ng.3752)

23. Anttila V *et al.* 2013 Genome-wide meta-analysis identifies new susceptibility loci for migraine. *Nat. Genet.* **45**, 912–917. (doi:10.1038/ng.2676)

24. Freilinger T *et al.* 2012 Genome-wide association analysis identifies susceptibility loci for migraine without aura. *Nat. Genet.* **44**, 777–782. (doi:10.1038/ng.2307)

25. Locke AE *et al.* 2015 Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206. (doi:10.1038/nature14177)

26. Mahajan A *et al.* 2014 Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* **46**, 234–244. (doi:10.1038/ng.2897)

27. Smoller JW. 2013 Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371–1379. (doi:10.1016/S0140-6736(12)62129-1)

28. Lee SH *et al.* 2013 Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994. (doi:10.1038/ng.2711)

29. Li YR *et al.* 2015 Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. *Nat. Med.* **21**, 1018–1027. (doi:10.1038/nm.3933)

30. Stuart PE *et al.* 2015 Genome-wide association analysis of psoriatic arthritis and cutaneous psoriasis reveals differences in their genetic architecture. *Am. J. Hum. Genet.* **97**, 816–836. (doi:10.1016/j.ajhg.2015.10.019)

31. Sudlow C *et al.* 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. (doi:10.1371/journal.pmed.1001779)

32. Matise TC *et al.* 2011 The next PAGE in understanding complex traits: design for the analysis of population architecture using genetics and epidemiology (PAGE) study. *Am. J. Epidemiol.* **174**, 849–859. (doi:10.1093/aje/kwr160)

33. Muñoz M, Pong-Wong R, Canela-Xandri O, Rawlik K, Haley CS, Tenesa A. 2016 Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank. *Nat. Genet.* **48**, 980–983. (doi:10.1038/ng.3618)

34. Bush WS, Oetjens MT, Crawford DC. 2016 Unravelling the human genome-phenome relationship using phenome-wide association studies. *Nat. Rev. Genet.* **17**, 129–145. (doi:10.1038/nrg.2015.36)

35. Verma A *et al.* 2016 eMERGE phenome-wide association study (PheWAS) identifies clinical associations and pleiotropy for stop-gain variants. *BMC Med. Genomics* **9**, 32. (doi:10.1186/s12920-016-0191-8)

36. Krapohl E *et al.* 2016 Phenome-wide analysis of genome-wide polygenic scores. *Mol. Psychiatry* **21**, 1188–1193. (doi:10.1038/mp.2015.126)

37. Liu J *et al.* 2016 Phenome-wide association study maps new diseases to the human major histocompatibility complex region. *J. Med. Genet.* **53**, 681–689. (doi:10.1136/jmedgenet-2016-103867)

38. Pendergrass SA *et al.* 2013 Phenome-wide association study (PheWAS) for detection of pleiotropy within the population architecture using genomics and epidemiology (PAGE) Network. *PLoS Genet.* **9**, e1003087. (doi:10.1371/journal.pgen.1003087)

39. Cai N *et al.* 2015 Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* **523**, 588–591. (doi:10.1038/nature14659)

40. Ripke S *et al.* 2013 A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18**, 497–511. (doi:10.1038/mp.2012.21)

41. Anttila AV *et al.* 2016 Analysis of shared heritability in common disorders of the brain. *bioRxiv* (doi:10.1101/048991)

42. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. 2014 Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383. (doi:10.1371/journal.pgen.1004383)

43. Pickrell J, Berisa T, Segurel L, Tung JY, Hinds D. 2015 Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, 709–717. (doi:10.1038/ng.3570)

44. Bulik-Sullivan B *et al.* 2015 An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241. (doi:10.1038/ng.3406)

45. Bhattacharjee S *et al.* 2012 A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of

heterogeneous traits. *Am. J. Hum. Genet.* **90**, 821–835. (doi:10.1016/j.ajhg.2012.03.015)

46. Liley J, Wallace C. 2015 A pleiotropy-informed Bayesian false discovery rate adapted to a shared control design finds new disease associations from GWAS summary statistics. *PLoS Genet.* **11**, e1004926. (doi:10.1371/journal.pgen.1004926)

47. Province MA, Borecki IB. 2013 A correlated meta-analysis strategy for data mining 'OMIC' scans. *Pacific Symp Biocomput.* 236–246. (doi:10.1142/9789814447973_0023)

48. Han B, Duong D, Sul JH, de Bakker PIW, Eskin E, Raychaudhuri S. 2016 A general framework for meta-analyzing dependent studies with overlapping subjects in association mapping. *Hum. Mol. Genet.* **25**, 1857–1866. (doi:10.1093/hmg/ddw049)

49. Lin D-Y, Sullivan PF. 2009 Meta-analysis of genome-wide association studies with overlapping subjects. *Am. J. Hum. Genet.* **85**, 862–872. (doi:10.1016/j.ajhg.2009.11.001)

50. Talmud PJ, Hingorani AD, Cooper JA, Marmot MG, Brunner EJ, Kumari M, Kivimaki M, Humphries SE. 2010 Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. *BMJ* **340**, b4838. (doi:10.1136/bmj.b4838)

51. Dudbridge F. 2013 Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* **9**, e1003348. (doi:10.1371/journal.pgen.1003348)

52. Evans DM, Visscher PM, Wray NR. 2009 Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. *Hum. Mol. Genet.* **18**, 3525–3531. (doi:10.1093/hmg/ddp295)

53. Palla L, Dudbridge F. 2015 A fast method that uses polygenic scores to estimate the variance explained by genome-wide marker panels and the proportion of variants affecting a trait. *Am. J. Hum. Genet.* **97**, 250–259. (doi:10.1016/j.ajhg.2015.06.005)

54. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P. 2009 Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752.

55. Euesden J, Lewis CM, O'Reilly PF. 2015 PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466–1468.

56. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. 2012 Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* **28**, 2540–2542. (doi:10.1093/bioinformatics/bts474)

57. Loh P-R *et al.* 2015 Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47**, 1385–1392. (doi:10.1038/ng.3431)

58. Furlotte NA, Eskin E. 2015 Efficient multiple-trait association and estimation of genetic correlation using the matrix-variate linear mixed model. *Genetics* **200**, 59–68. (doi:10.1534/genetics.114.171447)

59. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, 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. (doi:10.1038/ng.3211)

60. Zheng J *et al.* 2017 LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279. (doi:10.1093/bioinformatics/btw613)

61. Gudmundsson J *et al.* 2007 Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. *Nat. Genet.* **39**, 977–983. (doi:10.1038/ng2062)

62. Huang J, Johnson AD, O'Donnell CJ. 2011 PRIMe: a method for characterization and evaluation of pleiotropic regions from multiple genome-wide association studies. *Bioinformatics* **27**, 1201–1206. (doi:10.1093/bioinformatics/btr116)

63. Berisa T, Pickrell JK. 2016 Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* **32**, 283–285. (doi:10.1093/bioinformatics/btv546)

64. Tang CS, Ferreira MAR. 2012 A gene-based test of association using canonical correlation analysis. *Bioinformatics* **28**, 845–850. (doi:10.1093/bioinformatics/bts051)

65. Seoane JA, Campbell C, Day INM, Casas JP, Gaunt TR. 2014 Canonical correlation analysis for gene-based pleiotropy discovery. *PLoS Comput. Biol.* **10**, e1003876. (doi:10.1371/journal.pcbi.1003876)

66. Cichonska A *et al.* 2016 metaCCA: summary statistics-based multivariate meta-analysis of genome-wide association studies using canonical correlation analysis. *Bioinformatics* **32**, 1981–1989. (doi:10.1093/bioinformatics/btw052)

67. Wang Y, Liu A, Mills JL, Boehnke M, Wilson AF, Bailey-Wilson JE, Xiong M, Wu CO, Fan R. 2015 Pleiotropy analysis of quantitative traits at gene level by multivariate functional linear models. *Genet. Epidemiol.* **39**, 259–275. (doi:10.1002/gepi.21895)

68. Lutz SM, Fingerlin TE, Hokanson JE, Lange C. 2016 A general approach to testing for pleiotropy with rare and common variants. *Genet. Epidemiol.* **41**, 163–170. (doi:10.1002/gepi.22011)

69. Casale FP, Rakitsch B, Lippert C, Stegle O. 2015 Efficient set tests for the genetic analysis of correlated traits. *Nat. Methods* **12**, 755–758. (doi:10.1038/nmeth.3439)

70. Mardis ER. 2017 DNA sequencing technologies: 2006–2016. *Nat. Protoc.* **12**, 213–218. (doi:10.1038/nprot.2016.182)

71. McCarthy S *et al.* 2016 A reference panel of 64 976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283. (doi:10.1038/ng.3643)

72. Asimit J, Zeggini E. 2010 Rare variant association analysis methods for complex traits. *Annu. Rev. Genet.* **44**, 293–308. (doi:10.1146/annurev-genet-102209-163421)

73. Lee S, Abecasis GR, Boehnke M, Lin X. 2014 Rare-variant association analysis: study designs and statistical tests. *Am. J. Hum. Genet.* **95**, 5–23. (doi:10.1016/j.ajhg.2014.06.009)

74. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. 2011 Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **89**, 82–93. (doi:10.1016/j.ajhg.2011.05.029)

75. Madsen BE, Browning SR. 2009 A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* **5**, e1000384. (doi:10.1371/journal.pgen.1000384)

76. Wu B, Pankow JS. 2016 Sequence kernel association test of multiple continuous phenotypes. *Genet. Epidemiol.* **40**, 91–100. (doi:10.1002/gepi.21945)

77. Lee S, Won S, Kim YJ, Kim Y, Kim B-J, Park T. 2016 Rare variant association test with multiple phenotypes. *Genet. Epidemiol.* **41**, 198–209. (doi:10.1002/gepi.22021)

78. Wang Z, Wang X, Sha Q, Zhang S. 2016 Joint analysis of multiple traits in rare variant association studies. *Ann. Hum. Genet.* **80**, 162–171. (doi:10.1111/ahg.12149)

79. Cotsapas C *et al.* 2011 Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet.* **7**, e1002254. (doi:10.1371/journal.pgen.1002254)

80. Zhu X *et al.* 2015 Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *Am. J. Hum. Genet.* **96**, 21–36. (doi:10.1016/j.ajhg.2014.11.011)

81. Turley P *et al.* 2017 MTAG: Multi-trait analysis of GWAS. *bioRxiv.* (doi:10.1101/118810)

82. Andreassen OA *et al.* 2013 Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. *PLoS Genet.* **9**, e1003455. (doi:10.1371/journal.pgen.1003455)

83. Asimit JL, Panoutsopoulou K, Wheeler E, Berndt SI, Cordell HJ, Morris AP, Zeggini E, Barroso I. 2015 A Bayesian approach to the overlap analysis of epidemiologically linked traits. *Genet. Epidemiol.* **39**, 624–634. (doi:10.1002/gepi.21919)

84. Chung D, Yang C, Li C, Gelernter J, Zhao H. 2014 GPA: a statistical approach to prioritizing GWAS results by integrating pleiotropy information and annotation data. *PLoS Genet.* **10**, 27. (doi:10.1371/journal.pgen.1004787)

85. Wei W, Ramos PS, Hunt KJ, Wolf BJ, Hardiman G, Chung D. 2016 GPA-MDS: a visualization approach to investigate genetic architecture among phenotypes using GWAS results. *Int. J. Genomics* **2016**, 1–6. (doi:10.1155/2016/6589843)

86. Kichaev G, Roytman M, Johnson R, Eskin E, Lindström S, Kraft P, Pasaniuc B. 2017 Improved methods for multi-trait fine mapping of pleiotropic risk loci. *Bioinformatics* **33**, 248–255. (doi:10.1093/bioinformatics/btw615)

87. Liu J, Wan X, Ma S, Yang C. 2016 EPS: an empirical Bayes approach to integrating pleiotropy and tissue-specific information for prioritizing risk genes. *Bioinformatics* **32**, 1856–1864. (doi:10.1093/bioinformatics/btw081)

88. Li Y, Kellis M. 2016 RiVIERA-MT: a Bayesian model to infer risk variants in related traits using summary statistics and functional genomic annotations. *bioRxiv.* (doi:10.1101/059345)

89. Evangelou E, Ioannidis JPA. 2013 Meta-analysis methods for genome-wide association studies and

beyond. *Nat. Rev. Genet.* **14**, 379 – 389. (doi:10.1038/nrg3472)

90. Wang Z *et al.* 2014 Imputation and subset-based association analysis across different cancer types identifies multiple independent risk loci in the TERT-CLPTM1 L region on chromosome 5p15.33. *Hum. Mol. Genet.* **23**, 6616 – 6633. (doi:10.1093/hmg/ddu363)

91. Park H, Li X, Song YE, He KY, Zhu X. 2016 Multivariate analysis of anthropometric traits using summary statistics of genome-wide association studies from GIANT consortium. *PLoS ONE* **11**, e0163912. (doi:10.1371/journal.pone.0163912)

92. Vuckovic D, Gasparini P, Soranzo N, Iotchkova V. 2015 MultiMeta: an R package for meta-analyzing multi-phenotype genome-wide association studies. *Bioinformatics* **31**, 2754 – 2756. (doi:10.1093/bioinformatics/btv222)

93. Andreassen OA *et al.* 2014 Identifying common genetic variants in blood pressure due to polygenic pleiotropy with associated phenotypes. *Hypertension* **63**, 819 – 826. (doi:10.1161/HYPERTENSIONAHA.113.02077)

94. Lu Q, Yao X, Hu Y, Zhao H. 2016 GenoWAP: GWAS signal prioritization through integrated analysis of genomic functional annotation. *Bioinformatics* **32**, 542 – 548. (doi:10.1093/bioinformatics/btv610)

95. Trynka G, Westra H-J, Slowikowski K, Hu X, Xu H, Stranger BE, Klein RJ, Han B, Raychaudhuri S. 2015 Disentangling the effects of colocalizing genomic annotations to functionally prioritize non-coding variants within complex-trait loci. *Am. J. Hum. Genet.* **97**, 139 – 152. (doi:10.1016/j.ajhg.2015.05.016)

96. Pickrell JK. 2014 Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94**, 559 – 573. (doi:10.1016/j.ajhg.2014.03.004)

97. Kichaev G *et al.* 2014 Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* **10**, e1004722. (doi:10.1371/journal.pgen.1004722)

98. Lange C *et al.* 2004 A family-based association test for repeatedly measured quantitative traits adjusting for unknown environmental and/or polygenic effects. *Stat. Appl. Genet. Mol. Biol.* **3**, article 17. (doi:10.2202/1544-6115.1067)

99. Klei L, Luca D, Devlin B, Roeder K. 2008 Pleiotropy and principal components of heritability combine to increase power for association analysis. *Genet. Epidemiol.* **32**, 9 – 19. (doi:10.1002/gepi.20257)

100. Ried JS *et al.* 2016 A principal component meta-analysis on multiple anthropometric traits identifies novel loci for body shape. *Nat Commun.* **7**, 13357. (doi:10.1038/ncomms13357)

101. Ferreira MAR, Purcell SM. 2009 A multivariate test of association. *Bioinformatics* **25**, 132 – 133. (doi:10.1093/bioinformatics/btn563)

102. Korte A, Vilhjálmsson BJ, Segura V, Platt A, Long Q, Nordborg M. 2012 A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat. Genet.* **44**, 1066 – 1071. (doi:10.1038/ng.2376)

103. Zhou X, Stephens M. 2014 Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* **11**, 407 – 409. (doi:10.1038/nmeth.2848)

104. Joo JWJ, Kang EY, Org E, Furlotte N, Parks B, Hormozdiari F, Lusis AJ, Eskin E. 2016 Efficient and accurate multiple-phenotype regression method for high dimensional data considering population structure. *Genetics* **204**, 1379 – 1390. (doi:10.1534/genetics.116.189712)

105. Liu J, Pei Y, Papasian CJ, Deng H-W. 2009 Bivariate association analyses for the mixture of continuous and binary traits with the use of extended generalized estimating equations. *Genet. Epidemiol.* **33**, 217 – 227. (doi:10.1002/gepi.20372)

106. Hartley SW, Sebastiani P. 2013 PleioGRiP: genetic risk prediction with pleiotropy. *Bioinformatics* **29**, 1086 – 1088. (doi:10.1093/bioinformatics/btt081)

107. Stephens M. 2013 A unified framework for association analysis with multiple related phenotypes. *PLoS ONE* **8**, e65245. (doi:10.1371/journal.pone.0065245)

108. Zhang H, Liu C-T, Wang X. 2010 An association test for multiple traits based on the generalized Kendall's tau. *J. Am. Stat. Assoc.* **105**, 473 – 481. (doi:10.1198/jasa.2009.ap08387)

109. O'Reilly PF, Hoggart CJ, Pomyen Y, Calboli FCF, Elliott P, Jarvelin M-R, Coin LJM. 2012 MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. *PLoS ONE* **7**, e34861. (doi:10.1371/journal.pone.0034861)

110. Guo X, Li Y, Ding X, He M, Wang X, Zhang H. 2015 Association tests of multiple phenotypes: ATeMP. *PLoS ONE* **10**, e0140348. (doi:10.1371/journal.pone.0140348)

111. Majumdar A, Witte JS, Ghosh S. 2015 Semiparametric allelic tests for mapping multiple phenotypes: binomial regression and Mahalanobis distance. *Genet. Epidemiol.* **39**, 635 – 650. (doi:10.1002/gepi.21930)

112. van der Sluis S, Posthuma D, Dolan CV. 2013 TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS Genet.* **9**, e1003235. (doi:10.1371/journal.pgen.1003235)

113. O'Brien PC. 1984 Procedures for comparing samples with multiple endpoints. *Biometrics* **40**, 1079 – 1087. (doi:10.2307/2531158)

114. Jostins L, McVean G. 2016 Trinculo: Bayesian and frequentist multinomial logistic regression for genome-wide association studies of multi-category phenotypes. *Bioinformatics* **32**, 1898 – 1900. (doi:10.1093/bioinformatics/btw075)

115. Lee PH, Bergen SE, Perlis RH, Sullivan PF, Sklar P, Smoller JW, Purcell SM. 2011 Modifiers and subtype-specific analyses in whole-genome association studies: a likelihood framework. *Hum. Hered.* **72**, 10 – 20. (doi:10.1159/000327158)

116. Zhang Q, Feitosa M, Borecki IB. 2014 Estimating and testing pleiotropy of single genetic variant for two quantitative traits. *Genet. Epidemiol.* **38**, 523 – 530. (doi:10.1002/gepi.21837)

117. Yang Q, Wang Y. 2012 Methods for analyzing multivariate phenotypes in genetic association studies. *J. Probab. Stat.* **2012**, 1 – 13. (doi:10.1155/2012/652569)

118. Galesloot TE, Van Steen K, Kiemeney LALM, Janss LL, Vermeulen SH. 2014 A comparison of multivariate genome-wide association methods. *PLoS ONE* **9**, 1 – 8. (doi:10.1371/journal.pone.0095923)

119. Shriner D. 2012 Moving toward system genetics through multiple trait analysis in genome-wide association studies. *Front. Genet.* **3**, 1. (doi:10.3389/fgene.2012.00001)

120. Ott J, Rabinowitz D. 1999 A principal-components approach based on heritability for combining phenotype information. *Hum. Hered.* **49**, 106 – 111. (doi:10.1159/000022854)

121. Zhang Z *et al.* 2010 Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **42**, 355 – 360. (doi:10.1038/ng.546)

122. Yu J *et al.* 2006 A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **38**, 203 – 208. (doi:10.1038/ng1702)

123. Hanley JA. 2003 Statistical analysis of correlated data using generalized estimating equations: an orientation. *Am. J. Epidemiol.* **157**, 364 – 375. (doi:10.1093/aje/kwf215)

124. Lange C, Silverman EK, Xu X, Weiss ST, Laird NM. 2003 A multivariate family-based association test using generalized estimating equations: FBAT-GEE. *Biostatistics* **4**, 195 – 206. (doi:10.1093/biostatistics/4.2.195)

125. Wang X, Lee S, Zhu X, Redline S, Lin X. 2013 GEE-based SNP set association test for continuous and discrete traits in family-based association studies. *Genet. Epidemiol.* **37**, 778 – 786. (doi:10.1002/gepi.21763)

126. Hartley SW, Monti S, Liu C-T, Steinberg MH, Sebastiani P. 2012 Bayesian methods for multivariate modeling of pleiotropic SNP associations and genetic risk prediction. *Front. Genet.* **3**, 176. (doi:10.3389/fgene.2012.00176)

127. Shim H, Chasman DI, Smith JD, Mora S, Ridker PM, Nickerson DA, Krauss RM, Stephens M. 2015 A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. *PLoS ONE* **10**, e0120758. (doi:10.1371/journal.pone.0120758)

128. Marchini J, Howie B, Myers S, McVean G, Donnelly P. 2007 A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906 – 913. (doi:10.1038/ng2088)

129. Majumdar A, Haldar T, Witte JS. 2016 Determining which phenotypes underlie a pleiotropic signal. *Genet. Epidemiol.* **40**, 366 – 381. (doi:10.1002/gepi.21973)

130. Yang Q, Wu H, Guo C-Y, Fox CS. 2010 Analyze multivariate phenotypes in genetic association studies by combining univariate association tests. *Genet. Epidemiol.* **34**, 444 – 454. (doi:10.1002/gepi.20497)

131. Morris AP, Lindgren CM, Zeggini E, Timpson NJ, Frayling TM, Hattersley AT, McCarthy MI. 2010 A powerful approach to sub-phenotype analysis in

population-based genetic association studies. *Genet. Epidemiol.* **34**, 335–343. (doi:10.1002/gepi.20486)

132. Ellinghaus D *et al.* 2016 Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat. Genet.* **48**, 510–518. (doi:10.1038/ng.3528)

133. Chasman DI *et al.* 2014 Selectivity in genetic association with sub-classified migraine in women. *PLoS Genet.* **10**, e1004366. (doi:10.1371/journal.pgen.1004366)

134. Bigdeli TB *et al.* 2016 Genome-wide association study reveals greater polygenic loading for schizophrenia in cases with a family history of illness. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **171**, 276–289. (doi:10.1002/ajmg.b.32402)

135. Huang J *et al.* 2010 Cross-disorder genomewide analysis of schizophrenia, bipolar disorder, and depression. *Am. J. Psychiatry* **167**, 1254–1263. (doi:10.1176/appi.ajp.2010.09091335)

136. Zhu H, Zhang S, Sha Q. 2015 Power comparisons of methods for joint association analysis of multiple phenotypes. *Hum. Hered.* **80**, 144–152. (doi:10.1159/000446239)

137. Aschard H, Vilhjálmsson BJ, Greliche N, Morange P-EE, Trégouët D-AA, Kraft P. 2014 Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. *Am. J. Hum. Genet.* **94**, 662–676. (doi:10.1016/j.ajhg.2014.03.016)

138. Porter HF, O'Reilly PF. 2017 Multivariate simulation framework reveals performance of multi-trait GWAS methods. *Sci. Rep.* **7**, 38837. (doi:10.1038/srep38837)

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

140. Thomas DC, Conti DV. 2004 Commentary: the concept of 'Mendelian randomization.' *Int. J. Epidemiol.* **33**, 21–25. (doi:10.1093/ije/dyh048)

141. 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. (doi:10.1002/sim.3034)

142. White J *et al.* 2016 Plasma urate concentration and risk of coronary heart disease: a Mendelian randomisation analysis. *Lancet Diabetes Endocrinol.* **4**, 327–336. (doi:10.1016/S2213-8587(15)00386-1)

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

144. Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, Davey Smith G, Sterne JAC. 2012 Using multiple genetic variants as instrumental variables for modifiable risk factors.

Stat. Methods Med. Res. **21**, 223–242. (doi:10.1177/0962280210394459)

145. Greco MFDel, Minelli C, Sheehan NA, Thompson JR. 2015 Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Stat. Med.* **34**, 2926–2940. (doi:10.1002/sim.6522)

146. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan N, Thompson J. 2017 A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat. Med.* **36**, 1783–1802. (doi:10.1002/sim.7221)

147. Bowden J, Smith GD, Burgess S. 2015 Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. *Int. J. Epidemiol.* **44**, 512–525. (doi:10.1093/ije/dyv080)

148. Burgess S, Butterworth A, Thompson SG. 2013 Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665. (doi:10.1002/gepi.21758)

149. Han B *et al.* 2016 A method to decipher pleiotropy by detecting underlying heterogeneity driven by hidden subgroups applied to autoimmune and neuropsychiatric diseases. *Nat. Genet.* **48**, 803–810. (doi:10.1038/ng.3572)

150. Zhu Z *et al.* 2016 Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487. (doi:10.1038/ng.3538)