GENOME-WIDE ASSOCIATION STUDIES

Meta-analysis methods for genome-wide association studies and beyond

Evangelos Evangelou[1] and John P. A. Ioannidis[2,3]

Abstract | Meta-analysis of genome-wide association studies (GWASs) has become a popular method for discovering genetic risk variants. Here, we overview both widely applied and newer statistical methods for GWAS meta-analysis, including issues of interpretation and assessment of sources of heterogeneity. We also discuss extensions of these meta-analysis methods to complex data. Where possible, we provide guidelines for researchers who are planning to use these methods. Furthermore, we address special issues that may arise for meta-analysis of sequencing data and rare variants. Finally, we discuss challenges and solutions surrounding the goals of making meta-analysis data publicly available and building powerful consortia.

Meta-analysis
A statistical method for the combination of different studies to provide a summary result.

[1]*Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina 45110, Greece.*
[2]*Stanford Prevention Research Center, Department of Medicine and Department of Health Research and Policy, Stanford University School of Medicine, Stanford, California 94305–5411, USA.*
[3]*Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, California 94305–5411, USA.*
Correspondence to J.P.A.I.
e-mail: jioannid@stanford.edu
doi:10.1038/nrg3472
Published online 9 May 2013

Genome-wide association scans allow millions of single-nucleotide polymorphisms (SNPs) to be genotyped or imputed and studied simultaneously using hypothesis-free agnostic approaches[1,2]. Although single genome-wide association studies (GWASs) have identified many variants associated with complex diseases, these variants currently explain little of the risk variability for most diseases. Genetic effects due to common alleles are small, and detection of signals requires large sample sizes[3–6]. As single GWASs are underpowered, meta-analysis — the statistical synthesis of information from multiple independent studies — increases power and reduces false-positive findings. Also, the techniques developed for meta-analysis can use summary data, diminishing the limitations that are imposed by restrictions on sharing individual-level data. Therefore meta-analysis has become a popular approach for the discovery of new genetic loci for common diseases and phenotypes. Most genetic risk variants discovered in the past few years have come from large-scale meta-analyses of GWASs[7]. Several hundred GWAS meta-analyses have already been published. In the period between the beginning of 2009 and 15th June 2012, 139 of these meta-analyses had sample sizes in the discovery phase exceeding 10,000 participants[8]. These efforts have dramatically increased the yield of discovered and validated genetic risk loci, thus making a case that the manpower and time required for such a design are justified. Large meta-analyses may continue to increase the yield of loci in proportion to total sample size[9,10].

As GWAS meta-analysis has become so popular and as new waves of even more extensive data are accumulating from sequencing efforts, it is timely to review the meta-analysis methods that have successfully been applied so far and also to highlight novel approaches that may be useful for the discovery of new variants. We cover the main models along with their extensions and the inferential tools that are being used, providing guidance for the most appropriate methods in particular situations. Moreover, we address issues of heterogeneity and describe methodological approaches that take into account heterogeneity introduced by various sources, such as phenotype definition, different ancestral groups, different genotyping platforms or imputation software. Given that research is increasingly shifting towards low-frequency and rare variants, we also describe methods for the meta-analysis of these variants.

Meta-analysis stages

Previous reviews have described in detail the organizational stages of a GWAS meta-analysis[7,11,12], and only a brief overview is provided here (FIG. 1). In BOX 1, we provide a more detailed description of different steps of meta-analysis, including setting up an analysis plan, dealing with heterogeneity, data storage, prioritization of variants for follow up and other secondary analysis[13–19]. Some successful high-profile consortia for which workflows can be followed and for which useful material describing their operations can be downloaded are summarized in TABLE 1.

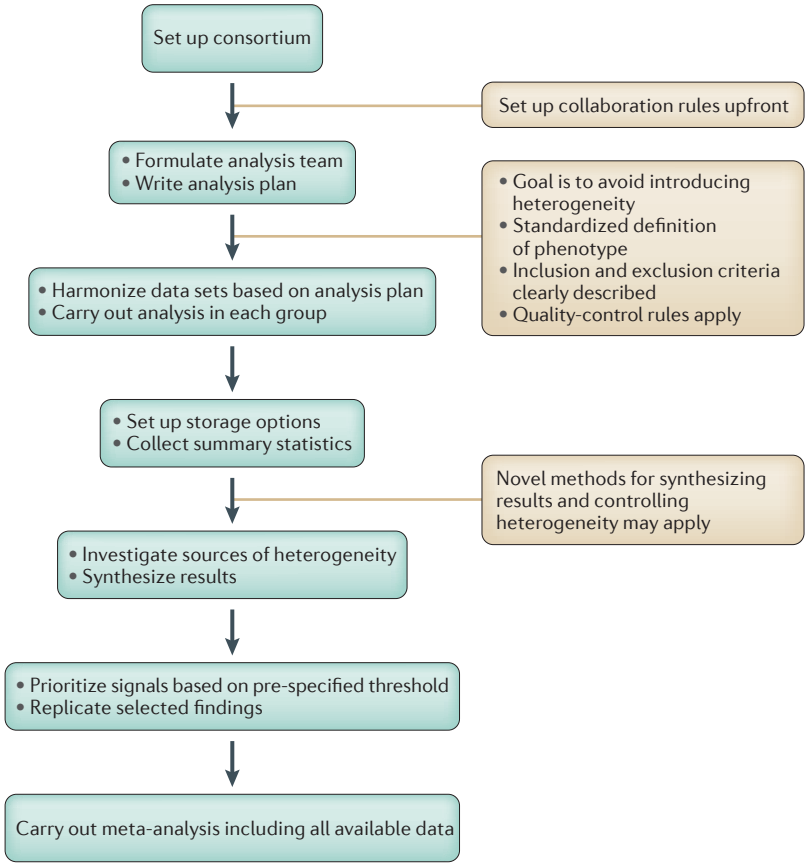

Figure 1 | **Stages in a meta-analysis.** A typical plan for a meta-analysis of genome-wide and next-generation sequence data.

Summary data
Data that present summary statistics of a population and are used in meta-analysis approaches without granting access to individual-level data.

Imputation
In genetics, the inference of genotypes of markers that have not been directly genotyped by making use of information from haplotype reference panels such as the HapMap or the 1000 Genomes panels.

Genome-wide significance
The significance threshold for rejecting the null hypothesis in genome-wide association studies.

Minor allele frequency
(MAF). The frequency of the less common allele of a polymorphic locus. It has a value that lies between 0 and 0.5 and can vary between populations.

Models for data synthesis

There are several approaches for GWAS meta-analysis. Here, we discuss methods that have been widely applied so far and alternative or new methods that have been proposed. For each model, issues such as weighing, power and ability to detect heterogeneity are discussed. TABLE 2 presents commonly used bioinformatics tools and software[20–23]. TABLE 3 presents the statistical properties of the most widely used methods.

P *values and* Z *scores.* Meta-analysis methods based on *P* values were widely used in different scientific fields until the 1980s but then became unpopular and were almost abandoned in biomedical sciences. This was because they suffered from several limitations, such as an inability to provide a summary effect and difficulties in addressing heterogeneity issues. The best known method of this type is Fisher's approach, but over a dozen other methods exist for combining *P* values in the traditional meta-analysis literature[24], although the large majority of these has never been applied with *P* values in the genome-wide significance range. In these approaches, the null hypothesis is that the true effect is null in each of the combined data sets (the alternative hypothesis is that there is a non-null association in at least one data set). The major disadvantage of *P* values meta-analysis is that it cannot provide an overall estimate of effect size. Moreover, between-data-set heterogeneity cannot be assessed, and there is disagreement on the optimal weighting of studies. Furthermore, combining *P* values may be spurious when the direction of effects in the combined studies is not consistent.

A closely related approach to Fisher's method is based on *Z* scores[24]. As Fisher's method is based on the average of $-\log(P_i)$ across all studies (where P_i is the *P* value of the i^{th} study) and as the *Z* scores method is based on the average of the Z_i values, these two approaches are highly correlated. One advantage of the *Z* score approach is that it takes into account the direction of the effect, and it is rather straightforward to introduce weights. A review has shown that meta-analysis of *Z* scores and *P* values (obtained using Fisher's method) has been applied in 14% and 3% of the published GWAS meta-analyses, respectively[25]. Both *P* values and *Z* scores meta-analysis might be useful tools for the synthesis of low-frequency variants (that is, those with a minor allele frequency (MAF) of 1–5%) or rare variants (that is, those with an MAF of <1%) when the association is based on tests that provide *P* values or *Z* test statistics rather than effect estimates[19,26] (see later for a discussion of methods for low-frequency and rare variants).

Fixed effects. Fixed effects meta-analysis is the most popular approach for synthesizing GWAS data and the most powerful approach for prioritizing and discovering phenotype-associated SNPs[27]. Fixed effects meta-analysis assumes that the true effect of each risk allele is the same in each data set. This is a rather tenuous assumption, but fixed effects models have the major advantage of maximizing discovery power compared to random effects models (see below)[28]. There are different models for fixed effects meta-analysis, but the inverse variance weighting[29], in which each study is weighted according to the inverse of its squared standard error, is predominantly used[7]. Another common method used in genetics is the Cochran–Mantel–Haenszel approach, which provides almost identical results to the inverse variance weighting method[24]. All bioinformatics tools that have been developed for GWAS meta-analysis carry out inverse variance meta-analysis[20–23].

Random effects. By far the most popular estimator of the between-study variance for the random effects approach is the DerSimonian and Laird estimator[30], but many others approaches also exist (for example, Sidik–Jonkman, Hedges–Vevea, Hunter–Schmidt and Schuster methods). The DerSimonian and Laird approach may give less robust results with small numbers (for example, rare variants)[24,31,32]. Random effects models are not used in discovery efforts owing to far more limited power than fixed effects models; however, random effects models are more appropriate than fixed effects models when the aim is to consider the generalizability of the observed association and estimate the average effect size of the associated variant and its uncertainty across different populations: for example, for predictive purposes[28,33].

Box 1 | **Stages in a GWAS meta-analysis**

Setting up an analysis plan

Each genome-wide association study (GWAS) meta-analysis initiative should be based on strong collaborative agreements and should be carefully designed and organized. An analysis team should design and draft a detailed plan that explicitly describes all of the steps of the anticipated analysis. Independent performance of some core statistical analyses by at least two analysts and using different software is not uncommon and may also allow for cross-verification and quality checks of processes and results. The analysis plan should be adopted by all teams, which should try to avoid deviations that introduce unnecessary between-study heterogeneity.

Dealing with heterogeneity

Despite careful planning to avoid heterogeneity, sometimes differences are inevitable, even in prospective designs: for example, when some samples have a family structure or when designs include extreme values[13]. Also, differences in phenotype definition may affect the estimated magnitude of the genetic effects, a factor that needs to be considered in terms of optimizing power for discovering new associations[6]. Ideally, phenotype definitions should be standardized according to stringent definitions applied in all data sets; if perfect standardization is impossible, participating teams should decide what kind of harmonization of definitions is desirable and feasible[14]. Inclusion and exclusion criteria of subjects and variants should be described in detail. Popular exclusion thresholds are >5% missing data, $P < 10^{-5}$ for Hardy–Weinberg equilibrium and quality index <0.3 for imputation metrics (BOX 2 summarizes the challenges of the imputation efforts using 1000 Genome Project panels[2,15]). Also, strand issues should be considered during quality control for the proper alignment of the alleles. Most GWAS meta-analyses to date focus on common variants and exclude variants with minor allele frequency <1%. However, this is likely to change as low-frequency and rare variants become the focus of interest. Statistical methods that account for between-study heterogeneity introduced by various sources are described in the main text.

Data storage

Data storage is an important aspect of meta-analysis as the individual-level data collected by each partner and also single participants' genotypes should be kept secured and unidentifiable. Most collaborative meta-analyses use online storage options to deposit summary data, giving access to members of the analysis team. This enables the partners to retain control of individual-level primary data. In most settings, summary data are statistically as efficient for meta-analysis as individual-level data[16]. The major drawback of working with summary data comes when more detailed investigations are required, such as conditional analyses, gene–gene interactions or adjusted analyses.

Prioritization of variants for follow-up and other secondary analyses

A GWAS meta-analysis may suggest variants to test in additional follow-up efforts. Prioritization of variants should follow rules that are pre-specified in the analysis plan. These typically include thresholds of statistical significance, but additional information may also be incorporated (for example, biological plausibility, evidence from other GWASs and other meta-analyses in the same field, text-mining-based methods or information on pathways[17]). The analysis plan should also specify how to choose among potentially multiple correlated variants based on linkage disequilibrium considerations.

When a variant passes the agreed level of genome-wide significance (typically set at $P < 5 \times 10^{-8}$), other secondary analyses might follow to assess the importance and the mode of action of the variant. For example, conditional analysis may be carried out for: other variants in the vicinity; adjustments for other traditional risk factors; gene–gene, gene–environment or protein–protein interactions; and diverse functional tests that may be carried out *in silico*, *in vitro* or *in vivo*[19].

Some new random effects methods have been proposed for improving discovery power in the presence of between-study heterogeneity in effect sizes. Han and Eskin[34] assume that there is no heterogeneity under the null hypothesis, in contrast to the traditional random effects models. Their statistic can be decomposed into two parts, one that is equal to the fixed effects statistic and another that is equal to the test statistic for heterogeneity. Further empirical validation of loci discovered by this method would be useful.

Optimal weights. Intuitively, it makes sense that studies with more data should count more (that is, have a larger weight) in the meta-analysis calculations, but the definition of weights depends on the statistical model used. The optimal weight for meta-analysis is the inverse variance weighting[35]. The favourable properties of such weighting in GWAS meta-analysis are well-known for initial screenings and discovery[28]. For imputed SNPs, simulations show that the optimal weight is proportional to the inverse variance and the expected value of the effect estimate[36], but the inverse variance modelling scheme approximates the optimal weights well. Other weights derived from score statistics can have practical advantages when only *P* values and direction of effects have been provided, and they might be easier to interpret when there are differences with respect to covariates or definitions of the phenotypes that are modelled. For small effect sizes, the optimal weights from score statistics have essentially the same power as inverse variance weighting[37].

Bayesian meta-analysis. Some consortia have applied Bayesian approaches for GWAS meta-analyses. Specifically, the Wellcome Trust Case Control Consortium has used the Bayes factor[38], whereas the Coronary Artery Disease Consortium has presented results as posterior probabilities that a variant is null[39].

Hardy–Weinberg equilibrium
A principle stating that the genetic variation in a population will remain constant from one generation to the next in the absence of disturbing factors.

Bayesian approaches
Fully probabilistic methods for describing models, parameters and data. They are so called because extensive use is made of Bayes' theorem to compute the probability distribution of model parameters given the experimental data.

Table 1 | **Examples of high-profile consortia for various disease phenotypes**

Consortium (acronym)	Phenotype (or phenotypes)	Publicly available genome-wide data?	Website
AMD	Age-related macular degeneration	Yes, accessible through the website	http://www.sph.umich.edu/csg/abecasis/public/amdgene2012
BCAC	Breast cancer	No	http://ccge.medschl.cam.ac.uk/consortia/bcac
CHARGE	Heart disease and ageing	No	http://web.chargeconsortium.com
GEFOS	Osteoporosis	Yes, accessible through the website	http://www.gefos.org
GIANT	Anthropometric traits	Yes, accessible through the website	http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium
GLGC	TC, HDL-C, LDL-C, triglycerides	Yes, accessible through the website	http://www.sph.umich.edu/csg/abecasis/public/lipids2010
IIBDGC	Inflammatory bowel disease	Yes, accessible through the website	http://www.ibdgenetics.org
IMSGC	Multiple sclerosis	Yes, accessible through the website	https://www.imsgenetics.org/
ISC	Schizophrenia	No	http://pngu.mgh.harvard.edu/isc
MAGIC	Glycaemic traits	Yes, accessible through the website	http://www.magicinvestigators.org
NARAC-III	Rheumatoid arthritis	No	http://www.naracstudy.org/NaracStudy/narac.aspx
TREATOA	Osteoarthritis	Yes, accessible through the website	http://treatoa.eu
WTCCC	Various phenotypes	Yes, accessible through the website	http://www.wtccc.org.uk

HDL-C: high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; TC, total cholesterol.

Bayesian methods have also been developed to characterize the best fitting inheritance model for variants emerging from GWAS meta-analyses[40], to characterize the polygenic architecture of complex traits for which there are thousands of top SNPs[41] and for multivariate methods (see relevant section below). Bayesian models are straightforward and intuitive, and the models used (for example, fixed, random effects) can be similar to the models used in the frequentist approaches described above. However, Bayesian models may depend on assumptions made about the prior distributions of parameters of interest, and their genome-wide implementation can become computationally intensive.

Inferential tools and their interpretation

P *values and genome-wide significance thresholds.* The most common measure used for flagging a SNP as 'noteworthy' is the *P* value[42]. The selection of the most suitable genome-wide significance threshold should take into account the multiplicity of comparisons carried out. Some early GWASs used a threshold of $P \leq 10^{-7}$, but the current practice, in which more SNPs are actually typed and the imputation quality is higher, accepts a threshold of $P \leq 5 \times 10^{-8}$ as reaching genome-wide significance. This roughly corresponds to a simple Bonferroni correction to maintain a 5% genome-wide type I error rate that is based on an estimated burden of 1 million independent comparisons for common sequence variation[18,43]. However, a study has shown that associations with borderline genome-wide significance (that is, $P \geq 5 \times 10^{-8}$ and $P \leq 10^{-7}$) are successfully replicated 73% of the time when additional data are accumulated[44]. The most suitable genome-wide significance threshold may vary for different populations, SNPs, MAFs, linkage disequilibrium (LD) patterns, arrays or different types of genetic data, but the differences are not major. For example, for populations with lower LD, such as Africans, stricter thresholds would be most appropriate[45].

Bayes factors. The Bayes factor[46] represents the ratio of the probability of the data under the null hypothesis, H_0, to the probability of data under the alternative hypothesis. Following the Bayes theorem, the posterior odds of H_0 = Bayes factor × prior odds of H_0. Bayes factor and *P* values will often provide similar rankings for MAFs in the range 0.1–0.5 (REF. 47). However, differences can be observed for SNPs with low MAFs, which will be more often assessed as next-generation sequencing data become more commonly used in GWASs[47].

Q *values (false discovery rate).* *Q* values and false discovery rate (FDR)[48] are very popular in 'omics' fields such as gene expression profiling, in which multiplicity issues are prominent, but surprisingly they have rarely been used in GWASs and meta-analyses thereof. The FDR estimates the proportion of associations that are seemingly discovered but are nevertheless expected to be false positives. The *Q* value is the minimum FDR that can be attained when declaring an association. *Q* values can be calculated for a list of *P* values using publicly available software called QVALUE[48]. An efficient process for selecting top SNPs using false-discovery rates following by further screening with least absolute shrinkage and selection operator (LASSO) regression has been proposed[49].

Reliability of signals from imputed genotypes. Imputation can allow data to be combined from studies using different genotyping platforms. Imputation is

Bonferroni correction
A method to counteract the problem of multiple comparisons. It is the simplest and most conservative approach to control for type I error.

Type I error
The probability of rejecting the null hypothesis when it is true. For genetic association studies, type I errors reflect false-positive findings of associations between allele or genotype and disease.

Linkage disequilibrium
The nonrandom association of alleles of different linked polymorphisms in a population.

Table 2 | **Comparison of meta-analysis software packages**

	METAL	GWAMA	MetABEL	PLINK	R packages
Ability to process files from GWAS analysis tools; software used	No	Yes; SNPTEST, PLINK	Yes; ABEL	Yes; PLINK	No
Fixed effects implemented?	Yes	Yes	Yes	Yes	Yes
Random effects implemented?	No	Yes	No	No	Yes
Heterogeneity metrics generated	Q, I^2	Q, I^2	Q, I^2	Q, I^2	Q, I^2
Graphical illustration of meta-analysis results	No	Manhattan and QQ plots	Forest plots	No	Yes

GWAS, genome-wide association study.

problematic for polymorphisms with low MAFs and those without at least modestly linked markers. Often, imputed variants with MAFs <5% (and certainly those with MAFs <1%) are disregarded and excluded from the analysis[11]. Imputation quality is usually measured by imputation accuracy. For most SNPs, imputation algorithms implemented in IMPUTE, MACH, MINIMAC or BEAGLE have very high imputation accuracy provided that the study population is similar to the reference population in terms of LD patterns and MAFs[50–52]. A metric to estimate the imputation quality is r^2, which is the ratio of the empirical variance of the imputed genotypes to the expected variance given the imputation estimate of the MAF. However, with low-frequency SNPs, the use of r^2 to evaluate the quality of the imputation can be misleading. Reliability of imputation in these cases can be assessed with imputation quality score (IQS)[53]. IQS takes chance agreement into account and thus controls for allele frequencies. Meta-analytical methods that take into account the errors in imputation by weighing not only in the inverse variance of each study but also in the accuracy of the imputed genotypes have been developed in each study; in some simulation studies, they may slightly outperform standard meta-analytical approaches[54], but this is not seen by all authors[37]. Challenges for imputations using panels from the 1000 Genomes Project are presented in BOX 2.

Dealing with heterogeneity

Phenotype-based heterogeneity. GWASs have been less successful for diseases in which phenotypes have been more difficult to define and to standardize, such as cognitive traits and mental-health-related diseases[55], behavioural traits[56] and osteoarthritis[57,58]. Evidence from other fields, such as obesity, also suggests that the establishment of associations may be dependent on phenotype definition[59] and that variability in definitions may cause heterogeneity in effect size or even spurious associations[60]. In some cases, harmonization of different phenotype definitions[61,62] is possible, whereas in other situations this may not be feasible: for example, if phenotypes have already been collected and it is not possible to go back and remeasure them. The process must balance the need to augment the sample size (to increase power for gene discovery) with the likelihood of increased heterogeneity, which dilutes the average genetic effect and thus leads to loss of power[6]. Methods have been proposed to improve power in the presence of heterogeneous traits[63,64] but have not been extensively used.

Ancestry-based heterogeneity. Synthesizing data from populations of different ancestry may increase the observed heterogeneity. The agnostic GWAS approach usually captures common markers that are likely to be in linkage only with the functional or causative culprits; the most strongly associated SNP may not be the functional or the causative variant. An assessment of GWAS-discovered variants shows modest correlation in MAF of the variants between ancestries and different genetic effects in different ancestries[65]. However, consistency across different ancestries may be higher for some common diseases[66]. A proposed transethnic meta-analysis approach takes into account the similarity in allelic effects between the most closely related populations while allowing for heterogeneity between more diverse ethnic groups[67]; this approach may occasionally improve power to detect a novel association and localize causal variants[67].

Other sources of heterogeneity. There are several other sources that can introduce heterogeneity in meta-analysis of genetic data. Population stratification may exist even in populations that are assumed to be fairly homogeneous: for example, Europeans[68–70]. One study showed different stratification patterns in the UK population for rare and low-frequency variants[71]. Adjusting for principal components can handle heterogeneity for common variants, but additional methods are required to address stratification issues for rare variants[71]. Heterogeneity can also be introduced by using different genotyping platforms, imputation software or quality-control criteria across studies[33]. Gene–gene interactions and gene–environment interactions with differential non-genetic environmental exposures across different populations may also introduce heterogeneity — see below for a discussion of meta-analysis of joint estimation of interaction effects.

Finally, sex differences may induce heterogeneity, and some studies suggest differential genetic effects with respect to sex for many common variants[72], although this has not been borne out in other large-scale empirical studies[73]. Methods that optimize power for meta-analysis of sex-specific GWASs have been proposed[74].

Population stratification
The presence of several population subgroups that show limited interbreeding. When such subgroups differ both in allele frequency and in disease prevalence, this can lead to erroneous results in association studies.

Principal components
A composite variable that summarizes the variation across a larger number of variables, each represented by a column of a matrix.

Table 3 | **Summary of methods for meta-analysis of genome-wide data**

Method	Description	Advantages	Disadvantages	Main software used
P value meta-analysis	Simplest meta-analytical approach	Allows meta-analysis when effects are not available	Direction of effect is not always available; inability to provide effect sizes; difficulties in interpretation	METAL, GWAMA, R packages
Fixed effects	Synthesis of effect sizes. Between-study variance is assumed to be zero	Effects readily available through specialized software	Results may be biased if a large amount of heterogeneity exists	METAL, GWAMA, R packages
Random effects	Synthesis of effect sizes. Assumes that the individual studies estimate different effects	Generalizability of results	Power deserts in discovery efforts; may yield spuriously large summary effect estimates when there are selection biases	GWAMA, R packages
Bayesian approach	Incorporates prior assessment of the genetic effects	Most direct method for interpretation of results as posterior probabilities given the observed data	Methodologically challenging; GWAS-tailored routine software not available; subjective prior information used	R packages
Multivariate approaches	Incorporates the possible correlation between outcomes or genetic variants	Increased power can identify variants that conventional meta-analysis do not reveal using the same data sets	Computationally intensive; software not available for all analyses; some may require individual-level data	GCTA for multi-locus approaches
Other extensions	A set of different approaches that allows for the identification of multiple variants across different diseases	Summary results of previous meta-analyses can be used	May need additional exploratory analyses for the identification of variants; prone to systematic biases	Software developed by the authors of the proposed methodologies

GCTA, genome-wide complex trait analysis; GWAS, genome-wide association study.

Statistical metrics of heterogeneity. Typical heterogeneity metrics that are extensively used in GWAS meta-analyses are Cochran's Q statistic and I^2. Cochran's Q statistic follows a χ^2 distribution with k–1 degrees of freedom, where k is the number of studies and is typically considered to be significant at $\alpha = 0.10$ (where α is the type I error)[35]. However, the test is often underpowered when there are few studies (for example, <20), which is a common scenario in GWAS meta-analysis. The power of the Q statistic can be estimated[75], and it can be used in the estimation of the posterior odds of heterogeneity in a Bayesian framework[76]. Q is also used in the estimation of the between-study variance; the ratio of the effect size over τ (which is defined in BOX 3) offers an estimate of how big the effect is against the typical variability seen across studies and populations. Finally, I^2 quantifies the heterogeneity by measuring the amount of heterogeneity that is not due to chance[77]. It ranges from 0–100% and is considered low, moderate, large and very large for values 0–25%, 25–50%, 50–75% and >75%, respectively[78]. Statistical properties of heterogeneity metrics are presented in BOX 3.

Other extensions

Cross-phenotype meta-analysis. Cross-phenotype meta-analysis (CPMA) aims to detect multiple associations at a single marker across different diseases that may share a common genetic background or that involve a common biological process, such as autoimmunity[79]. CPMA assumes that the P values used for the individual traits come from different non-overlapping cohorts; as such, it cannot be applied in the case of large consortia that investigate many phenotypes but usually share the same control samples[79]. Modest overlap of the control samples (<50%) is tolerable, but large overlaps erode the power of this statistic. CPMA has been applied in the study of immune-mediated and autoimmune diseases; of 107 previously identified variants across seven diseases, almost half were shared across diseases[79]. The CPMA statistic is agnostic to the direction of effect in each disease and can detect markers for at least some, but not necessarily all, phenotypes. It has one degree of freedom as it measures a deviation in P value behaviour instead of testing all possible combinations of diseases for association to each SNP, and therefore provides high power to reject the null hypothesis. However, power comes at the price of not knowing to which phenotypes the marker is associated, thus requiring further evidence.

Exploratory cross-phenotype checks. A common practice in many GWAS meta-analyses is to carry out cross-phenotype checks. Here, SNPs that have been identified as being significantly associated with one phenotype are also tested for association with other phenotypes in the same or other consortia. Typically, related phenotypes are examined: for example, SNPs that have been found to be associated with coronary artery disease may be specifically tested for association with phenotypes that represent diverse components of the metabolic syndrome[80]. Such checks could offer interesting insights about the pathophysiological paths implicated. The risk is that these analyses are exploratory, and thus their reporting may be less than comprehensive, potentially leading to selective reporting bias.

Joint meta-analysis for main and interaction effects. Main effects analysis may fail to identify additional susceptibility loci that interact with environmental factors

Box 2 | **Challenges for imputation using data from the 1000 Genomes Project**

Genotype imputation has widely been used in meta-analysis of genome-wide association studies (GWASs). In the first generation of GWASs, most imputation analysis used panels from Phase 2 of the International HapMap project, which contained a total of 210 unrelated individuals with ancestry from West Africa, East Asia and Europe. The 1000 Genomes Project[15] has now extended this resource by applying whole-genome shotgun sequencing to an even larger number of individuals (2,500) sampled from around 25 different locations. Direct sequencing can identify variants that are not represented on the genotyping arrays that were used in the HapMap Project and therefore can include more reference panels with most of the new variants occurring at low population frequencies. However, this might create enormous databases that require more exhaustive computer power to be analysed.

For original single-nucleotide polymorphism (SNP) sets from the 1000 Genomes Project, the estimated imputation quality criteria should be treated with caution and should be more conservative, given that early haplotype sets were of lower quality. However, in mid-2012, the 1000 Genomes Project released a powerful reference panel known as 'Phase I v3'. Different approaches for imputing the panels are proposed from publicly available imputation software such as IMPUTE2, MACH and BEAGLE.

if the genetic effects for distinct levels of an environmental exposure are not investigated. A method of joint analysis of SNP and SNP–environment interaction coefficients has been proposed using a multivariate approach that provides confidence intervals for the two estimates and a test of homogeneity across samples. This method performs better when both main and interaction effects are present[81], when joint meta-analysis may have a potential power gain of over 50% compared with classical meta-analysis[82]. Application of joint meta-analysis is hampered by a lack of sufficient environmental exposure data in most genetic data sets.

Meta-analysis of multiple correlated phenotypes. Consideration of several correlated phenotypes together in the same studies can boost power[83]. Different methods for meta-analyses of correlated phenotypes have been proposed[84,85]. The principle is that of bivariate meta-analysis, which is adapted for handling the correlation between outcomes[84]. The method can be applied both to binary and to continuous outcomes. One of the drawbacks is that individual-level data might be needed; however, methods for calculation of the covariance between two correlated effect sizes from summary data have also been proposed[85]. These approaches have not been applied yet in GWAS data sets, but they may offer the advantage of increased rigor to exploratory cross-phenotype checks within the same consortium.

Meta-analysis of multiple correlated variants. This multivariate approach is also known as multi-point, multi-locus, multi-marker or joint multiple-SNP analysis. The causal variant may not necessarily emerge as the top SNP in a GWAS meta-analysis, and there may be multiple causal variants at the same locus, each independently contributing to genetic association with the phenotype[86,87]. Conditional analysis has been used as a tool to detect secondary independent signals at a locus, but individual-level data are required, which are usually unavailable for large-scale meta-analysis of GWASs.

An approximate conditional approach using summary data from a meta-analysis and LD correlations between SNPs estimated from a reference sample (here, a subset of the meta-analysis sample) was successfully applied in a meta-analysis for height and body mass index[88]. This method identified 36 loci with multiple associated variants for height adding 49 additional SNPs on top of the already known variants. In this approach, a genome-wide stepwise selection procedure selects SNPs on the basis of conditional *P* values and estimates the joint effects of all selected SNPs after the model has been optimized. The method assumes that the reference sample is from the same population as the samples from which the genotype–phenotype associations are estimated (that is, that linkage correlation estimates in the reference sample are unbiased). Simulations show that *P* values from this approximate approach are consistent with those from conditional analysis using individual-level data[88]. A reference sample of at least 2,000 is required to estimate linkage correlations with little error.

Mendelian randomization in GWAS meta-analysis. Mendelian randomization entails the use of genetic variants as proxies ('instrumental variables') for modifiable or environmental exposures under investigation[89]. Meta-analysis methods have been proposed for Mendelian randomization studies[90]. Meta-analysis can help to bypass the problem that most genetic variants used as instrumental variables have small effects with large uncertainty in small studies, and thus they also leave large uncertainty about the presence or absence of causal relationships. For example, a large Mendelian randomization study based on a GWAS meta-analysis managed to demonstrate conclusively that high-density lipoprotein (HDL) is not causally related to coronary artery disease, as variants that affect specifically HDL levels do not have an impact on coronary artery disease risk[91].

Sequence data and low-frequency variants

Low event rates and zero counts. As discussed above, it has been common practice in GWAS meta-analyses to exclude low-frequency and rare variants. The main reason for these exclusions was related to imputation and genotyping quality, but this practice also worked as a safety net for avoiding false-positive or other spurious findings that would mostly be statistical artefacts. This situation is now changing as projects such as the 1000 Genomes Project[15] and the UK10K, along with the advent of next-generation sequencing technologies[92],

Main effects
The effects of a variable assuming no dependency or conditionality of other variables.

Bivariate meta-analysis
Joint synthesis of two phenotypes by using their correlation.

Box 3 | **Statistical properties of common GWAS meta-analysis approaches**

The simplest genome-wide association study (GWAS) meta-analysis approach is to combine *P* values using Fisher's method. The formula for the statistic is

$$X^2 = -2\sum_{i=1}^{k}\log(P_i)$$

where P_i is the *P* value for the i^{th} study, and *k* is the number of studies in the meta-analysis. Under the null hypothesis, X^2 follows a χ^2 distribution with $2k$ degrees of freedom. The *Z* scores meta-analysis can be implemented using the equation

$$Z = \frac{\sum_i Z_i w_i}{\sqrt{\sum_i w_i^2}}$$

where w_i is the square root of sample size of the i^{th} study and

$$Z_i = \Phi^{-1}\left(1-\frac{P_i}{2}\right)^{*}(\text{effect direction for study } i)$$

where Φ is the standard normal cumulative distribution function. For fixed effects models, inverse variance weighting is widely used. The weighted average of the effect sizes can be calculated as

$$\hat{\theta}_F = \frac{\sum_i w_i\hat{\theta}_i}{\sum_i w_i}$$

and the variance is

$$\mathrm{var}(\hat{\theta}_F) = \frac{1}{\sum_i w_i}$$

where $\hat{\theta}_i$ is the i^{th} study normalized effect (for example, logarithm of odds ratio or β-coefficient for a logistic regression for a binary phenotype or mean difference or standardized mean difference for a continuous phenotype), and w_i is the reciprocal of the estimated variance of the effect study. The random effects model incorporates the between-study variance of heterogeneity, and therefore the weight for the random effects model is calculated as

$$w_i^R = \frac{1}{\left(\frac{1}{w_i} + \hat{\tau}^2\right)}$$

where

$$\tau^2 = \frac{(Q-(k-1))}{\left(\sum_i w_i - \left(\frac{\sum_i w_i^2}{\sum_i w_i}\right)\right)}$$

and *Q* is Cochran's *Q* statistic, which is given by

$$Q = \sum_i w_i(\hat{\theta}_i - \hat{\theta}_F).$$

Another popular heterogeneity metric, I^2, is given by

$$I^2 = \frac{100^{*}(Q-(k-1))}{Q}.$$

The multivariate meta-analysis approaches are based on the calculation of a variance–covariance matrix for the correlated phenotypes or the single-nucleotide polymorphisms in linkage disequilibrium that will allow the calculation of the marginal effects. In cross-phenotype meta-analysis, the developed statistic measures the likelihood of the null hypothesis, given the data. The test is asymptotically distributed as

$$\chi^2_{df=1}.$$

will facilitate the study of low-frequency and rare variants. This translates, however, to larger sample size requirements to robustly identify associations involving variants in these classes that have modest effects. The problem is magnified for uncommon phenotypes. Asymptotic assumptions may not hold for rare variants or even low-frequency ones. Zero counts in 2 × 2 tables may become a common scenario. Typically used solutions use continuity corrections: for example, adding 0.5 to all four cells of the 2 × 2 table. However, most meta-analysis methods, including inverse variance fixed effects, DerSimonian and Laird random effects, and Mantel–Haenszel, perform poorly with such corrections to zero counts. The Peto method seems the least biased[93], but this method is also problematic when the two groups that are compared are unequal in terms of observed events and sample sizes. Other approaches besides continuity corrections may be considered (for example, arcsine transformation[94]), but there is no empirical evidence yet that relates to their application in genetics. Mega-analysis (which is the pooling of all individual-level data from all studies into a common single data set; see below) may also be useful, but in this case, analyses need to account carefully for ancestry and other sources of potential heterogeneity (as discussed above).

Merging rare variation. Single-point analysis of low-frequency and rare variants requires hundreds of thousands of individuals in order to detect weak effect sizes. Several approaches for the aggregate analysis of multiple low-frequency or rare variants across a locus that may exhibit heterogeneity have been proposed[95–102]. Locus-based methods that incorporate variant quality scores available from next-generation sequencing data using a regression-based collapsing approach and an allele-matching method have been also proposed[103]. These quality-incorporating approaches are more powerful than their unweighted counterparts, when the causal variants are of high quality[103]. A sequencing kernel association test (SKAT) seems to outperform previously proposed collapsing methods[104] when a large fraction of the variants in a region are non-causal or the effects of causal variants are in different directions. Collapsing methods seem to be more powerful when most variants in a region are causal, and the effects are in the same direction. Of course, further studies are needed to address power issues, given that few true-positive associations with rare variants have been identified so far. SKAT-O is another interesting unified approach with application in small case–control sequencing studies; it maximizes power by adaptively using the data to combine optimally the burden test and non-burden sequence kernel tests[105]. Besides the known approaches for single-variant tests, such as the typical logistic Wald test, other exact approaches could be applied, such as the Firth-bias-corrected likelihood ratio test[106].

Most of the methods described above provide a *P* value or a test statistic as a result for a locus-based association, and these can easily be combined in a meta-analysis when more than one data set is available, as described above. Software has been developed for

Asymptotic assumptions
When the sample size in a data set grows indefinitely, then the distribution of the estimators becomes approximately normal.

2 × 2 tables
A 2 × 2 table that describes the cross-classification of data that are divided into two groups with two categories in each.

Collapsing approach
Statistical methods for association analysis in which multiple low-frequency or rare variants are collapsing into a single locus.

Lambda inflation factor
A metric used in genetic association studies to correct for spurious associations (which may arise owing to population stratification) by estimating the extent of inflation in the statistical evidence and appropriately down-weighting this inflation.

the meta-analysis of rare variants (such as MetaSKAT, SKAT-O and Rare-METAL). However, it has been shown that meta-analyses based on *P* values or test statistics for rare variants lose important information. To retain full efficiency, the test can be reformulated using summary statistics from each cohort on both the per-variant score test and the genotype covariance matrix[107]. The individual-variant test statistics are routinely shared for meta-analysis in GWASs anyhow, and the genotype covariance matrix contains no information about genotype–phenotype associations. Thus, data sharing for these quantities should be no more difficult to arrange than in GWASs.

Another important issue to consider is that for rare variants, the identity of the carriers and their privacy may be jeopardized. A method has been developed (MetaSeq) for conducting the meta-analysis in which privacy is maintained through cryptographic means[108].

Genome-wide significance thresholds for sequencing. With the inclusion of low-frequency and rare variants from the 1000 Genomes Project imputations and sequencing efforts, in which up to ~50 million markers may be typed, the number of the independent loci under study will be significantly larger than the 1 million markers estimated for GWASs. Therefore, it may be argued that an even more stringent threshold, compared to the typical 5×10^{-8}, may be required to secure robust findings that will be considered replicated. Additional studies are needed to indicate the optimal threshold. This threshold may depend on sample size, asymptotical properties of the tests, number of low-frequency variants known and other factors.

Issues related to data sources

Publicly available data. The benefits of data sharing are obvious. The National Center for Biotechnology Information has created a public repository (called dbGaP) for individual-level phenotype, exposure, genotype and sequence data and the associations between them[109]. Moreover, several funders in Europe and the United States promote sharing of summary data from already published GWASs or meta-analyses of GWASs. This facilitates a wide range of secondary research, both methodological and applied, by investigators who can access these data. There is evidence that participants are in favour of providing a re-consent for submission of their data to dbGaP[110].

Meta-analyses of meta-analyses. For some phenotypes of interest, several independent or partially overlapping consortia may exist, such as those that are funded by different mechanisms and that are assembled by different initiatives. The successful cooperation between different consortia will allow for mega-analyses, which can include both previously published data and data from meta-analyses that are conducted prospectively. This will significantly increase the power to detect weak signals, especially for low-frequency and rare variants. Consortia studying anthropometric traits such as body mass index have now accumulated almost a quarter of a million participants[111], and this number is expected to increase markedly in the near future as it has been shown that meta-analysis sample size requirements increase steeply when small genetic effects are considered[4]. Big data in terms of sample sizes may be merged from existing and emerging research teams, consortia, biobanks and other population-level efforts.

The impressive increase in the numbers of included participants may result in an artificial increase of the lambda inflation factor, in many cases with values >2, thus reducing the novel variants assessed in the discovery effort if a genomic control for the observed value is taken in account. Approaches that correct for this inflated lambda and that are based on mixed models are being developed, such as EMMAX[112]. Furthermore, in such efforts, the researchers should always be cautious when they combine data with different sequencing depths and from different platforms. Given that expectations for finding many low-frequency and rare variants with large effects have not been validated in sequencing studies to date, meta-analytical methods will probably continue to be the key approach taken towards exploring further a genetic architecture that is characterized mostly by small effects.

Conclusions and future directions

Multiple methods have been developed for the meta-analysis of available genome-wide data improving our understanding of complex diseases and explaining part of the missing heritability that we were not able to identify using simple association metrics. Merging of existing research teams and consortia may further allow for the identification of novel variants and represent a significant advance in our understanding of genetic susceptibility. iGOGS, a recent example of such a collaborative effort including five consortia studying hormone-mediated cancers, revealed 74 new susceptibility loci for these cancers[113–117]. Such multi-consortium efforts may become a necessity in the future to identify more low-frequency and rare variants with small or even moderate effects. More empirical and simulation studies are needed to assess the advantages and disadvantages of some of the proposed techniques that were presented here and to delineate their optimal application and interpretation of the results that they produce.

1. McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Rev. Genet.* **9**, 356–369 (2008).
This is a comprehensive Review of challenges in the discovery of associations using GWASs.
2. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nature Rev. Genet.* **11**, 499–511 (2010).
3. Ioannidis, J. P., Trikalinos, T. A. & Khoury, M. J. Implications of small effect sizes of individual genetic variants on the design and interpretation of genetic association studies of complex diseases. *Am. J. Epidemiol.* **164**, 609–614 (2006).
4. Moonesinghe, R., Khoury, M. J., Liu, T. & Ioannidis, J. P. Required sample size and nonreplicability thresholds for heterogeneous genetic associations. *Proc. Natl Acad. Sci. USA* **105**, 617–622 (2008).
5. Chapman, K., Ferreira, T., Morris, A., Asimit, J. & Zeggini, E. Defining the power limits of genome-wide association scan meta-analyses. *Genet. Epidemiol.* **35**, 781–789 (2011).
6. Evangelou, E. *et al.* Impact of phenotype definition on genome-wide association signals: empirical evaluation in human immunodeficiency virus type 1 infection. *Am. J. Epidemiol.* **173**, 1336–1342 (2011).

7. Zeggini, E. & Ioannidis, J. P. Meta-analysis in genome-wide association studies. *Pharmacogenomics* **10**, 191–201 (2009).
8. Panagiotou, O. A., Willer, C. J., Hirschhorn, J. N. & Ioannidis, J. P. A. The power of meta-analysis of genome-wide association studies. *Annu. Rev. Genom. Hum. Genet.* (in the press).
9. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* **88**, 294–305 (2011).
In this paper, a method is presented for estimating the proportion of variation in disease liability that is captured in GWAS by simultaneously considering all SNPs.
10. Kutalik, Z. *et al.* Novel method to estimate the phenotypic variation explained by genome-wide association studies reveals large fraction of the missing heritability. *Genet. Epidemiol.* **35**, 341–349 (2011).
11. de Bakker, P. I. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **17**, R122–128 (2008).
12. Thompson, J. R., Attia, J. & Minelli, C. The meta-analysis of genome-wide association studies. *Brief. Bioinform.* **12**, 259–269 (2011).
13. Estrada, K. *et al.* Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nature Genet.* **44**, 491–501 (2012).
14. Seminara, D. *et al.* The emergence of networks in human genome epidemiology: challenges and opportunities. *Epidemiology* **18**, 1–8 (2007).
15. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
This is the first description of the 1000 Genomes Project.
16. Lin, D. Y. & Zeng, D. Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. *Genet. Epidemiol.* **34**, 60–66 (2010).
17. Cantor, R. M., Lange, K. & Sinsheimer, J. S. Prioritizing GWAS results: a review of statistical methods and recommendations for their application. *Am. J. Hum. Genet.* **86**, 6–22 (2010).
18. Dudbridge, F. & Gusnanto, A. Estimation of significance thresholds for genomewide association scans. *Genet. Epidemiol.* **32**, 227–234 (2008).
19. Ioannidis, J. P., Thomas, G. & Daly, M. J. Validating, augmenting and refining genome-wide association signals. *Nature Rev. Genet.* **10**, 318–329 (2009).
20. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
21. Magi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).
22. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23**, 1294–1296 (2007).
23. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
24. Cooper, H., Hedges, L. V. & Valentine, J. C. (eds) *The Handbook of Research Synthesis and Meta-Analysis* (Russell Sage Foundation, 2009).
25. Begum, F., Ghosh, D., Tseng, G. C. & Feingold, E. Comprehensive literature review and statistical considerations for GWAS meta-analysis. *Nucleic Acids Res.* **40**, 3777–3784 (2012).
26. Lawrence, R., Day-Williams, A. G., Elliott, K. S., Morris, A. P. & Zeggini, E. CCRaVAT and QuTie-enabling analysis of rare variants in large-scale case control and quantitative trait association studies. *BMC Bioinformatics* **11**, 527 (2010).
27. Pfeiffer, R. M., Mitchell, H. G. & Pee, D. On combining data from genome-wide assocition studies to discover disease-associated SNPs. *Statist. Sci.* **24**, 547–560 (2009).
28. Pereira, T. V., Patsopoulos, N. A., Salanti, G. & Ioannidis, J. P. Discovery properties of genome-wide association signals from cumulatively combined data sets. *Am. J. Epidemiol.* **170**, 1197–1206 (2009).
29. Kavvoura, F. K. & Ioannidis, J. P. Methods for meta-analysis in genetic association studies: a review of their potential and pitfalls. *Hum. Genet.* **123**, 1–14 (2008).
30. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control Clin. Trials* **7**, 177–188 (1986).
31. DerSimonian, R. & Kacker, R. Random-effects model for meta-analysis of clinical trials: an update. *Contemp. Clin. Trials* **28**, 105–114 (2007).
32. Shuster, J. J. Empirical versus natural weighting in random effects meta-analysis. *Stat. Med.* **29**, 1259–1265 (2010).
33. Ioannidis, J. P., Patsopoulos, N. A. & Evangelou, E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS ONE* **2**, e841 (2007).
34. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88**, 586–598 (2011).
35. Cochran, W. G. The combination of estimated from different experiments. *Biometrics* **10**, 101–129 (1954).
36. Jiao, S., Hsu, L., Hutter, C. M. & Peters, U. The use of imputed values in the meta-analysis of genome-wide association studies. *Genet. Epidemiol.* **35**, 597–605 (2011).
37. Zhou, B., Shi, J. & Whittemore, A. S. Optimal methods for meta-analysis of genome-wide association studies. *Genet. Epidemiol.* **35**, 581–591 (2011).
38. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
39. Samani, N. J. *et al.* Genomewide association analysis of coronary artery disease. *N. Engl. J. Med.* **357**, 443–453 (2007).
40. Salanti, G. *et al.* Underlying genetic models of inheritance in established type 2 diabetes associations. *Am. J. Epidemiol.* **170**, 537–545 (2009).
41. Stahl, E. A. *et al.* Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nature Genet.* **44**, 483–489 (2012).
42. Wacholder, S., Chanock, S., Garcia-Closas, M., El Ghormli, L. & Rothman, N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J. Natl Cancer Inst.* **96**, 434–442 (2004).
43. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M. J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* **32**, 381–385 (2008).
44. Panagiotou, O. A., Ioannidis, J. P. & The Genome-Wide Significance Project. What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. *Int. J. Epidemiol.* **41**, 273–286 (2012).
45. Palmer, N. D. *et al.* A genome-wide association search for type 2 diabetes genes in African Americans. *PLoS ONE* **7**, e29202 (2012).
46. Goodman, S. N. Toward evidence-based medical statistics. 2: the Bayes factor. *Ann. Intern. Med.* **130**, 1005–1013 (1999).
47. Wakefield, J. Bayes factors for genome-wide association studies: comparison with *P*-values. *Genet. Epidemiol.* **33**, 79–86 (2009).
48. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100**, 9440–9445 (2003).
49. Shi, G. *et al.* Mining gold dust under the genome wide significance level: a two-stage approach to analysis of GWAS. *Genet. Epidemiol.* **35**, 111–118 (2011).
50. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genet.* **39**, 906–913 (2007).
51. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genet.* **44**, 955–959 (2012).
A method is presented here for genotype imputation in GWASs using large reference panels.
52. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816–834 (2010).
53. Lin, P. *et al.* A new statistic to evaluate imputation reliability. *PLoS ONE* **5**, e9697 (2010).
54. Zaitlen, N. & Eskin, E. Imputation aware meta-analysis of genome-wide association studies. *Genet. Epidemiol.* **34**, 537–542 (2010).
55. Sabb, F. W. *et al.* Challenges in phenotype definition in the whole-genome era: multivariate models of memory and intelligence. *Neuroscience* **164**, 88–107 (2009).
56. The Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genet.* **42**, 441–447 (2010).
57. Evangelou, E. *et al.* Large-scale analysis of association between GDF5 and FRZB variants and osteoarthritis of the hip, knee, and hand. *Arthritis Rheum.* **60**, 1710–1721 (2009).
58. Evangelou, E. *et al.* Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. *Ann. Rheum. Dis.* **70**, 349–355 (2011).
59. Kring, S. I. *et al.* Genotype-phenotype associations in obesity dependent on definition of the obesity phenotype. Obes Facts **1**, 138–145 (2008).
60. Heid, I. M. *et al.* Meta-analysis of the INSIG2 association with obesity including 74,345 individuals: does heterogeneity of estimates relate to study design? *PLoS Genet.* **5**, e1000694 (2009).
61. Kerkhof, H. J. *et al.* Recommendations for standardization and phenotype definitions in genetic studies of osteoarthritis: the TREAT-OA consortium. *Osteoarthritis Cartilage* **19**, 254–264 (2011).
62. Bennett, S. N. *et al.* Phenotype harmonization and cross-study collaboration in GWAS consortia: the GENEVA experience. *Genet. Epidemiol.* **35**, 159–173 (2011).
63. Bhattacharjee, S. *et al.* 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 (2012).
64. Behrens, G., Winkler, T. W., Gorski, M., Leitzmann, M. F. & Heid, I. M. To stratify or not to stratify: power considerations for population-based genome-wide association studies of quantitative traits. *Genet. Epidemiol.* **35**, 867–879 (2011).
65. Ntzani, E. E., Liberopoulos, G., Manolio, T. A. & Ioannidis, J. P. Consistency of genome-wide associations across major ancestral groups. *Hum. Genet.* **131**, 1057–1071 (2012).
66. Waters, K. M. *et al.* Consistent association of type 2 diabetes risk variants found in Europeans in diverse racial and ethnic groups. *PLoS Genet.* **6**, e1001078 (2010).
67. Morris, A. P. Transethnic meta-analysis of genomewide association studies. *Genet. Epidemiol.* **35**, 809–822 (2011).
68. Tian, C. *et al.* Analysis and application of European genetic substructure using 300 K SNP information. *PLoS Genet.* **4**, e4 (2008).
69. Paschou, P. *et al.* Tracing sub-structure in the European American population with PCA-informative markers. *PLoS Genet.* **4**, e1000114 (2008).
70. Clayton, D. G. *et al.* Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nature Genet.* **37**, 1243–1246 (2005).
71. Babron, M. C., de Tayrac, M., Rutledge, D. N., Zeggini, E. & Genin, E. Rare and low frequency variant stratification in the UK population: description and impact on association tests. *PLoS ONE* **7**, e46519 (2012).
72. Heid, I. M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nature Genet.* **42**, 949–960 (2010).
73. Orozco, G. Ioannidis, J. P., Morris, A., Zeggini, E. & The DIAGRAM Consortium. Sex-specific differences in effect size estimates at established complex trait loci. *Int. J. Epidemiol.* **41**, 1376–1382 (2012).
74. Magi, R., Lindgren, C. M. & Morris, A. P. Meta-analysis of sex-specific genome-wide association studies. *Genet. Epidemiol.* **34**, 846–853 (2010).
75. Jackson, D. The power of the standard test for the presence of heterogeneity in meta-analysis. *Stat. Med.* **25**, 2688–2699 (2006).
76. Pereira, T. V., Patsopoulos, N. A., Salanti, G. & Ioannidis, J. P. A. Clinical interpretation of Cochran's *Q* test depends on power and prior assumptions about heterogeneity. *Res. Synthesis Methods* **1**, 149–161 (2010).
77. Higgins, J. P. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21**, 1539–1558 (2002).
78. Higgins, J. P., Thompson, S. G., Deeks, J. J. & Altman, D. G. Measuring inconsistency in meta-analyses. *BMJ* **327**, 557–560 (2003).
79. Cotsapas, C. *et al.* Pervasive sharing of genetic effects in autoimmune disease. *PLoS Genet.* **7**, e1002254 (2011).
80. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genet.* **43**, 333–338 (2011).

81. Manning, A. K. *et al.* Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP x environment regression coefficients. *Genet. Epidemiol.* **35**, 11–18 (2011).
82. Aschard, H., Hancock, D. B., London, S. J. & Kraft, P. Genome-wide meta-analysis of joint tests for genetic and gene-environment interaction effects. *Hum. Hered.* **70**, 292–300 (2010).
83. Riley, R. D., Abrams, K. R., Lambert, P. C., Sutton, A. J. & Thompson, J. R. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Stat. Med.* **26**, 78–97 (2007).
84. Bagos, P. G. A unification of multivariate methods for meta-analysis of genetic association studies. *Stat. Appl. Genet. Mol. Biol.* **7**, 31 (2008).
85. Bagos, P. G. On the covariance of two correlated log-odds ratios. *Stat. Med.* **31**, 1418–1431 (2012).
86. Galarneau, G. *et al.* Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. *Nature Genet.* **42**, 1049–1051 (2010).
87. Sanna, S. *et al.* Fine mapping of five loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability. *PLoS Genet.* **7**, e1002198 (2011).
88. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genet.* **44**, 369–*375* (2012).
89. Smith, G. D. & Ebrahim, S. Mendelian randomization: prospects, potentials, and limitations. *Int. J. Epidemiol.* **33**, 30–42 (2004).
90. Burgess, S., Thompson, S. G. & CRP CHD Genetics Collaboration. Methods for meta-analysis of individual participant data from Mendelian randomisation studies with binary outcomes. *Stat. Methods Med. Res.* 19 Jun 2012 (doi:10.1177/0962280212451882).
91. Voight, B. F. *et al.* Plasma HDL cholesterol and risk of myocardial infarction: a Mendelian randomisation study. *Lancet* **380**, 572–580 (2012).
92. Day-Williams, A. G. & Zeggini, E. The effect of next-generation sequencing technology on complex trait research. *Eur. J. Clin. Invest.* **41**, 561–567 (2011).
93. Bradburn, M. J., Deeks, J. J., Berlin, J. A. & Russell Localio, A. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Stat. Med.* **26**, 53–77 (2007).
94. Rucker, G., Schwarzer, G., Carpenter, J. & Olkin, I. Why add anything to nothing? The arcsine difference as a measure of treatment effect in meta-analysis with zero cells. *Stat. Med.* **28**, 721–738 (2009).
95. Li, B. & Leal, S. M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am. J. Hum. Genet.* **83**, 311–321 (2008).
96. Madsen, B. E. & Browning, S. R. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* **5**, e1000384 (2009).
97. Morris, A. P. & Zeggini, E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genet. Epidemiol.* **34**, 188–193 (2010).
98. Han, F. & Pan, W. A data-adaptive sum test for disease association with multiple common or rare variants. *Hum. Hered.* **70**, 42–54 (2010).
99. Zawistowski, M. *et al.* Extending rare-variant testing strategies: analysis of noncoding sequence and imputed genotypes. *Am. J. Hum. Genet.* **87**, 604–617 (2010).
100. Price, A. L. *et al.* Pooled association tests for rare variants in exon-resequencing studies. *Am. J. Hum. Genet.* **86**, 832–838 (2010).
101. Bhatia, G. *et al.* A covering method for detecting genetic associations between rare variants and common phenotypes. *PLoS Comput. Biol.* **6**, e1000954 (2010).
102. Shriner, D. & Vaughan, L. K. A unified framework for multi-locus association analysis of both common and rare variants. *BMC Genomics* **12**, 89 (2011).
103. Asimit, J. L., Day-Williams, A. G., Morris, A. P. & Zeggini, E. ARIEL and AMELIA: testing for an accumulation of rare variants using next-generation sequencing data. *Hum. Hered.* **73**, 84–94 (2012).
104. Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **89**, 82–93 (2011).
A SKAT is described here for the identification of rare variants associated with continuous of dichotomous traits.
105. Lee, S. *et al.* Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am. J. Hum. Genet.* **91**, 224–237 (2012).
106. Firth, D. Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 12 (1993).
107. Lumley, T., Brody, J., Dupuis, J. & Cupples, A. Meta-analysis of a rare-variant association test. *Stat Tech* [online], http://stattech.wordpress.fos.auckland.ac.nz/files/2012/11/skat-meta-paper.pdf (2012).
108. Singh, A. P., Zafer, S. & Pe'er, I. Metaseq: privacy preserving meta-analysis of sequencing-based association studies. *Pac. Symp. Biocomput.* **2013**, 356–367 (2013).
109. Mailman, M. D. *et al.* The NCBI dbGaP database of genotypes and phenotypes. *Nature Genet.* **39**, 1181–1186 (2007).
110. Ludman, E. J. *et al.* Glad you asked: participants' opinions of re-consent for dbGap data submission. *J. Empir. Res. Hum. Res. Eth.* **5**, 9–16 (2010).
111. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genet.* **42**, 937–948 (2010).
112. Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nature Genet.* **42**, 348–354 (2010).
113. Garcia-Closas, M. *et al.* Genome-wide association studies identify four ER negative-specific breast cancer risk loci. *Nature Genet.* **45**, 392–398 (2013).
114. Eeles, R. A. *et al.* Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nature Genet.* **45**, 385–391 (2013).
115. Pharoah, P. D. *et al.* GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nature Genet.* **45**, 362–370 (2013).
116. Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nature Genet.* **45**, 353–361 (2013).
This is a multi-consortium effort that led to the identification of numerous novel loci associated with breast cancer. Separate papers described the identification of additional loci were found for prostate and ovarian cancer.
117. Bojesen, S. E. *et al.* Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nature Genet.* **45**, 371–384 (2013).

Acknowledgements
E.E. is partially funded by the GEFOS (FP7-HEALTH-F2-2008-201865-GEFOS) and the TREATOA (FP7-HEALTH-F2-2008-200800-TREATOA) projects.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
1000 Genomes Project: http://www.1000genomes.org
dbGaP: http://www.ncbi.nlm.nih.gov/gap
GWAMA: http://www.well.ox.ac.uk/gwama
IMPUTE2: http://mathgen.stats.ox.ac.uk/impute/impute_v2.html
metABEL: http://www.genabel.org/packages/MetABEL
METAL: http://www.sph.umich.edu/csg/abecasis/metal
metaSKAT/SKAT-O: http://people.hsph.harvard.edu/~sglee/MetaSKAT
PLINK: http://pngu.mgh.harvard.edu/~purcell/plink
QVALUE: http://genomine.org/qvalue
Rare-METAL: http://genome.sph.umich.edu/wiki/Rare-Metal
SKAT: http://www.bios.unc.edu/~mcwu/software/index.html#SKAT
UK10K: http://www.uk10k.org
Wellcome Trust Case Control Consortium: http://www.wtccc.org.uk

ALL LINKS ARE ACTIVE IN THE ONLINE PDF