Rostrum

Investigating asthma heterogeneity through shared and distinct genetics: Insights from genome-wide cross-trait analysis

Zhaozhong Zhu, ScD,[a,b,c] Kohei Hasegawa, MD, MPH,[b] Carlos A. Camargo, Jr, MD, DrPH, FAAAAI,[a,b] and Liming Liang, PhD[a,d] *Boston, Mass*

Asthma is a heterogeneous respiratory disease reflecting distinct pathobiologic mechanisms. These mechanisms are based, at least partly, on different genetic factors shared by many other conditions, such as allergic diseases and obesity. Investigating the shared genetic effects enables better understanding of the mechanisms of phenotypic correlations and is less subject to confounding by environmental factors. The increasing availability of large-scale genome-wide association study (GWAS) for asthma has enabled researchers to examine the genetic contributions to the epidemiologic associations between asthma subtypes and those between coexisting diseases and/or traits and asthma. Studies have found not only shared but also distinct genetic components between asthma subtypes, indicating that the heterogeneity is related to distinct genetics. This review summarizes a recently compiled analytic approach—genome-wide cross-trait analysis—to determine shared and distinct genetic architecture. The genome-wide cross-trait analysis features in several analytic aspects: genetic correlation, cross-trait meta-analysis, Mendelian randomization, polygenic risk score, and functional analysis. In this article, we discuss in detail the scientific goals that can be achieved by these analyses, their advantages, and their limitations. We also make recommendations for future directions: (1) ethnicity-specific asthma GWASs and (2) application of cross-trait methods to multiomics data to dissect the heritability found in GWASs. Finally, these analytic approaches are also applicable to complex and heterogeneous traits beyond asthma. (J Allergy Clin Immunol 2020;■■■:■■■-■■■.)

Key words: *asthma, heterogeneity, GWAS, shared genetics, cross-trait, allergic diseases, obesity*

Abbreviations used

ASSET:	Association analysis based on SubSETs
CAAPA:	Consortium on Asthma among African-Ancestry Populations in the Americas
CPASSOC:	Cross-phenotype association
ENCODE:	Encyclopedia of DNA elements
eQTL:	Expression quantitative trait locus
GTEx:	Genotype-Tissue Expression Project
GWAS:	Genome-wide association study
ImmGen:	Immunological Genome Project
LD:	Linkage disequilibrium
LDSC:	Linkage disequilibrium score regression
SNP:	Single-nucleotide polymorphism

EPIDEMIOLOGIC FACTORS ASSOCIATED WITH ASTHMA

Asthma is a common chronic respiratory disease that affects approximately 340 million individuals worldwide.[1] Epidemiologic studies have identified many factors associated with asthma, including the following: genetic factors; demographics (eg, age and sex); personal and family history of allergic diseases and comorbid illnesses (eg, obesity and mental illnesses); and environmental (eg, air pollution), nutritional, and lifestyle (eg, physical activity) factors.[2]

Of the various factors associated with development of asthma, allergic diseases such as atopic dermatitis and allergic rhinitis play a pivotal role. The coexistence of allergic diseases with asthma in children and adolescences is high (eg, as high as 60% to 80% for allergic rhinitis).[3] In both children and adults, allergic sensitization to allergens has been found to be an important risk factor for asthma and bronchial hyperresponsiveness.[4-8] It is likely that the frequent coexistence of allergic diseases and asthma is due to their similar pathobiologic mechanisms.[9]

Another major factor associated with asthma is obesity in both children and adults.[10,11] This relation is complex: "obese-asthma syndrome" consists of multiple subgroups (eg, *de novo* asthma, asthma modified by obesity, and obesity predisposed by asthma).[10] Yet, studies have suggested the causal link from (anthropomorphically defined) overweight or obesity to asthma inception.[10,12,13] Emerging evidence also suggests the role of adiposopathy—"sick fat" or adipose tissue dysfunction—in the pathogenesis of complex disease conditions, including asthma.[10] Adiposopathy is characterized by impaired adipogenesis, altered lipid metabolism, and adipose and/or systemic inflammation (eg, upregulated IL-6, T_H1 polarization, and T_H17 pathways).[10]

From [a]the Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, [c]the Department of Environmental Health, and [d]the Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, and [b]the Department of Emergency Medicine, Massachusetts General Hospital, Harvard Medical School, Boston.

Disclosure of potential conflict of interest: The authors declare that they have no relevant conflicts of interest.

Received for publication March 28, 2020; revised July 3, 2020; accepted for publication July 9, 2020.

Corresponding author: Liming Liang, PhD, Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, 655 Huntington Ave, Boston, MA, 02115. E-mail: lliang@hsph.harvard.edu.

0091-6749/$36.00

© 2020 American Academy of Allergy, Asthma & Immunology

https://doi.org/10.1016/j.jaci.2020.07.004

Furthermore, studies have also reported that mental health disorders, such as attention-deficit/hyperactivity disorder,[14] anxiety, and major depressive disorder, are a comorbidity of asthma.[15] Yet, studies have suggested that such associations are potentially bidirectional.[16,17] For example, anxiety can induce asthma symptoms,[16] whereas living with an asthma condition (eg, poor asthma control and worse asthma-related quality of life) may have mental health implications.[17] The exact mechanisms that underlie these mental health disorder–asthma associations remain uncertain.

ASTHMA HETEROGENEITY AND ITS RELATED GENETICS

Understanding the exact pathobiology of asthma involves several major challenges—the identification of causal mechanisms, the effect of multiple environmental factors (eg, diet, physical activity, air pollution, and environmental microbiome), and the heterogeneity of asthma itself. Although asthma had been considered a single disease for decades, a growing body of literature has revealed that asthma comprises a range of heterogeneous subtypes differing in presentation and disease course[18] and that the heterogeneity is based, at least partly, on different genetic factors for asthma subtypes (eg, childhood vs adult asthma, allergic vs nonallergic asthma).[12,19,20] Accordingly, examinations of subtype-specific genetics in conjunction with shared genetic factors between coexistent diseases or traits (eg, allergic diseases and obesity) and asthma should inform research on the heterogeneity in asthma and provide insight into corresponding pathology (Fig 1).[21,22]

The genetic effect of asthma is significant, with the heritability estimates ranging from 35% to 95%.[23] Genome-wide association studies (GWASs) have been widely applied to complex diseases for more than 2 decades, with a greatly increased sample size. However, according to Schoettler et al, in the GWAS of asthma, a larger sample size with heterogeneous subtypes is not necessarily better than a smaller sample size for homogeneous subtypes to identify the relevant genetic variants because the genetic background between asthma subtypes may be different.[24] Thus, investigating the shared genetic contribution to coexistent diseases or traits (eg, allergic disease, obesity) and specific asthma subtypes (eg, allergic asthma, obesity-associated asthma phenotype) would boost the power to detect subtype-specific variants that would have been masked by a traditional single-disease GWAS (Fig 1). A comprehensive characterization of these shared genetic architectures would improve understanding of the multiple dimensions of asthma pathobiology.

Traditionally, examining the phenotypic correlation or coexistence of other factors is a useful way to investigate the heterogeneity of asthma. However, this approach may have residual confounding and provide insufficient biologic insight as to which underlying mechanism(s) drive the association. A major advantage in going from phenotypic correlations to genetic correlations is improved understanding of the mechanism(s): shared genetic components can be identified at different levels from the whole genome to individual variants, providing insights into the reasons why asthma and coexistent diseases or traits are correlated. Furthermore, genetic correlations are less subject to confounding by environmental factors for several reasons. After adequate. control for population ancestry, genetic correlation would occur only if the germline genetic variant is causal or in linkage disequilibrium (LD) with the causal variant of both traits. A purely environmental confounding factor (eg, air pollution) would not lead to genetic correlation because it is not associated with any genetic variant (Fig 2, *A* and *B*). In contrast, if an environmental factor is an intermediary step between the genetic variant and the trait, it is in the causal pathway and is not considered a confounder (ie, it does not create a false genetic correlation between the 2 traits) (Fig 2, *C*). Population stratification is arguably the only confounding factor in GWAS, but it can be effectively controlled by using principal components from genome-wide genetic markers.[25] Once the genetic effect on diseases and traits has been robustly established, the genetic correlation between diseases and traits can be reliably estimated and replicated.[26-29] In the following sections, we will discuss a range of detailed analyses that can be used to compile a comprehensive investigation between asthma and other coexistent diseases or traits.

GENOME-WIDE CROSS-TRAIT ANALYSIS STUDY DESIGN

With the increasing availability of large-scale genetic data for asthma, such as the GABRIEL Consortium,[30] the Trans-National Asthma Genetic Consortium,[31] and the UK Biobank,[12,21,22,26,32] as well as the advancement of genetic epidemiology and statistical genetics methods, researchers are now able to examine the genetic contribution to the epidemiologic associations between asthma subtypes and those between coexistent diseases or traits and asthma. For example, to understand the genetics of asthma heterogeneity, 2 recent studies examined the genetic overlap between asthma subtypes—childhood asthma and adult asthma—by using the UK Biobank and 23andMe data.[19,20] They both found substantially shared (eg, *IL1RL1*, *HLA–DQA1*) but also distinct (eg, *ORMDL3* specific for childhood asthma) genetic components between these 2 subtypes, supporting the idea that the heterogeneity is related to distinct genetics.[19,20] These fundamental studies largely depend on single-trait analysis, and they can be further extended by our recently implemented study design called genome-wide cross-trait analysis, which is broadly applicable to asthma and many other diseases and/or traits. The design has been successfully applied to the UK Biobank and GWAS consortia data sets and has determined the shared genetic architectures between asthma and allergic diseases,[22] obesity,[12] and mental health disorders,[21] which were reproducible in other studies.[26-29] A genome-wide cross-trait analysis features several analyses: genetic correlation, cross-trait meta-analysis, Mendelian randomization, polygenic risk score, and GWAS functional analysis. Each component is discussed in more detail in subsequent sections and depicted in Fig 3. A glossary of the cross-trait GWAS terminology may be found in Table I. A summary of genome-wide cross-trait analysis methods may be found in Table II.

PLEIOTROPY AND CAUSALITY

Genetic pleiotropy refers to the same gene simultaneously influencing multiple traits.[33] A gene can be associated with more than 1 trait in 2 ways: horizontal pleiotropy and vertical pleiotropy (Fig 4). Horizontal pleiotropy, which is often simplified as pleiotropy, is defined as 1 genetic variant having independent effects on multiple traits. Vertical pleiotropy, which

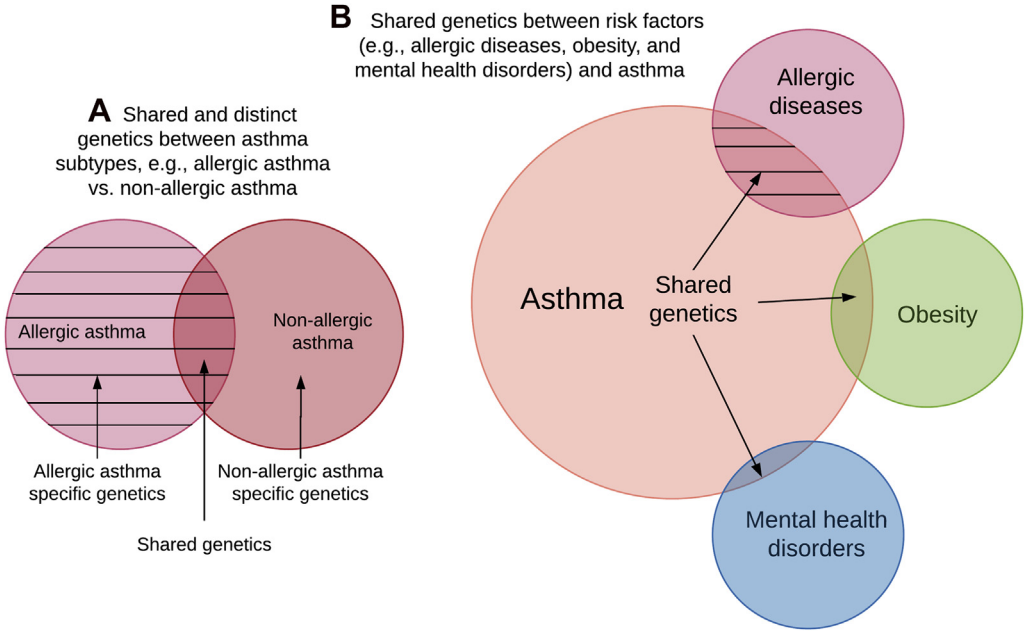

FIG 1. Venn diagram of shared and distinct genetics between asthma subtypes and those between coexistent diseases or traits and asthma. **A,** Shared and distinct genetics between asthma subtypes (eg, allergic asthma vs nonallergic asthma). **B,** Shared genetics between coexistent diseases/traits (eg, allergic diseases, obesity, mental health disorders) and asthma. The overlapping size between 3 example coexistent diseases/traits and asthma are based on the order of their genetic similarities, with allergic diseases sharing the most genetic components with asthma, followed by obesity and mental health disorders. **A** and **B**, The area covered by horizontal cross-lines indicates similar underlying genetic components/mechanisms in allergic asthma and shared genetics of allergic diseases and asthma.

is often called genetic causality, is defined as a genetic variant having an effect on a trait via its genetic effect on an intermediate trait. Identification of pleiotropy may improve the understanding and utility of disease-gene biology in multiple ways. First, there is a potential to detect the broad biologic impact of a gene, such as through phenome-wide association studies.[34] Second, if a pharmacologic genetic target could affect multiple traits or diseases, it might allow a drug developed for 1 disease to be repurposed for other diseases. For example, 3-hydroxy-3-methyl-glutaryl–coenzyme A reductase inhibitors, commonly known as statins, have also been found to have several molecular actions beyond cholesterol reduction,[35] such as reducing the development of type 2 diabetes.[36] Third, knowing the genetic causality can help develop disease prevention strategies via intervention in relation to nongenetic modifiable factors. For example, because the obesity-asthma genetic association suggests obesity-to-asthma effects, reduction of body mass index in patients with obesity might counteract the genetic effect, thereby potentially preventing the development of asthma. Therefore, distinguishing horizontal pleiotropy from vertical pleiotropy in cases in which both contribute to genetic correlations is important and can be challenging. We discuss methods for these analyses in the following sections on cross-trait meta-analysis and Mendelian randomization.

GENETIC CORRELATION

The genetic correlation between 2 traits (eg, A and B) measures the Pearson correlation between the genetic variant effect on traits A and B.[37,38] It could be the result of horizontal pleiotropic action of genes on trait A and B or a causal link between A and B. Genetic correlation indicates intrinsic correlations between 2 traits unaffected by environmental confounders, which are common and often unavoidable in conventional epidemiologic studies. Several methods have been developed to estimate genetic correlations, such as linkage disequilibrium score regression (LDSC/stratified-LDSC),[37] genome-wide complex trait analysis (GCTA/GCTA-linkage disequilibrium and minor allele frequency stratified-I),[38] and SumHer/BLD-LDAK from Linkage-Disequilibrium Adjusted Kinships (LDAK).[39] Among these, LDSC becomes one of the most commonly used methods for estimating genetic correlations because it uses GWAS summary statistics, which largely reduces computational burden and accounts for confounding in single-trait heritability (such as population stratification) and confounding in genetic correlation (such as shared study subjects).[22,40] The genetic correlation estimate (the Rg value) ranges from –1 to 1, where –1 indicates a perfect negative genetic correlation and 1 indicates a perfect positive genetic correlation. For example, Ferreira et al recently reported that the genetic correlation between childhood and adult asthma is 0.67 using same set of control samples, indicating both shared and distinct genetics between 2 asthma subtypes at the genome-wide level.[19] A recent study also conducted sensitivity analyses to examine the potential bias in LDSC due to overlapping subjects; the study used the following 4 scenarios while maintaining the same sample size: (1) no case overlap or control overlap, (2) case overlap and no control overlap, (3) no case overlap and control overlap, and (4) case overlap and control overlap. The sensitivity analyses showed that the Rg estimate from LDSC is unbiased to overlapping cases and/or controls.[22]

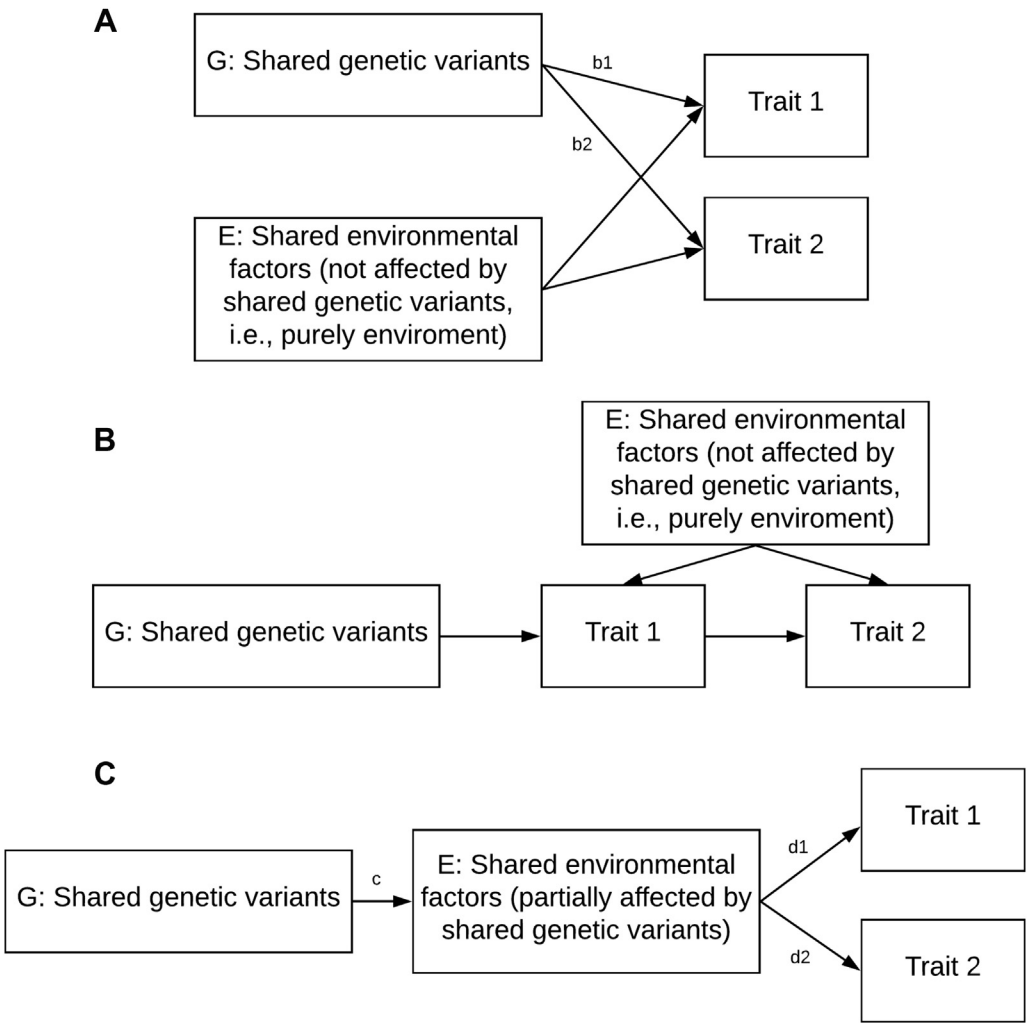

FIG 2. Directed acyclic graphs of relationship between shared genetic or environmental factors with traits. **A,** Shared environment factors (not affected by shared genetic variants) will not bias genetic correlation. After appropriately control for population ancestry, the genetic effects *b1* and *b2* are unrelated to *E*, and therefore the genetic correlation—correlation *b1* and *b2*—is not related to *E*. **B,** Another situation in which shared environment factors (not affected by shared genetic variants) will not bias genetic correlation. **C**, Shared environment factors (partially affected by shared genetic variants). In this case, *E* is not considered a confounder; rather, it is considered a mediator in the causal pathway of interest; *c* represents the effect of shared genetic variants on environmental factors, and *d1* and *d2* represent the effect of shared environmental factors on traits.

CROSS-TRAIT META-ANALYSIS

Genetic correlation depicts the genome-wide average sharing of genetic effect between traits. To identify genetic variants with pleiotropic effects, cross-trait GWASs have used a range of conventional and recently developed meta-analysis methods. Typically, summary statistics of distinct but potentially related traits are combined in a meta-analysis framework to detect specific loci with shared associations. Such univariate approaches do not require access to individual-level genotype data and thus are readily applicable to existing GWAS results. Combining results across studies of different traits also improves the statistical power of detecting modest cross-trait genetic effects that may not have reached genome-wide significance for any single trait.

Many cross-trait meta-analysis methods are available.[41] Two specific methods, association analysis based on SubSETs (ASSET)[42] for binary traits and cross phenotype association (CPASSOC)[43] for continuous or binary traits, have been shown to outperform a range of alternatives for detecting pleiotropy effects shared in all traits or a subset of traits.[41] ASSET combines an all-subsets fixed-effects GWAS meta-analysis with a bayesian method to evaluate the best-fit configuration of genotype-phenotype associations.[42] Similar to the ASSET, CPASSOC assumes that effects may exist only within a subset of traits.[43] This feature is not only useful for detecting overall pleiotropy but is also important for detecting subset-specific effect (eg, variants shared with only some but not all asthma subtypes because of heterogeneity). Unlike ASSET, CPASSOC identifies the subset of studies with effects by sequentially adding a trait by an incremental order of their association significance. Among the sequentially examined subsets, the one with the highest meta-statistics is selected. Both methods are designed to correct for inflation due to overlapping subjects. However, our previous

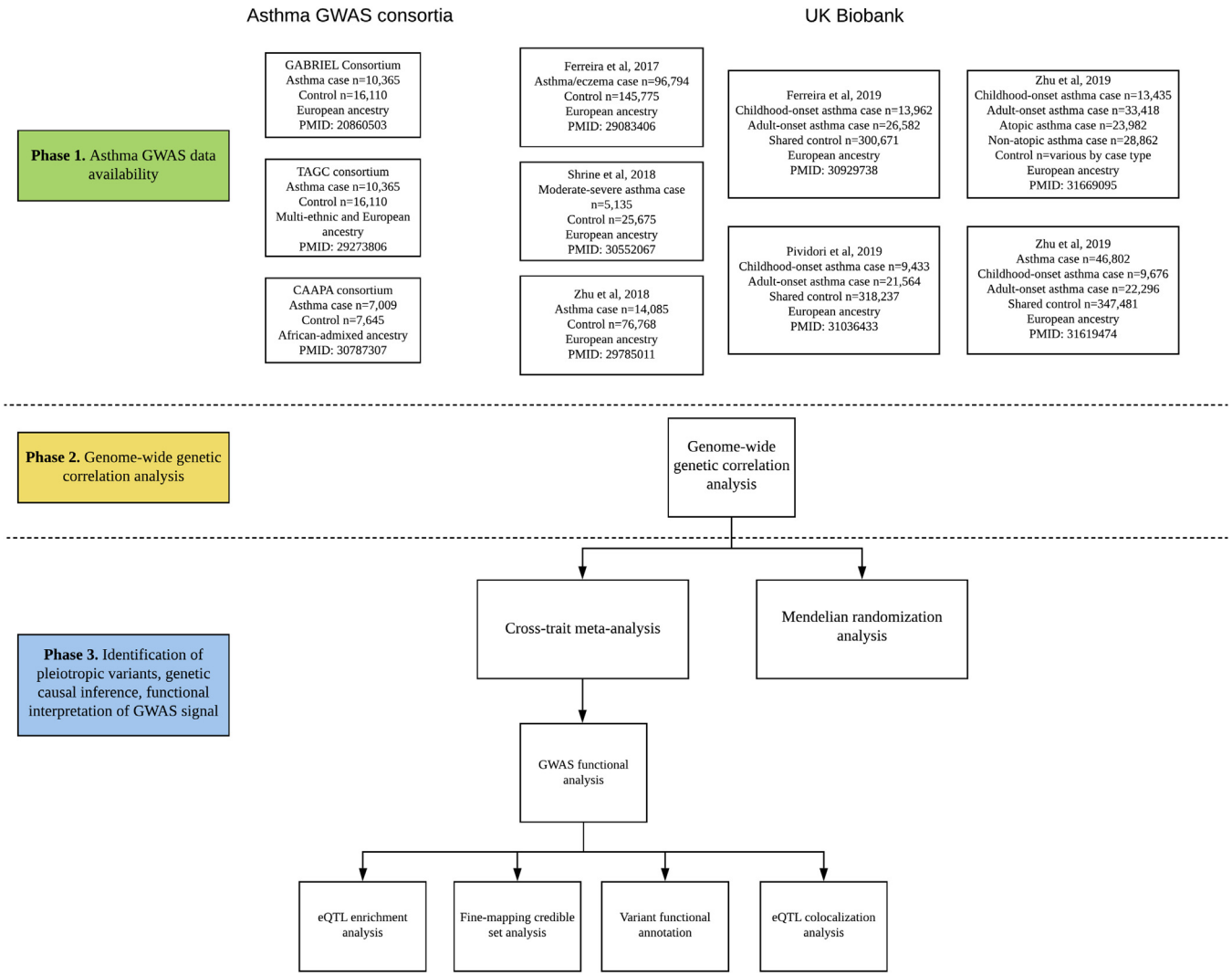

FIG 3. Data availability for GWAS of asthma and study design of genome-wide cross-trait analysis. Genome-wide genetic correlation analysis is used to examine the genetic correlation between a pair of traits by using genome-wide SNPs. Cross-trait meta-analysis is used to determine the shared genetic variants between multiple traits. Mendelian randomization analysis is used to examine the causal effect of the exposure trait on the other trait by using the genetic variants for exposure trait as the instrument variables. eQTL enrichment analysis is used to determine the enrichment of genetic variants associated with complex traits in eQTL. Fine mapping credible set analysis is used to examine whether there is a potential causal variant at each locus. Variant functional annotation is used to predict the functional effect of an individual SNP on a transcript. eQTL colocalization analysis is used to determine the shared causal variants between GWAS signals and eQTL signals. *CAAPA*, Consortium on Asthma among African-ancestry Populations in the Americas; *TAGC*, Trans-National Asthma Genetic Consortium.

simulation analysis showed that caution is advised for CPASSOC when almost all controls were shared between studies.[41] Another caveat is that it is sometimes difficult to retrieve the exact number of overlapping subjects between 2 studies, which is required for ASSET. A recently developed method called multitrait analysis of GWAS[44] describes a promising strategy to address the problem by using LDSC.[37] However, this method assumes that all single-nucleotide polymorphisms (SNPs) share the same variance-covariance matrix among all traits, which could be violated when some SNPs are associated only with a subset of traits.[44]

In a previous cross-trait analysis of allergic diseases and asthma,[22] ASSET was used to determine 38 shared loci, some of which (such as variations in *EVI5*, *NRROS,* and *C11orf30*) are newly discovered from the cross-trait meta-analysis. In the cross-trait analysis of mental health disorders and asthma,[21] the use of ASSET identified 7 loci that are jointly associated with attention-deficit/hyperactivity disorder and asthma, 1 locus that is jointly associated with anxiety disorder and asthma, and 10 loci that are jointly associated with major depressive disorder and asthma. Of note, the *HLA* region (chromosome 6, 25-34 Mb) was found to be shared in the cross-trait meta-analysis of allergic disease and asthma and in that between major depressive disorder and asthma. The *HLA* region was commonly reported to have important pleotropic effects.[45] However, because of its high gene density and extensive LD, GWAS signals within the *HLA*

TABLE I. Glossary of terms related to genome-wide cross-trait analysis

Term	Definition
Cross-trait meta-analysis	A meta-analysis testing the null hypothesis that none of the traits being examined is associated with the genetic variant. One genetic variant is tested at a time.
eQTLs	Genetic variants that are associated with the gene expression levels.
Genetic correlation	Assuming that all genetic variants have some effect on a trait and that their effect size follows a gaussian distribution (called the infinitesimal model), the genetic correlation between 2 traits (A and B) measures the Pearson correlation between the genetic variant effect on traits A and B.
GWAS	An analytic method that tests the association between each genetic variant and a specific phenotype (a disease status or a quantitative trait). One genetic variant is tested at a time.
HLA/MHC region	A genomic region of an approximately 3.6-Mb genome sequence located on the chromosome 6p21, which is mainly known for its pervasive pleiotropic effect and immune-related function. The extended MHC region is at 25 to 34 Mb on chromosome 6.
Horizontal pleiotropy	A genetic variant or gene having independent effects on multiple traits that do not have a causal effect on each other.
Instrumental variables	Variables that are associated with the modifiable exposure or risk factor of interest and affect the outcome only through the exposure or risk factor.
Mendelian randomization	An analytic approach that examines the causality of an observed association of a modifiable exposure or risk factor with an outcome of interest by using ≥1 genetic instrumental variables.
Polygenic risk score	A score based on a set of disease and/or trait-associated genetic variants, commonly defined as the weighted sum of their genotypes. Weights are chosen by their association effect on the disease and/or trait, directly from GWAS or further modified on the basis of a suitable statistical model incorporating all genetic variants on the genome.
Vertical pleiotropy (genetic causality)	A genetic variant or gene having an effect on a trait that has causal effect on another trait.

region are difficult to map in fine resolution. Thus, a considerable amount of work, which was recently developed in the sequencing of the entire *HLA* region,[46] has established its central role in the biology of the immune system and in predisposition to a large number of inflammatory diseases. However, the extent to which these diseases share the same causal risk variants or genes remains unclear. Finally, in the cross-trait analysis of obesity and asthma, CPASSOC has been used to determine potential shared loci and variations in *MYL6* and *ACOXL* and replicated them in a mouse model.[12]

Many cross-trait effects are not surprising. For example, variants in the *C11orf30*, *IL1R1*, and *FLG* genes, as well as in the *HLA* region, are associated with allergic diseases and asthma,[22] as was found in previous studies with individual traits.[26] Others are perhaps less intuitive and can shed light on hitherto unknown connections between traits. For example, variants in the *POLI* gene have been shown to affect risk of both major depressive disorder and asthma. These seemingly unrelated diseases share pathways involved in immune processes.[21] These results demonstrated a key benefit of cross-trait meta-analysis, namely, discovery of shared loci that have not been reported as having genome-wide significance ($P < 5 \times 10^{-8}$) from the individual-trait GWAS. As another example, whereas the *RERE* gene was found to be associated only with asthma in the single-trait GWAS,[22] the top variant, rs301817, within *RERE* gene was found to be shared by major depressive disorder and asthma,[21] although the biologic role of *RERE* between asthma and major depression needs further exploration.

MENDELIAN RANDOMIZATION

When 2 traits are known to be correlated (phenotypically or genetically), identification of causality between the traits is imperative. Here, GWAS data for these traits provide a unique opportunity to make robust causal inference by using GWAS summary statistics without requiring individual-level data for both traits in the same set of subjects. Mendelian randomization refers to an analytic approach that uses the genetic variants for 1 trait (exposure [eg, body mass index]) as the instrument variables to examine the causal effect of the exposure trait on the other trait (outcome [eg, asthma]). This method is based on the theory that the germline alleles of these variants are randomly allocated, which can be seen analogously to the randomized treatment assignment in a randomized controlled trial resulting in an unconfounded exposure-outcome relationship.[47] It provides a valuable tool—especially when randomized controlled trials are not feasible and observational studies provide a biased estimate for causal effects because of unmeasured confounding, model misspecification, or reverse causation. Of note, the validity of Mendelian randomization depends on 3 key assumptions: (1) the relevance assumption, (2) the independence assumption, and (3) the exclusion restriction, which have been thoroughly described in a previous review.[47] Briefly, relevance assumption refers to the genetic variants that are strongly associated with the exposure of interest; independence assumption refers to the effect of genetic variants on the outcome of interest that is unconfounded; and exclusion restriction refers to the genetic variants that affect the outcome only through the exposure of interest.

With the availability of large-scale GWASs, it has become easier to find genetic variants that are strongly associated with the exposure trait (satisfying the relevance assumption). Although population stratification is arguably the only potential confounder between the genetic variants and the outcome, it can be effectively accounted for in standard GWAS analysis. Accordingly, the independence assumption is not difficult to satisfy. The exclusion restriction assumption is the nonexistence of other pathway(s) from the genetic variant to the outcome except through the exposure. This assumption is difficult to verify and hence has led

TABLE II. Summary of genome-wide cross-trait analysis methods

Analysis method	Software	Advantages	Disadvantages	Examples of application in asthma or complex traits/PMID
Genetic correlation	LDSC/S-LDSC	Requires only GWAS summary statistics; computationally efficient; accounts for additive confounding in single-trait heritability (such as population stratification) and confounding in genetic correlation (such as overlapping samples); can allow relatively flexible heritability architecture in MAF, LD, and functional categories (the authors LDSC recommended S-LDSC)	Is sensitive to other genetic architectures not captured by the baseline LD model; requires that the reference panel LD and GWAS summary statistics be computed from the same population	Childhood asthma and adult asthma/30929738, allergic diseases and asthma/29785011, obesity and asthma/31669095, mental health disorders and asthma/31619474
	GCTA/GCTA-LDMS	Estimates genetic correlation with high accuracy; the LD and association effect are computed from the same genotype data; accounts for different genetic architectures by MAF and LD categories (the authors of GCTA recommended GCTA–LDMS-I)	Requires genotype data; computation is infeasible for an extremely large data set	Complex traits/ 21167468
	SumHer/BLD-LDAK	Is similar to LDSC but assumes a specific parametric model for MAF/LD-dependent genetic architecture and multiplicative inflation bias due to population stratification or family relatedness; can allow the same baseline LD categories as in LDSC (the authors of LDAK recommended BLD-LDAK/BLD-LDAK-alpha)	Is sensitive to other genetic architecture deviated from the assumed parametric model; requires that the reference panel LD and GWAS summary statistics be computed from the same population	Complex traits/ 32203469
Cross-trait meta-analysis	ASSET	Accounts for overlapping samples	Is applicable only to binary traits	Allergic diseases and asthma/29785011, mental health disorders and asthma/31619474
	CPASSOC	Is applicable to both binary and continuous traits	Yields potential false positives due to overlapping samples	Obesity and asthma/31669095
	MTAG	Accounts for possibly unknown sample overlap	Requires an assumption that all variants share the same genetic correlation across all traits (ie, no subset-specific effect is assumed)	Complex traits/29292387
Mendelian randomization	Inverse variance–weighted approach	Is applicable when the genetic variants' pleiotropic effects (genetic variant–outcome direct effect) happen to cancel out	Accounts for only the designed scenario; requires independent variants	Asthma and cancer/32006205
	Egger regression	Is applicable when the genetic variant–exposure association is independent of the pleiotropic effect; appears to protect false positives in several simulation studies	Accounts for only the designed scenario; requires independent variants; when the outcome GWAS is low-power, its power to detect causal effect could be substantially smaller than that of other methods	Asthma and cancer/32006205
	Weighted median estimator	<50% (counts or total weights) of the genetic variants are invalid instruments	Accounts for only the designed scenario	Asthma and cancer/32006205
	Weighted mode-based estimate (weighted MBE)	Is applicable when the variants satisfying the exclusion restriction assumption give a causal effect estimate that is the majority among the effect estimates from all variants in the analysis; appears to protect false positives in several simulation studies	Accounts for only the designed scenario	Asthma and cancer/32006205

(Continued)

TABLE II. *(Continued)*

Analysis method	Software	Advantages	Disadvantages	Examples of application in asthma or complex traits/PMID
	GSMR	Accounts for LD between variants; detects and accounts for outliers that could violate the exclusion restriction assumption	Requires sufficient numbers of GWAS significant variants; requires a genetic variant–exposure association that is independent of the pleiotropic effect	Obesity and asthma/31669095, mental health disorders and asthma/31619474
	MR-PRESSO	Detects and accounts for outliers that could violate the exclusion restriction assumption	Requires independent and sufficient numbers of GWAS significant variants; requires a genetic variant–exposure association that is independent of the pleiotropic effect	Asthma and cancer/32006205
	LCV	Distinguishes genetic correlation from genetic causality; uses all genome-wide variants	Does not estimate causal effect size, but does provide a scale parameter with higher magnitude indicating that it is closer to causality; requires that the LD reference match the study populations	Obesity and asthma/31669095
	CAUSE	Allows genetic correlation and genetic causality for different variants; can estimate causal effect size; uses all genome-wide variants; has better power avoid false positive trade-off than other methods	Requires independent variants by pruning GWAS results; may have a higher than expected false-positive rate than Egger regression and MBE when a large fraction of variants affect the exposure and outcome through a strong shared factor; in addition, the power for both exposure and outcome GWAS is high	Complex traits/32451458
Polygenic risk score	LDpred	Accounts for LD between SNPs	Is computation-intensive when the number of SNPs is more than a couple of million	Asthma/32522462
	MTGBLUP	Simultaneously estimates genetic effect and genetic correlation for multiple traits	Assumes an infinitesimal genetic architecture. Is computation-intensive when the number of SNPs is more than a couple of million and becomes prohibitive for more than hundreds of thousands of samples	Complex traits/25640677
	MTAG	Provides improved polygenic prediction thanks to a consistent estimator, and its effect estimates always have a lower genome-wide mean squared error than the corresponding single-trait GWAS estimates do	Requires an assumption that all variants share the same genetic correlation across all traits (ie, no subset-specific effect is assumed); power might be reduced when the genetic correlations between traits are not very high	Complex traits/29292387
	CTPR	Optimizes the prediction accuracy for the primary trait of interest by taking advantage of shared genetic effects among multiple traits; secondary traits of GWAS can be individual-level data or summary statistics	Requires individual-level data for the primary trait of interest	Complex traits/30718517

BLD, Background linkage disequilibrium; *CAUSE*, causal analysis using summary effect estimates; *CTPR*, cross-trait penalized regression; *GCTA*, genome-wide complex trait analysis; *GSMR*, generalized summary data–based Mendelian randomization; *LCV*, latent casual variable; *LDMS-I*: linkage disequlibrim and minor allele frequency stratified-I; *MAF*, minor allele frequency; *MR-PRESSO*, Mendelian randomization pleiotropy residual sum and outlier; *MTAG*, multitrait analysis of GWAS; *MTGBLUP*: multi-trait genomic best linear unbiased prediction; *PMID*, PubMed identifier; *S-LDSC*, stratified LDSC; SumHer, single-nucleotide polymorphism heritability.

to the development of various Mendelian randomization methods that relax this assumption in 3 ways: (1) by using less stringent assumptions, such as an inverse variance weighted approach,[48] Mendelian randomization–Egger regression,[49] a weighted median estimator,[50] and a mode-based estimate[51]; (2) by removing genetic variants that violate the assumption, such as generalized summary data–based Mendelian randomization[52] and Mendelian randomization pleiotropy residual sum and

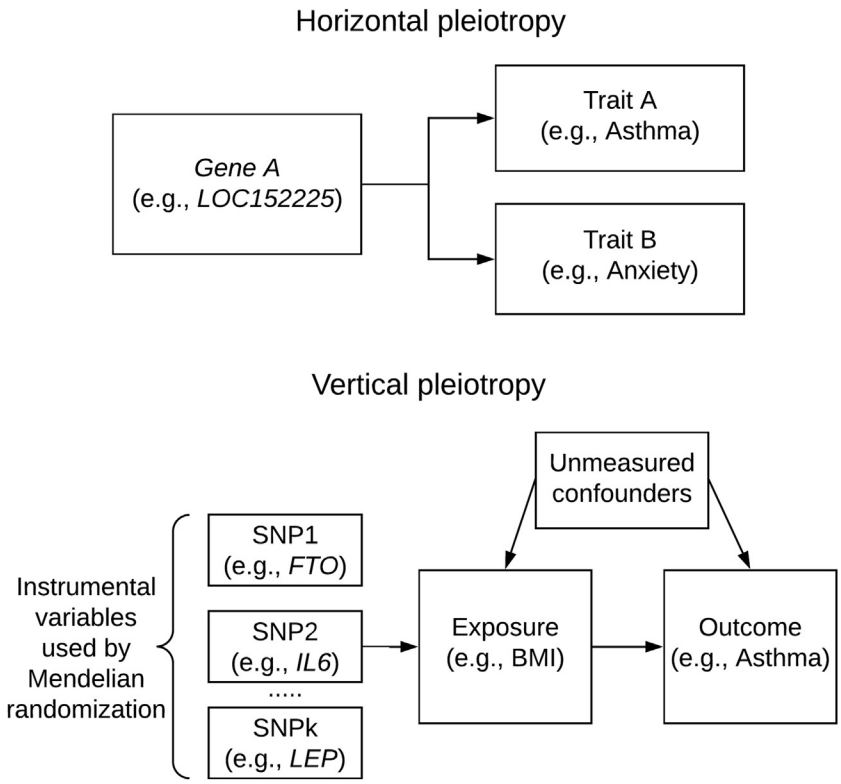

FIG 4. Diagram of horizontal pleiotropy and vertical pleiotropy with examples for asthma. *BMI,* Body mass index.

outlier[53]; and (3) by estimating the likelihood from pure horizontal pleiotropy to pure causality, such as the latent causal variable[54] and causal analysis using summary effect estimates.[55] The major features of each method can be found in Table II. Users can consider the likely pathobiology of the exposure and outcome to choose the most appropriate assumption and pick the most appropriate method. Often, however, which assumptions have been met is not clear. In such situations, we suggest conducting the Mendelian randomization by using multiple methods as a sensitivity analysis.[56] If multiple methods arrive at a similar conclusion, then we would consider the results robust. However, when the conclusions reached by using these methods do not agree, we recommend close comparisons of their results, including for consistency of the direction for causal effect estimates, the magnitude of the effect, and the statistical significance (related to the power and sample size). Later methods (eg, latent causal variable, causal analysis using summary effect estimates) might address some of the drawbacks of the methods developed earlier (eg, the inverse variance–weighted method). Thus, their disagreement would become more interpretable. Accordingly, we recommend cautious interpretation of the results when different methods yield discordant results.

POLYGENIC RISK SCORE

A genetic risk score aggregates genetic variants aiming to predict disease risk or trait level, and it can thereby help guide early prevention, targeted intervention, and characterization of subtypes for complex diseases (including asthma).[57] With the arrival of large-scale publicly available GWAS summary statistics, a polygenic risk score integrating genome-wide variants regardless of statistical significance is a promising approach to realize its clinical utility and potential public health benefit.[58,59] Among the many available polygenic risk models, those that are widely used include LD pruning followed by *P* value thresholding (probability + threshold) and LDpred, a bayesian framework that estimates posterior mean causal effect sizes from GWAS summary statistics by assuming a prior for the genetic architecture and LD information from a reference panel.[60] This method was recently applied to investigate the association between the asthma polygenic risk scores and COVID-19.[61] More recent methods have further improved prediction accuracy by combining effects across multiple genetically related traits; these methods include multi-trait genomic best linear unbiased prediction,[62] multitrait analysis of GWAS,[44] and cross-trait penalized regression (which attempts to optimize the prediction accuracy for the primary trait of interest by taking advantage of shared genetic effects among multiple traits through a multivariate penalized least-squares method). These methods can use either GWAS summary statistics or individual-level genotype data and can thus provide an opportunity to integrate many GWAS results.[63] The cross-trait design has demonstrated an advantage over a single-trait design in terms of prediction accuracy and model calibration.[63] One limitation of most of the current polygenic risk score methods is the generalizability of the algorithms to non-European racial/ethnic groups, which may exacerbate health disparities.[64] Large-scale application of polygenic risk scores in asthma studies is still ongoing and remains to be seen

in future publications, especially in studies involving non-European racial/ethnic groups.

CROSS-TRAIT GWAS FUNCTIONAL ANALYSIS

Cross-trait meta-analysis methods are helpful to identify novel genetic loci that are associated with multiple diseases and traits. However, the functional impact of the shared loci on disease risks or traits remain to be clarified. In the past decade, there has been development of large-scale studies for genomic functions annotation, including the Encyclopedia of DNA Elements[65] and National Institute of Health Roadmap Epigenomics Project,[66] which can be readily used to categorize the functions for GWAS-discovered loci. Another useful source of functional annotation is use of expression quantitative trait loci (eQTLs), namely, the discovery of genetic variants that explain the variations in gene expression levels, providing annotation to the regulatory regions outside coding sequence.

As of now, several eQTL databases are publicly available for GWAS functional analysis. The databases that are especially useful for asthma and other inflammatory diseases include the Genotype-Tissue Expression (GTEx) project[67] and Immunological Genome Project (ImmGen).[68,69] The GTEx project is a resource database for studying the relationship between genetic variation and gene expression in different human tissues.[67] GTEx version 8 contains 54 types of human tissues and 17,382 RNA-sequencing samples from 948 donors. Among these, 49 tissues have eQTL data. The application of GTEx eQTL in asthma can be useful. For example, in a cross-trait analysis, the use of GTEx revealed that shared genes between allergic diseases and asthma (eg, *FLG*) are most significantly enriched in epithelial tissues, such as skin.[22] Traditionally, researchers look to eQTL databases to identify gene expression associated with disease variants. However, a recent GTEx study has also pointed out that nearly all common variations show some associations with expression of at least 1 gene in at least 1 tissue and that care is required when using their eQTLs as demonstration of a causal gene.[67] Thus, we also recommend colocalization analysis[70] to determine the shared causal variants between GWAS signals and eQTL signals. Colocalization analysis[70] has adapted a bayesian model to compute the posterior probabilities for 4 scenarios that the 2 traits sharing causal variants at a specific locus: (1) no causal variant exists for either trait, (2) causal variants exist for only 1 trait, (3) causal variants exist for both traits, but they are different variants, and (4) causal variants exist for both traits, and they are the same variants. For example, in the cross-trait meta-analysis of allergic disease and asthma,[22] the variant rs34290285 in the 2q14 region was related to both *GAL3ST2* and *D2HGDH* genes, with both genes showing eQTL signals in multiple tissues in GTEx. However, the colocalization analysis has shown that asthma might share causal variants at the rs34290285 locus with *GAL3ST2* and in multiple tissues of GTEx but not with *D2HGDH*.

In addition to GTEx, other publicly available eQTL resources focus on a specific area, such as the immune-related eQTL database ImmGen.[68] The ImmGen contains gene expression data of 292 specific immune cell types from mice. It has established the baseline measurements of variations in the hematopoietic transcriptome that allow for many eQTL analyses. For example, Finucane et al recently developed a method based on LDSC to investigate heritability enrichment in specific eQTL data,[71] including ImmGen. They found that both eczema and asthma exhibited enrichment in T_H and natural killer T cells but not in other immune cell types. In addition to eQTL, quantitative trait locus mapping for other functional data such as methylation[69] would also provide important functional insight into the shared genetic variants discovered.

In addition to being applicable in eQTL analyses, genome-wide cross-trait analysis is also applicable to other types of functional investigation. For example, in addition to identifying shared causal variants between disease GWAS signals and functional data as already mentioned, colocalization analysis is also applicable to cross-disease analysis. For example, a recent study showed that 33 loci share causal variants (scenario 4) between allergic diseases and asthma.[22] These shared loci can then be prioritized for the further functional investigation and pathway analysis[72] to identify shared biologic pathways. For analysis involving more than 2 intermediate traits, it is possible to examine the mediating role of these intermediate traits between exposure and disease by using GWAS summary statistics based on 2-step Mendelian randomization.[73]

DISCUSSION AND FUTURE DIRECTIONS

Genome-wide cross-trait analysis offers opportunities to investigate shared genetics among complex traits by using large-scale publicly available GWAS data. In this review, we have summarized the current status of cross-trait genetics studies for the shared and distinct genetic effects between asthma subtypes and their shared genetics with coexistent diseases or traits.[12,19-22,26-29,31] We have surveyed a broad range of major analytic methods at each phase of the genome-wide cross-trait analysis, namely, genetic correlation, cross-trait meta-analysis, Mendelian randomization, polygenic risk score, and functional analysis. We have also discussed scientific goals for each phase, as well as the advantages and limitations for these methods. Major challenges to future genome-wide cross-trait studies involve the availability of comprehensive and consistent phenotype data, combining data from different GWAS imputation panels, cross-ethnic genetic analysis, and integration of multiomics data sets. We discuss these challenges and new opportunities in the following paragraph.

First, although we listed several publicly available asthma GWAS data in Fig 3, other uncommon phenotypes may not be available in large-scale GWASs, such as food allergy, which may also contains subtypes.[74] Additionally, consistency between the same phenotype across different studies is a challenge. For example, the definition of childhood and adult asthma in the study by Ferreira et al[19] is different from that in the study by Pividori et al.[20] Such inconsistency in phenotype definition may lead to different interpretations of cross-trait results. Second, there are challenges in combining data from different GWAS platforms and imputation panels used. We note that this is a major limitation for many cross-trait meta-analysis methods. For example, conducting cross-trait meta-analysis between a GWAS imputation based on 1000 Genome and HapMap panels will usually lead to use of only common SNPs between the 2 panels, which may greatly reduce the chance of identifying cross-trait signals. Thus, standardization and imputation of GWASs is strongly recommended before cross-trait analysis. Third, most asthma

GWASs have been performed primarily in European populations despite the widely recognized racial/ethnic disparities in asthma.[75] It is important to extend asthma GWASs to non-European racial/ethnic groups, thereby providing opportunities to gradually map the causal variants for the diseases and traits by leveraging population specific LD, develop a more accurate polygenic risk prediction model, and provide population-matched GWAS summary statistics for 2-sample Mendelian randomization. Several asthma GWASs in other non-European populations are ongoing. For example, one of the largest non-European asthma GWASs was recently conducted by the Consortium on Asthma among African-Ancestry Populations in the Americas, which identified 2 loci that are potentially specific to asthma risk in the population with African ancestry.[76] Yet the sample size and diversity in population ethnicity are still far from sufficient. Fourth, GWASs have successfully uncovered genetic variants related to asthma in the past decades.[12,19-23,26,31,32] However, these variants alone account for a limited proportion of the phenotypic variance in asthma.[77] The total additive heritability provides an upper bound estimate for how much asthma risk can be explained by the genetics. This has generated interest in alternative sources of phenotypic variance (eg, host-microbiome-environment interrelationships).[78] Integrating GWAS with other omics data, such as epigenomics, transcriptomics, proteomics, metabolomics, and microbiomics data, will provide a potential for defining the pathobiology of asthma and its subtypes. Additionally, we also note that to develop novel prevention and treatment strategies, it is imperative to apply these approaches integrated with causal inference methods to multiomics data. Examples of these causal inference methods are cross-trait methods (including Mendelian randomization, as presented in this review), causal mediation methods,[79] and causal structure learning.[80]

In summary, this review has illustrated the utility of large-scale genetic data coupled with advanced statistical genetic tools to understand the shared and distinct components between asthma subtypes as well as between coexistent diseases or traits and asthma. The success of current cross-trait studies in asthma also suggests useful applications to other complex and heterogeneous traits beyond asthma, including diseases of major interest to the allergy and immunology community, such as atopic dermatitis, food allergy, and allergic rhinitis.

REFERENCES

1. Gobal Initiative for Asthma. Global strategy for asthma management and prevention. 2018. Available at: www.ginasthma.org. Accessed December 19, 2019.
2. Subbarao P, Mandhane PJ, Sears MR. Asthma: epidemiology, etiology and risk factors. CMAJ 2009;181:E181-90.
3. de Groot EP, Duiverman EJ, Brand PL. Comorbidities of asthma during childhood: possibly important, yet poorly studied. Eur Respir J 2010;36:671-8.
4. Priftis KN, Mermiri D, Papadopoulou A, Papadopoulos M, Fretzayas A, Lagona E. Asthma symptoms and bronchial reactivity in school children sensitized to food allergens in infancy. J Asthma 2008;45:590-5.
5. Plaschke P, Janson C, Norrman E, Bjornsson E, Ellbjar S, Jarvholm B. Association between atopic sensitization and asthma and bronchial hyperresponsiveness in Swedish adults: pets, and not mites, are the most important allergens. J Allergy Clin Immunol 1999;104:58-65.
6. Vink NM, Postma DS, Schouten JP, Rosmalen JG, Boezen HM. Gender differences in asthma development and remission during transition through puberty: the tracking adolescents' individual lives survey (trails) study. J Allergy Clin Immunol 2010;126:498-504.e1-6.
7. Macsali F, Real FG, Plana E, Sunyer J, Anto J, Dratva J, et al. Early age at menarche, lung function, and adult asthma. Am J Respir Crit Care Med 2011;183:8-14.
8. Larsen GL. Differences between adult and childhood asthma. J Allergy Clin Immunol 2000;106:S153-7.
9. Simons FER. What's in a name? The allergic rhinitis–asthma connection. 2003;3:9-17.
10. Peters U, Dixon AE, Forno E. Obesity and asthma. J Allergy Clin Immunol 2018;141:1169-79.
11. Camargo CA Jr, Weiss ST, Zhang S, Willett WC, Speizer FE. Prospective study of body mass index, weight change, and risk of adult-onset asthma in women. Arch Intern Med 1999;159:2582-8.
12. Zhu Z, Guo Y, Shi H, Liu CL, Panganiban RA, Chung W, et al. Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK biobank. J Allergy Clin Immunol 2020;145:537-49.
13. Granell R, Henderson AJ, Evans DM, Smith GD, Ness AR, Lewis S, et al. Effects of BMI, fat mass, and lean mass on asthma in childhood: a Mendelian randomization study. PLoS Med 2014;11:e1001669.
14. Cortese S, Sun S, Zhang J, Sharma E, Chang Z, Kuja-Halkola R, et al. Association between attention deficit hyperactivity disorder and asthma: a systematic review and meta-analysis and a swedish population-based study. Lancet Psychiatry 2018;5:717-26.
15. Scott KM, Von Korff M, Ormel J, Zhang MY, Bruffaerts R, Alonso J, et al. Mental disorders among adults with asthma: results from the world mental health survey. Gen Hosp Psychiatry 2007;29:123-33.
16. Chen E, Miller GE. Stress and inflammation in exacerbations of asthma. Brain Behav Immun 2007;21:993-9.
17. Lavoie KL, Cartier A, Labrecque M, Bacon SL, Lemiere C, Malo JL, et al. Are psychiatric disorders associated with worse asthma control and quality of life in asthma patients? Respir Med 2005;99:1249-57.
18. Borish L, Culp JA. Asthma: a syndrome composed of heterogeneous diseases. Ann Allergy, Asthma Immunol 2008;101:1-9.
19. Ferreira MAR, Mathur R, Vonk JM, Szwajda A, Brumpton B, Granell R, et al. Genetic architectures of childhood- and adult-onset asthma are partly distinct. Am J Hum Genet 2019;104:665-84.
20. Pividori M, Schoettler N, Nicolae DL, Ober C, Im HK. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. Lancet Respir Med 2019;7:509-22.
21. Zhu Z, Zhu X, Liu CL, Shi H, Shen S, Yang Y, et al. Shared genetics of asthma and mental health disorders: a large-scale genome-wide cross-trait analysis. Eur Respir J 2019;54.
22. Zhu Z, Lee PH, Chaffin MD, Chung W, Loh PR, Lu Q, et al. A genome-wide cross-trait analysis from UK Biobank highlights the shared genetic architecture of asthma and allergic diseases. Nat Genet 2018;50:857-64.
23. Ober C, Yao TC. The genetics of asthma and allergic disease: a 21st century perspective. Immunol Rev 2011;242:10-30.
24. Schoettler N, Rodriguez E, Weidinger S, Ober C. Advances in asthma and allergic disease genetics - is bigger always better? J Allergy Clin Immunol 2019.
25. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. Nat Rev Genet 2010;11:459-63.
26. Ferreira MA, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. Nat Genet 2017;49:1752-7.
27. Pickrell JK, Berisa T, Liu JZ, Segurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. Nat Genet 2016;48:709-17.
28. Melen E, Himes BE, Brehm JM, Boutaoui N, Klanderman BJ, Sylvia JS, et al. Analyses of shared genetic factors between asthma and obesity in children. J Allergy Clin Immunol 2010;126:631-7.e1-8.
29. Lehto K, Pedersen NL, Almqvist C, Lu Y, Brew BK. Asthma and affective traits in adults: a genetically informative study. Eur Respir J 2019;53.
30. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 2010;363:1211-21.
31. Demenais F, Margaritte-Jeannin P, Barnes KC, Cookson WOC, Altmuller J, Ang W, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. Nat Genet 2018;50:42-53.
32. Shrine N, Portelli MA, John C, Soler Artigas M, Bennett N, Hall R, et al. Moderate-to-severe asthma in individuals of european ancestry: a genome-wide association study. Lancet Respir Med 2019;7:20-34.
33. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. Nat Rev Genet 2013;14:483-95.
34. Denny JC, Bastarache L, Roden DM. Phenome-wide association studies as a tool to advance precision medicine. Annu Rev Genomics Hum Genet 2016;17:353-73.
35. Takemoto M, Liao JK. Pleiotropic effects of 3-hydroxy-3-methylglutaryl coenzyme a reductase inhibitors. Arterioscler Thromb Vasc Biol 2001;21:1712-9.

36. Tonolo G, Melis MG, Formato M, Angius MF, Carboni A, Brizzi P, et al. Additive effects of simvastatin beyond its effects on LDL cholesterol in hypertensive type 2 diabetic patients. Eur J Clin Invest 2000;30:980-7.
37. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 2015; 47:1236-41.
38. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011;88:76-82.
39. Speed D, Holmes J, Balding DJ. Evaluating and improving heritability models using summary statistics. Nat Genet 2020;52:458-62.
40. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 2015;47:291-5.
41. Zhu Z, Anttila V, Smoller JW, Lee PH. Statistical power and utility of meta-analysis methods for cross-phenotype genome-wide association studies. PLoS One 2018;13:e0193256.
42. Bhattacharjee S, Rajaraman P, Jacobs KB, Wheeler WA, Melin BS, Hartge P, 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 2012; 90:821-35.
43. Zhu X, Feng T, Tayo BO, Liang J, Young JH, Franceschini N, et al. Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. Am J Hum Genet 2015;96:21-36.
44. Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. Nat Genet 2018;50:229-37.
45. Sivakumaran S, Agakov F, Theodoratou E, Prendergast JG, Zgaga L, Manolio T, et al. Abundant pleiotropy in human complex diseases and traits. Am J Hum Genet 2011;89:607-18.
46. Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. Nature 1999;401:921-3.
47. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ 2018;362:k601.
48. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG, EPIC-Interact Consortium. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. Eur J Epidemiol 2015;30:543-52.
49. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. Int J Epidemiol 2015;44:512-25.
50. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol 2016;40:304-14.
51. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol 2017;46:1985-98.
52. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. Nat Commun 2018;9:224.
53. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet 2018;50:693-8.
54. O'Connor LJ, Price AL. Distinguishing genetic correlation from causation across 52 diseases and complex traits. Nat Genet 2018;50:1728-34.
55. Morrison J, Knoblauch N, Marcus JH, Stephens M, He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. Nat Genet 2020.
56. Jiang X, Dimou NL, Zhu Z, Bonilla C, Lewis SJ, Lindstrom S, et al. Allergy, asthma, and the risk of breast and prostate cancer: a Mendelian randomization study. Cancer Causes Control 2020;31:273-82.
57. Belsky DW, Sears MR, Hancox RJ, Harrington H, Houts R, Moffitt TE, et al. Polygenic risk and the development and course of asthma: an analysis of data from a four-decade longitudinal study. Lancet Respir Med 2013;1:453-61.
58. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 2018;50:1219-24.
59. Sugrue LP, Desikan RS. What are polygenic scores and why are they important? JAMA 2019;321:1820-1.
60. Vilhjalmsson BJ, Yang J, Finucane HK, Gusev A, Lindstrom S, Ripke S, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. Am J Hum Genet 2015;97:576-92.
61. Zhu Z, Hasegawa K, Ma B, Fujiogi M, Camargo CA Jr, Liang L. Association of asthma and its genetic predisposition with the risk of severe COVID-19. J Allergy Clin Immunol 2020. https://doi.org/10.1016/j.jaci.2020.06.001.
62. Maier R, Moser G, Chen GB, Ripke S, Cross-Disorder Working Group of the Psychiatric Genomics C, Coryell W, et al. Joint analysis of psychiatric disorders increases accuracy of risk prediction for schizophrenia, bipolar disorder, and major depressive disorder. Am J Hum Genet 2015;96:283-94.
63. Chung W, Chen J, Turman C, Lindstrom S, Zhu Z, Loh PR, et al. Efficient cross-trait penalized regression increases prediction accuracy in large cohorts using secondary phenotypes. Nat Commun 2019;10:569.
64. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet 2019; 51:584-91.
65. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57-74.
66. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015;518:317-30.
67. GTEx Consortium. The genotype-tissue expression (GTEx) project. Nat Genet 2013;45:580-5.
68. Heng TS, Painter MW, Immunological Genome Project Consortium. The immunological genome project: networks of gene expression in immune cells. Nat Immunol 2008;9:1091-4.
69. Huan T, Joehanes R, Song C, Peng F, Guo Y, Mendelson M, et al. Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease. Nat Commun 2019;10:4267.
70. Wallace C. Statistical testing of shared genetic control for potentially related traits. Genet Epidemiol 2013;37:802-13.
71. Finucane HK, Reshef YA, Anttila V, Slowikowski K, Gusev A, Byrnes A, et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat Genet 2018;50:621-9.
72. Khatri P, Sirota M, Butte AJ. Ten years of pathway analysis: current approaches and outstanding challenges. PLoS Comput Biol 2012;8:e1002375.
73. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. Int J Epidemiol 2012;41:161-76.
74. Rona RJ, Keil T, Summers C, Gislason D, Zuidmeer L, Sodergren E, et al. The prevalence of food allergy: a meta-analysis. J Allergy Clin Immunol 2007;120: 638-46.
75. Oraka E, Iqbal S, Flanders WD, Brinker K, Garbe P. Racial and ethnic disparities in current asthma and emergency department visits: findings from the national health interview survey, 2001-2010. J Asthma 2013;50:488-96.
76. Daya M, Rafaels N, Brunetti TM, Chavan S, Levin AM, Shetty A, et al. Association study in African-admixed populations across the americas recapitulates asthma risk loci in non-African populations. Nat Commun 2019;10:880.
77. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature 2009;461: 747-53.
78. DeVries A, Vercelli D. Epigenetic mechanisms in asthma. Ann Am Thorac Soc 2016;13(Suppl 1):S48-50.
79. VanderWeele TJ. Mediation analysis: a practitioner's guide. Annu Rev Public Health 2016;37:17-32.
80. Heinze-Deml C, Maathuis MH, Meinshausen N. Causal structure learning. 2018;5: 371-91.