Perspective and a brief overview of genome-wide association studies in moderate to severe asthma

Md Monirul Hoque
Department of Pathobiology, College of Veterinary Medicine, Auburn University, Alabama, USA

Abstract
Asthma is a common chronic respiratory disease that shares phenotypic heritability and shows clusters of symptoms among the relatives. A large number of studies have been conducted to examine the genetic susceptibility of asthma over the past three decades. In the last decade, genome-wide association studies (GWAS) have readdressed the perspective of viewing asthma and have identified some novel genes associated with the susceptibility of asthma. However, few genetic studies have been conducted focusing the moderate to severe asthma, and the molecular targets explain a small proportion of asthma heritability. This review focuses on the principal findings of the genomic studies investigating the genome-wide association of moderate to severe asthma and how it is transitioning the phenotype-based approach towards the fundamental genomic studies. It further illustrates the integrative perspectives aimed towards the translation of the findings in precision medicine. Therefore, a better understanding of asthma pathogenesis would focus the individual at the center of asthma care.

Introduction
Asthma is a complex, non-communicable disease of the airways characterized by recurrent episodes of shortness of breath, cough, wheezing, reversible airflow obstruction, bronchial hyper-responsiveness, mucus overproduction, and abnormal inflammation of the respiratory mucosa. According to the WHO report, 262 million people suffered from asthma in 2019 worldwide. There were 383,000 deaths due to asthma in 2015 and this death toll jumped to 461,000 in 2019, most of which occurred in low and lower-middle income countries. Asthma is a common disease among children and cases are increasing at a rate of 50 percent every ten years [1]. An increase in the incidence of asthma has been associated with urbanization. Recurrent attacks of asthma symptoms are responsible for frequent sleeplessness, daytime fatigue, reduced activity, and absenteeism from school and work. Moreover, in some patients, airflow may be intractably compromised and the airway may be remodeled irreversibly making them refractory to the conventional treatment options with high dose inhaled corticosteroids (ICS) and long-acting β2-adrenergic receptor agonists (LABA) [2]. This subgroup of patients requires a different management approach for their treatment. Researches on the genetic basis of asthma have been evidenced as a promising field of study to develop newer treatment modalities as well as novel preventive protocols for severe asthmatic patients. There are significant differences in asthma prevalence not only among different countries and populations but also among different ethnic groups within the same country. Complex interactions among genetic and environmental factors are responsible for these variations, where genetic factors are assumed to contribute to 35-95% of the susceptibility to develop asthma [3]. Active researches are going on to find out the fundamental causes and underlying pathobiological

*Correspondence: Md Monirul Hoque, Department of Pathobiology, College of Veterinary Medicine, Auburn University, Alabama, USA. Email: hoquemonir@yahoo.com
pathways responsible for the development of asthma. Several genomic approximations have been done to find out the genes underlying the pathogenesis of asthma. The expedition started with linkage analysis studies followed by positional cloning and later by candidate-gene association studies. High-throughput polymorphism genotyping led to the development of methods for much denser genomic scans and initiated the era of genome-wide association studies (GWAS) [4].

Moffatt MF et al. conducted the first GWAS of asthma in 2007 [5]. Ninety three papers were found reported in the GWAS catalog (https://www.ebi.ac.uk/gwas/) till June 1, 2021, on 51 asthma or asthma-related traits. Among them, 9 are GWASs of severe asthma or asthma exacerbations. Validation of the genomic findings from GWAS through the studies of biological mechanistic pathways is opening up the prospect of discovery of potential targets and newer biological drugs, which can modify the progression of disease and prevent the development of severe diseases. This will lead to a paradigm shift in the management approach of asthma that will prioritize the endotype than the phenotype of the disease. It will ultimately lead to a better understanding of asthma heterogeneity and progression, and will help to develop new targeted treatments.

Phenotypic view of asthma

Asthma has long been managed conventionally based on the phenotypic characteristics which are diagnosed by different clinical parameters, such as, history of the patient, lung function test, spirometry, FEV1 (forced expiratory volume) and chest X-ray. Different studies have been conducted to precisely classify asthma so that management protocol can be tailored according to the requirement of the patient cohort. Moore et al. studied asthmatic patients of over 12 years old (726 patients), registered with the Severe Asthma Research Program (SARP) of National Heart Lung and Blood Institute (NHLBI). They conducted cluster analysis using different respiratory function tests and other parameters and categorized the asthmatic patients into 5 clusters emphasizing the clinical course and treatment response for better compliance and greater outcome. Although their algorithm was used for the differential diagnoses of asthma in research studies, it could not be applied in different levels of asthma severity [6,7].

Even though newer drugs are being discovered to combat asthma, the mainstay of treatment remains the inhaled corticosteroids, β2-adrenoceptor agonists, and cholinergic antagonists. None of these drugs prevents or cure asthma, though patients get some level of symptomatic relief but a large proportion continues to suffer [8,9]. A genetic basis can explain this discrepancy of response to the drugs. Asthma susceptibility genes cause mild or intermittent asthma by interacting with environmental factors. Later, different genes lead to disease progression by interacting with other environmental exposures. Thus genetic profiles combined with environmental factors create the platform of different pathophysiological abnormalities and lead to varied clinical asthma patterns. So, a prospective approach is required to categorize the severity of asthma and improve asthma control by personalization of asthma management and identifying the patients at risk for adverse outcomes [5].

Insight into the genetic epidemiology of asthma and associated contributing factors

Asthma is not merely a single disease rather it is an umbrella for multiple diseases with similar clinical features. It has different genetic and environmental contributors. The risk of developing asthma in a person depends not only on his/her degree of genetic relatedness to his/her relative with the disease but also on the severity and the age of onset of asthma in that relative. There is more chance of the development of asthma among the offspring of the asthmatic parents. This supports the genetic predisposition of asthma. The risk of developing asthma in children is 25% if one parent is affected, but it becomes 50% if both parents are affected. Studies on twin further support the genetic basis of asthma. The recurrence risk of asthma is much higher among monozygotic twins than in dizygotic twins [10,11]. However, the concordance of asthma in monozygotic twins is 75% rather than 100%. Even though the monozygotic twins share all their genes, this
discordance of asthma among them points out that not only the genetic factors but also environmental risk factors play an important role in asthma [10]. Hence, although family background plays an important role in the development of asthma, the phenotypic expression of asthma may be influenced by environmental and other genetic factors. A small number of genes are responsible for setting the individual risk background which is then acted upon by another set of modifying genes and environmental factors.

Markus J. Ege et al. conducted a genome-wide interaction analysis for candidate genes of asthma and atopy in a farming environment and they found 5 SNPs (Single Nucleotide Polymorphism) interact with farm-related exposures [12]. This indicates the presence of a potential interaction between the genotype and the environmental factors. Classic GWASs without considering the environmental exposures may not detect the involved SNPs. However, gene-environment interaction provides the opportunity to unravel genetic effects masked by environmental exposures. Moreover, the non-linear expression of the asthma phenotypes makes it even more variable. This adds more difficulty in the prediction of asthma status for a genotype or combination of genotypes. Asthma is more prevalent in the Western population (up to 20%) whereas it is around 1% in the developing world [13]. People in urban areas suffer more from asthma than rural people. Occurrence of symptom frequency, degree of airway responsiveness, level of lung function, and airway inflammation has been found to aggregate within families. So, a person is prone to develop severe asthma if he has a positive family history of severe asthma. A better understanding of the causative factors for the variation of diseases along with the host-related differences in genetic makeup would facilitate the personalized treatment [14].

**Approaches for discovering asthma genes: the study of molecular genetics**

Over a hundred genes have been found associated with the development of asthma and the list is still growing. Different experimental approaches have been used to unravel the genetic determinants of asthma. Technology has continuously evolved over the time to overcome the limitation of existing techniques to reach the desires goal. The candidate gene approach and genome-wide approaches are the principal approaches. Candidate gene association study is conducted in a case-control manner and enrichment of a marker allele (SNP) or haplotype (the group of alleles) are compared among the cases and controls. Another approach is the candidate gene analysis where course of the disease is investigated in the cohort and cross-sectional study designs. Candidate genes are selected based on their known function and the role they play in pathogenesis. As a result, these studies become biased towards studies of immune-related genes and they are unable to discover novel genes or pathways by themselves. These approaches are confined to what we already know or what we think we know about disease pathogenesis and gene functions [15,16]. To overcome the drawback of the candidate gene analysis study, the genome-wide approach was adopted. In this gene discovery approach, the whole genome is taken into consideration without any prior hypotheses about the location of the most important genetic contributors to disease risk. So this approach is called the “hypothesis-free” or “hypothesis-generating” approach. Genome-wide approaches can discover novel genes and pathways involved in the pathogenesis of the disease. Thus this approach introduces new targets and potential pathways for further exploration of the diseases process. Genome-wide linkage study and genome-wide association study (GWAS) are the two methods of genome-wide approach.

Genome-wide linkage studies require the availability of families with at least two affected relatives (affected sibling pairs) where the disease locus co-segregates within the families. The susceptibility loci are also shared among affected relatives more often than expected by chance. This linkage disequilibrium (LD) is utilized in such studies. Linkage studies require relatively few genetic markers and they reveal multiple rare alleles that confer risk for disease, even if the specific variant differs among families. However, they identify very broad regions that contain hundreds of genes and this limits the resolution of the method. Moreover, these studies have low
power to detect risk variants with modest effect sizes on disease risk [15]. The next approach that came into play to deal with the shortcomings of linkage studies is GWAS. GWAS has excellent resolution and good power to detect risk variants with modest effect sizes. It does not need to study families for linkage analysis. It extends the candidate gene approach to include markers that tag all common variations in the genome. GWAS can test for associations with more than a million Single Nucleotide Polymorphisms (SNPs). To achieve genome-wide levels of significance, very large sample sizes and very stringent thresholds of significance (typically with p<10^{-7}) are required to deal with the statistical issues while performing millions of SNP association analyses. To fulfill the criteria researchers collaborate at national and international levels, combine different smaller GWASs and conduct meta-analysis to increase the power of the study [15]. Several GWASs have been conducted to find out the genetic basis of severe asthma.

**Perspectives of genome-wide association studies**

GWAS is a form of genetic association study where hundreds of thousands of SNPs are assessed in a large group of subjects (in a case-control manner) for relationships to a specific phenotype (such as asthma) or a disease-related phenotype (such as IgE level). Unbiased interrogation of the whole genome is the main driving power of GWAS. It is viewed in the context of the Human Genome Project as a whole [17]. The GWASs enable the detection of previously un-described and unsuspected genetic components. But, variants detected as significantly associated in a GWAS do not certify that they are pathogenic. These variants might be in linkage disequilibrium with other rarer and untyped variants [18]. Moreover, the relationship between the genotype and disease is moderated by early environmental exposures, including tobacco smoke, respiratory infection [19], and place of residence. In particular, gene-environment interaction in childhood may determine the platform of risk factors so that associations become apparent only in the exposed individuals. Such gene-environment interactions are common in asthma, and they are very difficult to detect in a GWAS. Thirty-eight loci have been found associated with asthma with a threshold of the genome-wide significance level. Among these loci, the cluster of genes on chromosome 17q12-21 is the most consistently replicated locus among the childhood-onset disease across a diverse range of ethnic backgrounds [20]. Variation at this locus is not associated with atopy, indicating that it is an asthma susceptibility locus and it acts through non-atopic pathways (non-IgE-mediated) [15]. Thus, GWASs are redefining the conventional view of looking at the disease and treatment by identifying novel findings. Genome-wide association studies perform genotyping arrays with up to millions of SNP markers in an unbiased manner throughout the genome to detect the underlying genetic variants responsible for the disease. It requires a very large sample size to maximize the statistical power to detect risk alleles with modest size effects. This requirement is achieved by pooling the samples from multiple independent investigations where the participating members get the chance to agree on standard methods of analysis. A well-developed plan for a large meta-analysis provides the platform for examining genetic factors that are common to or variable between various studies [21].

Meta-analyses of asthma GWASs have been conducted by the GABRIEL (A Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community) and the EVE (a collection of US-based investigators assembled to investigate asthma-susceptibility genes in ethnically diverse populations) consortiums. Subjects only from European ancestry were included in GABRIEL meta-analysis, whereas the EVE study included racially and ethnically diverse subjects from the U.S. and Mexico. The combined results of these two large studies showed highly replicable ethnic or race-specific as well as ethnically diverse associations with asthma [15]. Novel genetic variants, as well as new biological pathways, came under the focus of study. Rose Du et al. conducted the first genome-wide association study of severe or exacerbated asthma among non-Hispanic white children in 2011 [22]. They found that the class I MHC-restricted T cell-associated molecule gene (CRTAM) expression (in the activated CD8+ and NK-T cells) was
associated with asthma exacerbation at a low level of vitamin D. This study referred to the importance of maintenance of an adequate level of vitamin D in the high-risk asthmatic patients. Another study to determine the genetic determinants of severe asthma found the role of ORMDL3/GSDMB locus on chromosome 17q12-21, which was identified as associated with mild to moderate asthma. Proper study design, control of the population heterogeneity with adequate sample size might discover variants responsible for severe asthma masked as a variant of mild asthma. This study found another two novel genes PRPS1L1 and intergenic associated with severe asthma [2]. These genes play biological roles in the pathogenic inflammation of asthma through dendritic cells or Th2 cytokines. Genetic variants identified underlying moderate to severe asthma are summarized in Table-1. All these variants are related to the biological pathways of asthma at different levels of pathogenesis. They will reveal fundamental information regarding severe asthma pathogenesis through a multi-dimensional approach.

The largest GWAS conducted among European ancestry for moderate to severe asthma was published in 2019. It identified 3 novel genes expressed in airway epithelium and in the blood eosinophil [21]. SEMA3D gene codes for a signaling protein for endothelial cell migration and angiogenesis which is responsible for airway remodeling and asthma exacerbation. It causes immune cell recruitment during inflammation. CTNNA3 plays a role in muscle cell coherence, which has brought bronchial smooth muscle under study for insight into its possible role in asthma exacerbation [24].

Target genes of asthma risk variants and functional study

The aim of asthma GWAS is to identify genetic variants (or another variant in strong linkage disequilibrium) associated with disease risk which affects the protein sequence or the transcription patterns of a gene (target gene) that ultimately plays a role in disease pathophysiology. Disease risk-associated variants highlight specific genes and molecular pathways which are dysregulated in asthma and they help understand the underlying cause of the development of asthma. There are at least 24 genes that are likely targets of severe asthma risk variants. Target genes of risk variants can be identified in two ways. Based on the published risk variants or variants in strong LD with them, a statistical approach, such as ANNOVAR is used to see if these variants are non-synonymous coding variants. Assessment of the variants for a reproducible association with asthma in the UK Biobank study can be performed. This approach identified 8 likely (non-synonymous variants) target genes: GSDMA, GSDMB, HLA-DQA1, HLA-DQB1, IL1RL1, IL6R, TLR1, and ZPB2 in asthma risk variants [26].

Expression quantitative trait loci (eQTL) analysis can be performed to see if the target genes or associated SNPs are in strong LD with the reported variants. For example, using the eQTL approach, 48 likely target genes were found in asthma susceptibility variants. This approach needs eQTL studies to be conducted in tissues relevant to asthma pathophysiology, for example, lung, epithelium, whole blood, different immune cell types, such as CD4+ T cells, Nk cells, dendritic cells, etc. These analyses show the importance of the variants as the determinants of cellular function and disease risk. However, these studies should be performed carefully to ensure that the effect of these variants on gene expression is not specific to tissues, cell types, and cellular conditions or the variants have a moderate effect on the genes to be detected with the sample sizes in the published eQTL studies. Otherwise, the findings will not interpret the true picture.

In such cases, the target genes could be identified by functional studies, such as experiments to determine the effect of risk variants on promoter enhancer chromatin interactions, transcription factor binding, or promoter activity [26]. Functional studies in mice produce a good result in this regard. For example, functional studies in transgenic (hORMDL3<sup>zp3-Cre</sup>) mice show that ORMDL3 expression increases airway remodeling (increased airway smooth muscle, sub-epithelial fibrosis, and mucus production) following airway inflammation [27]. Another study using hGSDMB<sup>zp5-Cre</sup> mice shows that increased expression of GSDMB causes a spontaneous increase in airway hyper-
responsiveness and an increase in airway remodeling (along with increased smooth muscle mass and fibrosis) in the absence of airway inflammation [28]. These functional and animal studies give us insights into the role of ORMDL3 and GSDMB in the pathophysiology of the disease. In another study, the expression of ORMDL3 in the lung influences the host response to fungal challenge and antiviral response to rhinovirus infection. It demonstrates the interaction between ORMDL3 (on locus 17q12-21) and early respiratory infection in the development of asthma. [16]. Thus, the validation of the findings of GWAS through statistical approach or functional studies would unravel the genetic basis of development and outcome of severe asthma.

Table 1: Genome-wide association studies identifying moderate to severe asthma risk variants*

<table>
<thead>
<tr>
<th>Study population</th>
<th>Phenotype</th>
<th>Chromosome</th>
<th>Reported gene</th>
<th>SNP</th>
<th>Alleles</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>395 non-Hispanic white children (CAMP cohort) with asthma. Replication: 584 Costa Rican children with asthma exacerbation [22].</td>
<td>Asthma exacerbation</td>
<td>11</td>
<td>CRTAM</td>
<td>rs7941607</td>
<td>C/T</td>
<td>2.8×10^{-4}</td>
</tr>
<tr>
<td>933 individuals of European ancestry with severe asthma and 3345 controls. Replication: 231 cases with severe asthma and 1,345 controls recruited from AAGC [2].</td>
<td>Severe asthma</td>
<td>2</td>
<td>IL18R1</td>
<td>rs3771166</td>
<td>G</td>
<td>6.6 × 10^{-9}</td>
</tr>
<tr>
<td>1,173 cases and 2,552 controls of Danish descent. Replication: 3,975 children from COPSAC-2000 and MAAS cohort [23].</td>
<td>Asthma with severe exacerbation</td>
<td>2</td>
<td>IL1R1</td>
<td>rs1558641</td>
<td>G</td>
<td>6.6 × 10^{-9}</td>
</tr>
<tr>
<td>1,347 children from GALAII and SAGE cohort. Replication: 1,697 individual from PiCA study [25].</td>
<td>Asthma exacerbation</td>
<td>22</td>
<td>APOBEC3B-APOBEC3C</td>
<td>rs5995653</td>
<td>A/G</td>
<td>4.80 × 10^{-6}</td>
</tr>
<tr>
<td>Stage 1: 5,135 cases and 25,675 controls from European ancestry. Stage 2: 5,414 cases and 21,471 controls of European ancestry [21].</td>
<td>Moderate to severe asthma</td>
<td>4</td>
<td>KIAA1109</td>
<td>rs560026225</td>
<td>G/A</td>
<td>3.06×10^{-9}</td>
</tr>
</tbody>
</table>

*Note that this list is not exhaustive. CAMP= Childhood Asthma Management Program, AAGC= Australian Asthma Genetics Consortium, COPSAC= Copenhagen Prospective Studies on Asthma in Childhood, MAAS= Manchester Asthma and Allergy Study, CARE= Childhood Asthma Research and Education, BioUV= Biobank of Vanderbilt University, GALAII= Genes-environments & Admixture in Latino Americans Study, SAGE= Study of African Americans, Asthma, Genes and Environments, PiCA= Pharmacogenomics in Childhood of Asthma consortium.
Translation of the genetic findings into clinical research or practice

The genetic risk factors identified until today, represent a small proportion of total asthma heritability. An important area of research is how much the individual response is influenced by genetic variation. B2-adrenergic receptor agonists (SABA and LABA) and inhaled corticosteroids (ICS) are the most commonly prescribed medication to treat asthma. Most asthma patients experience a decrease in the disease symptoms on treatment with these drugs but a wide range of responses is reported among individuals and populations [4,18]. In spite of high dosages, approximately 5–10% of all patients remain symptomatic and/or at the risk of severe exacerbation of asthma. Subsets of patients continue to suffer despite maintenance treatment even in clinical trials where patients groups are carefully selected and adherence to treatment is closely monitored. This evidence insinuates that there is a role of genetic variance underlying these varied responses [29]. There might be rapid or slow metabolizers among the asthmatics, target proteins or receptors might be expressed in different ways or in different numbers in different individuals, even the biological signaling pathways or epistasis or gene-environment interaction might work differently among them. Characterization of these issues would enable the knowledge gained in GWASs to translate into clinical treatment. For example, pharmacogenetics studies on coding variants in the B2-adrenergic receptor gene (ADRB2) to test the long-term responses of bronchoconstriction in asthma with short-acting B-agonist (SABA) showed an association with polymorphism at the Gly16Arg locus. The results showed that Arg16 homozygotes improved on treatment with intermittent SABA therapy but they did not show any improvement in regular SABA treatment. However, Gly16 homozygotes had improved response to regular SABA therapy, with similar genotype-specific effects [30]. Genetic variation in the STIP1 gene (encodes a protein which is involved in the activation of the glucocorticoid receptor) in candidate gene study has shown differences in FEV1 in response to inhaled corticosteroid treatment [30]. So, information regarding this genetic profile or knowledge of the polymorphisms might bring radical changes in the disease management approach and will help treatment optimization.

Most GWAS of asthma has limited statistical power due to small sample sizes or underrepresentation of genetically diverse populations. However, large consortia-based study approaches might solve this problem by gathering many small studies from different countries around the world through the contribution of patients from multiethnic populations [4]. The risk SNPs identified in GWAS will help understand how genetic variation at different loci affects gene expression, cellular function, disease pathophysiology and ultimately lead to its application in clinical treatment. A clinical study with tocilizumab in participants with mild to moderate asthma was motivated directly by the results from a GWAS. Another study, based on the genetic findings and the mouse studies, shows asthmatic patients with CT or TT genotype for rs4129267 (who have increased levels of sIL-6R) are expected to be benefited [26].

Individual at the center of the health care system: precision medicine for asthma

Genome-wide association studies are changing the idea of cluster-based symptomatic and maintenance treatment for asthma and bringing forth the idea of precision medicine, where the individual is at the center of the management protocol rather than the group of patients (sharing the common phenotype). Conventional clinical guidelines and some biomarkers have long been used to diagnose asthma. But these diagnostic tools cannot accurately categorize the patients due to the wide variety of molecular mechanisms underlying the different asthma phenotypes.

In this respect, integrative approaches such as a combination of complete clinical data and the biological sources namely genomic modifications, gene transcription, protein levels and chemical modifications, endogenous and exogenous metabolites, and metagenetics could contribute to better characterize the biological processes underlying asthma pathophysiology. This will ultimately help to define asthma subtypes (endotypes) and improve the prediction of severity and treatment response [4] (Figure 1).
For genome-based management, the patients are genotyped before enrollment in the treatment so that a sufficient number of patients can be studied with the important genotype [31]. The “genotype first” based approach identifies patients of more homogenous endotypes. The whole population is stratified based on an individual’s polygenic risk scores or genetic scores. A patient’s age, ancestry, family history, BMI, childhood infection, local environmental factors (exposure to tobacco smoke, air pollution, etc.) along with predictive biomarkers (the combination of multiple risk genetic variants) will determine his/her genetic and pharmacogenetics profile. The management protocol that is the preventive, diagnostic and therapeutic strategies will be tailored according to the need of the individual. The individual patient will be regarded as the unit of therapeutic approach and this will ultimately transform conventional disease-based medicine into personalized medicine.

**Conclusion**

Each complex disease has its own genetic architecture of common and rare variants. Gene-gene interaction and gene-environment interactions add another layer of individuality. Phenotypic heterogeneity and multifactorial etiology have made asthma a complex syndrome rather than a single disease entity. Asthma GWASs are identifying multiple genome-wide significant associations to genes that were not considered as key players in asthma pathogenesis previously. Biological pathway studies, multiomics approaches, cutting-edge statistical analyses are exploring the insights generated from genome-wide studies and paving the way to comprehensive management of asthma. Proper harnessing of this information in an integrated way, together with the analysis of mass of data will ultimately lead to personalized health care.
Conflict of interest: None

References


