Clinical reviews in allergy and immunology

Series editors: Donald Y. M. Leung, MD, PhD, and Dennis K. Ledford, MD

Application of genetic/genomic approaches to allergic disorders

Tesfaye M. Baye, PhD,[a] Lisa J. Martin, PhD,[b] and Gurjit K. Khurana Hershey, MD, PhD[a,c] *Cincinnati, Ohio*

INFORMATION FOR CATEGORY 1 CME CREDIT

Credit can now be obtained, free for a limited time, by reading the review articles in this issue. Please note the following instructions.

Method of Physician Participation in Learning Process: The core material for these activities can be read in this issue of the Journal or online at the JACI Web site: www.jacionline.org. The accompanying tests may only be submitted online at www.jacionline.org. Fax or other copies will not be accepted.

Date of Original Release: September 2010. Credit may be obtained for these courses until August 31, 2012.

Overall Purpose/Goal: To provide excellent reviews on key aspects of allergic disease to those who research, treat, or manage allergic disease.

Target Audience: Physicians and researchers within the field of allergic disease.

Accreditation/Provider Statements and Credit Designation: The American Academy of Allergy, Asthma & Immunology (AAAAI) is accredited by the Accreditation Council for Continuing Medical Education (ACCME) to provide continuing medical education for physicians. The AAAAI designates these educational activities for a maximum of 1 *AMA PRA Category 1 Credit*™. Physicians should only claim credit commensurate with the extent of their participation in the activity.

List of Design Committee Members: Tesfaye M. Baye, PhD, Lisa J. Martin, PhD, and Gurjit K. Khurana Hershey, MD, PhD

Activity Objectives

1. To differentiate between the research methodologies currently used to study disorders that are at least in part heritable.
2. To increase awareness of current major international studies of inherited disorders.
3. To understand different types of genetic variation.

Recognition of Commercial Support: This CME activity has not received external commercial support.

Disclosure of Significant Relationships with Relevant Commercial Companies/Organizations: L. J. Martin has received research support from the National Institutes of Health. The rest of the authors have declared that they have no conflict of interest.

Completion of the human genome project and rapid progress in genetics and bioinformatics have enabled the development of large public databases, which include genetic and genomic data linked to clinical health data. With the massive amount of information available, clinicians and researchers have the unique opportunity to complement and integrate their daily practice with the existing resources to clarify the underlying cause of complex phenotypes, such as allergic diseases. The genome itself is now often used as a starting point for many studies, and multiple innovative approaches have emerged applying genetic/genomic strategies to key questions in the field of allergy and immunology. There have been several successes that have uncovered new insights into the biologic underpinnings of allergic disorders. Herein we will provide an in-depth review of genomic approaches to identifying genes and biologic networks involved in allergic diseases. We will discuss genetic and phenotypic variation, statistical approaches for gene discovery, public databases, functional genomics, clinical implications, and the challenges that remain. (J Allergy Clin Immunol 2010;126:425-36.)

Key words: *Gene, allergy, database, browser, genome, common variants, rare variants, HapMap, imputation*

Human genome variation encompasses all of the genetic characteristics observed within the human species. Genetic variation occurs both within and among populations and is the basis for natural selection. Insights regarding the distribution of genetic variants among human populations have recently become available.[1] Interestingly, human genetic diversity decreases in native populations as the migratory distance from Africa increases, presumably because of limitations in human migration.[2]

Nucleotide diversity is based on single mutations called single nucleotide polymorphisms (SNPs) that occur at a rate of 1 SNP per 1,000 bp.[3] Currently, there are more than 12 million SNPs deposited in GenBank, 6.5 million of which have been validated (http://www.ncbi.nih.gov/SNP). The bulk of variations at these nucleotide levels are not visible at the phenotypic level. A better understanding of the basis of genetic diversity was gained with the publication of full sequences of individuals genomes.[4,5] The Human Genome Project and a parallel project by Celera Genomics yielded 2 haploid sequences; however, analysis of diploid sequences has revealed that non-SNP variation accounts for much more human genetic variation than single nucleotide diversity. Non-SNP variation includes copy number variation (CNV) and

From [a]Cincinnati Children's Hospital Medical Center, Division of Asthma Research; [b]the Division of Epidemiology and Biostatistics; and [c]the Division of Allergy and Immunology, Department of Pediatrics, University of Cincinnati.

Supported by National Institutes of Health grants U19A170235 (G. K. K. H. and L. J. M.) and P30HL10133 (T. M. B.).

Received for publication March 11, 2010; revised April 28, 2010; accepted for publication May 7, 2010.

Available online July 20, 2010.

Reprint requests: Gurjit K. Khurana Hershey, MD, PhD, Division of Asthma Research, Cincinnati Children's Hospital Medical Center, Cincinnati, OH. E-mail: Gurjit.Khurana.Hershey@cchmc.org.

0091-6749/$36.00

doi:10.1016/j.jaci.2010.05.025

Abbreviations used
AD: Atopic dermatitis
AIM: Ancestry-informative markers
CNV: Copy number variation
CNVR: Copy number variation region
FLG: Filaggrin
GO: Gene Ontology
GWAS: Genome-wide association study
LD: Linkage disequilibrium
ORMDL3: ORM1-like protein 3 gene
SNP: Single nucleotide polymorphism

results from deletions, inversions, insertions and duplications.[5] Copy number variation regions (CNVRs) have been found in 12% of the genome. CNVRs can be markedly different between populations and contain hundreds of genes, disease loci, functional elements, and segmental duplications.[5] Taking into account this variation, as well as SNPs, human-to-human genetic variation is estimated to be approximately 0.5%. This 0.5% difference amounts to a significant number of distinct genetic traits that uniquely distinguish the genome of every person and contribute to unique and distinct risks for diseases, responses to environmental exposures (including nutrition), and responses to pharmacologic treatment.

EPIGENETIC VARIATION IN ALLERGIC DISORDERS

Epigenetic variation does not affect the underlying DNA code but rather modifies how it is expressed through covalent modifications, including DNA methylation, histone modifications, and microRNAs. It is the structural adaptation of chromosomal regions so as to register, signal, or perpetuate altered activity states.[6] Detailed analysis of methylation across several chromosomes has demonstrated that the promoter regions of nearly 20% of genes are methylated, many of which influence transcription.[7] Progressive accumulation of phenotypic differences between genetically identical monozygotic twins illustrates how pollution, smoking, mold, diet, habits, or, in general, environment can shape phenotype and disease susceptibility. Monozygotic twins are epigenetically indistinguishable early in life but exhibit substantial differences with age, particularly when they have led different lifestyles and spent less of their lives together.[8,9] Therefore monozygotic twin discordance for many common disorders could be interpreted as the result of external environmental factors that modulate susceptibility through a change in the profile of epigenetic modifications that ultimately determine gene function. The field of epigenetics has emerged to explain how cells with the same DNA can differentiate into alternative cell types and how a phenotype can be passed from one cell to its daughter cells. It is now well established that epigenetic mechanisms are important to control the pattern of gene expression during development and the cell cycle and in response to biologic or environmental changes.[10-13] Unlike genetic alterations, which are permanent and usually affect all cells, epigenetic modifications are cell type specific.[14] Epigenetic regulation of the immune system occurs at many levels, including the differentiation of T cells.[6,15-19] Epigenetic effects on gene expression can persist even after the removal of the inducing agent and can be passed on through mitosis to subsequent cell generations, constituting a heritable epigenetic change. In a somatic cell a heritable change can generate a dysfunctional clone of cells with phenotypic consequences (eg, a tumor). In a germ-line cell a heritable change can be transmitted to the germ cells themselves (sperm or ova) and potentially to the next generation. In this model epialleles can be in linkage disequilibrium (LD) with SNPs that are genotyped in genome-wide association studies (GWASs). The role of epigenetics in allergic disease is becoming increasingly evident. One recent study showed that epigenetic reprogramming involving aberrant DNA methylation of a 5′-CpG island in acyl-CoA synthetase long-chain family member 3 *(ACSL3)* was significantly associated with asthma risk in children born to mothers exposed to air pollutants, such as traffic-related combustion emissions.[20] Another study found that neonates of allergic mothers are born with substantial changes in DNA methylation in their splenic dendritic cells and that these dendritic cells show enhanced allergen-presenting activity *in vitro*.[21] Current knowledge of epigenetics in allergic diseases is limited, and novel applications of epigenetic approaches, including genome-wide approaches to allergic diseases, are necessary to uncover the role of epigenetics.

DEFINING PHENOTYPIC VARIATION IN ALLERGIC DISEASE

The phenotype is defined as the observable characteristics of an organism, as determined by both genetic makeup and environmental influences, including individual physical, psychosocial, and environmental exposures (Fig 1). The genotype is the descriptor of the genome, which is the set of physical DNA molecules inherited from the organism's parents, whereas phenotype is the descriptor of the phenome, the manifest physical properties of the organism, including its physiology, morphology, and behavior.

Although single-gene disorders in classical Mendelian inheritance result in direct genotype-phenotype correspondence, the relationship between genotype and phenotype in traits of multifactorial (complex) inheritance is complicated. In complex diseases with a multifaceted phenotype, such as asthma, a given genotype can result in many different phenotypes, and there are different genotypes corresponding to a given phenotype. Although a subject's genotype is fairly stable over a lifetime, his or her phenotype is dynamic, influenced by both the environment and the underlying genotype, including interactions between them.[22] The definition, measurement, and validity of phenotyping need to be standardized to increase the quality of research and the reproducibility of genetic studies.[22] Indeed, recently, the National Institutes of Health launched an initiative (Consensus Measures for Phenotypes and Exposures [PhenX]) to address the need standardized phenotype and environmental exposure measures for cross-study comparison in genetic studies.[23] These measures do not include information for allergic diseases; however, the National Institute of Allergy and Infectious Diseases recently partnered with the National Heart, Lung, and Blood Institute; the National Institute of Environmental Health Sciences; the National Institute of Child Health and Human Development; the Agency for Healthcare Research and Quality; the Merck Childhood Asthma Network; and the Robert Wood Johnson Foundation to host an Asthma Outcomes Workshop. The objective of this workshop was to develop standardized definitions and data collection methodologies for established and validated asthma outcomes measures. The goal is that these outcomes will be broadly used in National Institutes of Health–funded studies.[24]