Systems biology of asthma and allergic diseases: A multiscale approach

Supinda Bunyavanich, MD, MPH,[a,b] and Eric E. Schadt, PhD[a] *New York, NY*

Systems biology is an approach to understanding living systems that focuses on modeling diverse types of high-dimensional interactions to develop a more comprehensive understanding of complex phenotypes manifested by the system. High-throughput molecular, cellular, and physiologic profiling of populations is coupled with bioinformatic and computational techniques to identify new functional roles for genes, regulatory elements, and metabolites in the context of the molecular networks that define biological processes associated with system physiology. Given the complexity and heterogeneity of asthma and allergic diseases, a systems biology approach is attractive, as it has the potential to model the myriad connections and interdependencies between genetic predisposition, environmental perturbations, regulatory intermediaries, and molecular sequelae that ultimately lead to diverse disease phenotypes and treatment responses across individuals. The increasing availability of high-throughput technologies has enabled system-wide profiling of the genome, transcriptome, epigenome, microbiome, and metabolome, providing fodder for systems biology approaches to examine asthma and allergy at a more holistic level. In this article we review the technologies and approaches for system-wide profiling, as well as their more recent applications to asthma and allergy. We discuss approaches for integrating multiscale data through network analyses and provide perspective on how individually captured health profiles will contribute to more accurate systems biology views of asthma and allergy. (J Allergy Clin Immunol 2015;135:31-42.)

Key words: *Systems biology, network, asthma, allergy, atopic, genome, transcriptome, epigenome, microbiome, metabolome, individual health profile, big data*

The complexity and heterogeneity of asthma and allergic diseases make both their clinical management and investigation challenging. Innumerable environmental and microbial exposures modulate variable degrees of genetic predisposition, yielding a spectrum of transcriptional and molecular sequelae, disease phenotypes, and treatment responses across individuals.

Abbreviations used

AD:	Atopic dermatitis
ChIP:	Chromatin immunoprecipitation
CpG:	Cytosine-phosphate-guanine
EoE:	Eosinophilic esophagitis
eQTL:	Expression quantitative trait loci
eSNP:	Expression single nucleotide polymorphism
FAIRE:	Formaldehyde-assisted isolation of regulatory elements
GWAS:	Genome-wide association studies
LC-MS:	Liquid chromatography mass spectrometry
miRNA:	MicroRNA
NMR:	Nuclear magnetic resonance
RNA-seq:	RNA sequencing

The increasing availability of high-throughput technologies has enabled system-wide profiling of the genome, transcriptome, epigenome, microbiome, and metabolome, providing fodder for systems biology approaches to examine asthma and allergy at a more holistic level (Fig 1). In systems biology, large data sets collected by multiple modalities in populations, ideally with multiple dimensions of data for each individual, are used to generate networks that link phenotypic information to interdependent genetic, regulatory, metabolite, and environmental profiles. The resulting networks are used to predict the behavior of the trait and generate novel and biologically relevant information.

In this review, we provide an overview of the technologies and approaches for system-wide profiling and review their more recent applications to the study of asthma and allergy. We focus on findings from the past few years and provide perspective on their integration with increasingly available personal health profiles toward a systems view of asthma and allergy.

GENOME

Since the completion of the Human Genome Project, oligonucleotide microarrays have enabled the simultaneous genotyping of millions of genetic variants scattered across the genome at low cost and with small amounts of starting DNA. The increasing technical and financial accessibility of this technology has led to direct-to-consumer genotyping services, expanding the pool of genotyped populations.[1] Genome-wide association studies (GWAS), which examine for associations between genotype and phenotype, enable unbiased identification of genetic loci for asthma and allergy disease risk.[2] The National Human Genome Research Institute provides a searchable catalog of published GWAS at www.genome.gov.[3]

More GWAS of asthma have been conducted than for other allergic diseases.[3] The 17q21 locus[4] has been associated with asthma with the greatest reproducibility.[4-7] Encompassing 4 genes (*ORMDL3*, *GSDMB*, *ZPBP2*, and *IKZF3*), this locus might affect endoplasmic reticulum–mediated Ca^{2+} homeostasis and protein folding, resulting in the unfolded protein response

From [a]the Department of Genetics and Genomic Sciences and the Icahn Institute for Genomics and Multiscale Biology and [b]the Division of Pediatric Allergy and Immunology, Department of Pediatrics, and the Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai.

Disclosure of potential conflict of interest: S. Bunyavanich has received research support from the National Institutes of Health/National Institute of Allergy and Infectious Disease (NIH K08AI093538). E. E. Schadt declares that he has no relevant conflicts of interest.

Received for publication August 21, 2014; revised October 10, 2014; accepted for publication October 15, 2014.

Available online November 21, 2014.

Corresponding author: Supinda Bunyavanich, MD, MPH, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, #1198, New York, NY 10029. E-mail: supinda@post.harvard.edu.

0091-6749/$36.00

http://dx.doi.org/10.1016/j.jaci.2014.10.015

as an endogenous inducer of inflammation.[8] Other asthma susceptibility loci identified through GWAS include genes related to recruitment or activation of inflammatory cells (*TSLP*,[5,9,10] *IL33*,[5,9] and *IL1RL1*[5,6,11]), T-cell response and differentiation (*HLA* genes,[9,10] *IL2RB*,[9] *DENND1B*,[12] and *IL6R*[7]), and, like the 17q21 locus, cell-signaling modulation (*PDE4D*,[13] *SMAD3*,[9] and *CDHR3*[14]). GWAS have also identified genetic variants associated with drug treatment response in asthmatic subjects.[15,16]

Although GWAS have uncovered common variation in numerous genes for asthma, a large proportion of genetic risk for asthma remains unexplained. Single nucleotide polymorphisms on genotyping arrays largely represent common variants that may tag rare variants incompletely. Sequencing can uncover rare variants, as it enables deeper coverage of genes and regions of interest, or of the entire genome in the case of whole-genome sequencing. However, sequencing is more cost prohibitive than genotyping arrays and therefore practical at this time to implement for targeted gene regions (eg, 9 candidate asthma genes in 965 subjects[17]) or for discovery in limited numbers of subjects (eg, whole-genome sequencing in 16 subjects with well-characterized asthma[18]).

GWAS of atopic dermatitis (AD) support that mutations in the *FLG* gene encoding the epidermal structural protein filaggrin are risk factors for AD in European and Asian subjects.[19-21] *FLG* mutations are not commonly found in those of African ancestry, as demonstrated by a study using whole-exome sequencing.[22] In whole-exome sequencing the subset of DNA that encodes proteins (exons) is sequenced. The whole exome represents about 1% of the human genome and has been more frequently sequenced in family-based studies of rare diseases (eg, primary immunodeficiencies[23]), with limited application for asthma and allergy.

Loci near *C11orf30*,[21,24] 5q22.1,[20] and 20q13.33[20] have also been identified as AD risk variants by GWAS. A meta-analysis of 16 population-based cohorts identified risk loci for AD near *OVOL1* and *ACTL9*, which are implicated in epidermal proliferation and differentiation; these loci were replicated in an independent study of Japanese subjects,[21] in which additional variants were also identified. GWAS of AD overlapping with other conditions, such as psoriasis and asthma, have also been performed.[25]

The *C11orf30* locus associated with AD was also found to be associated with allergic rhinitis through GWAS.[26] A genome-wide association meta-analysis of white subjects[27] examined the associations between genotype and self-reported allergies by using data from a direct-to-consumer genotyping service[1] combined with a more traditionally recruited study cohort. Several loci overlapped with those for asthma, and there was 1 locus near *HLA-DQA1* at 6p21.32 specifically associated with self-reported cat allergy.[27] GWAS of ethnically diverse North American subjects identified distinct genome-wide significant loci among Latinos, with integrative genomic analyses showing enrichment of the identified GWAS loci for mitochondrial pathways.[28]

GWAS have also been performed for allergen sensitization. A European study focusing on grass sensitization specifically found that the AD- and AR-associated gene *C11orf30* was also associated with grass sensitization, as were *HLA-DRB4* and a locus near *TMEM232* and *SLC25A46*.[26] A subsequent larger-scale GWAS of food and environmental allergen sensitization combined data from 16 cohorts and identified 10 loci associated with sensitization to any allergen. These loci were scattered across the genome, with 6 previously associated with established

GLOSSARY

EPIGENETICS: Modifications to DNA that affect gene expression and occur without direct alteration of the DNA sequence. Epigenetic changes can include DNA methylation, DNA hydroxymethylation, histone modification, and miRNAs. These changes can be heritable but are also reversible because the genetic code remains unchanged.

GENOME: The complete set of genetic information for an organism, including genes and noncoding sequences. The genome contains the information needed to build and maintain the organism. The human genome is more than 3 billion DNA base pairs.

GENOME-WIDE ASSOCIATION STUDY: A study in which DNA variants across the genome are simultaneously analyzed for association with a trait of interest.

HUMAN MICROBIOME: The collection of commensal, symbiotic, and pathogenic microorganisms and their genomes that are found in the human body.

METABOLOMICS: High-throughput characterization of metabolites found in an organism. NMR or LC-MS are typically used. Metabolites can be specific to certain body fluids, such as urine, serum, plasma, and exhaled breath condensate.

MICROARRAY: A technology used to study genotype, gene expression, methylation, miRNAs, and chromatin marks of thousands of genes at once. Known specific nucleic acid sequences (probes) are studded on a glass slide. A sample containing cDNA or cRNA is then placed. Complementary base pairing between the sample and the sequences on the chip yield fluorescence that is measured and correlates with the amount of nucleic acid present.

NETWORK: A framework for exploring the context in which genes, gene products, metabolites, and other variables operate. Networks are mathematic models comprised of nodes and edges that model the connectivity and complex interactions between variables in a system that associate with one another. Nodes in a network typically represent genes, gene products, metabolites, or other important molecular entities. Edges (or links) between any 2 nodes indicate a relationship between the 2 entities.

SYSTEMS BIOLOGY: An approach to understanding living systems that focuses on modeling diverse types of high-dimensional interactions to develop a more comprehensive understanding of complex phenotypes manifested by the system. High-throughput molecular, cellular, and physiologic profiling of populations is coupled with bioinformatic and computational techniques to identify new functional roles for genes, regulatory elements, and metabolites in the context of the molecular networks that define biological processes associated with system physiology.

TRANSCRIPTOME: The set of RNA molecules produced in 1 cell or a population of cells.

WEARABLE DEVICE: Clothing and accessories incorporating sensors and electronic technologies to convey data. Health-related wearable devices currently on the market include bracelets and watches that monitor activity, nutrition, and sleep; tattoos that monitor pH and lactate content in sweat; diapers with urinalysis monitors; ingestible sensors that track medication adherence and response; and clothing that measures electrocardiographic, electroencephalographic, and electromyographic signals.

WHOLE-EXOME SEQUENCING: The process of determining the sequence of an organism's protein-encoding DNA (exons). The whole exome represents about 1% of the human genome.

WHOLE-GENOME SEQUENCING: The process of determining the complete sequence of an organism's chromosomal and mitochondrial DNA at a single time.