Environmental genomics: a key to understanding biology, pathophysiology and disease

David A. Schwartz1,2,3,4,*, Jonathan H. Freedman2,3,4 and Elwood A. Linney2,3

1Department of Medicine and 2Department of Molecular Genetics and Microbiology, Duke University Medical Center, 3Nicholas School of the Environment and 4Durham VAMC, Durham, North Carolina, USA

Received June 7, 2004; Revised and Accepted July 19, 2004

Recent advances in human and molecular genetics provide an unparalleled opportunity to understand how genes and genetic changes interact with environmental stimuli to either preserve health or cause disease. The fields of environmental genetics and environmental genomics has enormous potential to affect our ability to accurately assess the risk of developing disease, identify and understand basic pathogenic mechanisms that are critical to disease progression, and to more precisely phenotype disease subtypes. However, the application of genetics and genomics to problems in environmental health is only the beginning yet, by itself, represents a potentially effective strategy to substantially impact morbidity and mortality. Collaborative approaches that team together environmental scientists with molecular biologists, geneticists, physiologists and physician scientists are critical to the investigation of environmental aspects of human health. Moreover, exploiting eukaryotic model systems (yeast, Caenorhabditis elegans, zebrafish, Drosophila and rodents) will accelerate our understanding of environmental exposures on human health.

Despite the tremendous inter-individual variability in the response to environmental toxicants, we simply do not understand why certain people develop disease when challenged with environmental agents, and others remain healthy. Although an emerging consensus suggests that many of the complex and prevalent diseases that humans develop occur as a result of multiple biologically unique gene–gene and gene–environment interactions, this conceptual framework is limited. In fact, the development of disease in humans, environmental and otherwise, is far more complex. Environmental exposures affect those that are vulnerable temporally (age) and by unique circumstance (co-morbid disease, nutritional status, economic status, race and genetics). Even this paradigm fails to address the complex interaction of endogenous and exogenous risks that ultimately interact to cause disease. Recent advances in human and molecular genetics provide an unparalleled opportunity to understand how genes and genetic changes interact with environmental stimuli to either preserve health or cause disease. Without accounting for the temporal, spatial and other unique components of an individual’s microenvironment, however, our understanding of environmental health will remain incomplete. Understanding the complex relationship between endogenous and exogenous risks in populations and within affected individuals is precisely the opportunity and challenge that face scientists involved in environmental genetics and genomics.

Our understanding of environmental health and genetics/genomics is evolving rapidly. It is now generally accepted that very few diseases are caused exclusively by sequence variations in the human genome. For instance, there is clear evidence that the phenotypes of even single-gene diseases, such as cystic fibrosis and phenylketonuria, are caused by multiple gene–gene and gene–environment interactions. Common complex diseases, such as atherosclerosis and type 2 diabetes, are caused by reversible behaviors or avoidable exposures at least 70% of the time (1). In fact, single gene mutations are rare and account for <5% of the cases of cancer and cardiovascular disease (1). In complex diseases, such as asthma, the interactions between genetic changes and environmental exposures are clearly evident. For example, the development of clinical asthma results from a complex biological interaction between environmental agents and multiple gene products. Genetic variation in one or more of these products can enhance susceptibility (2,3). Although everyone, to a certain extent, is exposed to aeroallergens, viruses, endotoxin, cigarette smoke and air pollution, only a minority of individuals develops asthma. Although Th1/Th2 balance explains some of the susceptibility to aeroallergens, Th1 responses drive the airway response to inhaled endotoxin

*To whom correspondence should be addressed at: Duke University Medical Center, PO Box 2629, Durham, NC 27710, USA. Tel: +1 9196680380; Fax: +1 9196680494; Email: david.schwartz@duke.edu

Human Molecular Genetics, Vol. 13, Review Issue 2 © Oxford University Press 2004; all rights reserved
and organic dusts (4–6), and have also been shown to enhance airway inflammation and even influence allergen induced asthma exacerbations (7–9). Thus, although the Th1/Th2 balance is important, this conceptual framework provides a limited explanation for asthma susceptibility and/or pathogenesis. Moreover, owing to the complex clinical phenotype of asthma (10,11), the polygenic inheritance pattern of this disease (3,12–16), and the substantial role of environmental exposures in the development and progression of asthma (3,14,17), major susceptibility genes for asthma have not been definitively identified (14,15). It is precisely this concern that demonstrates the need to incorporate environmental exposures into many of the studies that are searching for genes that contribute to the risk of developing a particular disease or a more specific phenotype.

As environmental exposures substantially contribute to the etiology of many common complex human diseases, environmental health scientists have a unique opportunity to focus on the interface between environmental exposures, basic biology/genetics, vulnerable populations and the common diseases that limit our longevity. In the genomic era of biomedical research, environmental health scientists can develop a leadership role in improving human health by using environmental toxicants to understand how genes work in biological systems, how genetic variants contribute to the development of disease, and why individuals with the same disease have very different clinical outcomes. Moreover, environmental health scientists are uniquely poised to develop sensitive pre-clinical markers of exposures and biological responses, to develop strategies to prevent disease in exposed and biologically responsive individuals/populations. Further, they can help to establish population-based cohorts to understand the impact of environmental exposures on human health, and to understand how environmental exposures affect the course and prognosis of a medical condition.

Genetic variability implies unpredictable vulnerability. Among the 3.2 billion base pairs in the human genome, <1% uniquely identifies each human being (18). Single nucleotide polymorphisms (SNPs) are predicted to occur at least every 300 base pairs. Although only 40% of these SNPs will result in amino acid changes, these SNPs provide the genetic diversity that underlies the variable susceptibility to environmental stimuli and the variable risk of disease development and progression (18). In this review, we will illustrate how gene–environment interactions can be used to identify key regulatory genes and pathways, discover susceptibility genes, and define disease phenotypes.

CLUES TO IDENTIFY KEY REGULATORY GENES AND PATHWAYS—COMPARATIVE GENOMICS WITH MODEL ORGANISMS

Although genetic diversity in the human population plays a large role in the relative vulnerability of any one individual to an environmental insult, the use of model organisms, some of which are not genetically diverse, can provide insights into biological responses to a specific environmental toxin. It is clear that there are a variety of shared genes, functions and pathways in different organisms. The field of comparative genomics is really at a very early stage of development (19) and characterization systems like Gene Ontology functional classifications (20) are helping one to make these comparisons between species. In mice, recombinant inbred strains (21) between two different strains that vary considerably between their phenotype and response to environmental insult allow one to map quantitative trait loci (QTLs) so that one has the ability to identify and localize on the genome different regions of genes that affect the characteristic being explored. This is a powerful approach and can be made more precise as the number of inbred members is increased for any different strain combination.

However, to translate information with any confidence from one species to another requires that the species have orthologous genes and pathways. Although there are many critical genes and pathways relevant in both the developing organism and the adult, the class of ligand-induced transcription factors can conceptually intersect many pathways in an organism. In this respect, one finds considerable homology between human, mouse, zebrafish and even Caenorhabditis elegans for these factors (e.g. humans, mice and zebrafish have estrogen receptors that can be responsive to environmental estrogen-like compounds; aryl hydrocarbon receptors that can use PCBs and TCDD as ligands for inducing a variety of cytochromes p450s involved in the processing of foreign molecules in the body; and retinoic acid receptors which play significant roles in the development of their embryos). In addition, C. elegans possesses orthologs of many of the receptor and cognate signal transduction pathways present in higher organisms. Biological and mutant evidence with these receptor systems in the different species confirm the overlap in functions. As each of these receptor systems can serve as ‘sensors’ for an environmental challenge, the homology amongst species then allows one to use each species to its own technical advantage.

As an example, the neural tube of humans, mice and zebrafish is generated during the segmentation phase of embryonic development. Transient structures, somites, form from paraxial mesoderm and give rise to vertebrae and ribs, skeletal muscle and dermis of the skin. They also provide the migration paths of neural crest cells and axons from spinal nerves. Somitic segments are added on caudally and the neural tube develops in this caudal fashion flanked by the somites. In retinoid signaling, a crucial enzyme, Raldh2 is expressed in the somites and is responsible for converting retinal to retinoic acid. This small ligand can then be transferred to the developing neural tube to signal events in its differentiation. At the growing caudal end of the organism, another enzyme, cytochrome p450RAI (or cyp26a1) is synthesized and can metabolize retinoic acid providing a retinoic acid free zone at the growing caudal end near the neural tube. Mouse mutants of cyp26a1 produce phenotypes which include an open caudal neural tube (spina bifida) and at a lower frequency caudal fusions and truncations (22,23). Crossing a heterozygous Raldh2 mutant allele into this homozygous mutant cyp26a1 background suppresses these phenotypes (24). This suggests that lowering the RA concentration 50% during the development of the neural tube can have a significant effect. In zebrafish, we and others have shown that the Raldh2 gene is itself repressed by retinoic acid and the cyp26a1 gene is induced by retinoic acid (25).
Anterior to the developing trunk neural tube retinoid signaling also plays significant roles. As the hindbrain begins to segment into structures (rhombomeres) another retinoid metabolizing enzyme cyp26b1 appears (26) and basically creates a retinoid-free zone in the middle of the hindbrain that can play a significant role both in the expression of retinoid responsive genes such as the homebox genes and also in the derived neural crest cells that migrate from the hindbrain region. Figure 1 is an illustration of the similarities in retinoic signaling between mouse and zebrafish. Transgenic mouse and zebrafish lines were constructed and, as shown, at similar developmental stages reveal retinoic acid activity in the developing neural tube. Therefore, these retinoid pathway genes plus additional ones yet to be as closely studied could be genes involved in neural tube defects or genes whose expression might be affected by environmental toxicants. Within this context, there is developing evidence of potential intersection of retinoid events and those mediated by the aryl hydrocarbon receptor (AhR). Although this has yet to be developed in embryos, in human airway epithelial cells exposed to the AhR ligand TCDD, a series of genes within the retinoid pathway are affected (29) though synthetic retinoids have been shown to have an impact upon AhR driven gene expression (30). A genomic approach using microarray analysis in different species is currently the best way to examine these interactions between these two receptor pathways and ligands that affect them.

Regarding estrogen signaling, there is a long history of environmental chemicals having estrogen-like activity. There is clear evidence of environmental effects upon animals of chemicals that affect estrogen signaling and have significant effects upon sexual identity and reproductive health in animals. There remains clear evidence to be developed regarding environmental effects upon humans though our knowledge and understanding of the roles estrogen plays in humans and other species continues to increase. Part of this knowledge and understanding has been through the use of gene knockout technology in mice to produce lines of mice deficient in both the estrogen receptor-alpha and the estrogen receptor-beta. As one works with these models of mice missing one or both the receptors, information concerning their roles in several pathways of mammals and other vertebrates is accumulating, and will also aid the derivation of pharmaceutical solutions to problems associated with the loss of estrogen in women after menopause (bone, heart and cancer complications). As estrogen is generated from aromatases, which use testosterone as a substrate, the study of the regulation of aromatases can tie in both or these hormones and the effects upon both men and women. Common pesticides have been shown to have inhibitory activity towards aromatase activity (31), and in other species this might play a role in reproduction and sexual identity, and meconium analyses of fetuses in the Philippines have indicated strikingly high levels of pesticides and metals that might play an adverse role in several processes.
(32). In this regard, having a vertebrate organism such as zebrafish with receptors and enzymes homologous to those identified in humans allows one to study some of these processes and their links in a visually available vertebrate model.

However, with transcription factors such as these ligand-inducible receptors, the complete molecular repertoire of interacting genes in any one species has yet to be completely defined. Therefore, a comparative genomics approach amongst species with microarray analysis and bioinformatics approaches that allow the hypothetical creation of pathways should allow one to determine the functional homology of these different regulatory genes. The zebrafish model will allow one to efficiently knock-down individual genes in the pathway to test its efficacy. This type of approach has been elegantly used to dissect developmental pathways in the sea urchin (33,34). However, for performing such analyses across species, it requires a considerable collaborative effort between laboratories having expertise human conditions and a working knowledge with each model system being compared.

**DISCOVERY OF SUSCEPTIBILITY GENES**

Asthma is a very common chronic disease, affecting nearly five million children and 10 million adults in the US (35). The prevalence (36) and severity (35–37) of asthma have continued to increase over the past decade, despite major advances in the recognition and treatment of this condition. Perhaps most alarming is the doubling of the mortality rate attributed to asthma, since 1980 for children between 5 and 24 years of age (35). Asthma is caused by a number of genes and specific environmental triggers that result in the expression of a broad phenotype (or syndrome)—acute and chronic forms of airflow obstruction, airway inflammation and airway remodeling. As such, asthma should be considered a syndrome of chest tightness and reversible airflow obstruction that is caused by many biologically unique gene–gene and gene–environment interactions. The broad clinical phenotype of asthma, the polygenic inheritance pattern of this disease, and the substantial role of environmental exposures in the development and progression of asthma have complicated the search for asthma susceptibility genes.

The dramatic increase in the prevalence and severity of asthma over the last 20 years suggests that changes in the environment play an important role in the development and progression of this disease. Asthma occurs more frequently in indoor air pollution (47,48), viruses (49), domestic (50–53) and occupational (54–61) exposure to endotoxin, or immunization against certain infectious diseases (62) play a particularly important role in the etiology and pathogenesis of this condition. Moreover, avoidance of allergens and cigarette smoke appears to decrease the risk of childhood asthma (63).

Although classical Mendelian patterns of inheritance do not apply in asthma, there is little doubt that inheritance plays a role in this disease (3,14,64). Community-based twin studies of asthma indicate that the estimated heritability for asthma ranges from 35 to 75% (65–70). Candidate gene/loci studies have reported linkage of atopy and non-specific airway hyperreactivity to chromosomes 5q (71,72), 6p (73), 11q (74,75), 12q (76) and 14q (77). Among atopic sib pairs with a 12% prevalence of asthma, a genome-wide screen identified linkage to 4q, 6p, 7p, 11q, 13q and 16q (78). A genome-wide screen for linkage with asthma in three racial groups (African Americans, Caucasians and Hispanics) yielded six novel regions (2q33, 5p15, 11p15, 17p11, 19q13 and 21q21) and confirmed several previous loci (5q23–31, 6p21–23, 12q14–24, 13q21–qter and 14q11–13) (79); and a genome-wide search in a founder population (Hutterites) identified linkage to five regions—5q23–31, 12q15–24.1, 16p, 19q13 and 21q21 (80). Although genetic loci and candidate genes within and outside of these loci have been found to be associated with asthma phenotypes, major susceptibility genes for asthma have not been definitively identified and confirmed between study populations (14,15). We and others believe that these inconsistencies result from the complex clinical phenotype of asthma (10,11), the polygenic inheritance pattern of this disease (3,12–16), and the substantial role of environmental exposures in the development and progression of asthma (3,14,17).

Given these complexities, we have decided to take a different approach to identify asthma susceptibility genes. Our plan has been to use environmental stimuli to narrow the pathophysiologic phenotype so that we are able to investigate specific types of asthma, types of asthma that are defined biologically by environmental toxicants, rather than by a diagnosis that represents a broad spectrum of disease phenotypes (chest tightness, inhaler use, airway hyperreactivity, airway eosinophils, etc.) that may have different underlying causes and biological mechanisms. By using a specific exposure to narrow the pathophysiologic phenotype, we anticipate identifying genes that are biologically and potentially genetically relevant to a specific type of asthma that may prove generalizable to other forms of asthma.

Our approach to discover genes involved in asthma is designed as a two-stage study (Fig. 2). In the first stage (the derivation stage), we will use highly selected study subjects to identify genes that are differentially expressed by the airway cells of asthmatics following subsegmental airway challenge with agents that induce airflow obstruction through either acquired or innate mechanisms of immunity. Individuals from each of the four highly selected groups of study subjects (atopic asthmatics, non-atopic asthmatics, atopic non-asthmatics and non-atopic non-asthmatics) will have saline, house dust mite allergen and lipopolysaccharide (LPS, a specific form of endotoxin) instilled in a separate subsegmental bronchi. Four hours after instillation, bronchoscopy and endobronchial brush biopsy lavage will be performed in each of the three subsegmental bronchi that were challenged with either saline, house dust mite allergen or LPS. RNA will be extracted from airway epithelia, and this RNA will be used to probe high-density microarrays to measure the level of expression of genes/ESTs. To limit the
number of candidate genes to \( \sim 100 \), we will require that the candidate genes/ESTs be over- or underexpressed in airway epithelia of asthmatics following specific subsegmental airway challenge of either house dust mite allergen or LPS, and that the differential expression is confirmed by quantitative PCR. We will give preference to candidates that map to a region of previous linkage from reported asthma genomic screens, have biologic plausibility, and are selectively expressed by epithelial cells. Polymorphic markers will be chosen from available genetic mapping resources. In the second stage (the test stage), we will determine whether the differentially expressed genes derived from the first stage of this investigation play a role in the genetics of asthma by using these closely linked markers to genotype an existing, well-characterized familial population of asthmatics. We will then perform a linkage and family-based association study to identify potential asthma susceptibility genes with the positive findings independently tested in a separate study population. Candidate genes that are genetically related to asthma in the association/linkage study will be investigated for trait-associated polymorphisms and/or mutations. This approach is designed to identify novel genes that are associated with both asthma pathogenesis (differentially expressed in the exposure-response study) and asthma susceptibility (genetically associated with asthma in a linkage/association study).

**DEFINE DISEASE PHENOTYPES**

In addition to providing mechanistic information that can be used by the clinician to help in identifying susceptible individuals or environmental agents that could contribute to the etiology of a disease, environmental genetics/genomic provides us with the opportunity to specifically define diseases with complex or varied symptoms. Autism is an example of this type of disease. Autism is a neurodevelopmental disorder characterized by significant disturbances in social, communicative, and behavioral functioning. The onset of autism is in early childhood (usually \(< 3 \) years of age) with symptoms that continue throughout life (81,82). The most recent prevalence estimates for autism suggests it has a frequency of 34/10 000 (1–300) (83) making it more frequent than Down syndrome \([1/1000; (85)]\). Autism is characterized by an increased male:female ratio of 3 or 4:1 (85,86). Autism is not a single disease, and it is typically referred to as Autism Spectrum Disorders. This umbrella term that includes classic autism (Kanner’s autism), Asperger’s syndrome and pervasive developmental disorder (PPD), Autism is considered as a spectrum disorder because of a variation in the intensity of the symptoms. To identify/classify children with autism clinical staff assess the level of sensory impairment (e.g. visual impairment or hearing loss) or motor impairments (i.e. failure to attain basic motor milestones of sitting and walking by 12 and 24 months, respectively), the presence of metabolic disorders and use Social Communication Questionnaires, a sensitive measure of symptoms suggestive of an autism spectrum disorder. The SCQ is a parent/caregiver screening instrument on the basis of the current diagnostic criteria for autism (87). The SCQ consists of 40 items, which assess communication skills and social functioning and can be used with all age groups. Many of these assessment tools are limited in that they are subjective and susceptible to interpretation bias.

Substantial evidence supports a genetic etiology for autism including (a) associations with single-gene disorders, (b) twin studies and (c) an increased recurrence risk for siblings of autistic probands, when compared with controls estimated to be 2.2% for classic autism and up to 5% when Asperger’s syndrome and other PDDs are included (88–92). Although the sibling recurrence risk is low, the relative risk is 50–100 times greater than that in the general population (90,91). Heritability is therefore estimated at 90%, which is among the highest for a psychiatric disorder (84). The high heritability in autism has driven efforts by several groups to identify susceptibility genes using genome-wide linkage screens in multiplex families (93–99). At least nine genomic screens and follow-up analyses (93–97,99–101) have been performed in search for autism candidate regions. Comparing the results from all studies identifies only a number of genomic regions in common across multiple studies including regions on 7q and 2q (~30 cm), but a prominent gene has not emerged as a positional candidate. Additional areas of interest include a linkages region on chromosome 19q (~40 cm), and chromosome 16 and a region on chromosome 15 near the GABA gene complex (102,103). Chromosome 2 and 19 are other areas of interest for autism susceptibility. Linkage to the chromosome 2q31–q33 region has been identified by four different studies (94,96,99,101). The complexity in identifying susceptibility loci underscores the likely genetic heterogeneity for this disorder. Although the data do not strongly endorse any one model for inheritance, twin and family studies support a multilocus etiology with as many as 10–20 interacting loci (91,95).

The role of environmental toxicants in the etiology of autism is controversial, and no significant environmental risk factors have been implicated in the development of autism. It has been proposed that mercury may be a contributing factor in the development of this disorder (104,105). Because
methymercury is a known neurotoxicant, prenatal exposure can contribute to impaired psychomotor and cognitive function (106). Methymercury exposure has been associated with blindness, cerebral palsy, spasticity and seizures. Many features of mercury poisoning overlap with features of autism including immune, sensory and behavioral problems (104). However, mercury also affects the peripheral nervous system and other organs that are normal in individuals with autism (107).

Because of the difficulty in identifying the multiple genetic loci and environmental factors that contribute to the etiology of autism, alternative eukaryotic model systems are being utilized. Clearly, the complex disorders in social, communicative and behavioral functioning cannot be replicated in zebrafish or nematodes. However, by using easily manipulated, genetically tractable model organisms environmental agents that affected neurological function (movement, feeding) and learning can be quickly identified. Further, target genes that are affected by these agents can be identified, and the relation between affected genes and neurological function quickly assessed. Ultimately, the orthologs of the toxicant responsive genes identified in the model organism will be identified in humans. These genes can serve as targets in lineage analysis to identify the loci associated with autism. Similar approaches can be used for other polygenic diseases including neural tube defects.

**FUTURE DEVELOPMENT OF ENVIRONMENTAL GENETICS/GENOMICS**

The fields of environmental genetics and environmental genomics has enormous potential to affect our ability to accurately assess the risk of developing disease, identify and understand basic pathogenic mechanisms that are critical to disease progression. In addition, they can help to biologically phenotype a disease so that we can accurately classify, stage and prognosticate the disease process for patients and their families. However, the application of genetics and genomics to problems in environmental health is only the beginning yet, by itself, represents a potentially effective strategy to substantially impact morbidity and mortality. Collaborative approaches that team together environmental scientists with molecular biologists, geneticists, physiologists and physician scientists are critical to the investigation of environmental aspects of human health. Moreover, exploiting eukaryotic model systems (yeast, *C. elegans*, zebrafish, *Drosophila* and rodents) will accelerate our understanding of environmental exposures on human health. Although this field is intellectually challenging, environmental genetics/genomics represents a unique opportunity for geneticists to work with other scientists to understand disease pathogenesis and improve human health.

**ACKNOWLEDGEMENTS**

This manuscript was supported by grants from the National Institute of Environmental Health Sciences (ES11375, ES012496 and ES011961) and the Department of Veterans’ Affairs (Merit Review).

**REFERENCES**


