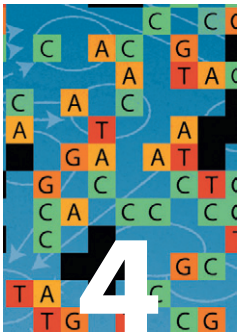

Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

CHAPTER




# Gene–Environment Interactions: Eco-Genetics and Toxicogenomics

*Gilbert S. Omenn*

## INTRODUCTION

The explosion of genome-wide association studies (GWAS) for diseases and for quantitative traits has generated a surprising and, to some, disappointing result: even with many associated genomic variants and even for conditions long known to have high "heritability", the population attributable risk accounted for by the variants identified thus far is very small, typically less than 10% (Omenn, 2010a; Thomas, 2010a, b). What could be the explanations?

The "missing heritability" is widely called the "dark matter" to connote our current lack of understanding and the relatively vague notion of what "heritable" factors might be responsible. For such conditions as diabetes, high blood pressure, high cholesterol values, obesity, or height, we know from epidemiological risk factor studies and from clinical management of patients that dietary and behavioral factors play on the inherited substrate to create the phenotypes. It is important to remember that, in the calculation of heritability, $h^2$ = genetic components (G) + gene–environment interactions (G × E), so that gene–environment interactions are "counted" as part of the heritable fraction of variation (Vogel and Motulsky, 1997). Endogenous but non-genetic factors, like hormone levels, pathways of inflammation, and neural signaling, and the enormous microbiome, interact with genetic variation. The broader environmental landscape of physical, chemical, and infectious exposures, the built environment, and social interactions surely must be considered as well. In addition, there are gene–gene interactions, transmitted epigenomic marks, copy-number variants, and alternative splicing events that must be considered in any models of heritability.

Another difficulty in the interpretation of GWAS for most diseases or traits is the fact that most genomic variants tested in such studies do not occur in the protein-coding regions of the genome. That should not be a surprise either, since only 1.2% of the genome accounts for the estimated 20,000 genes that code for proteins. The rest of the genome offers many possibilities of gene regulatory mechanisms, including promoter regions, enhancers, small RNAs, transposons, and even endogenous retroviral sequences. Figure 4.1 shows a scheme from the University of Michigan-based National Center for Integrative Biomedical Informatics for exploiting multi-level *'omics* datasets about the genome, gene regulation, and gene, protein, and metabolite expression that reflect interactions with external variables to produce various phenotypes. The small yield of variants occurring in protein-coding regions, let alone producing non-synonymous, function-changing mutations in those proteins, limits the interpretation of the functional consequences of these variants using bioinformatics and data mining techniques.

*Genomic and Personalized Medicine, 2nd edition*
*by Ginsburg & Willard. DOI: http://dx.doi.org/10.1016/B978-0-12-382227-7.00004-5*

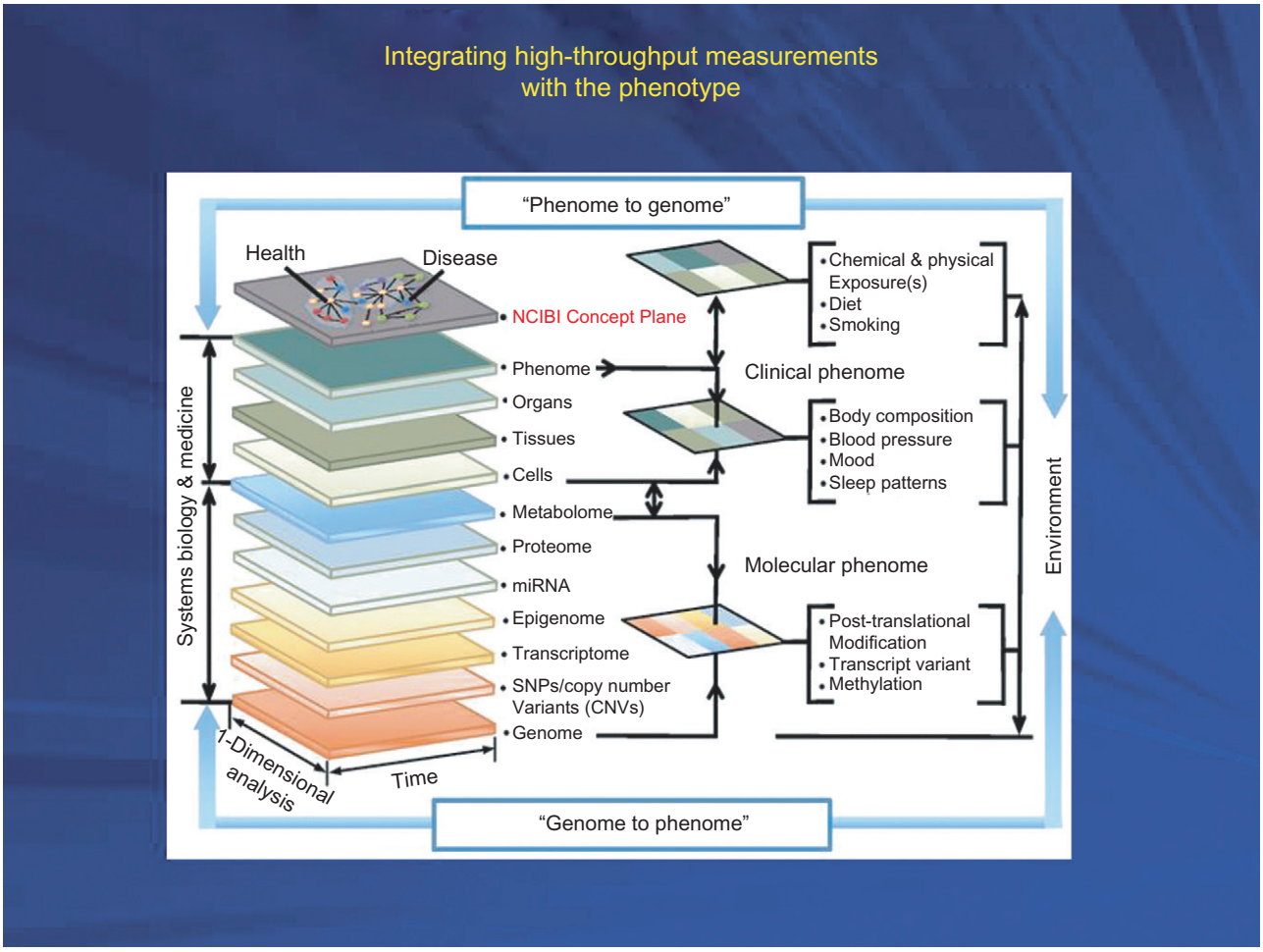



Figure 4.1 From genome to phenome and phenome to genome. Integration of high-throughput measurements of the genome, transcriptome, proteome, and metabolome with phenotypes at cellular, organ, and whole-person levels can capture both genetic variation and behavioral and environmental exposure variables (GS Omenn & BD Athey, University of Michigan).

## ECO-GENETICS

"Eco-genetics" is the term introduced by Brewer (1971) and carried forward by Motulsky and Omenn (Omenn and Motulsky, 1978) and many others to denote the interactions of genes with many categories of environmental factors, as outlined in Table 4.1. There are numerous examples in each category, some of which will be highlighted in this chapter.

As noted by Olden in the first edition of this book (Olden, 2009), the concept of gene–environment interactions dates back more than a century to Archibald Garrod's report of the first inborn errors of metabolism and his prediction that individuals would vary in their responses to diet and drugs (Garrod, 1902). In the 1950s, Motulsky, Kalow, Lehman, and Vogel recognized the importance of genetic variation in metabolism of drugs and target-based responses to drug action, and introduced the term "pharmacogenetics." Studies of monozygotic versus dizygotic twins showed striking heritability for rates of drug metabolism involving diverse drugs (Vesell, 1991). Now there are chips with probes for polymorphisms of the cytochrome P450 xenobiotic metabolizing enzymes and for other drug-metabolizing enzymes, which can be useful in choosing initial doses and modulating doses of the relevant drugs (see Chapter 31). Under Olden's leadership

Table 4.1 Eco-genetics: categories of environmental agents with known inherited predispositions

| |
|---|
| Foods and food additives |
| Infectious agents |
| Inhaled air pollutants |
| Pesticides |
| Drug chemicals |
| Sensory stimuli |
| Allergic and sensitizing agents |

in the 1990s, the National Institute for Environmental Health Sciences (NIEHS) of the US National Institutes of Health embarked upon the Environmental Genome Project and established the Center for Toxicogenomics.

## METHODS TO STUDY G × E INTERACTIONS

Thomas has thoroughly reviewed a broad range of study designs for G × E interaction protocols (Thomas, 2010a, b), updating earlier methods reviews (Kraft and Hunter, 2005; Yang and Khoury, 1997). Among the standard case-control, cohort, nested case-control, and case-cohort epidemiological study designs, cohort studies have the desirable feature of longitudinal exposure data collection. Counter-matching and two-phase designs permit stratification jointly on disease and exposure status, based on sampling probabilities. A variant that counter-matches each case with three controls on surrogates for both exposure and genotype can be substantially more efficient than counter-matching only on exposure, only on genotype, or neither. Family-based designs, including case and parents triads, generate a higher proportion of genotype-concordant, exposure-discordant case-control pairs than in usual case-control series. Mendelian randomization separately tests the associations of an environmentally modifiable intermediate variable and a disease with a gene that appears to influence it. For example, the gene *MTHFR* governs a critical step in folate metabolism; the genotype acts as an instrumental variable under the assumption that the gene has no effect on the disease independent of the intermediate phenotype (the enzyme activity). Emerging methods focus on multiple genes and multiple environmental factors related to biomarkers, pathways, modules, and functional systems. Hierarchical Bayes modeling is well suited to these studies. Applications include the WECARE study of radiotherapy-associated risk of second breast cancers in relation to DNA damage response pathways, and colon cancer candidate gene studies in relation to folate.

Challenge studies in exposure chambers are feasible to study some kinds of environmental exposures. For example, atopic individuals have been studied for effects of diesel exhaust particles and allergens on immunologic markers, using a double-blind factorial randomized crossover design; individuals with either *GSTM1* null or *GSTP* I105 genotype had higher IgE levels in response to either exposure, and still higher levels if they had both genotypes (Gilliland et al., 2004). Ancillary studies in randomized clinical trials can be efficient designs as well, since both treatment and genotype are randomly assigned. Subsets can be genotyped based on their treatments or outcomes, or can be selected for biomarker measurements based on treatment and genotype.

Exposure assessment is critical to G × E studies, as it is for environmental epidemiology generally. Exposure variables are often complex, with continuous scales of measurement that are time-dependent, multidimensional, spatially correlated, and sometimes available only at an aggregate level, as with ambient air pollutants. With ambient measurements, exposure doses for individuals must be imputed or predicted with models using time–activity and dispersion patterns, potential chemical modifications, subsample data, and transport and fate mapping. Much larger samples are required for interactions than for detection of main effects, typically at least four times larger (Thomas, 2010b). Measurement errors and uncertainties of exposure estimates put conclusions about G × E interactions at risk. Another complication is the use of individual single nucleotide polymorphisms (SNPs), because individual genes often have multiple variants. Thus, haplotype analysis is more reliable and more efficient.

Genetic, environmental, and behavioral variables jointly influence the risks of many diseases and traits. It is hard enough to discover, confirm, validate, and generalize across populations robust single-factor disease associations. Dose–response relationships for genotypes and especially for environmental exposures are critical to obtaining credible evidence.

In quantitative terms, two or more different risk factors (genetic or non-genetic) can appear to interact in several ways:

- They may each give statistically and clinically significant increases in risk of onset (or severity of progression) of the disease; the effects of the interaction between factors $a$ and $b$ may then be additive ($a + b$) or synergistic ($>>a + b$), as, for example, in the case of cigarette smoking and occupational asbestos exposure in causing lung cancers.
- Neither of the risk factors may be detectable as increasing incidence rates for the disease, but the two together may do so, as in the case of red meat and *NAT2* (see later).
- One of the factors may be predisposing silently, while the other may increase the risk, as with alcohol or smoking.

## NUTRIGENOMICS

It is obvious to laypeople that individuals differ in their metabolic and behavioral responses to the same or similar diet (Simopoulos, 2010). GWAS have been performed for type 1 and type 2 diabetes and cardiovascular disorders, and for quantitative traits like stature, blood pressure, blood glucose, and blood lipid levels. In general, dietary and nutrient information is not collected in these studies despite the high probability that diet is influential in these complex diseases. Similarly, G × E interactions are typically ignored in official recommendations of daily requirements for nutrients, individual medical evaluations, and public health campaigns. For example, general guidance to increase the polyunsaturated content of the diet to decrease plasma cholesterol levels and coronary heart disease risk may not be appropriate for women with the *APOE E3/E2* phenotype. Increasing omega-3 fatty acids (EPA, DHA)

and decreasing omega-6 fatty acids (arachidonic acid) may be protective against coronary heart disease and prostate cancers, interacting with the 5-lipoxygenase promoter polymorphism and with *COX2* variants, respectively. Intensive exercise, nutrition, and lifestyle changes can modify gene expression patterns, so it would be wise to analyze both gene expression and SNPs or haplotypes. The gene *FTO* (Cecil et al., 2008) affects adipose tissue mass, increases body mass index, and thus predisposes to higher diabetes risk.

During human evolution, drastic changes in the diets of most people have altered seven critical nutritional characteristics of ancient diets: glycemic load, fatty acid composition, macronutrient composition, micronutrient density, acid–base balance, sodium–potassium ratio, and fiber content (Cordain et al., 2005). Most of the food types that dominate present diets were introduced only recently: dairy products, cereal grains, refined sugars, alcoholic beverages, salt, refined vegetable oils low in omega-3 and high in omega-6, and mammalian red meats rich in omega-6. These foods displaced wild plant and animal foods of our predecessors, and contribute to obesity, diabetes, atherosclerosis, high blood pressure, dyslipidemias, osteoporosis, bowel disorders, inflammatory and autoimmune diseases, and several cancers. All of these diseases are rare among hunter-gatherers even today; most are associated with insulin resistance. Presumably, natural selection over many generations acted upon numerous gene variants available in our genome. The most studied G × E nutritional relationship involves evolution of the persistence of intestinal lactase activity, facilitating tolerance for lactose ingestion (from milk) after the time of weaning (Omenn, 2010a; Simopoulos, 2010).

## PATHOGENS AND HOST SUSCEPTIBILITY OR RESISTANCE GENE VARIANTS

Studies of infections reveal many host–pathogen interactions, with genetic variation in both the host and the pathogen genomes (Garantziotis and Schwartz, 2010). Many susceptibility genes are part of the innate or adaptive immune systems. The innate immune system notably recognizes common, conserved microbial antigens, now called pathogen-associated molecular patterns (PAMPs) through toll-like receptors. The TLR2 receptor recognizes lipoteichoic acid and peptidoglycans on Gram-positive bacterial or fungal walls, while TLR4 recognizes endotoxins in Gram-negative walls. G × E studies have had practical value in diagnosis, epidemiological surveillance, and targeted therapies of multiple infectious diseases. Genomics has played a critical role in understanding the evolution and pathogenicity of antigenic variation of influenza strains, origins of HIV/AIDS retroviruses, transmission of severe acute respiratory syndrome (SARS) (and other viruses) from animals to humans, spread of drug-resistant tuberculosis, and highly differential susceptibility to the vivax form of malaria (Omenn, 2010b). Hyporesponsive *TLR4* polymorphisms and a *CD14* SNP associated with higher circulating levels of CD14 are associated with increased mortality from sepsis in adults and from respiratory syncytial virus (RSV) in high-risk infants due to prematurity or congenital heart disease. SNPs in the vitamin D receptor gene *VDR* are associated with increased susceptibility to severe RSV and tuberculosis (TB). Conversely, an S180L *TIRAP* polymorphism is associated with reduced response to TLR2 and TLR4 activation *in vitro* and risk for developing TB. As more is learned about predisposition or resistance to infections, screening of potentially or definitely exposed animal-care and healthcare workers may be an important application.

The coagulation system is important in responses to infections. Plasminogen activator inhibitor (*PAI-1*) polymorphisms can increase or decrease susceptibility to community-acquired pneumonias (CAP) and to invasive aspergillosis in immune-compromised patients. SNPs in the heat shock protein gene *HSP70* and lymphotoxin alpha (LTA) or human leukocyte antigen (HLA) genes have been associated with increased susceptibility to severe CAP or TB. A polymorphism in the adenosine receptor *P2X7* gene has been reported to be associated with dissemination of TB.

The normal lung has long been described as sterile (Garantziotis and Schwartz, 2010). Now genomic analysis of unculturable microbes shows that there is a rich ecology of microorganisms, whose composition changes drastically in the course of chronic obstructive pulmonary disease (COPD) (enriched for *Pseudomonas*).

Observant clinicians have long recognized that some individuals are remarkably resistant to infections despite extensive exposure. Resistance to infectious disease can be a powerful positive selection factor in the emergence of particular genotypes. Childhood survival in areas with endemic *Plasmodium falciparum* malaria is enhanced if the child carries a gene for hemoglobinopathies (HbS or HbC), thalassemia, or glucose-6-phosphate dehydrogenase deficiency, all of which make the erythrocyte less able to host the malaria parasite. Susceptibility to infection by HIV and progression to AIDS is another salient example. A homozygous 32-amino-acid deletion in the co-receptor CCR5 renders the individuals impervious to HIV penetration of T cells to initiate the infection (Dean et al., 1996; Smith et al., 1997). This same deletion has been found in small numbers of individuals in many parts of the world. Multiple other gene variants are associated with faster or slower progression of HIV.

McNicholl and colleagues discussed host–pathogen interactions for five very important pathogens and the resulting diseases: *Mycobacterium* TB, HIV-1, hepatitis B, *Plasmodium*, and *Vibrio cholera* (McNicholl et al., 2000). Genetic variation in the human genome has been emphasized above; however, research on the genetics of the pathogens is equally important, especially for the design of vaccines and drugs effective and selective for particularly pathogenic strains.

## INHALED CHEMICAL AIR POLLUTANTS AND ALLERGENS AND SENSITIZING AGENTS

The lung is highly exposed to air pollutants from industry and mobile vehicles, aerosol toxins, dusts, fumes, and infectious agents, in numerous occupational, medical, indoor, and outdoor environments. Risks of pulmonary impairment and disease depend upon susceptibility, exposure, clearance, immune responses, and repair mechanisms. G × E interactions are important in essentially all respiratory diseases. Epidemiologic analyses of gene–particulate interactions can be combined with toxicologic assays of biological effects of particle samples on cell lines with the same genes modified (Thomas, 2007); this study of exposures to complex mixtures is an example of the NIEHS Genes, Environment, and Health Initiative (http://www.gei.nih.gov).

Asthma has frequent familial occurrence and high estimates of heritability. Much improved hygiene and delayed exposure to sensitizing agents have increased susceptibility over recent centuries. Striking increases in incidence, prevalence, and severity of asthma in recent years, especially in developed countries, are presumed to be due to environmental exposures, ranging from mites and cockroaches to chemicals. Prenatal exposures to diesel exhaust particles and environmental tobacco smoke are associated with increased risk of asthma, while maternal ingestion of fruits, vegetables, and oily fish appears to be associated with lower risk; exposure to an environment rich in microbial compounds seems to reduce the development of atopy and down-regulate toll-like receptors. Genomic imprinting and immune interactions may account for the four-fold higher risk of mother-to-child transmission of atopic disease than father-to-child. At least 10 genome-wide linkage and association studies have been completed, identifying potential associations with genes on 20 different chromosomes; however, the effect sizes are small, and few thus far have been confirmed.

In the US, the Clean Air Act requires protection of sensitive subpopulations from adverse health effects of criteria air pollutants. For example, in setting the ambient air standard for photochemical smog (ozone), the US Environmental Protection Agency (EPA) stated that its aim was to protect subpopulations with asthma, emphysema, or chronic bronchitis. Asthma has a prevalence in the US of 22 million adults and 9 million children. Note that EPA did not propose to protect the much smaller (and unstudied) population with cystic fibrosis. EPA has issued an *Interim Policy Guidance on Genomics* (EPA, 2004). For asthma, predisposing gene variants have been noted in genes for IL-4, IL-13, TNF-alpha, β2-ADR, Fc-RI-β, and IL-4R, mediated by influences on inflammation, hyperresponsiveness to airway irritants, and airway constriction and obstruction. There is an association among asthma symptoms, exposure to cigarette smoke, and SNPs in the genes for the xenobiotic enzymes EPHX1, CYP1B1, and CYP2D6 (Kramer et al., 2006).

For COPD (emphysema and chronic bronchitis), smoking is the dominant risk factor in the US, while similar chemical exposures from stoves and open fires are more responsible in less developed countries. Still, many highly exposed individuals do not develop COPD, so there must be substantial variation in susceptibility. Outliers with very high susceptibility may be rare homozygotes for alpha-1 anti-trypsin deficiency; still, they require cigarette smoke exposure to develop severe emphysema. SNPs apparently predisposing to emphysema include inflammatory genes *TNF-alpha* and *TGF-beta*; antioxidant genes *GSTM1*, *GSTP1*, and *HMOX-1*; and the xenobiotic metabolizing enzyme gene *EPHX1* (Garantziotis and Schwartz, 2010). Meanwhile, genes predisposing to nicotine/smoking addiction include nicotine-metabolizing P450 enzymes, nicotine receptors, and genes in dopamine and serotonin pathways. Interestingly, genes predisposing to addictive behaviors also increase the relevant environmental exposure.

Idiopathic pulmonary fibrosis (IPF) and interstitial lung diseases constitute a highly heterogeneous group of serious lung disorders with highly variable prognosis; many biomarkers are under investigation. Familial interstitial pneumonia occurs in several types of interstitial lung disease, with interaction between cigarette smoke and several predisposing genes, such as *ELMOD2*, *TERT*, and *TERC*, surfactant proteins C and A2, and various chromosomal regions. Metal- and wood-dust exposures can also trigger IPF.

## EXAMPLES OF RISK FACTORS

### *Cigarette Smoking, Bladder Cancers, and Colorectal Cancers*

One of the best-established G × E interactions involves cigarette smoking and polymorphism in the N-acetyltransferase (*NAT2*) gene in predisposition to urinary bladder cancers. A Spanish case-control study reported a 1.6-fold interaction odds ratio (OR) comparing NAT2 slow acetylators versus rapid or intermediate (heterozygous) acetylators, and smokers versus non-smokers (Garcia-Closas et al., 1998). The slope of the dose–response relationship for pack-years of smoking was modified by *NAT2* variants in an intensity-related manner. Four other European case-control studies have confirmed this interaction, mediated by detoxification of heterocyclic and aryl amines in smoke (as well as in hair dyes) (Thomas, 2010b). NAT2 much earlier was shown to influence strongly the susceptibility to occupational beta-naphthylamine exposures in the dye industry in multiple countries.

A less striking interaction has been reported for colorectal cancers, starting with a weak main effect of well-done red meat, a source of heterocyclic amines, together with the predisposing gene polymorphisms low *CYP1A2* and slow *NAT2*, with or without cigarette smoking, which can induce CYP1A2 activity. An OR of 8.8 was found for those who were heaviest smokers and genetically susceptible, compared with never-smokers not predisposed (Le Marchand et al., 2001).

### Heavy Metals

Hu and colleagues hypothesized that low-level lead exposures and elevated free iron levels might contribute to cognitive decline in older people through promotion of oxidative damage. They investigated prevalent polymorphisms in the gene for predisposition to hemochromatosis (*HFE*) and iron overload. Higher lead levels were associated with steeper cognitive decline among participants in the Normative Aging Study who had at least one *HFE* variant allele, compared with men with only wild-type alleles, and this was steeper with additional variant alleles (Wang et al., 2007). In the same population, there was evidence that *ALAD* genotype may modify the effect of lead on the renal excretion of uric acid as well as on overall renal function among middle-aged and elderly men who had community (nonoccupational) exposures to lead (Wu et al., 2003). Cantonwine and colleagues investigated the role of *HFE* C282Y, *HFE* H63D, and transferrin (*TF*) P570S gene variants in modifying the association of lead and infant birth weight in a cohort of Mexican mother–infant pairs. Interaction models indicated that maternal *HFE* H63D variant carriers had a negative association between tibia lead and birth weight (Cantonwine et al., 2010).

### Pesticides

Just as with pathogens, people differ remarkably in their susceptibility to organophosphate (OP) pesticides and even nerve gases such as Sarin (Costa and Eaton, 2006). About 25% of Asians and 10% of Caucasians have a variant of the gene that codes for paraoxonase (*PON1*), the enzyme that converts the OP pesticide parathion to the active neurotoxic metabolite paraoxon. The variant has 10-fold higher activity to detoxify the oxon of various OP pesticides to harmless end-metabolites (Costa and Furlong, 2002). This very interesting enzyme circulates in the blood within particles carrying high-density lipoproteins, and is associated with risk of cardiovascular disease.

## EPIGENOMICS AND THE ENVIRONMENT

Much more powerful than statistical associations are demonstrated mechanisms. A whole new category of mechanism has emerged from the study of epigenetics, the analysis of numerous covalent modifications in histones, the proteins in nucleosomes and chromatin, and methylation of cytosines in the DNA itself (see Chapter 1). External environmental agents can mediate these modifications, which control the transcriptional activity of specific genes, at specific points in time, in specific organs.

Unlike "alleles," which are defined in a purely genetic sense, "epi-alleles" do not differ in genome sequence; their information resides in self-propagating molecular signatures that provide a memory of previously experienced stimuli, without irreversible changes in the genetic information of DNA. For example, the methylation pattern of six genes is associated with development of lung cancer (Belinsky et al., 2006). Certain drugs can modify or reverse these modifications, opening avenues to potential therapies and chemoprevention (Andersen et al., 2010). For example, two DNA methylation inhibitors, 5-aza-deoxycytidine and 5-azacytidine, have been approved by the US Food and Drug Administration for treatment of myelodysplastic syndrome, a preleukemic disease. Compounds that inhibit histone deacetylase (HDACi) have proapoptotic and antitumor properties, and are undergoing phase I trials.

As one example, Schwartz has summarized knowledge about epigenetics in the context of respiratory diseases, especially COPD and asthma (Schwartz, 2010). Epigenetic mechanisms have profound effects on cellular, tissue, and whole-organism phenotypes. Hypermethylation of CpG motifs, particularly at promoter and enhancer sites, silences gene transcription, while hypomethylation of these motifs enhances transcription. Histone posttranslational modifications alter chromatin structure. Together with non-coding RNAs and protein transcription factor binding patterns, these mechanisms regulate the expression of specific genes, at specific stages of development, and in response to specific endogenous and exogenous stressors. A convincing experimental demonstration of the profound effect that epigenetic mechanisms can have on *in vivo* phenotypes is the coat color of the Agouti mouse, with yellow governed by a transposable element and brown due to methylation of the gene promoter, driven by a diet rich in methyl donor compounds (Waterland and Jirtle, 2003). Epigenetic mechanisms might well account for many of the differences between identical twins for various traits or diseases, especially as they age. Epigenetic marks (e.g., methylation sites) can persist from generation to generation, although the mechanisms are unclear.

In humans with obstructive lung disease, bronchial biopsies and alveolar macrophages in lavage fluid show increased histone acetyltransferase (HAT) activity and reduced histone deacetylase (HDAC) activity (Cosio et al., 2004; Ito et al., 2005); treatment with steroids reverses this pattern and reduces airway inflammation. *In utero* supplementation with methyl donors alters locus-specific DNA methylation and predisposes mice to allergic airway disease by directing the differentiation of T lymphocytes, with a skewing toward a Th2 phenotype (Hollingsworth et al., 2008). A total of 82 distinct loci were differentially methylated. This "epi-phenotype" could be reversed with demethylating agents, consistent with epigenetic plasticity. However, before we consider restricting methyl groups in pregnant women's diets based on results in mice, we must recognize that there is definite evidence of striking benefit from folate supplementation and fortification of foods in preventing neural-tube-closure congenital defects.

These findings are supported by a study of a birth cohort of 32,077 children. Perinatal folic acid supplements were associated with an increased risk of wheezing at 18 months of age (Haberg et al., 2009). Tobacco smoke is another *in utero*

exposure that is associated with childhood asthma, and can modify gene expression through DNA hypermethylation (Digel and Lubbert, 2005). In contrast, maternal exposure to a farming environment exerts a strong protective effect on asthma and allergy development in the offspring (Ege et al., 2008).

## EFFECTS ON PROTEIN FOLDING

Protein folding is "the final step in the decoding of genetic information," and prions represent what has been called "epigenetics in the extreme" (Halfmann and Lindquist, 2010). Prions exist in very different stable co-conformational states, and mediate inheritance of environmentally acquired traits. They occasionally fold into a conformation that replicates itself by templating the conformational conversion of other molecules of the same protein. Originally conceived to explain kuru and "mad cow disease," *de novo* prion formation proceeds through a high-energy oligomeric nucleus to a well-ordered fibrillar protein polymer, which is severed into smaller, actively growing pieces by protein remodeling factors such as disaggregases, and disseminated to daughter cells. The switch to the prion state can modify cell adhesion, nutrient use, and resistance to various toxins or antibiotics, with variation according to strain backgrounds. The response to stress constitutes an evolved bet-hedging strategy, allowing a fraction of cells to try new phenotypes that may prove beneficial and are transmissible and heritable. In mammals, prions may mediate cell-remodeling processes and memory formation. Protein misfolding is at the heart of major neurodegenerative diseases and of pro-insulin toxicity in pancreatic beta-cells.

## TOXICOGENOMICS AND PREDICTIVE TOXICOLOGY

The National Institute for Environmental Health Sciences (NIEHS) in the 1990s created a Center for Toxicogenomics to examine the effects of environmental exposures, especially industrial, agricultural, and pharmaceutical chemical exposures, on the genome, mRNA gene expression, and protein expression (toxicoproteomics) (Ramos and Olden, 2008). The combination is a systems and pathways approach to toxicology (Figure 4.1).

Most toxicologic studies are performed in experimental animals, especially rats and mice. A clever use of the large number of inbred strains of mice is the demonstration of 26 population-wide biomarkers of acetaminophen-induced hepatotoxicity, in which the changes in gene expression were significant for the treatment and the extent of necrosis, but not associated with any of the individual strains (Harrill et al., 2009). Pathway analyses showed many of the biomarkers were part of intracellular signaling for apoptosis. Studies in rats showed correlation with human toxicity reflected in gene-expression changes in peripheral blood cells (Bushel et al., 2007). The same approach has been applied to developmental toxicity (Daston and Naciff, 2010). Another approach is to compare profiles of differential gene expression in target and non-target organs, illustrated with methapyrilene (Auman et al., 2007).

The Comparative Toxicogenomics Database (CTD; http://ctd.mdibl.org) is a manually curated, public resource of the triad of chemical–gene, chemical–disease, and gene–disease relationships, integrated to construct chemical–gene–disease networks. As of July 2010, CTD contained 1.4 million triad data points, and analytical tools like GeneComps and ChemComps to find comparable genes and chemicals that share toxicogenomic profiles, enriched Gene Ontology terms, and Venn diagram tools to discover overlapping and unique attributes of any set of chemicals, genes, or disease, and enhance gene pathways data. CTD is indexed at numerous other databases, including PubChem, PharmGKB, UniProt, T3DB, GAD, ChemID, and TOXNET (Davis et al., 2011). Toxicogenomics also needs to be linked with databases of gene variants and eco-genetic relationships. Datasets from microarray and proteomics studies in various species are available at CEBS, the Chemical Effects in Biological Systems knowledgebase (Waters et al., 2008) (http://cebs.niehs.nih.gov/cebs).

As part of the Gene–Environment Initiative, NIEHS has put a major emphasis on biosensors and other technologies for automated quantitative monitoring of ambient exposures and body burdens (Weis et al., 2005). As noted earlier, such improvements in exposure assessment are critical for G × E interactions.

## WILL PERSONALIZED GENOMIC RISK PROFILES MOTIVATE PEOPLE TO ADOPT MORE HEALTHFUL BEHAVIORS?

Commercial genotyping services are becoming quite popular (see Chapter 6) and have attracted the attention of state-level and federal healthcare regulators. They will be superseded soon by whole-exome or whole-genome screening services for those willing to pay. In fact, knowing one's genotype in advance before starting on a variety of important drugs with pharmacogenomic (PGx) variants affecting efficacy or safety could provide critical guidance for initial dosage and for monitoring for adverse effects. There is a valuable data resource at PharmKGB (Hansen et al., 2009). In the example of the blood anti-coagulant coumadin (warfarin) for heart and stroke patients, variants in two major genes (*CYP2C9* and *VKORC1*) lead to much prolonged time until anticoagulation is in the effective range. Most knowledgeable physicians currently continue to administer and monitor coumadin without ordering the PGx tests. If the genotypes were available at the first decision point, the information would surely be used and would be expected to be critical in a small percentage of cases who are extreme outliers (Omenn, 2009). The same might be true

for the anti-clotting drug clopidogrel (plavix); known polymorphisms greatly affect the activity of this drug, used to prevent strokes and heart attacks (Hansen et al., 2009).

As promising as some of these advances are, the starting point for testing must be credible predictive information about disease risks. At present nearly all the disease-associated genomic variants discovered with GWAS contribute only a very small increased risk alone or even in aggregate. Practical public health questions about the utility of personalized genetic and genomic information and the cognitive capacity of individuals and families to process and benefit from such knowledge have been anticipated (McBride et al., 2010). There are many potential benefits: lifestyle changes that reduce disease risk, timely screening, improved adherence to disease prevention medications or other programs, and willingness to engage family members in similar actions. There could be negative impacts, such as fatalism or reduced sense of personal control, although these have not been documented as yet. An excellent test case might be whether evidence of genetic predisposition to smoking and smoking-related diseases can induce significant behavior change to reduce risks.

## FUTURE PROSPECTS: BALANCING THE GxE EQUATION

Discovering robust G × E interactions is a formidable challenge, and replicating or validating the findings is even harder, based on experience to date and methodological considerations (IOM, 2012). SNPs, haplotypes, and now high-throughput exome and whole-genome sequencing provide much more power on the genetic side of the equation. The 1000 Genomes Project and other applications of new sequencing methods have enabled the comprehensive identification and mapping of common and rare variants in individual genomes. The map currently includes the location, allele frequency, and local haplotype structure of over 24 million SNPs, 1 million short insertion–deletion polymorphisms, and thousands of structural variants, many of them novel. As of October 2010, >1000 genomes have been sequenced and >95% of the accessible variants in any individual genome are available for analysis (Durbin et al., 2010). The results show that low-coverage sequencing of many individuals can produce highly accurate individual genotypes at shared sites, demonstrate that the hitch-hiking effect has a marked effect on genetic variation around genes, and open new avenues for analysis of complex disease phenotypes.

Much better methods for measuring and monitoring ambient and individual exposure levels and modeling transport and fate are needed to enhance the environmental side of the equation. Moreover, a concerted effort is required to link monitoring data about air pollution, water pollution, and radon, pesticide, and other chemical exposures with genotyping studies. The Center for Disease Control and Prevention National Health and Nutrition Examination Study (NHANES) collects lots of exposure and health status data, and assays numerous metabolites and chemicals in thousands of individuals; genotyping those selected for appropriate ancillary studies could be productive. The EPA and state health and environmental protection agencies collect extensive monitoring data. Multiple large cohort studies have been launched (US Children's Health Study; UK BioBank Initiative) or proposed (Collins, 2004).

Several approaches must be integrated to increase understanding and applications of eco-genetics: comprehensive genomic analyses, multimodality longitudinal characterization of environmental stimuli and stressors, statistical and computational modeling of interactions, judicious use of well-designed animal models, and linkage of datasets for genomic, environmental, and behavioral variables important to specific diseases.

## REFERENCES


Andersen, J.B., Factor, V.M., Marquardt, J.U., et al., 2010. An integrated genomic and epigenomic approach predicts therapeutic response to zebularine in human liver cancer. Sci Transl Med (2) 54ra77.

Auman, J.T., Chou, J., Gerrish, K., et al., 2007. Identification of genes implicated in methapyrilene-induced hepatotoxicity by comparing differential gene expression in target and nontarget tissue. Environ Health Perspect 115, 572–578.

Belinsky, S.A., Liechty, K.C., Gentry, F.D., et al., 2006. Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high-risk cohort. Cancer Res 66, 3338–3344.

Brewer, G.J., 1971. Annotation: Human ecology, an expanding role for the human geneticist. Am J Hum Genet 23, 92–94.

Bushel, P.R., Heinloth, A.N., Li, J., et al., 2007. Blood gene expression signatures predict exposure levels. Proc Natl Acad Sci USA 104, 18,211–18,216.

Cantonwine, D., Hu, H., Tellez-Rojo, M.M., et al., 2010. HFE gene variants modify the association between maternal lead burden and infant birthweight: A prospective birth cohort study in Mexico city, Mexico. Environ Health 9, 43.

Cecil, J.E., Tavendale, R., Watt, P., Hetherington, M.M., Palmer, C.N.A., 2008. An obesity-associated FTO gene variant and increased energy intake in children. N Engl J Med 359, 2558–2566.

Collins, F.S., 2004. The case for a US prospective cohort study of genes and environment. Nature 429, 475–477.

Cordain, L., Eaton, S.B., Sebastian, A., et al., 2005. Origins and evolution of the Western diet: Health implications for the 21st century. Am J Clin Nutr 81, 341–354.

Cosio, B.G., Mann, B., Ito, K., et al., 2004. Histone acetylase and deacetylase activity in alveolar macrophages and blood mononocytes in asthma. Am J Respir Crit Care Med 170, 141–147.

Costa, L.G., Eaton, D.E. (Eds.), 2006. Gene–Environment Interactions: Fundamentals of Eco-Genetics, Wiley. New York.

Costa, L.G., Furlong, C.E., 2002. Paraoxonase (PON1) in Health and Disease: Basic and Clinical Aspects. Kluwer Academic Publishers, Boston.

Daston, G.P., Naciff, J.M., 2010. Predicting developmental toxicity through toxicogenomics. Birth Defects Res C: Embryo Today: Reviews 90, 110–117.

Davis, A.P., King, B.L., Mockus, S., et al., 2011. The comparative toxicogenomics database: Update 2011. Nucleic Acids Res 39, D1067–D1072.

Dean, M., Carrington, M., Winkler, C., et al., 1996. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia growth and development study, multicenter AIDS cohort study, Multicenter hemophilia cohort study, San Francisco city Cohort, ALIVE study. Science 273, 1856–1862.

Digel, W., Lubbert, M., 2005. DNA methylation disturbances as novel therapeutic target in lung cancer: Preclinical and clinical results. Crit Rev Oncol Hematol 55, 1–11.

Durbin, R.M., Abecasis, G.R., Altshuler, D.L., et al., 2010. 1000 Genomes Project Consortium: A map of human genome variation from population-scale sequencing. Nature 467, 1061–1073.

Ege, M.J., Herzum, I., Buchele, G., et al., 2008. Prenatal exposure to a farm environment modifies atopic sensitization at birth. J Allergy Clin Immunol 122, 407–412.

EPA, 2004. Potential Implications of Genomics for Regulatory and Risk Assessment Applications at EPA. Genomics Task Force Workgroup, EPA Science Policy Council, Washington, DC.

Garantziotis, S., Schwartz, D.A., 2010. Ecogenomics of respiratory diseases of public health significance. Annu Rev Public Health 31, 37–51.

Garcia-Closas, M., Thompson, W.D., Robins, J.M., 1998. Differential misclassification and the assessment of gene–environment interactions in case-control studies. Am J Epidemiol 147, 426–433.

Garrod, A.E., 1902. About alkaptonuria. Med Chir Trans 85, 69–78.

Gilliland, F.D., Li, Y.F., Saxon, A., Diaz-Sanchez, D., 2004. Effect of glutathione-S-transferase M1 and P1 genotypes on xenobiotic enhancement of allergic responses: Randomised, placebo-controlled crossover study. Lancet 363, 119–125.

Haberg, S.E., London, S.J., Stigum, H., Nafstad, P., Nystad, W., 2009. Folic acid supplements in pregnancy and early childhood respiratory health. Arch Dis Child 94, 180–184.

Halfmann, R., Lindquist, S., 2010. Epigenetics in the extreme: Prions and the inheritance of environmentally acquired traits. Science 330, 629–632.

Hansen, N.T., Brunak, S., Altman, R.B., 2009. Generating genome-scale candidate gene lists for pharmacogenomics. Clin Pharmacol Ther 86, 183–189.

Harrill, A.H., Ross, P.K., Gatti, D.M., Threadgill, D.W., Rusyn, I., 2009. Population-based discovery of toxicogenomics biomarkers for hepatotoxicity using a laboratory strain diversity panel. Toxicol Sci 110, 235–243.

Hollingsworth, J.W., Maruoka, S., Boon, K., et al., 2008. In utero supplementation with methyl donors enhances allergic airway disease in mice. J Clin Invest 118, 3462–3469.

IOM, Institute of Medicine Committee on Omics-Based Predictive tests, 2012. Evaluation of Traditional Omics, eds. Micahel, C.C., Nass, S., Omenn, G.S. National Academy Press, Washington, D.C.

Ito, K., Ito, M., Elliott, W.M., et al., 2005. Decreased histone deacetylase activity in chronic obstructive pulmonary disease. N Engl J Med 352, 1967–1976.

Kraft, P., Hunter, D., 2005. Integrating epidemiology and genetic association: The challenge of gene–environment interaction. Philos Trans R Soc Lond B 360, 1609–1616.

Kramer, C.B., Cullen, A.C., Faustman, E.M., 2006. Policy implications of genetic information on regulation under the Clean Air Act: The case of particulate matter and asthmatics. Environ Health Perspect 114, 313–319.

Le Marchand, L., Hankin, J.H., Wilkens, L.R., et al., 2001. Combined effects of well-done red meat, smoking, and rapid N-acetyltransferase 2 and CYP1A2 phenotypes in increasing colorectal cancer risk. Cancer Epidemiol Biomarkers Prev 10, 1259–1266.

McBride, C.M., Koehly, L.M., Sanderson, S.C., Kaphingst, K.A., 2010. The behavioral response to personalized genetic information: Will genetic risk profiles motivate individuals and families to choose more healthful behaviors? Annu Rev Public Health 31, 89–103.

McNicholl, J.M., Downer, M.V., Udhayakumar, V., Alper, C.A., Swerdlow, D.L., 2000. Host–pathogen interactions in emerging and re-emerging infectious diseases: A genomic perspective of tuberculosis, malaria, human immunodeficiency virus infection, hepatitis B, and cholera. Annu Rev Public Health 21, 15–46.

Olden, K., 2009. Human health and disease: Interaction between the genome and the environment. In: Willard, H.F., Ginsburg, G.S. (Eds.), Genomic and Personalized Medicine. Elsevier, Burlington.

Omenn, G.S., 2009. From human genome research to personalized health care. Issues Sci Technol Summer, 51–56.

Omenn, G.S., 2010a. Overview of the symposium on public health significance of genomics and eco-genetics. Annu Rev Public Health 31, 1–8.

Omenn, G.S., 2010b. Evolution in health and medicine Sackler colloquium: Evolution and public health. Proc Natl Acad Sci USA 107 (Suppl. 1), 1702–1709.

Omenn, G.S., Motulsky, A.G., 1978. Eco-genetics: Genetic variation in susceptibility to environmental agents. In: Cohen, B.H., Lilienfeld, A.M., Huang, P.C. (Eds.), Genetic Isssues in Public Health and Medicine. CC Thomas, Springfield, IL.

Ramos, R.G., Olden, K., 2008. Gene–environment interactions in the development of complex disease phenotypes. Int J Environ Res Public Health 5, 4–11.

Schwartz, D.A., 2010. Epigenetics and environmental lung disease. Proc Am Thorac Soc 7, 123–125.

Simopoulos, A.P., 2010. Nutrigenetics/nutrigenomics. Annu Rev Public Health 31, 53–68.

Smith, M.W., Dean, M., Carrington, M., et al., 1997. Contrasting genetic influence of CCR2 and CCR5 variants on HIV-1 infection and disease progression. Hemophilia growth and development study (HGDS), Multicenter aids cohort study (MACS), Multicenter hemophilia cohort study (MHCS), San Francisco city cohort (SFCC), ALIVE study. Science 277, 959–965.

Thomas, D., 2010a. Gene–environment-wide association studies: Emerging approaches. Nat Rev Genet 11, 259–272.

Thomas, D., 2010b. Methods for investigating gene–environment interactions in candidate pathway and genome-wide association studies. Annu Rev Public Health 31, 21–36.

Thomas, D.C., 2007. Viewpoint: Using gene–environment interactions to dissect the effects of complex mixtures. J Expo Sci Environ Epidemiol 17 (Suppl. 2), S71–S74.

Vesell, E.S., 1991. Genetic and environmental factors causing variation in drug response. Mutat Res 247, 241–257.

Vogel, F., Motulsky, A.G., 1997. Human evolution. In: Vogel, F., Motulsky, A.G. (Eds.), Human Genetics: Problems and Approaches, third ed. Springer-Verlag, Berlin.

Wang, F.T., Hu, H., Schwartz, J., et al., 2007. Modifying effects of the HFE polymorphisms on the association between lead burden and cognitive decline. Environ Health Perspect 115, 1210–1215.

Waterland, R.A., Jirtle, R.L., 2003. Transposable elements: Targets for early nutritional effects on epigenetic gene regulation. Mol Cell Biol 23, 5293–5300.

Waters, M., Stasiewicz, S., Alex Merrick, B., et al., 2008. CEBS – chemical effects in biological systems: A public data repository integrating study design and toxicity data with microarray and proteomics data. Nucleic Acids Res 36, D892–D900.

Weis, B.K., Balshaw, D., Barr, J.R., et al., 2005. Personalized exposure assessment: Promising approaches for human environmental health research. Environ Health Perspect 113, 840–848.

Wu, M.T., Kelsey, K., Schwartz, J., et al., 2003. A delta-aminolevulinic acid dehydratase (ALAD) polymorphism may modify the relationship of low-level lead exposure to uricemia and renal function: The normative aging study. Environ Health Perspect 111, 335–341.

Yang, Q., Khoury, M.J., 1997. Evolving methods in genetic epidemiology. III. Gene–environment interaction in epidemiologic research. Epidemiol Rev 19, 33–43.

## RECOMMENDED RESOURCES

http://www.genome.gov/gwastudies
http://www.ncbi.nih.gov
http://www.niehs.nih.gov
http://www.1000genomes.org
http://www.ccmb.med.umich.edu
http://www.gei.nih.gov
http://ctd.mdibl.org
http://www.niehs.nih.gov/research/resources/databases/cebs/
http://www.pharmgkb.org