Nutritional genomics: the next frontier in the postgenomic era

Jim Kaput and Raymond L. Rodriguez

Laboratory for High Performance Computing and Informatics, Section of Molecular and Cellular Biology, University of California at Davis, Davis, California 95616

Submitted 1 July 2003; accepted in final form 18 August 2003

Kaput, Jim, and Raymond L. Rodriguez. Nutritional genomics: the next frontier in the postgenomic era. Physiol Genomics 16: 166–177, 2004;10.1152/physiolgenomics.00107.2003.—The interface between the nutritional environment and cellular/genetic processes is being referred to as “nutrigenomics.” Nutrigenomics seeks to provide a molecular genetic understanding for how common dietary chemicals (i.e., nutrition) affect health by altering the expression and/or structure of an individual’s genetic makeup. The fundamental concepts of the field are that the progression from a healthy phenotype to a chronic disease phenotype must occur by changes in gene expression or by differences in activities of proteins and enzymes and that dietary chemicals directly or indirectly regulate the expression of genomic information. We present a conceptual basis and specific examples for this new branch of genomic research that focuses on the tenets of nutritional genomics: 1) common dietary chemicals act on the human genome, either directly or indirectly, to alter gene expression or structure; 2) under certain circumstances and in some individuals, diet can be a serious risk factor for a number of diseases; 3) some diet-regulated genes (and their normal, common variants) are likely to play a role in the onset, incidence, progression, and/or severity of chronic diseases; 4) the degree to which diet influences the balance between healthy and disease states may depend on an individual’s genetic makeup; and 5) dietary intervention based on knowledge of nutritional requirement, nutritional status, and genotype (i.e., “individualized nutrition”) can be used to prevent, mitigate, or cure chronic disease.

PROGRESS IN THE BATTLE against human disease and suffering is being accelerated by the availability of genomic information for humans, mice, and other organisms. The techniques and knowledge emerging from these genome projects have revolutionized the process of localizing and identifying genes involved in disease. To date, almost 1,000 human disease genes have been identified and partially characterized, 97% of which are now known to cause monogenic diseases (75). However, most cases of obesity, cardiovascular disease (CVD), diabetes, cancer, and other chronic diseases are due to complex interactions between several genes and environmental factors. It is not surprising, therefore, that the strategies for characterizing and identifying monogenic diseases have been unsuccessful when applied to chronic diseases. Despite the more than 600 association studies published as of 2002 (reviewed in Ref. 65), the molecular basis of chronic diseases remains elusive. Such results led to the development of the “common disease/common variant hypothesis” (i.e., CDCV hypothesis; Refs. 24 and 91), which states that chronic diseases are caused by sets of gene variants that collectively contribute to disease initiation and development. The complexity of genetic interactions and the number and spacing of mapping markers explain why it has been difficult for molecular epide-

Demological studies to localize genes associated with chronic diseases. More complete single-nucleotide polymorphism (SNP) and haplotype maps (31, 58, 76, 77, 173) will create additional resources for identifying genes involved in diseases. These genome-centric approaches, however, usually fail to take into account the most important variable in expression of genetic information and a major contributor to disease development, namely, dietary chemicals.

The interface between the nutritional environment and cellular/genetic processes is being referred to as nutritional genomics or “nutrigenomics.” Nutrigenomics seeks to provide a genetic understanding for how common dietary chemicals (i.e., nutrition) affects the balance between health and disease by altering the expression and/or structure of an individual’s genetic makeup. The conceptual basis for this new branch of genomic research can best be summarized with the following five tenets.

1) Common dietary chemicals act on the human genome, either directly or indirectly, to alter gene expression or structure.

2) Under certain circumstances and in some individuals, diet can be a serious risk factor for a number of diseases.

3) Some diet-regulated genes (and their normal, common variants) are likely to play a role in the onset, incidence, progression, and/or severity of chronic diseases.

4) The degree to which diet influences the balance between healthy and disease states may depend on an individual’s genetic makeup.

5) Dietary intervention based on knowledge of nutritional requirement, nutritional status, and genotype (i.e., “individual-
Dietary chemicals can affect gene expression directly or indirectly. At the cellular level, nutrients may: 1) act as ligands for transcription factor receptors (32, 70); 2) be metabolized by primary or secondary metabolic pathways, thereby altering concentrations of substrates or intermediates; or 3) positively or negatively affect signal pathways (22, 42). This is shown schematically in Fig. 1. Fatty acids, for example, are metabolized via the β-oxidation pathways to produce cellular energy (Fig. 1B). Altering intracellular energy balance may indirectly alter gene expression through changes in cellular NAD homeostasis (reviewed in Ref. 97). NAD reoxidation is associated with mitochondrial electron transport activity and is a cofactor for proteins involved in chromatin remodeling (59, 104). Chromatin remodeling processes have short- and long-term consequences for gene regulation due to reactions such as histone acetylation or DNA methylation that alter access to, and therefore regulation of, eukaryotic genes (reviewed in Ref. 43).

Some dietary chemicals also are ligands for nuclear receptors (Fig. 1A). Many, but not all genes involved in fatty acid metabolism are regulated by one of the three members of the peroxisome proliferator-activated receptor family (PPARα, PPARδ, PPARγ) family (reviewed in Ref. 9). The surprising finding (at the time) was that the fatty acids, palmitic (16:0), oleic (18:1 n9), linoleic (18:2 n6), and arachidonic (20:4 n6) acid (41, 63, 133), and the eicosanoids, 15-deoxy-Δ12,14-prostaglandin J2 and 8-(S)-hydroxyeicosatraenoic acid (54, 80), are ligands for PPARs (reviewed in Ref. 81). That is, these nuclear receptors act as sensors for fatty acids. Lipid sensors usually heterodimerize with retinoid X receptor (RXR), whose ligand is derived from another dietary chemical, retinol (vitamin A) (32). Some dietary chemicals, such as genistein, vitamin A, and hyperforin, bind directly to nuclear receptors and influence gene expression (Table 2). Other transcription factors are

Table 1. Composition of corn oil

<table>
<thead>
<tr>
<th>Fatty Acids (≈96 g/100 g corn oil)</th>
<th>Percent of Total FA, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>C16:0</td>
<td>10.9</td>
</tr>
<tr>
<td>C18:0</td>
<td>2.0</td>
</tr>
<tr>
<td>C18:1</td>
<td>24.9</td>
</tr>
<tr>
<td>C18:2</td>
<td>60.4</td>
</tr>
<tr>
<td>C18:3</td>
<td>0.9</td>
</tr>
<tr>
<td>C20:0</td>
<td>0.4</td>
</tr>
<tr>
<td>C20:1</td>
<td>0.2</td>
</tr>
<tr>
<td>C22:0</td>
<td>0.1</td>
</tr>
<tr>
<td>C24:0</td>
<td>0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sterols mg/100 g</th>
</tr>
</thead>
<tbody>
<tr>
<td>Campesterol</td>
</tr>
<tr>
<td>Stigmasterol</td>
</tr>
<tr>
<td>β-Sitosterol</td>
</tr>
<tr>
<td>Obtusifoliol</td>
</tr>
<tr>
<td>Unknown A</td>
</tr>
<tr>
<td>Cycloartenol</td>
</tr>
<tr>
<td>24-Methylene cycloartenol</td>
</tr>
<tr>
<td>Unknown B</td>
</tr>
<tr>
<td>Unknown C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fatty Acid Sterols mg/100 g</th>
</tr>
</thead>
<tbody>
<tr>
<td>Campesterol palmitate</td>
</tr>
<tr>
<td>β-Sitosteryl palmitate</td>
</tr>
<tr>
<td>Cycloartenyl palmitate</td>
</tr>
<tr>
<td>24-Methylene cycloartenol palmitate</td>
</tr>
<tr>
<td>Campesterol oleate</td>
</tr>
<tr>
<td>Campesterol linoleole</td>
</tr>
<tr>
<td>Stigmasterol oleate</td>
</tr>
<tr>
<td>β-Sitosteryl oleate</td>
</tr>
<tr>
<td>β-Sitosteryl linoleole</td>
</tr>
<tr>
<td>Cycloartenyl oleate</td>
</tr>
<tr>
<td>Cycloartenyl linoleole</td>
</tr>
<tr>
<td>24-Methylene cycloartenol linoleole</td>
</tr>
<tr>
<td>Other</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tocols ppm</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Tocopherol</td>
</tr>
<tr>
<td>β-Tocopherol</td>
</tr>
<tr>
<td>γ-Tocopherol</td>
</tr>
<tr>
<td>δ-Tocopherol</td>
</tr>
<tr>
<td>γ-Tocotrienol</td>
</tr>
</tbody>
</table>

Triglycerides 13 varieties

Analyses kindly performed by M. McClelland and L. Romanczyk, M&M Mars, Inc., using standard chemical analyses.

Common Dietary Chemicals Can Alter Gene Expression Or Structure

Epidemiological studies repeatedly show associations between food intake and the incidence and severity of chronic diseases (reviewed in Refs. 74 and 165), but the concept that food contains bioactive chemicals is not apparent from the design of many molecular and genetic association studies or laboratory animal or cell culture experiments. As an example of the complexity of a “simple” food, the constituents of corn oil are shown in Table 1. The variety and concentrations of fatty acids, triglycerides, sterols, sterol esters, and tocopherols are likely to have many and diverse effects on physiology since dietary chemicals have several fates upon entering a cell.
indirectly regulated by dietary chemicals. The sterol regulatory element binding proteins (SREBPs) are activated by protease cleavage, an event regulated by low levels of oxysterols and changes in insulin/glucose and polyunsaturated fatty acids (PUFA; reviewed in Ref. 44). The carbohydrate-responsive element-binding protein (ChREBP) is activated in response to high glucose levels and is regulated by reversible phosphorylation events (reviewed in Ref. 152).

Metabolic conversion of dietary chemicals also serves as a control mechanism for gene expression (109). The level of steroid hormones, which are ultimately derived from cholesterol, is regulated by the activities of the combined 10 steps in the steroid biosynthetic pathway. In addition, various intermediates branch into other metabolic pathways. Degradative pathways will also influence the overall intracellular concentrations of intermediates and end products (Fig. 1B, and Metabolism, below). Hence, the concentration of any given ligand (109) will be greatly influenced by specific combinations of alleles for the enzymatic steps in these assorted pathways. That a specific pair of alleles may be heterozygous and vary in frequency from one subpopulation to another is a fundamental precept of nutritional genomics.

Dietary chemicals also can directly affect signal transduction pathways (Fig. 1C). Green tea contains the polyphenol, 11-epigallocatechin-3-gallate (EGCG). EGCG inhibits tyrosine phosphorylation of Her-2/neu receptor and epidermal growth factor receptor, which, in turn, reduces signaling via the phosphatidylinositol 3-kinase (PI-3) → Akt kinase → NF-kB pathway (101, 119). Activation of the NF-kB pathway is associated with some virulent forms of breast cancer. Platelet-derived growth factor receptor phosphorylation is also inhibited by EGCG and derivatives (129). Grains such as rice contain inositol hexaphosphate (InsP6), which inhibits TPA- or EGF-induced cell transformation through its effects on PI-3 kinase (37). Resveratrol, phenethyl isothiocyanate (PEITC), genistein, and retinoids (vitamin A and metabolites) also affect signal transduction pathways (reviewed in Ref. 38).

The fact that dietary chemicals play such key roles in regulating gene expression beyond their well-known roles of producing energy and affecting insulin levels is consistent with evolutionary theory. Given that the human genome is so exquisitely responsive to its nutritional environment, it is reasonable to conclude that many human genes evolved in response to the plant- and animal-derived dietary chemicals we consume.

Diet Can Be a Risk Factor for Disease

The idea that adverse diet/genome interactions can cause disease is not new. The first example was the discovery of galactosemia by F. Goppart in 1917 (http://www.ncbi.nlm.nih.gov/htbin-post/Omim/disp?104000). Galactosemia is a rare recessive defect in galactose-1-phosphate uridylyltransferase (GALT). The lacks of GALT results in the accumulation of galactose in the blood, causing a number of health problems including mental retardation. Phenylketonuria (PKU), another recessive trait, was discovered in 1934 by Asbjørn Følling (http://www.ncbi.nlm.nih.gov/htbin-post/Omim/disp?108700). PKU is a defect in the enzyme phenylalanine hydroxylase that results in an accumulation of phenylalanine in the blood. The accumulation of high levels of phenylalanine can cause neurological damage. Both PKU and galac-
PKU and galactosemia are single gene traits and thus easy to identify and treat by changes in diet. Many chronic diseases are polygenic in nature and result from interactions of a subset of genes with environmental factors. The association between specific food intake and chronic disease was first recorded in 1908 and was based upon observations of Ignatovski that rabbits fed meat, milk, and eggs developed arterial lesions resembling atherosclerosis in humans (reviewed in Ref. 26). The associations of cholesterol with hypercholesterolemia and hypercholesterolemia with atherosclerosis focused much attention on the link between amount of calories (reviewed in Ref. 165) and/or the levels and types of vitamins (50), fat (e.g., 84; and reviewed in Refs. 84 and 167), and carbohydrates (reviewed in Ref. 73) with atherosclerosis, diabetes, obesity, cancer, and other chronic diseases. Although some associations have been confirmed by subsequent genetic or biochemical studies in laboratory animals or humans, others remain controversial. The much-debated association between level and type of dietary fat with breast cancer incidence (reviewed in Refs. 139 and 166) illustrates the difficulty of epidemiological studies to prove causation.

The limitations of epidemiological studies may be in design, dietary assessment tools, measurement errors, statistical methods, sample size, and nutritionally insignificant differences in fat intake among study populations (96). In addition, although epidemiological methods often include family histories, individual genotypes are not usually analyzed. Nevertheless, the underlying assumption of nonmolecular epidemiology studies is that individual genetic variation is insignificant and that each individual responds similarly to their environment. As demonstrated by the human genome (92, 156) and SNP projects (58, 93, 128, 134), there are potentially millions of base pair differences between individuals, and some of these differences could affect the way one individual responds to their nutritional environment relative to another individual. Assumptions, therefore, that fail to take into account these genetic differences are unlikely to reveal meaningful connections between specific nutrients and chronic diseases. With the advent of genomic sequence information and high-throughput technologies and reagents, analyzing SNPs or other polymorphisms in multiple genes is now possible. Analyzing patterns of SNPs with patterns of disease subphenotypes will require sophisticated statistical tools and large populations or many family studies.

A related confounding factor is the differences in allele frequencies among human subpopulations. Following the migrations from Africa, humans became geographically isolated, limiting genetic exchanges and fixing distinct alleles and haplotypes (148). As one example, the gene encoding arylamine N-acetyltransferase, NAT2 (64, 125), is polymorphic. Variants encode a fast acetylator allele and several subtypes of slow acetylators. These allele types are represented differently among populations in different geographic regions: slow allele subtypes are found in 72% of the Caucasians in the United States but only in 31% of Japanese. Individuals with slow acetylator NAT2 allele are more susceptible to bladder cancer when exposed to procarcinogens (125). The allele frequencies of NAT2 illustrate the importance of knowing allele distributions in populations and exposure to environmental influences, since molecular epidemiology studies rely upon statistical associations of certain haplotypes or SNPs (alleles) with disease phenotypes, incidence, and/or severity. False-positive associations of SNPs (or haplotype) and disease may be found because of the distribution of alleles within the population studied rather than a true association between an allele and a phenotype or disease. Determining allele frequencies requires resequencing genes in ethnically different individuals (greater than or equal to ~90), a costly enterprise that will ultimately be done for all genes. In the meantime, ethnic difference markers (25, 103, 114, 135) can determine the origin of chromosomal regions, which may encode ethnic-specific alleles and may be used to assess population substructures in epidemiological studies. Although population substructures confound statistical association studies, analyzing diverse admixtures may allow for the identification of genes contributing to certain specific chronic diseases (25, 103, 114, 135, 148), since certain ethnic populations are at increased risk for chronic diseases.

Analyzing genotype as a variable in association studies of disease phenotypes or subphenotypes will eliminate confounding due to population heterogeneity (e.g., 96, 116, 117, 134, 167). Monitoring or measuring dietary intakes must also be done, since epidemiological and laboratory animal studies have consistently demonstrated an effect of diet on disease initiation and progression. The literature linking diet to disease is too voluminous to review herein but certain recent advances are summarized below.

**Micronutrients.** Approximately 40 micronutrients are required in the human diet. Suboptimal intakes of specific micronutrients have been associated with CVD (B vitamins, vitamin E, carotenoids), cancer (folate, carotenoids), neural tube defects (folate), and bone mass (vitamin D) (50). B6, B12, and folate deficiencies, for example, are associated with increased serum homocysteine levels. Hyperhomocysteinemia is a risk factor and marker for coronary artery disease, but the mechanism(s) is not understood at the molecular level (51) although several theories have been proposed to explain its action (e.g., 132). Many of these conclusions are based upon cohort, randomized trials, and meta analyses wherein the cause of the disease cannot be conclusively determined.

Deficiency of vitamins B12, folic acid, B6, niacin, C, or E, or iron or zinc appears to mimic radiation in damaging DNA by causing single- and double-strand breaks, oxidative lesions, or both (5) (Table 3). Nutrient deficiencies are orders of magnitude more important than radiation because of constancy of exposure to milieu promoting DNA damage (4, 5, 7). Folate deficiency breaks chromosomes due to substantial incorporation of uracil in human DNA (4 million uracil/cell) (14). Single-strand breaks in DNA are subsequently formed during base excision repair, with two nearby single-strand breaks on opposite DNA strands leading to chromosome fragmentation. Micronutrient deficiency may explain why the quarter of the US population that consume less than the recommended five portions a day of vegetables and fruits has approximately twice the rate for most types of cancer compared with the quarter with the highest intake (5). A number of other degenerative diseases of aging are also associated with low fruit and vegetable intake. Progress is also being made in determining specific mechanisms for the role of certain minerals (calcium, magnesium, manganese, copper, and selenium) and vitamins in.
heart disease from work done in humans and in cell culture systems (reviewed in Ref. 168).

Macronutrients: Fats. Unbalanced intake of any of the three major macronutrients, fat, carbohydrates, or protein, contributes to the initiation, development, progression, and/or severity of chronic diseases. Intake of saturated fatty acids (SFA) is correlated with increased levels of low-density lipoprotein (LDL) cholesterol, the principal target of intervention for coronary disease risk reduction (86). In addition to CVDs, SFA may contribute to obesity and diabetes (87) because these diseases are also characterized by dyslipidemias (35, 39, 71, 96). Unbalanced fat intake may contribute to obesity and diabetes (87) because these diseases are also characterized by dyslipidemias (35, 39, 71, 96). Laboratory animal studies in which colorectal (reviewed in Ref. 60), and breast (26, 96) cancers are used to examine the role of dietary fats in cancer risk. In case control studies, these associations were observed after taking into account other risk factors, such as smoking, alcohol consumption, and physical activity. However, it is not clear whether these associations are causal or due to confounding factors.

Macronutrients: Carbohydrates. Dietary guidelines continue to emphasize diets low in saturated and total fat for reducing the risk of obesity and its related comorbidities: diabetes and CVD (86). Spurred by the increase in obesity and diabetes while fat intake barely decreased (36.4 to 34.1% of total calories) between 1970 and 1990 (48), a new focus of a number of epidemiological studies is carbohydrates. Simple and complex carbohydrates are metabolized at different rates and therefore have differential effects on blood glucose concentrations. The glycemic index (GI) is a quantitative measure of foods based upon postprandial blood glucose response (72). GI is expressed as the percentage of the response to an equivalent carbohydrate portion of white bread or glucose (169). Glycemic load (GL), the product of the average dietary GI and total carbohydrate intake, is a measure of total insulin demand (131). For example, if glucose has a GI of 100, potatoes have a GI of 87 and tomatoes, 9. Refined simple sugars or some polysaccharides that are cleaved rapidly to glucose produce higher blood glucose levels and a greater demand for insulin (8). High GI would increase insulin production and, at the same time, decrease synthesis of insulin receptors.

All but one cohort study of type 2 diabetes and GI (8) showed an association between GI and type 2 diabetes or coronary artery disease and colon or breast cancer and GI in case control studies. These associations were observed after multiple adjustments (e.g., for fiber or other dietary variables) and usually for the 5th quintile of GI (8). Molecular epidemiological analyses of the associations of candidate genes and nutritional variables are needed to determine the biochemical effects of GI on physiology.

Macronutrients: Protein. Analyzing the effect of protein intake on health is difficult because fat and macronutrients are significant and variable components of meat. Broiling, frying, and baking differentially alters the chemical composition of meat and other foods (1) and in some cases creates nitrosamines and other carcinogens (e.g., 130). Different ways of preparing foods may produce different amounts and types of natural or heat-generated dietary chemicals, thereby introducing confounding into epidemiological analyses. With these caveats, meat consumption appears to be associated with increased chronic disease risk (reviewed in Refs. 3 and 82) including bowel (e.g., 12) and colorectal (e.g., 57) cancers and type 2 diabetes (e.g., 153). Molecular epidemiology suggests that certain genes, for example, epoxide hydrolase (151), glutathione-S-transferase (28), and other detoxifying enzymes (130), may modify the effect of meat on disease risk.

Increased metabolism of protein also will increase the production of urea, with the corresponding increase in membrane-permeable ammonia (NH₃) and its ionized form NH₄⁺. Ammonia released in the alimentary tract of animals by microbial enzymes can disrupt metabolic pathways (158), alter the gastrointestinal mucosa (53), inhibit rates of growth in animals (e.g., 160), alter brain function (52), and promote cancer (e.g., 23).

Caloric restriction. Early epidemiological studies neglected to account for the differences in energy content between carbohydrates and proteins (each at ~4 kcal/g) and lipids (~9 kcal/g). Virtually all association studies show an increased risk for common diseases with increased energy intake (78). Laboratory animal studies have consistently shown that reducing caloric intake is the most effective means to reduce the incidence and severity of chronic diseases, retard the effects of aging, and increase genetic fidelity (reviewed in Refs. 149 and 163). Experiments in Saccharomyces cerevisiae suggest that caloric restriction may produce its largest effects by increasing respiration with the concomitant increase in membrane-permeable ammonia (NH₃) and its ionized form NH₄⁺.
tor for Sir2, a histone deacetylase involved in chromatin silencing of nucleolar rDNA, telomere, and mating type locus (59, 104). In mammals, other cellular targets, such as uncoupling proteins, and neuroendocrine peptides (e.g., leptin) of the central nervous system (CNS), are potential targets of regulation by caloric restriction.

Summary. The hunt for a single macronutrient or micronutrient that will prevent chronic diseases is destined to fail. It is more likely that dietary imbalances, from micronutrient deficiencies to overconsumption of macronutrients or dietary supplements, are the modifiers of metabolism and potentiators of chronic disease. Although the complexity of food and genotypic variations appears daunting, molecular and genetic technologies may provide the means for identifying causative genes (or their variants) and the nutrients that regulate them.

Some Diet-Regulated Genes Can Play a Role in Chronic Diseases

The progression from a healthy phenotype to a chronic disease phenotype must occur by changes in gene expression or by differences in activities of proteins and enzymes. Since dietary chemicals are regularly ingested and participate indirectly and directly in regulating gene expression, it follows that a subset of genes regulated by diet must be involved in disease initiation, progression, and severity (79, 113). The clearest example of genotype-diet interactions in chronic disease is type 2 diabetes, a condition that frequently occurs in sedentary, obese individuals and certain minority groups (13, 15). Once diagnosed with type 2 diabetes, some individuals can control symptoms by increasing physical activities and by reducing caloric (and specific fat) intake (108), i.e., expression of genomic information is changed by changing environmental (i.e., dietary) variables. Other individuals are refractory to such environmental interventions and require drug treatments. Many chronic diseases do not show the phenotypic plasticity seen in some type 2 diabetics; that is, symptoms are not reversible after some initiating event. Chromatin remodeling and changes in DNA methylation induced by unbalanced diets are possible mechanisms that contribute to irreversible gene expression changes. Nevertheless, genotype \times diet interactions contribute to the incidence and severity of obesity, atherosclerosis, many cancers, asthma, and other chronic conditions (106, 124, 145, 167).

Molecular approach. One approach to understanding the molecular mechanisms whereby diet alters health is to identify diet-regulated genes that cause or contribute to disease process. This can be done by examining the expression of a candidate gene or groups of genes (e.g., 142) in response to diets, an approach pioneered by Goodridge and coworkers (105; reviewed in Ref. 62). Many laboratories characterize the expression of candidate genes in a variety of tissues in laboratory animals in response to dietary variables (49, 61; reviewed in Refs. 21, 29, 33) and caloric restriction (17, 94, 95, 121). DNA and oligo-array technologies have extended this approach to multiple genes within a pathway (36) or all genes on an array (17, 94, 162; reviewed in Refs. 66, 141, 145, 155). Changes in gene expression are then associated with phenotype and can be explained by genetic variants in nuclear receptors, \textit{cis}-acting elements in promoters, or differences in metabolism that produce altered concentrations of transcriptional ligands.

The limitations of assessing regulation of individual or multiple genes by diet are 1) determining cause from effect for each gene; that is, what are the subset of causative genes for a given phenotype?; and 2) gene expression patterns in one strain (or genotype) may be unique to that genotype. The results from inbred mouse strains (Kaput J, Klein KG, Reyes EL, Vishek WJ, Kibbe WA, Jovanovic B, Cooney CA, and Wolf G, unpublished observations) would suggest that individual humans (170) may have unique patterns of gene expression depending upon their genotype and diet. Such individual qualitative and quantitative differences will complicate attempts to find patterns in gene expression results for dietary intake. Since diet recall is imprecise and controlling diets difficult in large population studies, identifying these complex interactions will be challenging.

A separate confounding influence on analyses of diet-induced changes in gene expression patterns is the health of the subject (laboratory animal or human). The presence of a disease can be considered an additional environmental influence that could affect gene expression patterns. For example, the presence of obesity unmasks additional type 2 diabetes loci in C57BL/6 and BTBR mice (143). Specifically, phenotypic expression of two interacting loci that affected fasting glucose and insulin levels was observed only in obese mice, and the alleles from the two parental strains (C57BL/6 and BTBR) had different effects on the diabetic subphenotypes. Hence, one would predict changes in gene expression based upon the presence or absence of disease processes and changes caused by dietary differences. Separating these variables will be an important component of future experimental designs for determining the effect of diet on susceptibility and disease progression.

Inbred strains of laboratory animals have proven useful for examining diet-disease interactions at the molecular level because 1) each individual member of a strain is genetically identical; 2) their environment can be rigorously controlled; 3) statistics can be applied to molecular, physiological, and genetic measurements; and 4) experiments can be repeated (e.g., 56, 98). The limitation of a unique genotype can be overcome by examining multiple strains of mice (159). One of our laboratories (79, 113) introduced a comparative method for laboratory animals that identifies genes regulated differently by dietary variables between two or more genotypes. The genotypes (or inbred strains) of mice are selected based upon their susceptibility to disease caused by diet. We found that certain genes were differentially regulated based upon genotype (in this case, A\textsuperscript{vy}/A obese yellow vs. A/a agouti mice) and/or on caloric intake (100% vs. 70% calories) or by the interaction between diet and genotype (Kaput et al., unpublished observations). The criteria for identifying a candidate disease gene are 1) genes must be differentially regulated by diet and/or 2) differentially regulated by genotype and 3) must map to chromosomal regions [e.g., quantitative trait loci (QTL)] associated with the disease (79, 113). This approach identifies candidate genes in an unbiased manner, and additional testing in humans or animal models is necessary to validate these.

Genetic approach. Strategies for identifying genes that cause chronic diseases in humans have been greatly influenced by the successes in identifying genes that cause monogenic diseases (75). The difficulty in identifying chronic disease genes (65) has been attributed to factors such as small sample size, poorly matched control groups, population stratification,
and overinterpreting data (among others, see: 18, 90, 127, 147). These methods and approaches are being improved to eliminate such errors and to reliably identify genes involved in chronic diseases (25, 34, 103, 114, 118, 122).

However, noticeably missing from discussions of the limitations of these genetic mapping techniques and gene association studies are the effects of environmental variables such as diet. Tanksley and coworkers (115) observed the importance of environment on expression of phenotypic traits in F2 and F3 generation tomato plants grown in Davis or Gilroy, CA, compared with plants grown near Rehovot, Israel. Only 4 of a total of 29 QTL were found at all three sites, and 10 QTL were found in only two sites. QTL identify multiple regions within all chromosomes that collectively contribute to a complex phenotype (126). Since an individual QTL encode many genes, identifying the causative genes within the QTL is challenging (147). So far, the QTL mapping that accounts for differences in diet has not been done, largely because controlling for diet in large population association studies is often not possible.

The complexities of gene-environment interactions are compounded by the same factors that affect molecular (gene expression) analyses: epigenetic interactions between genes (the obesity example in mice, Ref. 143), in utero effects, diet-gene interactions, and the “environmental history”; that is, the life-long exposure to changing diets may alter expression of genetic information later in life (136). Maternal nutrition during pregnancy also has been linked to altered phenotypes in laboratory and farm animals (e.g., 27, 30, 69, 164), and such epigenetic phenomena will likely affect gene-disease association studies. Maternal exposure to different nutrients and an organism’s exposure during its lifetime to changing diets is likely to also produce different health outcomes because the exposure to food and xenobiotics may act upon the genome to alter gene expression. An excellent discussion of the complexity of genotype × environmental history is presented by Sing and colleagues (136).

The Balance Between Health and Disease States May Depend on an Individual’s Genetic Makeup

All humans are 99.9% identical at the gene sequence level. The 0.1% variations in sequence, however, produce the differences in phenotypes (hair and skin color, height, weight, etc.) and an individual’s susceptibility to disease or health. Alterations in phenotype result from differences in gene expression or altered macromolecular activities.

1) A striking and simple example of how SNPs alter gene expression is a polymorphism that alters tolerance to dietary lactose (milk). Adult mammals are typically lactose intolerant. A mutation occurred ~9,000 years ago in Northern Europeans that allowed expression of the lactase-phlorizin hydrolase gene (LCH locus) to continue into adulthood. Although there are 11 polymorphisms in this gene clustered into 4 (A, B, C, U) prevalent haplotypes (>0.05%), a C13910T SNP located 14 kb upstream of the LCH (47) is highly associated with lactase persistence (lactose tolerance). This polymorphism is thought to alter regulatory protein-DNA interactions controlling expression of the gene (67). The A haplotype conferring lactose tolerance has an 86% frequency in the Northern European population, but only 36% in Southern European populations. The persistence of this variant in populations may confer selective advantages that include improved nutrition, prevention of dehydration, and improved calcium absorption. Regulatory SNPs (rSNPs) in other promoters are likely to play a role in regulating gene expression (e.g., 11, 16, 89).

2) SNPs may also alter splicing. Two insulin receptor splice variants differ in the presence (type B) or absence (type A) of exon 11. The type A isoform is associated with hyperinsulinemia (68, 144). Over 30,000 alternative splice sites have been identified in a genome-wide analysis of humans (93).

3) A subset of coding SNPs (cSNPs) will alter biochemical activities of enzymes, proteins, and cellular processes. As one example, steroid 5α-reductase (designated SRD5A2), a key enzyme in androgen metabolism in the prostate, has 13 naturally occurring variants in the human population. Nine of these variants reduce SRD5A2 activity by 20% or more, and three increase activity by more than 15% (99). Since SRD5A produces dihydrotestosterone (DHT) and DHT regulates genes in the prostate, variants in steroid 5α-reductase may affect incidence or severity of prostate cancer (123). Although these studies focused on coding SNPs, polymorphisms in 5' or 3' regulatory regions or splice sites may also alter the level of expression of a gene or the variant produced (e.g., 88, 170).

Dietary Intervention Based on Individualized Nutrition

Dietary intervention based on knowledge of nutritional requirements, nutritional status, and genotype (i.e., “individualized nutrition”) can be used to prevent, mitigate, or cure chronic disease. This assertion is obvious for nutritional deficiencies such as scurvy and beriberi or the potential harm from dietary phenylalanine for phenylketonurics. Less obvious are treatments for ~50 genetic diseases in humans caused by variants in enzymes (6). As many as one-third of enzyme variants are due to increased Km for a coenzyme, resulting in a lower rate of reaction (6). The Michaelis-Menten constant, Km, is a measure of binding affinity of an enzyme for its ligand (substrate or coenzyme) and is defined as the concentration of ligand required to fill half of the ligand-binding sites. Ames et al. (6) proposed the “Km hypothesis” to describe the effects of polymorphisms on enzymatic activity. Intracellular concentrations of coenzyme may be increased by high doses of the corresponding vitamin, which would partially restore enzymatic activity and potentially ameliorate the phenotype. Changing substrate concentrations may be a general approach to circumvent decreased coenzyme binding or decreased enzymatic activities caused by a given cSNP. Some examples include the following (6).

1) Glucose-6-phosphate dehydrogenase A44G (DNA: C131G) with NADP as cofactor. Defects in this gene are involved in favism and hemolytic anemia. Increased dietary intake of nicotinic acid or nicotinamide might increase NADPH coenzyme concentrations enough to alter the equilibrium of GDPH ↔ GPDH + NADPH.

2) NAD is a cofactor for aldehyde dehydrogenase (ALDH), an enzyme involved in alcohol intolerance and linked to Alzheimer’s and cancer. A cSNP causes E487K, which increases the Km 150-fold. This variant could not be treated with diet because the NAD substrate concentration could not be increased sufficiently to overcome the increased Km.

Ames and coworkers (46) have established a web site entitled “Km Mutants” (http://www.kmmutants.org/), which
summarizes nutritional information for a large number of enzymes requiring coenzymes.

Directed dietary intervention for prevention or treatment of chronic diseases in an individual is inherently more challenging because multiple genes interacting with each other and with environmental variables contribute to disease etiology. Identifying genes that contribute to the most to chronic disease initiation or progression and understanding their regulation by dietary variables is a likely first step in this process. A number of candidate gene-disease-diet association studies (reviewed in Refs. 100, 111, 157, 171) have shown the promise of this approach, as follows.

Hypertension. The amount of circulating angiotensinogen (ANG) is associated with increased blood pressure. A SNP, designated AA, at nucleotide position −6 of the ANG gene is linked with the level of circulating ANG protein. Individuals with the AA genotype who eat the Dietary Approaches to Stop Hypertension (DASH) diet show reduced blood pressure, but the same diet was less effective in reducing blood pressure in individuals with a GG genotype. A large percentage (~60%) of African Americans have the AA variant, and the remainder are heterozygotic (AG) at this position (146).

Cardiovascular health. Apo-A1 plays a central role in lipid metabolism and coronary heart disease. The G-to-A transition in the promoter of APOA1 gene is associated with increased HDL cholesterol concentrations. The A allele (or variant) was associated with decreased serum HDL levels (112). For example, women who eat more PUFA relative to saturated fats (SF) and monounsaturated fats (MUFA) have increased serum HDL levels. This type-of-fat effect is significant in men when alcohol consumption and tobacco smoking were considered in the analyses.

Individuals with small, dense LDL particles (phenotype B) have an increased risk of coronary artery disease relative to those individuals exhibiting large, less dense LDL particles (phenotype A) (reviewed in Ref. 86). In a classic crossover experiment, Krauss and coworkers (40) showed that the LDL patterns are influenced by low-fat diets. Thirty-eight men exhibiting phenotype A LDL were switched from a 32% fat diet to a diet containing 10% fat. Twelve of these 38 exhibited phenotype B LDL after 10 days on the low-fat diet (40), suggesting that for these 12, low-fat diets were not beneficial. Although not directly analyzed, these results suggest three distinct genotypes. Two genotypes produce either the A or B phenotype. A third genotype produces the A phenotype when these individuals eat a diet containing 32% fat, but a B phenotype when fed 10% fat, a result that can be explained by a genotype × environment interactions.

Cancer. Methylenetetrahydrofolate reductase (MTHFR) is a key gene in one-carbon metabolism and, indirectly, in all methylation reactions. Several laboratories have noted that the C667T polymorphism (Ala to Val), which reduces enzymatic activity, is inversely associated with occurrence of colorectal cancer (e.g., 20, 138, 150) and acute lymphocyte leukemia (137). Low intake of folate, vitamin B12, vitamin B6, or methionine was associated with increased risk for cancer among those with the MTHFR TT genotype. MTHFR variants are also implicated in CVD (83).

The effects of dietary chemicals on these polymorphisms raise the possibility that candidate gene or SNP association studies may be more accurate if diets and nutritional status are included as variables in the analyses. Dietary histories are notoriously inaccurate, and nutritional status is difficult to assess. Nevertheless these environmental factors are likely to alter results of association studies.

Summary

Recent advances in the field of pharmacogenomics underscore the importance of genotype × environment interactions by showing how individual genetic variation in human populations can affect a drug’s efficacy and the severity of undesirable side effects (102, 110). For this reason, pharmaceutical companies are incorporating genotyping as part of their clinical trials to predict drug safety, toxicity, and efficacy. By relating phenotype to genotype, drug companies are designing and developing better drugs with fewer adverse side effects. By identifying the nonresponding subpopulations, pharmacogenomics can also develop new drugs from compounds previously thought too toxic for human use.

The concept of “personalized” medicine is now being extended to the field of nutrition (85, 107, 161). It is now accepted that nutrients (i.e., macronutrients, micronutrients and antagonists) alter molecular processes such as DNA structure, gene expression, and metabolism, and these in turn may alter disease initiation, development, or progression. Individual genetic variation can influence how nutrients are assimilated, metabolized, stored, and excreted by the body. The same tools and methods used in pharmacogenomics (SNP analysis, gene expression profiling, proteomics, metabolomics, and bioinformatics) are being used to examine an individual’s response to his or her nutritional environment. In the near future, quick, low-cost, point-of-care tests will be available to assist patients and physicians to achieve, manage, and prolong health through dietary intervention. The desired outcome of nutritiongenomics is the use of personalized diets to delay the onset of disease and optimize and maintain human health.

The studies and examples cited here and by others (107) provide a conceptual basis for the emerging field of nutritional genomics. Building on this foundation will be challenging and will likely focus on the following broadly defined questions.

1) What are the quantitative nutritional requirements to produce optimal metabolism, particularly for the macronutrients?

2) How can we optimize nutrient intake for each individual, given the genetic diversity and complexity of common dietary chemicals?

3) How can we link dietary chemicals to subtle, long-term regulation of metabolism?

4) How can we assess the changing nutritional needs of an individual from birth through death, given the available molecular and genomic technologies?

5) How do we ensure that nutritional genomic information is used in a socially responsible manner, particularly as it relates to health disparities in subpopulations, such as ethnic racial minorities, the poor, and the uninsured?

NOTE ADDED IN PROOF

The following articles address issues related to nutritional genomics:


**ACKNOWLEDGEMENTS**

We thank George Wolff, Willard Vishek, Steven Watkins, Ted Powers, and Su Ju Lin for helpful discussion and critical review of the manuscript. We also thank Willard Vishek, for providing information and references for the discussion on the role of protein in chronic diseases, and Steven Watkins, who contributed the questions that end the article.

**GRANTS**

This work was supported by National Center for Minority Health and Health Disparities Center of Excellence in Nutritional Genomics Grant MD-00222.

**DISCLOSURES**

J. Kaput is the Chief Scientific Officer of NutraGenomics, Inc., a new biotechnology company that is investigating the science of nutritional genomics.

**REFERENCES**


Physiol Genomics • VOL 16 • www.physiolgenomics.org


