Influence of human genetic variation on nutritional requirements

Patrick J Stover

ABSTRACT

Genetic variation is known to affect food tolerances among human subpopulations and may also influence dietary requirements, giving rise to the new field of nutritional genomics and raising the possibility of individualizing nutritional intake for optimal health and disease prevention on the basis of an individual’s genome. However, because gene-diet interactions are complex and poorly understood, the use of genomic knowledge to adjust population-based dietary recommendations is not without risk. Whereas current recommendations target most of the population to prevent nutritional deficiencies, inclusion of genomic criteria may indicate subpopulations that may incur differential benefit or risk from generalized recommendations and fortification policies. Current efforts to identify gene alleles that affect nutrient utilization have been enhanced by the identification of genetic variations that have expanded as a consequence of selection under extreme conditions. Identification of genetic variation that arose as a consequence of diet as a selective pressure helps to identify gene alleles that affect nutrient utilization. Understanding the molecular mechanisms underlying gene-nutrient interactions and their modification by genetic variation is expected to result in dietary recommendations and nutritional interventions that optimize individual health.

KEY WORDS Single-nucleotide polymorphisms, thiopurine S-methyltransferase, methylene tetrahydrofolate reductase, homocysteine, folate, candidate gene approach

INTRODUCTION

Nutritional requirements are not usually generalized to a population as a whole; rather, they are tailored to population subgroups, for example, the elderly or women who are pregnant or breastfeeding. To date, knowledge of human genetic variation has not contributed significantly to the identification of subpopulations that would benefit from individualized nutritional requirements, although there are examples of genetic variation influencing food intolerances (1, 2). The recent availability of genomic data and our increased understanding of the relations among genetic variations and diet permits a quantitative examination of the contribution of genetic variation to nutritional requirements.

The Human Genome Project is one of the key factors that enables the study of gene-environment interactions. It provides the list of the 25 000 to 30 000 genes in human DNA (3), in essence, a “parts list” for the proteins and molecules that constitute the processes and pathways that perform all cellular functions, whether they be conversion of food into energy or the generation of thoughts. Having this parts list, we can assemble these networks and understand how they function, how they interact, and, most importantly, how they are regulated. With this information in hand, molecular pathways can be manipulated for benefit by exposure to exogenous agents, among the most potent being dietary components and pharmaceutical agents.

The next phase of the Human Genome Project is focused on cataloging and classifying all the variation that exists within the human genome. Each human is unique and phenotypically distinct, not only in physical appearance but also in physiology and response to environmental stimuli. Single-nucleotide polymorphisms (SNPs) are a primary component of human genetic variation and constitute a molecular basis for phenotypic variation. SNPs are differences in the DNA blueprint and can be single-nucleotide base pair insertions, deletions, or substitutions of one base pair for another. There are an estimated 7 million SNPs in the human genome (4), but only a small percentage of these have a functional effect. The goal is to identify which SNPs influence cellular networks and thereby create a variation in phenotype. SNPs contribute to complex traits, including susceptibility to chronic diseases and drug efficacy, and this has resulted in the evolution of the field of pharmacogenomics, which is a simpler model for the field of nutrigenomics.

PHARMACOGENOMICS

Pharmaceutical products can elicit a wide range of effects. For some patients, these agents are effective and beneficial and have the desired effect, whereas in others they have no effect at all. In some instances, pharmaceuticals can cause unintended and even unanticipated harm. The field of pharmacogenomics aims to understand the relation between the genetic makeup and responses to a specific pharmaceutical product, with the goal of better matching the drug to the individual to achieve the intended outcome without incurring the risk of an adverse consequence or side effect.

The importance of pharmacogenomics is well illustrated by the link between genetic polymorphism of the enzyme thiopurine S-methyltransferase (TPMT) and toxicity associated with the use of thiopurine monotherapy. The human genome contains approximately 800000 SNPs, each of which is a single-nucleotide variation in the DNA sequence. The allele frequencies of these SNPs vary across different populations, and the impact of each SNP on gene function can be determined by measuring the effect of the SNP on the activity of the gene product. In the case of TPMT, SNPs that affect the enzyme activity can lead to an increased risk of toxicity when using thiopurine-based therapies, such as those used in the treatment of leukemia and Crohn’s disease.

1 From the Division of Nutritional Sciences, Cornell University, Ithaca NY.
3 Supported by the Division of Nutritional Sciences, Cornell University.
4 Reprints not available. Address correspondence to PJ Stover, Cornell University, Division of Nutritional Sciences, 315 Savage Hall, Ithaca, NY 14853. E-mail: pjs13@cornell.edu.
drug 6-thiopurine (5, 6). 6-Thiopurine is an immunosuppressant that is also used in the treatment of childhood acute lymphoblastic leukemia. On entering a cell, 6-thiopurine is converted into a nucleoside and interferes with nucleic acid synthesis. In most patients, TPMT, which is found in the liver and in red blood cells, inactivates 6-thiopurine, and the standard dose is set high enough to account for this effect (7). However, TPMT activity exhibits genetic polymorphism, and about 1 in 300 people inherit TPMT deficiency as an autosomal recessive trait (5). If treated with a standard dose of 6-thiopurine, TPMT-deficient patients develop severe, and in some cases, lethal myelosuppression; in these patients, a 10- to 15-fold lower dose is needed for successful treatment (5, 6). Assays are now available for the 3 polymorphisms that account for most mutant alleles, and this is one example where genetic testing has a real role to play in determining the appropriate use of a drug (7).

NUTRITIONAL GENOMICS

The field of nutritional genomics aims to identify the genetic variations that account for why some individuals react differently to dietary components, in much the same way the pharmacogeneticist aims to identify the polymorphisms that affect drug efficacy and safety. The impact of genetic variation on nutritional requirements is inevitably more subtle than that of pharmaceutical agents for 2 reasons. First, for each nutrient, there is a window of intake between the Recommended Dietary Allowance (RDA), which is defined as the dietary intake that is sufficient to meet the requirement of 97% of healthy individuals in a particular life stage and sex group, and the tolerable upper limit (UL), which is the highest nutrient intake that can be achieved without incurring risk of adverse health effects for most individuals in the general population (Figure 1; 8). If the data used to set the RDA and the UL are established with the use of a diverse population, these benchmarks should accommodate most of the genetic variation in existence. If a genetic variation were identified that required a different RDA or UL, these parameters could be adjusted to accommodate the requirements of these individuals. The concept of the generalized nutritional requirement becomes compromised only when the RDA of a nutrient for one individual encroaches on the UL of another.

Second, the impact of nutrition takes place over a lifetime, whereas that of pharmaceutical agents occurs over a short period of time. Therefore, any genetic variation that confers an atypical nutritional requirement will almost certainly be incompatible with life and early development and therefore not viable. For example, single-nucleotide polymorphisms that affect folate utilization are risk factors for miscarriage and are not in Hardy-Weinberg equilibrium in populations that otherwise are in equilibrium (9–13). In other words, a woman whose fetus carries 2 alleles that prevent sufficient utilization of a given nutrient is more likely to miscarry than a woman whose fetus carries the more common functional variants.

Nonetheless, several genes and alleles have been found to affect nutrient utilization. A well-studied polymorphism (Ala222Val) in the methylene tetrahydrofolate reductase (MTHFR) gene has been shown to alter folate metabolism quite severely (14, 15), so that risk is increased for neural tube defects (NTDs) and cardiovascular disease (CVD) but decreased for colon cancer (16–20). The biochemical disruptions and disease risk can be ameliorated with increased folate intake (21). To date, this may be the best example of a genetic variation that can influence an RDA and supports the concept that genetic variation modifies nutrient utilization and potentially dietary requirements. Other polymorphisms have also been found to alter homocysteine metabolism as well as folate uptake and transport (9, 22–24).

Polymorphisms for enzymes that utilize and metabolize vitamin B-12 have been associated with NTDs and the development of Down syndrome and colon cancer, and this suggests a potential to affect nutrient requirements (25–27). Vitamin D receptor polymorphisms have been associated with childhood and adult asthma (28). A common polymorphism in the HFE gene (Cys282Tyr) is associated with the iron storage disease hereditary hemochromatosis in Europeans, and this might affect the UL for iron intake (29–32). Polymorphisms in other genes also affect lipid pathways (33–37), alcohol metabolism (38), and lactose metabolism (2). Interestingly, many of these SNPs are associated with specific ethnic groups or geographic ancestral subpopulations and display genomic signatures for positive selection, indicating that although these SNPs present risk of adverse
outcomes today, they were likely beneficial in the ancestral environment where they first arose (39–41).

Interestingly, polymorphisms for the aldolase B enzyme, which metabolizes fructose, have been discovered (1). The gene for this enzyme is highly polymorphic, but until recently, many of these polymorphisms were considered silent. Only when fructose was added in high quantities to the food supply as a sweetener did these polymorphisms present as disease alleles (42). This is an example where a change in environment has challenged a normally silent allele to the degree that it begins to present as a disease allele.

IDENTIFICATION OF GENETIC POLYMORPHISMS AFFECTING DIETARY REQUIREMENTS

Genetic variations that affect dietary requirements may be identified by using 2 approaches: the candidate gene approach and evolutionary genomics. The candidate gene approach looks for associations between a particular pathologic condition and genetic polymorphisms that affect the metabolic pathways known to be associated with that condition. This approach has been successful in many cases, but it is problematic in nutrition for 3 reasons:

Most nutrition-related diseases are complex diseases; that is, they are both multigenic and modifiable by multiple environmental factors. The penetrance and prevalence of a polymorphism affect the sample size needed to identify associations between the allele of interest and the pathologic condition.

Many functional SNPs are not coding SNPs but instead fall in the promoter regions of genes. In fact, it has been estimated that most of the genetic variation that influences physiology one way or another is actually in the promoter regions of genes (43, 44). Two-thirds of polymorphisms identified in human promoter regions affect transcription rates by two-fold or greater (43). Promoter regions can span many thousands of base pairs and tend to be more heterogeneous than the coding regions of genes (43).

The candidate gene approach is limited by knowledge of the pathways and networks contributing to a particular pathologic condition: if all the players in the pathway or a network are not known, it will not be possible to identify all the genes that affect it.

Evolutionary genomics is based on the recent availability of genomic sequences from humans and nonhuman primates that enables the identification of sequence “signatures” for positive selection (39, 41, 44–46). The molecular basis for this approach is change in the DNA sequence, which in the absence of selection pressure takes place at a background rate of about $2.5 \times 10^{-8}$ mutations per nucleotide site (47, 48). However, regions of the sequence that are under selection pressure evolve more quickly, and the rate can be as much as 400 times as fast (49). A rapidly evolving gene is an adaptive sequence that enables an organism to match its environmental challenges, and these challenges can occur in distinct geographic regions. In the case of nutritional genomics, nutrients and dietary components act as environmental pressures, shaping today’s human genome though millennia of selection pressure. This highly selected genome in turn may affect nutrient requirements. When a genome that has adapted to a certain environment is put in a different context, adaptive genes may become disease alleles, which must then be managed or modified through nutritional intervention (50–52).

The aim of evolutionary genomics is to identify the genes that evolve rapidly, within and among species. Once these high-variable genes are identified, we can infer selective pressures on the basis of their cellular function. In fact, genes that rapidly evolve tend to cluster as immunity genes, metabolism genes, or genes involved in reproduction (41, 46).

What would be the signs that genes are subjected to positive selection? Variants would be expected to concentrate in specific geographic regions or ethnic groups where a common selective pressure exists and to be absent in other places or in other ethnic groups. We would see an excess of rare variants, large allele frequency differences among the population, and a common haplotype that remains intact over many generations (41, 45).

Many genes identified through the use of evolutionary genomics were previously identified by use of the candidate gene approach, for example, the lactase gene polymorphism (53). This variant arose in northern Europeans and in people inhabiting the arid regions of northern Africa, areas where there might have been a selective benefit in being able to drink milk and consume dairy products as a source of nutrition. The HFE Cys282Tyr polymorphism shows evidence of positive selection (54), as does the alcohol metabolism gene, alcohol dehydrogenase (55), and the Hbs allele in the β-globin gene (56). The A+ and MED alleles of glucose-6-phosphate dehydrogenase exhibit positive selection as well because of their conferred protection against malaria (3, 46, 53, 57).

MTHFR POLYMORPHISM AND NUTRITIONAL REQUIREMENTS FOR FOLATE

One example suggesting that genetic variation can affect nutritional requirements is the MTHFR Ala222Val polymorphism that affects folate metabolism. Folate is a B vitamin whose function is to carry and activate one-carbon units (58). The one-carbon units are carried on the N-5 or N-10 of tetrahydrofolate. One-carbon metabolism is required for the de novo synthesis of purine nucleotides and thymidylate and for the remethylation of homocysteine to methionine (Figure 2). Methionine can be adenylylated to form S-adenosylmethionine, a cofactor for numerous methylation reactions including those that affect gene regulation (58).

The MTHFR Ala222Val polymorphism is a C-to-T transition in the coding region, resulting in the conversion of an alanine to a valine in the protein (14). This has 2 effects on one-carbon metabolism: First, it impairs remethylation of homocysteine to methionine, which thus alters DNA methylation and gene expression (59, 60). Second, because the 2 pathways compete, it also increases the conversion of deoxyuridine monophosphate (dUMP) to deoxothymidine monophosphate (dTMP), which leads to more folate-dependent thymidine biosynthesis (61).

Folate deficiency increases the risk of several diseases and anomalies, including NTDs and CVD, and folate supplementation can ameliorate or completely negate the effects of this polymorphism. Therefore, the risk of NTDs is most apparent in individuals who are folate deficient (16–19).

The same polymorphism is actually protective against several cancers, most notably colon cancer (20). The Physician’s Health Study examined the associations of MTHFR mutation, plasma folate concentrations, and their interaction with risk of colon cancer in 202 colorectal cancer cases and 326 cancer-free controls matched by age and smoking status (20). The study showed...
that in men with normal folate concentrations (serum folate > 3 ng/mL) who were homozygous for the mutant allele, the risk of colon cancer was one-third that in carriers of the wild-type allele (odds ratio: 0.32; 95% CI: 0.15, 0.68). In men who were folate-deficient, however, this protection was absent. The same polymorphism and downstream effect on metabolism presents a risk of NTD but is simultaneously highly protective in colon cancer (62).

**EFFECT OF GENOMICS ON DEFINING DIETARY REQUIREMENTS**

Can we use genomics approaches to determine dietary requirements? Persons who are homozygous for the rarer MTHFR T allele (T/T) need a higher folate intake than do carriers of the C allele to lower their risk of folate-related pathologies. Although there is no report indicating that this polymorphism has been subject to positive selection, there is wide geographic variation in terms of the prevalence of the T/T genotype (Table 1: 63–68). This has the potential to influence public health policy in that fortification with folate may not be required in populations where the T/T genotype is rare. However, the current RDA actually may cover both MTHFR genotypes, and there is no definitive evidence to date that the current RDA for folate should be modified for persons who are homozygous for the MTHFR 677T allele (69).

On the other hand, the HFE gene Cys282Tyr polymorphism has already affected government policy, with 2 countries in northern Europe halting their iron fortification policies, in part because of a potential risk to persons at risk of hereditary hemochromatosis (70, 71). However, no definitive evidence exists suggesting that the UL for persons at risk of hereditary hemochromatosis encroaches on the RDA. For the time being, it appears there is no need to individualize iron intake recommendations according to genotype (72).

Genetics is also driving a reevaluation of how we define nutritional inadequacy. Currently, the RDA for a particular nutrient is defined by the development of deficiency diseases, but with advances in the field of genomics, there is a view that we should be more sophisticated and use biomarkers to define inadequacy and safe upper limits of intake. An example of this would be the epigenetic effects of nutrients. Epigenetics refers to the inheritance of traits that are not linked to DNA sequence, but rather to modifications of DNA, and among these modifications is DNA methylation, which affects gene expression.

In 2002 Cooney et al (73) illustrated how nutritional intake can affect epigenetics very dramatically. Those authors showed that in inbred mice, the folate intake of the mother during pregnancy affects the coat color of the offspring through an epigenetic effect on expression of the agouti gene. Expression of the agouti protein produces yellow-furred mice. However, methylation of the agouti gene promoter region during gestation blocks agouti expression and the offspring have darker fur (73). The more methylation of the promoter region that takes place, the less yellow fur there is on the mouse. Methylation patterns established during gestation remain metastable throughout the lifetime of the animal; thus, by affecting methylation levels during gestation, it is possible to manipulate permanently the coat color of the offspring.

Folate metabolism is critical for DNA methylation, and in the pups of mothers whose diet while pregnant included a high intake of folate, the promoter region was methylated and the agouti gene

---

**TABLE 1**

The prevalence of the MTHFR 667T/T genotype in various ethnic groups

<table>
<thead>
<tr>
<th>Population</th>
<th>Prevalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mexicans (64, 68)</td>
<td>27–35</td>
</tr>
<tr>
<td>Brazilians (68)</td>
<td>7</td>
</tr>
<tr>
<td>Sub-Saharan Africans (63, 68)</td>
<td>0–1.5</td>
</tr>
<tr>
<td>African Americans (66)</td>
<td>2</td>
</tr>
<tr>
<td>Yemenite Jews (65)</td>
<td>2</td>
</tr>
<tr>
<td>Muslim Arab Israelis (65)</td>
<td>16</td>
</tr>
<tr>
<td>Toscanians (Italy) (67)</td>
<td>30</td>
</tr>
<tr>
<td>Irish (64)</td>
<td>6</td>
</tr>
<tr>
<td>Dutch (68)</td>
<td>5</td>
</tr>
<tr>
<td>Europeans (68)</td>
<td>17</td>
</tr>
<tr>
<td>Japanese (64)</td>
<td>19</td>
</tr>
<tr>
<td>Pakistanis (68)</td>
<td>4</td>
</tr>
<tr>
<td>Chinese (68)</td>
<td>16</td>
</tr>
</tbody>
</table>

1 MTHFR, methylenetetrahydrofolate reductase.
was not well expressed. However, less DNA methylation took place during gestation in pups whose mothers were more folate deficient during pregnancy, which allowed the agouti to be expressed and the appearance of the characteristic yellow coat color (73, 74). This example illustrates how maternal nutrition can permanently affect how genes are expressed in the fetus and ultimately the offspring, with potentially lifelong consequences that may alter various health outcomes.

Another such biomarker is genomic stability. Studies on twins indicate that \( \approx 25\% \) of the variation in life span can be attributed to genetic differences (75). Therefore, environmental factors play a key role in determining longevity (73, 76, 77). As we age, mutations accumulate in our genomes (78) and epigenetics seems to become dysregulated (79). Several B vitamins, as well as some antioxidants, are known to affect the mutation rate for both chromosomal and mitochondrial DNA. When determining the RDA, perhaps we should consider what concentration of individual vitamins will lower the mutation rate sufficiently to prevent disease (77).

**RATIONAL DESIGN OF NUTRITIONAL REGIMENS TO PREVENT DISEASE**

Many SNPs confer both advantage and risk, depending on the health outcome of interest, as illustrated by the *MTHFR* allelic variants. Understanding the physiologic and biochemical consequences of specific gene variants and the mechanisms that confer disease protection or risk enables the rational design of nutritional approaches that can give maximal benefit to all individuals by precise manipulation of metabolic networks. Our laboratory seeks to understand the mechanism whereby impairments in homocysteine remethylation, as occurs in individuals who are homozygous for the *MTHFR 677T* allele, lowers risk of cancer, as described previously (20). The changes in metabolism conferred by this polymorphism confer protection against colon cancer while increasing the risk of NTDs. Folate supplementation can lower the risk of NTDs associated with the *T* allele, but how can we provide the benefit of the *T* allele to carriers of the common *C* allele in terms of colon cancer risk?

Colon cancer is an epithelial cancer, the risk of which increases with age (80). Folate deficiency is a risk factor for colon cancer (81), and patients with cancer exhibit folate deficiency due to increased folate turnover (82). DNA hypomethylation induced by folate deficiency affects the mammalian genome, both in terms of gene expression and mutation rate (80). One of the first biomarkers for the transformation of normal epithelium into a metastatic tumor in colon cancer is a methylation defect (83). The *MTHFR 677T* polymorphism may function to lower DNA mutation rates by increasing the efficiency of dTMP synthesis or may somehow lessen the penetration of the methylation defect that is characteristic of cellular transformation.

Metabolic alterations resulting from *MTHFR* genotypes can be recapitulated by altering the expression of other genes in the folate metabolic network, namely, cytoplasmic serine hydroxymethyltransferase (cSHMT). This enzyme is expressed in a tissue-specific manner and increases the flux of one-carbon units through the thymidylate pathway, thereby suppressing the homocysteine methylation pathway by sequestering 5-methyltetrahydrofolate (Figure 2; 84). This is similar to the effect of the mutant *MTHFR* allele on the homocysteine-methionine cycle.

As mentioned earlier, the gene for cSHMT is not expressed in all tissues. In the mouse embryo, it is expressed in the neural tube, hind brain, midbrain, craniofacial structures, and crypt cells of the colon, all of which are areas associated with folate-related pathologies. This pattern of expression is consistent with a signaling or patterning gene rather than a metabolic gene.

The expression and activity of the cSHMT gene is regulated robustly by several nutrients, including folate, zinc (51), and ferritin (85). Within the cell, ferritin sequesters iron and chelates it. It also degrades folate; cancer cells have lower concentrations of folate because they increase expression of ferritin (58). Ferritin also up-regulates cSHMT expression at the level of translation (85), which has the same metabolic effect as the *MTHFR* mutant allele. Ferritin expression is also increased during inflammation and it is nuclear factor-\( \kappa \)B sensitive.

Therefore, the cSHMT gene is affected robustly by several environmental stimuli, and alterations in cSHMT expression can mimic the metabolic states that result from all *MTHFR* allelic variants. Currently, we are using mice that lack or overexpress the cSHMT gene to investigate its effects on DNA stability and gene transcription and to determine the effects of altered cSHMT expression on susceptibility to NTDs and colon cancer. Thus far, we have established that cSHMT robustly regulates the expression of \( \approx 100 \) genes that cluster in pathways that are known to be associated with cancer (PJ Stover, unpublished observations, 2004). This information will be used to better understand how the *MTHFR* mutant allele reduces the risk of colon cancer and to develop rational strategies to reproduce this effect by manipulating the folate metabolic network through diet.

**EFFECT OF NUTRITIONAL INTAKE ON GENOMICS**

One cautionary note must be raised. Elevating dietary requirements may hide the phenotypic effects of a mutation—the concept of genetic rescue—thereby allowing it to be inherited and become established in a population. In transgenic mice bred with a dysfunctional *HOX* gene, for example, the effect of the genetic mutation on bone development—these mice are usually born with bones so fragile the ribcage cannot withstand the pressures of breathing—can be masked by high concentrations of folate in their mother’s diet (86).

A study in humans suggests that the frequency of a polymorphism can be affected by dietary intake: \( \approx 25 \) y ago, the Spanish government started recommending that women of child-bearing years take folate. In 2002 Reyes-Engle et al (10) tracked the frequency of the *MTHFR 677C*\( \rightarrow \)T polymorphism in Spanish men and women of differing ages and found that the allelic frequency was stable in those born >24 y ago. In those aged \( \leq 24 \) y, however, the prevalence of the mutant *T* allele, which is a risk factor for miscarriage and spontaneous abortion as well as birth defects (87–89), was higher than in the older populations.

This study suggests that folate not only prevents birth defects but may also rescue embryos that normally would not be viable. Although this study has not been replicated in other populations and has been criticized for errors such as population bias (90), the concept is still valid: we may be able to override certain genetic lesions by changing nutrient intake and thereby increase the frequency of the disease alleles within the population.
CONCLUSION

Genetic variation certainly has an important influence on human nutritional requirements, and the introduction of genomics has both highlighted the complexity of the interaction between genes and diet and offered opportunities to reevaluate the criteria used to determine RDAs and the contribution of genetic variation to optimal nutrition for individuals. As the interactions between genetic variation and nutritional requirements become more fully understood, it will allow dietary recommendations to be individualized according to genotype to ultimately reduce our risk of degenerative diseases and increase health and well-being in old age.

The author had no conflicts of interest to report.

REFERENCES

10. Reyes-Engel A, Munoz E, Gaitan MJ, et al. Implications on human genetic variation and nutritional requirements become more fully understood, it will allow dietary recommendations to be individualized according to genotype to ultimately reduce our risk of degenerative diseases and increase health and well-being in old age.


