April 22, 2003
Page One Editorial Services
DRAFT meeting summary

Summary of the Symposium on

Genetic Variation and Gene Environment Interaction in Human Health and Disease

April 16, 2003

Sponsored by
National Institute of Environmental Health Sciences (NIEHS)
National Human Genome Research Institute (NHGRI)
and
National Institute of Alcohol Abuse and Alcoholism (NIAAA)
of the
National Institutes of Health

Organized by
Samuel Wilson, NIEHS
Lisa Brooks, NHGRI
Faye Calhoun, NIAAA

In April 2003, the National Human Genome Research Institute (NHGRI) and the National Institutes of Health (NIH) celebrated the 50th anniversary of Watson and Crick's Nobel Prize-winning work on the double-helical structure of DNA and the completion of a final version of the sequence of the human genome. The celebration featured a NHGRI-sponsored ***Scientific Symposium: From Double Helix to Human Sequence – and Beyond***, and several satellite symposia on related topics.

On April 16th, the National Institute of Environmental Health Sciences (NIEHS), along with NHGRI and NIAAA hosted a half-day satellite symposium entitled ***Genetic Variation and Gene Environment Interaction in Human Health and Disease.*** This symposium provided an opportunity to focus on DNA variation, gene-environment interactions, and their implications for human health and disease. Researchers presenting at the symposium included Drs. Lynn Jorde, Deborah Nickerson, Jeffrey Trent, Charles Rotimi, Clement Furlong, David Crabb, Mary-Claire King and Martyn Smith.

This symposium focusing on gene-environment interactions also provided an opportune moment for the NIEHS to celebrate completion of the first phase of the Environmental Genome Project.
The Environmental Genome Project (EGP) was initiated by the National Institute of Environmental Health Sciences (NIEHS) in 1998 to improve understanding of human genetic susceptibility to environmental exposure. The EGP catalogues single nucleotide variation (also called SNPs) in environmentally-responsive genes in the human genome and carries out research on the biological implications of such genetic variation. The EGP is resequencing 554 environmentally-responsive genes in subgroups according to a phased timeline. The first phase of resequencing, now complete, extended from 1998 to 2003 and focused on polymorphic variants in approximately 200 DNA repair and cell cycle genes. The second phase of resequencing will include genes regulating metabolism, signal transduction and apoptosis. Symposium speakers highlighted work undertaken under the umbrella of the NIEHS Environmental Genome Project.

Dr. Samuel Wilson (Deputy Director, NIEHS) opened the Symposium and chaired the first Symposium session. Dr. Lisa Brooks (NHGRI) chaired the second session of the Symposium.

Session I

Patterns of Human Genetic Variation

Lynn Jorde, University of Utah

Dr. Lynn Jorde presented an overview of the patterns of human genetic variation. In opening his talk, Dr. Jorde answered the question "why we are interested in studying human genetic variation?" Dr. Jorde pointed out that knowledge of human genetic variation is important because it has applications in medicine and forensics, and because it can be informative about the evolutionary history of the human population.

Variation in the human genome can be considered at the level of the individual (*i.e.*, how does the genome of one person differ from that of another person) and at the level of human populations (*i.e.*, how do the genomes of all persons in the U.S. differ from the genomes of all persons in Africa?).

Dr. Jorde studied genetic variation in human populations using a global genomic sample from Africans, Europeans and Asians in a ratio of 150:120:700. The amount of variation in this sample was assessed by measuring many types of genetic variations including Alu insertions, tandem repeats (microsatellites), restriction fragment length polymorphism, Y chromosome variants and alterations in autosomal gene regions. On average, variants are found at the rate of approximately 1 in 1000 bp. Dr. Jorde observed that the amount of diversity is higher within a geographically defined population (*i.e.*, between individuals living on one continent) than between two geographically defined populations (*i.e.*, between Americans and Africans). However, different amounts of variation are observed between any two continental populations. In addition, the African population has the highest level of diversity, suggesting it is more ancient than other human populations. This is consistent with the idea that the human race evolved first on the African continent and then spread to other continents.

Dr. Jorde analyzed variation in the CYP1A2 gene, comparing its haplotype in African, European and Asian populations. A haplotype is a group DNA polymorphisms that tend to segregate together. The most common CYP1A2 haplotypes are found in all populations, but rare haplotypes are sometimes unique to one or two populations. This result is consist with the hypothesis that humans evolved in a single geographical location (*i.e.*, Africa) and then dispersed to other geographical locations. Dr. Jorde also infers from the data that the African population is the oldest human population.

Alu insertions were examined in up to 160 genetic loci in individuals of different geographic and ethnic origins. Dr. Jorde then asked whether the pattern of genomic variation in one individual is sufficient to determine the ethnic origin (or race) of that person. Although it may be possible to identify geographic origin based on DNA variation, Dr. Jorde concluded that it is not possible to reliably infer ethnic origin from these data, because human populations are not "pure."

The rate at which the human genome evolves is strongly influenced by selective pressure(s) in the environment. Dr. Jorde analyzed the effect of selective pressure on genetic variation in the CYP1A2 gene in 108 individuals from Africa, Europe and Asia. CYP1A2 plays a role in metabolism of xenobiotic compounds including caffeine, acetominophen and other drugs. Variation was catalogued and compared in both the coding and non-coding regions of CYP1A2. The coding region had several common variants and an excess of rare variants, while the non-coding region had primarily rare variants. This result suggests that different selective forces on the coding and non-coding portions of the gene may influence the distribution of variants in the two genomic regions. Dr. Jorde also described genetic variants associated with hemochromatosis that occur at high frequency in some European populations. Although these variants are now common in these populations, they evolved very recently. Because newly evolved neutral variants are introduced into a common haplotype

slowly, their frequency in the population increases very slowly. Therefore, the data indicate that the variants associated with hemochromatosis were subject to positive selective pressure. These two examples demonstrate that natural selection can be used as an indicator of gene-environment interactions.

SNPing in the Human Genome

Deborah Nickerson, University of Washington

The availability of the complete human genome sequence challenges us to try to understand how genotype is linked to phenotype. This relationship can be very complex, reflecting multiple genetic components and/or several environmental influences. Dr. Deborah Nickerson is carrying out systematic polymorphism discovery by resequencing candidate genes that may play a role in environmentally-induced disease. Nickerson described the results of this resequencing effort, including studies on associations between genetic variants, and insights into the structure-function relationships of several candidate genes.

There are many types of genetic variation in the human genome, but single nucleotide polymorphisms (SNPs) are the most common. SNPs are nucleotide positions that differ on two copies of the same human chromosome. A sequence variant is called a SNP if it occurs at frequency of 1% or higher. Simple heritable diseases can be studied with a small number of genetic markers in a small number of individuals; the variants that cause these diseases are rare but their penetrance is very high. In contrast, common complex diseases often are caused by multiple genes with low penetrance, some or all of which interact with environmental factors. These disease are difficult to study, requiring large numbers of subjects and a large number of genetic markers. SNPs, which occur approximately 1 per 1300 bp in the human genome, are useful markers for such association studies of common complex diseases. Approximately 3.7 of the predicted 11 million SNPs in the human genome have been identified to date. Most of these SNPs were identified by data mining, but data mining detects primarily high frequency SNPs (>5%), and produces a dataset that is depleted for low frequency SNPs (1-5%).

To link a specific SNP with disease susceptibility, it is necessary to determine if and how the SNP alters biological function. However, there are 5 to 7 million common SNPs in the human genome, and it is therefore not possible to determine the functional consequences of all human SNPs. As a feasible alternative method to identifying functionally important human SNPs, the NIEHS Environmental Genome Project (EGP) has adopted a candidate gene approach. The EGP is resequencing and cataloguing SNPs in the coding and upstream regulatory regions of well-characterized DNA repair and cell cycle genes. To date, more than 17,000 SNPs (including >1000 cSNPs in gene coding regions) have been identified in 214 candidate genes by resequencing 90 DNA samples from an ethnically diverse population sample (http://locus.umdnj.edu/nigms). The EGP SNP data are deposited into a publicly-accessible integrated database called GeneSNPs (http://www.genome.utah.edu/genesnps), which has a graphical user interface, extensive gene annotation, and a suite of bioinformatic tools for functional analysis. SNP data are also being analyzed for associations within the gene or within several genes in the same pathway.

Dr. Nickerson described how SNPs in the DNA repair gene 8-oxoguanine DNA glycosylase 1 (OGG1) were studied in the context of other database-accessible information on this enzyme. The GeneSNPs database provides access to the published genomic and cDNA sequences of the gene, a map of OGG1 protein functional domains, references in the literature characterizing variants in OGG1, information linking OGG1 to other DNA repair enzymes, and images of the OGG1 3-dimensional

protein structure. All of this information is used to evaluate the functional significance of novel SNPs identified in OGG1 by resequencing.

The functional significance of a polymorphic variant can be determined using a direct or indirect approach. In the direct approach, only SNPs in the coding region (cSNPs), which have the potential to directly alter protein function, are analyzed. In the indirect approach, coding and non-coding SNPs are analyzed for linkage disequilibrium. Dr. Nickerson pointed out that SNPs in both the coding and non-coding regions provide important information about gene structure, evolution and function and showed how common SNPs can be correlated throughout a gene due to linkage disequilibrium. Awareness of this phenomenon has led the Human Genome Project to initiate a new effort to map blocks of correlated SNPs, known as haplotype blocks. Haplotype blocks can be identified using a subset of the SNPs within the block. These identifying SNPs are called Tag SNPs. Dr. Nickerson uses Tag SNPs and a graphical analysis method called visual genotyping to facilitate analysis of polymorphism in candidate genes.

Dr. Nickerson presented a detailed analysis of SNP distribution and association in the breast cancer gene BRCA1, which provides new insights into the structure of this gene. Approximately 84% (72 kb) of the coding and non-coding regions of BRCA 1 was resequenced, identifying 301 SNPs, 10 non-synonymous cSNPs, many insertions and repeat sequence elements and one complex insertion/deletion mutation . The complex mutation contains a putative GATA1 transcription factor binding site. Nickerson also showed evidence of extensive linkage disequilibrium covering a large fraction of the BRCA1 gene. This result suggests that the structure of BRCA1 is highly conserved.

Influence of DNA Variation on Gene Expression

Jeffrey Trent, Translational Genomics Research Institute

Dr.. Jeffrey Trent discussed the relationship between gene expression and genetic variation. Dr. Trent explored the possibility that global gene expression profiles might yield informative signatures for phenotype (including the phenotype of a complex disease such as cancer) and genotype (*i.e.*, DNA polymorphism). Dr. Trent also described a global high resolution method for determining DNA copy number.

The gene expression profiles of different cell types are often quite unique, with distinct patterns of gene expression that can be readily discriminated. To understand the significance of these gene expression patterns, it is first necessary to identify key genes (discriminators) that differentiate one pattern from another, and then to assess the function of the discriminator genes.

Dr. Trent used global gene expression analysis to differentiate metastatic melanoma cells that would or would not respond to IL-2 immunotherapy. IL-2 is an important, but highly toxic, therapeutic option for metastatic melanoma patients whose cancer does not respond to other anti-cancer treatments. However, the response rate to IL-2 immunotherapy is only 10-15%, and it would be advantageous to identify responders prior to initiating IL-2 therapy. Dr. Trent analyzed gene expression in 40 melanoma patients including 8 IL-2-responsive patients. Although several discriminators were readily identified, Dr. Trent indicated that it was hard to rank their value. To facilitate pattern recognition in this data set, Dr. Trent used the Genes@Work software package (developed by the IBM Computational Biology Center), which is a pattern discovery algorithm based on supervised learning that identifies statistically significant gene expression patterns in microarray data sets. Using Genes@Work, Dr. Trent identified 32 candidate discriminator genes in IL-2-responsive metastatic melanoma cells, which will be further analyzed in future studies.

Cancer phenotype is frequently associated with DNA copy number changes. Comparative genomic hybridization (CGH), in which DNA from experimental and control cells are labeled and hybridized to metaphase chromosomes, is a method currently used to measure DNA copy number changes throughout the genome. CGH has high copy number resolution but low positional resolution. Dr. Trent described a high positional resolution variation on this method called array-CGH, in which DNA samples are hybridized to cDNA probes on a microarray chip. Array-CGH can be used to identify gene-level amplifications or deletions; thus, array-CGH has extremely high positional resolution, and is useful for genotypic analysis of cancer cells.

Dr. Trent used global gene expression profiling and array-CGH to analyze familial non-BRCA1/2 (BRCAx) breast cancer cells. Gene expression profiling successfully identified unique expression signatures for BRCA1 and BRCA2 cancer cells, which are distinct from the expression patterns of BRCAx cancer cells. Multidimensional scaling analysis of microarray data for the BRCAx cells further subdivided them into two discrete groups with distinct signatures. Cells from each of these groups were analyzed by array-CGH, revealing significant differences in patterns of gene amplification. This result suggests a possible correlation between gene expression signature and DNA copy number change. It also provides experimental data that may be useful in defining specific genetic changes that underlie the BRCAx breast cancer phenotype.

Relating Variation to Phenotype
Charles Rotimi, Howard University

Dr. Charles Rotimi emphasized the important roles of cultural and social factors as components of the environmental milieu that influences human health. The environment is highly complex including factors such as diet, physical activity, social stressors, trauma, income, educational status, politics and religion. In addition, the environment changes continuously over time, adding enormously to the complexity of the environmental variable.

Dr. Rotimi conceptualized the complexity of gene-environment interactions in the following equation:

$$G_n + E_n + G_nxE_n + D_{noise}$$

where G_n are strictly genetic effects, E_n are strictly environmental effects, G_nxE_n are gene-environment interactions and D_{noise} is developmental noise. It is only possible to separate and quantify genetic and environmental effects in a highly defined system in which the environmental factor is not allowed to change (*i.e.*, at a fixed moment in time). Thus, human health and disease can only be understood in a context-dependent manner.

Some genes confer a dosage-dependent risk of disease, which can be further modulated by one or more environmental factors. For example, the risk of Alzheimer's disease is lowest in individuals who do not carry the Apolipoprotein E4 (ApoE4) gene, highest in individuals with 2 copies of ApoE4 and intermediate in individuals with 1 copy of ApoE4. The frequency of ApoE4 is different in different ethnic groups (New Guineans > African-Americans ≈ Polynesians ≈Africans > Europeans). Thus ethnic origin can be considered a risk factor for Alzheimer's Disease. However, individuals with both ApoE4 and a history of head injury show a dramatically increased risk of Alzheimer's Disease.

It has been proposed that the human genome evolved when human lifestyle was characterized by physical activity and the environment was characterized by relative scarcity and famine (*i.e.*, traditional hunter-gatherer lifestyle). The human genome has not had time to adapt to "modern" lifestyle in developed countries, which is characterized by sedentary living and an abundance of food and material goods. Thus, a large percentage of humans living the U.S. and other developed

countries are at risk for several complex diseases such as obesity, diabetes and hypertension. However, the prevalence of these (and other diseases) varies greatly in different ethnic groups and in different geographical locations. For example, African Americans have a 1.5- to 2-fold higher risk of heart disease, stroke, diabetes, prostate cancer and colorectal cancer than Caucasian Americans. Urban living (as opposed to rural living) also tends to increase risk of some diseases including hypertension. These patterns reflect the complexity of gene-environment interactions and the complexity of environmental influences, as indicated above.

Dr. Rotimi pointed out that race, ethnicity and social class are extremely confounded in the U.S. Thus, attempts to ascribe health disparities in American ethnic minority populations to genetic factors may be ill-founded. For one thing, group identity (*i.e.*, the African American identity or the Hispanic American identity) is much more complex than self-identity, so it is not possible to define any ethnic group by its genome. Dr. Rotimi encouraged scientists to embrace a new way of approaching these problems based on an integrated understanding of a complex reality. That complex reality includes important cultural and social factors, some of which are relevant and important environmental risk factors for disease.

In closing, Dr. Rotimi suggested that human diversity should be embraced and that our differences, though worthy of study, should not be the basis of boundaries or barriers that divide or separate us.

Session II

Functional Genomics of Paraoxonase (PON1) Polymorphisms

Clement Furlong, University of Washington

Dr. Clement Furlong (University of Washington) described genetic variation in the human paraoxonase (PON1) gene and the effects of PON1 variants on sensitivity to the organophosphorus insecticides chlorpyrifos and diazoxon. PON1 is an HDL-associated plasma enzyme that metabolizes toxic organophosphates, oxidized lipids and some pharmaceutical agents including statins. During normal metaobolism, the primary substrates of PON1 may be oxidized lipids. However, PON1 plays an important role in the cytochrome P450 pathway that facilitates detoxification after exposure to organophosphate compounds.

Common insecticides including parathion, diazoxon and chlorpyrifos are relatively non-toxic. However, they are bioactivated by cytochrome P450 enzymes to neurotoxic oxon forms which inhibit acetyl cholinesterase. In some cases, humans are exposed directly to neurotoxic oxon organophosphates. Nerve gas agents sarin and soman are neurotoxic without bioactivation and can be hydrolyzed directly by PON1, although these compounds are relatively poor substrates for PON1.

Dr. Furlong presented a detailed analysis of polymorphism in PON1 and its impact on organophosphate sensitivity and cardiovascular disease. The human PON1gene has two common SNPs in the coding region, L55M and Q192R, 5 promoter SNPs, 1 intronic polymorphic CA repeat and 4 3'-UTR SNPs. The L55M SNP does not alter enzyme activity or phenotype significantly, but the Q192R SNP has significant functional consequences. Homozygotes for the R192 allele have more paraoxonase activity than heterozygotes or QQ homozygotes. The increased activity is, however, substrate specific, and diazoxonase activity is not increased in R192 homozygotes. The PON1 promoter region has 3 common haplotypes, which influence the amount of PON1 plasma activity. The promoter SNPs are in linkage disequilibrium with the L55M SNP. Recent analyses by Dr. Nickerson's group identified a number of additional SNPs in PON1 coding and non-coding regions which have not yet been well characterized.

The role of PON1 in organophosphate sensitivity was explored using a PON1 -/- knockout mouse. PON1-deficient mice die rapidly after exposure to chlopyrifos-oxon or diazoxon, while wild-type and heterozygote mice have high or moderate resistance, respectively. Surprisingly, PON1 genotype is not a determinant for paraoxon toxicity. These results were confirmed by measuring organophosphate resistance in PON1 knockout mice that had been injected with human Q192 or R192 PON1. Resistance to chlopyrifos-oxon or diazoxon was restored by injection of human PON1, but resistance to paraoxon was not. R192 PON1 provided much greater protection against chlopyrifos-oxon than Q192 PON1, but the two isoforms protected equally well against diazoxon toxicity.

PON1 activity also plays an important role in preventing oxidation of lipids and high density lipoprotein. Consistent with this role, previous studies suggested that altered PON1 expression or activity may be involved in carotid artery disease. Dr. Furlong examined the Q192R status of PON1 in a large set of carotid artery disease cases and controls. No correlation was observed between PON1 status at Q192R and disease susceptibility. However, Dr. Furlong also showed that the level of PON1 activity was significantly lower in cases carrying homozygous Q192 or heterozygote QR PON1 genotype. This suggests that secondary factors may reduce PON1 activity in these individuals, which increases their susceptibility to carotid artery disease.

In closing, Dr. Furlong discussed some implications of these studies on PON1 for environmental health. PON1 activity plays a role in detoxification of organophosphates by hydrolyzing neurotoxic oxon compounds. Thus, sensitivity to oxons correlates with PON1 status. It may be important to consider direct environmental testing for oxon compounds, instead of testing only for parent organophosphates. PON1 genotype may play a role in response to chemical warfare agents, and it may also influence susceptibility to Gulf War Syndrome. Future studies should also examine the roles of PON2, 3, and 10 in normal metabolism and in response to xenobiotic compounds.

Gene-Environment Interaction Related to Alcohol Use and Its Consequences

David Crabb, Indiana University

Many individuals consume alcohol who do not develop the disease pathology associated with alcholism, but the determinants of susceptibility to alcoholism have only recently begun to be understood. Dr. David Crabb discussed this phenomenon from the perspective of gene-environment interactions. In this case, environment includes access to alcohol and exposure to alcoholic behavior, and genetic factors include enzymes that metabolize alcohol. Sex is also a factor in susceptibility to alcoholism.

Studies of monozygotic and dizygotic twins suggest that a genetic factor or factors play a role in alcoholism. For example, the rate of concordance for alcoholism was 2-fold higher in dizygotic than in monozygotic twins. Adoption studies also document environmental influences that promote alcoholism. Adopted children had a 9-fold higher risk of alcoholism if their adoptive mother was alcoholic and a 14-fold higher rate of abstention if both adoptive parents were alcoholic. Environmental factors also influence the age at which children begin to drink, and a 50% higher risk of alcoholism is associated with onset of drinking before the age of 14.

Alcohol dehydrogenase (ADH) converts alcohol to acetaldehyde and aldehyde dehydrogenase (ALDH) converts acetaldehyde to acetate. ADH2 exists in one slow (ADH2*1)and two fast isoforms.(ADH2*2 or 2*3). ALDH2*1 is a faster isoform than ALDH2*2. ADH2*2, ADH2*3 or ALDH2*2 are associated with increased levels of acetaldehyde after alcohol consumption.

The ALDH2*2 allele, common in Asian populations, is associated with flushing after alcohol consumption. In most cases, this allele is a strong protective factor against alcohol consumption and it is present at decreased frequency in alcoholics from Japan, Taiwan, China and Korea. However, recent trends indicate increasing number of ALDH2*2 carriers are becoming alcoholics in Japan. ADH2*2 may also be protective against alcoholism, as it is found a lower frequency in alcohol abusive patients in many countries worldwide. Cultural factors may also play a role in alcoholism in some populations including Jews, Israelis, Russian immigrants to the U.S. and African American women.

Dr. Crabb discussed the potential genetic and environmental factors involved in liver disease (ALD) in alcoholic women. Although women alcoholics start drinking later and drink less per day than male alcoholics, they die from liver disease at a much younger age than alcoholic men. This suggests that women are more sensitive to this pathological manifestation of alcohol abuse. Susceptibility to ALD does not correlate with incidence or severity of hepatitis C infection. Because women have a lower average body weight and higher average body fat content than men, the same amount of alcohol is a higher effective dose per kilogram body weight and per kilogram body water in women than in men. Nevertheless, the alcohol elimination rate per lean body mass is slightly higher in women than men, and the alcohol elimination rate per liver volume is similar in women and men. Crabb also indicated that women have a lower capacity to handle fatty acid load than men. This may contribute to liver hypoxia, high plasma endotoxin, more free radical adducts and more infiltrating neutrophils in alcoholic women than in alcoholic men.

Infants born to alcoholic women develop fetal alcohol syndrome as a consequence of alcohol exposure *in utero*. Some evidence suggests that maternal ADH2*3 may protect against fetal alcohol syndrome. It was reported that ADH2*3 is more common and that ADH2*2 is less common in mothers who drink heavily during pregnancy and in infants with fetal alcohol syndrome.

Alcohol metabolizing genes have also been linked to alcohol-related cancers including esophageal oropharyngeal, laryngeal and stomach cancer. ALDH2*2 may be a risk factor for esophageal cancer and ADH2*1 and ALDH2*1 may be risk factors for oropharyngeal and laryngeal cancer. Alcohol use *per se* is a major risk factor for esophageal cancer in East Asians.

Gene-Environment Interactions in BRCA Related Breast Cancer

Mary-Claire King, University of Washington

Dr. Mary-Claire King described new approaches to identifying genetic factors involved in sporadic breast cancer. Although BRCA1 and BRCA2 are major risk factors for a large fraction of heritable breast cancer, the genetic and environmental causes of most non-BRCA1/2 breast cancer remain poorly understood. Dr. King suggested that gene-environment interactions are likely to play a major role in many cases of breast cancer.

Some of the environmental factors promoting breast cancer may be related to social and behavioral change in women. Strong evidence indicates that early age at menarche and late age of first pregnancy are strong risk factors for breast cancer. Early age of menarche is promoted by obesity and high body mass index in young women. Late pregnancy may reflect prosperity and the increased number of women with professional careers. These factors have led to large recent increases in women experiencing both of these risk factors and these trends may be directly involved in the increase in breast cancer incidence in women in the U.S.

BRCA1/2 are linked to 70% of breast cancer in high-risk families, but to <10% of all breast cancer cases. Nevertheless, many non-BRCA1/2 breast cancer cases are associated with a family history of

breast cancer, suggesting that there are unknown factors that increase breast cancer risk and cause clusters of breast cancer in families. Dr. King considered several possible explanations for this pattern. The first possibility is that the breast cancer clusters occur by chance, which is very unlikely. Other alternative explanations are shared exposure or lifestyle, cryptic mutations in BRCA1/2, mutations in other known genes (*i.e.*, CHEK2, p53, PTEN, ATM), or gene-environment interactions involving unknown genes with variable penetrance. Dr. King favors the last of these possibilities.

Efforts to identify environmentally-responsive genes involved in non-BRCA1/2 breast cancer have proven very difficult. This fact likely reflects the involvement of multiple linked genetic regions, allelic heterogeneity, gene-gene and gene-environment interactions, which make these factors difficult to analyze with available genetic methods. In addition, relatively few families have a large number of cases, and analyses of these families are unlikely to identify common genes. Dr. King indicated that large-scale genome-wide searches for novel genetic factors will be needed to understand non-BRCA1/2 breast cancer.

Dr. King initiated studies of several breast cancer families in which no genetic alteration had been detected in BRCA1/2 CHEK2, p53, PTEN or ATM. These studies used a novel method called fosmid library allele separation haplotyping (FLASH) to identify BRCA1/2 alleles that are not detected by available screens. This method was developed by Maynard Olson (University of Washington). FLASH uses STS-based hybridization probes generated by PCR to correlate haplotype in an affected sibling to their phenotype. Fosmid libraries are prepared from the DNA of a proband, her siblings and parents. Probes are initially hybridized to large fosmid pools, followed by smaller pools and finally to individual fosmids. Dr. King described a FLASH experiment targeted to 160 kb of BRCA1 genomic DNA. This experiment identified several novel BRCA1 alleles, including a deletion of exon 20, a deletion of exons 20-22, a deletion of exon 3, a deletion of exon 22, and a deletion of exons 14-20 with a stop codon at 1829. This approach can be extended to non-BRCA1/2 candidate genes for breast cancer susceptibility.

Gene-Environment Interactions in Human Leukemia

Martyn Smith, University of California at Berkeley

Dr. Martyn Smith described recent studies of the genetic and environmental factors in blood-related cancers including leukemia, lymphoma and myeloma. In the U.S., these cancers account for as many as 620,000 cancer cases and 58,000 cancer-related deaths per year. They cause more deaths in children under the age of 14 than any other disease. There are four types of leukemia: acute, chronic, myeloid and lymphocytic.

Approximately 20% of leukemia cases are thought to be induced by known environmental factors, which include benzene, radiation and chemotherapeutic agents. Genetic factors are also thought to play a significant role, especially in pathways controlling oxidative DNA damage. These pathways include the NAD(P)H:quinone acceptor oxidoreductase (NQO) pathway and pathways that regulate folate metabolism. Dietary intake of substrates metabolized by these two pathways may also play a role in the etiology of leukemia.

Dr. Smith indicated that further genetic analyses of potential susceptibility factors would be useful and would help narrow attention to relevant environmental exposures that increase leukemia risk. Possible genetic approaches include candidate gene analysis in individual patients, pathway analysis in individual patients or using DNA pools (*i.e.*, all cases *vs.* all controls), or genome-wide scans in DNA pools.

NQO1 plays an important role in preventing oxidative damage caused by exogenous or endogenous quinones. There is a common polymorphism (prevalence 5-20%) in the NQO1 gene that strongly reduces (heterozygote) or eliminates (homozygote) NQO1 activity. This polymorphism is aC609T variation, which results in a Pro187Ser amino acid change and a complete loss of enzyme activity. Case-control studies indicate a 1.5- to 2.5-fold increased odds-ratio for several types of leukemia in association with the 609T variant. This effect is relatively small, but adverse environmental exposure may interact with this genetic variant, leading to increased risk of disease. Genetic variation in NQO1 is however not expected to play a role in leukemia risk after exposure to non-quinone-type oxidative stressors.

Folate metabolism is critical to cell survival and stability, because it provides precursors for DNA synthesis, DNA repair and methylation. Some of the critical enzymes in folate metabolism are methylenetetrahydrofolate reductase (MTHFR), serine hydroxymethyl transferase (SHMT) and thymidylate synthase (TS) There is a fork in the pathway which shunts folate metabolites into DNA synthesis/repair (TS) or methylation (MTHFR). There are several functional SNPs in the enzymes of folate metabolism . These variants reduce leukemia risk 2- to 3-fold in heterozygotes and 3- to 10-fold in homozygotes. Dietary folate also protects 2- to 3-fold against leukemia . Dr. Smith also presented some evidence for gene-gene interactions involving polymorphic variants of SHMT and TS.

Dr. Smith discussed the pros and cons of using DNA pools for population-based studies of leukemia risk. Pooling DNA samples is advantageous because it reduces labor and reagent costs, conserves DNA samples, and makes large scale studies more feasible. The disadvantages of using DNA pools are loss of power and a high rate of false positive or false negative results. In addition, it is not possible to study haplotypes, perform longitudinal outcome studies or analyze gene-environment interactions when DNA pools are used instead of individual DNA samples. Dr. Smith analyzed the allele frequency of several DNA repair genes (candidate pathway approach) in pooled DNA from 444 acute myeloid leukemia (AML) cases and 828 matched controls. Allele-specific PCR showed significant enrichment for one allele of Fas-associated death domain protein (FAS) in cases, but not in controls; a similar amount of allele enrichment (odds-ratio of 2.9 or 1.8, respectively) was observed for pooled DNA or for individual DNA samples. The hypothesis that FAS plays a role in the etiology of leukemia is supported by other evidence, including the fact that FAS knockout mice develop AML if they overexpress BCL2. One polymorphic variant of FAS may alter gene expression by interfering with SP1 transcription factor binding to the FAS promoter region.