DDT Vol. 9, No. 14 July 2004

Using germ-line genetic variation to investigate and treat cancer

Sharon A. Savage and Stephen J. Chanock

For many years, there has been spirited debate as to the relative importance of environmental and genetic factors in the pathogenesis of cancer. Current efforts to annotate the human genome for germ-line genetic variants should establish the foundation for dissecting the contribution of genetics to the risk for cancer susceptibility. Population-based studies should be conducted to determine the influence of germline genetic variation on cancer outcomes, including the efficacy of anti-cancer drugs and the risk for life-threatening toxicities. Although we are early in the investigation of the influence of germline genetics on cancer outcomes, it is likely that, in the future, it will be possible to individualize therapeutic interventions. In turn, knowledge of genetic risk factors could afford opportunities for prevention, early intervention and minimization of deleterious toxicities associated with cancer therapy.

Sharon A. Savage
Stephen J. Chanock*
Section of Genomic Variation
Pediatric Oncology Branch
Center for Cancer Research
National Cancer Institute
National Institutes of Health
Bethesda, MD, USA
*e-mail: sc83a@nih.gov
Stephen J. Chanock
Core Genotyping Facility
National Cancer Institute
National Institutes of Health
Bethesda, MD, USA

▼ Although great effort has been directed at elucidating basic steps in cancer progression, it has become evident that cancer is a complex set of diseases. Classically, specific types of cancer have been defined by pathological criteria. Investigation into the molecular signatures of cancer have added a new dimension to the categorization of cancer, as well as to the understanding of underlying disturbances in crucial cellular pathways. The profile of genetic changes that contribute to the development and outcomes in cancer includes both germ-line and somatic mutations. Germ-line mutations can be rare but strongly linked to a cancer (e.g. *BRCA1* and breast cancer) or a commonly observed variation associated with a higher risk for cancer (e.g. *NAT2* and bladder cancer) [1–3]. In parallel, the identification of somatic mutations has focused attention on crucial cellular pathways that are altered during malignant transformation; for instance, somatic mutations in *TP53* occur with variable frequencies in cancers of the breast, ovaries and colon [4]. Furthermore, we can now detect signature chromosomal abnormalities and characteristic translocations (e.g. *EWS-FLI1* in Ewing Sarcoma or *BCR-ABL* in chronic myelogenous leukemia). Overall, cancer has emerged as a complex set of diseases, which require further investigation to dissect the importance of genetic changes and the interaction of these genetic changes with environmental factors [5] (Figure 1).

Classical studies in cancer epidemiology have highlighted marked differences in the incidence of a particular pathological type of cancer according to geographical regions. For example, males living in Japan have an incidence of stomach cancer that is 22 times greater than men living in Kuwait [5]. Differences in the incidence in prostate cancer also vary by geographical regions; for instance, African–American men in the USA have an incidence 70-times greater than that observed in men in China [5]. These striking differences in incidence highlight differences in both environmental factors and underlying population genetics (which reflect differences in underlying patterns of germ-line genetic variation). It is also important to recognize that there are additional factors that contribute to differences in cancer incidence, such as environmental exposures, socioeconomic status and access to healthcare. In the USA, it is plausible that these factors contribute to differences in outcomes for all cancers and can partially explain the fact that men of African–American background have a higher rate of mortality than men of Caucasian background [6]. Even in a rare pediatric tumor, Ewings sarcoma, there is a substantial difference in incidence according to ethnic background: the disease is ~100 times more common in children of Caucasian ancestry compared to children of African ancestry in the USA and Europe [7].

1359-6446/04/$ – see front matter PII: S1359-6446(04)03179-4

Pharmacogenomics is the study of the inherited basis of inter-individual differences in response to known drugs [8,9]. The outcomes surveyed in pharmacogenomic studies can be separated into studies that search for genetic variants that are associated with: (1) severe adverse effects, which, in turn, can be used to screen for individuals who should not receive the drug in question or (2) the ability to predict the efficacy of a drug. The promise of pharmacogenomics is that it will generate genetic markers that will guide therapeutic choices in the future, in a manner dubbed as 'individualized medicine' [10]. To date, few studies have unequivocally shown the importance of germ-line variants as either prognostic markers or determinants of drug response or toxicity. This is because only a small fraction of known genes have been adequately studied.

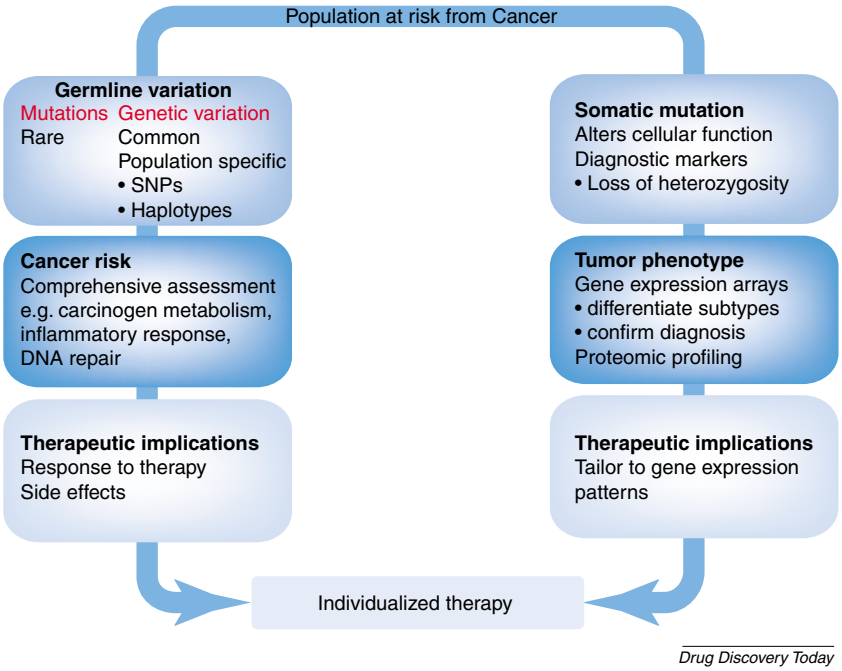

Figure 1. This figure depicts the flow of information required to proceed from population-based studies to application in individual cases. For both germ-line and somatic mutations, 'at risk' is defined by previous studies that have examined variation as a risk factor for cancer as well as a predictor for therapeutic outcomes, namely toxicity or efficacy.

Genetic variation in the human genome

The generation of a draft sequence of the human genome in 2001 has resulted in an opportunity to survey genetic variation across the entire human genome [11,12]. The scope of genetic variation in the human genome is substantial [13]. Historically, variable tandem repeat units (of two, three or four nucleotides), also known as microsatellites, were catalogued and subsequently used to map genetic diseases in family pedigrees, and rarely in studies of unrelated subjects. Microsatellites are present throughout the genome but at a density far below that observed for single nucleotide polymorphisms (SNPs). The most common genetic variant, the SNP, is a stable, single-base substitution; by definition, the frequency of a SNP should exceed 1% in at least one studied population [11–13]. SNPs are different from clinically important mutations in that the latter are rare and result in a significant phenotypic change; for example, mutations in the *RB* gene are strongly linked with familial retinoblastoma and rare, familial germ-line mutations in the *TP53* gene are linked with the familial cancer-predisposition syndrome, Li-Fraumeni Syndrome [14–16]. It has been estimated that there are as many as 10–15 million common SNPs, which still represents less than 0.1% of the total 2.9 billion base pairs of the genome [13]. Furthermore, ~5 million SNPs have a frequency of 10% or greater in a given population [17,18]. Many SNPs appear to be specific to one population; current estimates based on SNP surveys suggest that as many as 20% of SNP in populations of African ancestry are specific to this background, whereas less than 5% of common SNPs are seen only in populations of North European Caucasian background [18,19]. These differences in allele frequencies probably reflect the evolution of mankind; the current data point towards the origin of mankind in Africa, which would explain the greater number of SNPs, including population-specific SNPs, in this group [20].

The majority of SNPs are silent, namely, they do not have a measurable effect on a phenotype, such as an effect on the expression or function of a gene. However, an estimated 50,000–250,000 SNPs do confer a biological effect, either an alteration in the function of the gene product or regulation of the gene [21]. In this regard, there has been intense interest in validating SNPs that either change an amino acid residue (known as nonsynonymous SNPs) or alter the expression of the gene, often clustered in the 5′ or 3′ untranslated region [22]. SNPs that do not alter the sequence of an amino acid residue, known as synonymous SNPs, can still have an important effect on the stability and expression of the gene; for example, a synonymous SNP in the human dopamine receptor D2, *DRD2,* gene functionally alters mRNA (mRNA) stability, leading to measurable differences in gene expression [23].

International efforts have been focused on the collection and cataloging of SNPs, such as the SNP Consortium (http://www.tsc.org), Human Genome Variation Base (http://hgvbase.cgb.ki.se/) and db-SNP (http://www.ncbi.nlm.nih.gov/SNP/). The Cancer Genome Anatomy Project of the National Cancer Institute (http://cgap.nci.nih.gov) created the SNP500 cancer database of SNPs in genes that are important in molecular epidemiology (http://snp500cancer.nci.nih.gov) [24]. This program seeks to validate SNPs by resequence analysis in 102 publicly available reference samples from four major ethnic groups (Caucasian, African, Hispanic and Pacific Rim) [24]. Parallel programs, such as the Variation Discovery Resource (SeattleSNPs, http://pga.gs.washington.edu and http://genevar.org) provide detailed sequence data on common SNPs and haplotypes in key pathways of genes, such as DNA repair genes or genes of innate immunity (including interleukins and cytokines) [25]. What we have learned so far, is that SNPs are not evenly distributed across the genome; moreover, nearby SNPs, lying within a small number of bases, can impact the fidelity of SNP genotype analysis [24]. The databases, especially db-SNP, are rapidly accumulating additional SNPs but the annotation is not complete for common SNPs across the genome [26].

Although studies of single SNPs in single genes are informative, it is important to recognize that SNPs are not inherited as an individual entity but instead in combination as blocks, often encompassing a portion of a gene or an entire gene. These blocks are defined by the presence of strong linkage disequilibrium (LD) (the non-random association between alleles of different loci on the same chromosome) between SNPs [27–29]. This results in the production of haplotypes or sets of SNPs inherited on a chromosome from generation to generation. Measurement of LD between SNPs can be used to define the components of a haplotype, which, in turn, reduce the numbers and complexity of SNPs needed to evaluate variation in a gene or chromosomal region [28]. Blocks of LD can vary greatly among populations because of differences in population history or geographical-selective pressures (e.g. endemic pathogens or dietary restrictions); the size of blocks within the genome can vary between two or three kilobases up to about 40 kilobases [27–29]. Block structure in African populations is smaller and includes a greater density of common variants [27,28]. It is possible to determine haplotypes in family pedigrees by tracing the pattern of linked SNPs or by computational approaches, such as the expectation-maximization algorithm to predict population-specific haplotypes [30,31]. Once determined, a haplotypic structure can be used to identify whether or not a locus is associated with disease outcome [32]. With this information, it is possible to reduce the number of SNPs analyzed for a given association study because only a subset of SNPs in strong LD is needed to capture the common variation across the gene (in the range of 95–98%) [33–35].

A major international effort to determine common SNPs, with a minor allele frequency of more than 20–30%, has been launched, known as the HapMap project (http://www.hapmap.org) [36]. Its purpose is to genotype 300,000 to 500,000 SNPs across the entire genome in a reference set of subjects (comprising West African, Caucasian and Pacific Rim subjects). This resource will be invaluable for understanding the structure of common variation across the genome through the analysis of LD. The long-term goal is to provide a resource to perform whole-genome scans in well-designed studies (see below).

Methods for studying genetic variation in cancer

In the search for genetic risk factors for a complex disease such as cancer, the challenge lies in using appropriate strategies to identify informative genetic variants, either by (i) linkage analysis, which is suitable for mapping highly penetrant genetic loci in family pedigrees, or (ii) genetic association studies, which examine the distribution of common genetic variants in unrelated subjects, usually through case control analysis [37,38].

Microsatellite markers

Microsatellite markers have been successful in identifying highly penetrant genes, particularly in monogenic Mendelian disorders, such as cystic fibrosis. For example, with large family pedigrees and dominant mutations, Hall *et al*. narrowed down a locus on 17q21 as a risk factor for familial breast and ovarian cancer [39], which was subsequently determined to be the *BRCA1* gene. Traditional methods for identifying risk factors for the development of cancer have included the investigation of environmental risk factors (e.g. smoking) and the use of LD between microsatellite markers to map chromosomal regions in which genes that are implicated in cancer reside. The inheritance pattern of the majority of cancers do not fit a familial autosomal dominant or an autosomal recessive model, therefore, it is less likely that microsatellties will be informative in studies of unrelated cases and controls, particularly as it relates to predicting toxicity or outcome associated with one or more drugs [9,22].

SNPs and the candidate gene approach

Historically, the candidate gene approach has investigated genes in which there is prior evidence that there are common genetic variants with functional properties; in this regard, the approach has been driven by a plausible disease

Table 1. Genetic variants that influence susceptibility or outcome in cancer

Mechanism	Gene	Association	Refs
Carcinogen metabolism	*N*-Acetyltransferase 2 (*NAT2*)	Bladder and colon cancer (increased risk for slow acetylators)	[2,3,42–44]
	Myeloperoxidase (*MPO*)	Lung cancer (decreased risk with G-463A promoter SNP)	[40,41]
Inflammation	Interleukin 1 (*IL1A*)	Gastric cancer (increased risk with regulatory SNP)	[74,75]
	Interleukin 10 (*IL10*)	Cervical cancer (increased risk associated with increased production)	[76]
		Graft versus Host Disease (increased risk by allele)	[77]
DNA repair	*XRCC2, XRCC3, Ligase IV*	Breast cancer (increased risk with set of variants)	[78]

model of a gene-environment interaction [2]. Because myeloperoxidase activity (encoded by the *MPO* gene) participates in the metabolism of tobacco smoke, several studies have examined a functional SNP in the *MPO* gene [40,41]. There is evidence that the risk for lung cancer is lower in individuals who carry the A allele at –463 of the proximal promoter in the *MPO* gene [40,41]. The G to A transition at –463 of the proximal promoter has been shown to decrease the mRNA expression of *MPO* and led some to hypothesize that less *MPO* gene product, which metabolizes oxidants such as by-products of tobacco smoke, could be protective against lung cancer. Intense investigation has focused on the contribution of the genetic variations in the *N*-acetyltransferase (*NAT2)* gene to risk for specific cancers, especially bladder cancer; in particular, differences in the activity of *NAT2*, encoded by so called rapid and slow acetylator genotypes could explain the association between the *NAT2* gene and tobacco smoke and subsequent risk for bladder cancer [2,3,42–44]. The slow acetylator phenotype is associated with an increased risk for bladder cancer compared with individuals with the fast acetylator phenotype, especially when combined with tobacco use [43,44]. Additional examples of genes studied and their association with cancer are shown in Tables 1 and 2.

The importance of studying individual populations cannot be underestimated because of the differences in the distribution of SNPs by geographical and population migration history. This has important consequences for applying SNPs in studies. For example, the BRCA1 mutation, associated with breast cancer, is highest in frequency in individuals of Ashkenazi Jewish descent [45]. A SNP in the proximal promoter of the myeloperoxidase gene (*MPO*), at –G463A has been associated with decreased lung cancer risk and is more common in Caucasian individuals (~20% mean minor allele frequency) than in individuals of African–American background (~5% mean minor allele frequency) [24,40,41].

Haplotype analysis

The number of genes with a well-defined, functional SNP is small, therefore, most studies rely upon the choice of common genetic variants that could either have potential function or contribute to a common haplotype structure [33,37]. By selecting SNPs that tag the common haplotypes, known as ht-SNPs, candidate gene studies can interrogate variation across a gene. If there is a positive signal for one or more haplotype, its constituents can be carefully studied to elucidate the functional basis of the association [33–35]. There are several advantages for the analysis of haplotypes. First, knowledge about the functional SNP is not necessary. Second, it can be determined if any variation in the gene or region is associated with an outcome, such as drug toxicity or efficacy. Third, if the common haplotypes (with a frequency greater than 3–5%) have been

Table 2. Genetic variants and pharmacologic endpoints

Metabolism of chemotherapy	Thiopurine S-methyltransferase (*TPMT*)	Increased thiopurine cytotoxicity (hematotoxicity during acute lymphoblastic leukemia therapy)	[47–55]
	Dihydropyrimidine dehydrogenase (*DPD*)	5-fluorouracil metabolism (increased toxicity)	[78–80]
	UDP-glucuronosyltransferase 1A1 (*UGT1A1*)	Irinotecan metabolism (increased neutropenia and diarrhea)	[56–58]
Efficacy of chemotherapy	Epidermal growth factor receptor *(EGFR)*	Mutations in non-small-cell lung cancer predict for response to Gefitinib	[60]

analyzed and no association found, provided that the study is sufficiently powered, it is possible to conclude that genetic variation in the gene is not associated [32,34,35]. In turn, investigators can turn to other genes for study. This paradigm can be expanded to look at dense sets of SNPs or ht-SNPs across key genes in the metabolism of a drug to determine if variation in the variation in genes that influence uptake, activation, degradation or excretion could be associated with efficacy or sever toxicity [8,9].

The same principle for haplotype analysis can be extended to include a sufficient density of SNP markers to capture most blocks across the genome. Upon completion of the HapMap project, there will be a common set of SNPs that could be used to generate a map of LD across the genome [36]. This map can be used to scan the entire genome in search of regions in which there is a disproportionate distribution of one or more haplotypes that could be associated with a drug toxicity or outcome, such as efficacy. The analysis of a set of microsatellites, spaced across the genome, will probably be supplanted by the dense map of SNPs generated by the HapMap [36]. However, the sheer number and cost of undertaking a study with SNPs chosen at regular intervals across the genome (i.e. every 2–5 kb) is currently prohibitive (the estimated cost to analyze 500 cases and 500 controls over 300,000 SNPs is over 3 million USD in 2004) [17,18,46]. Because intense analytical requirements have not yet been established for the study of the ~300,000 SNPs needed to analyze the entire genome, not to mention the high cost, most investigators will continue to pursue the candidate gene approach and chose genes, either individually or in groups based on inclusion in a biological pathway, such as a drug metabolism pathway.

Genetic variation and individualization of treatment

The genomic revolution has generated a rich resource for the investigation of genetic factors that could influence both efficacy and toxicity associated with anticancer therapy (see Table 2). Before individualization of therapy can be developed, even with single SNPs or haplotypes, it will be necessary to conduct well-designed and sufficiently powered studies that, by necessity, are population-based. Consequently, what we learn from large-scale studies, will, in turn, be translated into clinical paradigms that will have to be applied to the individual patient. So far, there have been a few notable examples of the importance of pharmacogenomics, as discussed below.

Predicting treatment toxicity

The thiopurine methyltransferase (*TPMT*) gene is important for toxicity associated with anti-metabolites, such as 6-mercaptopurine, and the risk for relapse in acute lymphoblastic leukemia, ALL [47–55]. *TPMT* catalyzes the S-methylation of thioguanine, thiopurines (used in maintenance therapy in childhood ALL) and azathioprine (an immunosuppressant used to minimize the risk for graft rejection following solid organ transplantation). The majority of the population has high enzyme activity of *TPMT*, however, 10% have intermediate activity and 0.3% have low or undetectable enzyme activity [47,49]. The annotation of the *TPMT* gene has identified a large catalog of SNPs, of which a subset alter the activity of the enzyme, as measured in red blood cell assays [49,50,52]. Individuals with low *TPMT* activity have severe hematopoietic toxicity after the administration of thioguanine, thiopurines and azathioprine [47]. The same *TPMT* variants that alter the risk for hematotoxicity during leukemia also correlate with event-free survival in childhood leukemia and decreased risk for secondary malignancies [53,55]. In summary, there is adequate data to recommend that children undergoing therapy for ALL should be assessed for common *TPMT* alleles to help guide therapeutic decisions.

Genetic variants in one of the genes responsible for the metabolism of irinotecan, a chemotherapy used to treat patients with colorectal cancer, can alter the risk for dose-limiting toxicities associated with irinotecan therapy [56–58]. The hepatic UDP-glucurosyltransferase 1A1 (*UGT1A1*) gene inactivates the active form of SN-38, the active metabolite of irinotecan. The expression of *UGT1A1* is altered by the number of dinucleotide, TA repeats, most commonly between five and eight in the promoter region [57]. A variant that includes seven TA repeats instead of six (which is observed in the majority of subjects), is associated with reduced *UGT1A1* expression and less SN-38 inactivation. Consequently, individuals with genetic variants that decrease enzymatic activity are at significantly higher risk for irinotecan-associated dose-limiting side effects, including diarrhea and neutropenia [56,58]. It is notable that this genetic variant is a dinucleotide repeat polymorphism in the functional promoter, which highlights the importance of genetic variants, other than SNPs.

Predicting efficacy of cancer therapy

To date, few studies have established the importance of germ-line variants as either prognostic markers or determinants of therapeutic response. Recently, activating mutations in the epidermal growth factor receptor (EGFR) [59,60] were shown to be crucial for responsiveness to Gefitinib, the first molecular targeted agent to be approved for the treatment of non-small-cell lung cancer. It had been noted that 10–19% of patients responded [60]. Resequence

analysis of the *EGFR* gene in patients receiving therapy with Gefitinib revealed that the responsive subgroup had specific mutations that lead to increased growth factor signaling. This preliminary study raises several important issues. First, screening of patients for the EGFR mutations should be considered once these studies are confirmed. Second, it suggests that newly diagnosed patients with non-small-cell lung cancer could be screened to avoid usage of an expensive therapy in those unlikely to respond. Alternative therapies could be considered sooner. Finally, the study highlights the importance of sequence analysis of candidate genes in cancer to discover somatic or germline changes that could be associated with efficacy or toxicity of anticancer drugs [61,62].

SNPs and cancer outcome

If SNPs are to become part of the clinical armamentarium for decision making, the identification of SNPs that alter outcome will certainly call attention to high risk cases. In this regard, the study of cancer susceptibility and outcomes should identify new targets for drug development and perhaps define risk factors for toxicity that will have to be balanced against poor prognosis. For instance, there has been great interest in SNPs in *CYP3A4,* which have been associated with long-term survival in prostate cancer. The A-290G in the *CYP3A4* proximal promoter of a gene required for the oxidation of testosterone to 2β-, 6β, or 15β-hydroxytestosterone is associated with more severe disease, even when controlled for stage and grade.

Designing studies of genetic variation and cancer

The study of cancer as a complex genetic disease presents a fundamental challenge: the balance between establishing a genetic variant in a single gene and its interaction with one or more other genes. In cancer susceptibility studies, it is likely that environmental factors, such as tobacco or industrial carcinogens, can interact with one or more variant [5]. As mentioned previously, the effect of tobacco smoke on bladder cancer is enhanced in individuals with the slow acetylator phenotype of *NAT2* [2,3]. The paradigm in cancer susceptibility studies is applicable to cancer pharmacogenomics [8,9,61]. However, the challenge of a drug or intervention can be studied in a well-defined case control or cohort setting. Still, it is likely that more than one variant will contribute to outcome and the problem lies in identifying informative variants, many of which will have modest effects on disease risk or modification of disease complications [22,63,64]. It is also notable that the investigation of the genetic contribution to different cancers might identify key genes or pathways that could be targeted in new therapies, either directed at the informative variant or at the metabolism pathway for the anti-cancer agent.

Challenges of study design

The interpretation of any single-gene association study must be guarded, especially because many association studies have failed to replicate. It is prudent to consider a SNP association real only after confirmatory studies have been published. There is an important caveat to consider – a finding in one population might not replicate in another population of different ancestry [63,65–68]. Understanding the genetic heterogeneity of a study population should be factored into the interpretation of the results and specifically before generation of either public health or individualized pharmacogenetic recommendations.

There are several reasons why studies do not replicate. Often, a poor study design might compare cases that are not appropriately matched with suitable controls; this can lead to a false association that is likely a result of other factors, such as different populations genetics or access to healthcare. The most common problem is the lack of sufficient power to detect a true association. Some have commented on the publication bias, also known as 'winner's curse', yet other biases within the study can confound the analysis [65–68]. There is also the problem of multiple tests conducted on a small dataset, in which borderline 'p' values are selectively reported [66,69]. Recently, an analytical process has been proposed to assess the likelihood that a finding is a false positive; an assessment of a prior probability that the gene, or its variant, is associated with the outcome in a study is considered along with the size of the study population and the allele frequency of the gene being studied [69].

Knowledge of the underlying population genetics is important for several reasons. The comparison of subjects from appropriately matched geographical and ethnic background should minimize the effect of population stratification [70–72]. This issue has been suggested as a contributing factor to the plethora of false positive reports in genetic association studies, yet careful analysis of both methodological issues and the examination of large studies do not provide support for the contention that this is a significant issue for studies examining Caucasians of European background (either in Europe or the USA). Most would agree that the problem is significant in groups in which there is substantial admixture of populations, such as in subjects of Hispanic or African–American background in the USA [70]. In addition, it is important to have knowledge of SNPs residing near to the SNP under study because neighboring SNPs can interfere with assay performance, leading to miscalled genotypes [24].

Future directions

For many years, there has been spirited debate as to the relative importance of environmental and genetic factors in the pathogenesis of cancer. With the new tools of human genomics (i.e. knowledge of the catalog of human genes and their annotation for genetic variation), it is now possible to look more closely at this issue, in each type of cancer. Moreover, the genomic revolution has set the stage for dissecting the genetic contribution to both drug response and drug toxicity, particularly restricted to each type of cancer. In the future, it might be possible to individualize therapeutic decisions, partly based on the profile of germ-line genetic variants. So far, we have learned that genetic variation is extensive across the genome and that the genetic contribution to most disease outcomes is complex, that is, it involves more than one gene.

As the search for genetic determinants of pharmacological outcomes advances, it is important to keep several factors in mind. First, genetic variation can vary greatly by populations, which means that careful annotation of gene variants is required before completing studies. Second, validation of SNPs in available databases of normal individuals is required before choosing SNPs to study in a disease population. Third, the determination that a SNP, or a particular haplotype, is associated with a drug toxicity or outcome necessitates additional information to understand the plausible mechanism underlying the link. In this regard, additional SNPs or variants in LD could either be functionally responsible for the effect observed or perhaps act in cooperation with the identified genetic marker. It could be that combinations of variation between genes of certain pathways will confer the greatest risk for the development of cancer. For instance, members of the *CYP2C* family of the P450 cytochrome family might represent suitable targets for analysis of the toxicity and outcome associated with administration of taxol, a chemotherapy agent used for treatment of female hormonal cancers of the breast and ovaries [73]. Finally, in general, SNPs alter the risk for response or toxicity to a drug, but are neither sufficient nor necessary for the outcome under study.

As more information is generated by the analysis of SNPs, and haploypes in crucial pathways, the difficult task of applying these genetic markers to clinical practice will become more urgent. Although insight into the contribution of germ-line variants (i.e. SNPs and haplotypes) to pharmacological outcomes holds the promise for developing individualized approaches to healthcare, two major tasks lie ahead. One is the complex analysis of integrating multiple genetic variants into decision-making algorithms that can guide healthcare providers. The second will require a shift in the practice of medicine to better incorporate the assessment and recommendation of drug intervention based on risk. In this regard, armed with information predictive for possible toxicity or deleterious outcomes, practitioners and patients together will have to make these important decisions. However, for this to occur, education of the public and healthcare network will need to take place; this will require a better understanding of the importance of multiple genes in any cancer outcome instead of the classical dictum of genetics 'a single gene leads to single outcome' [64]. Finally, the genomic revolution could improve our capabilities to assess those at greatest risk for developing malignancy and to address the response to therapy more specifically. This will provide the foundation to improve the effectiveness of preventive medicine by earlier identification of individuals at risk, which, in turn, can lead to earlier intervention strategies. These can be either avoidance of a strong factor, such as tobacco smoke, or intervention with a chemopreventive agent.

References

1 Fitzgerald, M.G. *et al.* (1996) Germ-line BRCA1 mutations in Jewish and non-Jewish women with early-onset breast cancer. *N. Engl. J. Med.* 334, 143–149

2 Marcus, P.M. *et al.* (2000) Cigarette smoking, N-acetyltransferase 2 acetylation status, and bladder cancer risk: a case-series meta-analysis of a gene-enviroment interaction. *Cancer Epidemiol. Biomarkers Prev.* 9, 461–467

3 Vineis, P. *et al.* (2001) Current smoking, occupation, N-acetyltransferase-2 and bladder cancer: a pooled analysis of genotype-based studies. *Cancer Epidemiol. Biomarkers Prev.* 10, 1249–1252

4 Børresen-Dale, A.L. (2003) *TP53* and breast cancer. *Hum. Mutat.* 21, 292–300

5 Rothman, N. *et al.* (2001) The use of common genetic polymorphisms to enhance the epidemiologic study of environmental carcinogens. *Biochim. Biophys. Acta* 1471, C1–10

6 Shavers, V.L. and Brown, M.L. (2002) Racial and ethnic disparities in the receipt of cancer treatment. *J. Natl. Cancer Inst.* 94, 334–357

7 Fraumeni, J.F., Jr and Glass, A.G. (1970) Rarity of Ewing's sarcoma among U.S. Negro children. *Lancet* 1, 366–367

8 Watters, J.W. and McLeod, H.L. (2003) Cancer pharmacogenomics: current and future applications. *Biochim. Biophys. Acta* 1603, 99–111

9 Roses, A.D. (2002) Genome-based pharmacogenetics and the pharmaceutical industry. *Nat. Rev. Drug Discov.* 1, 541–549

10 Collins, F.S. *et al.* (2003) A vision for the future of genomics research. *Nature* 422, 835–847

11 Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921

12 Venter, J.C. *et al.* (2001) The sequence of the human genome. *Science* 291, 1304–1351

13 Botstein, D. and Risch, N. (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.* 33(Suppl), 228–237

14 Friend, S.H. *et al.* (1986) A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* 323, 643–646

15 Malkin, D. *et al.* (1992) Germline mutations of the p53 tumor-suppressor gene in children and young adults with second malignant neoplasms. *N. Engl. J. Med.* 326, 1309–1315

16 Frebourg, T. *et al.* (1992) Germ-line mutations of the p53 tumor suppressor gene in patients with high risk for cancer inactivate the p53 protein. *Proc. Natl. Acad. Sci. U. S. A.* 89, 6413–6417

17 Kruglyak, L. (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* 22, 139–144
18 Carlson, C.S. *et al.* (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am. J. Hum. Genet.* 74, 106–120
19 Marth, G. *et al.* (2003) Sequence variations in the public human genome data reflect a bottlenecked population history. *Proc. Natl. Acad. Sci. U. S. A.* 100, 376–381
20 Tishkoff, S.A. and Verrelli, B.C. (2003) Patterns of human genetic diversity: Implications for human evolutionary history and disease. *Annu. Rev. Genomics Hum. Genet.* 4, 293–340
21 Chanock, S. (2001) Candidate genes and single nucleotide polymorphisms (SNPs) in the study of human disease. *Dis. Markers* 17, 89–98
22 Risch, N. (1999) Searching for genetic determinants in the new millenium. *Nature* 405, 847–856
23 Duan, J. *et al.* (2003) Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum. Mol. Genet.* 12, 205–216
24 Packer, B. *et al.* (2004) SNP500cancer: a public resource for sequence validation and assay development for genetic variation in candidate genes. *Nucleic Acids Res.* 32, D528–D532
25 Cambien, F. *et al.* (1999) Sequence diversity in 36 candidate genes for cardiovascular disease. *Am. J. Hum. Genet.* 65, 183–191
26 Reich, D.E. *et al.* (2003) Quality and completeness of SNP databases. *Nat. Genet.* 33, 457–458
27 Daly, M.J. *et al.* (2001) High-resolution haplotype structure in the human genome. *Nat. Genet.* 29, 229–232
28 Gabriel, S.B. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science* 296, 2225–2229
29 Stephens, J.C. *et al.* (2001) Haplotype variation and linkage disequilibrium in 313 human genes. *Science* 293, 489–493
30 Stephens, M. *et al.* (2001) A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68, 978–989
31 Excoffier, L. and Slatkin, M. (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.* 12, 921–927
32 Fallin, D. *et al.* (2001) Genetic analysis of case/control data using estimated haplotype frequencies: application to APOE locus variation and Alzheimer's disease. *Genome Res.* 11, 143–151
33 Thompson, D. *et al.* (2003) Haplotype tagging single nucleotide polymorphisms and association studies. *Hum. Hered.* 56, 48–55
34 Stram, D.O. *et al.* (2003) Modeling and E-M estimation of haplotype-specific relative risk from genotype data for a case-control study of unrelated individuals. *Hum. Hered.* 55, 179–190
35 Chapman, J.M. *et al.* (2003) Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power. *Hum. Hered.* 56, 18–31
36 The International HapMap Consortium (2003) The international HapMap project. *Nature* 426, 789–796
37 Taylor, A. *et al.* (2001) Using genetic variation to study human disease. *Trends Mol. Med.* 7, 507–512
38 Collins, A. *et al.* (1999) Genetic epidemiology of single-nucleotide polymorphisms. *Proc. Natl. Acad. Sci. U. S. A.* 96, 15173–15177
39 Hall, J.M. *et al.* (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 250, 1684–1689
40 Le Marchand, L. *et al.* (2000) Association of the myeloperoxidase -463G→a polymorphism with lung cancer risk. *Cancer Epidemiol. Biomarkers Prev.* 9, 181–184
41 London, S.J. *et al.* (1997) Myeloperoxidase genetic polymorphism and lung cancer risk. *Cancer Res.* 57, 5001–5003
42 Golka, K. *et al.* (2002) The enhanced bladder cancer susceptibility of NAT2 slow acetylators towards aromatic amines: a review considering ethnic differences. *Toxicol. Lett.* 128, 229–241
43 Hein, D.W. (2002) Molecular genetics and function of NAT1 and NAT2: role in aromatic amine metabolism and carcinogenesis. *Mutat. Res.* 506-507, 65–77
44 Marcus, P.M. *et al.* (2000) NAT2 slow acetylation and bladder cancer risk: a meta-analysis of 22 case-control studies conducted in the general population. *Pharmacogenetics* 10, 115–122
45 Hopper, J.L. (2001) Genetic epidemiology of female breast cancer. *Semin. Cancer Biol.* 11, 367–374
46 Carlson, C.S. *et al.* (2003) Additional SNPs and linkage disequilibrium analyses are necessary for whole-genome asssociations tudies in humans. *Nat. Genet.* 33, 518–521
47 Krynetski, E.Y. *et al.* (1996) Genetic polymorphism of thiopurine S-methyltransferase: clinical importance and molecular mechanisms. *Pharmacogenetics* 6, 279–290
48 Krynetski, E.Y. *et al.* (1995) A single point mutation leading to loss of catalytic activity in human thiopurine S-methyltransferase. *Proc. Natl. Acad. Sci. U. S. A.* 92, 949–953
49 McLeod, H.L. *et al.* (2000) Genetic polymorphism of thiopurine methyltransferase and its clinical relevance for childhood acute lymphoblastic leukemia. *Leukemia* 14, 567–572
50 McLeod, H.L. *et al.* (1995) Polymorphic thiopurine methyltransferase in erythrocytes is indicative of activity in leukemic blasts from children with acute lymphoblastic leukemia. *Blood* 85, 1897–1902
51 McLeod, H.L. *et al.* (2000) Genetic polymorphism of thiopurine methyltransferase and its clinical relevance for childhood acute lymphoblastic leukemia. *Leukemia* 14, 567–572
52 Evans, W.E. *et al.* (2001) Preponderance of thiopurine S-methyltransferase deficiency and heterozygosity among patients intolerant to mercaptopurine or azathioprine. *J. Clin. Oncol.* 19, 2293–2301
53 Relling, M.V. *et al.* (1999) Prognostic importance of 6-mercaptopurine dose intensity in acute lymphoblastic leukemia. *Blood* 93, 2817–2823
54 Hon, Y.Y. *et al.* (1999) Polymorphism of the thiopurine S-methyltransferase gene in African-Americans. *Hum. Mol. Genet.* 8, 371–376
55 Schmiegelow, K. *et al.* (2003) Intensification of 0mercaptopurine/methotrexate maintenance chemotherapy may increase the risk of relapse for some children with acute lymphoblastic leukemia. *J. Clin. Oncol.* 21, 1332–1339
56 Gupta, E. *et al.* (1994) fate of irinotecan in humans: correlation of glucuronidation with diarrhea. *Cancer Res.* 54, 3723–3725
57 Innocenti, F. and Ratain, M.J. (2003) Irinotecan treatment in cancer patients with UGT1A1 polymorphisms. *Oncology (Huntingt.)* 17 (Suppl 5), 52–55
58 Iyer, L. *et al.* (1998) Genetic predisposition to the metabolism of irinotecan (CPT-11). Role of uridine diphosphate glucuronosyltransferase isoform 1A1 in the glucuronidation of its active metabolite (SN-38) in human liver microsomes. *J. Clin. Invest.* 101, 847–854
59 Paez, J.G. *et al.* (2004) EGFR mutations in lung cancer: correlation with clinical response to Gefitinib therapy. *Science* 304, 1497–1500
60 Lynch, T.J. *et al.* (2004) Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* 350, 2129–2139
61 Strausberg, R.L. *et al.* (2003) Sequence-based cancer genomics: progress, lessons and opportunities. *Nat. Rev. Genet.* 4, 409–418
62 Iqbal, S. and Lenz, H.J. (2003) Targeted therapy and pharmacogenomic programs. *Cancer* 97(Suppl. 8), 2076–2082
63 Lohmueller, K.E. *et al.* (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat. Genet.* 33, 177–182
64 Chanock, S.J. and Wacholder, S. (2002) One gene and one outcome? No way. *Trends Mol. Med.* 8, 266–269
65 No author (1999) Freely associating. *Nat. Genet.* 22, 1–2
66 Colhoun, H.M. *et al.* (2003) Problems of reporting genetic associations with complex outcomes. *Lancet* 361, 865–872
67 Hirschhorn, J.N. *et al.* (2002) A comprehensive review of genetic association studies. *Genet. Med.* 4, 45–61
68 Ioannidis, J.P. *et al.* (2001) Replication validity of genetic association studies. *Nat. Genet.* 29, 306–309

69 Wacholder, S. *et al.* (2004) Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J. Natl. Cancer Inst.* 96, 434–442
70 Wacholder, S. *et al.* (2000) Population stratification in epidemiologic studies of common genetic variants and cacner:quantification of bias. *J. Natl. Cancer Inst.* 92, 1151–1158
71 Marchini, J. et al. (2004) The effects of human population structure on large genetic association studies. *Nat. Genet.* 36, 512–517
72 Freedman, M.L. *et al.* (2004) Assessing the impact of population stratification on genetic association studies. *Nat. Genet.* 36, 388–393
73 Goldstein, J.A. (2001) Clinical relevance of genetic polymorphisms in the human CYP2C subfamily. *Br. J. Clin. Pharmacol.* 52, 349–355
74 El Omar, E.M. *et al.* (2000) Interleukin-1 polymorphisms associated with increased risk of gastric cancer. *Nature* 404, 398–402
75 El Omar, E.M. *et al.* (2003) Increased risk of noncardia gastric cancer associated with proinflammatory cytokine gene polymorphisms. *Gastroenterology* 124, 1193–1201
76 Stanczuk, G.A. *et al.* (2001) Cancer of the uterine cervix may be significantly associated with a gene polymorphism coding for increased IL-10 production. *Int. J. Cancer* 94, 792–794
77 Cavet, J. *et al.* (1999) Recipient tumor necrosis factor-alpha and interleukin-10 gene polymorphisms associate with early mortality and acute graft-versus-host disease severity in HLA-matched sibling bone marrow transplants. *Blood* 94, 3941–3946
78 Kuschel, B. *et al.* (2002) Variants in DNA double-strand break repair genes and breast cancer susceptibility. *Hum. Mol. Genet.* 11, 1399–1407
79 Diasio, R.B. *et al.* (1988) Familial deficiency of dihydropyrimidine dehydrogenase. Biochemical basis for familial pyrimidinemia and severe 5-fluorouracil-induced toxicity. *J. Clin. Invest.* 81, 47–51
80 Lu, Z. *et al.* (1993) Dihydropyrimidine dehydrogenase activity in human peripheral blood mononuclear cells and liver: population characteristics, newly identified deficient patients, and clinical implication in 5-fluorouracil chemotherapy. *Cancer Res.* 53, 5433–5438