Genetic Variations as Cancer Prognostic Markers: Review and Update

Sevtap Savas* and Geoffrey Liu1,2
1Department of Medical Biophysics, Division of Applied Molecular Oncology, Ontario Cancer Institute, Toronto, Ontario, Canada; 2Department of Medicine, Division of Medical Oncology and Hematology, Princess Margaret Hospital, Toronto, Ontario, Canada

Communicated by David E. Goldgar
Received 9 March 2009; accepted revised manuscript 11 June 2009.
Published online 26 June 2009 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/humu.21078

ABSTRACT: Cancer molecular epidemiology traditionally studies the relationship between genetic variations and cancer risk. However, recent studies have also focused on disease outcomes. The application and design of disease outcome studies have been an extension of disease risk assessment. Yet there are a number of unique considerations important in outcome assessments. We review how genetic approaches used for disease susceptibility, such as candidate gene and genome-wide association study (GWAS) approaches, can be adapted carefully to systematically identify cancer prognostic and predictive alleles. We discuss the interrelatedness among the disease susceptibility, treatment response, and prognosis at the genetic level and focus on how the emerging technologies and approaches can uniquely benefit the genetic prognosis studies.

KEY WORDS: cancer; prognosis; susceptibility; treatment response; genetic variation; genome-wide association studies; candidate gene approach

Introduction

Methodological approaches to the study of the genetic basis of cancer susceptibility have been developed for several decades. Recently, similar efforts have been undertaken to identify the genetic basis of variable treatment response and cancer prognosis. The primary aim in identification of these genetic factors is to translate this knowledge into the clinic, where appropriate and feasible treatments are used for patients based on genetic makeup (i.e., personalized medicine).

Cancer predisposition, treatment response, and prognosis are interrelated at the genetic level (Fig. 1, Box 1), and, thus, genetic studies applied to study the genetic predisposition to cancer may be adapted to study the genetics of cancer outcome, including pharmacogenetics (genetics of treatment response). Yet merely applying the risk methodologies is simplistic at best and misleading at worst. This review details the interactions of the genetic basis of cancer susceptibility, treatment response, and prognosis, and how to adapt established methodologies to genetic prognostic studies.

Prognostic Risk Factors

Understanding the difference between risk and outcome studies requires a clear definition of prognosis. The National Cancer Institute defines prognosis as “giving an idea of the likely course and outcome of a disease” (www.cancer.gov/cancertopics/factsheet/support/prognosis-stats). Therefore, prognosis provides a forecast on chances of recovery from cancer or its recurrence, irrespective of treatment. It is likely that numerous social, psychological, biological, genetic, and environmental factors work together to determine prognosis in patients, and their identification has important implications in clinical practice. Among the well-known prognostic factors in cancer are the tumor-related and morphological factors (e.g., type, location, histology, grade, stage, tumor extent, lymph-node status, and venous invasion), demographic factors (e.g., patient age, sex, general health, socioeconomic status, and access to health care), psychological factors (e.g., level of social support and effective ways to handle impacts of the disease), clinical factors (e.g., performance status, metastatic disease, and disease symptoms), lifestyle factors (e.g., diet, exercise, and smoking), and molecular factors (e.g., somatic and inherited genetic variations, epigenetic variations, protein markers, and cytogenetic abnormalities) [Galsky and Vogelzang, 2007; Hallek and German CLL Study Group, 2008; Zlobec and Lugli, 2008; Thomas and Davies, 2007; Bouchardy et al., 2006].

In clinical practice, prognostic factors may allow stratification of patients into groups with distinct outcomes [Galsky and Vogelzang, 2007]. For example, patients with lymph node invasion or distant metastasis are predicted to have poor outcome when compared to patients with localized cancer. Patients grouped into worse prognosis cohorts may then be treated with more aggressive treatment regimens, such as combination therapies. Thus, prognostic factors are also critical in planning treatment [Galsky and Vogelzang, 2007; Hallek and German CLL Study Group, 2008]. Moreover, genetic factors that alter the treatment response may also affect the disease prognosis and outcome. Inherited genetic factors are thus reasonable candidates as potential cancer prognostic factors. Distinct from prognostic factors are predictive factors, in which case the genetic variant modifies a specific treatment response and outcome. Thus a predictive factor is not useful in the absence of the specific treatment, while a prognostic factor is independently associated with outcome regardless of treatment.

© 2009 WILEY-LISS, INC.
Treatment Response, and Prognosis

Cancer development

![Figure 1](image)

Figure 1. Interrelatedness of the biology behind the predisposition, treatment (drug and/or radiotherapy) response, and outcome in cancer. Genetic risk factors that predispose individuals to develop cancer can also contribute to variable treatment response. Additionally, genetic risk factors can also determine the prognosis-outcome. Treatment response also can determine the outcome.

Box 1. Parallels Between Disease Susceptibility, Variable Treatment Response, and Prognosis

(a) The variable treatment response and prognosis among patients is also attributable to a number of low-penetrant alleles.
(b) Both somatic and inherited genetic variations have a role in variable prognosis and outcome.
(c) Disease susceptibility and treatment response genes may have direct roles in outcome.
(d) Candidate gene, candidate pathway and GWAS approaches are and can be applied in outcome analysis.
(e) Like disease risk and treatment response, prognosis and outcome is likely to have a multifactorial nature characterized by dynamic interactions between genetic and environmental factors.

Common Elements Among Genetic Susceptibility, Treatment Response, And Prognosis: The Basis For Applying Similar Methodologies

Role of Tumor Somatic Mutations in Carcinogenesis, Treatment Response, and Prognosis

Carcinogenesis is characterized by clonal selection of cells in which somatic mutations that confer growth advantage are positively selected [Canerari et al., 2006; Hanahan and Weinberg, 2000], and in some cases even lead to yet more mutations (mutator phenotype) [Loeb, 1991]. Acquired somatic changes of the genetic material have been studied extensively as biological prognostic factors in cancer [Mitsudomi and Yatabe, 2007; Haber et al., 2005]. Recent large-scale efforts in identification of somatic mutations [Forbes et al., 2008] will substantially increase our understanding of somatic driver mutations in cancer predisposition [Greenman et al., 2007; Ding et al., 2008; Yeang et al., 2008] as well as in treatment response [Mitsudomi and Yatabe, 2007; Haber et al., 2005]. Though somatic mutations represent a key factor in driving the cancer risk, treatment response, and prognosis, focus in this review is the effects of inherited genetic factors. Additionally, other factors, such as epigenetic alterations and loss-of-heterozygosity, that are frequently observed in cancer and are potential confounders of the association between inherited genetic polymorphisms and prognosis, are also out of scope of this review article.

Low-Penetrance Is Common to Risk, Treatment Response, and Prognosis

Familial or inherited high-penetrant genes, such as TP53 (MIM # 191170) in the case of Li-Fraumeni syndrome (MIM # 151623) and BRCA1 (MIM # 113705) and BRCA2 (MIM # 600185) in familial breast cancer (MIM # 114480) [Miki et al., 1994; Taftvigian et al., 1996; Malkin et al., 1990], account for only a portion of cancer cases. In the remaining sporadic cases, low-to-moderate penetrant alleles, in combination with other genetic, epigenetic, and environmental factors, predispose individuals to cancer [Chakravarti, 1999]. Low-penetrance alleles are characterized by modest relative risk and are not sufficient alone to lead to cancer. Therefore, cancer is a polygenic disease and many low-penetrance alleles with additive effects are needed to develop it.

Low-penetrance may also be a common genetic feature in treatment response and cancer prognosis. Radiation and a large array of drugs are utilized to treat cancer, albeit with certain rates of resistance and serious side effects, which are attributable to both environmental and low-penetrant genetic factors [Bartsch et al., 2007; Madhusudan and Middleton, 2005]. Findings from previous prognosis and outcome analyses also suggest the presence of several modest-effect, low-penetrance alleles in determining certain phenotypes [Cerhan et al., 2007; Habuchi, 2006; Hsieh et al., 2006; Hunter et al., 2003; Yasui et al., 2005; Zhou et al., 2006]. For example, in pharmacogenetics of 5-Fluorouracil treatment response, the germline mutations in the dihydropyrimidine dehydrogenase gene, DPYD (MIM # 274270), cause severe toxicity, yet these germline mutations are extremely rare in populations [Yamaguchi et al., 2001]. Thus, low-penetrant alleles may also explain some of the variable treatment response and outcome in cancer.

A low-penetrant allele has a biological role in human phenotypes due to its direct or indirect effects on functional genetic units, such as coding and noncoding genes, by means of influencing their expression or function. These alleles can be any kind of genetic variation in the form of single-nucleotide polymorphisms (SNPs), copy-number variations (CNVs; i.e., variable copies of up to 1-million-base-long genomic regions), small insertions and deletions, and other genomic rearrangements. In cancer research, SNPs have been mostly utilized to identify disease genes and alleles; however, an increasing number of studies are now also focusing on the role of CNVs in complex phenotypes [McCarroll, 2008] (see CNVs section below).

Cancer Predisposition, Treatment Response, and Outcome Have Common Biology

Cancer-related phenotypes, namely cancer predisposition, treatment response, and outcome, are likely to share certain biological characteristics.

Molecular changes that drive carcinogenesis have been clearly outlined [Canerari et al., 2006; Hanahan and Weinberg, 2000]. In brief: (1) cells are no longer dependent to growth signals for growth; (2) cells no longer respond to growth inhibition mechanisms/signals; (3) apoptotic mechanisms are inhibited; (4) cells develop the potential for unlimited growth and division; (5) angiogenesis is activated; (6) tissue invasion and metastasis become possible; (7) abnormal tissue repair and stem cell renewal is observed; and (8) metabolic alterations for better tumor survival develop. The same features may also drive a biologic response to therapy and aggressiveness of disease, thus affecting the cancer prognosis.

Supporting this hypothesis, there is evidence that one gene can be biologically linked to either individual phenotypes
(predisposition, treatment response, and outcome) or their combinations (Table 1). Examples in literature include: TYMS (MIM 188350) [Lurje et al., in press; Ulrich et al., 2005], MTHFR (MIM 607093) [Capitain et al., 2008; Kono and Chen, 2005], and TP53 (MIM 191170) [de Jong et al., 2002; Munro et al., 2005], which are associated with predisposition, treatment response, and outcome in colorectal cancer; BRCA1 (MIM 113705), which is associated with both predisposition and treatment response in breast cancer [Chen et al., 2006]; ABCB1 (MIM 171050), which is associated with treatment response and outcome in lung cancer [Bartsch et al., 2007] and with predisposition and prognosis in hematological cancers [Jamroziak and Robak, 2004]; MDM2 (MIM 164785), which is associated with both susceptibility and outcome in pancreatic cancer [Asomaning et al., 2008]; and drug metabolism genes that affect treatment response and prognosis [Yang et al., 2006]. Thus, some genes may underlie one, two, or all of the three phenotypes of genetic predisposition, treatment response, and overall outcome in cancer [Spitz et al., 2005].

Figure 3 depicts several models we propose for this concept of “integrative epidemiology” [Spitz et al., 2005], showing variation and overlap between therapy, risk, and prognosis. Model 1 assumes that certain groups of genes can have roles in all three phenotypes, in two phenotypes, or in one specific phenotype only. Model 2 assumes that there is an overlap between predisposition and treatment response and the genes in predisposition and treatment response collectively determine the outcome. Well-known examples in this category include many xenobiotic metabolism and DNA repair genes that are involved in response to environmental factors as well as chemotherapeutic agents/radiotherapy. These genes can also explain Model 3. Model 3 assumes that there is an overlap between predisposition and treatment response genes; this group of genes are part of the outcome-related genes. In this model, the relationship of predisposition with outcome is restricted with the genes determining the treatment response. Model 4 assumes that in the absence of intervention by treatment, the genetic predisposition factors are also prognostic factors. Examples in this category are not likely to be present in scientific literature, as almost all studies investigate patient cohorts with some type of medical intervention (surgery, chemotherapy, or radiotherapy). In this model, it is also not clear whether there are prognosis-outcome-specific genes (Model 4a) or not (Model 4b).

Genes or genetic variations that can ease, promote, or contribute to cellular oncologic hallmarks are candidate genetic factors for cancer predisposition [Canevari et al., 2006] and for other oncologic phenotypes. For example, the individual carrying

### Table 1. Examples of Genes That Can Affect One or More of P, TR, and O in Cancer

<table>
<thead>
<tr>
<th>Known examples</th>
<th>Cancer site</th>
<th>Gene’s role in cancer</th>
<th>Applicable model in Figure 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>TYMS (MIM 188350)</td>
<td>Colorectal cancer</td>
<td>Associated with P, TR, and O [Lurje et al., in press; Ulrich et al., 2005]</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>MTHFR (MIM 607093)</td>
<td>Colorectal cancer</td>
<td>Associated with P, TR, and O [Capitain et al., 2008; Kono and Chen, 2005]</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>TP53 (MIM 191170)</td>
<td>Colorectal cancer</td>
<td>Associated with P, TR, and O [de Jong et al., 2002; Munro et al., 2005]</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>BRCA1 (MIM 113705)</td>
<td>Breast cancer</td>
<td>Associated with P and TR [Chen et al., 2006]</td>
<td>1, 2, 3</td>
</tr>
<tr>
<td>ABCB1 (MIM 171050)</td>
<td>Lung cancer; hematological cancers</td>
<td>Associated with TR and O [Bartsch et al., 2007]; associated with P and O [Jamroziak and Robak, 2004]</td>
<td>1, 2, 3, 1, 2, 3, 4a, 4b</td>
</tr>
<tr>
<td>MDM2 (MIM 164785)</td>
<td>Pancreatic cancer</td>
<td>Associated with P and O [Asomaning et al., 2008]</td>
<td>1, 2, 3, 4a, 4b</td>
</tr>
<tr>
<td>Genes in glutathione metabolic pathway</td>
<td>Lung cancer</td>
<td>Associated with TR and O [Yang et al., 2006]</td>
<td>1, 2, 3</td>
</tr>
</tbody>
</table>

P, predisposition; TR, treatment response; O, outcome/prognosis.
an angiogenesis gene variant that promotes tumor formation may also be more likely to have a more aggressive type of tumor when compared to the other individual who does not carry it. Such genetic factors can also be directly involved in therapy response and outcome [Canevari et al., 2006]. Genes involved in oncologic phenotypes, such as the DNA repair, mutagen metabolism, antioxidant mechanisms, and gatekeeper genes, are among the top candidates that affect all these clinical events.

**Adapting Disease Susceptibility Principles To Prognosis And Drug Response Studies**

Gene expression signatures [Zlobec and Lugli, 2008; van de Vijver et al., 2002], somatic cytogenetic and molecular lesions such as loss-of-heterozygosity, microsatellite instability, and mutational status of oncologic genes [Hallek and German CLL Study Group, 2008; Greenman et al., 2007], and now the inherited genetic variations, are being assessed as predictors of cancer outcome. Most of these studies on genetic variations have used candidate gene approaches. Among the well-studied candidate genes in prognosis are those that are involved in cell growth, angiogenesis, tissue invasion and metastasis, and immune system [Cerhan et al., 2007; Habuchi, 2006; Hsieh et al., 2006; Asomaning et al., 2008]. Van Ness et al. [2008] applied a pathway-based chip assay consisting of ~1,000 genes to identify novel loci associated with early clinical relapse in multiple myeloma by investigating the role of a large number of germline genetic polymorphisms in cancer prognosis. More, large-scale systematic studies on prognosis and outcome are needed by adaptive methodologies initially developed for cancer susceptibility evaluation.

There are, however, study design caveats in cancer prognostic studies. Prognostic studies, unlike risk studies, depend on having a well-described disease phenotype state. In cancer, this means that the major clinical prognostic variables, such as stage, presence of comorbidities, treatments, and treatment strategies, are uniform among patients. Further, disease outcome and progression, in addition to overall survival, should follow standard clinical guidelines, as should treatment and toxicity variables. In addition to considerations in study design, analyses should include all potential clinical and treatment variables. Specific diseases will have unique prognostic variables (e.g., *KRAS*; MIM# 190070 mutations in colorectal cancer) [Russo et al., 2005]. Therefore, a major recommendation is to conduct all prognostic studies with a clinician consultant familiar with clinical trials.

Below we discuss specific considerations for genetic analysis in cancer prognostic studies.

**Candidate Gene Approach**

This traditional approach is easily applicable to outcome studies. In a “functional candidate gene approach,” genes that are involved in carcinogenesis-related molecular processes, such as DNA repair, oxidative stress, cell cycle, apoptosis, and signal transduction, are the prime candidates as cancer susceptibility genes. Moreover, the genes that are somatically mutated (deleted or amplified in tumors) are the “positional candidate genes” that we often investigate in cancer susceptibility studies [Imreh et al., 2003; Oldenburg et al., 2007]. Utilization of such approaches has elucidated many cancer susceptibility alleles: *XRCC3* (MIM# 600675) [Han et al., 2006; Manuguerra et al., 2006] and *ERCC2* (MIM# 126340) (Manuguerra et al., 2006) (two DNA repair genes associated with risks of several cancer sites); the polymorphisms from *CYP1B1* gene (MIM# 601771) (involved in estrogen metabolism) [Paracchini et al., 2007]; *MTHFR* (MIM# 607093) (a folate metabolism gene) [Macis et al., 2007]; and *AKAP9* (MIM# 604001) (a signal transduction gene) [Frank et al., 2008], all of which are associated with breast cancer risk. These candidates are also all reasonable prognostic factors to evaluate.
Pharmacogenetics is the field that investigates the genetic basis of interindividual variability in drug responses [Katz, 2006]. Most pharmacogenetics studies have focused on candidate gene approaches, and among the genes extensively studied are the ones involved in the pharmacokinetics (PK) and the pharmacodynamics (PD) of the drug. Biological PK processes underlie the drug absorption, metabolism, transportation, and excretions, and examples include the cytochrome p450 (CYP) family members and the drug transporters, such as the ABC class of transporters [Bosch, 2008; Rochat, 2005]. PD, on the other hand, refers to the mechanism of action of the drug in the body (such as induction of apoptosis, halting DNA synthesis, etc.). For example, gemcitabine halts the DNA synthesis, resulting in apoptosis [Hilbig and Oettle, 2008], whereas taxanes disrupt the microtubules, resulting in arrest of mitosis [Ganansia-Leymarie et al., 2003]. In the case of variable radiation therapy (RT) response, the best candidate genes are the DNA repair genes, which are implicated in both RT efficacy (damage in the tumor) and the toxic side effects (damage in the tumor-surrounding tissue) [Gossage and Madhusudan, 2007; Carles et al., 2006; Damaraju et al., 2006]. Thus the candidate gene approach can be useful to assess pharmacogenetic pathways (treatment response) where the focus is not only on toxicity but also on drug efficacy, in addition to the toxigenomics of risk.

**Candidate Pathway Approach**

Although the candidate gene approach is highly useful, it lacks the comprehensive and systematic nature of the candidate pathway approach. A candidate pathway approach identifies all the genes important to a biological or pharmacogenetic pathway, and then selects for polymorphisms of this gene set. The need for a pathway-based approach in molecular epidemiology risk studies was previously argued [Thomas, 2005]. Luckily, the availability of increasing amounts of biological and molecular information, progress in bioinformatic analytical tools and databases, and reduction in genotyping cost have all helped researchers to switch to the candidate pathway approach. This approach has been utilized in cancer research where, for example, the role of angiogenesis [Schneider et al., 2008] and estrogen-metabolism pathways [Justenhoven et al., 2008] in breast cancer risk, and the role of DNA repair pathway in colon cancer risk [Schaafmayer et al., 2007] were investigated. Adaptations can lead to pathway analyses of outcome in cancer [Wu et al., 2006].

A major caveat of applying candidate gene and candidate pathway approaches to cancer prognosis studies is that cancer phenotypes are often more complicated than cancer risk. Toxicity phenotype depends on cumulative dosing, comorbidities, and patient’s tolerance to pain, and is, therefore, semiquantitative. Time to recurrence can depend on location of tumor, patient symptoms, and frequency of routine patient follow-up. Thus studies involving large-scale pathway analyses and genome-wide association studies (GWASs) should be developed around disease states and patient populations with clear outcome phenotype.

Lately, development of new generation, cost-effective high-throughput technologies, offered by different commercial entities such as Affymetrix (Santa Clara, CA; www.affymetrix.com) and Illumina (San Diego, CA: www.illumina.com), has opened the way for the whole-genome scans to identify low-penetrant alleles.

**GWASs**

GWASs are association studies in which large numbers of cases and controls are genotyped for dense genetic markers covering the genome, usually in the form of SNPs and the CNVs. In GWASs of risk, statistically significant differences in allele and genotype frequencies between the case and control groups point to potential candidate genomic regions harboring the susceptibility alleles. However, the same approach can be applied to treatment response and clinical outcomes. GWASs benefited mostly from the results of HapMap project, which genotyped more than 4 million common SNPs in human populations from Asian, African, and Caucasian samples [International HapMap Consortium et al., 2007]. The results of the HapMap project also identified the linkage disequilibrium structure of human genome in different populations and this information helped determining tagSNPs as well as population-specific sets of SNPs needed in GWAS analysis. Currently, GWAS genotyping platforms for African, Asian, and Caucasian populations with different genome coverage are offered by both Affymetrix and Illumina [Motsinger et al., 2006].

GWASs have their own advantages and disadvantages [Motsinger et al., 2006]. The ultimate advantage of this approach is that, in contrast to candidate gene approach, it does not require a priori knowledge on the genes/genomic regions to be investigated for the susceptibility/outcome alleles. Therefore, novel and previously unidentified susceptibility genes can be identified by GWAS. However, GWAS design and data analysis require multiple considerations, such as sample size issue, level of statistical significance, correction for multiple testing, population stratification, marker density, and replication of the results by independent studies [Hemminki and Bernas, 2007; Miyagawa et al., 2008; Ziegler et al., 2008]. Sample size and availability of independent datasets of similar stage and therapy is a major limitation of its applicability to prognostic studies. In addition, this approach is designed to detect common variants with modest biological effects [Chakravarti, 1999], and is therefore likely to miss the rare associated alleles [Reich and Lander, 2001]. In risk, so far more than 100 cancer susceptibility alleles have been identified using GWAS [Chanock and Hunter, 2008], including the cancers of prostate (MIM# 176807) [Thomas et al., 2008], lung (MIM# 211980) [Hung et al., 2008], and breast (MIM# 114480) [Easton et al., 2007]. In contrast, there are a few GWASs used in identification of drug response genes. For example, Huang et al. [2008] combined and correlated the GWAS data on the HapMap samples with that of gene expression and cell toxicity assays to identify genes involved in toxicity to the chemotherapeutic agent daunorubicin. We expect that GWASs will be utilized soon to identify novel genes associated with variable drug and radiation response, and general prognosis in cancer patients.

Caveats of GWASs in prognostic studies stem from inability to recruit adequate number of patients for these studies. If we assume 100 patients with breast cancer, perhaps only 40 are of a specific disease stage relevant to the question, and within this group only 25 are treated uniformly. Although we would be able to utilize all 100 patients for a GWAS of risk, only 25 could be used for GWAS in outcome. Alternatively, so many important prognostic variables will need adjustment in the analyses if all 100 patients were included that the genetic variant relationship with outcome would be compromised.

**Emerging Concepts In Cancer Research**

**CNVs**

CNVs are the newly-appreciated structural genomic variations that involve duplication or deletion of genomic segments encompassing ~1 kilobase (kb) up to millions of bases (Mb) [McCarroll, 2008]. Preliminary estimates on the number of CNVs are around 11,000 [Ionita-Laza et al., 2009]; however, as better
technological approaches are undertaken, this number is likely to increase. CNVs can be either inherited or formed de novo by new mutations and can be presented in populations as rare or common variations. Potential biological impacts of CNVs may exceed that of the SNPs [Tuzun et al., 2005]. In addition, CNVs usually contain full or partial gene sequences and thus may disrupt gene structure, function, and expression. Therefore, CNVs are highly likely to contribute to phenotypic variation in humans. This potentially pathologic feature of the CNVs makes them exciting candidates as genetic factors in complex disease susceptibility, treatment response, and prognosis studies. In fact, CNVs of certain drug metabolism genes, such as CYP2D6 (MIM # 124030), are already known to have a role in variable drug response [Ingelman-Sundberg et al., 2007].

Initial detection of CNVs was based on hybridization techniques using the fosmid and bacterial artificial chromosome (BAC) clones as probes in array-based comparative genomic hybridization (CGH) experiments [Tuzun et al., 2005; Carter, 2007]. The resolution and specificity achieved by this approach is, however, low; thus, it has been replaced by other techniques, including one that utilizes long oligonucleotide probes in CGH reactions, which seems to have better sensitivity and resolution than the earlier approach [Ionita-Laza et al., 2009; Carter, 2007]. A third technique in CNV detection uses the SNP arrays. In this case, the CNVs are detected either by means of the intensity of the signal obtained in the genotyping experiment, or by integrating CNV-detecting probes into the platform [Ionita-Laza et al., 2009; Carter, 2007; Sharp, 2009]. Currently, some of the GWAS platforms offer both SNP and CNV detection and genotyping; therefore, it is possible to investigate the role of these structural variants in human health and diseases genome-wide.

As in the case of SNPs, GWAS CNV analysis has its own challenges as well, such as the current statistical and technical issues [Ionita-Laza et al., 2009; Sharp, 2009], as well as the probably currently incomplete list of the CNVs that can be analyzed this way. Nevertheless, a few numbers of large-scale studies has already shown that both inherited and de novo as well as common and rare CNVs are associated with a variety of human conditions including an inherited cancer syndrome [McCarroll, 2008; Ionita-Laza et al., 2009]. As more information becomes available on the content and structure of the CNVs, this will doubtlessly benefit the complex phenotype studies, including the ones involved in the prognosis and outcome of cancer.

Next-Generation Sequencing

The recently developed next-generation sequencing (NGS) technologies offer sequencing of colossal genomic sequences in a cost-effective way and thus represent a hallmark in genomic research [Shendure and Ji, 2008]. Briefly, NGS involves creation of DNA fragments, which are then amplified and/or sequenced following sophisticated methods, such as pyrosequencing and ligation-based sequencings, in a massively parallel manner [Mardis, 2008; Voelkerding et al., 2009]. The sequence reads are then assembled. Currently, a few different NGS platforms, which differ in their sequencing chemistry and overall approach, are offered by commercial companies [Mardis, 2008; Voelkerding et al., 2009].

A wide array of applications is possible using the NGS technologies [Shendure and Ji, 2008; Voelkerding et al., 2009], yet two of them are particularly relevant to genomics studies in treatment response and prognosis. One such application is the sequencing of the complete individual genome, as exemplified by the genome sequences of the three individuals published recently [Wang et al., 2008; Wheeler et al., 2008; Levy et al., 2007]. Also, in a breakthrough study, Ley et al. [2008] recently reported sequencing and comparison of tumor and normal DNA of an individual affected by acute myeloid leukemia, which revealed novel genes and somatic mutations in etiology of this disease. The results obtained from these studies significantly contribute to our understanding of the human genome sequence variations (polymorphisms and mutations alike), though cost will likely prohibit large-scale sample analysis.

A more feasible application of NGS is the targeted resequencing approach, in which specific genomic regions are resequenced to elucidate the entire spectrum of the genetic variations [Voelkerding et al., 2009]. An exciting example undertaken lately in targeted resequencing is the resequencing of 623 cancer-related genes in lung adenocarcinoma, which elucidated the patterns of somatic mutations as well as the genes and pathways somatically mutated in this type of cancer [Ding et al., 2008]. This important information then can be used to develop to strategies for effective therapeutic intervention.

Gene–Gene And Gene–Environment Interactions In Cancer

Cancer susceptibility and variable treatment response and prognosis among patients are likely to be multifactorial in nature, determined by multiple genetic and environmental factors. Therefore, in cancer risk, variable treatment response, and prognosis, while the identification of individual low-penetrant alleles are important (through studying the relationship of one genetic polymorphism with the phenotype), even more important is the identification of the gene–gene and gene–environment interactions (through studying the relationship of more than one genetic polymorphisms/environmental factors with the phenotype) that can help explain the complete etiology of these phenotypes [Schmidt, 2007; Moore, 2003].

The gene–environment interaction refers to the modification of the impacts of the environmental factors by genetic factors in presentation of a phenotype. Perhaps the best well-known gene–environment interaction is observed in the cases of tobacco-related cancers, such as lung cancer, where carcinogenic impacts of exposure to cigarette smoking are modified by polymorphisms in the carcinogen metabolism genes [Taioli, 2008]. Similarly, the gene–gene interaction refers to the interaction between two or more genes at different loci, where the phenotypic expression of one variation is either masked or reinforced by the other. Several examples in literature highlight the gene–gene interactions in etiology of cancer, such as xenobiotic genes in lung cancer [Vineis et al., 2007], DNA repair genes in bladder cancer [Andrew et al., 2008], and the alcohol metabolizing genes in head and neck cancer [Hiraki et al., 2007].

Statistical analyses that investigate various genetic variations/environmental factors in relation to a phenotype can reveal gene–gene and gene–environment interactions in cancer predisposition, treatment response, and prognosis. However, the multiple testing problem generated by inclusion of many different genetic variations and/or environmental factors together in the statistical test is a challenge. Luckily, a variety of statistical approaches have been developed to overcome this problem, such as logistic regression and nonparametric multifactor dimensionality reduction methods [Brilhais et al., 2007; Heidema et al., 2006]. These methods are characterized by certain strengths and
null


