

Genetic association studies in cancer: Good, bad or no longer ugly?

Sharon A. Savage¹ and Stephen J. Chanock^{2,3*}

¹ Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

² Pediatric Oncology Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

³ Core Genotyping Facility, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

* Correspondence to: Tel: +1 301 435 7559; Fax: +1 301 402 3134; E-mail: chanocks@mail.nih.gov

Date received (in revised form): 11th April 2006

Abstract

For some time, investigators have appreciated that genetic association studies in cancer are complex because of the multi-stage process of cancer and the daunting challenge of analysing genetic variants in population and family studies. Because of recent technological advances and annotation of common genetic variation in the human genome, it is now possible for investigators to study genetic variation and cancer risk in many different settings. While these studies hold great promise for unravelling multiple genetic risk factors that contribute to the set of complex diseases called cancer, it is also imperative that study design and methods of interpretation be carefully considered. Replication of results in sufficiently large, well-powered studies is critical if genetic variation is to realise the promise of personalised medicine — namely, using genetic data to individualise medical decisions. In this regard, the plausibility of validated genetic variants can only be realised by the study of gene–gene and gene–environment interactions. The genetic association study in cancer has come a long way from the days of restriction fragment length polymorphisms, and now promises to scan an entire genome ‘agnostically’ in search of genetic markers for a disease or outcome. Moreover, the application and interpretation of these studies should be conducted cautiously.

Keywords: single nucleotide polymorphism, haplotype, association study, genome-wide scan, linkage disequilibrium, cancer risk

Introduction

The promise of analysing common germ-line genetic variation and cancer risk has been accelerated by knowledge gained from annotating the draft sequence of the human genome. Genetic variation in different populations can now be used to search for genetic markers that associate with cancer risk, therapeutic response and outcome. This new paradigm, the study of complex diseases, such as cancer, by the analysis of common genetic variation represents the first step in surveying the genome comprehensively. It is also known that the differences between individual human genomes additionally includes other types of variation, such as microsatellite markers, insertions and deletions (from a single base to large regions of thousands of bases) and copy number variation, but, nevertheless, the first large-scale maps have been generated for single nucleotide polymorphisms (SNPs).^{1,2} Since most common SNPs (with a minor allele frequency greater than 5 per cent in a studied population) are silent and have no apparent function, currently, testing for SNPs is directed at identifying markers of disease risk or outcome.^{3–5} There is a

subset of SNPs, however, that have functional consequences which can result in a subtle change in gene function, such as alteration of a transcription factor binding site in the promoter of a gene or in the coding sequence of a gene product.

One of the first major steps towards identifying the common SNPs for study was the establishing of the International HapMap Project, which has developed a fine-scale haplotype map of the human genome.⁶ This project has genotyped more than 2.6 million SNPs in three distinct continental populations. In parallel, other initiatives have begun to sequence genes of great biological interest in search of common and uncommon SNPs. Sequence verification, while slower and more costly, has provided important insights into the spectrum of common and uncommon single-base nucleotide substitutions in the genome. For example, the National Cancer Institute SNP500 Cancer project is validating SNPs in genes implicated in cancer biology (<http://snp500cancer.nci.nih.gov>),⁷ and the National Heart, Lung and Blood Institute's Seattle SNPs project (<http://pga.mbt.washington.edu/>) has focused on candidate genes and pathways that underlie the inflammatory response.

Genotyping technology has advanced significantly and it is now possible to genotype hundreds of thousands of SNPs in accurate, high-throughput platforms at lower prices.^{8,9} In fact, there are commercial products available for interrogation of common genetic variation across the ‘whole genome’, utilising a strategy of surrogacy testing. Based on the HapMap Phase 2 data,⁶ it is possible to take advantage of linkage disequilibrium across the genome by choosing a set of tagging SNPs as markers of genetic variation across the human genome. It is estimated that with this approach, at least 500,000 SNPs would be required to survey common genetic variation.^{10,11} This significant expansion of knowledge of normal human genetic variation, together with technical advances, has created an opportunity to interrogate the genetic basis of cancer risk, response to therapy and outcome. There are many issues in study design and analysis that must carefully be considered.

The complexities of genetic association studies in cancer

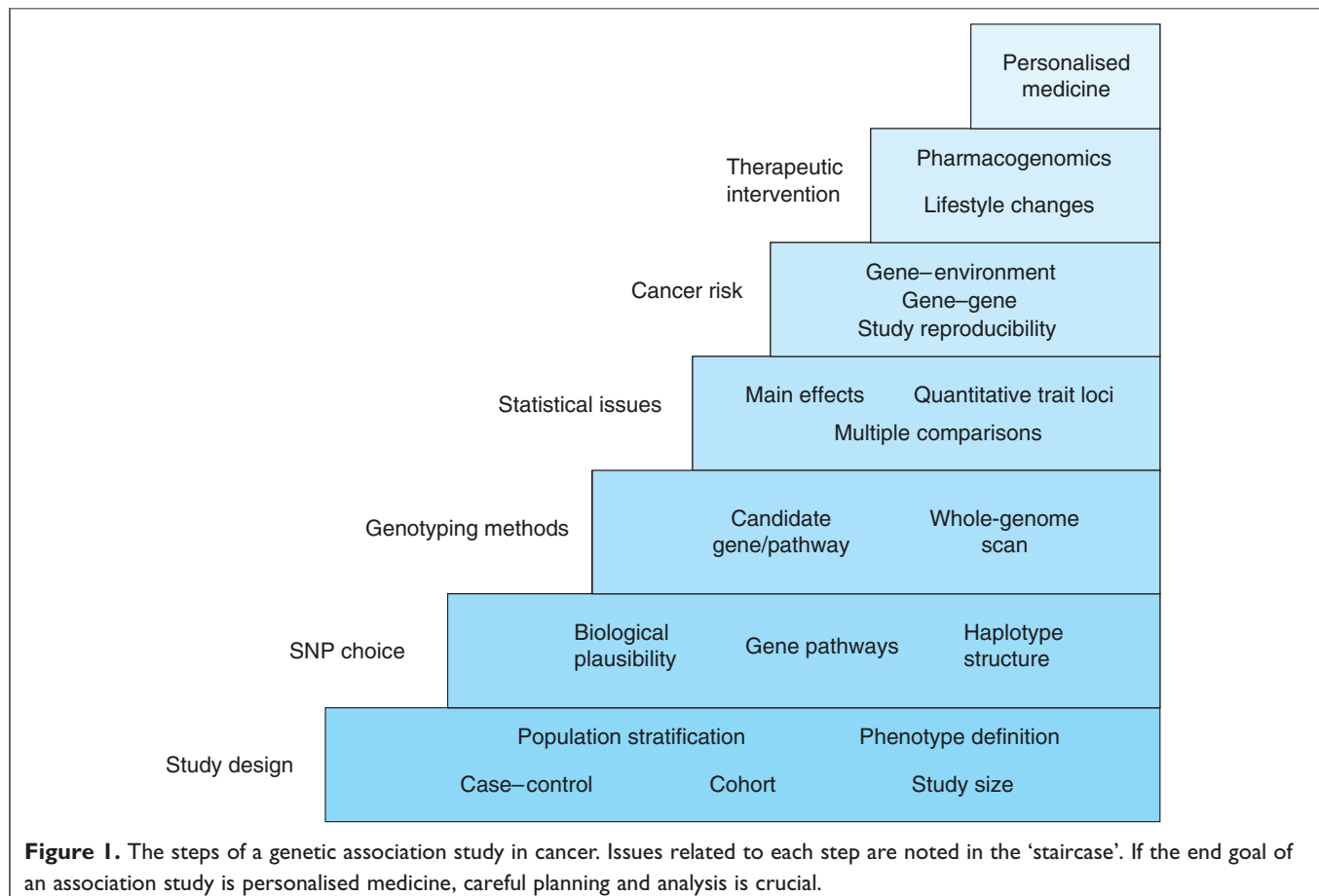
The study of genetic variation and its contribution to cancer risk is a daunting undertaking because of the need to combine

large population-based studies with dense genetic analyses. Figure 1 shows many of the steps to consider in designing and interpreting a genetic association study in cancer.

Although the complexity of cancer as a disease has been described by others, the interaction between genes and the environment has not yet been explored in detail.^{12,13} In any one type of cancer, there are often significant differences in age of onset, rapidity of tumour growth, presence of metastases, pathological appearance, gene expression patterns, somatic genetic changes, response to therapy and familial risk. Thus, the task of searching for common factors that associate with genetic markers has to carefully consider well-designed studies that address specific hypotheses.

Cancer genetics

Studies of familial cancer have provided great insights into cancer biology by mapping rare familial mutations that have been subsequently evaluated in the laboratory, thus adding plausibility to the observed disruption in function due to a mutation in one or more genes. These observations have also led to insights in sporadic cancers. In this regard, studies in rare paediatric cancers have yielded important insights. For example, the *RB* gene was the first tumour suppressor gene



identified through a genetic association study.¹⁴ Knudson's original description of the inheritance of retinoblastoma became the foundation of an excellent understanding of the role that *RB* plays as a tumour suppressor and transcriptional regulator.¹⁵ Another such example is the Li–Fraumeni syndrome, which is characterised by family pedigrees with high rates of sarcoma and breast cancer, as well as leukaemia, brain tumours and adrenocortical carcinomas.^{16,17} Subsequently, the identification of mutations in the *TP53* gene in a majority of, but not all, patients with Li–Fraumeni syndrome led directly to an understanding of *TP53* and its role as a critical transcription factor in normal cell growth, apoptosis and DNA repair.^{18–20}

The identification of familial breast cancer pedigrees through careful epidemiological study identified the *BRCA1* and *BRCA2* genes;^{21,22} in turn, follow-up studies have generated important insights into the function of these genes in DNA repair. Mutations in these genes in family pedigrees are highly penetrant and are associated with a significant risk for breast and ovarian cancers. Common genetic variation in *BRCA1* and *BRCA2* also appears to contribute to the risk for sporadic breast cancer, albeit with a substantially smaller effect. For example, genetic variation in *BRCA2* was shown to result in an increased risk for sporadic breast cancer in the Multiethnic Cohort (MEC).²³ Specifically, a single SNP in intron 24 was associated with a two-fold increased risk for breast cancer. This suggests that, even in the absence of a mutation that could change protein function or regulation, more subtle variants can serve as markers for increased risk for cancer.

SNPs as disease markers

Although the early history of SNP analysis was predicated on choosing candidate SNPs with known functional consequences, currently no functional information is available for the vast majority of SNPs. In fact, it is unlikely that most SNPs have functional consequences.²⁴ SNPs in certain genomic regions, such as promoters or intron–exon splice sites, could result in significant functional alterations in gene regulation, but the effort to validate this in the laboratory is arduous. It has been suggested by others that, in choosing SNPs for a genetic association study, one should cull from high-priority lists of SNPs with functional implications.²⁵ This approach has the potential to find the more highly penetrant SNPs in an association study, but is limited because it underutilises SNPs as genetic markers and, in particular, other untested SNPs that could be in linkage disequilibrium with the positive marker SNP. Until recently, many studies focused on non-synonymous SNPs because of potential amino acid changes that could affect protein structure and function. Non-synonymous SNPs contribute to the genetic diversity seen in the immune system²⁶ and potentially change the structure or function of the protein of interest; however, a large number of non-synonymous SNPs may be

conservative and have minimal or no effect on gene function.^{27,28}

SNPs that change gene regulation have also been described. Examples include an SNP in the promoter of *MDM2*, a negative regulator of p53, which was shown to increase the affinity of the transcriptional activator Sp1, resulting in higher levels of *MDM2* RNA and protein,²⁹ and synonymous variants in the human dopamine receptor 2 (*DRD2*) gene, which affect mRNA stability and translation.³⁰ Other such functional variants have been recently described, especially in pharmacogenomics.³¹ It is possible that SNPs that result in subtle changes in gene regulation are of minimal consequence in the short term, but, over the life span of an individual, accumulated changes could be significant. It is quite likely, however, that even when the nuances of gene regulation are fully understood, the majority of SNPs will still serve best as genetic markers of disease.

An understanding of population-specific genetic variation in healthy individuals is critical in choosing SNPs to investigate in a study of cancer risk. It has been well established that the distribution of the incidence of specific cancers can vary greatly across the global populations. While some of this has been ascribed to different environmental factors, it is also plausible that differences in the genetic variation of distinct populations could also contribute. In many ways, large association studies in cancer are designed to analyse genetic profiles of common variation that has been shaped by unrelated factors. In this regard, the molecular evolution of SNPs reflects the specific history of populations — in particular the admixture of different populations over time. This latter issue has been exploited by some in the use of admixture markers to investigate cancers with a disparate incidence between populations.^{32,33} Throughout evolution, humans have been subjected to different selective pressures (ie endemic pathogens or dietary needs), resulting in genetic variants which have been 'fine-tuned' in their ability to fight infection, reproduce and respond to other challenges.^{27,34,35} This results in genetic differences between different populations around the world. Differences in the origin of groups within a study can be significant enough to generate sufficient population stratification and thus add a potential confounding factor in the genetic epidemiology of complex disease.^{36–38}

Multiple interactions

Other biomarkers and environmental influences which contribute to the multi-factorial nature of cancer, as well as other complex diseases, further complicate the study of genetic association and cancer risk. Gene–gene interactions are also crucial to cancer risk assessment. The recent report from the InterLymph Consortium showed the greatest risk for non-Hodgkin lymphoma to be in individuals homozygous for the *TNF-308A* allele and carrying at least one *IL10-3585A* allele (odds ratio [OR] 2.13).³⁹ The importance of gene–gene

interactions was also demonstrated in a study of gastric cancer and cytokine gene SNPs.⁴⁰ Individuals with multiple polymorphisms of interleukin-(IL-) 1 receptor antagonist, tumour necrosis factor A and IL-10 had the greatest risk for gastric cancer, with ORs of 2.8 for one, 5.4 for two and 27.3 for three or four high-risk genotypes.

Gene–environment interactions add complexity to the interpretation of genetic association studies. One example is the investigation which has focused on the contribution of the genetic variations in the N-acetyltransferase (*NAT2*) gene to the risk for specific cancers, especially bladder and lung cancer.⁴¹ In particular, differences in the activity of *NAT2* (ie rapid and slow acetylator genotypes) could explain the association between the *NAT2* gene and tobacco smoke and subsequent risk for bladder cancer. 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.^{42,43} Interestingly, the type of tobacco appears to be important; for example, so-called black tobacco is more strongly associated with the observed effect of *NAT2* genotypes.^{42,43}

Genetic association and other clinical studies often assess only two outcomes: affected or unaffected. This approach is useful in cancer studies because cancer is usually an all or none diagnosis at the time of the study. When intermediate precursors or quantitative traits of disease are added to the analysis, however, the complexity significantly increases. Mendelian randomisation is a concept that attempts to bring together independent inheritance of individual traits with modifiable environmentally modifiable exposures.^{44,45} By using independent inheritance of traits, it is possible to reduce the confounding in studying exposure–disease associations.⁴⁶ Examples include studies of serum cholesterol, cancer risk and the *APOE* gene;⁴⁷ folate, homocysteine, coronary heart disease and the *MTHFR* gene;⁴⁴ and the relationship between alcohol, variation in the *ALDH2* gene and oesophageal cancer.⁴⁸

Study design

Subject selection and sample size

In designing a study of genetic variation and cancer risk in a population, there are a number of critical factors to consider, such as sample size, population stratification, allele frequencies of the SNPs of interest, environmental risk factors and phenotype definition. In particular, a careful definition of the cancer phenotype to be studied is crucial. Genetic factors that contribute to low-grade prostate cancer could be different to those that contribute to high-grade prostate cancer. If so, a study in which low- and high-grade diseases are grouped together could miss a potential genetic contribution for one form of the disease.^{49,50} While it may be difficult to ensure a study population that is as homogeneous as possible, it is crucial to limit confounding due to background genetic

differences. Differences in genetic variation between ethnic groups have been well described and are due to a combination of evolutionary history, migration and admixture.^{36,38,51} Efforts to avoid population stratification also need to be taken to provide cases and controls with genetic backgrounds as similar as possible.

To address some of these issues, large cohort studies, such as the MEC,⁵² are being established to create the large sample sizes needed. One strength of the MEC is that exposure and biomarker data on individuals from five different ethnic groups in Hawaii and California have been collected. This study is an immense resource for genetic epidemiology. Another such study is the National Cancer Institute (NCI)'s Breast and Prostate Cancer Cohort Consortium, consisting of over 5,000 breast cancer and 8,000 prostate cancer cases. The consortium's goal is to study genetic variation in genes in key pathways.⁵³ The Network of Investigator Networks,⁵⁴ sponsored by the Human Genome Epidemiology Network, seeks to pool analysis from multiple investigations for critical analysis and to address reproducibility issues.^{55–57}

SNP choice and interpreting the results

In order fully to understand the results of a genetic association study, all of the study endpoints described above must be considered to design a study with sufficient power to detect a measurable effect. So far, the majority of genetic association studies with common SNPs in cancer have reported modest associations, with ORs typically between 1 and 2. Examples of meta-analyses that found ORs in this range in lung cancer include *XPD* 751GG (OR 1.27)⁵⁸ and *CYP1A1* exon 7 polymorphism (OR 1.15),⁵⁹ in breast cancer include *XRCC3* T241M (OR 1.16) and *BRCA2* N372H (OR 1.13)⁶⁰ and in gastric cancer include an approximately twofold increased risk for the IL8-251A allele.^{61–63}

These studies illustrate the fact that the likelihood of finding a significant association (ie OR > 2) in a large study of a sporadic cancer is low, even for candidate genes with a strong prior. Since, by definition, SNPs are common genetic variants, individuals with a particular risk allele may never develop disease. Instead, it has become apparent that a large number of variants will each have a small contribution, perhaps evident in its population-attributable risk of 1–2 per cent per SNP. The consequence of searching for alleles with a moderate effect, namely an OR less than 1.8, is that studies have to be large and can, with rare exception, only address high frequency SNPs (ie SNPs greater than 5 per cent). Moreover, the opportunity to examine gene–environment interactions should be considered as an important reason for conducting a study.

Biological plausibility is a critical step in choosing genes for either a candidate gene or pathway approach. So far, less than 2 per cent of genes have been studied, but with the advent of new tools of whole-genome scans, there is now an opportunity to look across the genome. Still, for many studies,

SNPs have to be selected based on knowledge of the pattern of linkage disequilibrium across the gene or chromosomal region. It is fortuitous that genetic association studies have increased rapidly in scope, moving away from a single SNP in a single gene to haplotype-tagging methods for SNP selection in pathways of genes or, in the near future, whole-genome scans of 500,000 or more SNPs per individual. In the end, whole-genome scans will identify markers that will need to be carefully mapped, similar to the approach for candidate gene studies.

One of the key issues in SNP association studies is replication of results. The literature is strewn with false-positive associations and reproducibility issues. One way to address the false-positive association problem is by using the probability of a false-positive report as a means to weight the likelihood that a SNP would be associated with disease based on knowledge of the gene and/or pathway.⁶⁴ The concept of false discovery rate (FDR) is an alternative, useful way of correcting for multiple testing comparisons without the stringent penalty stipulated by the Bonferroni correction.⁶⁵ The expected proportion of false rejections of the null hypothesis among the total number of rejections is used as a measure of global error. This method has been applied successfully to studies of qualitative⁶⁵ and quantitative⁶⁶ traits. Due to linkage disequilibrium between SNPs, however, the Bonferroni correction — which tests each SNP as an individual entity — may be too stringent, and an FDR approach may be more conducive to multiple testing concerns in genetic association studies.

The whole-genome association study is based on the extremely high-throughput methods of genotyping hundreds of thousands of SNPs in each individual in the study. An advantage of this method is that the extent of genetic variation across the entire human genome can be evaluated at one time, in an 'agnostic manner'; namely without prior knowledge of the putative functional importance of a region. The NCI Cancer Genetic Markers of Susceptibility Strategic Initiative, (<http://cgems.cancer.gov>) is a programme designed to conduct whole-genome scans in breast and prostate cancer, separately, and make the data available to the public. Built into the study is the availability of nearly 7,000 cases and 7,000 controls for each disease to conduct rapid replication of findings based on an initial scan of 1,200 cases and 1,200 controls per disease, drawn from prospective, cohort studies. Over 500,000 SNPs will be analysed per subject. The choice of SNPs is based on tagging bins of SNPs using a pairwise correlation ($r^2 > 0.8$) in the North European group in HapMap Phase 2.¹¹

Reproducibility

As mentioned above, one of the most challenging aspects of genetic association studies in cancer is replication of study results. This is essential for a more thorough understanding of biological mechanisms and the development of preventive or treatment strategies. Many studies resulting in a possible

association of a particular genetic variant with an increased risk of cancer have failed to be reproduced. Some of the reasons for this include small study size, population stratification, gene–environment interactions, linkage disequilibrium around the variant studied and other intrinsic study biases. For example, in an analysis of 201 studies of complex disease of 25 different associations, Lohmueller *et al.* found evidence for replication in just less than half of the studies.⁶⁷ Another review of genetic association studies in complex diseases also showed low reproducibility.⁶⁸

Meta-analyses and large investigator networks are crucial to address these issues. Recent meta-analyses have shown reproducibility of both positive and negative associations. These include a null association of *GSTM1* deficiency in breast cancer,^{69–71} prostate cancer⁷² and in colorectal cancer,⁷³ but positive associations of *GSTM1* in leukaemia⁷⁴ and bladder cancer.⁷⁵ Meta-analyses have confirmed positive associations of *IGF1* promoter [CA]_n repeats in breast cancer,⁷⁶ the *NAT2* slow-acetylator phenotype in bladder cancer⁷⁵ and polymorphisms in DNA-repair genes in breast.^{60,77} The InterLymph Consortium investigated SNPs in key immune pathway genes, *TNF* and *IL10*, in non-Hodgkin's lymphoma (NHL) and showed an increased risk for NHL in *TNF-308A* and *IL10-3575A* allele carriers.³⁹

A hopeful future

Well-designed, well-powered studies of genetic association in cancer hold great promise for advancing knowledge of cancer biology, genetic risk factors for cancer, therapeutic response and outcome. SNPs have the potential to be used as markers of disease risk, even in the absence of understanding the functional implications of the SNP. Studies of mutations in genes such as *BRCA1* and *TP53* in families have made profound impacts on our understanding of molecular and cellular biology.^{19,22} While SNPs may not be associated with cancer risk to the same degree as a highly penetrant mutation in familial cancer, they will still contribute significantly to an understanding of a pathway or process in cancer biology. SNPs may confer as yet unknown subtle changes in gene function, transcription, intron–exon splicing or protein folding that, in the context of the right environmental exposure and/or in the appropriate genetic background of other variants, could have a significant effect on disease risk or outcome.

The public health implications of genetic association studies in cancer and other complex diseases are just beginning to emerge.⁴⁴ An excellent example of this is a study of age-related macular degeneration in which the population-attributable risk of genetic variation in the complement factor H gene is approximately 50 per cent.^{78–80} A population-attributable risk for genetic variation in cancer this significant has yet to be described, but it is possible. This, in combination with improved understanding of gene–gene and

gene–environment interactions, will provide the basis for early diagnosis, intervention and prevention of cancer. It should be pointed out, however, that the promise of studying genetic variation in cancer cannot be realised without the careful collection and annotation of cases and controls in sufficiently large studies. For the low penetrant SNPs, replication of results will have to be followed by demonstration of plausibility before entering clinical testing.

In conclusion, the tools for looking at common genetic variation are now available. Moreover, there is the opportunity to sequence large portions of the genome in many cases and controls on the horizon. The genetic opportunities will best be realised when studies that include outcome and co-variables have been carried out, especially those that reflect the environmental contributions to cancer.

Note

This is a US Government work, and, as such, is in the public domain of the United States of America.

References

- Brookes, A.J. (1999), 'The essence of SNPs', *Gene* Vol. 234, pp. 177–186.
- Kong, A., Gudbjartsson, D.F., Sainz, J. *et al.* (2002), 'A high-resolution recombination map of the human genome', *Nat. Genet.* Vol. 31, pp. 241–247.
- Collins, A., Lonjou, C. and Morton, N.E. (1999), 'Genetic epidemiology of single-nucleotide polymorphisms', *Proc. Natl. Acad. Sci. USA* Vol. 96, pp. 15173–15177.
- Cardon, L.R. and Abecasis, G.R. (2003), 'Using haplotype blocks to map human complex trait loci', *Trends Genet.* Vol. 19, pp. 135–140.
- Martin, E.R., Lai, E.H., Gilbert, J.R. *et al.* (2000), 'SNPping away at complex diseases: Analysis of single-nucleotide polymorphisms around APOE in Alzheimer disease', *Am. J. Hum. Genet.* Vol. 67, pp. 383–394.
- Altshuler, D., Brooks, L.D., Chakravarti, A. *et al.* (2005), 'A Haplotype map of the human genome', *Nature* Vol. 437, pp. 1299–1320.
- Packer, B.R., Yeager, M., Burdett, L. *et al.* (2006), 'SNP500Cancer: A public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes', *Nucleic Acids Res.* Vol. 34, pp. D617–D621.
- Kwok, P.Y. and Chen, X. (2003), 'Detection of single nucleotide polymorphisms', *Cur. Issues Mol. Biol.* Vol. 5, pp. 43–60.
- Steemers, F.J., Chang, W., Lee, G. *et al.* (2006), 'Whole-genome genotyping with the single-base extension assay', *Nat. Methods* Vol. 3, pp. 31–33.
- de Bakker, P.I., Yelensky, R., Pe'er, I. *et al.* (2005), 'Efficiency and power in genetic association studies', *Nat. Genet.* Vol. 37, pp. 1217–1223.
- Carlson, C.S., Eberle, M.A., Rieder, M.J. *et al.* (2004), 'Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium', *Am. J. Hum. Genet.* Vol. 74, pp. 106–120.
- Balmain, A., Gray, J. and Ponder, B. (2003), 'The genetics and genomics of cancer', *Nat. Genet.* Vol. 33(Suppl.), pp. 238–244.
- Chen, Y.C. and Hunter, D.J. (2005), 'Molecular epidemiology of cancer', *CA Cancer J. Clin.* Vol. 55, pp. 45–54.
- Knudson Jr., A.G. (1971), 'Mutation and cancer: Statistical study of retinoblastoma', *Proc. Natl. Acad. Sci. USA* Vol. 68, pp. 820–823.
- Zhu, L. (2005), 'Tumour suppressor retinoblastoma protein Rb: A transcriptional regulator', *Eur. J. Cancer* Vol. 41, pp. 2415–2427.
- Li, F. and Fraumeni Jr., J.F. (1969), 'Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome?', *Ann. Intern. Med.* Vol. 71, pp. 747–752.
- Li, F.P., Fraumeni Jr., J.F., Mulvihill, J.J. *et al.* (1988), 'A cancer family syndrome in twenty-four kindreds', *Cancer Res.* Vol. 48, pp. 5358–5362.
- Sengupta, S. and Harris, C.C. (2005), 'p53: Traffic cop at the crossroads of DNA repair and recombination', *Nat. Rev. Mol. Cell Biol.* Vol. 6, pp. 44–55.
- Artandi, S.E. and Attardi, L.D. (2005), 'Pathways connecting telomeres and p53 in senescence, apoptosis, and cancer', *Biochem. Biophys. Res. Commun.* Vol. 331, pp. 881–890.
- Vogelstein, B. and Kinzler, K.W. (2004), 'Cancer genes and the pathways they control', *Nat. Med.* Vol. 10, pp. 789–799.
- Hopper, J.L. (2001), 'Genetic epidemiology of female breast cancer', *Semin. Cancer Biol.* Vol. 11, pp. 367–374.
- Yoshida, K. and Miki, Y. (2004), 'Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage', *Cancer Sci.* Vol. 95, pp. 866–871.
- Freedman, M.L., Penney, K.L., Stram, D.O. *et al.* (2004), 'Common variation in BRCA2 and breast cancer risk: A haplotype-based analysis in the Multiethnic Cohort', *Hum. Mol. Genet.* Vol. 13, pp. 2431–2441.
- Chanock, S. (2001), 'Candidate genes and single nucleotide polymorphisms (SNPs) in the study of human disease', *Dis. Markers* Vol. 17, pp. 89–98.
- Tabor, H.K., Risch, N.J. and Myers, R.M. (2002), 'Opinion: Candidate-gene approaches for studying complex genetic traits: Practical considerations', *Nat. Rev. Genet.* Vol. 3, pp. 391–397.
- Hughes, A.L., Packer, B., Welch, R. *et al.* (2005), 'High level of functional polymorphism indicates a unique role of natural selection at human immune system loci', *Immunogenetics* Vol. 57, pp. 821–827.
- Fredman, D., Sawyer, S.L., Stromqvist, L. *et al.* (2006), 'Nonsynonymous SNPs: Validation characteristics, derived allele frequency patterns, and suggestive evidence for natural selection', *Hum. Mutat.* Vol. 27, pp. 173–186.
- Johnson, M.M., Houck, J. and Chen, C. (2005), 'Screening for deleterious nonsynonymous single-nucleotide polymorphisms in genes involved in steroid hormone metabolism and response', *Cancer Epidemiol. Biomarkers Prev.* Vol. 14, pp. 1326–1329.
- Bond, G.L., Hu, W., Bond, E.E. *et al.* (2004), 'A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans', *Cell* Vol. 119, pp. 591–602.
- Duan, J., Wainwright, M.S., Comeron, J.M. *et al.* (2003), 'Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor', *Hum. Mol. Genet.* Vol. 12, pp. 205–216.
- Watters, J.W. and McLeod, H.L. (2003), 'Cancer pharmacogenomics: Current and future applications', *Biochim. Biophys. Acta* Vol. 1603, pp. 99–111.
- Reich, D. and Patterson, N. (2005), 'Will admixture mapping work to find disease genes?', *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* Vol. 360, pp. 1605–1607.
- Patterson, N., Hattangadi, N., Lane, B. *et al.* (2004), 'Methods for high-density admixture mapping of disease genes', *Am. J. Hum. Genet.* Vol. 74, pp. 979–1000.
- Voight, B.F., Kudravalli, S., Wen, X. and Pritchard, J.K. (2006), 'A map of recent positive selection in the human genome', *PLoS Biol.* Vol. 4, p. e72.
- Akey, J.M., Eberle, M.A., Rieder, M.J. *et al.* (2004), 'Population history and natural selection shape patterns of genetic variation in 132 genes', *PLoS Biol.* Vol. 2, p. e286.
- Ioannidis, J.P., Ntzani, E.E. and Trikalinos, T.A. (2004), '"Racial" differences in genetic effects for complex diseases', *Nat. Genet.* Vol. 36, pp. 1312–1318.
- Cardon, L.R. and Palmer, L.J. (2003), 'Population stratification and spurious allelic association', *Lancet* Vol. 361, pp. 598–604.
- Marchini, J., Cardon, L.R., Phillips, M.S. and Donnelly, P. (2004), 'The effects of human population structure on large genetic association studies', *Nat. Genet.* Vol. 36, pp. 512–517.

39. Rothman, N., Skibola, C.F., Wang, S.S. *et al.* (2006), 'Genetic variation in TNF and IL10 and risk of non-Hodgkin lymphoma: A report from the InterLymph Consortium', *Lancet Oncol.* Vol. 7, pp. 27–38.
40. El Omar, E.M., Rabkin, C.S., Gammon, M.D. *et al.* (2003), 'Increased risk of noncardia gastric cancer associated with proinflammatory cytokine gene polymorphisms', *Gastroenterology* Vol. 124, pp. 1193–1201.
41. Hein, D.W., Doll, M.A., Fretland, A.J. *et al.* (2000), 'Molecular genetics and epidemiology of the NAT1 and NAT2 acetylation polymorphisms', *Cancer Epidemiol. Biomarkers Prev.* Vol. 9, pp. 29–42.
42. Golka, K., Prior, V., Blaszkewicz, M. and Bolt, H.M. (2002), 'The enhanced bladder cancer susceptibility of NAT2 slow acetylators towards aromatic amines: A review considering ethnic differences', *Toxicol. Lett.* Vol. 128, pp. 229–241.
43. Hein, D.W. (2002), 'Molecular genetics and function of NAT1 and NAT2: Role in aromatic amine metabolism and carcinogenesis', *Mutat. Res.* Vol. 506–507, pp. 65–77.
44. Davey, S.G. and Ebrahim, S. (2003), "'Mendelian randomization": Can genetic epidemiology contribute to understanding environmental determinants of disease?', *Int. J. Epidemiol.* Vol. 32, pp. 1–22.
45. Davey, S.G., Ebrahim, S., Lewis, S. *et al.* (2005), 'Genetic epidemiology and public health: Hope, hype, and future prospects', *Lancet* Vol. 366, pp. 1484–1498.
46. Smith, G.D. and Ebrahim, S. (2004), 'Mendelian randomization: Prospects, potentials, and limitations', *Int. J. Epidemiol.* Vol. 33, pp. 30–42.
47. Katan, M.B. (1986), 'Apolipoprotein E isoforms, serum cholesterol, and cancer', *Lancet* Vol. 1(8479), pp. 507–508.
48. Lewis, S.J. and Smith, G.D. (2005), 'Alcohol, ALDH2, and esophageal cancer: A meta-analysis which illustrates the potentials and limitations of a Mendelian randomization approach', *Cancer Epidemiol. Biomarkers Prev.* Vol. 14, pp. 1967–1971.
49. Sorlie, T., Tibshirani, R., Parker, J. *et al.* (2003), 'Repeated observation of breast tumor subtypes in independent gene expression data sets', *Proc. Natl. Acad. Sci. USA* Vol. 100, pp. 8418–8423.
50. Staudt, L.M. (2003), 'Molecular diagnosis of the hematologic cancers', *N. Engl. J. Med.* Vol. 348, pp. 1777–1785.
51. Carlson, C.S., Eberle, M.A., Rieder, M.J. *et al.* (2003), 'Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans', *Nat. Genet.* Vol. 33, pp. 518–521.
52. Kolonel, L.N., Altshuler, D. and Henderson, B.E. (2004), 'The multi-ethnic cohort study: Exploring genes, lifestyle and cancer risk', *Nat. Rev. Cancer* Vol. 4, pp. 519–527.
53. Hunter, D.J., Riboli, E., Haiman, C.A. *et al.* (2005), 'A candidate gene approach to searching for low-penetrance breast and prostate cancer genes', *Nat. Rev. Cancer* Vol. 5, pp. 977–985.
54. Ioannidis, J.P., Gwinn, M., Little, J. *et al.* (2006), 'A road map for efficient and reliable human genome epidemiology', *Nat. Genet.* Vol. 38, pp. 3–5.
55. Benhamou, S. and Sarasin, A. (2005), 'ERCC2 /XPD gene polymorphisms and lung cancer: A HuGE review', *Am. J. Epidemiol.* Vol. 161, pp. 1–14.
56. Hung, R.J., Hall, J., Brennan, P. and Boffetta, P. (2005), 'Genetic polymorphisms in the base excision repair pathway and cancer risk: A HuGE review', *Am. J. Epidemiol.* Vol. 162, pp. 925–942.
57. Masson, L.F., Sharp, L., Cotton, S.C. and Little, J. (2005), 'Cytochrome P-450 1A1 gene polymorphisms and risk of breast cancer: A HuGE review', *Am. J. Epidemiol.* Vol. 161, pp. 901–915.
58. Hu, Z., Wei, Q., Wang, X. and Shen, H. (2004), 'DNA repair gene XPD polymorphism and lung cancer risk: A meta-analysis', *Lung Cancer* Vol. 46, pp. 1–10.
59. Le Marchand, L., Guo, C., Benhamou, S. *et al.* (2003), 'Pooled analysis of the CYP1A1 exon 7 polymorphism and lung cancer (United States)', *Cancer Causes Control* Vol. 14, pp. 339–346.
60. Garcia-Closas, M., Egan, K.M., Newcomb, P.A. *et al.* (2006), 'Polymorphisms in DNA double-strand break repair genes and risk of breast cancer: Two population-based studies in USA and Poland, and meta-analyses', *Hum. Genet.* Vol. 119, pp. 376–388.
61. Savage, S.A., Abnet, C.C., Mark, S.D. *et al.* (2004), 'Variants of the IL8 and IL8RB genes and risk for gastric cardia adenocarcinoma and esophageal squamous cell carcinoma', *Cancer Epidemiol. Biomarkers Prev.* Vol. 13, pp. 2251–2257.
62. Taguchi, A., Ohmiya, N., Shirai, K. *et al.* (2005), 'Interleukin-8 promoter polymorphism increases the risk of atrophic gastritis and gastric cancer in Japan', *Cancer Epidemiol. Biomarkers Prev.* Vol. 14, pp. 2487–2493.
63. Ohyauchi, M., Imatani, A., Yonechi, M. *et al.* (2005), 'The polymorphism interleukin 8 -251 A/T influences the susceptibility of *Helicobacter pylori* related gastric diseases in the Japanese population', *Gut* Vol. 54, pp. 330–335.
64. Wacholder, S., Chanock, S., Garcia-Closas, M. *et al.* (2004), 'Assessing the probability that a positive report is false: An approach for molecular epidemiology studies', *J. Natl. Cancer Inst.* Vol. 96, pp. 434–442.
65. Sabatti, C., Service, S. and Freimer, N. (2003), 'False discovery rate in linkage and association genome screens for complex disorders', *Genetics* Vol. 164, pp. 829–833.
66. Weller, J.I., Song, J.Z., Heyen, D.W. *et al.* (1998), 'A new approach to the problem of multiple comparisons in the genetic dissection of complex traits', *Genetics* Vol. 150, pp. 1699–1706.
67. Lohmueller, K.E., Pearce, C.L., Pike, M. *et al.* (2003), 'Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease', *Nat. Genet.* Vol. 33, pp. 177–182.
68. Hirschhorn, J.N., Lohmueller, K., Byrne, E. and Hirschhorn, K. (2002), 'A comprehensive review of genetic association studies', *Genet. Med.* Vol. 4, pp. 45–61.
69. Egan, K.M., Cai, Q., Shu, X.O. *et al.* (2004), 'Genetic polymorphisms in GSTM1, GSTP1, and GSTT1 and the risk for breast cancer: Results from the Shanghai Breast Cancer Study and meta-analysis', *Cancer Epidemiol. Biomarkers Prev.* Vol. 13, pp. 197–204.
70. Sull, J.W., Ohrr, H., Kang, D.R. and Nam, C.M. (2004), 'Glutathione S-transferase M1 status and breast cancer risk: A meta-analysis', *Yonsei Med. J.* Vol. 45, pp. 683–689.
71. Vogl, F.D., Taioli, E., Maugard, C. *et al.* (2004), 'Glutathione S-transferases M1, T1, and P1 and breast cancer: A pooled analysis', *Cancer Epidemiol. Biomarkers Prev.* Vol. 13, pp. 1473–1479.
72. Ntasi, C., Polycarpou, A. and Ioannidis, J.P. (2005), 'Association of GSTM1, GSTT1, and GSTP1 gene polymorphisms with the risk of prostate cancer: A meta-analysis', *Cancer Epidemiol. Biomarkers Prev.* Vol. 14, pp. 176–181.
73. Smit, K.M., Gaspari, L., Weijnenberg, M.P. *et al.* (2003), 'Interaction between smoking, GSTM1 deletion and colorectal cancer: Results from the GSEC study', *Biomarkers* Vol. 8, pp. 299–310.
74. Ye, Z. and Song, H. (2005), 'Glutathione s-transferase polymorphisms (GSTM1, GSTP1 and GSTT1) and the risk of acute leukaemia: A systematic review and meta-analysis', *Eur. J. Cancer* Vol. 41, pp. 980–989.
75. Garcia-Closas, M., Malats, N., Silverman, D. *et al.* (2005), 'NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: Results from the Spanish Bladder Cancer Study and meta-analyses', *Lancet* Vol. 366, pp. 649–659.
76. Wen, W., Gao, Y.T., Shu, X.O. *et al.* (2005), 'Insulin-like growth factor-I gene polymorphism and breast cancer risk in Chinese women', *Int. J. Cancer* Vol. 113, pp. 307–311.
77. Zhang, Y., Newcomb, P.A., Egan, K.M. *et al.* (2006), 'Genetic polymorphisms in base-excision repair pathway genes and risk of breast cancer', *Cancer Epidemiol. Biomarkers Prev.* Vol. 15, pp. 353–358.
78. Klein, R.J., Zeiss, C., Chew, E.Y. *et al.* (2005), 'Complement factor H polymorphism in age-related macular degeneration', *Science* Vol. 308, pp. 385–389.
79. Edwards, A.O., Ritter, III, R., Abel, K.J. *et al.* (2005), 'Complement factor H polymorphism and age-related macular degeneration', *Science* Vol. 308, pp. 421–424.
80. Haines, J.L., Hauser, M.A., Schmidt, S. *et al.* (2005), 'Complement factor H variant increases the risk of age-related macular degeneration', *Science* Vol. 308, pp. 419–421.