Wünsch Filho, Victor; Zago, Marco A

Modern cancer epidemiological research: genetic polymorphisms and environment

Revisão   *Review*

Modern cancer epidemiological research: genetic polymorphisms and environment
Pesquisa epidemiológica contemporânea em câncer: polimorfismos genéticos e ambiente

Victor Wünsch Filho[a] **and Marco A Zago**[b]

[a]*Departamento de Epidemiologia. Faculdade de Saúde Pública. Universidade de São Paulo (USP). São Paulo, SP, Brasil.* [b]*Departamento de Medicina. Faculdade de Medicina. USP. Ribeirão Preto, SP, Brasil*

Abstract

Individual cancer susceptibility seems to be related to factors such as changes in oncogenes and tumor suppressor genes expression, and differences in the action of metabolic enzymes and DNA repair regulated by specific genes. Epidemiological studies on genetic polymorphisms of human xenobiotics metabolizing enzymes and cancer have revealed low relative risks. Research considering genetic polymorphisms prevalence jointly with environmental exposures could be relevant for a better understanding of cancer etiology and the mechanisms of carcinogenesis and also for new insights on cancer prognosis. This study reviews the approaches of molecular epidemiology in cancer research, stressing case-control and cohort designs involving genetic polymorphisms, and factors that could introduce bias and confounding in these studies. Similarly to classical epidemiological research, genetic polymorphisms requires considering aspects of precision and accuracy in the study design.

Resumo

A suscetibilidade individual ao câncer envolve fatores como mudanças na expressão dos oncogenes e genes supressores de tumores e, também, diferenças em genes específicos que regulam a ação de enzimas metabólicas e de reparo do DNA. Estudos epidemiológicos relacionando polimorfismos genéticos e câncer têm mostrado riscos relativos tênues. Pesquisas abordando simultaneamente a prevalência de polimorfismos genéticos e a exposição a fatores ambientais podem conduzir a resultados relevantes para a compreensão da etiologia do câncer e dos mecanismos da carcinogênese e, ainda, trazer novas contribuições sobre o prognóstico da doença. No presente artigo, discute-se a lógica da epidemiologia molecular na pesquisa em câncer, ressaltando-se o exame de polimorfismos genéticos nos estudos caso-controle e de coorte, bem como os fatores que podem introduzir vieses e confusão. Assim como na pesquisa epidemiológica clássica, na epidemiologia molecular é necessário avaliar aspectos relacionados à precisão e à acurácia na fase de planejamento do estudo.

Correspondence to:
Victor Wünsch Filho
Departamento de Epidemiologia - FSP/USP
Av. Dr. Arnaldo, 715
01246-904 São Paulo, SP, Brasil
E-mail: wunsch@usp.br

INTRODUCTION

There is now a consensus that environmental factors often exert their influence through genetic mechanisms to promote disease. The large variations of cancer incidence among populations living in different regions of the world suggest that the majority of malignant tumors are related to environmental factors.[25] There is no thorough mapping of cancer incidence, since population-based tumor registries do not cover some regions and information on mortality could be poor. Nevertheless, differences between tumor rates observed in a population compared to another with minimum rates have been assumed as the attributable fraction of cancer caused by environmental factors, even though these populations could differ greatly according to gene polymorphisms frequency. According to these assumptions, 80% to 90% of all cancers have determinants such as viruses, soot and fumes inhaled, food and pollutants ingested, as well as chemical substances and radiations, which affect human body.[27] Such inferences are consistent with studies on migrant groups. The first-generation of Japanese immigrants in the city of São Paulo revealed incidence patterns for some cancers distinct from those of the Japanese people and closer to the Brazilian population.[33] The contribution of hereditary factors in cancer is considered minor. Certain familial cancer clusters with monogenic pattern are very rare and account for a relatively small proportion (less than 5%) of total cancer.[26] The role of heredity on occurrence of common cancers with non-Mendelian pattern is unclear.[38]

Several environmental carcinogens have been identified through epidemiological studies during last century, but until recently they altogether disregarded inter-individual variations in response to exposure or, in other words, there were no available technologies in the past to evaluate individual susceptibility. The rapid progress achieved by molecular biology has provided basic tools for geneticists and epidemiologists to go deeper within genotype characteristics and has as well given fresh insights of this influence on cancer distribution within populations.

For this review the first step was the scientific bibliography systematic assessment from the Medline database using the following keywords: genetic polymorphisms and cancer; gene-environment interaction and cancer. For the first keywords searched were established some limits: only items in English and with abstracts available from 1/1/1990 to 12/31/2003. Only three publications types were examined: clinical trials (n=87); randomized controlled trials (n=29); and meta-analysis (n=24). For gene-environment interactions and cancer keywords all publications types had their abstracts reviewed (n=93). From the abstracts reading, only the relevant complete papers were selected considering their insights on the current rationality of molecular epidemiology in cancer research, that of simultaneously evaluate environmental exposures and genetic markers in cancer etiology, particularly focusing the use of common genetic polymorphisms. Additional papers were searched from the most important bibliography references included in those papers.

ENVIRONMENT AND GENETIC FACTORS IN CANCER

Tobacco smoking is well recognized worldwide as the most important exogenous isolated risk factor for cancer. About 90% of lung cancer cases can be attributed to cigarette smoking, but no more than 20% of smokers develop this type of cancer.[39] From the pioneering study of Tokuhata & Lilienfeld[32] on familial aggregation of lung cancer, several studies have consistently demonstrated an increased prevalence of cancer among relatives of lung cancer patients. Table 1 shows data on smoking from a case-control study[40] and as can be noted, cigarette smoking has distinct effects on those with or without reported cancer cases in first-degree relatives. A joint effect was detected

Table 1 - Joint effect of smoking and cancer in first-degree relatives on lung cancer.*

Smoking (Pack-years)	Any cancer in relatives**	Cases (N=334)	Controls (N=578)	OR*** (95% CI)****
No smoker	No	28	162	1.00
	Yes	9	47	1.08 (0.47-2.47)
1-20	No	39	88	3.36 (1.87-6.03)
	Yes	6	24	1.86 (0.68-5.10)
21-40	No	50	85	5.18 (2.80-9.58)
	Yes	17	19	7.17 (3.19-16.13)
41-60	No	55	49	8.93 (4.78-16.65)
	Yes	29	18	11.73 (5.44-25.25)
61 and more	No	58	45	10.03 (5.32-18.91)
	Yes	30	15	14.90 (6.68-33.25)

*Source: Wünsch-Filho et al,[40] with permission
**Parents and siblings
***Odds ratio adjusted for gender, age and socio-economic status (four levels)
****95% confidence interval

between the variables smoking and cancer in relatives on the risk of lung cancer.

Some authors have suggested that genetic predisposition may contribute to familial aggregation of lung cancer.[6,39] For example, interactions were observed between pack-years of smoking and the combined GSTM1 null, GSTM3 AA, GSTP1 (AG or GG) genotype. The combination of three at risk genotypes conferred an increased risk of lung cancer among smokers with a history of at least 35 pack-years, but not in lighter smokers,[13] suggesting that this genotype increases susceptibility to lung cancer only in smokers. On the other hand, results of familial aggregation studies may also be explained by the fact that people from the same family tend to share similar habits, such as tobacco smoking, alcohol intake or occupation.

Disentangle genetic and environmental components on the risk of cancer is an old puzzle for epidemiologists.[30] Lichtenstein et al[19] reported from their twin cohort some distinctions on environment and genetics in cancer etiology. The rates of sporadic cancers were similar in monozygotic and dizygotic pairs. For five tumors (stomach, colorectal, pancreas, lung and prostate) the estimated effects of genetic factors varied from 26% to 42%, the highest for prostate cancer. The effects of environmental factors ranged from 58% to 64%, and authors detected the highest effect due to environmental components for breast cancer (73%). Despite these results from Scandinavia, there are some practical limitations, which are, in general, inherent to studies concerning twins:[11] the study did not address information on screening practices and specific types of exposure (for example, tobacco and alcohol). Then no interaction among genes and environmental factors could be evaluated. Understanding mechanisms of inherited genetic predisposition, responsible for inter-individual variation in response to xenobiotics, will lead to a better comprehension of carcinogenesis, motivating preventive policies and advances in development of new intelligent drugs against cancer.[38] Research in molecular biology has given continuous facts concerning specific function of biomarkers, which could have complementary acting mechanisms. From current conceptions, cancer arises from a stepwise accumulation of genetic changes. Efforts in molecular biology to unravel the nature of neoplastic transformation, through experimental studies in animals and epidemiological observational studies in humans, indicate the existence of a common set of mechanisms involved in carcinogenesis.[8]

Genetic factors in cancer encompass a gradient, which involves two extremes, the first of which comprise high-penetrance gene mutations and individuals with alleles at risk approaches to 100% of disease incidence, such as familial retinoblastoma – a childhood cancer inherited in an autosomal dominant pattern. On the opposite, genetic polymorphisms, whose variations in gene sequence show low-penetrance, such as GSTM1 null polymorphism and lung cancer.[36,37] Table 2 exhibits comparisons of this spectrum.

Gene mutations linked with recognized familial cancers are very specific and rare, and is a characteristic particularly of those tumors incident at younger ages (retinoblastoma). Strongly predisposing mutations in

Table 2 - Spectrum of genetic variables according to gene penetrance* in cancer.

Characteristic	High penetrance gene mutation	Low penetrance gene mutation
Genetic variable	Single gene is the predominant factor on cancer Simple Mendelian inheritance (Single-gene diseases)	Genetic variants (polymorphisms) with modest effect on cancer No Mendelian pathway inheritance (Multifatorial diseases)
Population prevalence	Rare	Common
Population effect	Small cancer proportion (Low attributable risk)	Potentially high contribution to cancer considering interaction with environmental variables (High attributable risk)
Individual effect	High risk in carriers	Low risk in carriers
Relation with environmental factors	Could act with no interactions with environmental factors	Gene-environment interactions are intrinsic to mode of action of low-penetrance gene polymorphisms
Study design approach	Genetic studies: • Twins studies • Segregation analysis • Linkage analysis	Epidemiological studies: • Case-control • Cohort
Example	BRCA1 and BRCA2 mutations and breast cancer Penetrance =6.08%** Population prevalence =0.4%***	GSTM1 null and lung cancer Relative risk =1.14**** Population prevalence =4.06%*****

*Penetrance: The lifetime risk of developing cancer (usually up to the age of 70 years)
Source:
**Narod & Foulkes[23] and Peto[26]
***Narod & Foulkes[23]
****Houlston[12]
*****Gattás & Soares-Vieira[5]

BRCA1 and BRCA2 genes might be responsible for most multiple-case families (which confer individual risks of around 60-80% by age 70), but for only 2% of all cases at population level.[26] The population prevalence of BRCA1 and BRCA2 genes mutations is about 0.4%.[23] In contrast, prevalence of genetic polymorphisms changes according to race[5] and population prevalence could range from 1% to 90%, but the effect on cancer causality seems to be low.

GENETIC POLYMORPHISMS AND CANCER

Evolution has rendered human species capable of metabolizing chemical substances, which if accumulated in cells could harm the health, and explain interethnic and inter-individual variability of chemical metabolism. Gene polymorphisms of drug-metabolizing enzymes are almost always the rule, not exception.[16] Several gene variants, such as single nucleotide polymorphisms, could be nominated. One first and essential research question is how to separate those important for evaluation of cancer etiology or cancer prognosis from those to be disregard.

Most of gene polymorphisms pointed thus far in cancer epidemiology codifies for metabolic enzymes involved in phases I or II of chemical metabolism and those responsible for the DNA repair, inflammatory response, cell adhesion and vascular grow as well. In general, environmental proto-carcinogens are converted into their reactive electrophilic intermediates through enzymes acting in phase I, much in same fashion as those of cytochrome P450 family (CYP). The effect on distinct tumors of CYP gene polymorphisms was examined in several populations and ethnic groups. Considering the association between CYP1A1 and lung cancer, studies appointed to odds ratios (OR) with variation from OR 2.08[41] to OR 7.31.[22] During phase II, molecule toxicity can be reduced through the action of other enzymes, codified by specific genes, which in turn transform xenobiotics in water soluble, easily excretable metabolites. A group of enzymes, as glutathione S-transferase (GST) and N-acetil-transferase (NAT), lead this phase of metabolization process.[31] Glutathione S-transferase (GST) family includes no fewer than four classes of genes: GSTalpha, mu (GSTM), pi (GSTP), and theta (GSTT), the alleles of which have been frequently associated with some types of tumors.[1] Studies of GST genes polymorphisms and lung cancer have found associations lower than three,[9] ranging from OR 0.8[24] to OR 2.7.[13] A meta-analysis of combined results of studies with GST genes revealed a tenuous risk, OR 1.14 (95% CI: 1.03-1.25).[12] However, considering incidence of lung cancer and polymorphisms prevalence the attributable risk may be considerable.[2]

Studies that merged enzymes acting in phase I and II of metabolism have showed higher associations with lung cancer, but variation of observed risks was wide. Nakachi et al[21] reported very high risks of lung cancer, OR 16.0 (95% CI: 3.7-68.0) for combination of CYPlA1 (Mspl) and GSTM1 deletion polymorphisms and high cumulative cigarette dose. Otherwise, Le Marchand et al[18] achieved an OR 3.1 (95% CI: 1.2-7.9) for this same conjunction.

Even in studies with statistically significant results, some not always seem biologically plausible.[34] The vast amount of data generated by human genome sequencing and the continuous use and development of arrays and other parallel genomic laboratory technologies[20] will assist the practice of epidemiologists. These may examine more objectively individual susceptibility and conduct studies adjusted to current research demands and in close consonance with biologists and clinicians.

Undoubtedly, one of the main defiance for cancer research within next decades will be unraveling genetic basis of multifactorial cancers, which do not show a Mendelian pattern of inheritance. For examining the genetic factors effect modification on the environmental factors in the genesis of cancer, it will be necessary to conduct studies with larger samples sizes in order to reach statistical power. Studies with considerable number of subjects imply attention to a few variables. For example, differences between samples of biological specimens may be significant, taking into account several difficulties, such as variations in sample collecting and laboratory procedures.[3] Additionally, error in gene polymorphisms identification may vary depending on laboratory technical details, however, the current genotyping method Taqman has 96% sensitivity and 98% specificity.[35]

CASE-CONTROL AND COHORT DESIGNS IN CANCER STUDIES INVOLVING GENE POLYMORPHISMS

Even though etiological studies relating individual gene polymorphisms and cancer have been disconcerting results, promising lines of research would include investigations on gene-gene and gene-environment interactions. Case-control studies have been the method of choice in assessment of risks related to genetic polymorphisms, which are analyzed in DNA obtained from leukocytes in cancer cases and respective controls, frequently matched by sex, age and ethnic group. As standard in this type of study, cases and controls are compared in statistical analysis by use of stratified and multivariate models.

Table 3 - Joint effect of alcohol and CYP1A1 gene polymorphisms. Oral, pharynx and larynx tumors case-control study.*

Alcohol drinking (g/L/day)**	CYP1A1	Cases (N=103)	Controls (N=102)	OR (95% CI)
≤5	Null	9	24	1.00
	Present	6	11	1.45 (0.14-5.11)
	Present in both alleles	0	3	0
>5	Null	55	38	3.86 (1.62-9.22)
	Present	29	25	3.09 (1.22-7.87)
	Present in both alleles	4	1	10.67 (1.05-108.69)

*Source: Gattás GJF, Siraque MS, Soares-Vieira JA, Kohler P, Guarniero FB, Carvalho MB, Eluf-Neto J, Wünsch-Filho V. Gene polymorphisms and oral and larynx cancer. (Research project FAPESP 03/07864-9. Unpublished data)
**Grams/Liter/Day

Hospital-based case-control study seems the most appropriate strategy to easily get participants' agreement to provide biological specimens, such as blood samples or fresh tumor tissue. Controls will also be much more receptive concerning their participation, since they are inside hospital. In contrast, controls selected from a population sample could potentially present higher rates of participation refusal, thus inducing selection bias. Large variations, inconclusive results and lack of biological plausibility derived from case-control studies on gene polymorphisms and cancer have been imputed to sample size and lack of power to show statistical significance, control selection and variations in accuracy of laboratorial methods.[2,4]

In order to have a study powerful enough to test gene polymorphisms effect, calculations of sample sizes depend upon population prevalence, as well as the level of effect to be estimated. If it is known that allele A increases susceptibility for a specific cancer and its prevalence in the population is 15%, a case-control study will need around 200 cases and similar number of controls to estimate the effect of allele A (assuming an odds ratio around 2.0), and considering a Type 1 error of 0.05 and a power of 80%. Moreover, testing interactions between genes and environment factors will require studies with some extra hundreds of participants.

Table 3 displays data from a study of head and neck tumors conducted in the city of São Paulo. It was considered the joint effect of alcohol drinking and CYP1A1 polymorphisms. Study subjects related to alcohol intake were split in two groups, those with low intake (equal or less than 5 grams/liter/day) and those with higher intake. These groups were examined according to deletion presence (heterozygous condition) or presence in both alleles (homozygous condition). The sample size of 103 cases and 102 controls hampered joint effect risk estimation for those with low alcohol drinking and CYP1A1 homozygous condition. Those with higher alcohol consumption and presence of CYP1A1 in both alleles have high risk, but the large confidence interval revealed the result inconsistencies and difficulties to accept it as valid.

In parallel to statistical power, misclassification, selection bias and confounding are problems to be considered in these studies. Misclassification of exposure status can seriously bias assessment of gene-environment interaction.[28] Selection bias may occur according to ethnic group. If response rates differ between cases and controls due to ethnicity, comparisons of gene frequency across groups will be threatened. Thus, any genetic or environmental exposure whose distribution differs between ethnic groups may seem to be disease related, even though there is no causal relationship. To counteract this difficulty, an alternative could be a nested case-control, but the issue is in setting cohort study. Confounding is another potential problem caused by ethnicity. The option is matching cases and controls during the study's recruitment phase or, else, simple population stratification by race in analysis to better control confounding. However, for multiracial populations such as the Brazilian – where crossbreeding of Europeans, Africans and Native Indian molds about 50% of population – confounding imputed to ethnicity may be minimum.[5]

Evolutions of case-control design have been proposed to study genetic polymorphisms. Case-only study is a choice when the main objective is joint effect of environment and genetic factors estimation. Inability to examine effects of environmental exposure or genetic polymorphism alone is a limitation of this approach.[2] Nevertheless, there are some advantages, gain in accuracy and cost reduction brought about by not having to test controls. Additionally, case-only design removes the control component from variance, and achieves the same statistical power as a study with a larger number of controls per case.[4] Family-based case-control studies have also been proposed to evaluate effects of gene polymorphisms in cancer. Families are recruited from a diseased member, and data on environmental exposures and biological samples are collected from all family members. The challenges are to collect individual data on gene polymorphisms and environmental exposures and other covariates on all family members in a sample size large enough to allow detecting interaction.[7] Some author's argument that this kind of approach

has no advantages on a well conducted matched case-control study.[2]

Cohort design seems a less attractive option for studies of genetic polymorphisms and cancer etiology. Cohort studies' main hindrances are defining a well-established population exposed to an environmental carcinogen, such as a group of workers in an occupational setting, and also, convincing participants to provide biological samples. Those who agree to be involved in the study can be rather small, thus introducing a selection bias, which might compromise the internal validity of the study and limit the generalization of the estimated associations. But, if a cohort succeeds in achieving a high degree of adherence by participants, there will be virtually no misclassification of exposure status. Cohorts allow planning nested case-control studies, which have various advantages on hospital-based case-control studies, the most important being the confident classification of environmental exposure.[17]

Cohort strategy could be used for follow-up of patients with tumors stratified according to cell subtype and presence of specific gene polymorphism. This use is probably the most effective application of cohort approach in studies of molecular epidemiology. The target is to examine implications of genetic polymorphisms or gene expression on disease prognosis considering death and other endpoints, such as treatment responses or clinical evolution. Characterization of a malignant disease by molecular markers could improve the understanding of clinical course of disease in individual patients and have insights on their prognosis. Some studies have examined the effect of genetic polymorphisms, such as those linked to genes of invasion and metastasis, cell cycle control, DNA repair, and inflammation, on survival of patients with different types of cancer.[29]

FUTURE PROSPECTS OF MOLECULAR EPIDEMIOLOGY IN CANCER RESEARCH

In recent years numerous single nucleotide polymorphisms (SNPs) have been identified within human population, some of them related to cancer susceptibility, cancer progress or host response to disease. Gene polymorphisms are involved in several mechanisms, such as inflammatory response genes, chemokine and chemokine-receptor genes, cell cycling, and cell metabolism. Research groups are joining efforts to identify novel polymorphisms and build testing confirmatory studies to measure the effect of identified gene polymorphisms. Selection of SNPs to be examined through epidemiological studies must be conducted conforming to a few criteria, such as

SNPs belonging to one type of polymorphism common in population (allele frequency higher than 5%) or SNPs of research interest from scientific literature.[34]

Some of identified SNPs through expanding, less expensive and more powerful genomic technologies will provide more alternatives for cancer etiology studies. A second fraction of them could be important to clarify prognosis of different types of cancer, while some others could guide research for development of new oncological drugs. Additionally, inherited differences in sensitivity to drugs may be another reasonable research goal, important for identification of different individual reactions to cancer therapy. This will lend guidance in administering lower dosages to sensitive patients.[10]

CONCLUSIONS

By long time it has been debated whether nature or nurture is responsible for individual variation to different diseases or, in other terms, how much disease is due to genetics as opposed to environment.[30] Cancer is a disease of complex etiology. Classical epidemiological investigations focused on evaluation of environmental agents and cancer seems to have reached the edge of discovering new carcinogens with important public health impact.[15] One of the proposals of molecular epidemiology in cancer is to know how environmental effects on cancer risk are modulated by genetic polymorphisms.

Single genes, necessary and sufficient to cause cancer, are rare, with high absolute and relative risks, have virtually no dependence on environmental exposures and have low population attributable risk. Conversely, genetic polymorphisms are common, have low absolute and relative risks, so usually exert small effects on cancer. Genetic polymorphisms modulate the individual susceptibility to cancer and are strongly dependent of environmental mutagens and therefore have high population attributable risks.[14]

The current epidemiological cancer research is driven to understand the intricate relationships of environment and genetics and improve one's comprehension on prevention, diagnosis and prognosis of cancer. Interactions of many genes and gene-environment are categories of analysis in these studies. In order to obtain statistical power to examine these interactions it is necessary to obtain large number of patients and controls. The strategy to have these large sample sizes needs to consider the conduction of multicentric studies and an integrated effort of clinicians, geneticists, molecular biologists, epidemiologists and biostatisticians.

REFERENCES

1. Awasthi YC, Sharma R, Singhal SS. Human glutathione S-transferases. *Int J Biochem* 1994;26:295-308.

2. Brennan P. Gene-environment interaction and etiology of cancer: what does it mean and how can we measure it? *Carcinogenesis* 2002;23:381-7.

3. Caporaso N, Rothman N, Wacholder S. Case-control studies of common alleles and environmental factors. *Natl Cancer Inst Monogr* 1999;26:25-30.

4. Cuzick J. Interaction, subgroup analysis and sample size. In: Vineis P, Malats N, Lang M, d'Errico A, Caporaso N, Cuzick J, Boffetta P, editors. Metabolic polymorphisms and susceptibility to cancer. Lyon: IARC Press; 1999. p. 109-21. [IARC Scientific Publications, 148]

5. Gattás GJF, Soares-Vieira JA. Cytochrome P450-2E1 and glutathione S-transferase mu polymorphisms among caucasians and mulattoes from Brazil. *Occup Med* 2000;50:508-11.

6. Gauderman WJ, Morrison JL. Evidence for age-specific genetic relative risks in lung cancer. *Am J Epidemiol* 2000;151:41-9.

7. Goldstein AM, Andrieu N. Detection of interaction involving identified genes: available study designs. *Natl Cancer Inst Monogr* 1999;26:49-54.

8. Hahn WC, Weinberg RA. Modelling the molecular circuitry of cancer. *Nat Rev Cancer* 2002;2:331-41.

9. Hatagima A. Genetic polymorphisms and metabolism of endocrine disruptors in cancer susceptibility. *Cad Saúde Pública* 2002;18:357-77.

10. Holtzman NA, Marteau TM. Will genetics revolutionize medicine? *N Engl J Med* 2000;343:141-4.

11. Hoover RN. Cancer. Nature, nurture, or both. *N Engl J Med* 2000;343:135-6.

12. Houlston RS. Glutathione S-transferase M1 status and lung cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 1999;8:675-82.

13. Jourenkova-Mironova N, Wikman H, Bouchardy C, Voho A, Dayer P, Benhamou S, Hirnoven A. Role of glutathione S-transferase GSTMl, GSTM3, GSTPl end GSTT1 genotypes in modulating susceptibility to smoking-related lung cancer. *Pharmacogenetics* 1998;8:495-502.

14. Kisberstis P, Roberts L. It's not just the genes. *Science* 2002;296:685.

15. Kuller LH. Circular epidemiology. *Am J Epidemiol* 1999;150:897-903.

16. Lang M, Pelkonen O. Metabolism of xenobiotics and chemical carcinogenesis. In: Vineis P, Malats N, Lang M, d'Errico A, Caporaso N, Cuzick J, Boffetta P, editors. Metabolic polymorphisms and susceptibility to cancer. Lyon: IARC Press; 1999. p. 13-22. [IARC Scientific Publications, 148]

17. Langholz B, Rothman N, Wacholder S, Thomas DC. Cohort studies for characterizing measured genes. *Natl Cancer Inst Monogr* 1999;26:39-42.

18. Le Marchand L, Sivaraman L, Pierce L, Seifried A, Lum A, Wilkens LR, Lau AF. Associations of CYP1A1, GSTM1, and CYP2E1 polymorphisms with hung cancer suggest cell type specificities to tobacco carcinogens. *Cancer Res* 1998;58:4858-63.

19. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K. Environmental and heritable factors in the causation of cancer – analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000;343:78-85.

20. Lockhart DJ, Winzeler EA: Genomics, gene expression and DNA arrays. *Nature* 2000;405:827-36.

21. Nakachi K, Imai K, Hayashi S, Watanabe J, Kawajiri K. Polymorphisms of the CYP1A1 and glutathione S-transferase genes associated with susceptibility to lung cancer in relation to cigarette dose in a Japanese population. *Cancer Res* 1993;53:2994-9.

22. Nakachi K, Imai K, Hayashi S, Watanabe J, Kawajiri K. Genetic susceptibility to squamous cell carcinoma of the lung in relation to cigarette smoking dose. *Cancer Res* 1991;51:5177-80.

23. Narod SA, Foulkes WD. BRCA1 and BRCA2: 1994 and beyond. *Nat Rev Cancer* 2004;4:665-76.

24. Nyberg F, Hou SM, Hemminki K, Lambert B, Pershagen G. Glutathione S-transferase M1 and N-acetyltransferase 2 genetic polymorphisms and exposure to tobacco smoke in nonsmoking and smoking lung cancer patients and population controls. *Cancer Epidemiol Biomarkers Prev* 1998;7:875-883.

25. Parkin DM, Whelan SL, Ferlay J, Teppo L, Thomas DB. Cancer incidence in five continents. vol. VIII. Lyon: IARC Press; 2003. [IARC Scientific Publications, 155]

26. Peto J. Cancer epidemiology in the last century and the next decade. *Nature* 2001;411:390-395.

27. Rothman N, Wacholder S, Caporaso NE, Garcia-Closas M, Buetow K, Fraumeni Jr JF. The use of common genetic polymorphisms to enhance the epidemiologic study of environmental carcinogens. *Biochim Biophys Acta* 2001;1471:C1-C10.

28. Rothman N, Garcia-Closas M, Stewart WT, Lubin J. The impact of misclassification in case-control studies of gene-interactions. In: Vineis P, Malats N, Lang M, d'Errico A, Caporaso N, Cuzick J, Boffetta P, editors. Metabolic polymorphisms and susceptibility to cancer. Lyon: IARC Press; 1999. p. 89-96. [IARC Scientific Publications, 148]

29. Schliephake H. Prognostic relevance of molecular markers of oral cancer: a review. *Int J Oral Maxillofac Surg* 2003;32:233-45.

30. Susser M. Separating heredity and environment. *Am J Prev Med* 1985;1:5-23.

31. Taningher M, Malacarne D, Izzotti A, Ugolini D, Parodis S. Drug metabolism polymorphisms as modulators of cancer susceptibility. *Mutat Res* 1999;436:227-61.

32. Tokuhata GK, Lilienfeld AM. Familial aggregation of lung cancer in humans. *J Natl Cancer Inst* 1963;30:289-312.

33. Tsugane S, de Souza JM, Costa ML Jr, Mirra AP, Gotlieb SL, Laurenti R, Watanabe S. Cancer incidence rates among Japanese immigrants in the city of Sao Paulo, Brazil, 1969-78. *Cancer Causes Control* 1990;1:189-193.

34. Tuma RS. Narrowing the field: identifying ways to improve SNP association studies. *J Natl Cancer Inst* 2004;96:94-5.

35. Vineis P. A self-fulfilling prophecy: are we understimating the role of the environment in gene-environment interaction researc? *Int J Epidemiol* 2004;33:945-6.

36. Whittemore AS, Nelson LM. Study design in genetic epidemiology: theoretical and practical considerations. *Natl Cancer Inst Monogr* 1999;26:61-9.

37. Wild CP, Law GT, Roman E. Molecular epidemiology and cancer: promising areas for future research in the post-genomic era. *Mutat Res* 2002;499:3-12.

38. Willet WC. Balancing life-style and genomics research for disease prevention. *Science* 2002;296:695-8.

39. Wood ME, Kelly K, Mullineaux LG, Bunn PA Jr. The inherited nature of lung cancer: a pilot study. *Lung Cancer* 2000;30:135-44.

40. Wünsch-Filho V, Boffetta P, Colin D, Moncau JE. Familial cancer aggregation and the risk of lung cancer. *São Paulo Med J* 2002;120:38-44.

41. Xu XP, Kelsey KT, Wiencke JK, Wain JC, Christiani DC. Cytochrome P450 CYP1A1 MspI polymorphism and lung cancer susceptibility. *Cancer Epidemiol Biomarkers Prev* 1996;5:687-92.