7

Models for Determining Genetic Susceptibility and Predicting Outcome

Peter W. Jones, Richard C. Strange, Sud Ramachandran, and Anthony Fryer

1. Introduction

In this chapter, we focus on the study of associations between disease and inheritance of particular genetic variants, a field described as genetic epidemiology. Genetic variation based on polymorphism is common in human populations and appears to be a critical factor in determining susceptibility to disease. Polymorphism describes the presence of variant forms of genes (alleles) that are inhherited from parents. Individuals within a population may therefore inherit none (homozygous wild-type), one (heterozygote), or two (homozygous mutant) copies of the variant allele. These combinations are referred to as genotypes. Many types of allelic variation have been described, including deletions and insertions of DNA bases or even whole genes. Recently, genetic variation derived from single nucleotide polymorphisms (SNPs) — single base changes thought to occur every 500–1000 nucleotides — have attracted considerable interest in the context of disease susceptibility. For the purpose of this review, we will use data collected in our laboratories on ploymorphisms in members of the glutathione S-transferase (GST) supergene family of enzymes (*see* Hayes and Strange ***[1]*** for a recent review).

The number of genetic epidemiology association studies, and the complexity of statistical analysis of the resulting data, have developed exponentially in recent years. Many earlier studies were based on standard cross tabulations (chi-squared analyses) in small numbers of cases and controls, often with case groups being clinically heterogeneous. In many instances, even basic confounding factors such as age at presentation and gender were ignored, and most focused on a single gene ***(2–6)***. Although this type of study still appears in journals, more recent work has identified the need for large, well-characterized

From: *Methods in Molecular Biology, vol. 184: Biostatistical Methods*
Edited by: S. W. Looney © Humana Press Inc., Totowa, NJ

patient cohorts with multivariate modeling to enable consideration of potential clinical heterogeneity within the total population *(7–11)*. This assessment of disease susceptibility continues to be plagued by problems of control selection and, in some cases, recruitment. Furthermore, much of the current success of this approach has highlighted the importance of examining outcome (e.g., disease severity, survival) rather than susceptibility alone (see the following paragraphs for examples).

2. General Modeling Concepts

In statistical terms, in any association study, we are given a series of candidate genes and possible confounders such as demographic (e.g., age, gender) and environmental (e.g., smoking habit, UV exposure) factors. Linear models are derived for determining whether (1) there are genetic components that affect susceptibility to disease and (2) within those who have the disease, outcome is also associated with the presence or absence of genetic risk factors (including gene–gene or gene–environment interactions) in the presence of the other nongenetic factors.

In multiple linear regression a dependent variable, outcome or response, y, is related to a set of independent variables, covariates or factors, $\mathbf{x} = (x_1, x_2, \ldots, x_p)$, where these are considered fixed, through a series of unknown parameters $\beta_1, \beta_2, \ldots, \beta_p$, such that the probability distribution of the random variable y has mean given by $\theta = \alpha + \Sigma \beta_i x_i$ and usually constant variance. A further assumption that is commonly made is that the y's are normally distributed. Some of the independent variables could be squares or cross-products of a smaller set of variables, so, for example, a linear regression can be constructed in which there is only one independent variable, x, and $\theta = \alpha + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$.

McCullagh and Nelder *(12)* have extended this regression model to the generalized linear model (GLM) to deal with situations where the observations, y, are not continuous or normal. However, these models do have a common assumption that their mean, θ, is linear in the unknown parameters. These ideas are exploited in this chapter to develop three linear models, one for measuring susceptibility effects and the others for exploring disease outcome. We have applied these modeling concepts to specific examples taken from a study of the effects of GST polymorphisms on skin cancer risk in renal transplant patients **(Table 1)**. All models produce estimates of genetic effects that may be interpreted as assessments of risk associated with genotypes but that are adjusted for other baseline characteristics.

3. Susceptibility

3.1. Clinical Considerations

Traditionally, case-control studies have compared the proportions of genotypes in a sample of cases with those in a sample of controls (often laboratory

Table 1
Study Examining the Association of Two Genes, *GSTM1* and *GSTP1*, with Risk of Cutaneous Squamous Cell Carcinoma (SCC) in Renal Transplant Patients

(a) Risk of any SCC (logistic regression)
logistic sccyn sex agetx M1n

Variable	OR	95% CI	*p* value
sex	3.8	1.0, 15.1	0.058
agetx	1.063	1.021, 1.106	0.003
M1n	3.2	1.1, 9.5	0.040

(b) Numbers of SCC lesions (negative binomial regression)
nbreg scc_no sex agetx P1AA, exposure (folup) irr

Variable	RR	95% CI	*p* value
sex	16.6	2.7, 101.7	0.002
agetx	1.081	1.035, 1.129	< 0.001
P1AA	6.9	2.1, 22.8	0.002

Likelihood ratio test of $\gamma = 0$ (overdispersion test): $\chi^2_1 = 86.39$, $p < 0.001$

(c) Time from transplantation to development of the first SCC (Cox's proportional hazards regression)
cox time sex agetx scor3 M1n M1nscor3, dead(censor)

Variable	HR	95% CI	*p* value
sex	3.7	0.9, 15.3	0.066
agetx	1.140	1.071, 1.213	< 0.001
scor3	0.17	0.03, .99	0.048
M1n	0.1	0.02, 0.9	0.041
M1nscor3	42.6	3.8, 479.7	0.002

Key

logistic	Stata command for logistic regression
sccyn	1 = SCC present, 0 = SCC absent
sex	1 = male, 0 = female
agetx	age at transplantation in years
M1n	1 = null genotype at *GSTM1* gene, 0 = other *GSTM1* genotypes
nbreg	Stata command for negative binomial regression
scc_no	number of SCC tumors
P1AA	1 = *AA* genotype at *GSTP1* gene 0 = other *GSTP1* genotypes
exposure	Stata command allowing normalization for exposure
folup	length of follow-up
irr	Stata command indicating display of incidence rate ratio
cox	Stata command for Cox's proportional hazards regression
time	time between transplantation and development of SCC (or last follow-up)
scor3	1 = sunbathing score ≥ 3, 0 = sunbathing score < 3
M1nscor3	1 = sunbathing score ≥ 3 and null genotype at *GSTM1* 0 = other genotype/sunbathing score combinations
dead	Stata command that specifies censor variable
censor	censor variable; 1 = SCC, 0 = no SCC

volunteers) by using chi-squared (χ^2) tests. For example, in an early study, we examined the frequency of genotypes at the *GSTM1* locus in hospital controls and patients with skin cancer using this approach ***(13)***. These data showed that the proportion of patients with the *GSTM1* null genotype was significantly higher in patients with multiple skin cancers of different histological types (32/45, 71%) than in controls (79/153, 52%, $p = 0.033$). From these data, it is possible to obtain an assessment of the degree of risk imparted by this genotype by calculation of the odds ratio (OR = 2.3) and its corresponding confidence interval (95% CI = 1.1–4.8). These data can then also be used to assess the relative importance of the gene in determining susceptibility by attributable risk calculations ***(14)***.

The importance of selection of suitable controls is demonstrated by studies of the sulfotransferase, *SULT1A1*, gene. A polymorphism in this gene has been shown to exhibit differences in genotype frequencies with increasing age in healthy controls ***(15)***. It is therefore critical when studying associations between polymorphisms that cases be age-matched (preferably one-to-one, but at least overall) to controls, thereby necessitating more advanced statistical approaches (*see* **Subheading 6.**).

More recently, case-control data have been handled using more sophisticated statistical approaches, including examining the interactions between genetic and environmental factors (discussed in detail in ***[16]***), with correction for confounding using multivariate models. For example, GST have traditionally been viewed as carcinogen detoxifying enzymes and several studies have examined the effect of GST genotype on lung cancer risk in smokers compared with nonsmokers ***(7,17)***. In addition, the use of multivariate models to correct for confounding factors can provide useful information on the potential mechanism of the observed genetic association ***(11,18)***. This is illustrated in **Table 1** (a) in which, using logistic regression analysis, the association of GSTM1 genotype with squamous cell carcinoma (SCC) risk in renal transplant recipients is corrected for the potential confounding effect of age at transplantation and gender, both of which are known SCC risk factors ***(19)***.

3.2. Modeling Aspects

Models for susceptibility are developed using data from case-control (retrospective) studies (*see* Schlesselman ***[20]*** and Breslow and Day ***[21]*** for a comprehensive discussion of this type of study). In these models, the dependent variable is binary, with $y = 0$ representing the controls and $y = 1$ representing the cases in a logistic model.

This model makes the assumption that the probability of being a case, p, is related to $\mathbf{x}$ through the log odds or logit transformation given by

$$\text{logit}(p) = \log[p/(1-p)] = \theta = \alpha + \Sigma\, \beta_i\, x_i$$

Suppose that we wish to estimate the effect on y of a change in a binary variable, x_k from 0 to 1. (In the context of genetic susceptibility, x_k represents the presence or absence of a genotype.) If all other x's remain the same, the parameter β_k is:

$$\beta_k = \log \text{odds}\,(x_k = 1) - \log \text{odds}\,(x_k = 0) = \log[\text{odds}\,(x_k = 1)/\text{odds}\,(x_k = 0)]$$

or the "log odds ratio." Rewriting, the odds ratio, which is a measure of the relative risk of being a case when an individual possesses a genotype, is e^{β_k}. The parameter β_k is estimated in the presence of all the other x's and may therefore be interpreted as a relative risk corrected for all other x's, unlike the uncorrected values in the previous subheading. These x's will have an effect on the estimate of risk and will be dealt with by the estimation procedure but may call into question the make up of the control group. In most disease association studies, a younger control group could become cases over time. Matching controls on key independent variables one-to-one with the cases is a possibility; this would lead to estimation using conditional as opposed to the unconditional logistic regression described above.

There are problems with applying this logistic model and indeed with the interpretation of simple odds ratios from 2 × 2 tables in case-control studies. As pointed out in Breslow and Day (***21***, Section 6.3.), the subjects are selected on the basis of whether they are cases or not, hence y cannot be regarded as a random variable while the x's, determined retrospectively, are random variables. This is the opposite of the usual situation. However, Breslow and Day show that, in most cases, the method of estimation of the logistic parameters which would be applied under the assumption that y is random or the data are generated from a cohort, or prospective, study gives similar numerical values of the parameter estimates and relative risks.

Table 1 illustrates the estimation of odds ratios (OR) for the GSTM1 null genotype using logistic regression modeling. The value of the odds ratio provides a quantitative estimate of the relative impact of the genetic effect on disease susceptibility. Many case-control studies on single genes in complex disorders such as cancer have often generated only modest odds ratios (e.g., 2.0–3.0). This suggests that other parameters must be taken into consideration. These include (1) the interaction of multiple genetic factors with each other and with environmental factors such as smoking, diet, or UV exposure and (2) the effect of genetic factors in genetically high risk subgroups diluted by a large number of cases with low genetic risk.

4. Outcome

4.1. Clinical Considerations

Studies in the literature on the same gene in the same disease group often produce conflicting data. For example, several groups have examined the

importance of polymorphism in the GSTT1 gene in mediating susceptibility to colon cancer. Chenevix-Trench et al. ***(22)*** found that the frequency of the GSTT1 null genotype was significantly increased in patients diagnosed before 70 yr of age, while Deakin et al. ***(23)*** showed the frequency of this genotype was increased in cases compared with controls, although no age effect was observed. In contrast, Gerdig et al. ***(24)*** and Katoh and Bell ***(25)*** failed to show any significant association. These studies have highlighted the role of clinical heterogeneity, as recent data from our laboratory suggest that these discrepancies may result from differences in the proportion of patients with advanced disease in the different study populations. Thus, it is possible that, for example, GSTT1 null genotype is associated with poor outcome in these patients and, consequently, its frequency is increased in patients with advanced tumors (as reflected in Dukes' stage, tumor site, or age at onset). Therefore, the relative proportions of patients with advanced vs early disease in a sample of cases may determine whether the frequency of the polymorphism differs significantly between the cases and controls.

These observations have led to many studies examining the effect of so-called ''modifier genes'' on disease outcome. In general, these studies have met with significantly greater success that those examining susceptibility, with larger odds ratios (or other effect sizes). For example, **Table 1** (b),(c) show the association of *GST* genotype with both number of tumors and time between transplantation and appearance of the first SCC in renal transplant recipients. Number of tumors was examined using negative binomial regression, which, using the Stata statistical software package, allows simultaneous normalization for follow-up time (or degree of exposure) as well as correction for potential confounding factors. In this case, *GSTP1 AA* genotype has a rate ratio *(RR)* of 6.9. This represents a very large increase in risk, which may have clinical, as well as statistical, significance.

In the data described in **Table 1** (c), time between transplantation and appearance of the first SCC in renal transplant recipients was modeled using Cox's proportional hazards regression, an approach often used in survival analysis ***(7,26)***. This example illustrates the examination of a gene–environment interaction between *GSTM1* null genotype and a high sunbathing score. This suggests that individuals who have high UV exposure and are *GSTM1* null genotype demonstrate a markedly reduced time from transplantation to development of their first SCC. Interestingly, when *GSTM1* null and sunbathing score were considered individually, neither was significant. This illustrates that an interaction cannot be predicted from the individual effects, a phenomenon referred to as *epistasis* ***(27)***.

Thus, examination of associations between polymorphic variants and outcome may require the application of other types of statistical modeling. Indeed, we have used several approaches for the analysis of trends (ordered

logistic regression) ***(10)***, count data (Poisson regression and negative binomial regression) ***(28,29)***, and time until an event (Cox's proportional hazards regression) ***(18,30)***. We discuss some of the modeling aspects in more detail in the following subheading.

4.2. Modeling Aspects

In this subheading, we focus on the two modeling approaches illustrated in **Table 1** (b) and (c) (i.e., the accrual rate of tumors over time and the time to the next tumor) in patients with differing follow-up times. The methods may be applied to any situation where the dependent variable is a count or the time between two events.

When the data are in the form of counts and it is required to model associations with a set of independent predictors, there are a number of alternatives. Linear regression could be used with the square root of the counts as the dependent variable, or either Poisson or negative binomial regression could be used, where the means of the probability distributions are again a linear function of unknown parameters, some of which may be interpreted as measuring risk; we will concentrate on the latter two methods (*see **[31]*** for an application of these models to tumor counts). In each of these models**,** the time of exposure or follow-up will be a determinant of the eventual count and will have to be controlled for in the final estimation procedure.

In Poisson regression, the rate at which incidents (tumors) occur, usually termed the incidence rate, is assumed to be

$$\lambda = e^{\theta} = e^{\alpha + \mathrm{S}\beta_i x_i}$$

for an individual with independent variable vector $\boldsymbol{x}$. It follows that the mean of the number of incidents will be λT, where T is the exposure of the individual. To measure the effect of x_k changing from 0 to 1 (or a unit increase in x_k) where all other x's remain the same, the incidence rate ratio *(irr)* may be calculated as (incidence rate when $x_k = 1$)/(incidence rate when $x_k = 0$). It follows that $irr = e^{\beta_k}$. This may now be used as a way of evaluating the risk, in the sense of the effect of a change in the value of x_k producing a decrease or increase in the mean number of incidents, since for the same exposure, T, the *irr* is the ratio of the (corrected) number of incidents in the two groups (those with $x_k = 1$ and those with $x_k = 0$). This may be applied directly to the skin cancer study in **Table 1** to evaluate the effect of the presence of a genotype, in the presence of other covariates, on tumor incidence.

The Poisson distribution has the property that the mean and variance are equal and so if samples of the number of incidents in a study produce means and variances which differ, then this suggests that the assumption of a Poisson

model for the counts is not tenable. More formally, a χ^2 goodness of fit test may be used to determine whether the model is suitable. In the situation where the variance is larger that the mean (or the process is exhibiting extra Poisson variation or overdispersion; underdispersion is rare), the negative binomial regression model can be used. This is a modification of the Poisson where the mean is multiplied by a random variable, z, whose distribution depends on a single overdispersion parameter γ (*see* **Table 1** [b]). This leads to a mean of $z\lambda T$.

If the measure of severity is the time between presentation with the first tumor and occurrence of the next tumor, then methods in survival analysis are appropriate. These methods have been developed to use data on all individuals, including those where the second event has not yet occurred but where a follow-up time is available. The parameter of interest here is the hazard rate, $\lambda(t)$, which measures the risk of an incident occurring at a particular point in time; another interpretation is that it is the instantaneous incident rate. In many practical situations this rate will vary with time. A full exposition of survival analysis may be found in Parmar and Machin ***(32)***, which includes many clinical examples. They give an example of a hazard function and its relationship with time for measuring the risk of infant mortality, which is known to be highest just after birth but thereafter declines rapidly.

Cox's proportional hazards regression (Parmar and Machin ***[32],*** Cox ***[33]***), may be used to assess the impact of genotypes on time to a further event, adjusted for other x's. It is assumed that the hazard takes the form

$$\lambda(t) = \lambda_0(t)\exp(\Sigma\beta_i x_i),$$

where $\lambda_0(t)$ is interpreted as an underlying hazard for all individuals that is adjusted by the x's to give the hazard for a individual. In the linear form employed above $\alpha = \ln\lambda_0(t)$. The risk associated with a change of x_k from 0 to 1, with all other x's unchanged, is measured by the hazard ratio (HR) given by (hazard when $x_k = 1$)/(hazard when $x_k = 0$), which is easily seen to be e^{β_k}. In **Table 1** (c), we present HRs for two main effects and an interaction. The main assumption behind the Cox model is that the hazards are proportional; this is usually tested by a visual inspection of graphs of the estimated hazard functions in the two groups based on the values of x_k. If these are approximately parallel, then it suggests that the assumption is reasonable.

Because all models in this section depend on a linear function of covariates then it is possible to use the estimates of these functions to derive prognostic indices or simple scoring models for outcome. Christiansen ***(34)*** gives a detailed description of using Cox's regression to derive a prognostic index (PI) for an individual with a given set of x's.

5. Software

Most commercial statistics packages include routines for the analysis of linear, logistic, and Cox's proportional hazards regression models. However, few offer the possibility of fitting Poisson and negative binomial regression models. Stata (Stata Corporation, College Station, TX) offers a programmable package that includes these routines; all the analyses presented here were performed using this package and **Table 1** gives the command code necessary to generate the results.

6. Extensions to the Basic Modeling Approach

Most of the examples given in this chapter and the Stata code in **Table 1** refer to only a single genotype. It is straightforward to extend this to large multivariate models with other genotype main effects and interactions between genotypes. Furthermore, most packages will allow the user to specify stepwise selection of predictors. This may be used in cases where it is useful to allow all predictors to compete to enable a best subset of them to be chosen.

This could be especially useful in obtaining an estimate of the relative importance of genetic *vis à vis* environmental factors. However, care should be exercised in using these routines, especially where there are missing values in the data, as inferences could be based on only a small percentage of the observations. In this case, we suggest refitting the model with all the data using the reduced set of predictors.

We have briefly discussed the difficulty of obtaining suitable controls and the need to match or control for key variables when looking at susceptibility. However, this does not address the problem of population stratification in which the genetic background of the cases may be different from that of the controls. One method that is used in genetic studies that reduces the impact of this effect is to use a family-based study design such as that employed in transmission disequilibrium testing (TDT; *see* **[ref. *35***, Section 4.7] for more details). This approach compares the proportions of transmitted and nontransmitted alleles in parents and affected offspring, thereby correcting for the potential effect of non-disease-associated genetic differences between cases and controls including ethnic, geographical, and in some cases exposure differences.

It is worth noting that the studies described in this chapter are association studies. Thus, it is possible that a genotype demonstrates a strong association with risk without directly affecting the disease process at all. Indeed, the genetic marker may simply be coinherited with a neighboring gene that is critical to the pathological process.

References

1. Hayes, J. D. and Strange, R. C. (2000) Glutathione S-transferase polymorphisms and their biological consequences. *Pharmacology* **61,** 154–166.
2. Davies, M. H., Elias, E., Acharya, S., Cotton, W., Faulder, G. C., Fryer, A. A., and Strange, R. C. (1993) GSTM1 null polymorphism at the glutathione S-transferase M1 locus: phenotype and genotype studies in patients with primary biliary cirrhosis. *Gut* **34,** 549–553.
3. Yu, M-W., Gladek-Yarborough, A., Chiamprasert, S., Santella, R. M., Liaw, Y-F., and Chen, C-J. (1995) Cytochrome P450 2E1 and glutathione S-transferase M1 polymorphisms and susceptibility to hepatocellular carcinoma. *Gastroenterology* **109,** 1266–1273.
4. Katoh, T., Inatomi, H., Nagaoka, A., and Sugita, A. (1995) Cytochrome P4501A1 gene polymorphism and homozygous deletion of the glutathione S-transferase M1 gene in urothelial cancer patients. *Carcinogenesis* **16,** 655–657.
5. El-Zein, R. A., Zwischenberger, J. B., Abdel-Rahman, S. Z., Sankar, A. B., and Au, W. W. (1997). Polymorphism of metabolising genes and lung cancer histology: prevalence of CYP2E1 in adenocarcinoma. *Cancer Lett.* **112,** 71–78.
6. Nazar-Stewart, V., Motulsky, A. G., Eaton, D. L., White, E., Hornung, S. K., Leng, Z. T., et al. (1993) The glutathione s-transferase-mu polymorphism as a marker for susceptibility to lung-carcinoma. *Cancer Res.* **53,** 2313–2318.
7. Goto, I., Yoneda, S., Yamamoto, M., and Kawajiri, K. (1996) Prognostic significance of germ line polymorphisms of the *CYP1A1* and glutathione S-transferase genes in patients with non-small cell lung cancer. *Cancer Res.* **56,** 3725–3730.
8. Brockmoller, J., Cascorbi, I., Kerb, R., and Roots, I. (1996) Combined analysis of inherited polymorphisms in arylamine N-acetyltransferase 2, glutathione S-transferases M1 and T1, microsomal epoxide hydrolase, and cytochrome P450 enzymes as modulators of bladder cancer risk. *Cancer Res.* **56,** 3915–3925.
9. Probst-Hensch, N. M., Haile, R. W., Ingles, S. A., Longnecker, M. P., Han, C. Y., Lin, B. K., et al. (1995). Acetylation polymorphism and prevalence of colorectal adenomas. *Cancer Res.* **55,** 2017–2020.
10. Fryer, A. A., Bianco, A., Hepple, M., Alldersea, J., Jones, P. W., Strange, R. C., and Spiteri, M. A. (2000) Polymorphism at the glutathione S-transferase, GSTP1, locus is associated with atopy/airway responsiveness. *Am. J. Respir. Crit. Care Med.* **161,** 1437–1442.
11. Clairmont, A., Sies, H., Ramachandran, S., Lear, J. T., Smith, A. G., Bowers, B., et al. (1999) Association of NAD(P)H:quinone oxidoreductase NQO1 null with numbers of basal cell carcinomas: Use of a multivariate model to rank the relative importance of this polymorphism and those at other relevant loci. *Carcinogenesis* **20,** 1235–1240.
12. McCullagh, P. and Nelder, J.A. (1989) *Generalised Linear Models*, 2nd ed. Chapman and Hall, London.

13. Heagerty, A. H., Fitzgerald, D., Smith, A., Bowers, B., Jones, P., Fryer, A. A., et al. (1994) Glutathione S-transferase GSTM1 phenotypes and protection against cutaneous tumors. *Lancet* **343,** 266–268.
14. Hutchinson, P. E., Osborne, J. E., Lear, J. T., Smith, A. G., Bowers, P. W., Jones, P. W., et al. (2000) Vitamin D receptor polymorphisms are associated with altered prognosis in patients with malignant melanoma. *Clin. Cancer Res.*, **6**, 498–504.
15. Coughtrie, M. W. H., Gilissen, R. A. H. J., Shek, B., Strange, R. C., Fryer, A. A., Jones, P. W., and Bamber, D. E. (1999) The phenol sulfotransferase *sult1a1* polymorphism: molecular diagnosis and allele frequencies in Caucasian and African populations. *Biochem. J*. **337,** 45–49.
16. Fryer, A. A. and Jones, P. W. (1999) Interactions between detoxifying enzyme polymorphisms and susceptibility to cancer, in: *Metabolic Polymorphisms and Cancer, vol. 148* (Boffetta, P., Caporaso, N., Cuzick, J., Lang, M., and Vineis, P. eds.) IARC, Lyon, pp. 303–322.
17. Kihara, M., Kihara, M., and Noda, K. (1995). Risk of smoking for squamous and small cell carcinomas of the lung modulated by combinations of CYP1A1 and GSTM1 gene polymorphisms in a Japanese population. *Carcinogenesis* **16**, 2331–2336.
18. Howells, R. E. J., Redman, C. W. E., Dhar, K. K., Sarhanis, P., Musgrove, C., Jones, P. W., et al. (1998) Association of glutathione S-transferase GSTM1 and GSTT1 null genotypes with clinical outcome in epithelial ovarian cancer. *Clin. Cancer Res.* **4,** 2439–2445**.**
19. Ramsay, H. M., Harden, P. N., Reece, S., Smith, A. G., Jones, P. W., Strange, R. C., and Fryer, A. A. (2000) Polymorphisms in glutathione S-transferases are associated with altered risk of non-melanoma skin cancer in renal transplant recipients: a preliminary analysis. *J. Invest. Dermatol.*, in press.
20. Schlesselman, J. J. (1982) *Case-Control Studies*. Oxford University Press, New York.
21. Breslow, N. E. and Day, N. E. (1980) *Statistical Methods in Cancer Research,* vol. 1: *Case Control Studies*. IARC, Lyon.
22. Chenevix-Trench, G., Young, J., Coggan, M., and Board, P. (1995) Glutathione S-transferase M1 and T1 polymorphisms: susceptibility to colon cancer and age of onset. *Carcinogenesis* **16,** 1655–1657.
23. Deakin, M., Elder, J., Hendrickse, C., Peckham, D., Baldwin, D., Pantin, C., et al. (1996) Glutathione S-transferase GSTT1 genotypes and susceptibility to cancer: studies of interactions with GSTM1 in lung, oral, gastric and colorectal cancer. *Carcinogenesis* **17,** 881–884.
24. Gertig, D. M., Stampfer, M., Haiman, C., Hennekens, C. H., Kelsey, K., and Hunter, D. J. (1998) Glutathione S-transferase GSTM1 and GSTT1 polymorphisms and colorectal cancer risk: a prospective study. *Cancer Epidemiol. Biomarkers Prev*. **7,** 1001–1005.
25. Katoh, T. and Bell, D. A. (1996) Glutathione S-transferase M1 and T1 genetic polymorphism and susceptibility to gastric and colorectal adenocarcinoma. *Proc. Am. Assoc. Cancer Res.* **37,** 257–258.

26. Ramsay, H. M., Fryer, A. A., Smith, A. G., and Harden, P. N. (2000) Clinical risk factors associated with non-melanoma skin cancer in renal transplant recipients. *Am. J. Kid. Dis.* **36,** 167–176.
27. Frankel, W. N. and Schork, N. J. (1996) Who's afraid of epistasis? *Nat. Genet*. **14,** 371–373.
28. Yengi, L., Inskip, A., Gilford, J., Alldersea, J., Bailey, L., Smith, A., et al. (1996) Polymorphism at the glutathione S-transferase, GSTM3 locus: interactions with cytochrome P450 and glutathione S-transferase genotypes as risk factors for multiple cutaneous basal cell carcinoma. *Cancer Res.* **56,** 1974–1977.
29. Ramachandran, S., Lear, J. T., Ramsay, H., Smith, A. G., Bowers, B., Hutchinson, P. E., et al. (1999) Presentation with multiple cutaneous basal cell carcinomas: association of glutathione S-transferase and cytochrome P450 genotypes with clinical phenotype. *Cancer Epidemiol. Biomarkers Prev.* **8,** 61–67**.**
30. Lear, J. T., Smith, A., Heagerty, A. H. M., Bowers, B., Jones, P. W., Gilford, J., et al. (1997) Truncal site and detoxifying enzyme polymorphisms significantly reduce time to presentation of next cutaneous basal cell carcinoma. *Carcinogenesis* **18,** 1499–1503.
31. Drinkwater, N. R. and Klotz, J. H. (1981) Statistical methods for the analysis of tumor multiplicity data. *Cancer Res*. **41,** 113–119.
32. Parmar, M. K. B. and Machin, D. (1995) *Survival Analysis: A Practical Approach*. John Wiley and Sons, Chichester.
33. Cox, D. R. (1970) Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34,** 187–220.
34. Christiansen, E. (1987) Multivariate survival analysis using Cox's regression model. *Hepatology* **7,** 1346–1358.
35. Sham, P. (1998) *Statistics in Human Genetics*. Arnold Press, London.