

Genetic polymorphism and clinical outcome: identification of individuals at risk of a poor clinical outcome

Susceptibility and outcome in complex disorders such as asthma and cancer appear to be determined, at least in part, by genetic polymorphism. However, while our ability to identify new allelic variants and study them in case and control populations has greatly improved, considerable difficulties remain in elucidating how many genes determine particular clinical phenotypes. This is because most studies have concentrated on study of single genes in relatively small study groups. The important issues of gene–gene interactions (epistasis) and high-risk subgroups have not yet been adequately addressed. We now describe a general approach, using patients with head and neck cancers as an example. Our purpose is to demonstrate candidate gene selection, statistical approaches, and identification of patient subgroups.

**R. C. Strange, J. Aldersea,
P. R. Hoban, A. A. Fryer**

Centre for Cell and Molecular Medicine,
School of Postgraduate Medicine, Keele University,
North Staffordshire Hospital, Stoke-on-Trent, UK

C. Matthias, V. Jahnke

Department of Otorhinolaryngology, Virchow-Klinikum,
Humboldt University, Berlin, Germany

P. W. Jones

Department of Mathematics, Keele University,
Staffordshire, UK

Key words: candidate genes; molecular epidemiology;
polygenic diseases; polymorphism.

Prof. R. C. Strange
Centre for Cell and Molecular Medicine
School of Postgraduate Medicine
Keele University
North Staffordshire Hospital
Stoke-on-Trent
UK
Tel. +44 (0) 1782554667
Fax: +44 (0) 1782554646

Polymorphism and polygenic disease

There is accumulating evidence that both individual susceptibility to inflammatory and neoplastic diseases and clinical outcome are mediated by allelic variation (1). However, while the importance of polymorphism in disease pathogenesis is clear, many aspects of how alleles confer altered risk remain unknown. For example, the number of genes that determines a particular clinical phenotype is unknown, as is the nature of the interaction between high- and low-risk alleles (1–3). Obtaining a better understanding of these phenomena will greatly help in the identification of high-risk individuals in screening programmes as well as in determining individual prognosis. We have used a candidate gene approach to study the influence of selected genes on susceptibility and outcome in polygenic diseases (3, 4). Selection of candidates is always problematic although, in the context of diseases such as asthma and cancer in which exposure to environmental pollutants is important, it is logical

to focus on loci involved in the detoxication of endogenous and exogenous toxins (1–9). Thus, various members of the phase I cytochrome P450 (CYP) and phase II glutathione S-transferase (GST) supergene families have attracted particular interest in various cancers, although corresponding studies in asthma have not been reported. Data from various cancers also suggest the importance of genes involved in control of the cell cycle. For example, a polymorphism in cyclin D1 (CCND1) has been associated with outcome in lung and head and neck cancers (10, 11). We now describe the general approach used to assess the involvement of polymorphic genes in determining disease outcome. We have used cancer of the upper aerodigestive tract (SCCHN) as a model of a polygenic disorder with a significant environmental component and have assessed the importance of selected CYP, GST, and CCND1 genes on three outcome parameters, tumour extension, histologic grade, and the presence of nodes (Table 1). This model is also applicable to other

Table 1. Candidate genes

GSTM1
– Substrates include polycyclic aromatic hydrocarbon (PAH) epoxides
– Polymorphism controversially associated with smoking-related cancers
GSTM3
– Overlapping substrate specificities with GSTM1
– Polymorphism in linkage disequilibrium with GSTM1
GSTT1
– Detoxification of PAH epoxides and halomethanes
– Polymorphism associated with colon cancer
GSTP1
– Widely expressed
– Detoxification of PAH epoxides and acrolein
– Association of polymorphism with bladder cancer
– Catalyses oxidation of PAH to epoxide and is inducible by PAH
– 3'-flanking region, m2 allele, associated with inducibility and lung cancer risk in Japanese smokers
– Exon 7, Val allele, associated with increased activity
CYP2D6
– Utilizes tobacco-specific nitrosamine, NNK
– 53 alleles: e.g., intron3/exon 4 splice site, exon 5 base pair deletion, gene deletion
– Associated with smoking-related cancers
CYP2E1
– Utilizes tobacco-specific nitrosamine, NNK
– Induced by alcohol
– Mutation in 5'-flanking region containing the Pst1 and Rsa1 digestion sites (<i>c1c1</i> wild-type homozygotes) and Dra1 RFLP (<i>DD</i> wild type homozygotes)
CCND1
– A/G polymorphism identified using amplification across the exon 4/intron 4
– This polymorphism appears to modulate splicing of mRNA
– Overexpression of <i>CCND1</i> associated with genetic instability

complex multigene disorders, and studies are underway in our laboratories to determine the relevance of these allelic variants in the context of asthma.

Cancers of the upper aerodigestive tract comprise about 5% of new cancers in Western countries and offer an important opportunity to study interactions between genes and the environment (12). Thus, chronic tobacco and alcohol consumption is a dose-dependent, causative factor, although not all patients have been exposed and many individuals with excessive consumption do not develop these lesions. This provides a model for the assessment of the influence of both individual genotypes and gene–gene interactions on outcome. Such studies involve the examination of several genetic variants in generally large case groups, and it is possible to generate stepwise models that include all the genes to determine, on the basis of *P* value and rate ratio, the relative importance of each genotype. A further factor is the general heterogeneity of pathologic mechanisms within such complex diseases as asthma and cancer. Thus, several clinically distinct subgroups may exist within a disease population. Since, in the context of head and neck cancer, the influence of polymorphisms on susceptibility varies in different sites (6–9), we separately analysed data from cases with oral cavity/pharyngeal and laryngeal squamous cell carcinoma (SCC). These studies demonstrate the ability of molecular epidemiologic studies to identify associations

between allelic variants and disease phenotypes but also the problems inherent in the approach: difficulties in obtaining adequate numbers of cases (particularly in subgroups), multiple testing, and the complex nature of gene–gene interactions (epistasis).

Patient recruitment and characterization

The assessment of multiple genetic polymorphisms in complex diseases such as asthma and cancer requires the recruitment of large numbers of well-characterized patients. Indeed, many genetic epidemiologic studies have failed to recruit sufficient numbers of patients even for the examination of the effect of individual genes. This is further compounded when the total patient population is subdivided into relevant subgroups and when gene–gene interactions are considered. In the study of genetic polymorphisms in head and neck cancer, we recruited 384 unrelated German Caucasians suffering one histologically confirmed oral cavity/pharyngeal (*n* = 125) or laryngeal SCC (*n* = 259) (8, 9, 11). They were recruited in 1994–6 and, importantly, comprised 91% of the cases seen during this period. Thus, no patient subgroups were omitted. Data on smoking history and alcohol consumption, the major environmental factors, were also collected.

Although many studies have examined associations between allelic variation and disease susceptibility by case-control approaches, there is increasing evidence that genetic polymorphism may be critical in modifying prognosis. Indeed, as described elsewhere in this volume, we have identified such genetic associations with severity of asthma by bronchial hyperresponsiveness. In head and neck cancer, relevant outcome measures included initial tumour extension (T1–T4) recorded with the TNM system at diagnosis, histologic differentiation (G1–G3), and lymph-node involvement of the neck.

Identification of genotypes

The use of polymerase chain reaction (PCR)-based methods has led to the rapid genotyping of multiple loci in large numbers of DNA samples. Indeed, newer technologies using lower ($\leq 5 \mu\text{l}$) reaction volumes allow many PCR reactions to be performed on DNA extracted from a single blood sample. In our cohort of head and neck cancer cases, we have obtained data on polymorphisms in the glutathione S-transferases, GSTM1, GSTT1, GSTP1, GSTM3, CYP2D6, CYP1A1, CYP2E1, and cyclin D1 (8, 9, 11).

Statistical approaches in multigene analysis

Many early genetic association studies used chi-square tests to examine homogeneity between cases and

controls. Clearly, many potential confounding factors exist in multifactorial diseases such as asthma. These include patient characteristics (age, sex) and markers of environmental exposure (exposure to aeroallergens, smoking, and drinking histories). The use of logistic regression analyses enables examination of associations between genetic factors with both susceptibility and outcome while simultaneously correcting for potential confounding factors. We have also used regression models to examine associations with nonbinary end points such as survival times (Cox's proportional hazards regression), number of lesions (negative binomial regression), and continuous variables such as age of onset or IgE level (linear regression with transformation to normality).

These regression models are also useful in the assessment of gene–gene interactions. Thus, in the context of head and neck cancer, we examined the data for pairwise gene–gene interactions between putative risk genotypes; *GSTT1 null*, *CCND1 GG*, *CYP2D6 PM*, *GSTM1 A* (or *GSTM1 null*), *GSTP1 Ile¹⁰⁵/Val¹⁰⁵*, *CYP2E1 c1c1* and *CYP2D6 EM*. The identification of a significant interaction term (subjects with both of the two risk genotypes, e.g., *GSTT1 null* and *CCND1 GG*) may not necessarily indicate that the two genes are synergistic. Indeed, they may influence risk via independent mechanisms. Thus, it is important to examine the interaction term by inclusion of the main effects (individual risk genotypes, e.g., *GSTT1 null* and *CCND1 GG* separately) in the regression model.

The relative importance of genes and clinical characteristics can be assessed by a stepwise routine which allows all genes and characteristics to compete in a regression model, thereby identifying the best set of predictors for each susceptibility/outcome parameter. Thus, all the genes were included in determining association with tumour grade, differentiation, and presence of nodes, together with age at diagnosis, sex, alcohol and tobacco consumption, and tumour site (2).

Relationship between clinical end points

Since the outcome parameters themselves may not be independent of each other, there is likely to be some overlap between the genes that are associated with each end point. For example, as shown elsewhere in this volume, we have identified associations between the *GSTP1* genotype and both atopic status and bronchial hyperresponsiveness. These measures are related to atopic individuals being more likely to be hyperresponsive. In our population of head and neck cancer patients, we confirmed that undifferentiated tumours (G3) were more likely to demonstrate lymph-node involvement ($P < 0.0001$) even if initial surgery was curative. Furthermore, patients with T3/T4 tumours

were more likely to suffer lymph-node metastasis ($P < 0.0001$) and undifferentiated tumours ($P < 0.0001$). However, these clinical markers represent different aspects of tumour behaviour, and therefore it is possible for some genes to be specific to one end point. These aspects are illustrated in our findings by assessment of both individual genotypes (Fig. 1) and stepwise multi-gene analyses in head and neck cancer.

Stepwise analysis in the head and neck cancer example

We used a stepwise model to determine the relative influence of the genotypes in predicting each of the outcome parameters. The model included several genetic polymorphisms, clinical characteristics, and exposure factors: *GSTM1 null*, *GSTT1 null*, *GSTM3 AA*, *GSTP1 Ile¹⁰⁵/Ile¹⁰⁵*, *CYP2D6 PM*, *CYP1A1 Ile/Val* and *Val/Val*, *CYP1A1 m1m1*, *CYP2E1 c1c1*, *CYP2E1 DD*, *CCND1 GG*, age at diagnosis, sex, tumour site, and alcohol and tobacco consumption. As illustrated below, the influence of the genetic factors differs from that shown for individual genes (Fig. 1) since they may demonstrate some associations with other factors in the model (i.e., not be completely independent risk factors). Thus, ethanol consumption was associated with T3/T4 tumours ($P = 0.003$), and tumour site (laryngeal SCC) with T1/T2 tumours ($P = 0.001$). *CCND1 GG* was the only genotype that was significantly associated with poor histologic differentiation ($P = 0.04$). Tumour site (laryngeal SCC) was also significantly associated with G1/G2 tumours. In a model predicting the presence of nodes, *CYP2D6 PM* and *GSTM1 null* were significant ($P = 0.006$ and $P = 0.049$).

Conclusions

As with many polygenic disorders, while the outcome parameters studied in the example given are related, they are not likely to be similarly associated with genetic factors. Differentiation describes the extent that tumour cells morphologically and functionally resemble normal cells. Metastases are a marker of malignancy, although cancers are heterogeneous with regard to metastatic potential. Thus, while large lesions with metastatic spread are traditionally considered to have a worse prognosis than small, locally restricted tumors, outcome can vary markedly among cases with tumors from the same site, stage, nodal status, and histologic grade.

We identified individual genes that were associated with outcome. *GSTT1 null* was associated with T3/T4 and G3 lesions in the oral/pharyngeal, but not laryngeal, SCC cases. *CCND1 GG* was associated with G3 tumors in the oral/pharyngeal and laryngeal SCC cases. We also used a stepwise routine to obtain the best set of predictors for each outcome parameter

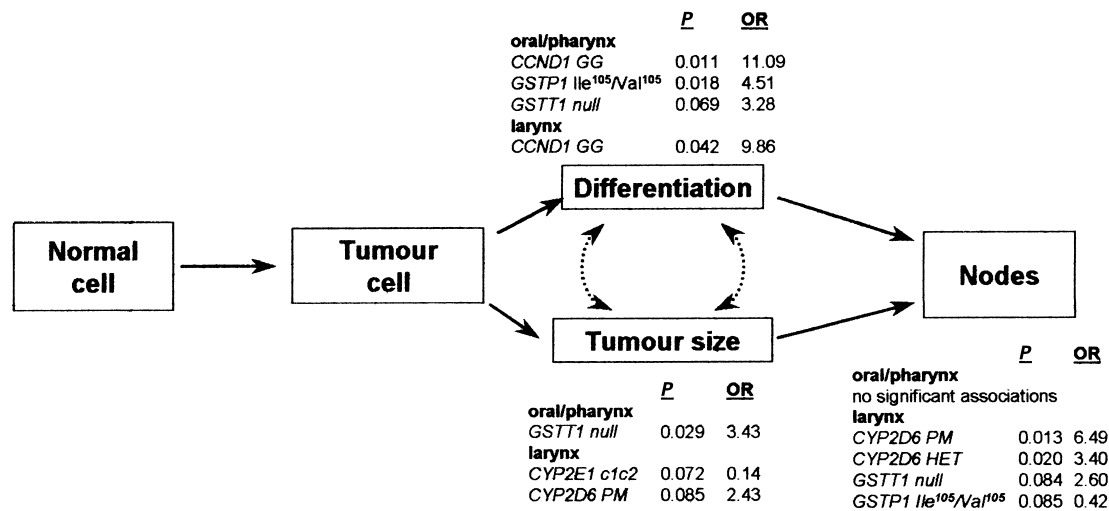


Figure 1. Relationship between outcome parameters in oral/pharyngeal and laryngeal SCC.

and to allow assessment of each genotype within this set. *GSTT1 null*, *CCND1 GG*, and *CYP2D6 PM* were associated with the appropriate parameter. Associations between outcome parameters and other genes were also significant, or approached significance. The data suggest that the *CYP2E1c2* allele is protective. *CYP1A1*, although a good candidate, was not associated with the outcome parameters. *GSTP1 Ile¹⁰⁵/Val¹⁰⁵* was associated with G3 tumours in oral/pharyngeal SCC cases. Associations between *GSTM1* genotypes and outcome were also identified. While most studies have considered *GSTM1*0* homozygotes to be at risk, the few studies that have identified *GSTM1 AB*, *GSTM1 A*, and *GSTM1 B* suggest these genotypes are not equally protective (4). Thus, in SCCHN cases, *GSTM1 AB*, but not *GSTM1 A* or *GSTM1 B*, is protective (9). Indeed, in some other cancers, *GSTM1 B* is associated with increased risk compared with *GSTM1 null* (4). In this study, *GSTM1 null* was associated with the absence of nodes in the stepwise routine. These data may reflect linkage disequilibrium between the mu *GST* genes. Thus, *GSTM1*A* is in linkage disequilibrium with *GSTM3*B*. However, we did not identify links between *GSTM1 B* or *GSTM3* genotypes and outcome.

The associations between *GST* and *CYP* genotypes and outcome presumably reflect their influence on the detoxication of mutagens. Indeed, associations between allelic variants and the formation of mutations in target genes have been reported in several cancers. Importantly, chronic tobacco and alcohol consumption is associated with a high frequency of p53 mutations in SCCHN. A further related possibility is that *GST* genes influence genomic stability. Thus, *GSTT1 null* is linked with a higher background rate of sister chromatid exchange. The association of *CYP2D6 PM* and *HET*

with laryngeal, but not oral/pharyngeal, SCC complements data showing that these genotypes are linked with the time to cervical lymph-node metastases in oral cavity SCC patients (7). *CCND1 GG* was associated with G3 tumours. Since the *G* allele splices less of the *b* transcript than the *A* allele, individuals with *CCND1 GG* may have different cellular levels of cyclin D1 protein from subjects with *CCND1 AA*. Such differences may be critical, since cells overexpressing this protein demonstrate loss of control of the G1-S checkpoint.

The genetic basis of complex traits such as susceptibility and outcome is unclear (3). Recent studies emphasize the importance of epistasis, which describes an interaction between genes that could not be predicted from the sum of the effects of the separate genes. We considered the effects of pairwise interactions between the genes on each outcome parameter. No interactions were found. We have studied associations between eight genes and three outcome measures in two sites in the upper aerodigestive tract. A large number of significance tests have been performed. Clearly, some of the observed associations could arise from multiple testing. To obtain significant results after such correction would require thousands of cases. Our data on head and neck cancer provide a model that is equally applicable to many inflammatory diseases such as asthma as well as malignant diseases. It emphasizes the effect of heterogeneity within the case group on the significance of the results obtained and the importance of statistical considerations such as sample size, multi-gene modelling, gene-gene interactions, and multiple testing. These factors should be investigated if genetic epidemiologic studies in multifactorial disorders are to provide reproducible results.

References

1. STRANGE RC, FRYER AA. The glutathione S-transferases: influence of polymorphism on susceptibility to cancer. In: BOFFETTA P, CAPORASO N, CUZICK J, LANG M, VINEIS P, editors. *Metabolic polymorphisms and cancer*. Lyon: IARC Scientific Publications, 1999;148:231–249.
2. CLAIRMONT A, SIES H, RAMACHANDRAN S, et al. 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* 1999;20:1235–1240.
3. CHAKRAVARTI A. Population genetics – making sense out of sequence. *Nat Gen* 1999;21:56–60.
4. RAMACHANDRAN S, LEAR JT, RAMSAY H, et al. Presentation with multiple cutaneous basal cell carcinomas: association of glutathione S-transferase and cytochrome P450 genotypes with clinical phenotype. *Cancer Epidemiol Biomarkers Prev* 1999;8:61–67.
5. TRIZNA Z, CLAYMAN GL, SPITZ MR, GOEPFERT H. Glutathione S-transferase genotypes as risk factors for head and neck cancer. *Am J Surg* 1995;170:499–501.
6. JOURENKOVA N, REINIKAINEN M, BOUCHARDY C, DAYER P, BENHAMOU S, HIRVONEN A. Larynx cancer risk in relation to glutathione S-transferase M1 and T1 genotypes and tobacco smoking. *Cancer Epidemiol Biomarkers Prev* 1998;7:19–23.
7. WORRALL SE, CORRIGAN M, HIGH A, et al. Susceptibility and outcome in oral cancer: preliminary data showing an association with polymorphism in cytochrome P450 *CYP2D6*. *Pharmacogenetics* 1998;8:433–439.
8. MATTHIAS C, BOCKMÜHL U, JAHNKE V, et al. The glutathione S-transferase *GSTP1* polymorphism: effects on susceptibility to oral/pharyngeal and laryngeal carcinomas. *Pharmacogenetics* 1998;8:1–6.
9. MATTHIAS C, BOCKMÜHL U, JAHNKE V, et al. Polymorphism in cytochrome P450 *CYP2D6*, *CYP1A1*, *CYP2E1* and glutathione S-transferase, *GSTM1*, *GSTM3*, *GSTT1* and susceptibility to tobacco-related cancers: studies in upper aerodigestive tract cancers. *Pharmacogenetics* 1998;8:91–100.
10. BETTICHER DC, THATCHER N, ALTERMATT HJ, HOBAN P, RYDER WDJ, HEIGHWAY J. Alternate splicing produces a novel cyclin D1 transcript. *Oncogene* 1995;11:1005–1011.
11. MATTHIAS C, BRANIGAN K, JAHNKE V, et al. Polymorphism within the cyclin D1 Gene is associated with prognosis in patients with squamous cell carcinoma of the head and neck. *Clin Cancer Res* 1998;4:2411–2418.
12. COPPER MP, JOVANOVIĆ A, NAUTA JJP, et al. Role of genetic factors in the etiology of squamous cell carcinoma of the head and neck. *Arch Otolaryngol* 1995;121:157–160.