Glycoconjugate Journal (1997) **14:** 107–111

MUC1 gene polymorphism and gastric cancer – an epidemiological study

F. Carvalho[1], R. Seruca[1], L. David[1], A. Amorim[1,2], M. Seixas[1], E. Bennett[3], H. Clausen[3] and M. Sobrinho-Simões[1]*

[1]*Institute of Molecular Pathology and Immunology of the University of Oporto, IPATIMUP, Dept. of Pathology, Hospital de S. João, 4200 Porto, Portugal*
[2]*Institute of Anthropology, Faculty of Sciences of the University of Oporto, 4000 Porto, Portugal*
[3]*Department of Oral Diagnostics, School of Dentistry, Faculty of Health Sciences, University of Copenhagen, 2200 Copenhagen, Denmark*

Gastric carcinoma is a major cause of cancer death worldwide and, like most human cancers, probably develops after environmental insults acting on normal individuals and/or individuals with increased genetic susceptibility. Mucins are attractive molecules to study the relationship between genetics and environment because they play an important role in the protection of gastric mucosa against environmental insults and exhibit a highly polymorphic genetic variation. We performed a case-control study using Southern blot analysis to evaluate the *MUC1* gene polymorphism in a series of blood donors (*n*=324) and in patients with gastric carcinoma (*n*=159). We found that the distribution of *MUC1* alleles is significantly different in the two populations and that small *MUC1* alleles and small *MUC1* genotypes are significantly more frequent in patients with gastric carcinoma than in controls. Individuals with small *MUC1* genotypes are at increased risk for gastric carcinoma development.

***Keywords*: *MUC1* gene polymorphism, gastric cancer**

Introduction

Gastric carcinoma is a major cause of cancer death worldwide [1] and, like most human cancers, probably develops after environmental insults acting on normal individuals and/or individuals with increased genetic susceptibility [2, 3]. The relative contribution of environmental exposure and genetics for the risk of developing gastric carcinoma is far from being established. Diet and infections, with particular emphasis on *Helicobacter pylori* infection, have been identified as exposure risk factors [4, 5]. In contrast to most Western countries, the mortality rates from gastric carcinoma have not declined in Portugal [6] raising the possibility that the Portuguese population may have some particular genetic susceptibility for gastric carcinoma development.

Mucins are attractive molecules to study the relationship between genetics and environment because they play an important role in the protection of gastric mucosa and exhibit a highly polymorphic genetic variation in their length. In fact, mucins are the major structural components of the mucus viscous gel covering the gastric mucosa, and represent the first line defence barrier against environmental aggressions. All mucins have in common the presence of extended arrays of tandemly repeated peptides rich in serine and threonine residues that are potential O-glycosylation sites. The variable number of tandem repeats accounts for the extensive polymorphism observed at DNA, RNA and protein levels [7–14]. *MUC1* [10, 15, 16] and *MUC2* [17] are the only two fully sequenced human mucin genes. Several other human mucin genes have been cloned and partially sequenced [12–14, 18–21]. The protein product of the *MUC1* gene (mammary/pancreatic mucin) has a molecular weight that correlates to the size of the mRNA and to the size of the DNA restriction fragments, suggesting an important gene effect on the final structure of the mucin [22]. All known mucins are secreted products except *MUC1* which has a transmembrane anchorage domain leading to a membrane-bound mucin.

MUC1 mucin is highly expressed on the gastric mucosa [23] and therefore we advanced the hypothesis that *MUC1* polymorphism, resulting in individual size differences with respect to VNTR (variable number of tandem repeats), might influence the function of *MUC1* mucins, i.e. could provide different protective effects against various environmental aggressions.

*To whom correspondence should be addressed. Tel: + 351-2-590591; Fax: + 351-2-5503940.

0282–0080

In order to verify our hypothesis we performed a case-control study using Southern blot analysis to evaluate *MUC1* gene polymorphism in patients with gastric carcinoma and in a control group of blood donors.

Materials and methods

We have evaluated, in a case-control study, the *MUC1* gene polymorphism in a Caucasian control group of blood donors ($n = 324$) and in a series of patients with gastric carcinoma ($n = 159$). DNA from the control population was isolated from blood samples (10 ml), collected after obtaining informed consent. Blood samples were pelleted and stored at $-70\,^{\circ}C$ until DNA extraction. DNA from patients with gastric carcinoma was isolated from samples of non-neoplastic gastric mucosa collected immediately after surgery. The collected tissue samples were frozen in liquid nitrogen and stored at $-70\,^{\circ}C$ until DNA extraction.

High-molecular weight DNA was isolated using a salt-chloroform extraction method, as previously described by Mullenbach *et al.* [24]. DNA samples from blood donors and gastric cancer patients were digested with *Eco*RI that recognizes restriction sites in the flanking regions of the tandem repeats of the *MUC1* gene, separated on 0.7% agarose gel by electrophoresis, for 17–19 h at approximately 53 V, and transferred to nylon membranes (Hybond-N, Amersham) by alkaline blotting [25].

The *MUC1* probe was constructed using primers to the flanking regions of the tandem repeats (Primer sense – 5′TGGGCTGGGGGGGCGGTGG3′: Primer anti-sense – 5′CGGTGTCACCTCGGCCCCGGACA3′). The PCR product of 94 bp (base pairs) was subcloned into *Eco* RI site of pT7T3 plasmid vector (Pharmacia). The total plasmid was PCR-labelled with [α-^{32}P] dCTP (Amersham) and used as a probe for the *MUC1* tandem repeat region. Following an initial denaturation of 94 °C for 3 min, the amplification was performed for 30 cycles with denaturation at 94 °C for 1 min, annealing at 50 °C for 2 min and extension at 72 °C for 3 min. An additional extension period of 10 min at 72 °C was performed in the end of the cycles. The labelled probe was purified from non-incorporated nucleotide by passage through Sephadex-G50 columns.

Prehybridization was performed for 3–4 h in a phosphate solution (0.5 M $NaHPO_4$, pH = 7.2, 1 mM EDTA, 7% SDS) at 65 °C. Hybridization was performed overnight at 65 °C. Washing of the blots was performed under high stringency conditions (first with 40 mM $NaHPO_4$, 1 mM EDTA, 5% SDS, and secondly with 40 mM $NaHPO_4$, 1mM EDTA, 1% SDS, at 65 °C). Membranes were exposed for autoradiography at $-70\,^{\circ}C$ with an intensifying screen. Autoradiography was developed after 7 days.

All bands from the autoradiograms were scored visually. The size of the different alleles of the *MUC1* gene, both in the blood donors and in the gastric carcinoma patients, was determined by comparison with the size of the fragments of the marker λ*Hind*III (Amersham). All the alleles identified in controls and patients were ranked and numbered from 1 to 15 according to their molecular weight. Allele 1 represents the allele with higher molecular weight and allele 15 represents the allele with lower molecular weight.

The comparison between cases and controls was performed using χ^2 statistics and Student's t test. The Monte Carlo test applied to the study of highly polymorphic genes, as described by Sham and Curtis [26], was performed using the computer program CLUMP (ftp.diamond.gene.ulc.ac.uk [directory/pub/packages/dcurtis]). Odds ratio and 95% confidence limits were determined using the BMDP 4F computer program (Statistical Package program BMDP; Los Angeles, CA, USA).

Results

Comparison between controls and cases regarding age, gender and ABO histo-blood groups

The age of the control population ($n = 324$) was 39.7 ± 11.0 and the age of the gastric carcinoma patients ($n = 159$) was 61.0 ± 12.3 ($p < 0.0001$). The male:female sex-ratio was 3.6:1 in the control population and 1.6:1 in the patients with gastric carcinoma ($p < 0.0001$). The blood group was known in all the donors and in 113 patients with gastric carcinoma. Both populations were in Hardy-Weinberg equilibrium regarding the ABO histo-blood groups. The differences between the two populations were not significant ($p > 0.05$), regarding the ABO system.

Both populations are Caucasian and originate from the same districts in the Northwest part of Portugal.

We considered the samples appropriate for a case-control study of the *MUC1* gene polymorphism based on the presence of Hardy-Weinberg equilibrium in the ABO system. The differences in age and gender of controls and patients with gastric carcinoma will be further analysed when comparing the two populations for the *MUC1* gene polymorphism.

Comparison between controls and cases regarding *MUC1* gene polymorphism

Fifteen alleles were identified in the 324 blood donors. The allele size ranged from 8.9 kb (kilobase pairs) to 12.6 kb – alleles 15 and 1 in the ranking system (Fig. 1). Thirteen alleles were identified in the 159 patients with gastric carcinoma, all of which were also present in the control population. Alleles 1 and 14 from the control population were not identified in patients with gastric carcinoma. Allelic frequencies are represented in Table 1 and Fig. 2.

Alleles 4 and 10 were the most common alleles both in the control group and in patients with gastric carcinoma (Fig. 2). Most of the alleles were represented in both populations and the modal alleles were the same in both populations. Despite this, a difference in the distribution of allelic frequencies between the two groups is clearly observed

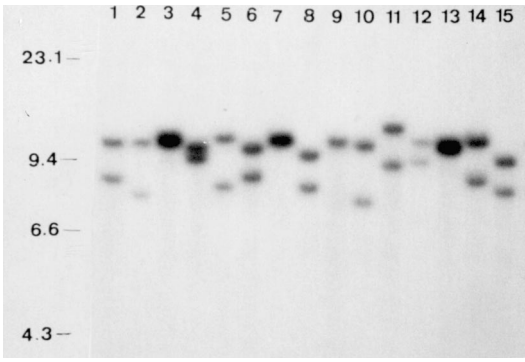

Figure 1. Detection of *MUC1* gene polymorphism. Human DNA samples from blood donors were digested with *Eco* RI, separated on 0.7% agarose gel, transferred to nylon membranes by alkaline blotting, and hybridized with the *MUC1* probe. Lanes numbered 1–15 correspond to the following genotypes: 4–8; 4–11; 4–4; 5–6; 3–10; 5–8; 3–3; 6–10; 3–3; 4–12; 2–7; 4–6; 4–4; 3–9; 6–10. Size marker is given in kilobases.

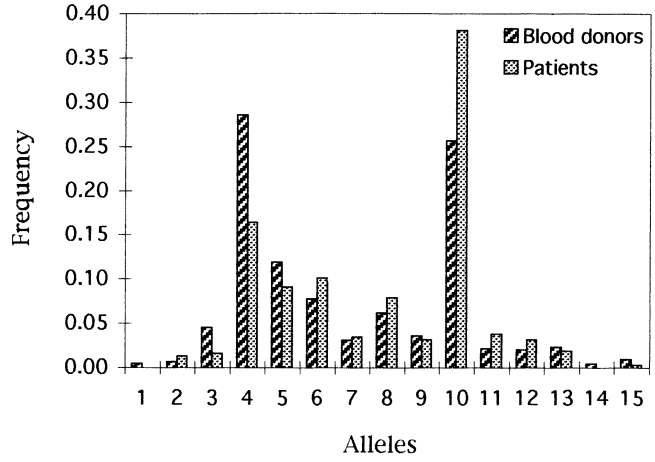

Figure 2. Allelic frequencies distribution of *MUC1* gene in blood donors and in patients with gastric carcinoma. Alleles are numbered according to their size by decreasing order of molecular weight (allele 1 is the heaviest).

Table 1. Distribution of *MUC1* allelic frequencies in blood donors and in patients with gastric carcinoma.

MUC1 *alleles*	*Blood donors* (n = *324*)	*Patients with gastric carcinoma* (n = *159*)
1	0.005	0.000
2	0.006	0.013
3	0.045	0.016
4	0.285	0.164
5	0.119	0.091
6	0.077	0.101
7	0.031	0.035
8	0.062	0.079
9	0.035	0.031
10	0.256	0.381
11	0.022	0.038
12	0.020	0.031
13	0.023	0.019
14	0.005	0.000
15	0.009	0.003

(Fig. 2). In fact the majority of the alleles with a relatively high molecular weight (alleles 1–5, except allele 2) were more frequent in the control group and alleles with a lower molecular weight (alleles 6–14, except alleles 9 and 13) were more frequent in patients with gastric carcinoma ($\chi^2 = 41.6$; $p < 0.001$). Due to the high number of alleles and the sparse representation of some of them we performed a Monte Carlo based test that confirmed a significantly ($p < 0.001$) different distribution of the *MUC1* alleles between patients and controls.

We further evaluated the *MUC1* genotypes distribution in the two populations. Observed heterozygosity was 72.2% in controls and 73.0% in patients with gastric carcinoma. The heterozygote 4–10 was the most frequent genotype in both populations: 44 of 324 controls (13.6%) and 28 of 159 patients with gastric carcinoma (17.6%). The homozygous genotype for allele 4 was more frequent in controls than in patients with gastric carcinoma – 36 individuals (11.1%) and three patients (1.9%), respectively – and the homozygous genotype for allele 10 was less frequent in controls than in patients with gastric carcinoma – 28 individuals (8.6%) and 24 patients (15.1%), respectively.

Sixty-six different genotypes were identified overall, most of them with a low number of individuals in both populations, which prevented the use of all genotypes for some statistical analyses. To group the genotypes we divided the alleles using the cut-point that corresponds to the median value of the blood donors population. The selected cut-point is also the one that maximizes the differences between the two populations: alleles 1–5 were grouped into a category of Large *MUC1* alleles (*L*) and alleles 6–15 were grouped into a category of Small *MUC1* alleles (*S*). Genotypes were thereafter recoded as having two large *MUC1* alleles (LL), two small *MUC1* alleles (SS) and one large *MUC1* allele and one small *MUC1* allele (LS).

In the control group of blood donors no differences were observed regarding the mean age ($p > 0.1$) and the male:female sex-ratio ($p < 0.1$) between the three different recoded genotypes: LL, SS, LS (data not shown). We assumed, based on these observations, that no major gender or age dependent selection of the different genotypes has occurred in the control population.

The comparison between patients and controls regarding the recoded *MUC1* genotypes is shown in Table 2. Significant differences are observed in the genotypic frequencies between patients and controls. Genotypes for large *MUC1* alleles (LL) are more frequent in controls (50.9%) than in patients with gastric carcinoma (29.3%) and genotypes for

Table 2. Distribution of recoded *MUC1* genotypes, LL (two large *MUC1* alleles), LS (one large *MUC1* allele and one small *MUC1* allele) and SS (two small *MUC1* alleles), in blood donors and in patients with gastric carcinoma.

Genotypes	*Blood donors* (n = *324*) n (%)	*Patients with gastric carcinoma* (n = *159*) n (%)
LL	67 (50.9%)	12 (29.3%)*
LS	162 (41.5%)	66 (50.0%)
SS	95 (7.5%)	81 (20.7%)*

The estimated relative risk (odds ratio) adjusted for age and gender, for gastric carcinoma development between individuals with SS genotype and individuals with LL genotype is 4.3 (95% confidence limits 1.8–10.5). *$p < 0.0001$.

small *MUC1* alleles (SS) are more frequent in patients with gastric carcinoma (20.7%) than in controls (7.5%) ($p < 0.0001$). Genotypes with one large *MUC1* allele and one small *MUC1* allele (LS) have a similar frequency in both groups (Table 2). The estimated relative risk (odds ratio), adjusted for age and gender, for gastric carcinoma development in individuals with SS genotype is 4.3 (95% confidence limits 1.8–10.5).

Figure 3 depicts the results obtained after grouping the genotypes LS together with the genotypes LL (Fig. 3A) or with the genotypes SS (Fig. 3B). In both scenaria the differences between controls and patients with gastric carcinoma are highly significant ($p < 0.0001$). The results obtained suggest that the presence of a small *MUC1* genotype is a risk factor for gastric carcinoma development, either in a LL or in a LS combination.

Discussion

The *rationale* of the present work was based on previous studies showing that the *MUC1* gene polymorphism is reflected both at the RNA and protein levels [7–14]. The evaluation of DNA polymorphism at the tandem repeat region of the *MUC1* gene was therefore used as a surrogate endpoint to the evaluation of protein polymorphism. In fact, individuals with *MUC1* alleles encompassing a lower number of tandem repeats are expected to code smaller *MUC1* proteins. We postulated therefore that such individuals might have a thinner mucus layer and might be more susceptible to environmental insults such as diet, toxins or microorganisms.

We found a signficantly different distribution of *MUC1* alleles between the control group of blood donors and the group of patients with gastric carcinoma. We have also found that the prevalence of small sized alleles is higher in patients with gastric carcinoma than in controls. The higher prevalence of allele 10 and the lower prevalence of allele 4 in patients with gastric carcinoma are the most striking differences between patients and controls. A similar tendency for a higher prevalence of smaller alleles in carcinomas was observed in respect to alleles 6, 7, 8, 11 and 12.

These results are in accordance with our hypothesis that individuals with small *MUC1* alleles are more susceptible to the development of gastric carcinoma.

The second approach we have undertaken was aimed at determining if the allele combination per individual, i.e. the genotype, was in keeping with this interpretation. In fact the expression of *MUC1* alleles is codominant [8] and the net effect of *MUC1* expression will be dependent on the polymorphism of both alleles.

The results observed with homozygous genotypes for the two most common alleles (alleles 4 and 10) followed the same tendency of allele distribution: patients with gastric carcinoma showed a significantly higher prevalence of homozygous genotype for the small sized allele (allele 10).

For the analysis of all genotypes we classified the cases as described in the Results section: LL, two large *MUC1* alleles; SS, two small *MUC1* alleles; and LS, one large *MUC1* allele and one small *MUC1* allele. We found that genotypes

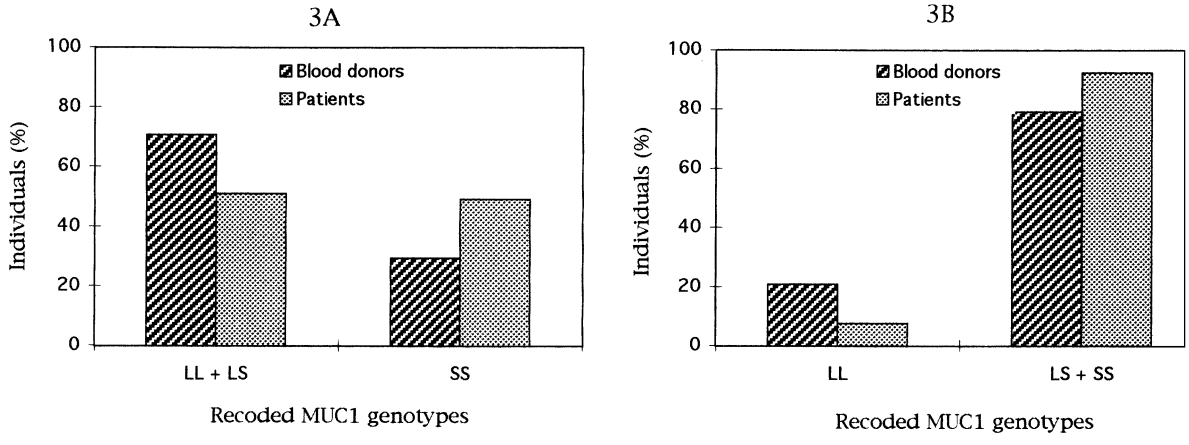

Figure 3. Distribution of recoded *MUC1* genotypes in blood donors and patients; (3A) Genotypes LS grouped with genotypes LL; (3B) Genotypes LS grouped with genotypes SS.

carrying small *MUC1* alleles, either combined with another small *MUC1* allele (SS) or with a large *MUC1* allele (SL) are more prevalent in patients with gastric carcinoma than in controls. The results suggest that the presence of a single small *MUC1* allele is sufficient to confer genetic susceptibility for the development of gastric carcinoma. The individuals with two small *MUC1* alleles (SS) carry an estimated relative risk for gastric carcinoma development of 4.3.

A few issues have to be addressed before we can recommend the study of mucin genes polymorphism for the identification of individuals at risk for the development of gastric carcinoma: (1) The young age of our control population may have introduced a bias in the statistical analysis, in the sense that we do not know how many individuals in this population will develop a gastric carcinoma in the future; (2) The DNA restriction fragments obtained after *Eco*RI digestion are of a quite large size contributing to the poor resolution of allele sizes. This may be responsible for an underestimation of the total number of alleles and for the high number of individuals classified as homozygous in both populations; (3) Finally, we need to get more information regarding the basis of our study. Namely, we need to confirm the assumption that the genetic polymorphism is directly reflected at the protein level and the genetic/protein polymorphism is causally involved in *in vivo* and *in vitro* models of gastric carcinogenesis.

If these shortcomings are resolved we will be able to claim that the newly proposed risk factor might be of major importance in the future. All the other reported risk factors for gastric carcinoma rely upon the identification of precursor conditions evaluated on histological slides, as is the case of chronic atrophic gastritis, intestinal metaplasia and dysplasia [27]. The detection of precursor conditions in symptomatic individuals forms the present basis for entering in follow-up protocols intended to achieve early diagnosis of gastric carcinoma. This strategy implicates invasive procedures with frequent endoscopies and biopsies. We except that the molecular epidemiology of mucin genes polymorphism will allow the establishment of non-invasive screening strategies for identification of individuals at risk for gastric carcinoma development in the general population.

Acknowledgements

This work was supported by Junta Nacional de Investigação Científica e Tecnológica (JNICT), PECS/SAU/255/95 and PRAXIS XX1 2/2.1/SAU/1356/95.

References

1 Munoz N (1988) In *Gastric carcinogenesis* (PI Reed and MJ Hill, eds) pp. 51–69. Amsterdam: Elsevier.
2 Knudson AG (1985) *Cancer Res* **45**: 1437–43.
3 Venitt S (1994) *Clin Chem* **40**: 1421–25.
4 La Vecchia C, Negri E, Decarli A, D'Avanzo B, Franceschi S (1987) *Int J Cancer* **40**: 484–89.
5 Parsonnet J, Friedman GD, Vandersteen DP, Chang Y, Vogelman JH, Orentreich N, Sibley RK (1991) *N Engl J Med* **325**: 1127–31.
6 La Vecchia C, Lucchini F, Negri E, Boyle P, Maisonneuve P, Levi F (1992) *Eur J Cancer* **28**: 132–235.
7 Karlson S, Swallow DM, Griffiths B, Corney G, Hopkinson DA (1983) *Ann Hum Genet* **47**: 263–69.
8 Swallow DM, Gendler S, Griffiths B, Corney G, Taylor-Papadimitrou J, Bramwell ME (1987) *Nature* **328**: 82–84.
9 Gendler S, Taylor-Papadimitrou J, Duhig T, Rothbard J, Burchell J (1988) *J Biol Chem* **263**: 12820–23.
10 Gendler SJ, Lancaster CA, Taylor-Papadimitrou J, Duhig T, Peat N, Burchell J (1990) *J Biol Chem* **265**: 15286–93.
11 Toribara NW, Gum JR, Culhane PJ, Lagace RE, Hicks JW, Peterson GM, Kim YS (1991) *J Clin Invest* **88**: 1005–13.
12 Porchet N, Van Cong N, Dufosse J, Audie JP, Guyonnet-Duperat V, Gross MS, Denis C, Degand P, Bernheim A, Aubert JP (1991) *Biochem Biophys Res Commun* **175**: 414–22.
13 Toribara NW, Roberton AM, Ho SB, Kuo W-L, Gum E, Hicks JW, Gum JR, Byrd JC, Siddiki B, Kim YS (1993) *J Biol Chem* **268**: 5879–85.
14 Bobek LA, Tsai H, Biesbrock AR, Levine MJ (1993) *J Biol Chem* **268**: 20563–69.
15 Lan MS, Batra SK, Qi W-N, Metzgar RS, Hollinsworth MA (1990) *J Biol Chem* **265**: 15294–99.
16 Lancaster CA, Peat N, Duhig T, Wilson D, Taylor-Papadimitrou J, Gendler SL (1990) *Biochem Biophys Res Commun.* **173**: 1019–29.
17 Gum JR, Hicks JW, Toribara NW, Siddiki B, Kim YS (1994) *J Biol Chem* **269**: 2440–46.
18 Gum JR, Hicks JW, Swallow DM, Lagace RL, Byrd JC, Lamport DTA, Siddiki B, Kim YS (1990) *Biochem Biophys Res Commun* **171**: 407–15.
19 Van Cong N, Aubert JP, Gross MS, Porchet N, Degand P, Frézal J (1990) *Hum Genet* **86**: 167–72.
20 Aubert JP, Porchet N, Crepin M, Duterque-Coquillaud M, Vergnes G, Mazzuca M, Debuire B, Petitprez D, Degand P (1991) *Am J Resp Cell Mol Biol* **5**: 178–85.
21 Meerzaman D, Charles P, Daskal E, Polymeropolus MH, Martin BM, Rose MC (1994) *J Biol Chem* **269**: 12932–39.
22 Ligtenberg MJL, Vos HL, Gennissen AMC, Hilkens J (1990) *J Biol Chem* **265**: 5573–78.
23 Ho SB, Niehans GA, Lyftogt C, Yan PS, Cherwitz DL, Gum ET, Dahiya R, Kim YS (1993) *Cancer Res* **53**: 641–51.
24 Mullenbach R, Lagoda PJL, Welter C (1989) *Trends Genet* **5**: 391.
25 Southern EM (1975) *J Mol Biol* **98**: 503–17.
26 Sham PC, Curtis D (1995) *Ann Hum Genet* **59**: 97–105.
27 Correa P (1992) *Cancer Res* **52**: 6735–40.

Received 8 September 1995, revised 28 December 1995