Oncogene (2002) 21, 8192–8195

www.nature.com/onc

Association of *CDH1* haplotypes with susceptibility to sporadic diffuse gastric cancer

Bostjan Humar[1], Francesco Graziano[2], Stefano Cascinu[3], Vincenzo Catalano[4], Anna M Ruzzo[5], Mauro Magnani[5], Tumi Toro[1], Tudor Burchill[1], Matthias E Futschik[6], Tony Merriman[7] and Parry Guilford*,[1]

[1]*Cancer Genetics Laboratory, University of Otago, Dunedin, Aotearoa, New Zealand;* [2]*Division of Medical Oncology, Hospital of Urbino, 61029 Urbino, Italy;* [3]*Medical Oncology, Hospital of Parma, 43100 Parma, Italy;* [4]*Medical Oncology, Hospital of Pesaro, 61100 Pesaro, Italy;* [5]*G Fornaini Institute of Biochemistry and Molecular Biology, University of Urbino, 61029 Urbino, Italy;* [6]*Knowledge Engineering Laboratory, University of Otago, Dunedin, Aotearoa, New Zealand;* [7]*Biochemistry Department, University of Otago, Dunedin, Aotearoa, New Zealand*

Truncating mutations in the gene for the cell to cell adhesion protein E-cadherin are the most consistent genetic alterations observed in sporadic and hereditary diffuse gastric cancer (DGC). In addition to these inactivating mutations, a *CDH1* promoter polymorphism at position −160 has been reported to lead to transcriptional downregulation of the gene *in vitro*. We therefore performed a case–control study to investigate whether this variant is associated with an increased susceptibility to DGC. The frequency of the −160A allele was significantly higher ($P<0.005$) in 53 diffuse gastric cancer cases compared to 70 matched controls. The odds ratio associated with the A-allele was 2.27 for CA-heterozygotes (95%CI 1.16–4.44) and 7.84 for AA-homozygotes (95%CI 2.89–21.24). Two additional polymorphisms (the 48+6T→C and the 2076C→T variant) were genotyped and shown to be equally distributed among cases and controls. Haplotype analysis with the three polymorphisms confirmed an association with disease ($P<0.004$). However, this analysis suggested the −160C→A *CDH1* promoter polymorphism may be in linkage disequilibrium with a distinct aetiological locus or acts in combination with other functional variants in or near the *CDH1* region.
Oncogene (2002) **21,** 8192–8195. doi:10.1038/sj.onc.1205921

Keywords: diffuse gastric cancer; E-cadherin; *CDH1*; polymorphism; genetics

Introduction

Over recent decades a steady decline in the incidence of gastric cancer has been observed. Nevertheless, gastric cancer remains a major health concern due to the high mortality and poor prognosis for this disease (Dunbier and Guilford, 2001; Howson *et al*., 1986). Gastric cancer is usually divided into two predominant histological forms, the intestinal and the diffuse type (Lauren, 1965). Intestinal gastric cancer is strongly associated with older age groups and its incidence appears to parallel the worldwide decrease in gastric cancer. The diffuse form is more prevalent in younger age groups and more likely to show familial clustering. Limited data suggests the incidence of diffuse gastric cancer (DGC) to be stable (Borch *et al*., 2000; Lauren and Nevalainen, 1993).

The *E-cadherin* gene (*CDH1*, OMIM 192090) encodes a homophilic transmembrane cellular adhesion protein that is expressed in epithelial tissues and found to be mutated in about 50% of sporadic DGCs (Berx *et al*., 1998). Significant familial clustering of the malignancy is attributable to truncating germline mutations in *CDH1* (Guilford *et al*., 1998) and underscores the central role of the gene as a tumour suppressor in DGC. Several polymorphic variants of *CDH1* exist in the population (Berx *et al*., 1998), albeit without any apparent biological consequences. Recently, Li *et al*. (2000) characterized a C to A polymorphism located 160 bp upstream from the *CDH1* transcription start site and found the A-allele to have reduced transcriptional factor binding strength and only 32% of the transcriptional activity of the C-allele *in vitro*. The −160C→A *CDH1* promoter variant may thus reduce *CDH1* expression *in vivo* and be regarded as a candidate low-penetrance cancer susceptibility polymorphism. However, recent epidemiological studies failed to demonstrate a correlation between the promoter *CDH1* variant and breast (Lei *et al*., 2002) or colorectal cancer (Porter *et al*., 2002). Moreover, the A-allele was suggested to be protective against gastric cancer (Wu *et al*., 2002). The study described here is aimed at clarifying the role of the −160C→A *CDH1* promoter polymorphism in DGC susceptibility.

In order to investigate whether the −160C→A *CDH1* promoter polymorphism (Table 1) may confer an increased risk to DGC, its frequency was

*Correspondence: P Guilford, Cancer Genetics Laboratory, Department of Biochemistry, PO Box 56, Dunedin, Aotearoa, New Zealand; E-mail: parry.guilford@otago.ac.nz
Received 9 May 2002; revised 30 July 2002; accepted 31 July 2002

determined in 53 sporadic DGC cases and in 70 controls matched for age, sex and derived from the same region in Italy. The allelic distribution was in Hardy–Weinberg equilibrium in the case and in the control group ($\chi^2 < 0.2$; d.f. = 2; $P > 0.91$). The A-allele was significantly over-represented ($\chi^2 = 10.9$; d.f. = 2; $P < 0.005$) in gastric cancer cases compared to controls (for frequencies see Table 2). The corresponding odds ratio was 2.27 for CA heterozygotes (95%CI 1.16–4.44) and 7.84 for AA homozygotes (95%CI 2.89–21.24) when compared to CC homozygotes.

Phenotypic data were compared among DGC cases subdivided according to their −160C→A genotype. Notably, the age at diagnosis was highest in CC homozygotes and lowest in AA homozygotes (mean ages ± s.d.: CC = 59 ± 11.1, CA = 52.6 ± 10.5, AA = 52.3 ± 8.5 years). The differences in age at diagnosis between the CC and CA or CC and AA genotypes were significant (one-tailed *t*-test for unequal variance, $P < 0.005$). As the age distribution within the individual genotypes did not appear to be normal (data not shown), a Fisher randomization test (10^5 random permutations; Fisher, 1951) was applied to examine the data under the assumption of non-Gaussian distribution. While the difference in age at diagnosis between the AA and CC genotype was of borderline significance ($P = 0.059$), the age at diagnosis was significantly lower in CA heterozygotes ($P < 0.04$) and in A-allele carriers (CA or AA genotype; $P < 0.02$) compared to CC homozygotes. No significant differences between the three genotypes were apparent regarding gender, tumour site, tumour grade and presence of metastasis.

The genotyping data suggests an association between the promoter −160A allele and an increased risk of sporadic DGC in the Italian study group. This is supported by the significantly younger age at diagnosis in carriers of the A-allele compared to CC homozygotes. Wu *et al*. (2002) investigated Taiwanese gastric cancer cases of mixed histology and found a reduced frequency of AA-homozygotes compared to controls. The protective role of the AA genotype was also apparent when only gastric cancer cases of diffuse histology were included in the analysis, which contrasts with the results of this study. Several reasons may account for this discrepancy. Firstly, the influence of a

Table 1 Summary of investigated *CDH1* polymorphisms

Polymorphism	*Change*	*Putative effect*	*Reference*
Promoter	−160C→A	Reduced transcription	Li *et al*. (2000)
Intron 1	48 + 6T→C	Unknown	Avizienyte *et al*. (2001)
Exon 13	2076C→T	Silent	Risinger *et al*. (1994)

PCR–RFLP was used to genotype all three polymorphisms. For the −160C→A polymorphism, a promoter fragment of 414 bp was amplified by PCR at 65°C annealing temperature and 1.5 mM Mg^{2+} using 5′AGTGAGCCCCATCTCCAAAA-3′ forward and 5′-TGACTTCCGCAAGCTCACA-3′ reverse primers. The product was digested at 37°C for 16 h with the enzyme *Afl*III (New England Biolab Ltd.), which cuts the A allele into a 211 and 203 bp fragment. DNA sequencing was applied to confirm the RFLP results. PCR amplified fragments containing the intron 1 and exon 13 polymorphic sites were digested by *Hpa*II and confirmed using *Msp*I enzymes (both New England Biolab Ltd.). Details of the procedures are given in Guilford *et al*. (1998) and Humar *et al*. (2002)

Table 2 Frequency of *CDH1* polymorphisms in DGC cases and controls

Variant	*Subject*	*Frequency*							*χ^2-test cases/controls*	
		CC	(%)	CA	(%)	AA	(%)	*n*	χ^2	*P*
−160	case	17	32.1	26	49.0	10	18.9	53	10.9	0.004
	control	40	57.1	27	38.6	3	4.3	70		
		TT		TC		CC		*n*	χ^2	*P*
48 + 6	case	41	89.1	4	8.7	1	2.2	46	1.791	0.408
	control	56	81.2	12	17.4	1	1.4	69		
		TT		TC		CC		*n*	χ^2	*P*
2076	case	12	26.1	22	47.8	12	26.1	46	0.317	0.853
	control	15	21.7	34	49.3	20	29	69		

Peripheral blood samples were collected from 53 consecutive patients (mean age ± s.d. 54.6 ± 10.6 years) with diffuse gastric cancer who were natives of the District of Pesaro-Urbino, Region Marche, Central Italy. After surgery, patients were referred to the local Medical Oncology Units for clinical evaluation and routine follow-up procedures. The diagnosis of diffuse gastric cancer was confirmed by two pathologists after independent review of tumour blocks. In addition to common clinico-pathologic features, patients completed a demographic sheet that included items and personal and familial cancer history. Subsequently, all these data were verified during interviews with the oncology physician and pedigrees were traced back for at least three generations and laterally to second- and third-degree relatives. Where necessary, cancer diagnoses and deaths in relatives were confirmed by medical or pathologic records. On the basis of this evaluation, none of the patients fulfilled the criteria for HDGC or other cancer syndromes such as HNPCC. The study was approved by the local Ethical Committee and all patients and healthy volunteers gave their written informed consent before their study entry. Control blood samples were obtained from 70 healthy blood donors (mean age ± s.d. 51.8 ± 11.11 years), who were evaluated for study entry at the transfusion units of the Hospitals of Urbino and Pesaro. The controls were natives from the same region as the cancer cases, had no personal history of any major disease and did not report familial history of cancer. The controls were matched according to the demographic characteristics of the patients; the two populations were comparable for age, sex, ethnicity and residential region. Mean ages were similar (*t*-test, $P = 0.16$) for cases and controls; female proportion was 54.7% in cases and 52.9% in controls; mean ages in males or females were similar within the cases and controls or when compared between cases and controls (*t*-test, $P > 0.12$). The significance of the difference in the distribution of the polymorphisms among different groups was calculated using χ^2-test with d.f. = 2. All allelic distributions were in Hardy–Weinberg equilibrium ($\chi^2 < 0.82$, $P > 0.66$) in cases and in controls. The distribution of the −160 *CDH1* polymorphism was similar ($\chi^2 < 3.83$, $P > 0.147$) between sexes in both groups. The gastric cancer risk associated with the −160 A genotypes was estimated by means of the odds ratio with 95% confidence limits

low-penetrance susceptibility gene on disease risk is likely to be affected by modifying genes and environmental factors. The different genetic background and local environment between the Italian and Taiwanese population may to some extent explain the different risk estimates associated with the A-allele. Furthermore, the case group investigated by Wu *et al*. (2002) does not appear to be in Hardy–Weinberg equilibrium (95CC, 102CA, 4AA; expected 106CC, 80CA, 15AA; $\chi^2 = 9.7$; d.f. = 2; $P = 0.0078$), suggesting that some form of population stratification and/or selective pressure acting differently between cases and controls may be present. The observed under-representation of AA-homozygotes among the Taiwanese cancer cases might hence be due to a reason other than A-allele conferred protection from gastric cancer.

Alternatively, the association observed in this study between the *CDH1* promoter polymorphism and gastric cancer risk may be secondary to linkage disequilibrium with an as yet unidentified, but tightly linked, DGC locus. To further examine this hypothesis, both the case and the control group were genotyped for an intronic polymorphism (48 + 6T→C; *CDH1* intron 1; Avizienyte *et al*., 2001) and a silent exonic polymorphism (2076C→T; *CDH1* exon 13; Risinger *et al*., 1994). Neither of the two polymorphisms were associated with DGC (Table 2). However, when a global association test (which takes into account all possible haplotypes of the three polymorphisms) was performed, the combination of the three *CDH1* polymorphisms was significantly associated with disease ($P < 0.004$; Table 3). Because individual haplotype frequencies were estimated (Table 3, legend), statistical parameters cannot be used to provide support for a significant association, or otherwise, of specific haplotypes with disease. Nevertheless, examination of the estimated haplotype frequencies indicates that the A-T-T haplotype appears to increase the risk for DGC, whereas the C-T-T haplotype decreases the risk (Table 3). If the −160C→A polymorphism was aetiological and the A variant a dominant disease-causing allele, all haplotypes containing the A allele would be expected to be positively associated with disease; this does not appear to be true (Table 3). Conversely, if the C allele was dominantly protective then all C containing haplotypes should be over-represented in controls; again this is not apparent (Table 3).

Table 3 Global three-marker haplotype analysis

Haplotype	*Controls* (%)	*Cases* (%)
C-T-T	37.76	21.61
A-T-T	6.32	28.39
C-C-C	7.85	6.52
C-T-C	28.18	28.39
C-C-T	2.29	0.00
A-T-C	17.59	15.09
Significance	$P = 0.00332$, d.f. = 5	
Marker order	5′-(−160)-(48 + 6)-(2076)-3′	

Haplotype frequencies were estimated using the expectation-maximization algorithm (Dempster *et al*., 1977) with the assumption of Hardy–Weinberg equilibrium. Haplotype phase of the three polymorphic markers genotyped could not be unambiguously assigned, because unrelated individuals were studied. The global test of association of the three markers with disease was done by measuring haplotype odds ratios across multiple categories and calculating a likelihood ratio test statistic of homogeneity (Koeleman and Dudbridge, 2000, Software for multilocus association analysis in unrelated subjects; available from ftp://ftp-gene.cimr.cam.ac.uk). Not all possible combinations of three-marker haplotypes were detected, as the short physical distance between the individual markers hinders free meiotic recombination. To assess linkage disequilibrium between haplotypes D′ values were calculated (Devlin and Risch, 1995). D′ values range from 1 (complete disequilibrium) through 0 (complete equilibrium) to −1 (alleles never found on same haplotype). In cancer cases the − 160 polymorphism was in linkage disequilibrium with each of 48 + 6 and 2076 (D′ = 1.0 and 0.34 respectively) and the 48 + 6 and 2076 polymorphism were in linkage disequilibrium (D′ = 1.0). In controls the respective D′ values were 1.0, 0.40 and 0.51

One explanation for these findings is that the −160 *CDH1* promoter polymorphism is not a determinant of DGC susceptibility, but instead is in linkage disequilibrium with a putative aetiological variant (either within *CDH1* or a neighbouring gene). In such a case, the discrepancy between our results and those of Wu *et al*. (2002) could be explained by either a single recombination event between this putative susceptibility locus and the promoter polymorphism, or a *de novo* mutation at the −160 position in an ancestral chromosome of one of the populations.

Alternatively, our findings may point to a polyallelic effect, with several tightly linked polymorphisms modulating DGC risk. Interestingly, examination of the data shown in Table 3 suggests the presence of at least three functionally distinct haplotypes at *CDH1* – a susceptibility haplotype (marked by A-T-T), a protective haplotype (C-T-T) and one or more neutral haplotypes (e.g. A-T-C). The −160A polymorphism may thus contribute to DGC risk in a contextual manner, with one or more additional polymorphisms within, or close to, the *CDH1* gene influencing DGC risk both positively or negatively; depending on the combination of polymorphic variants, the influence of the −160A-allele may be masked due to the presence of other, as yet unidentified alleles involved in DGC susceptibility. Hence separate disease-associated haplotypes may account for the discrepancy between the roles of the −160C→A polymorphism in Caucasian (this study) and Taiwanese (Wu *et al*., 2002) cohorts.

Based on these results, we propose that the *CDH1* −160A, 48 + 6T and 2076T haplotype is a marker for DGC susceptibility in the Italian study group. Assuming the estimated frequencies (Table 3) to be representative, the A-T-T and C-T-T haplotypes are associated with a relative risk of 1.89 and 0.65 respectively. At least in the Italian region from where the cases and controls were sampled, the A-T-T haplotype might prove a useful marker for DGC susceptibility and may also increase the penetrance of disease in families carrying truncating germline *CDH1* mutations. Notably, we have previously described a hereditary DGC kindred from this Italian region (Humar *et al*., 2002), where the A-T-T haplotype appears to cosegregate with disease (data not shown), supporting the putative utility of the A-T-T haplotype

allele as a disease marker. It is important to stress, however, that the haplotype results need to be replicated both in larger data sets and in family-based studies (e.g. using the transmission disequilibrium test (Spielman *et al*., 1993)) so that disease-associated and protective haplotypes can be unambiguously identified.

In conclusion, we have identified a significant association between a polymorphic *CDH1* haplotype and an increased risk to DGC. The data presented in this study suggests either tight linkage of the −160C→A *CDH1* polymorphism to an unidentified aetiological DGC locus or a contribution of the promoter variant to a polyallelic cumulative DGC susceptibility locus. The associated risk estimates suggest a significant role of the susceptibility locus in the incidence of sporadic DGC. These results emphasize the importance of haplotype analysis of closely linked markers in disease association studies.

Acknowledgements
We thank all the patients and volunteers who contributed to this study. We also thank Natalie J Kerr and Anita Dunbier for critical reading of the manuscript. B Humar was supported by a grant from the Swiss Cancer League (BIL SKL 993-02-2000) and the Swiss National Foundation.

References

Avizienyte E, Launonen V, Salovaara R, Kiviluoto T and Aaltonen LA. (2001). *J. Med. Genet.,* **38,** 49–52.
Berx G, Becker KF, Hofler H and van Roy F. (1998). *Hum. Mutat.,* **12,** 226–237.
Borch K, Jonsson B, Tarpila E, Franzen T, Berglund J, Kullman E and Franzen L. (2000). *Br. J. Surg.,* **87,** 618–626.
Dempster AP, Laird NM and Rubin DB. (1977). *J. R. Stat. Soc. B,* **39,** 1–38.
Devlin B and Risch N. (1995). *Genomics,* **29,** 311–322.
Dunbier A and Guilford P. (2001). *Adv. Cancer Res.,* **83,** 55–65.
Fisher RA. (ed). (1951). *The design of experiments*. London: Oliver and Boyd Ltd, pp. 1–244.
Guilford P, Hopkins J, Harraway J, McLeod M, McLeod N, Harawira P, Taite H, Scoular R, Miller A and Reeve AE. (1998). *Nature,* **392,** 402–405.
Howson CP, Hiyama T and Wynder EL. (1986). *Epidemiol. Rev.,* **8,** 1–27.
Humar B, Toro TT, Graziano F, Muller H, Dobbie Z, Kwang-Yan H, Eng C, Hampel H, Gilbert D, Winship I, Parry S, Ward R, Findlay MP, Christian A, Tucker M, Tucker K, Merriman T and Guilford P. (2002). *Hum. Mutat.,* **19,** 518–525.
Koeleman BP, Dudbridge F, Cordell HJ and Todd JA. (2000). *Ann. Hum. Genet.,* **64,** 207–213.
Lauren P. (1965). *Acta Pathol. Microbiol. Scand.,* **64,** 31–49.
Lauren PA and Nevalainen TJ. (1993). *Cancer,* **71,** 2926–2933.
Lei H, Sjoberg-Margolin S, Salahshor S, Werelius B, Jandakova E, Hemminki K, Lindblom A and Vorechovsky I. (2002). *Int. J. Cancer,* **98,** 199–204.
Li LC, Chui RM, Sasaki M, Nakajima K, Perinchery G, Au HC, Nojima D, Carroll P and Dahiya R. (2000). *Cancer Res.,* **60,** 873–876.
Porter TR, Richards FM, Houlston RS, Evans DG, Jankowski JA, MacDonald F, Norbury G, Payne SJ, Fisher SA, Tomlinson I and Maher ER. (2002). *Oncogene,* **21,** 1928–1933.
Risinger JI, Berchuck A, Kohler MF and Boyd J. (1994). *Nat. Genet.,* **7,** 98–102.
Spielman RS, McGinnis RE and Ewens WJ. (1993). *Am. J. Hum. Genet.,* **52,** 506–516.
Wu MS, Huang SP, Chang YT, Lin MT, Shun CT, Chang MC, Wang HP, Chen CJ and Lin JT. (2002). *Cancer,* **94,** 1443–1448.