Quantitative trait locus analysis reveals two intragenic sites that influence O6-alkylguanine-DNA alkyltransferase activity in peripheral blood mononuclear cells

Geoffrey P. Margison1, Jim Heighway1, Steven Pearson, Gail McGown, Mary R. Thorncroft, Amanda J. Watson, Kathryn L. Harrison2, Sarah J. Lewis3, Klaus Rohde2,3, Philip V. Barber4, Paul O’Donnell1, Andrew C. Povey2 and Mauro F. Santibáñez-Koref5

Cancer Research-UK Carcinogenesis Group, Paterson Institute for Cancer Research, Christie Hospital NHS Trust, Manchester, UK, 1University of Liverpool Cancer Research Centre, Liverpool, UK, 2Centre for Occupational and Environmental Health, University of Manchester, Manchester, UK, 3Max-Delbrück-Centrum für Molekulare Medizin, Berlin-Buch, Germany, 4North West Lung Cancer Centre, Wythenshawe Hospital, Manchester, UK and 5Institute of Human Genetics, University of Newcastle, Newcastle upon Tyne, UK

The repair of specific types of DNA alkylation damage by O6-alkylguanine-DNA alkyltransferase (MGMT) is a major mechanism of resistance to the carcinogenic and chemotherapeutic effects of certain alkylating agents. MGMT expression levels vary widely between individuals but the underlying causes of this variability are not known. To address this, we used an expressed single nucleotide polymorphism (SNP) and demonstrated that the MGMT alleles are frequently expressed at different levels in peripheral blood mononuclear cells (PBMC). This suggests that there is a genetic component of inter-allelic variation of MGMT levels that maps close to or within the MGMT locus. We then used quantitative trait locus (QTL) analysis using intragenic SNPs and found that there are at least two sites influencing inter-individual variation in PBMC MGMT activity. One is characterized by an SNP at the 3' end of the first intron and the second by two SNPs in the last exon. The latter are in perfect disequilibrium and both result in amino acid substitutions—one of them, Ile143Val, affecting an amino acid close to the Cys145 residue at the active site of MGMT. Using in vitro assays, we further showed that while the Val143 variant did not affect the activity of the protein on methylated DNA substrate, it was more resistant to inactivation by the MGMT pseudosubstrate, O6-(4-bromo)moethylnuamine. These findings suggest that further investigations of the potential epidemiological and clinical significance of inherited differences in MGMT expression and activity are warranted.

Introduction

Cellular DNA is constantly undergoing damage by a variety of endogenous and exogenous processes and agents that produce a wide range of adducts and other modifications that can result in mutation, chromosome fragmentation and cell death. In order to maintain genome integrity and reproductive capacity, a complex network of DNA damage detection and processing mechanisms has evolved. One of these systems, which processes specific types of alkylation damage relevant to both carcinogenesis and cancer chemotherapy, is the DNA repair protein O6-alkylguanine-DNA alkyltransferase [O6-methylguanine-DNA methyltransferase; MGMT, E.C.2.1.1.63 for recent reviews see (1–4)]. This protein constitutes the first line of cellular defense against the genotoxic effects of alkylation of DNA at the O6-position of guanine. MGMT reverses O6-alkylation damage by the covalent transfer of the alkyl group to the protein itself, leading to its inactivation, ubiquitination and proteasome-mediated degradation.

The clearest demonstration that MGMT is a critical determinant for the genotoxic effects of alkylating agents stems from in vivo models. MGMT null mutant mice are more susceptible both to tumour induction by alkylating carcinogens and to the toxic effects of alkylating antitumour agents (5,6). In addition, MGMT and other alkyltransferase-transgenic mice that express increased levels of repair activity are more resistant to alkylating agent-induced carcinogenesis and also have a lower frequency of spontaneous tumors (7–10). Evidence indicating the importance of MGMT in carcinogenesis also comes from studies on human tumors. Tumors lacking MGMT show a high frequency of mutations in genes critical for tumorigenesis, such as KRAS2 and TP53 and the spectrum of these mutations is consistent with the mutagenic effect of unreppaired O6-alkylguanine lesions (for a review of the mechanisms of carcinogenesis involving O6-alkylguanine adducts see ref. 11).

The human MGMT gene is located on chromosome band 10q26, is ~300 kb in size and consists of 5 exons of which the last four are coding. The promoter spans the first exon and part of the first intron; it contains CG-rich regions but lacks TATA or CAAT boxes and is similar to that of housekeeping genes. The transcript is ~0.95 kb long and no splice variants have been described (12). MGMT activity is present at different levels in different normal tissues (13) and in those tissues that have been examined considerable inter-individual differences in activity levels (14) are noticed. MGMT activity in peripheral blood mononuclear cells (PBMC) taken from the same individuals at different time points also shows extensive variation, although the extent of inter-individual variability is higher than that of intra-individual variability (15). The basis of such intra- and inter-individual variation in MGMT expression levels is unknown. In rats, and to a lesser extent in other rodents, increased MGMT expression has been observed in response to a range of treatments, including exposure to genotoxic agents such as ionizing radiation or to γ-interferon, and even partial hepatectomy (11). However, in humans, definite evidence of inducibility is lacking.

Abbreviations: AEI, allelic expression imbalance; MGMT, O6-alkylguanine-DNA alkyltransferase (O6-methylguanine-DNA methyltransferase; E.C.2.1.1.63); PBMC, peripheral blood mononuclear cell; QTL, quantitative trait locus; SNP, single nucleotide polymorphism.

Carcinogenesis vol.26 no.8 © Oxford University Press 2005; all rights reserved.

1473
The chemotherapeutic exploitation of the \(O^2\)-alkylating agents is based on the toxicity of \(O^2\)-alkylguanine which is mediated, for the methylating agents, by the post replication mismatch repair system and for the chloroethylating agents, by DNA interstrand crosslinks (1). The dose-limiting toxicity of these agents is almost always myelosuppression likely to be a consequence of the generally low levels of expression of MGMT in bone marrow cells (16). Despite this, there is increasing interest in the use of MGMT pseudosubstrates, mostly low molecular weight analogues of \(O^2\)-methylguanine, to inactivate MGMT in tumor cells and increase their sensitivity to chemotherapy (17). Patient management would benefit if a convenient method could be found to predict the efficacy of such pseudosubstrates and the variable toxicity.

Such observations in humans have led to the assumption that inter-individual differences in MGMT activity have a genetic component. This has prompted a series of case-control studies attempting to establish an association between different forms of cancer and polymorphisms, mainly in the coding region but more recently also in the promoter region of the gene. The rationale for such studies is the hypothesis that MGMT expression levels and/or the functional activity of the expressed protein may influence cancer risk, in particular of cancers where environmental exposure to alkylating agents may play a role. Early investigations focused on variations affecting the coded protein and led to the identification of rare variants with altered sensitivity to inhibitors or reduced protein half-life (18,19). However, the variants studied so far are too infrequent to account for a significant proportion of the differences in MGMT activity between individuals. If MGMT activity levels do indeed influence cancer risk, establishing an association between cancer risk and intragenic polymorphisms can be facilitated by identifying first sites that are associated with variation of activity across the population. Knowledge of these polymorphisms and how they relate to protein activity can then be used to predict protein activity in individuals where the relevant polymorphisms have been typed. This then allows an examination of possible associations with cancer risk or the extent of side effects in patients receiving chemotherapy.

Genetic factors affecting message levels can be located either in trans-acting, e.g. transcription factors, or in cis-acting, e.g. the binding sites of such factors. The latter can result in unequal message levels of both alleles (allelic expression imbalance, AEI) in heterozygous individuals. The presence of AEI is, therefore, consistent with the fact that polymorphisms affecting elements acting in cis are involved in the variation of message levels and hence protein activity across the population. AEI is not uncommon: Yan et al. (20) were able to detect various degrees of AEI in more than half of the genes they examined. Similar results have recently been reported by Bray et al. (21) in normal brain tissue. It has previously been reported that there are allelic expression differences in MGMT in lung tissue (22), suggesting that polymorphisms in cis-acting elements may modulate expression levels and therefore MGMT activity in that tissue.

Here, we investigate the genetic basis of inter-individual differences in MGMT activity in PBMC. We first examine whether allelic expression differences can also be observed in these cells: these would be consistent with polymorphisms in cis-acting elements influencing activity levels in PBMC. We then analyse a number of intragenic sites for their association with differences in MGMT activity and investigate the consequences on protein function and stability of two polymorphisms, one of them located near the active site of the protein, that result in amino acid substitutions and are associated with differences in MGMT activity.

Materials and methods

Study population

The material used in these studies was collected for an investigation of lung cancer susceptibility in relation to DNA adduct formation, DNA repair and sequence variation of genes involved in DNA repair and in the metabolism of genotoxic agents (23). Patients attending bronchoscopy clinics at the North West Lung Centre, Wythenshawe Hospital, who were over the age of 18 and who were well enough to take part were asked to participate. Of the total of 250 eligible patients, 184 were asked to participate: 4 refused giving an overall response rate of 98%.

Determination of MGMT activity in PBMC

Blood samples were obtained from 180 individuals recruited for the studies described above. PBMC were isolated by a standard Ficoll (Amersham Bioscience) method. For 151 samples, there was sufficient material to determine the MGMT specific activity in PBMC sonicates using a standard MGMT assay (24). Of the 151 patients, 62 were female and 89 male; 58 had cancer of the lung and 10 had cancer at other sites.

Nucleic acid extraction

Total mRNA was extracted from PBMC using a standard Trizol (GIDCO Biotechnology) protocol. DNA was extracted using the QIAamp DNA Blood Midi Kit (Qiagen) from the whole blood. Material for DNA extraction was available for 138 of the original 180 individuals included in the analysis of MGMT. For the remaining individuals all material had been used for the MGMT assay. RNA was extracted from 21 samples of which there was sufficient material. These 21 samples were not selected by any criterion other than availability of material for RNA extraction.

Allelic expression imbalance

cDNA was synthesized from 1 μl aliquots of total RNA in a 20 μl poly T-primed reaction using a Promega Reverse Transcription System (as instructed by the manufacturer). A stock reaction mix was prepared according to the number of tubes + 1 consisting (per tube) of 42 μl distilled water, 0.5 μl of 10× Taq reaction buffer, 0.5 μl of each primer, 0.5 μl of a 250 mM dNTP mix and 1.25 U of Roche Taq polymerase. The stock mix was vortexed after Taq addition and 49 μl added to each tube. Reactions were immediately transferred to a Perkin Elmer 9600 thermocycler and heated at 94°C for 2 min, followed by 36 cycles of 58°C for 1 min, 74°C for 1 min, and 94°C for 1 min and finally 58°C for 2 min and 74°C for 10 min.

Relative transcript levels of the MGMT alleles were determined by RT-PCR-RFLP in individuals heterozygous for a polymorphism in the fifth exon of the gene (designated as ExSb or Lys178Arg in Table 1). cDNA-specific primers were used to amplify the cDNA: 5′ AGCCTGGCTGAAAT-GCCCTATTTC (in exon 3, sense) and 5′ TGACGCTCTCtCCCAACCGCAG (in exon 5, reverse). The latter creates a StuI RFLP at the codon 178 polymorphism by virtue of an internal primer mismatch (underlined). The analysis of allelic balance in genomic DNA, used as equimolar control, was carried out by substituting the cDNA sense primer with an MGMT intron 4 primer (5′ TCCATGCTGAGACATAGCTGAC). The amplified fragments were visualized by gel electrophoresis after digestion with StuI. The relative abundance of each allele was quantified according to absolute fragment concentration using an Agilent 2100 bioanalyzer running a DNA 1000 LabChip (Agilent). We determined the ratios between the peak heights of the larger cut band (corresponding to the G or Arg178 allele) to the uncut band (corresponding to the A or Lys178 allele) for each of the heterozygous cDNA samples and for 12 genomic DNA samples. Analyses were performed in triplicate. The average ratio from the genomic samples was used to normalize the results. Following the criteria from Yan et al. (20) samples where one allele was over- or under-expressed by >30% from the mean were scored as showing AEI. Examples are presented in Figure 2.

Polymorphisms

Single nucleotide polymorphisms (SNPs) –60 kb apart at both intronic and exonic locations were identified from public databases or the Celera human SNP database and confirmed by sequencing the regions of interest in 10 individuals. Sequencing led to the identification of additional polymorphisms that were included in the analysis. We also included transcribed SNPs with a rare allele frequency >0.05 (14). Details are listed in Table I. Snp 1 is located

---

**G.P. Margison et al.**

---
in the promoter of the gene (25). SNPs located in the exons are designated by the prefix Exon followed by the number of the exon in which they are located.

**Genotyping**

Sequences around the SNPs of interest were amplified by PCR. The primers are listed in Table I. The PCR conditions were as follows (final concentrations): 1× Taq buffer (Promega Corporation), 100 μM dNTP, 1.0 mM MgCl₂, 0.5 μM each primer and 0.2 U Taq polymerase (Promega). DNA was amplified in a standard hot-start PCR for 35 cycles each consisting of denaturation steps of 1 min at 94°C, 1 min at the appropriate annealing temperature (see Table I) and an extension step of 1 min at 72°C. The products were visualized on a 2% agarose gel (Flowgen). Sequencing was carried out on an ABI 3100 genetic analyser using ABI Prism Big Dye Terminator Cycle Sequencing chemistry (Applied Biosystems). The genotypes were determined independently by GMG-C and MT, who did not know the case status of the subjects. Repeat analyses were performed when an ambiguous sequence was obtained.

**Quantitative trait locus (QTL) analysis**

Analysis of variance was used to ascertain associations between alleles at each of the loci genotyped and the ATase activity in PBMC. The effects in individual alleles were assumed to be additive.

**Generation of pMAL-2c-MGMTwt construct**

The MGMT cDNA (Accession no. M29971) was PCR amplified using primers (a) 5′-CGGATCCATGGAAGATGTTGAAATGACG-3′ and (b) 5′-CGGATCCATGGAAGATGTTGAAATGACG-3′. PCR products were digested and cloned into the pMAL-2c (NEB) bacterial expression vector using BamHI and EcoRI restriction sites. PCR amplifications were carried out using 1×Vent polymerase (NEB), 5 µM MgCl₂, 0.5 mM DTT, 1 mg/ml BSA, 100 mM NaCl, 60 mM Tris-Cl (pH 8.8), 20 mM MgSO₄ and 1% Triton X-100, 1 mM DnTPs (Promega), 15 pmol of each primer and 50 ng of DNA template in a total volume of 50 µl. The cycling conditions were: 1 cycle of 1 min at 95°C, followed by 25 cycles of 45 s at 95°C, 45 s at 55°C and 1 min at 72°C.

**Expression and purification of recombinant MBP proteins**

Constructs were transformed into competent XL-1 blue Escherichia coli (Novagen). Fresh cultures grown in Luria broth (500 ml with ampicillin 50 μg/ml; Sigma) were induced using 0.4 mM IPTG (Sigma) for 3 h at 37°C. Bacterial cell pellets were resuspended in 20 ml of binding buffer (20 mM Tris-HCl (pH 7.5), 5 mM EDTA, 150 mM NaCl, 3 mM DTT, 10% Glycerol and protease inhibitor cocktail) (Sigma) and extracts were prepared by sonication. The soluble fraction was applied to amylose-resin (NEB) and incubated for 2 h at 4°C. The resin was washed three times in binding buffer and the fusion proteins eluted in 10 ml of binding buffer supplemented with 10 mM maltose (Sigma). A VIVAspin 20 ml centrifugal concentrator (10 000 molecular weight cut-off Vivascience–Sartorius Group) was used to concentrate the proteins.

**Functional activity of polymorphic variants**

The kinetics of methyl group transfer from methylated substrate DNA to MGMT was determined over a 4 h period at 37°C. The effect of increasing concentrations of non-methylated DNA on methyl transfer from methylated substrate DNA was determined at 37°C. The thermal stability of the MGMT proteins was determined by pre-incubation at 51°C for up to 1 h followed by the addition of methylated DNA substrate and incubation for a further 1 h. Inactivation of MGMT by the pseudosubstrate O′-(4-bromothenyl)guanine (PaTrin-2, Patrin, Lomeguatrib) was determined by pre-incubation of MGMT protein with 10 µM PaTrin-2 for 1 h at 37°C followed by the addition of excess methylated substrate DNA and further incubation for 1 h. Samples were then processed as in the standard MGMT assay.

**Ethical approval**

Ethical approval for the studies described here was granted by South Manchester Research Ethics Committee (ER/95/217, SOU/98/157).

**Results**

**MGMT activity in PBMC**

The distribution of activities is represented in Figure 1. The mean activity is 10.3 fmol/µg DNA and the standard deviation 13.6 fmol/µg DNA. Two individuals with values of 118.8 and 125.8 fmol/µg DNA were considered outliers and excluded. The difference between the lowest (2.14 fmol/µg) and the highest activity (23.8 fmol/µg) included in the analysis is ~10-fold. There was no significant association of MGMT activity in PBMC either with gender (P = 0.5; two-sided t-test) or cancer status (P = 0.43).

**Allelic expression imbalance**

Seven of 21 blood DNA samples analysed were heterozygous for the polymorphism Ex5b and the corresponding RNA was subjected to RT-PCR analysis. Figure 2A shows that both
alleles were expressed in all samples. However, there were substantial differences in the relative messages levels (Figure 2A). Quantitative analysis revealed four cases in which the Lys178 allele was over-represented, in one case by up to 3-fold, and in two cases the Arg178 allele was over-represented. One case was scored as not showing AEI (Figure 2B) but in all of the other analyses, AEI was detected and repeat analyses of the same samples showed that this was statistically significant ($P < 0.05$).

**SNP genotyping**

Table I shows details of the polymorphisms used. We did not detect any significant departure from Hardy-Weinberg equilibrium in our samples for any of these polymorphisms. The data in Table II suggest that in the 5' half of the gene, linkage disequilibrium extends over comparatively large regions and is significant for markers >180 kb apart. This contrasts with the 3' end of the gene where significant linkage disequilibrium is only detectable for markers <60 kb apart.

**QTL analysis of intragenic polymorphisms**

Initially 100 individuals were typed for all polymorphisms. The results of the analysis for each polymorphism are separately presented in Figure 3 (black bars). The polymorphisms at Snp 2, Snp 3, Snp 9, Exon 5a and Exon 5b showed the most significant association with MGMT activity in PBMC. The remaining individuals were typed for these markers (Figure 3, grey bars, Snp 9 has a $P$-value of 0.0412) Significant associations are seen in two regions. One (Snp 1--Snp 5) is characterized by markers at the 5' end of the gene and in the first intron and the other (Snp 9, Exon 5a, Exon 5b) at the 3' end. Analysis of marker combinations shows that the data are significantly better described when sites in the 3' and in the 5' region are included, suggesting that these reflect independent influences on MGMT activity. The best fit is achieved by the combination of one of the exon 5 polymorphisms and Snp 3 ($P = 3 \times 10^{-6}$, multiple $R^2 = 0.19$). The mean, standard deviation and range of MGMT values by genotype for these two polymorphisms are shown in Table III. There is no significant evidence for an interaction term. No significant improvement can be achieved by considering other markers in the 5' half in addition to Snp 3, or additional makers in the 3' half in addition to an exon 5 marker, suggesting that markers within these regions may reflect the same functionally important sites.

**Functional activity of polymorphic variants**

The variant MGMT proteins showed very closely similar functional characteristics including methyl transfer rates (Figure 4A), inhibition by non-methylated DNA (Figure 4B).
and thermal stability at 51°C (Figure 4C). However, the Ile\textsuperscript{143}-Lys\textsuperscript{178} and Ile\textsuperscript{143}-Arg\textsuperscript{178} proteins were significantly more sensitive to inactivation by PaTrin-2 than the Val\textsuperscript{143}-Lys\textsuperscript{178} and Val\textsuperscript{143}-Arg\textsuperscript{178} alleles (Figure 5, \(P < 0.001\)).

Discussion

The variation in MGMT activity levels observed in our material is similar to that reported in other studies (reviewed in ref. 4). Allelic expression analysis of MGMT in PBMC showed that in most of the informative samples the two alleles were not expressed at the same level. This is analogous to earlier observations in normal lung tissues (22) and suggests that differences in message levels may be one of the genetically determined mechanisms generating the inter-individual variation in MGMT activity. It also indicates that at least some of this variation maps close to or within the MGMT locus and acts in \textit{cis}. The fact that the same marker allele was over- or under-represented in different samples suggests that either the determinants of allele specific differences in expression are not in strong linkage disequilibrium with the marker used or that more than one \textit{cis}-acting site is involved in the regulation

### Table II. Linkage disequilibrium for sites across the MGMT locus: the values of \(\text{LD}^0\) are presented

<table>
<thead>
<tr>
<th>Position(^a)</th>
<th>Snp 2</th>
<th>Snp 3</th>
<th>Snp 4</th>
<th>Snp 5</th>
<th>Snp 6</th>
<th>Snp 7</th>
<th>Snp 8</th>
<th>Exon 3a</th>
<th>Exon 3b</th>
<th>Snp 9</th>
<th>Snp 10</th>
<th>Exon 5a</th>
<th>Exon 5b</th>
</tr>
</thead>
<tbody>
<tr>
<td>1539</td>
<td>0.93</td>
<td>0.63</td>
<td>0.32</td>
<td>0.79</td>
<td>0.60</td>
<td>0.25</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.07</td>
<td>0.13</td>
<td>0.14</td>
<td>0.17</td>
</tr>
<tr>
<td>60368</td>
<td>0.89</td>
<td>0.72</td>
<td>0.84</td>
<td>0.66</td>
<td>0.12</td>
<td>0.36</td>
<td>0.05</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
<td>0.11</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>60408</td>
<td>0.77</td>
<td>0.96</td>
<td>0.82</td>
<td>0.34</td>
<td>0.08</td>
<td>0.14</td>
<td>0.14</td>
<td>0.11</td>
<td>0.03</td>
<td>0.13</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>114259</td>
<td>1.00</td>
<td>0.87</td>
<td>0.17</td>
<td>0.24</td>
<td>0.03</td>
<td>0.03</td>
<td>0.24</td>
<td>0.22</td>
<td>0.43</td>
<td>0.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>174006</td>
<td>0.03</td>
<td>0.08</td>
<td>0.03</td>
<td>0.03</td>
<td>0.04</td>
<td>0.25</td>
<td>0.22</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>234002</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>234045</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>241261</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.15</td>
<td>0.08</td>
<td>0.16</td>
<td>0.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>241352</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>294780</td>
<td>1.00</td>
<td>0.85</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>294879</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300133</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300239</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The values of \(\text{LD}^0\) are presented from ref. 36. Values not significantly different from zero at the 0.05 level are in grey. \(^{a+}\): as distance from Snp 1.

### Table III. MGMT values by genotype for polymorphisms Snp 3 and Exon 5a (codon 143)

<table>
<thead>
<tr>
<th>Polymorphism</th>
<th>Genotype</th>
<th>Number of individuals</th>
<th>MGMT (fmol/μg DNA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Snp 3</td>
<td>A/A</td>
<td>62</td>
<td>10.1</td>
</tr>
<tr>
<td></td>
<td>A/G</td>
<td>54</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td>G/G</td>
<td>11</td>
<td>7.2</td>
</tr>
<tr>
<td>Exon 5a (codon 143)</td>
<td>Ile/Ile</td>
<td>94</td>
<td>8.4</td>
</tr>
<tr>
<td></td>
<td>Ile/Val</td>
<td>34</td>
<td>9.8</td>
</tr>
<tr>
<td></td>
<td>Val/Val</td>
<td>3</td>
<td>15.5</td>
</tr>
</tbody>
</table>
of expression. However allelic expression differences can be due to a variety of causes. These include epigenetic mechanisms, such as imprinting or allelic exclusion, and somatic mutations. Depending on when and where such changes occur they can give rise to imbalanced expression in a range of tissues. However if AEI were exclusively a consequence of such mechanisms, there would be no relationship between expression levels and variation at the DNA level. We therefore used QTL analysis to establish whether or not there is such a relationship.

QTL analysis using intragenic polymorphisms reveals that there are at least two sites significantly associated with MGMT activity. This implicates variation at the DNA levels as a cause of allelic expression differences although it does not exclude the involvement of epigenetic mechanisms in mediating the effect of sequence variation on expression. The first of the two regions is characterized by markers in the first intron and the second by markers in the fifth exon. Together they account for 19% (95% CI 8–33%) of the variance observed in our sample.

Janssen et al. analysed MGMT activity in a group of individuals sampled repeatedly over a period of up to 120 days (15), Figure 1). According to their data, inter-individual variance represents 40% of the total variance (95% CI 31–71%). Together with our data, this suggests that intragenic polymorphisms account for a substantial proportion of the genetic variance.

Our results indicate that linkage disequilibrium at the 5' end of the gene can be detected for markers separated by >180 kb. In this region, the marker with the strongest association with activity is located in the 3' end of the first intron. However, since the MGMT promoter, as defined by Harris et al. (12), includes elements in the first exon and in the adjacent intron, and given the extent of linkage disequilibrium in that region, we cannot exclude the possibility of the causative change being in the promoter region of the gene. The effect of such a causative change may be mediated by methylation or other epigenetic modifications. This is a particularly interesting proposition given that methylation of the MGMT promoter is
found in a variety of human tumours. However, the association of expression levels with polymorphisms in this region indicates that the primary cause resides in the variation of the DNA sequence and intra-individual differences in expression are not solely a consequence of epigenetic modifications.

The two SNPs at the 3' end of the gene that are most strongly associated with activity both lead to amino acid changes. According to the terminology proposed in (26), they are in perfect linkage disequilibrium and represent essentially a single biallelic system with the alleles Ile143-Lys178 and Val143-Arg178. In our sample we found only one Val143-Lys178 and no Ile143-Arg178 alleles. The most common allele has isoleucine at position 143. This is close to the cysteine residue at position 145 that acts as alkyl group acceptor. This region is strongly conserved and isoleucine can be found at orthologous positions in species as distantly related to mammals such as Fugu rubripes and even Drosophila melanogaster. However, valine, the residue present in the alternative human allele can be found in Saccharomyces cerevisiae, in some of the bacterial MGMT genes, and in one of the two Caenorhabditis elegans MGMT homologues (14,27). A study by Kaur et al. reported an association between the MGMT genotype at position 143 and lung cancer risk (28), but this was of borderline significance and awaits confirmation; no associations have been reported for other cancer types (25). Ma et al. reported a higher frequency of the Val143-Arg178 allele among melanoma patients who did not respond to chemotherapy, but the difference was not statistically significant (29). The frequency of Val143 seems to vary widely: Kaur et al. failed to detect it in 35 probands of Asian origin and reported a frequency 0.03 in African Americans (81 probands) and of 0.07 in Caucasians. In a Swedish control population (76 samples), Egyhazi et al. (25) reported a frequency of 0.11 and in our series, the frequency was higher at 0.16 (130 samples).

The question of whether or not the 143/178 polymorphism by itself has a bearing on the function of the MGMT protein was recently addressed by Ma et al., who found no differences between the alleles using an E.coli MNNG survival assay (29). Mijal et al. reported no significant differences in the ability to process 6-benzyl, butyl or [4-oxo-4-(3-pyridyl)butyl]guanine (30). We generated the four variant proteins by site directed mutagenesis and found that their kinetics of methyl transfer, inhibition by non-methylated DNA and thermal stability were indistinguishable under the assay conditions that we used. In contrast, Val143 alleles were significantly more resistant to inactivation by the pseudosubstrate PaTrin-2 than Ile143 variants, irrespective of the residue at position 178. While the differences were slight, they suggest that the active site pocket of MGMT may be affected by the change and this may have an impact on the processing of certain types of DNA lesions (see below) or of other substrates, such as the inhibitors used in chemotherapy (1,16,17). In our analysis, in a purely additive model, the Val143 allele is associated with 'higher' MGMT specific activity than the common allele (see Table III). This would be consistent with the higher activity being the result of more efficient and continuous inactivation of the Ile143 protein by as yet unidentified endogenous substrates. However, we cannot exclude the possibility that genetic variation in this region also affects the levels of activity through other mechanisms. These may include modulation of the stability or processing of the protein or of the transcript, alteration of the efficiency of transcription or that some other polymorphic site is responsible for differences in specific activity: that site would be in linkage disequilibrium with the markers we used at the 3' end of the gene and would affect some unknown regulatory element. However, even without the knowledge of the causative mechanisms, our results allow the inference of MGMT activity levels based on the genotype.

Recently several groups have investigated associations between cancer risk and the codon 143 polymorphism or the tightly linked polymorphism at codon 178. While some of these studies find associations (31,32), others fail to do so (33,35). This could reflect genuine biological differences. For example, differences in the ability of different alleles to process specific substrates may lead to an association between MGMT genotype and cancer susceptibility only when particular carcinogens are involved. However, our results also indicate that at least two sites influence MGMT activity levels emphasizing the need for studies based on series large enough to detect associations even in the presence of this additional complexity.

Previous studies have identified human MGMT variants with reduced protein half-life (19) and sensitivity to O6-benzylguanine, a commonly used MGMT inhibitor (18), but these are rare, in most studies reporting a frequency well below 0.01 (1,14,19,34). In contrast the Val143 allele is comparatively common being carried by 28% of the probands in our study. In view of the chemotherapeutic use of O6-alkylating agents, and of the possible introduction of MGMT inactivators into clinical practice, potential differences in the processing of cytotoxic lesions in DNA, or in the sensitivity to pseudosubstrates, such as PaTrin-2, may be of clinical relevance. It, therefore, seems reasonable to suggest that this should be considered in ongoing and future clinical trials of these agents.

Acknowledgements

The authors would like to thank the staff at the Bronchoscopy Unit, North West Lung Centre, Wythenshawe Hospital. This work was supported by Cancer Research UK, the British Lung Foundation, the Roy Castle Lung Cancer Foundation and the Colt Foundation. Published by his colleagues in memory of Kevin de Freyne in recognition of his friendship and dedication to his work.

Conflict of Interest Statement: None declared.

References


Received January 19, 2005; revised March 15, 2005; accepted March 27, 2005