A gene transcribed from the bidirectional ATM promoter coding for a serine rich protein: amino acid sequence, structure and expression studies

Philip J. Byrd*, Paul R. Cooper+, Tatjana Stankovic, Harjit S. Kullar, Giles D. J. Watts, Paula J. Robinson and A. Malcolm R. Taylor

CRC Institute for Cancer Studies, The University of Birmingham Medical School, Birmingham B15 2TJ, UK

Received June 21, 1996; Revised and Accepted August 1, 1996 DDBJ/EMBL/GenBank accession no. X97186

In an earlier report we showed that the 5′ end of the gene for ataxia telangiectasia ATM is within 700 bp of the 5′ end of a novel gene E14, and suggested that the CpG island that separates these genes functions as a bidirectional promoter. We have now determined the complete amino acid sequence of the E14 protein, defined the exon/intron structure of the gene and estimate that the complete gene is more than 55 kb in length. The E14 gene appears to be a housekeeping gene that is expressed in all tissues, including all parts of the brain. The E14/ATM promoter organisation is conserved in man, monkey and mouse, although the mouse promoter is more compact and appears to lack two of the four putative Sp1 boxes found in the human promoter. Reporter gene constructs showed that the human and mouse E14/ATM promoters were indeed bidirectional, that the ATM side of the human promoter was three times stronger than the E14 side, and that the mouse promoter (in human cells) directed transcription with equal efficiency in both directions, but at a lower level than the human promoter. Analysis of a small number of A-T patients for mutations in the promoter region or the E14 coding sequence did not provide evidence to suggest that E14 contributes to the A-T phenotype.

INTRODUCTION

In the course of completing the 5′ half of the cDNA for the gene for ataxia telangiectasia ATM (1–3) we identified a CpG island immediately upstream of the end of a 5′ ATM RACE product. Within this island and ∼700 bp upstream of the 5′ end of the ATM RACE product we found the first exon of a gene that we had isolated from this region by exon trapping (2). This novel gene, which we called E14, appeared to be transcribed in the opposite direction to the ATM gene from an intergenic region that had features of a bidirectional promoter. Since the proteins encoded by some divergently transcribed genes have been found to interact or to be involved in the same biochemical pathway (4–6), we decided to characterise the E14 protein as a protein that potentially interacts with the ATM protein or which is involved in the same pathway as the ATM protein.

At the present time there is very little known about the biochemical properties or function of the ATM protein, beyond a superficial knowledge of its role in the response to DNA damage and its apparent involvement in the maintenance of the stability of the genome (7). The ATM protein has a phosphatidylinositol 3-kinase (PI 3-kinase) like domain at its C-terminus (1), which is a feature common to a family of yeast, Drosophila and mammalian genes that are variously involved in cell cycle control, DNA repair, mitotic chromosome stability and meiotic recombination (8–11). The yeast TOR2 and mammalian RAFT1 genes are, so far, the only members of this family which have been shown to have lipid kinase activity (12,13). It has been suggested that the ATM protein is functionally related to DNA-PK, a member of the PI 3-kinase family which has serine/threonine protein kinase activity, on the basis that both are involved in aspects of the repair of ionizing radiation damaged DNA and both have a role in the recombination of immune system genes (7,11). By analogy with DNA-PK, which functions as a heterotrimer of a 450 kDa catalytic subunit DNA-PKcs and the 70 kDa and 80 kDa Ku antigens (14), the ATM protein might be expected to function as a complex with several other proteins, one of which could possibly be the E14 protein.

In this report we present the complete amino acid sequence of the E14 protein, the exon/intron structure of the gene, and expression analysis that suggests that it is a housekeeping gene. We also compare the E14/ATM promoter sequences of man, monkey and mouse and evaluate the bidirectional activity of two of these promoters. Finally, we report our investigation to identify mutations in the promoter region and the E14 coding sequence in a small number of A-T patients.

RESULTS

The E14 gene was identified originally (2) as an exon which we isolated from cosmids that we cloned from an ATM region YAC. Initially a series of cDNA clones were assembled into a 1200 bp contig which was found to have an open reading frame (ORF) which started 35 bp in from the 5′ end but encountered no in frame

*To whom correspondence should be addressed
+Present address: Department of Human Genetics, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263, USA
stop codons downstream. Subsequently, this contig was extended by a combination of RACE and the isolation of additional cDNA clones into a sequence of ∼4.5 kb which contained a 4281 bp ORF. Further cDNA library screens for non-coding sequences at the 3′ end identified two polyadenylated overlapping clones, each of which contained canonical polyadenylation sequences AA TAAA 30–40 bp upstream of their poly-A tracts. The complete E14 cDNA sequence is 5867 bp (GenBank accession no. X97186) excluding the poly-A tail. The shorter sequence, produced by the internal polyadenylation signal, is 5672 bp. The ORF is predicted to encode a serine rich protein of 1427 amino acids (Fig. 1) with a molecular weight of ∼157 kDa. The function of this gene is unknown and a protein database search failed to suggest a possible function because no significant matches were found to any motif or domain in any other protein for which a function is known.

Analysis of the exon/intron structure of the gene identified 16 exons (Table 1). The largest exon was found to be the last, containing 1210 bp of coding sequence and 1552 bp of untranslated sequence. At 1653 bp exon 13 was found to be the next largest, accounting for 39% of the coding sequence. The remaining exons ranged in size from 41–225 bp, giving an average of 104 bp. The smallest intron was found to be 82 bp whilst the largest amplifiable intron was ∼10 kb; we could not amplify the first intron which must be in excess of 12 kb. Taken together the length of the coding sequence and the lengths of the introns that we could size indicates that the E14 gene is at least 43.5 kb and probably a minimum of 55 kb if our conservative estimate of the size of the first intron is included. The donor and acceptor splice sites all conform to the gt-ag consensus (15). Of the intron types 0, 1 or 2 (16), type 1 is the most common (8/15), whilst type 0 and type 2 are found in 3/15 and 4/15, respectively. Introns, and therefore exons, of the same type tend to occur in groups due to exon shuffling in the evolution of the gene (17), and as such may define domains or structural features. The consecutive introns 11–14 are all type 1, thus exons 12–14 are spliced symmetrically at the type 1 position and as such may represent a functional domain encompassing nearly 50% of the E14 protein.

Hybridisation of a zoo-blot with a probe representing the entire E14 coding sequence showed that the E14 sequence has been conserved through mammalian evolution with hybridizing bands being apparent in monkey, rat, mouse, dog, cow and rabbit (data not shown). A very weakly hybridising band was just discernible in chicken DNA after 15 days’ exposure, suggesting that E14 related sequences might be present in birds. A more obvious discrete band was identified in yeast DNA after equally long exposure, raising the possibility that E14 sequences have been conserved in lower eukaryotes. However, further Southern blotting experiments of yeast DNA digested with five different restriction endonucleases showed that the hybridizing bands, which again could be seen only after a long exposure, were those that could also be seen on the ethidium bromide stained gel prior to blotting (data not shown). This suggested that the E14 probe was hybridising to sequences of a repetitive nature in the yeast genome. Additionally, a search of the complete yeast DNA database failed to identify potentially homologous sequences at either the DNA or protein level, other than those with a very high serine content.

Analysis of the expression of the E14 gene showed that it was expressed in spleen, thymus, prostate, testis, ovary, small intestine, colon and peripheral blood leukocytes with expression apparently being highest in testis, taking the loading of mRNA in each track into account (Fig. 2). Expression was also found in heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas (data not shown), in fact in all tissues examined. Interestingly, expression was evident in all parts of the brain including the cerebellum (Fig. 2), which is of particular importance in terms of A-T. E14 therefore appears to be a housekeeping gene of unknown function. The E14 probe detected mRNA species of ∼8.8 kb, ∼6.25 kb and ∼5.3 kb by northern blotting. The 6.25 kb species was the most abundant, followed by the 5.3 kb, whilst the 8.8 kb mRNA was present at the lowest levels. The 6.25 kb species approximates in size to the 5.9 kb polyadenylated cDNA whilst the 5.3 kb species is smaller than expected for the 5.7 kb polyadenylated cDNA; the origin of the least abundant 8.8 kb mRNA is unclear. The relative abundance of these three mRNA species was approximately the same in different tissues.

In an earlier report we presented the sequence of a CpG island which we found upstream of the end of a 5′ ATM RACE product
Table 1. Exon/intron organisation of E14

<table>
<thead>
<tr>
<th>Exon No.</th>
<th>5' Intron sequence</th>
<th>Exon first base</th>
<th>Exon length (bp)</th>
<th>Exon last base</th>
<th>3' Intron sequence</th>
<th>Size of 3' intron (Kb)</th>
<th>Intron type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ttttttttttgg</td>
<td>CTTCTC</td>
<td>71</td>
<td>TATGCG</td>
<td>gctggggg</td>
<td>&gt;12</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>ccccccctttttttttt</td>
<td>GAGGC</td>
<td>71</td>
<td>ATTGTTTTT</td>
<td>ccccccctttttttttt</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>ccccccctttttttttt</td>
<td>AAGACT</td>
<td>71</td>
<td>GTTCCTTTC</td>
<td>acagtgtgtgtgtgtgtgt</td>
<td>1.4</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>ccccccctttttttttt</td>
<td>GGCACC</td>
<td>71</td>
<td>CTTTGTTTTT</td>
<td>ccccccctttttttttt</td>
<td>1.4</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>ccccccctttttttttt</td>
<td>GAGGC</td>
<td>71</td>
<td>ATTGTTTTT</td>
<td>ccccccctttttttttt</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>ccccccctttttttttt</td>
<td>AAGACT</td>
<td>71</td>
<td>GTTCCTTTC</td>
<td>acagtgtgtgtgtgtgtgt</td>
<td>1.4</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>ccccccctttttttttt</td>
<td>GGCACC</td>
<td>71</td>
<td>CTTTGTTTTT</td>
<td>ccccccctttttttttt</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>ccccccctttttttttt</td>
<td>GAGGC</td>
<td>71</td>
<td>ATTGTTTTT</td>
<td>ccccccctttttttttt</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>ccccccctttttttttt</td>
<td>AAGACT</td>
<td>71</td>
<td>GTTCCTTTC</td>
<td>acagtgtgtgtgtgtgtgt</td>
<td>1.4</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>ccccccctttttttttt</td>
<td>GGCACC</td>
<td>71</td>
<td>CTTTGTTTTT</td>
<td>ccccccctttttttttt</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>ccccccctttttttttt</td>
<td>GAGGC</td>
<td>71</td>
<td>ATTGTTTTT</td>
<td>ccccccctttttttttt</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>ccccccctttttttttt</td>
<td>AAGACT</td>
<td>71</td>
<td>GTTCCTTTC</td>
<td>acagtgtgtgtgtgtgtgt</td>
<td>1.4</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>ccccccctttttttttt</td>
<td>GGCACC</td>
<td>71</td>
<td>CTTTGTTTTT</td>
<td>ccccccctttttttttt</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>ccccccctttttttttt</td>
<td>GAGGC</td>
<td>71</td>
<td>ATTGTTTTT</td>
<td>ccccccctttttttttt</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>ccccccctttttttttt</td>
<td>AAGACT</td>
<td>71</td>
<td>GTTCCTTTC</td>
<td>acagtgtgtgtgtgtgtgt</td>
<td>1.4</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>ccccccctttttttttt</td>
<td>GGCACC</td>
<td>71</td>
<td>CTTTGTTTTT</td>
<td>ccccccctttttttttt</td>
<td>1.0</td>
<td>1</td>
</tr>
</tbody>
</table>

Sequence for introns is in lowercase and for exons, uppercase. Intron sequences in bold type are intron splice sites (5'→3') that conform to the gt-ag consensus. Intron types are defined by the position of the splice relative to the codon; type 0 is between codon boundaries; type 1, after the first base and type 2, is after the second base.

and suggested that this represented a bidirectional promoter controlling the ATM and E14 genes (2). Primers directed to the human E14 and ATM genes, which were designed to PCR the putative promoter, were found to amplify a sequence of the same size from rhesus monkey DNA. The homologous mouse sequence was isolated from a lambda genomic clone that was identified by plaque hybridisation using the human promoter as a probe. Comparison of the monkey and human promoter sequences showed that they were essentially identical, the only major difference being an 18 bp deletion in the monkey that includes a putative Sp1 box (Sp1 box IV) in the human; the Sp1 box in the monkey sequence appears to be in an adjacent 18 bp region where the identity between the human and monkey sequences drops to 75% (data not shown). In both the monkey and mouse sequences the E14 and ATM genes were found to be arranged in a head to head configuration either side of a CpG rich region. The mouse and human sequences were found to be 81% identical in a 247 bp region starting 17 bp into the first exon for E14 and extending back into the promoter sequence (Fig. 3). Within this region the Sp1 boxes I and II, which also contain insulin responsive elements, are entirely conserved, as is a CCAAT box. Additionally, 12 out of 14 bp in a region containing overlapping homologies for several transcription factor binding sites were found to be identical. Beyond this region, on the ATM gene side of the promoter region the homology between the mouse and the human sequences was poorer. The mouse sequence is 63 bp shorter in the region between the box containing the transcription factor binding sites and the initiation codon for the ATM gene, and possesses little homology to the human Sp1 boxes III and IV. The CCAAT box identified in this region in the human sequence is perfectly conserved in the mouse sequence.

Functional analysis of promoter activity was assessed using a luciferase reporter gene assay in an EBV negative Burkitt lymphoma cell line DG75. The human sequence (from –33 to –680 relative to the ATG initiation codon) was three times more active when orientated such that the ATM side of the promoter was closest to the luciferase gene than in the reverse, E14 orientation (Fig. 4A). Transcription of the luciferase gene was also 50% more efficient in the ATM orientation compared with that under the control of the reference SV40 promoter. In the same human cell line the mouse promoter (from –35 to –593) directed transcription of the reporter gene with equal efficiency in either orientation. The human promoter was 6.8-fold more active in the ATM orientation and 2.2-fold more active in the E14 orientation, compared with the mouse promoter in either orientation (Fig. 4B). The mouse promoter appeared to initiate transcription with equal efficiency in either orientation in mouse NIH3T3 cells (Fig. 4C). Because transcription from the reference SV40 promoter was much lower in NIH3T3 cells than DG75 cells, the absolute level of mouse promoter activity in these cell lines could not be compared.
The E14/ATM promoter region was analysed for mutations in five A-T patients who, by haplotype analysis, we knew to be compound heterozygotes (as are the vast majority of A-T patients). Three of these patients have an ATM splicing mutation that confers a milder form of A-T (18), but no mutations have been found in their other ATM alleles. Only one mutated ATM allele had been identified in one of the two remaining patients analysed in this group. Of the six ATM alleles in which we have so far not identified mutations, none was found to contain promoter region mutations. The only sequence variation we did identify within the promoter sequence was a SacII restriction enzyme site polymorphism that was also found in normal individuals (data not shown). A different set of five patients, in whom no mutations have been demonstrated in the ATM gene, were analysed for mutations in the E14 coding sequence. Each patient had two different haplotypes for markers across the region, suggesting that like patients with ATM mutations, they would also be compound heterozygotes. However, REF analysis using three restriction enzymes for each fragment analysed failed to show any mutations in the 10 different E14 alleles in these patients.

**DISCUSSION**

This report describes the amino acid sequence of the E14 gene that is transcribed divergently from a promoter region that is shared with the gene for A-T. The molecular weight of this sequence of 1427 amino acids is predicted to be 157 kDa. The major feature of interest in the exon/intron structure of the gene is the length of exon 13, which at 1653 bp encodes over a third of the E14 protein. In this respect the E14 gene resembles the BRCA 1 (19) and BRCA 2 (20) genes in which exon 11 is very much larger than the other exons (3426 bp and 4932 bp respectively). Two major E14 mRNAs, which were identified by northern analysis, may be the result of alternative poly-A signal usage because two different poly-adenylated E14 cDNAs were identified in a cDNA library screen. Both major mRNA species and a less abundant larger transcript were expressed in all tissues analysed, indicating that this is a housekeeping gene important to all cells and, on the basis of phylogenetic analysis, to all mammalian species. A very weakly hybridising band was barely discernible in chicken DNA after a long exposure time, suggesting that E14 might also be conserved in birds. Imai et al. (23) identified a relatively weak but more obvious signal in chicken using NPA T (nuclear protein mapped to the A-T locus; E14 by a different name) as a probe. In the same study they also identified a relatively weak but positive signal in yeast and their conclusion was that the NPAT gene has been conserved throughout eukaryotic evolution. Our analysis suggests that the hybridisation to yeast DNA is spurious but that the conservation of E14 might extend to birds.

Fourteen per cent of the E14 amino acids are serine residues and in protein database searches proteins with an even higher serine content, e.g. yeast glucosamylase S1/S2 precursor protein (21) and yeast suppressor protein SRP40 (22), are identified as the best matches. However, the matches are purely on the basis of amino acid content and not on the basis of the sequence of amino acids...
in the E14 and ‘matched’ proteins. Taken together, the serine and threonine residues comprise 21% of the E14 protein, suggesting that this protein might be a target for a member of the protein-serine/threonine kinase family. Our analysis of the E14 protein sequence for motifs or domains that have been found in proteins or protein families with known function failed to reveal any significant matches, precluding speculation on the function of E14. However Imai et al. (23) performed similar analyses identifying weak homologies with protein-tyrosine phosphatase zeta precursor and human nucleoporin NUP214, and possible motif similarities to phosphorylation sites for CDK4/E2F, CDK6/E2F, PKC and cAMP/cGMP. They also identified two putative ATP/GTP-binding domains and raised the possibility that two basic amino acid rich regions at the C-terminus might represent a nuclear localization signal. The significance of these hypothetical domains is, however, unclear because the match to the consensus sequence is limited, and whilst the putative nuclear localization signal sequences are provocative, the subcellular localization of E14 protein needs to be determined experimentally.

Furthermore, this was found to be true of the human and the mouse promoters. For the human promoter the ATM side initiated transcription 50% more efficiently than the SV40 reference promoter whilst the E14 side was 50% less efficient than the reference. In contrast the mouse promoter was equally efficient in either orientation but the level of transcription initiation was only 20% that of the SV40 promoter. Whilst we cannot compare promoter strength in NIH3T3 and DG75 cells directly, the relative transcription efficiencies of the E14 and ATM sides of the human and mouse promoters. The presence of four Sp1 boxes in the human promoter, compared with only two in the mouse promoter, might account for the strength of the ATM side of the human promoter, a possibility that could be investigated by deletion mutagenesis of the human promoter in a reporter vector. Interestingly, the two Sp1 boxes that are retained in the mouse are in the side of the promoter region that is 81% conserved between mouse and man, and are those that contain the putative insulin receptor elements. This side of the intergenic region might conceivably be responsible for the basal levels of transcription of both E14 and ATM, whilst the other side might be more intimately involved in perhaps regulating tissue specific levels of ATM expression or in determining the level of ATM transcription in response to ionising radiation induced damage. In terms of the effect of possible promoter mutations, this hypothesis would predict that mutations in the conserved region on the E14 side of the intergenic region would be the most severe. Further investigations are required to identify regulatory sequence elements e.g. initiator elements (24) that are important for promoter activity, and to address the question of whether this intergenic region represents an integrated bidirectional promoter or fused independent promoters (4).

The close proximity of the E14 gene to the ATM gene and the shared use of an intergenic promoter region raises the possibility that the proteins encoded by these genes interact directly, are involved in different steps of the same pathway or have, perhaps, different regulatory functions in a common process. There are precedents for genes that are divergently transcribed from a bidirectional promoter being involved in the same pathway, e.g. GAT and AIC in purine nucleotide synthesis (4), in the same protein complex, e.g. α1(IV) and α2(IV) collagen genes (5), or being functionally linked in the same process, e.g. TAP1 and LMP2 in antigen processing (6). Also, different genes which are in close proximity on the chromosome can contribute to the phenotype of the same disease, e.g. the SMN and NAIP genes in spinal muscular atrophy (25,26), or produce the same phenotype when mutated, e.g. the RAG-1 and RAG-2 genes involved in V(D)J recombination.
recombination (27). If the E14 and ATM genes do function as a complex or do function in the same pathway, a proportion of A-T patients might be expected to have mutations in E14, particularly those in which no ATM mutations can be found. The limited screen performed in this study found no evidence of mutations either in the E14/ATM promoter or in the E14 cDNA sequence in 10 A-T patients. Our assumption has been that there is only one promoter for the A-T gene, but the identification of multiple 5′ coding exons (28) might suggest the existence of one or more additional promoters, and possible sites for mutations in the A-T patients we investigated in this study. Analysis of a larger number of patients would be necessary to rule out the possibility that mutation of the E14 gene could be the cause of A-T in some individuals. And biochemical analysis, involving the use of immunoprecipitating antibodies, will be necessary to investigate the possible involvement of the ATM and E14 proteins in a functional complex or at different points in a common pathway or process.

MATERIALS AND METHODS

RACE and cDNA library screening

A human foetal thymus cDNA library cloned in λDR2 (Clontech) was arrayed and pooled in a 96-well format (33) for screening by PCR. Clones specific to exons that were trapped from cosmids which derived from the A-T region were identified and partially purified by PCR of the arrayed thymus cDNA library, and isolated in a final hybridization step. Several exons identified cDNA clones that formed a contig that was initiated by exon E14. The sequence of this contig was determined and extended in the 3′ direction by RACE using RACE Ready human foetal brain and placental cDNA libraries (Clontech). Clones overlapping the 3′ end of the RACE sequence were identified in the arrayed thymus library by PCR. The sequence of the extension to the contig was obtained from products of a PCR in which an E14 specific primer was used in conjunction with a vector specific primer. Subsequently the sequence of the 3′ end of the E14 cDNA sequence was determined from clones isolated from a human foetal brain cDNA library (Stratagene).

DNA sequencing and exon/intron determination

The sequence of the E14 cDNA was determined using a vectorette sequencing strategy described previously (2). The same method was used to define the exon/intron boundaries, but in this case the substrate for vectorette PCR sequencing was restriction enzyme digested, vectorette ligated aliquots (10–20 ng) of a 150 kb PAC clone 51A6 that was found to be positive for the E14A-T promoter region and the centromeric flanking gene ACAT. Gel purified, agarased PCR products were sequenced using an applied Biosystems 373A DNA sequencer, and sequence assembly and analysis was performed using applications of the GCG program as described previously (2).

Analysis of E14 gene conservation and expression

Zoo blots (Clontech; no. 7753–1) and multiple tissue northern blots (Clontech; no. 7755–1, no. 7759–1 and no. 7760–1) were hybridised with the complete E14 cDNA sequence and washed according to the manufacturers recommendations. Hybridisation signals were detected using Kodak X-OMAT AR film; exposures for up to 15 days were required to detect the weakest signals.

Construction of plasmids

Human E14/ATM promoter fragments were amplified using primers 8740: 5′-TTGTGAGATCTGTACAAACACAG-CAGGAAACCACAATAAG-3′ and 9024: 5′-GTAAGTGAGATCTCTGACATCGGAAGGGTTCAAGTTACATCA-3′. Both primers contained 5′/3′ ends to facilitate cloning. The human E14/ATM promoter was cloned initially into the BamH I site of pUC19 and then subcloned directionally on SacI/Hind III fragments, in either the ATM or E14 orientation, upstream of the firefly luciferase gene of pGL2-Basic (Promega); the resulting constructs were called pHATP and pHE14P. Mouse E14/ATM promoter fragments were amplified using the primers 9445: 5′-GTAAGTGAGATCTGTACAAACACAGC-3′ and 9446: 5′-GTAAGTGAGATCTGACATCGGAAGGGTTCAAGTTACATCA-3′ and cloned into pGL2-Basic using the same strategy as above; the resulting constructs were designated pMATP and pME14P.

The orientation and sequence of each construct was checked by DNA sequencing. The pCMV-βgal construct (34) was a gift from Prof. Martin Rowe. The reference plasmid SV40-LUC was obtained from a commercial source (Promega). Qiagen Plasmid Maxi kits were used to purify plasmids for transfection experiments.

DNA transfection and luciferase assay

The EBV negative Burkitt cell line DG75 was maintained in suspension in RPMI 1640 growth medium supplemented with 10% bovine serum and antibiotic gentamycin sulphate. NIH3T3 murine fibroblast cells were maintained in DME supplemented with 10% bovine serum and the antibiotics penicillin and streptomycin. Cells were washed and suspended at 2.5 × 10^7 cells/ml in HEPES buffered DME growth medium supplemented with 10% bovine serum, prior to transfection. 10^7 cells were co-transfected with 10 μg of reporter plasmid DNA (pHATP, pHE14P, pMATP, pME14P, pGL2-Basic or pSV40-LUC) and 2 μg of pCMV-βgal (an internal control used to monitor transfection efficiency). Transfection was by electroporation (Bio-Rad Gene Pulser) using 0.4 cm electrode gap cuvettes (Bio-Rad) under conditions optimised for each cell line; 270 V, 300 V and 1000 μF for DG75, 300 V and 1000 μF for NIH3T3. After electroporation the cells were incubated at 37°C in growth medium for 48 h. Luciferase activity was assayed according to Huen et al. (34). Cells washed in phosphate buffered saline were lysed in 500 μl of 100 mM HEPES pH 8.0, 2 mM MgCl2, 5 mM DTT and 2% Triton X-100. Cell debris was pelleted, 50 μl supernatant was mixed with 100 μl of luciferase assay reagent (Promega) and light emission was measured immediately in a Berthold LB 9501 luminometer. Light release was integrated for 10 s. β-gal activity was assayed (34) using 20 μl of the same cytosolic extract.

Mutation analysis

E14/ATM promoter mutation analysis was performed on DNA extracted from lymphoblastoid cell lines. Promoter sequences were amplified from genomic DNA using the primers 8434: 5′-CAAGC-CCGCGCTACGTCCGGAAGGTGACAAACACAGCAGGAAACCACAATAAG-3′ and 8435: 5′-CTGCACTCCGGAAGGTGACAAACACAGCAGGAAACCACAATAAG-3′. PCR products were analysed for mutations using the restriction endonuclease fingerprinting (REF) method (35). Restriction
enzymes used: BrUI, CfoI, Ddel, HaeIII, MnlI, Mspl. DNA sequencing was used to identify the basis of any differences.

Mutation analysis of E14 was performed by RT-PCR and REF. The complete E14 coding sequence was amplified in four overlapping fragments using the primer pairs: 8409: 5′-CTTATTGTTGCTTCTGTTGTTTGTATCATG-3′ and 8784: 5′-GGTTGTATTCTGTGGTTGTTAGTAC-3′; 9644: 5′-GTACCTAAGCAACAAGATAAACAACC-3′ and 8971: 5′-GCTAATGGAAGACCTCCCAGGA-3′; 8697: 5′-TGGCCTGAAATTCTGACCTATG-3′ and 8488: 5′-GTTCTAGAAATGGCTGCCCTTGAGAAATC-3′.

Each amplified fragment was digested with three enzymes from a set which included Alul, Apol, Ddel, HinfI, MnlI and RsaI. After 5′ end labelling the digested fragments were run on SSCP gels as described in the original method.

ACKNOWLEDGEMENTS
We thank the Cancer Research Campaign, the Wellcome Trust, the Medical Research Council, the UK A-T Society, and the A-T Research and Support Trust for their continued support. We acknowledge the Medical Research Council, the UK A-T Society, and the A-T Research and Support Trust for their continued support. We thank Andrew Bell for help with Luciferase assays and Julian Parkhill for help with computer analysis of DNA sequences.

REFERENCES