

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
20 February 2003 (20.02.2003)

PCT

(10) International Publication Number
WO 03/014390 A2

(51) International Patent Classification⁷: **C12Q 1/68**

(21) International Application Number: PCT/GB02/03591

(22) International Filing Date: 2 August 2002 (02.08.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
01189950 3 August 2001 (03.08.2001) GB

(71) Applicant (*for all designated States except US*): **UNIVERSITY OF WALES COLLEGE OF MEDICINE**
[GB/GB]; Heath Park, Cardiff CF14 4XN (GB).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **SAMPSON, Julian, Roy** [GB/GB]; Institute of Medical Genetics, University of Wales College of Medicine, Heath Park, Cardiff CF14 4XN (GB). **CHEADLE, Jeremy, Peter** [GB/GB]; Institute of Medical Genetics, University of Wales College of Medicine, Heath Park, Cardiff CF14 4XN (GB).

(74) Agents: **NEWELL, William, Joseph** et al.; Wynne-Jones Laine & James, Morgan Arcade Chambers, 33 St. Mary Street, Cardiff CF10 1AF (GB).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SCREENING METHODS AND SEQUENCES RELATING THERETO

(57) Abstract: A screening method for identifying an individual having a pre-disposition towards having a cancer is disclosed, which screening method comprises the steps of : (a) obtaining a test sample comprising a nucleotide sequence comprised in a gene in a base excision repair (BER) pathway of the individual or an amino acid sequence of a polypeptide expressed thereby; and (b) comparing a region of the test sample sequence with the corresponding region of the wild type sequence, whereby a difference between the test sample sequence and the wild type sequence signifies that the individual is pre-disposed to having the cancer; and wherein the difference comprises a specified variation. The specified variation can be the known mutation in the human MYH protein, G382D-hMYH or a nucleotide sequence encoding it, or it can be one or more novel variations, namely, Y165C, E466X, and Y90X, or the respective corresponding nucleotide sequences. The method is particularly suitable for determining a pre-disposition towards bowel cancer.



WO 03/014390 A2

Screening Methods and Sequences relating thereto

The present invention relates to inherited variations in genes believed to be involved in base excision repair (BER) pathways of humans; to their use in screening patients for defects in BER and thereby for cancers or predisposition to cancers. The invention further relates to screening methods using the variations, and to a diagnostic kit for use in the screening methods.

BER pathways play a major rôle in the repair of mutations caused by reactive oxygen species that are generated during aerobic metabolism, as described in Nature 362, 709-715(1993). Oxidative DNA damage has been implicated in the aetiology of degenerative diseases, ageing and cancer (Mutat. Res. 250, 3116 (1991), but evidence linking inherited deficiencies of BER to these diseases has been lacking.

8-Oxo-7,8-dihydrodeoxyguanine (8-oxoG), the most stable product of oxidative DNA damage, is highly mutagenic, since it readily mispairs with A residues (Nature 349, 431-434 (1991)), leading to an increased frequency of spontaneous G:C→T:A transversion mutations in repair-deficient bacteria and yeast cells. In *E. coli*, three enzymes, *mutM*, *mutY* and *mutT*, function synergistically to protect cells from the deleterious effects of guanine oxidation (J Bacteriol. 174, 6321-6325(1992)). The *mutM* DNA glycosylase removes the oxidised base from 8-oxoG:C base pairs in duplex DNA; the *mutY* DNA glycosylase excises A misincorporated opposite unrepaired 8-oxoG during replication; and *mutT* is an 8-oxo-dGTPase preventing incorporation of 8-oxo-dGMP into nascent DNA. Human *mutM*, *mutY* and *mutT* homologues have been identified and termed *hOGG1* (Proc. Natl. Acad.

Sci. (USA) 94, 8016-8020 (1997)), *hMYH* (J. Bactiol. 178, 3885-3892 (1996)) and *hMTH* (J. Biol. Chem. 268, 23524-23530 (1993)), respectively. Patent specification no. WO 97/33903 also discloses a human MutY polypeptide and DNA
5 encoding it, together with its potential use in diagnosing a cancer or a susceptibility to a cancer.

Until now, inherited, as distinguished from somatic, defects of BER have not been associated with any human genetic disorder, although mutations of the *Escherichia coli* BER genes *mutM* and *mutY* lead to increased G:C→T:A transversions (Proc. Natl. Acad. Sci. (USA) 85, 2709-2713 (1988); J. Bacteriol. 174, 6321-6325 (1992); Mol. Gen. Genet. 239, 72-76 (1993); and Mol. Gen. Genet. 254, 171-178 (1997)).

15 We now provide evidence that inherited defects of human BER genes might also lead to increased spontaneous (somatic, as opposed to hereditary) G:C→T:A transversions in other genes, which control cellular growth and so predispose the individual to cancer. Such other genes
20 include the APC gene, a known tumour suppressor gene for colorectal adenomas and carcinomas. Somatic mutations of APC have been found to occur in most such cancers. Accordingly, we further provide evidence that inherited defects in human BER genes can indicate a predisposition
25 towards cancers in humans.

We studied a family (hereinafter, 'family N') having multiple colorectal adenomas and carcinoma, and excluded an inherited mutation of APC as is seen in familial adenomatous polyposis (FAP) (Hum Mol Genet 10 721-733
30 (2001)). Eleven tumours from three affected siblings contained eighteen somatic APC mutations. Fifteen were G:C→T:A transversions, a significantly greater proportion than reported in sporadic or FAP associated tumours. Analysis of *hMYH* revealed that the siblings were compound

heterozygotes for the non-conservative missense variants, Y165C (an A to G substitution at nucleotide 494 in exon 7) and G382D (a G to A substitution at nucleotide 1145 in exon 13). G382D has been identified (by Shinmura K et al in Cancer Letters 166 65-69 (2001)) in lung cancer samples from the tumour itself; however, it was not identified as an inherited change that might be useful in the detection of predisposition towards lung (or any other) cancer.

These two changes affect residues that are conserved in *E. coli mutY*, namely, y82 and g253; y82 lies within the pseudo HhH motif and has been predicted to function in mismatch specificity (Nat Struct Biol 5 1058-1064 (1998)). Assays of adenine glycosylase activity of y82c and g253d mutants with a G:A substrate showed 90- and 6-fold reduction compared to wild type.

These data link the inherited variants in *hMYH* to the pattern of somatic *APC* mutation in family N and implicate defective BER in tumour predisposition in humans and other animals. Subsequent studies have revealed additional inherited variants in *hMYH*. Therefore, we have been able to provide the identity of such inherited variants and a screening method for identifying an individual having a pre-disposition towards having a cancer as a result of inherited variants in the genes involved in BER.

The present invention therefore provides a variant of *hMYH*, suitable for use in a screening method of the invention, comprising a polypeptide variant selected from:

- (i) Y165C, as defined herein [SEQ ID NO:1];
- (ii) E466X, as defined herein [SEQ ID NO:3]; and
- (iii) Y90X, as defined herein [SEQ ID NO:4].

The present invention further provides a nucleotide sequence corresponding to any one of the above polypeptide variations, being a nucleotide sequence encoding the polypeptide variant of this invention.

5 Accordingly, the present invention further provides a nucleotide sequence

(a) encoding a polypeptide variant according to the invention;

(b) a sequence substantially homologous to or that
10 hybridises to sequence (a) under stringent conditions; or

(c) a sequence substantially homologous to or that hybridizes under stringent conditions to the sequence (a) or (b) but for the degeneracy of the genetic code; or

(d) an oligonucleotide specific for any of the sequences
15 (a), (b) or (c).

Such homologous sequences as are referred to in (b) and (c) also display the functional and biological activity of the variation according to the invention. Preferably, such homologous sequences are at least 90%
20 identical to the sequence (a).

Preferably, there is provided a variant of hMYH, comprising Y165C, or a nucleotide sequence encoding the polypeptide, as defined herein [SEQ ID NO:1].

Such polypeptide and corresponding nucleotide
25 variants are herein collectively referred to as 'variations according to this invention'.

Accordingly, the present invention provides a method for diagnosing susceptibility to cancer comprising determining, from a sample derived from a patient, a
30 mutation comprising a variation according to this invention. In particular, there is provided a screening method for identifying an individual having a pre-disposition towards having a cancer, which screening method comprises the steps of:

(a) obtaining a test sample comprising a nucleotide sequence comprised in a gene in a base excision repair (BER) pathway of the individual or an amino acid sequence of a polypeptide expressed thereby; and

- 5 (b) comparing a region of the test sample sequence with the corresponding region of the wild type sequence, whereby a difference between the test sample sequence and the wild type sequence signifies that the individual is pre-disposed to having the cancer; and
- 10 wherein the difference comprises a variation according to this invention, or the known G382D-hMYH or the nucleotide sequence encoding it.

For example, where the variant polypeptide sequence is Y165C-hMYH, this means the human MYH protein in which
15 the 165th amino acid (Y in the wild type) is replaced by C. The corresponding difference in the nucleotide sequence means the gene variant, which is the nucleotide sequence coding for the variant polypeptide (protein) sequence. In the case of Y165C-hMYH, the corresponding
20 nucleotide variant is 494A→G-hMYH, which means the human MYH gene in which the 494th base (A in exon 7 in the wild type) is replaced by G, namely [SEQ ID NO: 1]. However, the variant may also comprise the nucleotide sequence encoding the known G382D-hMYH; or the variant may
25 comprise that encoding E466X-hMYH (1396G→T) or Y90X-hMYH (270C→A).

Preferably, the BER pathway gene is *hMYH*.

Preferably, in the screening method of the invention, the BER pathway gene is involved in the repair
30 of another gene and protects against G:C→T:A transversion mutations in that gene. More preferably, these G:C→T:A transversion mutations occur at regions of the sequence

wherein the G is followed by AA in the sequence and therefore comprise GAA→TAA mutations.

The invention therefore further provides an alternative screening method for identifying an

5 individual having an inherited defect in a BER gene and/or *hMYH* and/or is pre-disposed to having a cancer, which screening method comprises the steps of:

(a) obtaining a test sample comprising a nucleotide sequence comprised in a marker gene of the individual or
10 an amino acid sequence of a polypeptide expressed thereby, the marker gene being a marker for a disease or condition correlated with the presence of a defect in a BER gene and/or *hMYH* or cancer; and

(b) comparing a region of the test sample sequence
15 with the following diagnostic sequence [SEQ ID NO: 2] or a peptide sequence encoded thereby:

5' -X-X1-A-A-X2-A-A- 3' [SEQ ID NO: 2]

20 wherein X is A or T; X1 is G or T; and X2 is G or A.

whereby identity between the region of the test sample sequence and the diagnostic sequence signifies that the individual has an inherited defect in a BER gene and/or
25 *hMYH* and/or is pre-disposed to having the cancer.

Another aspect of the invention is a diagnostic sequence suitable for use in the alternative screening method, which sequence comprises:

30 (a) 5' -X-X1-A-A-X2-A-A- 3' [SEQ ID NO: 2]

wherein X is A or T; X1 is G or T; and X2 is G or A;

(b) a sequence substantially homologous to or that hybridises to sequence (a) under stringent conditions; or

(c) a sequence substantially homologous to or that hybridizes under stringent conditions to the sequence (a) or (b) but for the degeneracy of the genetic code; or
(d) an oligonucleotide specific for any of the sequences
5 (a), (b) or (c).

Accordingly, this invention further provides for the use of:

(a) a diagnostic sequence according to the invention, or a polypeptide encoded thereby,

10 (b) a sequence substantially homologous to or that hybridises to nucleotide sequence (a) under stringent conditions; or

(c) a sequence substantially homologous to or that hybridizes under stringent conditions to the nucleotide
15 sequence (a) or (b) but for the degeneracy of the genetic code; or

(d) an oligonucleotide specific for any of the sequences (a), (b) or (c)

in a therapeutic, diagnostic or detection method,
20 especially for the determination of susceptibility to a disease, such as cancer.

Especially preferred is when the damaged gene is a known marker for the cancer, such as APC (a marker for bowel cancer).

25 Therefore, in the screening method according to the invention, the cancer is most preferably bowel cancer and/or the damaged gene is APC.

In accordance with the above findings, the present invention therefore still further provides for use of:

30 (a) a BER gene variant, or a polypeptide encoded thereby, selected from variants of this invention and the known G382D-hMYH or the nucleotide sequence encoding it;

(b) a sequence substantially homologous to or that hybridises to nucleotide sequence (a) under stringent conditions; or

5 (c) a sequence substantially homologous to or that hybridizes under stringent conditions to the nucleotide sequence (a) or (b) but for the degeneracy of the genetic code; or

(d) an oligonucleotide specific for any of the sequences (a), (b) or (c)

10 in a therapeutic, diagnostic or detection method.

Preferably, the BER gene is *hMYH* and therefore a preferred use is of:

(a) a variant of the *hMYH* gene, or the *hMYH* polypeptide encoded thereby, selected from variants of this invention
15 and the known G382D-*hMYH* or the nucleotide sequence encoding it;

(b) a sequence substantially homologous to or that hybridises to nucleotide sequence (a) under stringent conditions; or

20 (c) a sequence substantially homologous to or that hybridizes under stringent conditions to the nucleotide sequence (a) or (b) but for the degeneracy of the genetic code; or

(d) an oligonucleotide specific for any of the sequences
25 (a), (b) or (c)
in a therapeutic, diagnostic or detection method.

A particularly preferred use is when the BER gene and/or *hMYH* is for the determination of susceptibility to a disease, such as cancer. Especially preferred is
30 wherein the corresponding wild type BER pathway gene or wild type *hMYH* gene acts to protect against G:C→T:A transverse mutations in a cancer marker gene, such as *APC*.

The present invention further provides a kit suitable for use in carrying out the screening method of the invention, which kit comprises one or more of:

- 5 (a) an oligo- or poly-nucleotide having a nucleic acid sequence corresponding to a region of a variant BER gene, which region incorporates at least one variation from the corresponding wild-type sequence selected from a variation according to this invention;
- 10 (b) an oligo- or poly-nucleotide having a nucleic acid sequence corresponding to the wild-type sequence in the region specified in (a); and/or
- (c) one or more reagents suitable for carrying out PCR for amplifying desired regions of the patient's DNA.

15 A kit suitable for use in carrying out an alternative screening method as described above comprises one or more of:

- (a) an oligo- or poly-nucleotide comprising a diagnostic sequence as described herein, or an amino acid sequence encoded thereby;
- 20 (b) one or more reagents suitable for carrying out PCR for amplifying desired regions of the patient's DNA; and/or
- (c) a 'surrogate marker' that is indicative of or correlated to the presence of a variant of a sequence
- 25 (a).

Reagents for the kit may include, for example, PCR primers corresponding to the exon of the BER gene, *hMYH* or a diagnostic sequence according to the invention; and/or primers mentioned herein; and/or other reagents
30 for use in PCR, such as *Taq* DNA polymerase.

Preferably, any oligonucleotides in the kit comprise in the range of from 5 to 25 base-pairs, such as 10-20 base-pairs for the variant sequences. In any case, the oligonucleotides must be selected so as to be unique for

the region selected and not repeated elsewhere in the genome.

Since, in the situation where it is desired to screen for multiple variations, such as in the range of
5 from 15 to 20 or more, a kit comprising up to 40 oligo- or poly-nucleotides or more would be required, in the alternative screening method, therefore, using DNA chip technology, the present invention provides a plurality of oligo- or poly-nucleotides as defined in kit component
10 (a) above immobilised on a solid support.

Other nucleotide detection methods could be used, such as signal amplification methods being pioneered in nanotechnology (such as Q-Dots). Also, single molecule detection methods could be employed (such as STM). In
15 which case, the kit according to this invention may comprise one or more reagents for use in such alternative methods.

Further, suitable, alternative screening methods according to this invention may further comprise
20 obtaining a test sample comprising a BER variant (eg a protein/peptide sequence comprising a variation of hMYH, such as one encoded by a variant of *hMYH* as identified above) or a peptide sequence encoded by a diagnostic sequence as defined herein that is identifiable by
25 conventional protein sequence methods (including mass spectroscopy, micro-array analysis, pyrosequencing, etc), and/or antibody-based methods of detection (eg ELISA), and carrying out one or more such protein sequencing method(s).

30 Alternatively, the screening method and corresponding kit according to this invention may be based on one or more so-called 'surrogate markers' that are indicative of or correlated to the presence of a variant of a BER gene, *hMYH* or a diagnostic sequence as

defined herein, or the polypeptide encoded thereby, such as proteins/amino acid sequences eg antibodies specific for a BER gene or protein.

Such a "surrogate marker" may therefore comprise:

- 5 (a) any biomolecule (including, but not limited to, nucleotides, proteins eg antibodies, sugars, and lipids);
(b) a chemical compound (including, but not limited to, drugs, metabolites thereof, and other chemical compounds); and/or
10 (c) a physical characteristic,
whose absence, presence, or quantity in an individual is measurable and correlated with the presence of a BER gene variant, *hMYH* or a diagnostic sequence as defined herein, or a protein encoded thereby.

15 In which alternative cases, the kit according to this invention may comprise one or more reagents for use in such alternative methods.

It will be evident to the person skilled in the art that, throughout the specification unless the context
20 indicates otherwise, the *hMYH* gene is classified as an example of a BER pathway gene, but even if *hMYH* was found to act in an alternative pathway, then the screening methods, kits and uses as described herein in relation to a 'BER gene' would nevertheless apply *mutatis mutandis* to
25 '*hMYH*'.

The basis for the invention will now be described in more detail with reference to the following Examples and Figures, in which:

30 **Figure 1** shows the pedigree of family N. II.1 and II.2 were found to have approximately 50 macroscopically visible adenomas at 59 and 55 years of age. II.3 died following discovery of a colonic adenocarcinoma and an adjacent adenoma at 46 years of age, but without full

assessment of the large bowel. II.4-7 were normal on colonoscopic assessment at between 36 and 49 years of age and III.1-III.5 were normal on colonoscopic assessment at between 24 and 33 years of age. APC haplotypes with the intragenic markers E1317Q, S2497L and the closely linked DP1 (CA)_n repeat are shown.

Figure 2 shows the identification of somatic G:C→T:A mutations of APC in colorectal tumours. Sequences of LD-PCR product clones were aligned. Variants in two or more clones from the same allele (I or II) were confirmed by an independent assay on a fresh PCR product. (a) G>T mutation at position 2602 (E868X) in adenoma A1 confirmed by direct sequencing of standard PCR products. (b) G>T mutation at position 4351 (E1451X) on the second APC allele from adenoma A1 confirmed by direct sequencing of LD-PCR products. (c) G>T mutation at position 3850 (E1284X) in adenoma B5 confirmed by restriction enzyme analysis. Arrows indicate the position of the G:C→T:A mutations on the sequencing gels and the mutant allele on *Bfr* I cleavage of a PCR product amplified from adenoma B5. M DNA size marker (ϕx174 *Hae* III), -ve wild type control DNA, +ve B5 adenoma DNA.

Figure 3 shows the identification and segregation of germline *hMYH* variants in family N. Direct sequencing of constitutional DNA from sibling II:1 revealed (a) an A to G substitution at nucleotide 494 in exon 7 corresponding to Y165C (arrow) and (b) a G to A substitution at nucleotide 1145 in exon 13 corresponding to G382D (arrow). (c) Screening for Y165C by ARMS and G382D by a *Bgl* II digest revealed that the three affected siblings (filled symbols) were compound heterozygotes for these *hMYH* missense variants, while normal family members (non-

filled symbols) were either heterozygous for one of the variants, or normal. N = normal ARMS reaction, M = mutant ARMS reaction. Arrows indicate the positions of the mutant alleles.

5

Figure 4 shows the evolutionary conservation of the variant residues in *hMYH*. Comparison of the variant residues (a) Y165C and (b) G382D in family N with *hMYH* homologues from *Homo sapiens* (*H.sap.*), *Mus musculus* (*M.mus.*), *Arabidopsis thaliana* (*A.tha.*), *Schizosaccharomyces pombe* (*S.pom.*), *Hemophilus influenzae* (*H.inf.*), *Vibrio cholerae* (*V.cho.*), *Salmonella typhimurium* (*S.typ.*) and *E.coli* using ClustalW. Arrows indicate the position of the variant residues. Identical, conserved and semi-conserved residues are shaded black, dark grey and light grey respectively. // indicates the position of 18 amino acids in *A.tha.* that are not present in the other organisms.

Figure 5 shows representative plots of single turnover adenine glycosylase assays. Wild type, Y82C, and G253D *mutY* were assayed for glycosylase activity at 2°C with a 20 nM duplex DNA substrate containing an 8-oxo G:A mismatch and 30 nM *mutY* (active site concentration). k_2 (min^{-1}) = 1.6 ± 0.2 for wild type, 0.26 ± 0.05 for G253D and <0.0006 (estimated) for Y82C. All values represent an average of at least four separate determinations and the error is reported as the standard deviation.

30

EXAMPLE 1: GENERAL METHODS AND PROTOCOLS RELATING TO APCSamples

Family N: Nucleic Acid was prepared from venous blood
5 samples and from normal colonic mucosa obtained at
surgery, using standard methods. Nine colorectal adenomas
were obtained at colonoscopy or surgery and one adenoma
and a carcinoma were obtained from archived tissues in
paraffin blocks. Histopathology was confirmed by
10 independent review. DNA and RNA were prepared from whole
tissues that had been snap frozen or from micro-dissected
tissue whose nature had been verified histologically on
the same or adjacent sections.

Patients with multiple colorectal adenomas: DNA was
15 extracted from venous blood samples from sixteen further
unrelated patients with multiple adenomas, with or
without co-existing carcinoma. All cases were shown to be
normal on sequencing of exon 4 and the alternatively
spliced region of exon 9 of APC, mutations in which are
20 associated with AFAP (Hum Mol Genet 10 721-733(2001)).
Archived tumour tissue was micro-dissected and DNA
extracted using standard methods.

Patients with colorectal cancer: DNA was extracted from
venous blood samples using standard methods.

25

Standard and long distance (LD-) PCR and microsatellite
analysis

Exons 1-3 and 5-14 of APC were amplified using published
primers (Cell 66,589-600 (1991)), and exon 4 was
30 amplified using ex4F (5'-TGCAGTCTTTATTAGCATTGTTT-3') and
ex4R (5'-TTCAAATAAGTTGTACTGCCAAG-3') which generated a
195bp product. For DNA extracted from paraffin embedded
blocks, exon 15 of APC was amplified as 40 overlapping
fragments of 162-285bp -----

(http://www.uwcm.ac.uk/study/medicine/medical_genetics/research/tmg/projects/hMYH.html). Exons 2-15 of *β-catenin* (Accession Nos.X89579, 13648651), 2-11 of *p53* (Accession No.U94788), 1-16 of *hMYH*, 1-8 of *hOGG1* (Accession Nos.AC066599, AC011610) and 2-5 of *hMTH* (Accession Nos.D38591-4) were amplified as 18, 11, 16, 11 and 4 fragments, respectively. Standard PCR was carried out in 50μl reaction volumes containing 100ng genomic DNA, 25pmole primers, 0.2mM dNTPs, 5μl reaction buffer and 1U AmpliTaq Gold DNA Polymerase (Applied Biosystems). Cycling parameters were 94°C 10 mins, followed by 32 cycles of 50-67°C 1 min, 72°C 1 min, 94°C 30 secs, and a final step of 72°C 10 mins. For DNA extracted from fresh tissue, exon 15 of *APC* was amplified either as a single 6.67kb LD-PCR fragment using N15F (1997) 5'-GCAAATCCTAAGAGAGAACAACCTGT-3' and N15R1 (8684) 5'-TCCAGAACAAAAACCTCTAACAAG-3', or as two overlapping LD-PCR fragments using N15F (1997) and NS15 4R (5571) 5'-CCTTCAATAGGCGTGTAATG-3' which generated a 3.59 kb product, and NS15 3F (3649) 5'-AAAGCAGTAAAACCGAACAT-3' and N15R (8698) 5'-TCAAATATGGCTTCCAGAACAAA-3' which generated a 5.07kb product. Exons 10 to 16 of *hMYH* were amplified as a 3.1kb LD-PCR fragment using Y10F1L (5'-GCTGATCCCAGCAGCACCCCTTGTTT-3') and Y16RL (5'-AATGGGGGCTTTCAGAGGTGTCACT-3'). The 50μl LD-PCR reaction mixes contained 100ng genomic DNA, 10pmol primers, 0.35mM dNTPs, 5μl reaction buffer 3 and 1.75U of Expand Long Template DNA Polymerase (Roche). Cycling parameters were 94°C 2mins followed by 33-35 cycles of 56°C 1min, 68°C 4-8mins (+20secs per cycle, after cycle number 10) and 94°C 20secs, and a final elongation step at 68°C for 4-8mins. DNA extracted from normal and tumour tissue was tested

for MSI using the markers D2S123, BAT 26, BAT 24, Mfd15, DP1(APC), D18S69 and BAT 25.

RT-PCR and expression analysis

5 100ng-2µg RNA was used for first strand cDNA synthesis using oligo (dT)₁₅ and Superscript II RNase H⁻ Transcriptase (Invitrogen Life Technologies). Second strand synthesis was carried out in 50µl reaction volumes using 1µl cDNA, 25pmol primers, 0.2mM dNTPs, 5µl reaction
10 buffer and 2-5U AmpliTaq Gold DNA Polymerase. Cycling parameters were 94°C for 10 mins, followed by 35-40 cycles of 94°C 1min, 50-54°C 1min, 72°C 1-3mins, and a final elongation step at 72°C for 10mins. Exons 1-14 of APC were amplified as a 1.958kb fragment, using the primers
15 previously described (Proc. Natl. Acad. Sci.(USA) 94, 2449-2453 (1997)). To determine the expression levels of individual APC alleles, the exon 11 polymorphism Y486 was assayed in recombinant RT-PCR product clones by restriction digestion or sequence analysis. To
20 characterise aberrant splicing associated with the 423-1G>T somatic mutation, exons 3-12 of APC were amplified by RT-PCR using APCFEx3 (5'-GAGGGTTTGTAATGGAAGCAG-3') and APCjREx11-12 (5'-CTCATGCAGCCTTTCATAGAGC-3'), cloned and sequenced. To quantify the expression level of the
25 hMYH allele harbouring G382D, normal colonic mucosa cDNA from sibling II:1 was amplified using rY12F (5'-GTGGTCAACTTCCCCAGAAA-3') and rY14R (5'-GGCCAGCCCATATACTTGAT-3'), cloned and assayed with a Bgl II digest.

30

Sequencing

Standard PCR products were sequenced manually using the ThermoSequenase cycle sequencing kit (Amersham), and

analysed on 6% polyacrylamide gels. For automated plasmid based sequencing, standard, LD- and RT-PCR products were purified using the PCR purification kit (Qiagen), cloned into pGEM-T Easy (Promega), and propagated in JM109 *E.coli*; at least twelve recombinant clones of each product were sequenced. Automated sequencing of RT-PCR product clones spanning exons 1-14 of APC was performed using two overlapping bi-directional sequencing reactions: (1) NS1_14F (39) 5'-ATGGCTGCAGCTTCATATGA-3' to NS1_14R2 (1049) 5'-GCTGTCTTGGGAGCTAGAC-3'; (2) NS1_14F2 (892) 5'-ACCATGAAACAGCCAGTGT-3' to NS1_14R (1978) 5'-CTGTGGTCCTCATTTGTAG-3'. Automated sequencing of LD-PCR products and clones spanning exon 15 of APC was performed using eight overlapping bi-directional sequencing reactions: (1) NS15 1F (1997) 5'-GCAAATCCTAAGAGAGAACA-3' to NS15 8R (3146) 5'-GACTTTGCCTTCCAGAGTTC-3'; (2) NS15 2F (2810) 5'-AAGCTCTGCTGCCCATACACA-3' to NS15 7R (3935) 5'-CTGCTATTTGCAGGGTATTA-3'; (3) NS15 3F (3649) 5'-AAAGCAGTAAAACCGAACAT-3' to NS15 3R (4775) 5'-TTGTTGGCATGGCAGAAATA-3'; (4) NS15 4F (4480) 5'-TTCTTCCAGATGCTGATACT-3' to NS15 4R (5571) 5'-CCTTCAATAGGCGTGTAATG-3'; (5) NS15 5F (5234) 5'-GCCCAAAGGGAAAGTCACA-3' to NS15 5R (6346) 5'-ATTTGCACCTTCCTGAATAG-3'; (6) NS15 6F (6015) 5'-CCTGACTCACAGGGAGAAC-3' to NS15 6R (7135) 5'-CTGTCTACCTGGAGATGTAT-3'; (7) NS15 7F (6807) 5'-GCCTCCAAAAGCCCTAGTGA-3' to NS15 2R (7920) 5'-AGCACCTGAGGAAACGGTCTG-3'; (8) NS15 8F (7552) 5'-GAAAACTCCCACCTAATCTC-3' to NS15 1R (8684) 5'-AACAAAAACCCTCTAACAAG-3'.

Primer nucleotide numbers in parentheses are cited according to Science 253, 661-5 (1991). Automated sequencing of LD-PCR product clones spanning exons 10-16 of *hMYH*, RT-PCR product clones spanning exons 3-12 of APC

and 12-14 of *hMYH*, and standard PCR product clones, was performed using M13 forward and reverse primers, as previously described (Hum. Mol. Genet 9, 1119-1129 (2000)). Sequence data for ≥ 12 clones was aligned (AlignIR v1.2, Li-Cor) and variants in two or more clones from the same allele were analysed by an independent assay on a fresh PCR product, to confirm that they represented real mutations and were not PCR or cloning induced errors.

Assays for sequence variants

In APC: Y486 (1458 C>T) in exon 11 was assayed using an *Rsa* I digest, as previously described (Cell 66, 589-600 (1991)). E1317Q (3949 G>C) in exon 15 was assayed using a *Pvu* II digest of a 503bp PCR product generated using E1317QLF (3652) 5'- GCAGTAAAACCGAACATATG-3' and E1317QR (4137) 5'- TGGACTTTTGGGTGTCTG-3'; DNA from paraffin embedded blocks was assayed using a 224bp PCR product generated with E1317QSF (3934) 5'- CTAATACCCTGCAAATAGCA-3' and E1317QR (4137). A545 (1635 A>G) in exon 13 and T1493 (4479 G>A), A1755 (5265 G>A), S1756 (5268 G>T), and S2497L (7491 C>T) in exon 15, were assayed by sequencing.

The somatic APC mutations E1284X (3850 G>T) and E1317X (3949 G>T) in exon 15 were assayed using a *Bfr* I digest of PCR products generated with E1317QLF and E1317QR. Somatic APC mutations were assigned to an allele by linking them to one of the polymorphic markers using either standard, RT-, or LD-PCR, followed by cloning and sequencing. The following missense variants in *hMYH* were assayed in 100 normal control patients: V22M (66 G>A) was assayed using an *Nco* I digest of exon 2 PCR products. Y165C (494 A>G) in exon 7 was assayed using normal (165N 5'-CGCCGGCCACGAGAATGGT-3') - or mutant (165M 5'-

CGCCGGCCACGAGAATTGC-3') specific ARMS primers together with the common primer 165C (5'-AGTGCTTCCCTGGAGGTGAGA-3'). R260Q (779 G>A) in exon 10 was assayed using normal (260N 5'-CTTGGTTGAAATCTCCTGCCC-3') or mutant (260M 5'-CTTGGTTGAAATCTCCTGACT-3') specific ARMS primers together with the common primer 260C (5'-CGAGCCATTGGTGCTGATC-3'). G382D (1145 G>A) was assayed using a *Bgl* II digest of exon 13 PCR products. S501F (1502 C>T) in exon 16 was assayed using normal (501N 5'-GCTTTTTCCGACTGCACGCAG-3') or mutant (501M 5'-GCTTTTTCCGACTGCACGAAA-3') specific ARMS primers together with the common primer 501C (5'-GCATTCCAGGCTAAGCCTAGC-3'). All ARMS reactions incorporated internal control primers (AJ31 and AJ32) to validate the assay. SSCP and dHPLC analysis at the RT_m and RT_m+2°C, was carried out as described by Jones *et al.* (Hum. Genet. 106, 663-668 (2000)).

Somatic APC mutation database and statistical analysis

We reviewed literature reports of characterised somatic APC mutations in colorectal tumours. This included publications cited in the APC mutation database (Nucleic Acids Res. 24, 121-124 (1996)) and publications from the period 1991-2001 identified through a PubMed (<http://www.ncbi.nlm.nih.gov>) search. Reports of truncating mutations that were inconsistent with the published cDNA sequence (Science 253, 661-5 (1991)) were excluded, as were putative missense mutations since the evidence for their pathogenicity was inconclusive. Data on 503 somatic mutations observed in sporadic tumours and 308 somatic mutations observed in FAP and AFAP associated tumours was retrieved. This included cases of allelic loss, nonsense, frameshift and splice site mutations at invariant GT-AG dinucleotides. Statistical analysis was carried out using Fisher's Exact test.

Site directed mutagenesis (SDM) and assays of mutY glycosylase activity

SDM with the primers y82c_F (5'-
5 GCGCGCGCGGGCGCAATAGCCAAGCCC-3') and g253d_F (5'-
CCGCCCCACAAGTCGCTCGGCGGACGC-3'), cloning, expression and
purification of wild type and mutant mutY, was carried
out as previously described (Nucleic Acids Res. 29, 553-
564 (2001)).

10 To determine the effect of the y82c and g253d
mutations on the intrinsic rate of adenine removal
compared to wild type, glycosylase assays were performed
under single turnover conditions ($[DNA] < [MutY]$) as
described (Biochem. 37, 14756-14764 (1998)) using a 30bp
15 duplex containing a centrally-located 8-oxo-G:A or G:A
base pair. The amount of active protein (wild type 39%,
y82c 53%, g253d 58%) was determined using active site
titration methods (Biochem 37, 14756-14764 (1998)). The
resulting data were fitted to the single exponential
20 equation: $[P]_t = A_0[1 - \exp(-k_{obs}t)]$. Under the conditions used
for these experiments, k_{obs} approximates k_2 (Biochem 37,
14756-14764 (1998)).

25 **EXAMPLE 2: Primers, conditions and methods for
amplification and analysis of the human BER genes hMYH,
hOGG1 and hMTH**

(1) Standard PCR

30 Exons 1-16 of hMYH, 1-8 of hOGG1 (Accession
Nos.AC066599, AC011610) and 2-5 of hMTH (Accession
Nos.D38591-4) were amplified as 16, 11 and 4 fragments,
respectively (Tables 2-4). Standard PCR was carried out
in 50 μ l reaction volumes containing 100ng genomic DNA,

25pmole primers, 0.2mM dNTPs, 5µl reaction buffer and 1U AmpliTaq Gold DNA Polymerase (Applied Biosystems). Cycling parameters were 94°C 10 mins, followed by 32 cycles of 50-67°C 1 min, 72°C 1 min, 94°C 30 secs, and a final step of 72°C 10 mins.

Table 2: Primers used for the amplification of *hMYH*

Exon	Primer name	Sequence	Product size	Annealing Temp.
1	Y1F Y1R	5' -GAAGCTGCGGGAGCTGAAA-3' 5' -ATCCCCGACTGCCTGAACC-3'	133 bp	60 °C
2	Y2F Y2R	5' -CTGCATTTGGCTGGGTCTTT-3' 5' -CGCACCTGGCCCTTAGTAAG-3'	263 bp	54 °C
3	Y3F Y3R	5' -AGCCTGTGCAGGGATGATTG-3' 5' -CAACCCCAGATGAGGAGTTAGG-3'	272 bp	57 °C
4	Y4F Y4R	5' -CTCATCTGGGGTTGCATTGA-3' 5' -GGGTTGGCATGAGGACACTG-3'	167 bp	57 °C
5	Y5F Y5R	5' -GGGCAGGTCAGCAGTGTC-3' 5' -TACACCCACCCCAAAGTAGA-3'	189 bp	57 °C
6	Y6F Y6R	5' -TACTTTGGGGTGGGTGTAGA-3' 5' -AAGAGATCACCCGTCAGTCC-3'	185 bp	54 °C
7	Y7F Y7R	5' -GGGACTGACGGGTGATCTCT-3' 5' -TTGGAGTGCAAGACTCAAGATT-3'	186 bp	54 °C
8	Y8F Y8R	5' -CCAGGAGTCTTGGGTGTCTT-3' 5' -AGAGGGGGCCAAAGAGTTAGC-3'	240 bp	57 °C
9	Y9F Y9R	5' -AACTCTTTGGCCCTCTGTG-3' 5' -GAAGGGAACACTGCTGTGAAG-3'	196 bp	57 °C
10	Y10F Y10R	5' -GTGCTTCAGGGGTGTCTGC-3' 5' -TGTCATAGGGCAGAGTCACTCC-3'	262 bp	57 °C
11	Y11F Y11R	5' -TAAGGAGTGACTCTGCCCTATG-3' 5' -GCCAAGAGGGGCTTTAGG-3'	248 bp	54 °C
12	Y12F Y12R	5' -AGCCCCTCTTGGCTTGAGTA-3' 5' -TGCCGATTCCCTCCATTCT-3'	298 bp	57 °C
13	Y13F Y13R	5' -AGGGCAGTGGCATGAGTAAC-3' 5' -GGCTATTCCGCTGCTCACTT-3'	242 bp	57 °C
14	Y14F Y14R	5' -TTGGCTTTTGAGGCTATATCC-3' 5' -CATGTAGGAAACACAAGGAAGTA-3'	256 bp	54 °C
15	Y15F Y15R	5' -TGAAGTTAAGGGCAGAACACC-3' 5' -GTTCACCCAGACATTCGTTAGT-3'	205 bp	54 °C
16	Y16F Y16R	5' -AGGACAAGGAGAGGATTCTCTG-3' 5' -GGAATGGGGGCTTTCAGA-3'	224 bp	54 °C

Table 3: Primers used for the amplification of *hOGG1*

Ex on	Primer name	Sequence	Product size	Annealing Temp.
1	M1F M1R	5'-CTTTGGGCGTCGACGAG-3' 5'-GAGGGGACAGGCTTCTCAG-3'	237 bp	57 °C
2	M2F1 M2R1 M2F2 M2R2	5'-ATTGAGTGCCAGGGTTGTCA-3' 5'-CGGAACCCCAGTGGTGATAC-3' 5'-TGTA TAGCGGATCAAGTAT-3' 5'-TGGCAAACTGAGTCATAG-3'	245 bp 286 bp	57 °C 50 °C
3	M3F1 M3R1 M3F2 M3R2	5'-GTCTGGTGTGCTTTCTCTAAC-3' 5'-GTGATGCGGGCGATGTT-3' 5'-TCTGCAGGTGTGCGACTGC-3' 5'-AGGAAGCCTTGAGAAGGTAACC-3'	229 bp 275 bp	50 °C 57 °C
4	M4F M4R	5'-GGAAGAACTTGAAGATGCCT-3' 5'-GCTCATTTCTGCTCTCC-3'	296 bp	55 °C
5	M5F M5R	5'-CCGGCTTTGGGGCTATA-3' 5'-GTTTCTACCATCCCAGCCCA-3'	279 bp	57 °C
6	M6F M6R	5'-TACTTCTGTTGATGGGTCAC-3' 5'-TGGAGGAGAGGAAACCTAG-3'	153 bp	55 °C
7	M7F M7R	5'-ACCTCCCAACACTGTCCTA-3' 5'-CCCTCCCCAACATGAGA-3'	265 bp	55 °C
8	M8F1 M8R1 M8F2 M8R2	5'-CTGTGGCCACGCACTTGTG-3' 5'-ACGTCCTTGGTCCAGCAGTGGT-3' 5'-GAGAGGGGATTACAAGGTG-3' 5'-GCCATTAGCTCCAGGCTTAC-3'	253 bp 287 bp	57 °C 55 °C

5 **Table 4: Primers used for the amplification of *hMTH***

Ex on	Primer Name	Sequence	Product size	Annealing Temp.
2	T2F T2R	5'-GCAAGGACAGAGGGCTTTCTG-3' 5'-CCAGCAGGCCATCAACTGAT-3'	249 bp	67 °C
3	T3F T3R	5'-GCACGTCATGGCTGACTCT-3' 5'-CTGGGAAAGCCGGTTCTAT-3'	246 bp	57 °C
4	T4F T4R	5'-TCCCTGGGCTGTGTGTAGAT-3' 5'-GAGATGGGACCCGCATAGT-3'	298 bp	57 °C
5	T5F T5R	5'-TGAAGTTTGGGTGACCTC-3' 5'-AGATGGTTTGCGGCTGTTC-3'	281 bp	57 °C

(2) Long-distance (LD-) PCR

Exons 10 to 16 of *hMYH* were amplified as a 3.1kb LD-PCR fragment using Y10F1L (5'-GCTGATCCCAGCAGCACCCTTGTTT-3') and Y16RL (5'-AATGGGGGCTTTCAGAGGTGTCAC-3'). The 50µl LD-PCR reaction mixes contained 100ng genomic DNA, 10pmol primers, 0.35mM dNTPs, 5µl reaction buffer 3 and 1.75U of Expand Long Template DNA Polymerase (Roche). Cycling parameters were 94°C 2mins followed by 33-35 cycles of 56°C 1min, 68°C 4-8mins (+20secs per cycle, after cycle number 10) and 94°C 20secs, and a final elongation step at 68°C for 4-8mins.

(3) RT-PCR and expression analysis

100ng-2µg RNA was used for first strand cDNA synthesis using oligo (dT)₁₅ and Superscript II RNase H⁻ Transcriptase (Invitrogen Life Technologies). Second strand synthesis was carried out in 50µl reaction volumes using 1µl cDNA, 25pmol primers, 0.2mM dNTPs, 5µl reaction buffer and 2-5U AmpliTaq Gold DNA Polymerase. Cycling parameters were 94°C for 10 mins, followed by 35-40 cycles of 94°C 1min, 50-54°C 1min, 72°C 1-3mins, and a final elongation step at 72°C for 10mins. To quantify the expression level of the *hMYH* allele harbouring G382D, normal colonic mucosa cDNA from sibling II:1 was amplified using rY12F (5'-GTGGTCAACTTCCCCAGAAA-3') and rY14R (5'-GGCCAGCCCATATACTTGAT-3'), cloned and assayed with a *Bgl* II digest.

(4) Sequencing

Standard PCR products were sequenced manually using the ThermoSequenase cycle sequencing kit (Amersham), and analysed on 6% polyacrylamide gels. For automated plasmid based sequencing, standard, LD- and RT-PCR products were

purified using the PCR purification kit (Qiagen), cloned into pGEM-T Easy (Promega), and propagated in JM109 *E.coli*; at least twelve recombinant clones of each product were sequenced. Automated sequencing of LD-PCR product clones spanning exons 10-16 of *hMYH*, RT-PCR product clones spanning exons 12-14 of *hMYH*, and standard PCR product clones, was performed using M13 forward and reverse primers.

10 (5) Single strand conformation polymorphism (SSCP) and
denaturing high performance liquid chromatography (dHPLC)
analysis

SSCP and dHPLC analysis at the R_{Tm} and $R_{Tm}+2^{\circ}C$, was carried out as previously described (Hum. Genet. 106, 663-668 (2000)).

(6) Assays for sequence variants

The missense variants in *hMYH* were assayed in 100 normal control patients (Table 5). All ARMS reactions incorporated internal control primers (AJ31 and AJ32, Hum. Genet. 106, 663-668 (2000)) to validate the assay.

Table 5: Assays for missense variants in *hMYH*

Variant	Exon	Assay
V22M (66 G→A)	2	NcoI digest
Y165C (494 A→G)	7	Normal ARMS (165N 5'-CGCCGGCCACGAGAATGGT-3') Mutant ARMS (165M 5'-CGCCGGCCACGAGAATTGC-3') Common (165C 5'-AGTGCTTCCCTGGAGGTGAGA-3')
R260Q (779 G→A)	10	Normal ARMS (260N 5'-CTTGGTTGAAATCTCCTGCCC-3') Mutant ARMS (260M 5'-CTTGGTTGAAATCTCCTGACT-3') Common (260C 5'-CGAGCCATTGGTGCTGATC-3')

H324Q (972 C→G)	12	Normal ARMS (324N 5'- CCAGCTCCCAACACTGGAGAC-3') Mutant ARMS (324M 5'- CCAGCTCCCAACACTGGAGAG-3') Common (324C 5'- CCCAGGCTGTTCCAGAACAC-3')
G382D (1145 G→A)	13	<i>Bgl</i> III digest
S501F (1502 C→T)	16	Normal ARMS (501N 5'- GCTTTTTCCGACTGCACGCAG-3') Mutant ARMS (501M 5'- GCTTTTTCCGACTGCACGAAA-3') Common (501C 5'- GCATTCCAGGCTAAGCCTAGC-3')

Example 3: Investigation of and Data from Family N

5 As summarised above, and using the methods and materials as described in Examples 1 and 2, we investigated a family N in which three siblings (II:1-3) were affected by multiple colorectal adenomas and carcinoma (Fig. 1). This was a Caucasian family but the
10 method is not limited to any particular ethnic grouping. DNA and RNA extracted from normal colonic mucosa from sibling II:1 and DNA extracted from normal colon tissue embedded within a paraffin block from the deceased sibling II:3, was amplified and sequenced for the 8532bp
15 open reading frame (ORF) of the APC gene (identification and characterisation of the APC gene is described in Cell 66, 589-600 (1991)).

Five silent base substitutions (1458 C>T [Y486], 1635 A>G [A545], 4479 G>A [T1493], 5265 G>A [A1755] and
20 5268 G>T [S1756]), and two missense variants (E1317Q and S2497L) were identified, but no clearly pathogenic change was found. Assays of the variants showed that none were present in all three affected siblings and that their shared wild type APC haplotype was also present in five
25 other family members who were phenotypically normal on

colonoscopic assessment (Fig. 1). Sequencing of RT-PCR products of exons 1-14 of APC in sibling II:1 confirmed equal expression of both alleles with alternate splicing of exons 9a and 10a, as previously reported and described in Human Mol Genet 10, 735-740 (2001). These data effectively excluded inactivation of APC as the primary inherited defect in family N.

Inherited mutations of the mismatch repair (MMR) genes cause hereditary non-polyposis colorectal cancer (HNPCC) characterised by micro-satellite instability (MSI) in the associated tumours. Assessment for MSI in DNA extracted from each of the available tumours that included five adenomas from sibling II:1, four adenomas from sibling II:2, and one adenoma and one carcinoma from sibling II:3, revealed instability with only one (Mfd15) of the seven markers tested in a single adenoma. This observation, and the multiple adenoma phenotype, provided evidence against the presence of a MMR gene defect in family N.

Since biallelic inactivation of APC occurs in most colorectal adenomas and carcinomas (Hum Mol Genet 10 721-733 (2001)), we sequenced the APC ORF to identify somatic mutations in each of the 11 tumours obtained from family N. Eighteen mutations were characterised, of which 15 were G:C→T:A transversions, including 14 nonsense changes and one splice site mutation (Table 6 and Fig. 2). The three remaining mutations were two C:G→T:A transitions at CpG dinucleotides and a case of allelic loss (Table 6).

Table 6: Somatic *APC* mutations identified in family N

	Sample [†] Sequence	Nucleotide change	Amino acid change	No. of clones (x/y) ^{a^}	context
5					
	A1	2602 G>T <u>AG</u> AAAAT	E868X	2/6	
10		4351 G>T <u>AG</u> AAGTA	E1451X	2/6	
	A2	721 G>T <u>AG</u> AAGCA	E241X	NA	
15		4381 G>T <u>TG</u> AAAAG	E1461X	2/6	
	A3	4717 G>T <u>TG</u> AAATA	E1573X	4/5	
20		NI	NI		
	A4	423-1 G>T [#] 4351 G>T <u>AG</u> AAGTA	NA E1451X	2/2 6/6	NA
25	A5	601 G>T <u>GG</u> AAGAA	E201X	NA	
		4348 C>T	R1450X	3/6	NA
30	B2	3331 G>T <u>AG</u> AAACA	E1111X	7/10	
		LOH	LOH		NA
	B4	3586 C>A <u>TG</u> AAAAT	S1196X	3/7	
35		3856 G>T <u>TG</u> AAATA	E1286X	4/5	
	B5	604 G>T <u>AG</u> AACAA	E202X	3/6	
40		3850 G>T <u>TG</u> AAGAT	E1284X	6/6	
	B6	2863 G>T <u>AG</u> AATAC	E955X	5/7	
45		3949 G>T <u>TG</u> AAGAT	E1317X	4/6	
	C2b	1495 C>T NI	R499X NI	3/6	NA
50					

Cla	NI	NI
	NI	NI

5

¶ Five adenomas from sibling II:1 (A1-5), four adenomas from sibling II:2 (B2,4-6), and one adenoma (C2b) and one carcinoma (Cla) from sibling II:3 were analysed for somatic APC mutations. Mutations were described according to the established nomenclature system. Biallelic mutations were proven to be on opposite alleles in all tumours, except A2 and A5. 423-1 G>T[#] was shown to cause skipping of exon 4 and predicted to terminate the reading frame at the 7th codon of exon 5.

10

15 ¶ Number of clones, where x represents the number with the mutation and y represents the total number from that allele. In general, mutations were found in only a proportion of clones. Non-mutated clones from the same allele most likely represent contaminating normal tissue. All mutations were confirmed by an independent assay on a fresh PCR product.

20

^ Sequence context surrounding the coding region G:C→T:A mutations (underlined) (the sequence of the non-transcribed strand is shown except for S1196X in B4).

25

NA = not applicable; NI = not identified.

The carcinoma did not contain any identified APC mutations despite re-sequencing of the ORF in DNA from a second micro-dissected tumour sample. Sequence analysis of the coding regions of β -catenin and p53 in DNA from this carcinoma also failed to identify any somatic mutations, suggesting involvement of an alternative tumourigenic pathway. We compared the proportion of G:C→T:A transversion mutations detected in tumours from family N to a database of 503 reported somatic APC

30

35

mutations from sporadic colorectal adenomas and carcinomas and 308 somatic mutations from FAP associated tumours. The excess of G:C→T:A transversions in family N was highly significant (15/18 vs. 49/503, $P=2.77 \times 10^{-12}$ and 5 15/18 vs. 30/308, $P=7.69 \times 10^{-12}$ respectively).

To determine if an inherited defect in one of the human BER genes (*hOGG1*, *hMYH* or *hMTH*) was responsible for the pattern of somatic G:C→T:A mutations in family N, DNA extracted from peripheral blood lymphocytes from sibling 10 II:1 was amplified and sequenced for the coding regions of *hOGG1*, *hMYH* and *hMTH*. Two amino acid variants were identified in *hMYH*, Y165C (an A to G substitution at nucleotide 494 in exon 7) and G382D (a G to A substitution at nucleotide 1145 in exon 13) (Fig. 3). No 15 missense variants or other likely pathogenic changes were identified in *hOGG1* or *hMTH*.

Both *hMYH* variants were assayed in blood DNA samples from all members of family N and 100 Caucasian control individuals with no history of colorectal adenoma or 20 carcinoma. In family N, the three affected siblings were compound heterozygotes for Y165C and G382D and the unaffected family members were either heterozygous for one of these variants or normal (Fig. 3). Each of the missense variants was also identified once in different 25 normal controls. Since the G to A substitution causing G382D was located at the first base in exon 13, its potential affect on splicing and expression was examined. Only 31 of 100 clones obtained by RT-PCR of normal colonic mucosa total RNA from sibling II:1 harboured the 30 G382D allele, although no aberrant splicing could be detected.

Somatic mutations of *hMYH* were sought in each of the eleven tumours by dHPLC and SSCP analysis of all exons and by screening for allelic loss by assay of the exon 7

and 13 missense variants. No somatic mutations were identified to suggest that *hMYH* might function as a tumour suppressor in a manner analogous to the MMR genes in HNPCC (Hum.Mol.Genet.10, 735-740 (2001)). Neither was there clear evidence for the Y165C or G382D variants being dominant to wild type, since heterozygotes for each were phenotypically normal. Rather, the occurrence of the multiple adenoma phenotype in only the three compound heterozygotes suggested transmission as an autosomal recessive trait.

We then searched for germline mutations of *hMYH*, *hOGG1* and *hMTH* by sequence analysis of their ORFs in sixteen unrelated patients with between 3 and ~50 colorectal adenomas, with or without carcinoma. *hMYH* mutations were also sought by dHPLC analysis of all exons in forty-two unrelated patients with colorectal cancer diagnosed at 40 years of age or less or with a family history of at least one first degree relative also affected by colorectal cancer. Several frequent missense polymorphisms were identified, S326C in *hOGG1* (Oncogene 16, 3219-32225 (1998)), and V22M, H324Q (J. Bactiol. 178, 3885-3892 (1996)), and S501F in *hMYH*, but their allele frequencies were not significantly different in the patient groups compared to 100 unaffected controls. One case, MA12, with three adenomas and a carcinoma, was a compound heterozygote for the unique *hMYH* missense variant R260Q (779 G>A) and the S501F polymorphism. Analysis of the *APC* ORF in the four tumours revealed two G:C→T:A transversions producing nonsense changes (E477X and S1344X), two frame-shift mutations and one case of allelic loss. Other family members were not available for study and the limited number of tumours precluded establishing a meaningful pattern of *APC* mutation.

Comparison of *hMYH* homologues in bacteria, yeast, plant and mammals revealed identical or similar amino acids at the positions of the non-conservative missense changes identified in family N (Fig. 4). To gain insight into the functional consequences of the missense variants, we assessed the effects of the equivalent *E coli mutY* mutations, y82c and g253d, on the intrinsic rate of adenine removal from a centrally located 8-oxo-G:A or G:A mismatch in a 30bp duplex.

The mutant proteins exhibited significantly slower rates of adenine removal than wild type (Fig. 5). The mutant proteins exhibited approximately ninety-fold (Y82C) and six-fold (G253D) slower rates of adenine removal from the G:A substrate at 37°C (k_2 (min⁻¹) = 1.6 ± 0.04 for wild type, 0.04 ± 0.01 for Y82C and 0.22 ± 0.04 for G253D). The high affinity of MutY for 8-oxo-G:A substrates results in reaction rates that are too fast at 37°C to be measured using our manual methods, and therefore the reaction rates with this duplex were analysed at 4°C (Fig 5). The G253D enzyme exhibited a 5-fold decreased rate of adenine removal; while the Y82C enzyme was so severely compromised in its catalytic activity that minimal conversion of substrate to product was observed during the time period that was monitored.

The dramatic effect of the y82c mutation is consistent with the findings of structural studies of mutY (Nat Struct Biol 5 1058-1064 (1998)), which locate y82 within the pseudo-HhH motif (79-gxgyya-84) and suggest a role in mismatch specificity and flipping of adenine into the base specificity pocket. The reduction in activity associated with g253d was similar to that observed with a truncated form of mutY that lacked the C-terminal third of the protein (Nucleic Acids Res. 29, 553-564 (2001)). In the colonic mucosa, the activity of

the hMYH G382D allele may be further compromised by the reduced expression we noted on RT-PCR analysis.

The activity of mutY on mismatched DNA substrates is influenced by the immediate sequence context, and methylation interference experiments have shown that mutY interacts with purines including the G:A mismatched bases and two bases each side (J Biol Chem 270 23582-23588 (1995)). Examination of the sequence surrounding the 14 coding region G:C→T:A mutations in family N revealed that the two bases immediately 3' to the mutated G were always AA. Furthermore, 13/14 sites matched three or all bases in a sequence extending one base 5' (A/T) and three bases 3' (G/A,A,A) to the mutated GAA (Table 6).

Inherited factors are thought to play a major rôle in at least 15% of colorectal cancer cases, but established predisposition genes account only for a minority of these (Cell 87, 159-170 (1996)). The sub-polymorphic frequency of the hMYH variants identified in family N, and the lack of evidence for pathogenic variants in the other multiple adenoma and colorectal cancer cases studied here, suggest that inherited defects of hMYH will prove to be an uncommon cause of colorectal tumour predisposition. The multiple adenoma phenotype in members of family N may reflect the number of somatic mutations required for initiation of adenoma development. Patients with FAP develop hundreds or thousands of adenomas, each requiring only a single somatic APC mutation. The compound heterozygotes in family N may be more comparable to patients with attenuated FAP (AFAP) who develop smaller numbers of adenomas that require two somatic APC mutations for tumour initiation (Nat. Genet. 20, 385-388 (1998)). By contrast, patients with HNPCC develop only one or a few adenomas or carcinomas that

require somatic inactivation of a wild type MMR allele and two somatic APC mutations.

Example 4: Investigation of and Data from Seven

Unrelated Patients

This example describes the identification of seven further, unrelated patients having multiple colo-rectal adenomas and biallelic germline *MYH* mutations, including four cases homozygous for truncating mutations. Colo-rectal tumours from these individuals exhibit a significant excess of somatic G:C→T:A mutations, as compared to sporadic and FAP-associated tumours, confirming that biallelic mutations in *MYH* predispose to CRC.

Methods

Samples

Seventeen unrelated cases of British descent and four unrelated cases of Indian or Pakistani descent, each having multiple colo-rectal adenomas and/or carcinoma were analysed. DNA was prepared from venous blood samples, and from adenoma and carcinoma tissue from colon that had been micro-dissected from paraffin blocks. The nature of all tissues was verified histologically.

PCR amplification

Exons 4 and 9 of *APC*, 1-16 of *MYH*, 1-8 of *OGG1* and 2-5 of *MTH1* as 2, 16, 11 and 4 fragments, were amplified as previously described (Al-Tassan et al, Nat Genet 30 227-232 (2002)). A ~2.8kb segment of *APC* (between codons 653 and 1589) was amplified, which encompassed the somatic

mutation cluster region, as eighteen overlapping fragments. Primer sequences are available at http://www.uwcm.ac.uk/study/medicine/medical_genetics/research/tmg/projects/hMYH.html.

5

Denaturing high performance liquid chromatography (dHPLC) analysis and fraction collection

dHPLC was carried out using the 3500HT WAVE nucleic acid fragment analysis system (Transgenomic Ltd, Crewe Hall
10 Weston Road, Crewe CW1 6UZ, UK). To enhance the formation of heteroduplexes prior to analysis, the PCR products were denatured at 94°C and re-annealed by cooling to 50°C at a rate of 1°C per minute. dHPLC was carried out at the melting temperatures predicted by Wavemaker
15 (version 4.1) software (Transgenomic) with a 12% acetonitrile (ACN) gradient over 3 minutes. Samples displaying aberrant dHPLC elution profiles were sequenced directly; those samples without clear sequence variations were re-analysed by isolating and sequencing dHPLC
20 separated heteroduplexes. Fraction collection of heteroduplexes was carried out using a Transgenomic FCW-200 in-line fragment collector and products were eluted in 8% CAN.

25 *Automated sequencing*

Amplification products were purified using the PCR purification kit (Qiagen, 28159 Avenue Stanford, Valencia, CA 91355, USA) and automated sequencing was carried out using the Big Dye Terminator Cycle Sequencing
30 kit (Applied Biosystems [ABI], Applied Biosystems, 850 Lincoln Center Drive, Foster City, CA 94404, USA) according to the manufacturer's instructions. Sequencing reactions were purified by isopropanol precipitation and analysed on an ABI PRISM 3100 Genetic Analyser.

Mutations were described according to the established nomenclature system (Antonarakis *et al*, Hum Mat 11 1-3 (1998)).

5 *Assays for sequence variants*

All germline mutations in MYH and somatic mutations in APC were confirmed by sequencing two independent PCR products and/or dHPLC separated heteroduplexes, in forward and/or reverse directions. The germline mutations Y90X, G382D and E466X in MYH were further confirmed by restriction enzyme digestion (using *RsaI*, *BglIII* and *ApoI*, respectively). The common polymorphism 972 C>G (H324Q) in exon 12 of MYH was assayed by sequencing.

15

Somatic APC mutation database and statistical analysis

A database of 503 somatic mutations observed in sporadic colo-rectal tumours, and 308 somatic mutations observed in FAP and AFAP associated colo-rectal tumours was compiled (Al-Tassan *et al* Nat Genet 30 227-232 (2002) and Example 1). Statistical analyses were carried out using Fisher's Exact and the chi-squared tests.

Results

25 *Biallelic germline mutations in MYH*

No germline APC truncating mutations were identified. Sequencing of the entire open reading frame (ORF) of MYH in these cases revealed seven patients with biallelic mutations (Table 7), six of which were presumed to be homozygous for MYH variants since no wild type allele could be detected upon sequence analysis.

One Pakistani case (MA27) was homozygous for the exon 3 nonsense mutation Y90X (270.C>A); two British cases (MA22

and MA34) were homozygous for the exon 7 missense mutation Y165C (494 A>G); one British case (MA25) was compound heterozygous for Y165C/G382D (1145 G>A); and three Indian cases (MA20, MA24 and MA26) were homozygous for the exon 14 nonsense mutation E466X (1396 G>T). No samples were available from any of the parents of these patients, all of whom were unaffected, or any other family members. No patients carried single mutant MYH alleles.

10

The recurrent MYH variants Y90X (2 mutations) and Y165C (5 mutations) were found in association with the G allele of the exon 12 polymorphism 972 C>G (H324Q), and the recurrent variant E466X (6 mutations) was always found in association with the C allele. Sequencing of the entire ORFs of OGG1 and MTH1 in the fourteen MYH negative cases did not identify any likely pathogenic changes.

15

Table 7: Biallelic germline MYH mutations in patients with multiple colo-rectal adenomas

Patient	Exon	Nucleotide change	Amino acid change	Ethnic background
5	MA27	3 270 C>A	Y90X	Pakistani
		3 270 C>A	Y90X	
	MA22	7 494 A>G	Y165C	British
		7 494 A>G	Y165C	
10	MA34	7 494 A>G	Y165C	British
		7 494 A>G	Y165C	
37	MA25	7 494 A>G	Y165C	British
		13 1145 G>A	G382D	
15	MA20	14 1396 G>T	E466X	Indian
		14 1396 G>T	E466X	
20	MA24	14 1396 G>T	E466X	Indian
		14 1396 G>T	E466X	
	MA26	14 1396 G>T	E466X	Indian
		14 1396 G>T	E466X	
25				

Identification of somatic G:C→T:A mutations in colo-rectal tumours

Using denaturing high performance liquid chromatography (dHPLC) analysis, we sought somatic mutations in the APC gene in colo-rectal tumours isolated from the patients with biallelic germline MYH mutations.

A region of APC was screened, spanning codons 653 to 1589 which encompassed the mutation cluster region (MCR, codons 1286 to 1513, Miyoshi et al, Hum Mol Genet 1 229-233 (1992)), a known hotspot for somatic mutations (Fearnhead et al, Hum Mol Genet 10 721-733 (2001)). In total, 50 somatic mutations were identified of which 49 were G:C→T:A transversion mutations.

The proportion of somatic G:C→T:A transversion mutations in APC that were detected in colo-rectal tumours from patients with biallelic MYH mutations were compared to a database of 503 reported somatic APC mutations from sporadic colo-rectal adenomas and carcinomas, and 308 somatic mutations from FAP associated colo-rectal tumours. The excess of somatic G:C→T:A transversions in patients with biallelic MYH mutations was highly significant.

Sequence surrounding the somatic G:C→T:A mutations

Examination of the sequence context surrounding the somatic G:C→T:A mutations revealed that the two bases immediately 3' to the mutated G were always AA, irrespective of the nature of the germline MYH mutations. The preponderance of G:C→T:A mutations at GAA sequences is highly significant, since other sequences that could undergo G:C→T:A mutation

to stop codons are equally prevalent in the APC coding region (216 GAA sites vs 213 non-GAA sites, $\chi^2=13.28$, $P=2.7\times 10^{-4}$).

5 **Conclusions**

In this study, another patient compound heterozygous for Y165C/G382D and two patients homozygous for Y165C have been identified. More significantly, four unrelated patients homozygous for nonsense mutations in MYH have been
10 identified. Homozygosity for MYH variants (as opposed to a point mutation on one allele and a large deletion on the other allele) is considered highly likely, since one of the six patients was shown to be from a family with known consanguinity, and four patients were from Indian families,
15 which are known to have a high frequency of first and second cousin marriages. None of the patients harboured truncating mutations in exon 4 or the alternatively spliced region of exon 9 of APC, which would have been consistent with AFAP, and none of the parents who carried single
20 mutant MYH alleles, showed signs of CRC. Therefore, these data unequivocally confirm that biallelic inactivation of MYH predisposes to colo-rectal adenoma and carcinoma.

The recurrent mutations Y90X, Y165C and E466X cannot be
25 readily explained in terms of the well characterised mechanisms of hypermutagenesis, and are associated with specific (and different) alleles of the polymorphism 972 C>G in exon 12 of MYH. These mutations are therefore probably not independent mutational events, but derived
30 from the same ancestral chromosomes. In total, four British families that are either homozygous for Y165C or compound heterozygous for Y165C/G382D, three Indian

families that are homozygous for E466X and a single Indian or Pakistani family that is homozygous for Y90X have been identified. Different mutations in MYH may therefore be more frequent in different ethnic populations, consistent with founder effects, and diagnostic screening strategies may have to be optimised accordingly.

In this study, an unusually high frequency of somatic G:C→T:A mutations was observed in colo-rectal tumours from patients with biallelic MYH mutations, and this confirms the pathogenicity of the germline variants.

Together with the earlier study, likely pathogenic variants in the BER genes OGG1 or MTH1 have not been detected in over thirty cases with multiple colo-rectal adenoma and carcinoma.

In conclusion, the above Example provides evidence for a novel autosomal recessive colo-rectal tumour pre-disposition syndrome, and this is the first reported inherited disorder of base excision repair in humans.

CLAIMS

1. A screening method for identifying an individual having a pre-disposition towards having a cancer, which screening method comprises the steps of:

(a) obtaining a test sample comprising a nucleotide sequence comprised in a gene in a base excision repair (BER) pathway of the individual or an amino acid sequence of a polypeptide expressed thereby; and

(b) comparing a region of the test sample sequence with the corresponding region of the wild type sequence, whereby a difference between the test sample sequence and the wild type sequence signifies that the individual is pre-disposed to having the cancer; and

wherein the difference comprises a variation as defined in any of claims 14 to 18, or G382D-hMYH or a nucleotide sequence encoding it.

2. A screening method for identifying an individual having a cancer or a pre-disposition towards having a cancer, which screening method comprises the steps of:

(a) obtaining a test sample comprising a nucleotide sequence comprised in the *hMYH* gene of the individual or an amino acid sequence of a polypeptide expressed thereby; and

(b) comparing a region of the test sample sequence with the corresponding region of the wild type sequence, whereby a difference between the test sample sequence and the wild type sequence signifies that the individual is pre-disposed to having the cancer; and

wherein the difference comprises a variation as defined in any of claims 14 to 18, or G382D-hMYH or a nucleotide sequence encoding it.

3. A method according to claim 1 or claim 2, wherein the BER pathway gene or the *hMYH* gene is involved in the repair of another gene and protects against G:C→T:A transverse mutations therein.

5

4. A method according to claim 3, wherein the other gene is a known marker for the cancer.

10

5. A method according to claim 3 or claim 4, wherein the other gene is *APC*.

6. A method according to any preceding claim, wherein the cancer is bowel cancer.

15

7. A method according to any of claims 3 to 6, wherein the BER pathway gene is *hMYH*.

20

8. A method according to claim 7, wherein the difference between the test sample sequence and the wild type sequence is Y165C-*hMYH* or a nucleotide sequence encoding therefor, as defined herein [SEQ ID NO:1].

25

9. A method according to claim 7, wherein the difference between the test sample sequence and the wild type sequence is E466X-*hMYH* or a nucleotide sequence encoding therefor, as defined herein [SEQ ID NO:3].

30

10. A method according to claim 7, wherein the difference between the test sample sequence and the wild type sequence is Y90X-*hMYH* or a nucleotide sequence encoding therefor, as defined herein [SEQ ID NO:4].

11. A method according to claim 7, wherein the difference between the test sample sequence and the wild type sequence is the known G382D-hMYH or a nucleotide sequence encoding therefor.

12. A method according to any preceding claim, wherein the test sample comprises genomic DNA or mRNA, or the polypeptide expressed thereby.

13. A screening method according to any preceding claim, wherein the step (b) comprises a protein or polypeptide sequencing method.

14. A variant of hMYH, suitable for use in a screening method according to any of claims 1 to 13, comprising a polypeptide variant selected from:

- (i) Y165C, as defined herein [SEQ ID NO:1];
- (ii) E466X, as defined herein [SEQ ID NO:3]; and
- (iii) Y90X, as defined herein [SEQ ID NO:4].

15. A nucleotide sequence

- (a) encoding the polypeptide variant according to claim 14
- (b) a sequence substantially homologous to or that hybridises to sequence (a) under stringent conditions; or
- (c) a sequence substantially homologous to or that hybridizes under stringent conditions to the sequence (a) or (b) but for the degeneracy of the genetic code; or
- (d) an oligonucleotide specific for any of the sequences (a), (b) or (c).

16. A variant according to claim 14 or 15, comprising Y165C or a nucleotide sequence encoding the polypeptide, as defined herein [SEQ ID NO:1].

5 17. A variant according to claim 14 or 15, comprising E466X or a nucleotide sequence encoding the polypeptide, as defined herein [SEQ ID NO:3].

10 18. A variant according to claim 14 or 15, comprising Y90X or a nucleotide sequence encoding the polypeptide, as defined herein [SEQ ID NO:4].

19. Use of:

- 15 (a) a BER gene variant, or a polypeptide encoded thereby,
(b) a sequence substantially homologous to or that hybridises to a nucleotide sequence (a) under stringent conditions; or
(c) a sequence substantially homologous to or that hybridizes under stringent conditions to the nucleotide
20 sequence (a) or (b) but for the degeneracy of the genetic code; or
(d) an oligonucleotide specific for any of the nucleotide sequences (a), (b) or (c)
in a therapeutic, diagnostic or detection method, wherein
25 the variant is as defined in any of claims 14 to 18, or G382D-hMYH or a nucleotide sequence encoding it.

30 20. Use according to claim 19 in the diagnosis or detection of a variation in hMYH and/or a condition of the APC gene.

21. Use according to claim 19 or claim 20, for the determination of pre-disposition or susceptibility to a disease, such as cancer.

5 22. Use according to any of claims 19 to 21, wherein the corresponding wild type BER pathway gene or wild type *hMYH* gene acts to protect against G:C→T:A transverse mutations in a cancer marker gene, including APC.

10 23. A kit suitable for use in carrying out a screening method according to any of claims 1 to 13, which kit comprises one or more of:

(a) an oligo- or poly-nucleotide having a nucleic acid sequence corresponding to a region of a variant BER gene
15 and/or a variant *hMYH*, or an amino acid sequence encoded thereby, which region incorporates at least one variation from the corresponding wild-type sequence as defined in claim 14 to ...;

(b) a nucleotide or amino acid sequence corresponding to
20 the wild-type sequence in the region specified in (a);

(c) one or more reagents suitable for carrying out PCR for amplifying desired regions of the patient's DNA; and/or

(d) a 'surrogate marker' that is indicative of or
25 correlated to the presence of a variant of a BER gene or the polypeptide encoded thereby.

24. A kit according to claim 23, wherein component (c) comprises one or more of: PCR primers corresponding to an exon of the BER and/or *hMYH* gene, and/or primers as defined
30 herein in the Examples; and/or *Taq* DNA polymerase.

25. A kit according to claim 23 or claim 24, wherein component (d) is selected from one or more of:

(a) any biomolecule, including nucleotides and proteins, including antibodies, sugars, and lipids;

5 (b) a chemical compound, including drugs and metabolites thereof; and/or

(c) a physical characteristic,

whose absence, presence, or quantity in an individual is measurable and correlated with the presence of a BER and/or
10 an *hMYH* gene variant, or a protein encoded thereby.

26. Use of a BER gene and/or *hMYH*, or a variant thereof, or a polypeptide encoded thereby, as defined in any of claims 14 to 18, or G382D-*hMYH* or a nucleotide sequence
15 encoding it, for the preparation of a therapeutic composition, diagnostics composition or kit, or detection kit.

27. A screening method for identifying an individual
20 having an inherited defect in a BER gene and/or *hMYH* and/or who is pre-disposed to having a cancer, which screening method comprises the steps of:

(a) obtaining a test sample comprising a nucleotide sequence comprised in a marker gene of the individual or an
25 amino acid sequence of a polypeptide expressed thereby, the marker gene being a marker for a disease or condition correlated with the presence of a defect in a BER gene and/or *hMYH* or cancer; and

(b) comparing a region of the test sample sequence
30 with the following diagnostic sequence [SEQ ID NO: 2] or a peptide encoded thereby:

5' -X-X1-A-A-X2-A-A- 3' [SEQ ID NO: 2]

wherein X is A or T; X1 is G or T; and X2 is G or A;
whereby identity between the region of the test sample gene
sequence and the diagnostic sequence signifies that the
individual has an inherited defect in a BER gene and/or
hMYH and/or is pre-disposed to having the cancer.

28. A method according to claim 27, wherein the marker
gene is *APC*.

29. A diagnostic sequence suitable for use in a screening
method according to any of claims 27 to 28, comprising:

(a) 5' -X-X1-A-A-X2-A-A- 3' [SEQ ID NO: 2]

wherein X is A or T; X1 is G or T; and X2 is G or A;

(b) a sequence substantially homologous to or that
hybridises to sequence (a) under stringent conditions; or

(c) a sequence substantially homologous to or that
hybridizes under stringent conditions to the sequence (a)
or (b) but for the degeneracy of the genetic code; or

(d) an oligonucleotide specific for any of the sequences
(a), (b) or (c).

30. Use of:

(a) a diagnostic sequence according to claim 29, or a
polypeptide encoded thereby,

(b) a sequence substantially homologous to or that
hybridises to nucleotide sequence (a) under stringent
conditions; or

(c) a sequence substantially homologous to or that hybridizes under stringent conditions to the nucleotide sequence (a) or (b) but for the degeneracy of the genetic code; or

5 (d) an oligonucleotide specific for any of the nucleotide sequences (a), (b) or (c)
in a therapeutic, diagnostic or detection method.

31. Use according to claim 30, for the determination of
.0 susceptibility to a disease, including cancer.

32. Use according to claim 30 or claim 31, wherein the diagnostic sequence is for detecting in a cancer marker gene, including APC.

.5 33. A kit suitable for use in carrying out a screening method according to claim 27 or claim 28, which kit comprises one or more of:

(a) an oligo- or poly-nucleotide comprising a diagnostic
20 sequence according to claim 29, or an amino acid sequence encoded thereby;

(b) one or more reagents suitable for carrying out PCR for amplifying desired regions of the patient's DNA; and/or

(c) a 'surrogate marker' that is indicative of or
25 correlated to the presence of a variant of a sequence (a).



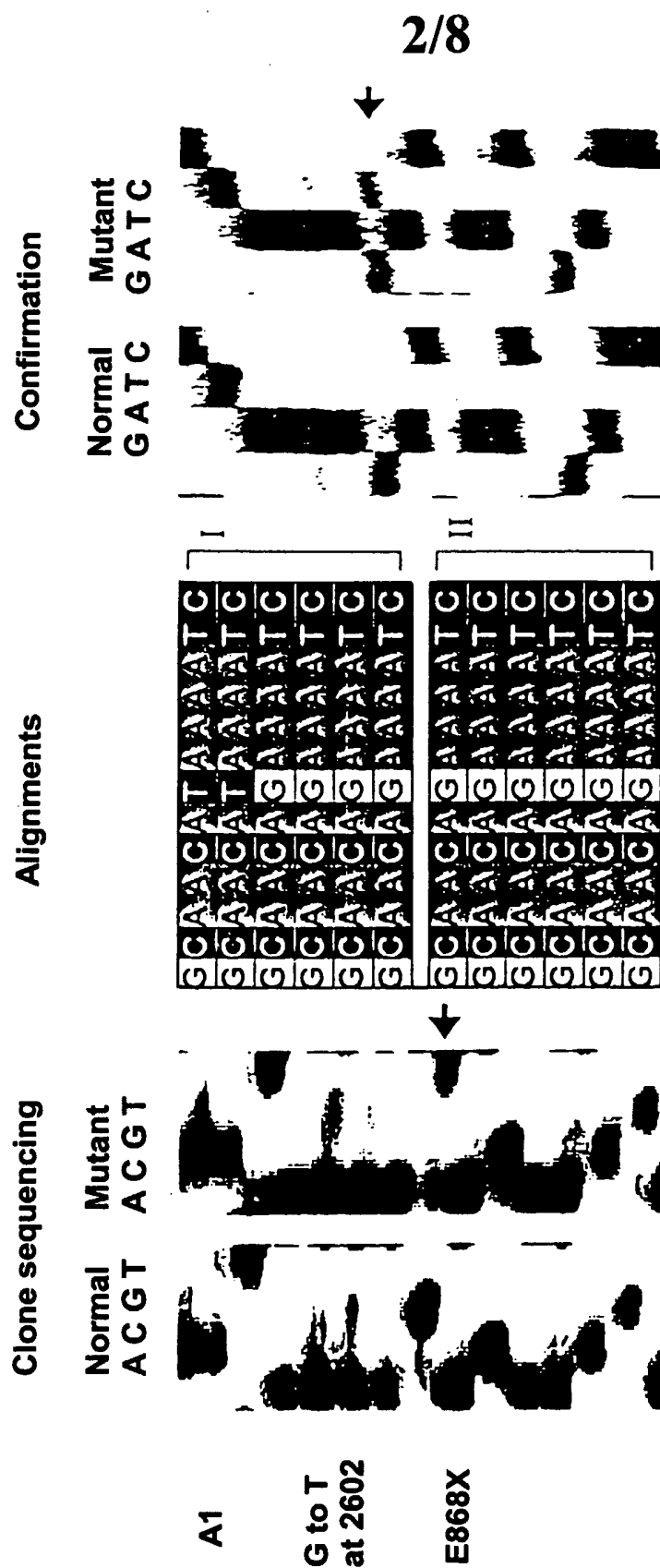


Fig. 2(a)

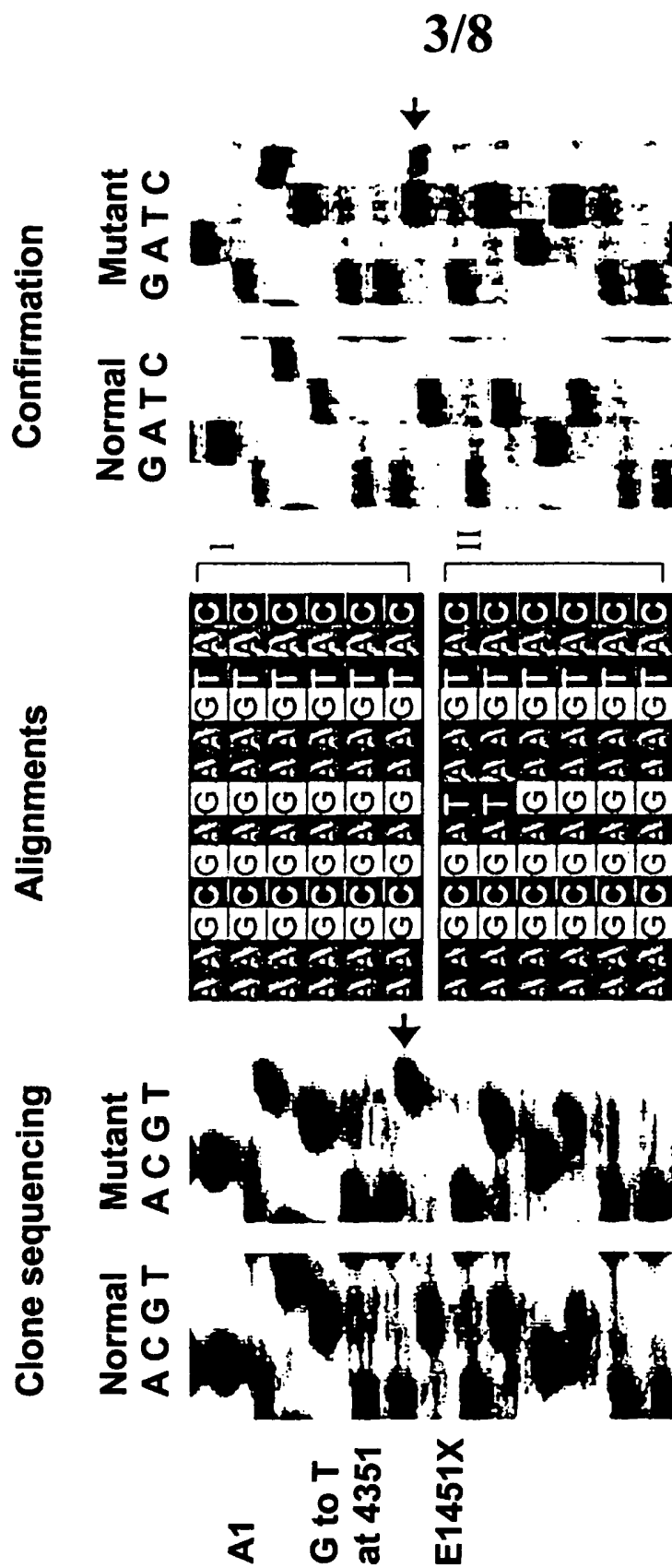


Fig. 2(b)

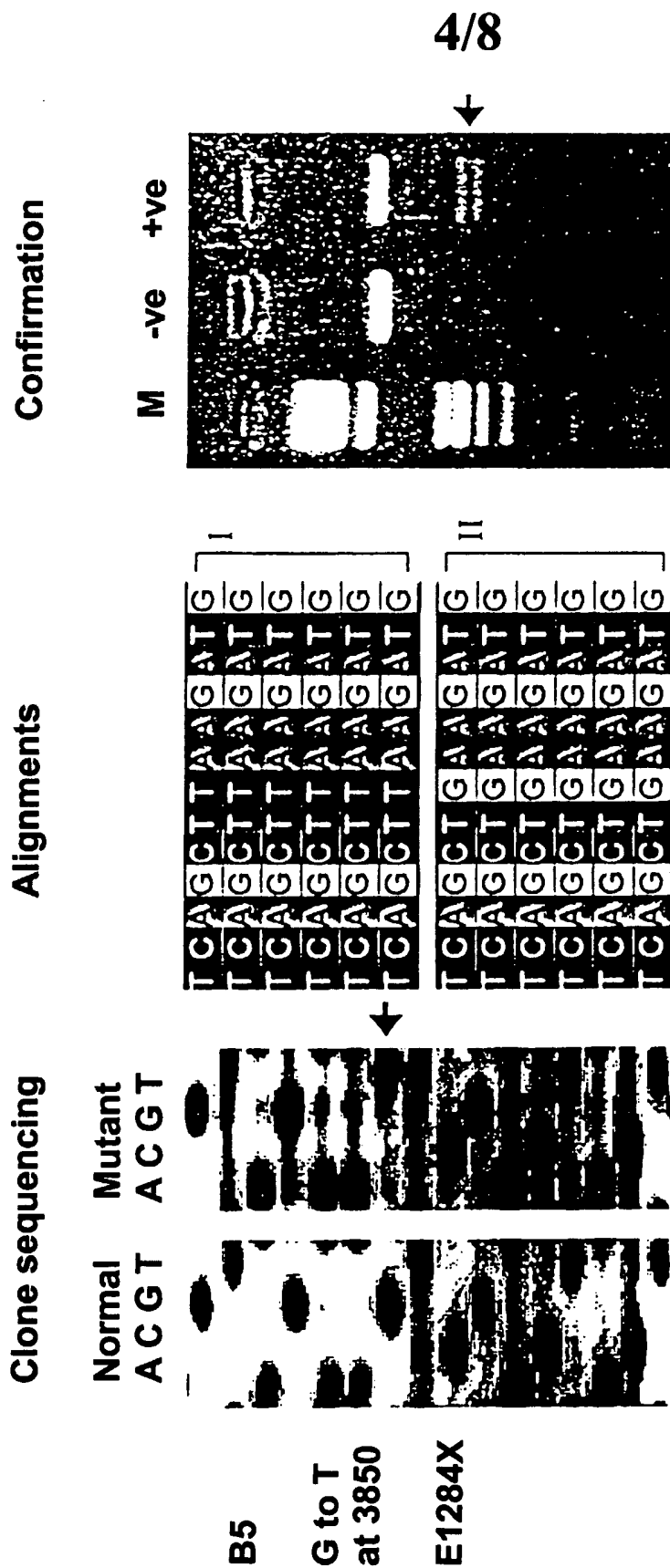


Fig. 2(c)

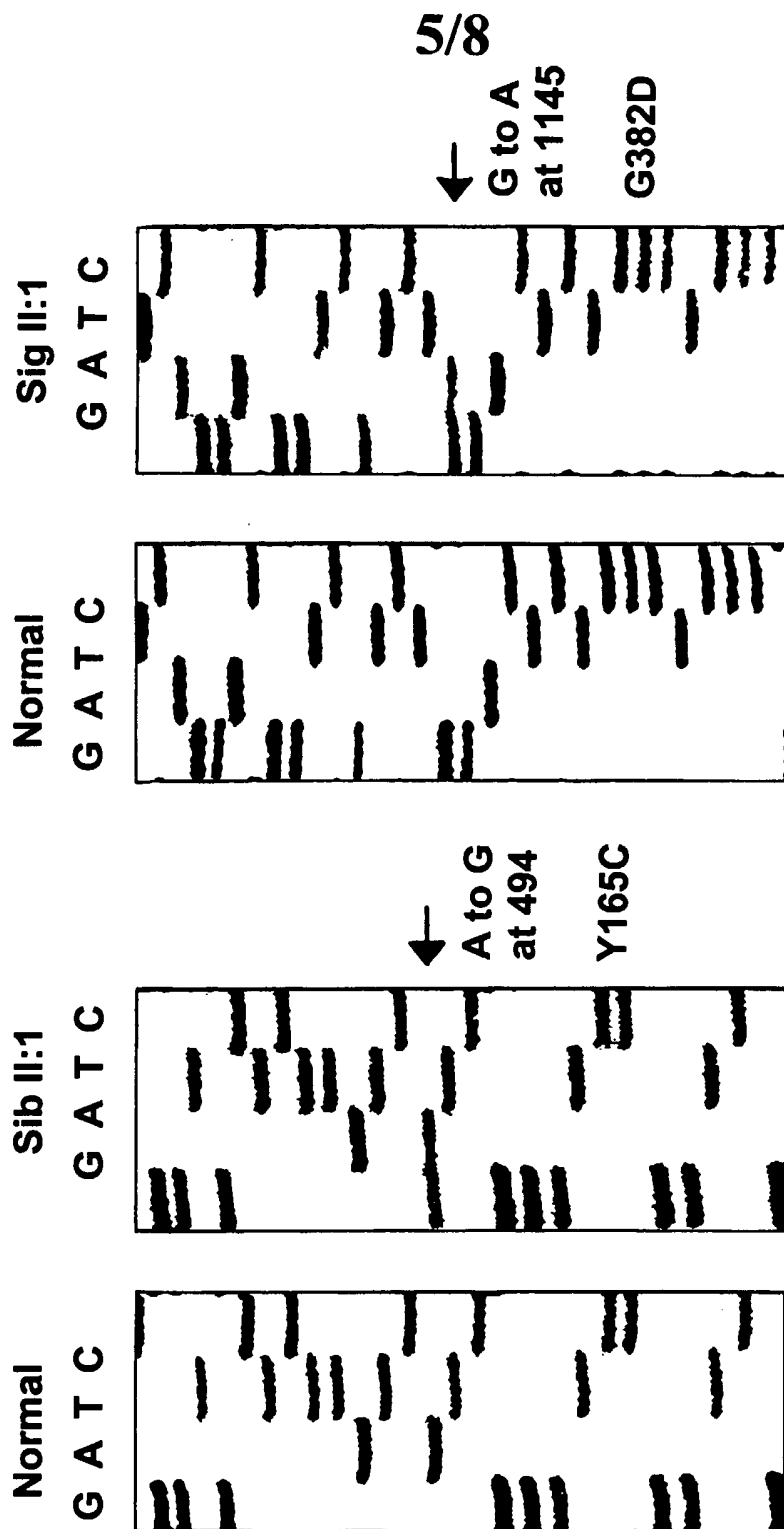


Fig. 3(b)

Fig. 3(a)

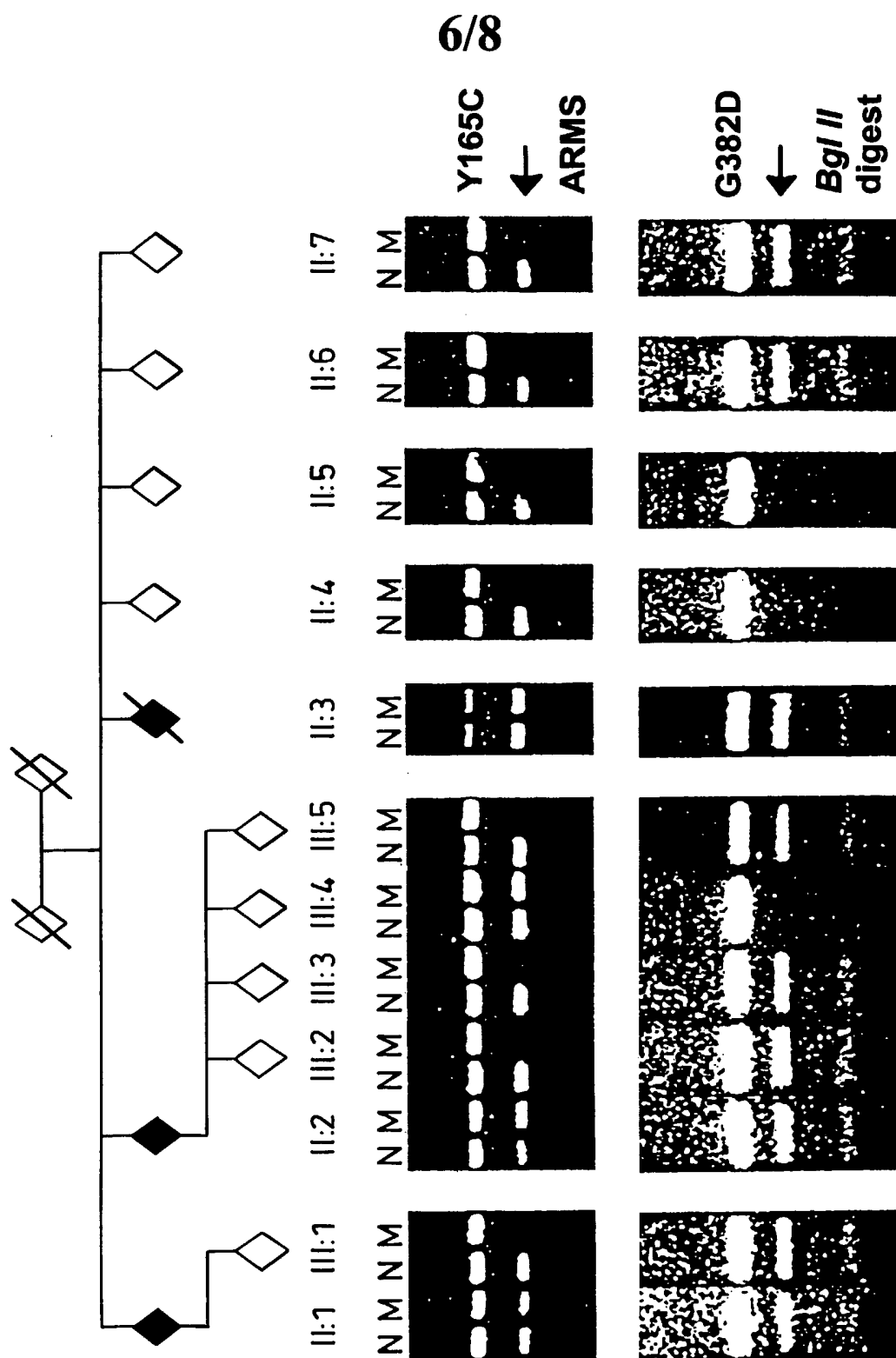


Fig. 3(c)

↓

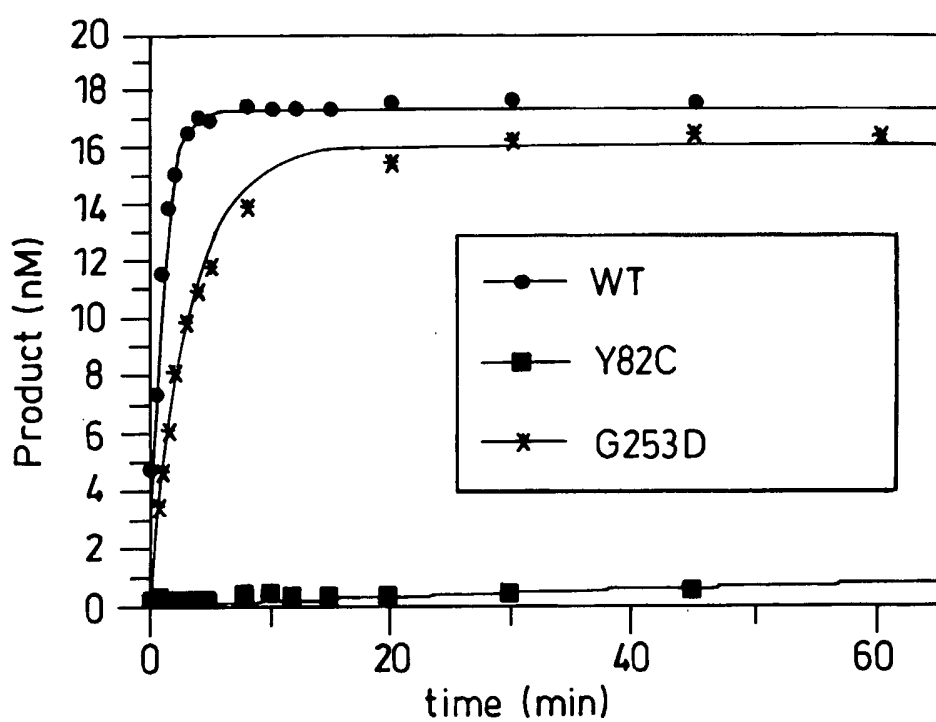
H. sap.	132	INNTGMMQKWP	TLQDLASASLE	----	EVNQLWAGIGHYNS	SRGRRLE	QEGARKVVEELGGHMP	RTAETLQQLIP	199
M. mus.	117	IDNTTRMMQKWP	KLQDLASASLE	----	EVNQLWMSGIGYNS	SRGRRLE	QEGARKVVEELGGHMP	RTAETLQQLIP	184
A. tha.	180	MKVMKRWQKWP	TIYDLGQASLEN	/	EVNEMWAGIGYARRARE	LELEGAKMVVAGTEG	-EPNQASSIMK	-VK	264
S. pom.	81	KRYNTKMETLP	TLKSCAEAEYNT	----	QVMPLWSGCGEYTRCKR	LHQACQHLAKLHPSEI	PPRTGDEWAKGIP		149
H. inf.	54	IPYFEREIKTFP	NI TALANASQD	----	EVLHLWTGIGYARARN	LHKAAOKVRDEFNGNEP	TNFEQVWA	-ILS	120
V. cho.	47	IPYFEREIERFP	TVHALAAPQD	----	EVLHEWTGIGYARARN	LHKAAQMVVSEYGGEEP	PTDLEQMNA	-ILP	113
S. typ.	49	IPYFEREMARFP	TVTDLANAPLD	----	EVLHLMTGIGYARARN	LHKAAQQVATLHGGEEP	QTFAEIAA	-ILP	115
E. coli	49	IPYFEREMARFP	TVTDLANAPLD	----	EVLHLMTGIGYARARN	LHKAAQQVATLHGGKEP	ETTFEEVAA	-ILP	115

Fig. 4(a)

↓

H. sap.	353	-PREESSATC	VLE--OP	----	GALGAQ--ILIV	-QRPNSGLLAGLWEEF	SVTW--E--PSE	----	QLORKA	---L	406
M. mus.	335	-PREEYSATC	VVE--OP	----	GAIGGPLVLIV	-QRPDSGLLAGLWEEF	SVTL--E--PSE	----	QHOKA	---L	389
A. tha.	361	-PRHDFCCVC	VLEIHNLER	NQSGGR-FVIV	-KRPEQGLLAGLWEEF	SVILN-E--EADS	--ATTRNAINVY				423
S. pom.	296	QR-EERALV	VIFQ-KT	----	DPSTKEKFFLIRKPS	AGLLAGLWDEFT	IEFGQESWPKDMDAE	FKSIA	-QW		360
H. inf.	233	MP-EKTTYFL	ILS-KN	----	GK-----VCDE	-QRENSGLWGLFCFP	--QF--E--DKS	----	SLH	-----F	278
V. cho.	226	KP-VKATWF	VMLY-HD	----	NA-----VWIE	-QRPQSGIWWGLYCFP	--Q--S--EIA	----	NIQT	-----T	270
S. typ.	228	LP-ERTGYFL	ILQ-HN	----	QE-----IFLA	-QRPSPGIWWGLYCFP	--QF--A--RED	----	ELRE	-----W	273
E. coli	228	LP-ERTGYFL	ILQ-HE	----	DE-----VLLA	-QRPSPGIWWGLYCFP	--QF--A--DEE	----	SLRQ	-----W	273

Fig. 4(b)

*Fig. 5*

SEQ ID NO:1 - Y165C Mutant sequence

Mutant hMYH sequence: 494 A to G (Y165C) mutant. The mutant codon and amino acid indicated in bold and underlined

(Single letter amino acid sequence above and cDNA sequence below).

5	M	T	P	L	V	S	R	L	S	R	L	W	A	13	
	ATG	ACA	CCG	CTC	GTC	TCC	CGC	CTG	AGT	CGT	CTG	TGG	GCC	39	
10	I	M	R	K	P	R	A	A	V	G	S	G	H	26	
15	ATC	ATG	AGG	AAG	CCA	CGA	GCA	GCC	GTG	GGA	AGT	GGT	CAC	78	
	R	K	Q	A	A	S	Q	E	G	R	Q	K	H	39	
	AGG	AAG	CAG	GCA	GCC	AGC	CAG	GAA	GGG	AGG	CAG	AAG	CAT	117	
20	A	K	N	N	S	Q	A	K	P	S	A	C	D	52	
	GCT	AAG	AAC	AAC	AGT	CAG	GCC	AAG	CCT	TCT	GCC	TGT	GAT	156	
	G	L	A	R	Q	P	E	E	V	V	L	Q	A	65	
	GGC	CTG	GCC	AGG	CAG	CCG	GAA	GAG	GTG	GTA	TTG	CAG	GCC	195	
25	S	V	S	S	Y	H	L	F	R	D	V	A	E	78	
	TCT	GTC	TCC	TCA	TAC	CAT	CTA	TTC	AGA	GAC	GTA	GCT	GAA	234	
	V	T	A	F	R	G	S	L	L	S	W	Y	D	91	
30	GTC	ACA	GCC	TTC	CGA	GGG	AGC	CTG	CTA	AGC	TGG	TAC	GAC	273	
	Q	E	K	R	D	L	P	W	R	R	R	A	E	104	
	CAA	GAG	AAA	CGG	GAC	CTA	CCA	TGG	AGA	AGA	CGG	GCA	GAA	312	
35	D	E	M	D	L	D	R	R	A	Y	A	V	W	117	
	GAT	GAG	ATG	GAC	CTG	GAC	AGG	CGG	GCA	TAT	GCT	GTG	TGG	351	
	V	S	E	V	M	L	Q	Q	T	Q	V	A	T	130	
	GTC	TCA	GAG	GTC	ATG	CTG	CAG	CAG	ACC	CAG	GTT	GCC	ACT	390	
40	V	I	N	Y	Y	T	G	W	M	Q	K	W	P	143	
	GTG	ATC	AAC	TAC	TAT	ACC	GGA	TGG	ATG	CAG	AAG	TGG	CCT	429	
	T	L	Q	D	L	A	S	A	S	L	E	E	V	156	
45	ACA	CTG	CAG	GAC	CTG	GCC	AGT	GCT	TCC	CTG	GAG	GAG	GTG	468	
	N	Q	L	W	A	G	L	G	<u>GGG</u>	<u>GGG</u>	Y	S	R	G	169
	AAT	CAA	CTC	TGG	GCT	GGC	CTG	GGC	<u>GGG</u>	<u>GGG</u>	TAT	TCT	CGT	GGC	507

	R	R	L	Q	E	G	A	R	K	V	V	E	E	182
	CGG	CGG	CTG	CAG	GAG	GGA	GCT	CGG	AAG	GTG	GTA	GAG	GAG	546
5	L	G	G	H	M	P	R	T	A	E	T	L	Q	195
	CTA	GGG	GGC	CAC	ATG	CCA	CGT	ACA	GCA	GAG	ACC	CTG	CAG	585
	Q	L	L	P	G	V	G	R	Y	T	A	G	A	208
10	CAG	CTC	CTG	CCT	GGC	GTG	GGG	CGC	TAC	ACA	GCT	GGG	GCC	624
	I	A	S	I	A	F	G	Q	A	T	G	V	V	221
	ATT	GCC	TCT	ATC	GCC	TTT	GGC	CAG	GCA	ACC	GGT	GTG	GTG	663
15	D	G	N	V	A	R	V	L	C	R	V	R	A	234
	GAT	GGC	AAC	GTA	GCA	CGG	GTG	CTG	TGC	CGT	GTC	CGA	GCC	702
	I	G	A	D	P	S	S	T	L	V	S	Q	Q	247
	ATT	GGT	GCT	GAT	CCC	AGC	AGC	ACC	CTT	GTT	TCC	CAG	CAG	741
20	L	W	G	L	A	Q	Q	L	V	D	P	A	R	260
	CTC	TGG	GGT	CTA	GCC	CAG	CAG	CTG	GTG	GAC	CCA	GCC	CGG	780
	P	G	D	F	N	Q	A	A	M	E	L	G	A	273
25	CCA	GGA	GAT	TTC	AAC	CAA	GCA	GCC	ATG	GAG	CTA	GGG	GCC	819
	T	V	C	T	P	Q	R	P	L	C	S	Q	C	286
	ACA	GTG	TGT	ACC	CCA	CAG	CGC	CCA	CTG	TGC	AGC	CAG	TGC	858
30	P	V	E	S	L	C	R	A	R	Q	R	V	E	299
	CCT	GTG	GAG	AGC	CTG	TGC	CGG	GCA	CGC	CAG	AGA	GTG	GAG	897
	Q	E	Q	L	L	A	S	G	S	L	S	G	S	312
	CAG	GAA	CAG	CTC	TTA	GCC	TCA	GGG	AGC	CTG	TCG	GGC	AGT	936
35	P	D	V	E	E	C	A	P	N	T	G	Q	C	325
	CCT	GAC	GTG	GAG	GAG	TGT	GCT	CCC	AAC	ACT	GGA	CAG	TGC	975
40	H	L	C	L	P	P	S	E	P	W	D	Q	T	338
	CAC	CTG	TGC	CTG	CCT	CCC	TCG	GAG	CCC	TGG	GAC	CAG	ACC	1014
	L	G	V	V	N	F	P	R	K	A	S	R	K	351
	CTG	GGA	GTG	GTC	AAC	TTC	CCC	AGA	AAG	GCC	AGC	CGC	AAG	1053
45	P	P	R	E	E	S	S	A	T	C	V	L	E	364
	CCC	CCC	AGG	GAG	GAG	AGC	TCT	GCC	ACC	TGT	GTT	CTG	GAA	1092
	Q	P	G	A	L	G	A	Q	I	L	L	V	Q	377
	CAG	CCT	GGG	GCC	CTT	GGG	GCC	CAA	ATT	CTG	CTG	GTG	CAG	1131

	R	P	N	S	G	L	L	A	G	L	W	E	F	390
	AGG	CCC	AAC	TCA	GGT	CTG	CTG	GCA	GGA	CTG	TGG	GAG	TTC	1170
5	P	S	V	T	W	E	P	S	E	Q	L	Q	R	403
	CCG	TCC	GTG	ACC	TGG	GAG	CCC	TCA	GAG	CAG	CTT	CAG	CGC	1209
	K	A	L	L	Q	E	L	Q	R	W	A	G	P	416
	AAG	GCC	CTG	CTG	CAG	GAA	CTA	CAG	CGT	TGG	GCT	GGG	CCC	1248
10	L	P	A	T	H	L	R	H	L	G	E	V	V	429
	CTC	CCA	GCC	ACG	CAC	CTC	CGG	CAC	CTT	GGG	GAG	GTT	GTC	1287
	H	T	F	S	H	I	K	L	T	Y	Q	V	Y	442
15	CAC	ACC	TTC	TCT	CAC	ATC	AAG	CTG	ACA	TAT	CAA	GTA	TAT	1326
	G	L	A	L	E	G	Q	T	P	V	T	T	V	455
	GGG	CTG	GCC	TTG	GAA	GGG	CAG	ACC	CCA	GTG	ACC	ACC	GTA	1365
20	P	P	G	A	R	W	L	T	Q	E	E	F	H	468
	CCA	CCA	GGT	GCT	CGC	TGG	CTG	ACG	CAG	GAG	GAA	TTT	CAC	1404
	T	A	A	V	S	T	A	M	K	K	V	F	R	481
	ACC	GCA	GCT	GTT	TCC	ACC	GCC	ATG	AAA	AAG	GTT	TTC	CGT	1443
25	V	Y	Q	G	Q	Q	P	G	T	C	M	G	S	494
	GTG	TAT	CAG	GGC	CAA	CAG	CCA	GGG	ACC	TGT	ATG	GGT	TCC	1482
	K	R	S	Q	V	S	S	P	C	S	R	K	K	407
30	AAA	AGG	TCC	CAG	GTG	TCC	TCT	CCG	TGC	AGT	CGG	AAA	AAG	1521
	P	R	M	G	Q	Q	V	L	D	N	F	F	R	520
	CCC	CGC	ATG	GGC	CAG	CAA	GTC	CTG	GAT	AAT	TTC	TTT	CGG	1560
35	S	H	I	S	T	D	A	H	S	L	N	S	A	533
	TCT	CAC	ATC	TCC	ACT	GAT	GCA	CAC	AGC	CTC	AAC	AGT	GCA	1599
	A	Q	*											535
40	GCC	CAG	TGA											1608

SEQ ID NO:3 - E466X Mutant sequence

Mutant hMYH sequence: 1396 G to T (E466X) mutant. The mutant codon and amino acid indicated in bold and underlined

(Single letter amino acid sequence above and cDNA sequence below).

10	M	T	P	L	V	S	R	L	S	R	L	W	A	13
	ATG	ACA	CCG	CTC	GTC	TCC	CGC	CTG	AGT	CGT	CTG	TGG	GCC	39
	I	M	R	K	P	R	A	A	V	G	S	G	H	26
15	ATC	ATG	AGG	AAG	CCA	CGA	GCA	GCC	GTG	GGA	AGT	GGT	CAC	78
	R	K	Q	A	A	S	Q	E	G	R	Q	K	H	39
	AGG	AAG	CAG	GCA	GCC	AGC	CAG	GAA	GGG	AGG	CAG	AAG	CAT	117
20	A	K	N	N	S	Q	A	K	P	S	A	C	D	52
	GCT	AAG	AAC	AAC	AGT	CAG	GCC	AAG	CCT	TCT	GCC	TGT	GAT	156
	G	L	A	R	Q	P	E	E	V	V	L	Q	A	65
	GGC	CTG	GCC	AGG	CAG	CCG	GAA	GAG	GTG	GTA	TTG	CAG	GCC	195
25	S	V	S	S	Y	H	L	F	R	D	V	A	E	78
	TCT	GTC	TCC	TCA	TAC	CAT	CTA	TTC	AGA	GAC	GTA	GCT	GAA	234
	V	T	A	F	R	G	S	L	L	S	W	Y	D	91
30	GTC	ACA	GCC	TTC	CGA	GGG	AGC	CTG	CTA	AGC	TGG	TAC	GAC	273
	Q	E	K	R	D	L	P	W	R	R	R	A	E	104
	CAA	GAG	AAA	CGG	GAC	CTA	CCA	TGG	AGA	AGA	CGG	GCA	GAA	312
35	D	E	M	D	L	D	R	R	A	Y	A	V	W	117
	GAT	GAG	ATG	GAC	CTG	GAC	AGG	CGG	GCA	TAT	GCT	GTG	TGG	351
	V	S	E	V	M	L	Q	Q	T	Q	V	A	T	130
	GTC	TCA	GAG	GTC	ATG	CTG	CAG	CAG	ACC	CAG	GTT	GCC	ACT	390
40	V	I	N	Y	Y	T	G	W	M	Q	K	W	P	143
	GTG	ATC	AAC	TAC	TAT	ACC	GGA	TGG	ATG	CAG	AAG	TGG	CCT	429
	T	L	Q	D	L	A	S	A	S	L	E	E	V	156
45	ACA	CTG	CAG	GAC	CTG	GCC	AGT	GCT	TCC	CTG	GAG	GAG	GTG	468
	N	Q	L	W	A	G	L	G	Y	Y	S	R	G	169
	AAT	CAA	CTC	TGG	GCT	GGC	CTG	GGC	TAC	TAT	TCT	CGT	GGC	507

	R	R	L	Q	E	G	A	R	K	V	V	E	E	182
	CGG	CGG	CTG	CAG	GAG	GGA	GCT	CGG	AAG	GTG	GTA	GAG	GAG	546
5	L	G	G	H	M	P	R	T	A	E	T	L	Q	195
	CTA	GGG	GGC	CAC	ATG	CCA	CGT	ACA	GCA	GAG	ACC	CTG	CAG	585
	Q	L	L	P	G	V	G	R	Y	T	A	G	A	208
10	CAG	CTC	CTG	CCT	GGC	GTG	GGG	CGC	TAC	ACA	GCT	GGG	GCC	624
	I	A	S	I	A	F	G	Q	A	T	G	V	V	221
	ATT	GCC	TCT	ATC	GCC	TTT	GGC	CAG	GCA	ACC	GGT	GTG	GTG	663
15	D	G	N	V	A	R	V	L	C	R	V	R	A	234
	GAT	GGC	AAC	GTA	GCA	CGG	GTG	CTG	TGC	CGT	GTC	CGA	GCC	702
	I	G	A	D	P	S	S	T	L	V	S	Q	Q	247
	ATT	GGT	GCT	GAT	CCC	AGC	AGC	ACC	CTT	GTT	TCC	CAG	CAG	741
20	L	W	G	L	A	Q	Q	L	V	D	P	A	R	260
	CTC	TGG	GGT	CTA	GCC	CAG	CAG	CTG	GTG	GAC	CCA	GCC	CGG	780
	P	G	D	F	N	Q	A	A	M	E	L	G	A	273
25	CCA	GGA	GAT	TTC	AAC	CAA	GCA	GCC	ATG	GAG	CTA	GGG	GCC	819
	T	V	C	T	P	Q	R	P	L	C	S	Q	C	286
	ACA	GTG	TGT	ACC	CCA	CAG	CGC	CCA	CTG	TGC	AGC	CAG	TGC	858
30	P	V	E	S	L	C	R	A	R	Q	R	V	E	299
	CCT	GTG	GAG	AGC	CTG	TGC	CGG	GCA	CGC	CAG	AGA	GTG	GAG	897
	Q	E	Q	L	L	A	S	G	S	L	S	G	S	312
35	CAG	GAA	CAG	CTC	TTA	GCC	TCA	GGG	AGC	CTG	TCG	GGC	AGT	936
	P	D	V	E	E	C	A	P	N	T	G	Q	C	325
	CCT	GAC	GTG	GAG	GAG	TGT	GCT	CCC	AAC	ACT	GGA	CAG	TGC	975
40	H	L	C	L	P	P	S	E	P	W	D	Q	T	338
	CAC	CTG	TGC	CTG	CCT	CCC	TCG	GAG	CCC	TGG	GAC	CAG	ACC	1014
	L	G	V	V	N	F	P	R	K	A	S	R	K	351
	CTG	GGA	GTG	GTC	AAC	TTC	CCC	AGA	AAG	GCC	AGC	CGC	AAG	1053
45	P	P	R	E	E	S	S	A	T	C	V	L	E	364
	CCC	CCC	AGG	GAG	GAG	AGC	TCT	GCC	ACC	TGT	GTT	CTG	GAA	1092
	Q	P	G	A	L	G	A	Q	I	L	L	V	Q	377
	CAG	CCT	GGG	GCC	CTT	GGG	GCC	CAA	ATT	CTG	CTG	GTG	CAG	1131

	R	P	N	S	G	L	L	A	G	L	W	E	F	390
	AGG	CCC	AAC	TCA	GGT	CTG	CTG	GCA	GGA	CTG	TGG	GAG	TTC	1170
5	P	S	V	T	W	E	P	S	E	Q	L	Q	R	403
	CCG	TCC	GTG	ACC	TGG	GAG	CCC	TCA	GAG	CAG	CTT	CAG	CGC	1209
	K	A	L	L	Q	E	L	Q	R	W	A	G	P	416
10	AAG	GCC	CTG	CTG	CAG	GAA	CTA	CAG	CGT	TGG	GCT	GGG	CCC	1248
	L	P	A	T	H	L	R	H	L	G	E	V	V	429
	CTC	CCA	GCC	ACG	CAC	CTC	CGG	CAC	CTT	GGG	GAG	GTT	GTC	1287
15	H	T	F	S	H	I	K	L	T	Y	Q	V	Y	442
	CAC	ACC	TTC	TCT	CAC	ATC	AAG	CTG	ACA	TAT	CAA	GTA	TAT	1326
	G	L	A	L	E	G	Q	T	P	V	T	T	V	455
	GGG	CTG	GCC	TTG	GAA	GGG	CAG	ACC	CCA	GTG	ACC	ACC	GTA	1365
20	P	P	G	A	R	W	L	T	Q	E		F	H	468
	CCA	CCA	GGT	GCT	CGC	TGG	CTG	ACG	CAG	GAG		TTT	CAC	1404
	T	A	A	V	S	T	A	M	K	K	V	F	R	481
25	ACC	GCA	GCT	GTT	TCC	ACC	GCC	ATG	AAA	AAG	GTT	TTC	CGT	1443
	V	Y	Q	G	Q	Q	P	G	T	C	M	G	S	494
	GTG	TAT	CAG	GGC	CAA	CAG	CCA	GGG	ACC	TGT	ATG	GGT	TCC	1482
30	K	R	S	Q	V	S	S	P	C	S	R	K	K	407
	AAA	AGG	TCC	CAG	GTG	TCC	TCT	CCG	TGC	AGT	CGG	AAA	AAG	1521
	P	R	M	G	Q	Q	V	L	D	N	F	F	R	520
35	CCC	CGC	ATG	GGC	CAG	CAA	GTC	CTG	GAT	AAT	TTC	TTT	CGG	1560
	S	H	I	S	T	D	A	H	S	L	N	S	A	533
	TCT	CAC	ATC	TCC	ACT	GAT	GCA	CAC	AGC	CTC	AAC	AGT	GCA	1599
40	A	Q	*											535
	GCC	CAG	TGA											1608

SEQ ID NO:4 - Y90X Mutant sequence

Mutant hMYH sequence: 270 C to A (Y90X) mutant. The mutant codon and amino acid indicated in bold and underlined

5

(Single letter amino acid sequence above and cDNA sequence below).

10	M	T	P	L	V	S	R	L	S	R	L	W	A	13
	ATG	ACA	CCG	CTC	GTC	TCC	CGC	CTG	AGT	CGT	CTG	TGG	GCC	39
	I	M	R	K	P	R	A	A	V	G	S	G	H	26
	ATC	ATG	AGG	AAG	CCA	CGA	GCA	GCC	GTG	GGA	AGT	GGT	CAC	78
15	R	K	Q	A	A	S	Q	E	G	R	Q	K	H	39
	AGG	AAG	CAG	GCA	GCC	AGC	CAG	GAA	GGG	AGG	CAG	AAG	CAT	117
	A	K	N	N	S	Q	A	K	P	S	A	C	D	52
20	GCT	AAG	AAC	AAC	AGT	CAG	GCC	AAG	CCT	TCT	GCC	TGT	GAT	156
	G	L	A	R	Q	P	E	E	V	V	L	Q	A	65
	GGC	CTG	GCC	AGG	CAG	CCG	GAA	GAG	GTG	GTA	TTG	CAG	GCC	195
25	S	V	S	S	Y	H	L	F	R	D	V	A	E	78
	TCT	GTC	TCC	TCA	TAC	CAT	CTA	TTC	AGA	GAC	GTA	GCT	GAA	234
	V	T	A	F	R	G	S	L	L	S	W	<u>Y</u>	D	91
	GTC	ACA	GCC	TTC	CGA	GGG	AGC	CTG	CTA	AGC	TGG	<u>TAA</u>	GAC	273
30	Q	E	K	R	D	L	P	W	R	R	R	A	E	104
	CAA	GAG	AAA	CGG	GAC	CTA	CCA	TGG	AGA	AGA	CGG	GCA	GAA	312
	D	E	M	D	L	D	R	R	A	Y	A	V	W	117
35	GAT	GAG	ATG	GAC	CTG	GAC	AGG	CGG	GCA	TAT	GCT	GTG	TGG	351
	V	S	E	V	M	L	Q	Q	T	Q	V	A	T	130
	GTC	TCA	GAG	GTC	ATG	CTG	CAG	CAG	ACC	CAG	GTT	GCC	ACT	390
40	V	I	N	Y	Y	T	G	W	M	Q	K	W	P	143
	GTG	ATC	AAC	TAC	TAT	ACC	GGA	TGG	ATG	CAG	AAG	TGG	CCT	429
	T	L	Q	D	L	A	S	A	S	L	E	E	V	156
	ACA	CTG	CAG	GAC	CTG	GCC	AGT	GCT	TCC	CTG	GAG	GAG	GTG	468
45	N	Q	L	W	A	G	L	G	Y	Y	S	R	G	169
	AAT	CAA	CTC	TGG	GCT	GGC	CTG	GGC	TAC	TAT	TCT	CGT	GGC	507

	R	R	L	Q	E	G	A	R	K	V	V	E	E	182
	CGG	CGG	CTG	CAG	GAG	GGA	GCT	CGG	AAG	GTG	GTA	GAG	GAG	546
5	L	G	G	H	M	P	R	T	A	E	T	L	Q	195
	CTA	GGG	GGC	CAC	ATG	CCA	CGT	ACA	GCA	GAG	ACC	CTG	CAG	585
	Q	L	L	P	G	V	G	R	Y	T	A	G	A	208
10	CAG	CTC	CTG	CCT	GGC	GTG	GGG	CGC	TAC	ACA	GCT	GGG	GCC	624
	I	A	S	I	A	F	G	Q	A	T	G	V	V	221
	ATT	GCC	TCT	ATC	GCC	TTT	GGC	CAG	GCA	ACC	GGT	GTG	GTG	663
	D	G	N	V	A	R	V	L	C	R	V	R	A	234
15	GAT	GGC	AAC	GTA	GCA	CGG	GTG	CTG	TGC	CGT	GTC	CGA	GCC	702
	I	G	A	D	P	S	S	T	L	V	S	Q	Q	247
	ATT	GGT	GCT	GAT	CCC	AGC	AGC	ACC	CTT	GTT	TCC	CAG	CAG	741
20	L	W	G	L	A	Q	Q	L	V	D	P	A	R	260
	CTC	TGG	GGT	CTA	GCC	CAG	CAG	CTG	GTG	GAC	CCA	GCC	CGG	780
	P	G	D	F	N	Q	A	A	M	E	L	G	A	273
25	CCA	GGA	GAT	TTC	AAC	CAA	GCA	GCC	ATG	GAG	CTA	GGG	GCC	819
	T	V	C	T	P	Q	R	P	L	C	S	Q	C	286
	ACA	GTG	TGT	ACC	CCA	CAG	CGC	CCA	CTG	TGC	AGC	CAG	TGC	858
	P	V	E	S	L	C	R	A	R	Q	R	V	E	299
30	CCT	GTG	GAG	AGC	CTG	TGC	CGG	GCA	CGC	CAG	AGA	GTG	GAG	897
	Q	E	Q	L	L	A	S	G	S	L	S	G	S	312
	CAG	GAA	CAG	CTC	TTA	GCC	TCA	GGG	AGC	CTG	TCG	GGC	AGT	936
35	P	D	V	E	E	C	A	P	N	T	G	Q	C	325
	CCT	GAC	GTG	GAG	GAG	TGT	GCT	CCC	AAC	ACT	GGA	CAG	TGC	975
	H	L	C	L	P	P	S	E	P	W	D	Q	T	338
40	CAC	CTG	TGC	CTG	CCT	CCC	TCG	GAG	CCC	TGG	GAC	CAG	ACC	1014
	L	G	V	V	N	F	P	R	K	A	S	R	K	351
	CTG	GGA	GTG	GTC	AAC	TTC	CCC	AGA	AAG	GCC	AGC	CGC	AAG	1053
	P	P	R	E	E	S	S	A	T	C	V	L	E	364
45	CCC	CCC	AGG	GAG	GAG	AGC	TCT	GCC	ACC	TGT	GTT	CTG	GAA	1092
	Q	P	G	A	L	G	A	Q	I	L	L	V	Q	377
	CAG	CCT	GGG	GCC	CTT	GGG	GCC	CAA	ATT	CTG	CTG	GTG	CAG	1131

	R	P	N	S	G	L	L	A	G	L	W	E	F	390
	AGG	CCC	AAC	TCA	GGT	CTG	CTG	GCA	GGA	CTG	TGG	GAG	TTC	1170
5	P	S	V	T	W	E	P	S	E	Q	L	Q	R	403
	CCG	TCC	GTG	ACC	TGG	GAG	CCC	TCA	GAG	CAG	CTT	CAG	CGC	1209
	K	A	L	L	Q	E	L	Q	R	W	A	G	P	416
10	AAG	GCC	CTG	CTG	CAG	GAA	CTA	CAG	CGT	TGG	GCT	GGG	CCC	1248
	L	P	A	T	H	L	R	H	L	G	E	V	V	429
	CTC	CCA	GCC	ACG	CAC	CTC	CGG	CAC	CTT	GGG	GAG	GTT	GTC	1287
	H	T	F	S	H	I	K	L	T	Y	Q	V	Y	442
15	CAC	ACC	TTC	TCT	CAC	ATC	AAG	CTG	ACA	TAT	CAA	GTA	TAT	1326
	G	L	A	L	E	G	Q	T	P	V	T	T	V	455
	GGG	CTG	GCC	TTG	GAA	GGG	CAG	ACC	CCA	GTG	ACC	ACC	GTA	1365
20	P	P	G	A	R	W	L	T	Q	E	E	F	H	468
	CCA	CCA	GGT	GCT	CGC	TGG	CTG	ACG	CAG	GAG	GAA	TTT	CAC	1404
	T	A	A	V	S	T	A	M	K	K	V	F	R	481
25	ACC	GCA	GCT	GTT	TCC	ACC	GCC	ATG	AAA	AAG	GTT	TTC	CGT	1443
	V	Y	Q	G	Q	Q	P	G	T	C	M	G	S	494
	GTG	TAT	CAG	GGC	CAA	CAG	CCA	GGG	ACC	TGT	ATG	GGT	TCC	1482
	K	R	S	Q	V	S	S	P	C	S	R	K	K	407
30	AAA	AGG	TCC	CAG	GTG	TCC	TCT	CCG	TGC	AGT	CGG	AAA	AAG	1521
	P	R	M	G	Q	Q	V	L	D	N	F	F	R	520
	CCC	CGC	ATG	GGC	CAG	CAA	GTC	CTG	GAT	AAT	TTC	TTT	CGG	1560
35	S	H	I	S	T	D	A	H	S	L	N	S	A	533
	TCT	CAC	ATC	TCC	ACT	GAT	GCA	CAC	AGC	CTC	AAC	AGT	GCA	1599
	A	Q	*											535
40	GCC	CAG	TGA											1608