Molecular genetic and phenotypic characteristics of patients with adenomatous colorectal polyposis

Submitted for the degree of Doctor of Philosophy
at Cardiff University

Natalie Jones

2008

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Summary

Familial adenomatous polyposis (FAP), *MUTYH*-associated polyposis (MAP) and Hereditary non-polyposis colorectal cancer (HNPCC) are rare inherited colorectal cancer (CRC) predisposition syndromes resulting from mutations in *APC*, *MUTYH* and mismatch repair (MMR) genes, respectively.

Mutational analysis of *APC* and *MUTYH* in 92 colorectal polyposis families living in Wales identified a genetic defect in 85 families (92%). Seventy unrelated cases (70/92, 76%) harboured pathogenic mutations in *APC* and biallelic *MUTYH* mutations were identified in 15 families. No pathogenic mutations could be detected in seven index patients without a dominant family history who had tens to hundreds of colorectal polyps with or without CRC.

A European collaborative project established a cohort of 237 MAP patients. Analysis of *MUTYH* mutations in 182 unrelated MAP index cases highlighted the need for ORF sequencing as 17% (31/182) did not carry either of the common non-Asian mutations (Y176C and G393D). Fifty-eight percent of the MAP patient cohort (138/237) developed CRC and 36% (49/138) had more than one CRC. Retrospective assessment of medical records indicated that MAP patients are not at increased risk of cancers outside the GI tract. The mean age at presentation and CRC diagnosis were found to be inversely correlated with the number of Y176C alleles (p=0.003 & p<0.001, respectively, linear regression).

Analysis of mortality and cancer risk in 350 obligate *MUTYH* mutation heterozygote parents of European MAP patients revealed they were at a two-fold increased risk of CRC compared with the general population.

To investigate whether germline mutations in other base excision repair or DNA damage protection genes predispose to colorectal adenomas, the ORFs of *TDG*, *MPG*, *SMUG1*, *MBD4*, *APE1*, *POLL* and *NUDT5* were sequenced in a cohort of 58 unrelated patients with multiple (>10) colorectal adenomas in whom *APC* or biallelic *MUTYH* mutations had not been demonstrated. No clearly pathogenic mutations were identified, suggesting that defects in these glycosylases are not frequently associated with colorectal polyposis.

Acknowledgements

I would like to thank the following people;

My supervisors Prof. Julian Sampson and Prof. Jerry Cheadle

Our European collaborators Prof. Frederik J. Hes, Dr. Maartje Nielsen, Dr. Stefan Aretz and Dr. Stefanie Vogt

Members of the All Wales Medical Genetics Service for their contributions to the work and use of equipment

All the clinicians involved, particularly Prof. Diana Eccles, Prof. D. Gareth. R. Evans and Prof. Eamonn R. Maher who acted as PIs and their colleagues Emma Tyler, Andrew Shenton and Kris Hepherd for help with data collection.

Dr. Petra Wark, Dr. Valentina Moskvina, Prof. Geraint Williams and Dr. John Gallacher for their expert advice

Dr. Anthony Dallosso, James Colley, Dr. Vikki Humphreys, Julie Maynard, Michelle Williams, Dr. Siân Jones, Cleo Bonnet and Shelley Idziaszczyk for their contributions to the lab work

Julian, for his guidance and support

My friends and family, especially Vikki and Rana, for putting up with me

Siân, for all her help and advice and Jean for incredible patience and understanding

Cancer Research Wales and The Wales Gene Park for supporting this work and all the patients who made it possible.

Abbreviations

A	Adenine
aa	Amino acid
AFAP	Attenuated FAP
ANOVA	Analysis of variance
APC	Adenomatous polyposis coli
APE	Apurinic/apyrimidinic endonuclease
ARMS	Amplification refractory mutation system
AWMGS	All Wales Medical Genetics Service
BCC	Basal cell carcinoma
BER	Base excision repair
bp	Base pair
BS	Brain specific
BSA	Bovine serum albumin
C	Cytosine
CHRPE	Congenital hypertrophy of the retinal pigment epithelium
CI	Confidence interval
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
CRC	Colorectal cancer
CTS	Contents trade secret
DGGE	Denaturing gradient gel electrophoresis
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
ddNTP	Dideoxynucleotide triphosphate
dNTP	Deoxynucleotide triphosphate
DOB	Date of birth
DOD	Date of death
EDTA	Ethylene diamine tetra acetic acid
ESE	Exonic splicing enhancer
FAP	Familial adenomatous polyposis
FGP	Fundic gland polyp
G	Guanine

GDP	Guanosine diphosphate
GEF	Guanine exchange factor
GI	Gastrointestinal
GTP	Guanosine triphosphate
HhH	Helix-hairpin-helix
HMPS	Hereditary mixed polyposis syndrome
HNPCC	Hereditary non-polyposis colorectal cancer
ICD	International classification of diseases
IPTG	Isopropyl-β-D-thiogalactopyranoside
IVS	Intervening sequence
LOH	Loss of heterozygosity
LREC	Local research ethics committee
MAF	Minor allele frequency
MAP	MUTYH-associated polyposis
MBD	Methyl-CpG-binding domain
MCR	Mutation cluster region
MLH	MutL homologue
MLPA	Multiplex ligation-dependent probe amplification
MMR	Mismatch repair
MPG	N-methylpurine-DNA glycosylase
MREC	Multi-centre research ethics committee
mRNA	Messenger RNA
MSH	MutS homologue
MSI	Microsatellite instability
MUTYH	Human MutY homologue (also known as MYH)
NEIL	Nei like glycosylase
NES	Nuclear export signal
NHS	National health service
NLS	Nuclear localisation signal
NMD	Nonsense mediated decay
NTHL	Nth endonuclease III-like
NUDT	Nudix (nucleoside diphosphate linked moiety X)-type motif
OGG	8-oxoG glycosylase

2-OH-A	2-hydroxyadenine
OR	Odds ratio
ORF	Open reading frame
8-oxoG	7,8-dihydro-8-oxoguanine
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PI	Principal investigator
PMS	Post-meiotic segregation increased
Pol	Polymerase
PTT	Protein truncation test
R&D	Research and development
RGS	Regional genetics service
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RPA	Replication protein A
RT	Reverse transcriptase
SDS	Sodium dodecyl sulphate
SIR	Standardized incidence ratio
SMR	Standardized mortality ratio
SMUG	Single-strand-selective monofunctional uracil-DNA glycosylase
SNP	Single nucleotide polymorphism
T	Thymine
TAE	Tris-acetate-EDTA
TDG	Thymine-DNA glycosylase
TGF	Transforming growth factor
UNG	Uracil-DNA glycosylase
UTR	Untranslated region
UV	Ultraviolet
X-gal	5-bromo-4-chloro-3-indoyl-D-galactoside

Contents

List of figures

List of tables

Chapter One

General Introduction

1.1 Colorectal cancer

Colorectal cancer (CRC) is one of the most common cancers worldwide. It is the second most prevalent form of the disease in developed countries (Weitz, J. *et al*, 2005) and in the UK alone, 35 006 new cases were diagnosed in 2003 (Cancer Research UK). Although mortality rates are now decreasing due to improved screening and treatment, it is still estimated that 492 000 CRC deaths occur globally each year (Weitz, J. *et al*, 2005).

Both environmental and genetic factors are important in the development of CRC and either can be the predominant cause of this disease. They also interact with each other, for example a genetic predisposition resulting in an increased susceptibility to particular carcinogens in the diet may induce cancer (Menko, F.H., 1993). The significance of environmental influences on colorectal cancer risk has been shown by studying individuals who have migrated from one country to another. The colorectal cancer risk of Japanese immigrants to Hawaii, for example, was found to change from that of their native country to that of the host country due to environmental differences (Kolonel, L.N. *et al*, 2004). A diet rich in meat and fat and deficient in fibre, folate and calcium can increase the risk of colorectal cancer. Other risk factors include; obesity, an inactive lifestyle, smoking and a high intake of alcohol (Weitz, J. *et al*, 2005). Twin studies have shown that genetic factors also play an important role in causing colorectal cancer. By comparing concordance for colorectal cancer in monozygotic twins with that in dizygotic twins, it has been estimated that genetic factors could account for up to 35% of CRC (Lichtenstein, P. *et al*, 2000). A small proportion of this (approximately 2-6% of all CRC) is the result of known genetic predisposition syndromes (Kemp, Z. *et al*, 2004).

1.2 Colorectal tumourigenesis

Regardless of the causative factor of CRC, the development of malignancy requires multiple mutations. The stages of tumourigenesis have been relatively

well-defined and often occur over several years (Figure 1.1) (Kinzler, K.W. & Vogelstein, B., 1996). Each stage results from mutations in genes involved in cell cycle control and signalling pathways. Different genes involved in the same pathway can be mutated in different colorectal cancers but these mutations are generally mutually exclusive i.e. only one gene in a particular pathway is mutated in any one CRC (Vogelstein, B. & Kinzler, K.W., 2004). An example of this is the transforming growth factor-β (TGF-β) pathway, where mutations in *TGF-βRII* and *SMAD4* are found separately in CRC, but not usually together in the same tumour (Woodford-Richens, K.L *et al*, 2001). The gene mutated in a particular pathway is influenced by the type of genomic instability present in the adenoma or carcinoma.

1.2.1 Genomic instability

Genomic instability plays a key role in tumourigenesis, contributing to the accumulation of mutations in genes involved in colorectal cancer (Grady, W.M. & Markowitz, S.D., 2002). This loss of genomic stability can be in the form of chromosomal instability (CIN) or microsatellite instability (MSI). CIN results in genetic alterations through chromosomal losses and gains whereas MSI causes genetic alterations through defective DNA mismatch repair (MMR) proteins (Söreide, K. *et al*, 2006).

MSI affects short repetitive sequences called microsatellites that are prone to misalignment errors during replication. In normal cells such errors are repaired by the MMR proteins. Defects in these proteins result in a germline microsatellite allele undergoing a somatic change in length by gaining or losing repeated units. MSI influences which genes are mutated in colorectal tumours with this form of genomic instability. Genes containing coding microsatellites are susceptible to such errors, resulting in frameshift mutations that can contribute to tumourigenesis (Söreide, K *et al*, 2006). The *TGF-βRII* gene is mutated in colorectal tumours with MSI and frameshift mutations occur at a polyadenine $(A)_{10}$ and a $(GT)_3$ tract within the coding region of this gene (Markowitz, S. *et al*, 1995).

CIN induces larger changes in DNA than MSI by increasing the rate of

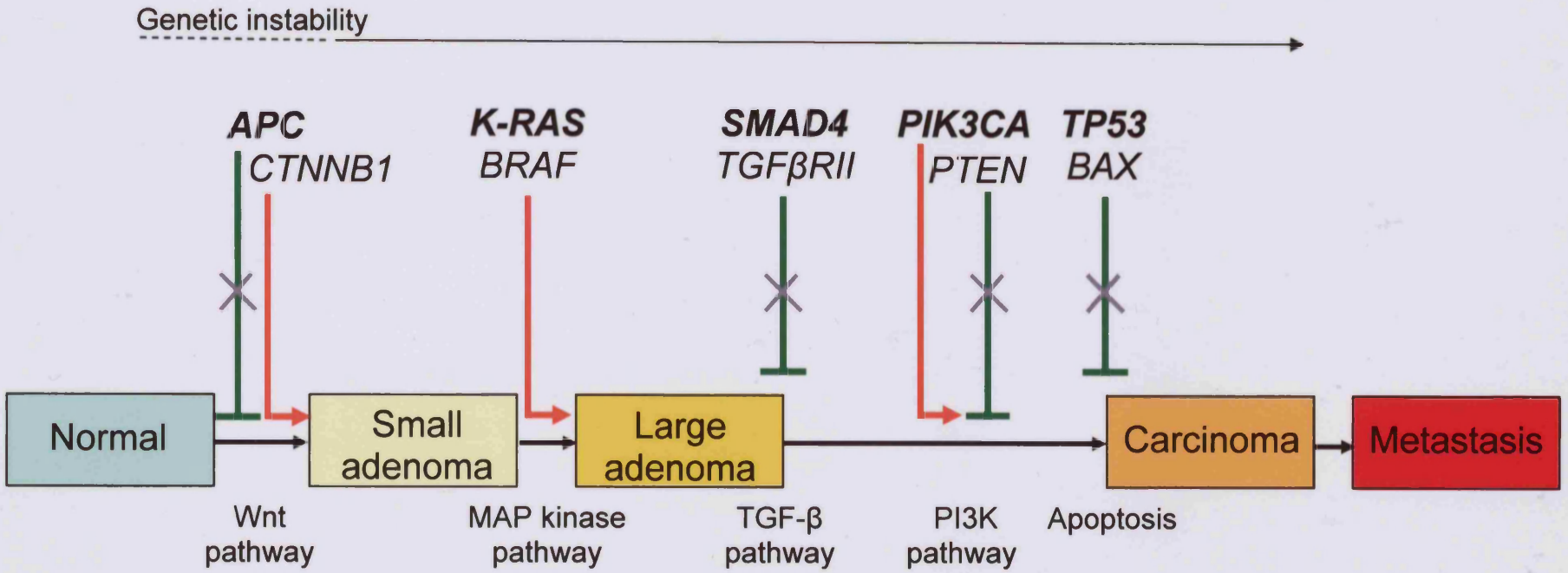


Figure 1.1 Colorectal tumourigenesis. The stages of tumour development in the colon and genes which, when mutated, drive this process. Tumour suppressors and oncogenes are attached to green and red lines respectively. Genes shown in bold are generally mutated in tumours with CIN and those not in bold are generally mutated in MSI tumours. The pairs of genes shown which act at the same point in the process are members of the same pathway, which is indicated below each pair. Genetic instability plays a key role in tumourigenesis, contributing to the accumulation of mutations in genes involved in colorectal cancer. Adapted from Kinzler, K.W. & Vogelstein, B., 1996.

chromosomal gains or losses (Rajagopalan, H. *et al*, 2003). Such alterations result from mitotic recombination or aberrant segregation of chromosomes at mitosis and lead to the aneuploidy seen in cancers with CIN (Kinzler, K.W. & Vogelstein, B., 1996). The chromosomal changes are not random but are driven by natural selection and chromosomal arms that are frequently lost in tumours with CIN are the loci of genes important in tumourigenesis. Allelic loss (loss of heterozygosity, LOH) at 17p, 18q and 5q often occurs and these are the loci of *p53*, *SMAD4* and *APC* respectively (Terdiman, J.P., 2000). Deletion at 1p and 8p is also common although the genes important for tumourigenesis at these loci have not yet been proven. The molecular mechanism responsible for CIN is not fully understood but this instability is thought to occur at an early stage of colorectal tumourigenesis (Shih, I-M. *et al*, 2001). Mutations in genes involved in the accurate mitotic segregation of chromosomes, including *hBUB1* (Cahill, D.P. *et al*, 1998), *APC* (Fodde, R. *et al*, 2001a) and *hCDC4* (Rajagopalan, H. *et al*, 2004), has been suggested as a possible cause of CIN.

The majority of CRCs (approximately 85%) show CIN and have gross chromosomal alterations rather than the subtle changes found in tumours with MSI (Kinzler, K.W. & Vogelstein, B., 1996). In a small proportion of cases however, neither CIN nor MSI is evident (Goel, A., et al, 2003). Epigenetic instability in the form of CpG island methylator phenotype (CIMP) has been identified in CRC. CIMP results in the aberrant hypermethylation of CpG islands which are often found in the promoter regions of genes. This instability promotes tumourigenesis by transcriptionally silencing genes that prevent tumour formation (Toyota, M. *et al*, 1999) such as CDKN2A which encodes p16, a protein involved in cell cycle control (Herman, J.G. *et al*, 1995). Genetic and epigenetic instability are not mutually exclusive; colorectal tumours with CIMP and CIN or (more frequently) MSI have been identified (Samowitz, W.S. *et al*, 2005). MSI in some CRC is in fact the result of epigenetic instability as promoter hypermethylation of *hMLH1* (which encodes a MMR protein) silences this gene leading to defective MMR (Herman, J.G. *et al*, 1998). Genomic instability in some rare CRC is a result of defective base excision repair (BER) due to mutations in the *MUTYH* gene. Following DNA replication, unrepaired

mismatches become G:C→T:A transversions and are found in genes including *APC* and k-*ras* (Sampson, J.R. *et al*, 2005).

1.2.2 Cancer genes

Genes that are mutated in CRC and directly contribute to the adenoma to carcinoma sequence fall into two classes; oncogenes and tumour suppressor genes. Mutations in oncogenes cause them to become constitutively active or active when the wild-type gene would not be. Such activation can be caused by chromosomal translocations, gene amplifications or subtle mutations within the gene that affect residues important for the activity of the protein (Vogelstein, B. & Kinzler, K.W., 2004). An oncogene often mutated in colorectal cancer is the *K-RAS* gene. The K-ras protein is involved in signal transduction and couples growth factors to the mitogen-activated protein kinase (MAPK) cascade which results in expression of genes involved in cell proliferation. K-ras is a monomeric GTPase which is inactive when bound to GDP but when GTP is bound, K-ras is activated and the MAPK cascade is stimulated. Normally, the GTPase activity of K-ras would then hydrolyse the bound GTP thus inactivating itself. The most common mutations in the *K-RAS* gene affect codons 12, 13 and 61 all of which result in disrupted GTPase activity, allowing K-ras to remain activated and stimulate cell proliferation (Grady, W.M. & Markowitz, S.D., 2002). Only one allele of an oncogene needs to be somatically mutated for it to have a tumourigenic effect. Other oncogenes involved in CRC tumourigenesis include *CTNNB1* (Morin, P.J. *et al*, 1997), *BRAF* (Davies, H. *et al*, 2002) and *PIK3CA* (Samuels, Y. *et al*, 2005)

Mutations in tumour suppressor genes cause inactivation of the gene product. This loss of function can result from mutations that cause a truncated protein or lead to nonsense mediated decay (NMD), missense mutations at functionally important residues, insertions, deletions or gene silencing by promoter hypermethylation (Vogelstein, B. & Kinzler, K.W., 2004). A tumour suppressor gene mutated in the majority of colorectal cancers is the *APC* gene. The APC protein has multiple cellular functions including an inhibitory role preventing over-activation of the Wnt signalling pathway. This pathway stimulates cell growth when activated. *APC* mutations are often protein-truncating mutations

and chromosomal deletions. These inactivate the APC protein so it can no longer regulate the Wnt signalling pathway leading genes involved in cell proliferation to be expressed (Grady, W.M. & Markowitz, S.D., 2002). Other tumour suppressor genes involved in CRC tumourigenesis include *TP53* (Hollstein, M. *et al*, 1991), *SMAD4* (Woodford-Richens, K.L *et al*, 2001), *TGFβRII* (Markowitz, S. *et al*, 1995), *PTEN* (Nassif, N.T. *et al*, 2004) and *BAX* (Rampino, N. *et al*, 1997). In general, both alleles of a tumour suppressor gene need to be mutated to have a tumourigenic effect. The two inactivating mutations occur somatically in sporadic CRC but in dominantly inherited CRC predisposition syndromes one mutation is found in the germline and the second mutation occurs somatically. An example of this is found in the *APC* gene which is mutated somatically in sporadic CRC while germline mutations cause Familial Adenomatous Polyposis (FAP). This disease predisposes to CRC since a second somatic mutation in the *APC* gene can initiate adenoma development (Fearnhead, N.S. *et al*, 2001). This is in accordance with Knudson's two-hit hypothesis that inherited and sporadic forms of the same type of cancer require two mutations in the same tumour suppressor gene; both are somatic in sporadic cancer but in dominantly inherited cancer predisposition syndromes the first is germline and the second is somatic (Knudson, A.G., 1996).

The majority of CRC cases are sporadic but inherited factors can contribute to up to 35% of CRC (Lichtenstein, P. *et al*, 2000), a small proportion of which (approximately 2-6% of all CRC) are accounted for by known genetic predisposition syndromes (Kemp, Z. *et al*, 2004). The study of inherited CRC syndromes can help elucidate mechanisms of colorectal tumourigenesis in both sporadic and familial CRC as many of the genetic changes required for tumour formation are common to both.

1.3 Inherited CRC Syndromes

1.3.1 Familial Adenomatous Polyposis (FAP)

1.3.1.1 Clinical presentation

FAP is characterized by the appearance of hundreds to thousands of colorectal adenomas, usually by the second or third decade of life. If left untreated, one or

more of these polyps inevitably progresses to cancer, usually by early to mid adult life and with CRC risk related to polyp number (Fearnhead, N.S. *et al*, 2001; Strate, L.L. & Syngal, S., 2005). FAP is a rare disease, affecting approximately 1 in 10 000 of the population. Less than 1% of all colorectal cancers are due to FAP and this disease affects both men and women equally (Strate, L.L. & Syngal, S., 2005).

Extracolonic features are also associated with FAP. Congenital hypertrophy of the retinal pigment epithelium (CHRPE) affects about 60% of FAP families. It has no detrimental effect on sight but can be used to identify family members who are at risk long before polyps appear. FAP patients frequently develop upper gastrointestinal tumours and for those patients who have undergone prophylactic colectomies, periampullary carcinoma in the duodenum is an increasingly common cause of death (Fearnhead, N.S. *et al*, 2001). Desmoid tumours are also a major cause of morbidity in FAP patients and often occur after surgery. The presence of multiple desmoid tumours without colonic polyposis is attributed to the FAP variant hereditary desmoid disease (Galiatsatos, P. & Foulkes, W.D., 2006). Gardner's syndrome is a largely redundant term describing a variant of FAP in which colonic polyps are associated with extracolonic manifestations that include epidermoid skin cysts and benign osteoid tumours of the mandible and long bones (Fearnhead, N.S. *et al*, 2001). Colorectal polyposis and a primary central nervous system tumour (most commonly medulloblastoma) are characteristics of the variant Turcot syndrome (Galiatsatos, P. & Foulkes, W.D., 2006). Other infrequent extracolonic malignancies that have been associated with FAP include thyroid cancer, adrenocortical tumours and hepatoblastoma (Fearnhead, N.S. *et al*, 2001).

Attenuated familial adenomatous polyposis (AFAP) is a phenotypically distinct, milder variant of FAP traditionally characterized by the presence of fewer than 100 colorectal adenomas. The mean age of CRC onset is delayed by ~15 years and CRC does not inevitably develop in AFAP patients although the risk of CRC remains high (69% by the age of 80 years) (Galiatsatos, P. & Foulkes, W.D., 2006). As in classical FAP, gastric fundic gland polyps and duodenal

adenomas are frequently present, but other extracolonic manifestations are rare (Knudsen, A.L. *et al*, 2003).

1.3.1.2 Genetics

FAP is an autosomal dominant disorder caused by specific classes of germline mutations in the adenomatous polyposis coli (*APC*) gene. These mutations attain virtually 100% penetrance (Fearnhead, N.S. *et al*, 2001). At least 10% of FAP cases appear to result from new *APC* mutations so the incidence of this disease is maintained despite its significant selective disadvantage.

1.3.1.3 The APC gene

Groden *et al* (1991) and Kinzler *et al* (1991) identified and characterized the *APC* gene found on chromosome 5q21. The 8.5kb transcript is encoded by 21 exons, the largest of which (exon 15) makes up more than three-quarters of the *APC* coding sequence (Fearnhead, N.S. *et al*, 2001). Many of the exons are subject to alternative splicing, giving rise to different mRNA transcripts (Figure 1.2). Exons 0.3, 0.1, 0.2 and BS are located upstream of the initiation codon in exon 1 (Horii, A. *et al*, 1993; Thliveris, A. *et al*, 1994). Despite exons 0.2, 0.3 and BS containing in-frame initiation codons, these open reading frames (ORFs) are not translated when found in transcripts that contain exon 1 because of a stop codon found 6bp upstream of the exon 1 initiation codon (Santoro, I.M. & Groden, J., 1997). In some alternative splice forms however, exon 1 is excluded and exons 0.3 or BS are included resulting in transcripts that are translated into the APC isoforms 0.3 APC and BS APC, respectively (Santoro, I.M. & Groden, J., 1997; Carson, D.J. *et al*, 2004). Transcripts excluding exon 1 but containing one or more of the upstream exons are expressed at higher levels in terminally differentiated tissues such as brain and heart than other tissues, suggesting these transcripts are associated with differentiation (Santoro, I.M. & Groden, J., 1997). No transcripts have been identified containing exon 0.3 in combination with either exon 0.1 or 0.2, indicating the presence of two independent promoters (Thliveris, A. *et al*, 1994). Promoter 1A is most commonly active (Fearnhead, N.S. *et al*, 2001) and lies upstream of exon 0.3 (Thliveris, A. *et al*, 1994), while promoter 1B lies upstream of exon 0.1 (Horii, A. *et al*, 1993).

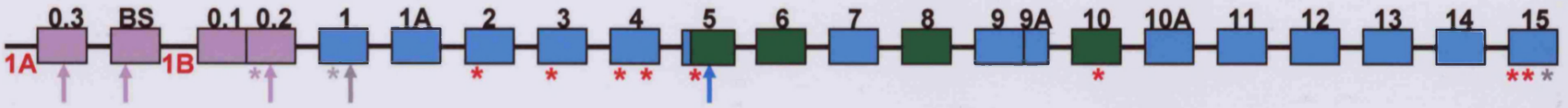


Figure 1.2 Alternative splicing of the *APC* gene. A schematic diagram of the *APC* gene showing its 22 exons and which of these undergo alternative splicing (not to scale). ■, alternatively spliced exon 5' to exon 1; ■, alternatively spliced exon; ■, constitutively expressed exon. Exon 1 encodes a translational start site (↑) and 5' exons 0.3, BS and 0.2 also contain initiation codons (↑) in-frame with exon 2. The 5' exons are only translated in transcripts lacking exon 1 due to a stop codon (*) in exon 1. Promoters 1A and 1B are located upstream of exons 0.3 and 0.1 respectively. Alternative splicing can alter the *APC* reading frame resulting in stop codons (*) but exon 5 encodes an alternative translational start site (↑).

Several exons downstream of exon 1 also undergo alternative splicing including exons 1-4 (Samowitz, W.S. *et al*, 1995; De Rosa, M. *et al*, 2007), 7 (Oshima, M. *et al*, 1993), 9 (Groden, J. *et al*, 1991; De Rosa, M. *et al*, 2007), 9A (De Rosa, M. *et al*, 2007), 10A (Xia, L. *et al*, 1995) and 11-15 (Sulekova, Z. *et al*, 1995; Horii, A *et al*, 1993) (Figure 1.2). Alternatively spliced transcripts which are translated from exon 1 result in a stop codon before codon 184 and could still result in a protein with an intact C-terminus through the use of an alternative translation start site at codon 184 in exon 5 (Heppner Goss, K. *et al*, 2002). Alternative splicing of some exons results in transcripts with an in-frame insertion (exon 10A, Xia, L. *et al*, 1995) or deletion (exons 7 or 9, Oshima, M. *et al*, 1993; Groden, J. *et al*, 1991) but several alternatively spliced transcripts contain premature termination codons (Sulekova, Z. *et al*, 1995) and may be subject to NMD (De Rosa, M. *et al*, 2007). *APC* can also be spliced to an adjacent gene, *SRP19*, which encodes a signal recognition particle (Horii, A. *et al*, 1993). The most common and ubiquitously expressed *APC* transcript includes exons 1, 2-8, 9, 9A, 10, 11-15 and encodes a 2843 amino acid protein.

1.3.1.4 The APC protein

APC is a large protein (312kDa) with several functional domains which allow it to interact with various other proteins (Figure 1.3) (Fodde, R. *et al*, 2001b). The oligomerisation domain of APC is found at the N-terminus and contains heptad repeats that can facilitate homodimerization (Su, L-K. et al, 1993a). The N-terminal and C-terminal regions of APC have recently been found to interact, suggesting intramolecular interactions may also occur (Li, Z. & Näthke, I.S., 2005). A set of seven armadillo repeats are found in the N-terminal portion of APC and this highly conserved region can bind to protein phosphatase 2A (PP2A), APC-stimulated guanine nucleotide exchange factor (Asef) and KAP3 (Näthke, I.S., 2004). The central region of APC contains three 15-amino acid (aa) repeats followed by seven 20-aa repeats, both of which can bind to β-catenin (Rubinfeld, B. *et al*, 1993; Su, L-K. *et al*, 1993b). The 15-aa repeats can also bind to α-catenin (Su, L-K. *et al*, 1993b) and C-terminal binding protein (CtBP) (Hamada, F. & Bienz, M., 2004). The 20-aa repeats contain a TPXXFSXXXSL motif (Groden, J. *et al*, 1991) and can bind glycogen synthase kinase 3β (GSK3β) (Rubinfeld, B. *et al*, 1996). Interspersed between the 20-aa

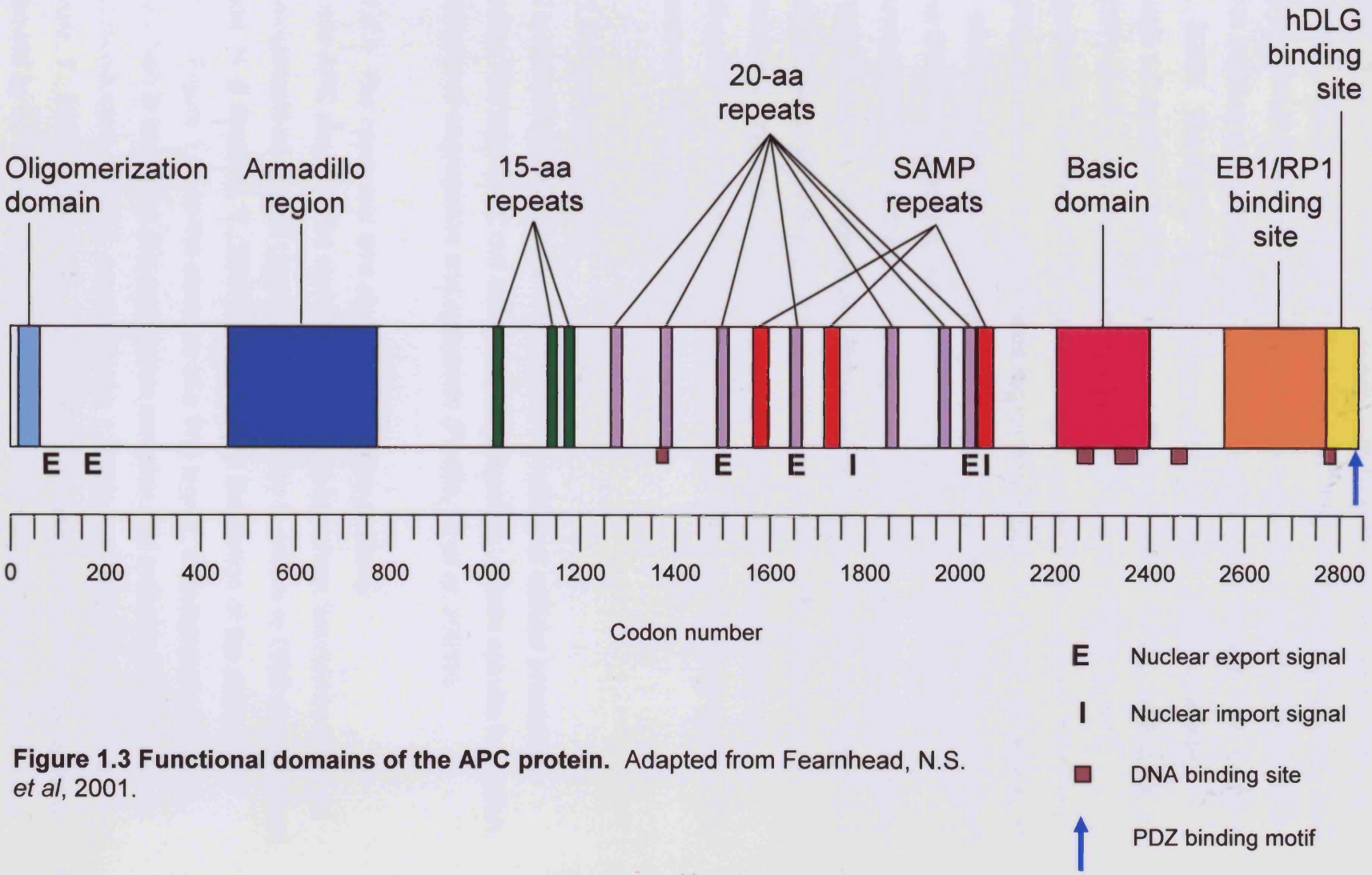


Figure 1.3 Functional domains of the APC protein. Adapted from Fearnhead, N.S. *et al*, 2001.

repeats are 3 stretches of 31-32 amino acids, each of which contains the sequence SAMP. These SAMP repeats can bind to axin and its homologue conductin (also known as axin2) (Behrens, J. *et al*, 1998; Hart, M.J. *et al*, 1998). The C-terminal region of APC contains a basic domain (Groden, J. *et al*, 1991) rich in arginine and lysine which can bind microtubules (Hanson, C.A. & Miller, J.R., 2005). The EB/RP family members EB1 and RP1 also bind to APC through a C-terminal domain (Juwana, J-P. *et al*, 1999). A PDZ binding motif (S/TXV) at the C-terminus allows binding of hDLG and PTP-BL to APC (Fearnhead, N.S *et al*, 2001). APC has five nuclear export signals (NESs) two of which are found in the N-terminal region and contain the motif LXXXLXXLXL. The other three signals are found in the third, fourth and seventh 20-aa repeats in the C-terminal portion of APC and contain the motif LXXLXL/I/M/V (Fearnhead, N.S *et al*, 2001). The two APC nuclear localization signals (NLSs) are found between the fourth and fifth 20-aa repeats and within the third SAMP repeat (Fearnhead, N.S *et al*, 2001). APC can bind to DNA through sites containing 3-5 repeats of the sequence S/TPXX. These can be found overlapping the second 20-aa repeat, within the basic domain and close to the C-terminus (Deka, J. *et al*, 1999).

1.3.1.5 Functions of APC

APC is a multifunctional protein involved in a variety of cellular processes including Wnt signalling, cell adhesion, cell migration, mitotic spindle formation, chromosomal segregation and apoptosis (Fodde, R. *et al*, 2001b).

1.3.1.5.1 The canonical Wnt signal transduction pathway

The role APC plays in Wnt signalling is its most important function in terms of tumourigenesis as loss of control of this pathway is critical to CRC development (Fodde, R. & Brabletz, T., 2007). A schematic illustration of the pathway is shown in Figure 1.4. In the absence of a Wnt signal, a multiprotein complex forms which is known as the destruction complex and includes APC, Axin1 or Axin2 (conductin), GSK3β, casein kinase 1 (CK1) and PP2A (Fodde, R. & Brabletz, T., 2007; Näthke, I.S., 2004). Free cytoplasmic β-catenin is recruited and bound by APC and axin, then primed by CK1 for GSK3β serine/threonine phosphorylation (Reya, T. & Clevers, H., 2005; Polakis, P., 2002). Within the

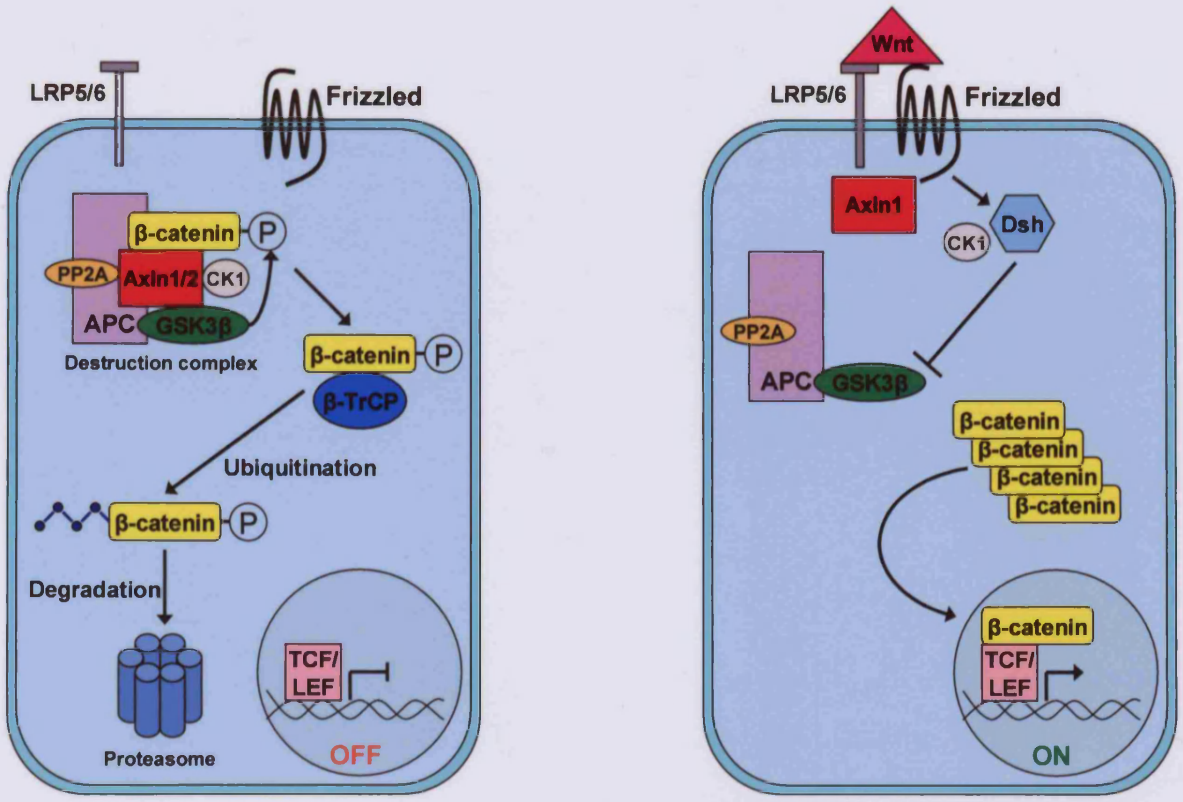


Figure 1.4 The canonical Wnt signalling pathway. In the absence of a Wnt signal, cytoplasmic β-catenin is bound by the destruction complex and phosphorylated by GSK3β. It is then recognised and targeted for proteasomal degradation. In the presence of a Wnt signal, β-catenin is not bound and GSK3β kinase activity is inhibited. β-catenin accumulates then translocates to the nucleus where it binds to transcription factors and activates transcription of Wnt target genes. Adapted from Reya, T. & Clevers, H., 2005.

complex, GSK3β phosphorylates APC, Axin and β-catenin (Rubinfeld, B. *et al*, 1996; Ikeda, S. *et al*, 1998a). The interactions between these proteins are affected by, and can affect, their phosphorylation. For example, phosphorylated APC binds to β-catenin (Rubinfeld, B. *et al*, 1996) and β-catenin enhances the phosphorylation of APC in the destruction complex (Ikeda, S. *et al*, 2000). Phosphorylation of APC within the complex can also be regulated by PP2A (Ikeda, S. *et al*, 2000) and CK1 (Rubinfeld, B. *et al*, 2001) which can bind to axin and dephosphorylate or phosphorylate APC respectively. Once β-catenin is phosphorylated by GSK3β it is recognised and ubiquitinated by a β-TrCP-containing E3 ubiquitin ligase. This targets β-catenin for proteasomal degradation (Fodde, R. *et al*, 2001b). Although both the 15-aa and 20-aa repeats of APC can bind β-catenin, it is the 20-aa repeats which are necessary for the downregulation of β-catenin (Rubinfeld, B. *et al*, 1997).

In the presence of a Wnt signal, the Frizzled and low-density lipoprotein receptor-related proteins (LRP) receptors cooperate to bind the Wnt ligand. This leads to inhibition of GSK3β activity in the destruction complex through the Dishevelled (Dsh) protein and/or the interaction between axin and LRP5/6 (Reya, T. & Clevers, H., 2005). The action of Dsh may be due to its phosphorylation by CK1 which promotes GBP (GSK3β binding protein) binding. Dsh activates GBP which can then inhibit GSK3β activity (Polakis, P., 2002; Fodde, R. *et al*, 2001b). Axin and APC are dephosphorylated as a result of their interaction with Dsh and PP2A, respectively. APC is dephosphorylated at GSK3β-phosphorylated sites and axin is less able to complex with APC and β-catenin (Fearnhead, N.S. *et al*, 2001). β-catenin cannot be phosphorylated by GSK3β, resulting in its stabilization and cytoplasmic accumulation. β-catenin then translocates to the nucleus where it binds to TCF/LEF transcription factors leading to the expression of target genes. In the absence of a Wnt signal, the TCF/LEF family of proteins are complexed with co-repressors so transcription of target genes is inhibited (Fodde, R. *et al*, 2001b). The main TCF family member expressed in intestinal epithelium is TCF-4 which upregulates genes including those involved in proliferation such as *c-myc* (Fearnhead, N.S. *et al*, 2001). Other Wnt target genes have roles in differentiation (e.g. ephrins) and adhesion (e.g matrilysin) (Schneikert, J. & Behrens, J., 2007). APC can also

sequester β-catenin in the nucleus by cooperating with CtBP therefore regulating the amount of nuclear β-catenin which is free to bind TCF and activate transcription (Hamada, F. & Bienz, M., 2004).

If *APC* is mutated, β-catenin is inappropriately stabilized in the absence of a Wnt signal. β-catenin can then translocate to the nucleus and constitutively activate transcription of the Wnt target genes (Fodde, R. & Brabletz, T., 2007). The sequestration of nuclear β-catenin by APC could also be reduced if *APC* is mutated, which would allow more of the nuclear β-catenin to bind TCF and activate transcription (Bienz, M. & Hamada, F., 2004).

1.3.1.5.2 Cell adhesion

In addition to its role in Wnt signalling, β-catenin is a component of adherens junctions which mediate epithelial cell adhesion. Newly synthesized β-catenin binds to E-cadherin and α-catenin, which is bound to the actin cytoskeleton. These interactions can be disrupted by tyrosine phosphorylation of β-catenin which also promotes transcription of Wnt target genes (Schneikert, J. & Behrens, J., 2007). APC promotes E-cadherin mediated cell adhesion but it is not clear exactly how this occurs at present (Hanson, C.A. & Miller, J.R., 2005). The interaction between APC and protein tyrosine phosphatase PTP-BL could promote adhesion through tyrosine dephosphorylation of β-catenin thus increasing the amount of β-catenin able to bind to E-cadherin (Erdmann, K.S. *et al*, 2000). It has been suggested that APC could have a role in the regulation of catenin-cadherin exchange at the plasma membrane by facilitating the exchange of newly synthesized β-catenin between its free and E-cadherin bound states (Bienz, M. & Hamada, F., 2004; Hanson, C.A. & Miller, J.R., 2005). APC also has a role in the regulation of cell adhesion through its interaction with Asef, a Rac-specific guanine nucleotide exchange factor (GEF). Mutant, but not wild-type, APC activates Asef which decreases E-cadherin mediated cell adhesion (Kawasaki,Y. *et al*, 2003).

1.3.5.1.3 Cell migration

APC is needed at the leading edge of cells for directed cell migration. Overexpression of APC in the intestine results in aberrant cell migration and

mutant APC causes decreased cell migration (Schneikert, J. & Behrens, J., 2007; Mahmoud, N.N. *et al*, 1997). APC is found at the plus ends of microtubules and regulates their stability by binding to unassembled tubulin and promoting microtubule assembly. APC is targeted to the microtubule tips by EB1 binding and the interaction between these proteins and mDia promotes cell migration by mediating microtubule stabilization and capture at the cell membrane (Hanson, C.A. & Miller, J.R., 2005). The movement of APC along microtubules to their plus ends is thought to involve interaction between APC and kinesin superfamily-associated protein 3 (KAP3) (Jimbo, T. *et al*, 2002).

APC also interacts with IQ motif containing GTPase activating protein 1 (IQGAP1) at the leading edge and this is thought to stabilize captured microtubules. IQGAP1 may link APC to the actin cytoskeleton as it can bind F-actin to promote actin polymerization (Hanson, C.A. & Miller, J.R., 2005).

As discussed above, APC interacts with Asef which is also involved in reorganisation of the actin cytoskeleton during lamellipodia formation and membrane ruffling (Hanson, C.A. & Miller, J.R., 2005). Mutant APC activates Asef and promotes cell migration (Kawasaki,Y. *et al*, 2003).

1.3.5.1.4 Mitotic spindle formation and chromosome segregation

Consistent with its localization at the plus ends of microtubules, APC is found at kinetochores during mitosis (Fodde, R *et al*, 2001a). Its interaction with EB1 at the plus ends of microtubules may regulate the association of microtubule tips with kinetochores and the cell cortex. Mutant APC acts in a dominant manner to interfere with the attachment of microtubule plus ends to kinetochores and the cell cortex, resulting in increased chromosome segregation defects (Green, R.A. & Kaplan, K.B., 2003). APC is also found at centrosomes and on the mitotic spindle (Näthke, I.S., 2004) which is anchored to the cell cortex at adherens junctions (Schneikert, J. & Behrens, J., 2007).

1.3.5.1.5 Apoptosis and cell cycle regulation

Mutant APC expression is associated with a decrease in apoptosis (Venesio, T. *et al*, 2003a) and inducible expression of wild type APC in cells containing only

mutant APC increases apoptosis (Morin, P.J. *et al*, 1996). This suggests a role for APC in the regulation of apoptosis. APC expression can upregulate the pro-apoptotic protein Clusterin and downregulate the anti-apoptotic factor Survivin. It is not clear how APC regulates apoptosis, but it may be through the canonical Wnt signalling pathway, as appears to be the case for Survivin regulation (Hanson, C.A. & Miller, J.R., 2005). There is also evidence, however, that APC can regulate apoptosis independently of the transcriptional activation of Wnt target genes (Venesio, T. *et al*, 2003a).

APC can also downregulate cell growth through its interaction with hDLG. These proteins form a complex that can regulate the G_1/S phase transition independently of the effect of β-catenin on the cell cycle. Mutant APC results in a weaker cell cycle blocking activity (Fearnhead, N.S. *et al*, 2001; Hanson, C.A. & Miller, J.R., 2005).

Changes in a number of processes involved in normal gut maintenance have to occur for tumours to develop and APC has roles in several of these processes (Näthke, I., 2004). The many functions of this tumour suppressor support its role as a gatekeeper of colonic epithelial cell proliferation (Kinzler, K.W. & Vogelstein, B., 1996).

1.3.1.6 Molecular genetics

1.3.1.6.1 Germline mutations

FAP is the result of germline mutations in the *APC* gene, most of which (>95%) are nonsense or frameshift mutations that result in a truncated protein. As shown in Figure 1.5, the vast majority of mutations are located in the 5' half of the gene, with two notable hotspots at codons 1061 and 1309 (Thierry Soussi database, personal communication). These commonly mutated codons are part of short repeat sequences which may increase their mutability (Segditsas, S. & Tomlinson, I., 2006). Large deletions encompassing the whole *APC* gene or exons have also been identified and are estimated to contribute to approximately 5% of FAP cases (Sieber, O. *et al*, 2002; Michils, G. *et al*, 2005). There still remain, however, 20% of FAP patients in whom no germline *APC* mutation can be identified by conventional screening methods such as the

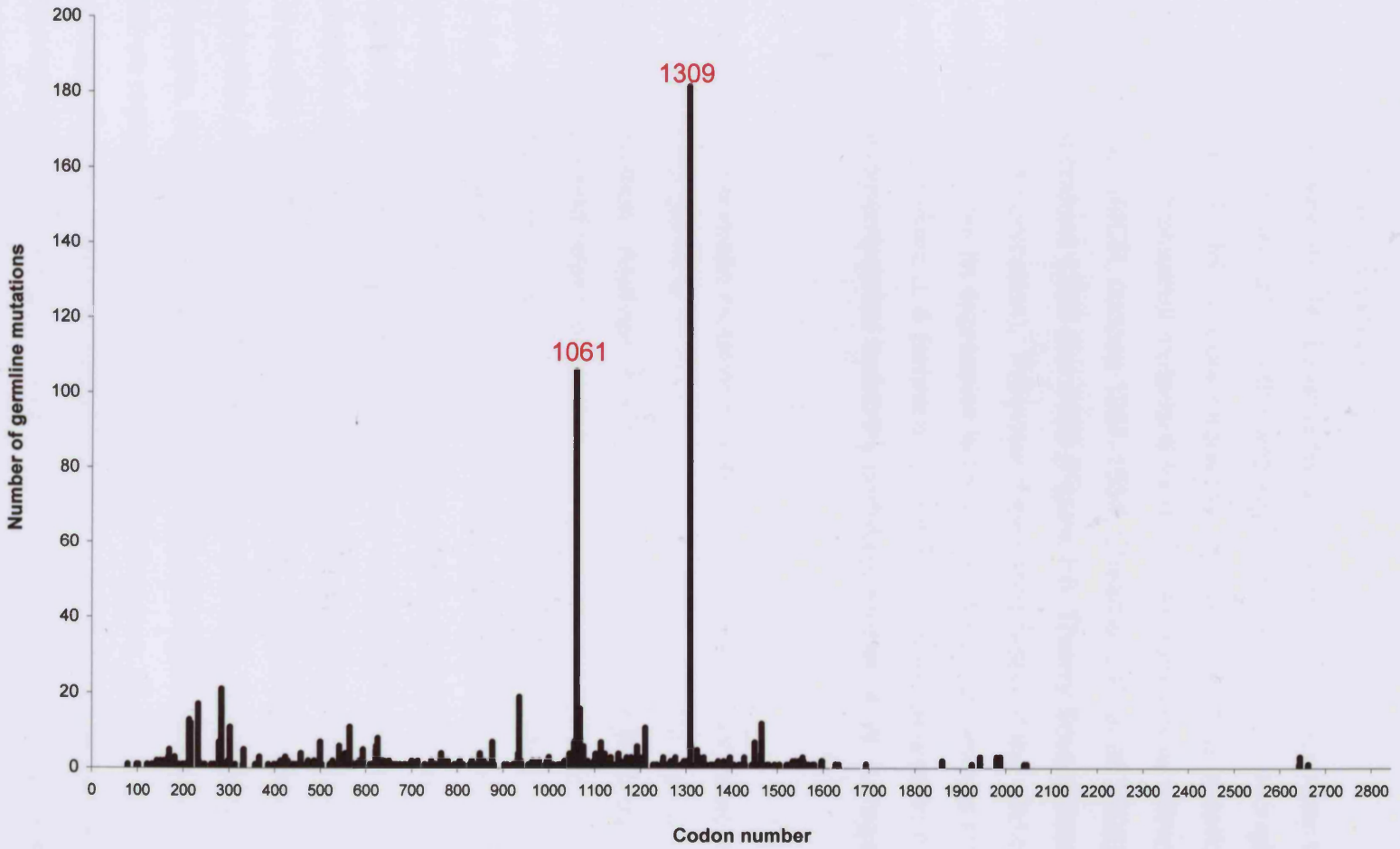


Figure 1.5 Distribution of germline *APC* mutations in FAP patients. Germline mutations cluster in the 5' half of the gene with hotspots at codons 1061 and 1309. Thierry Soussi database, personal communication.

protein truncation test (PTT) or direct DNA sequencing of the coding region (de la Chapelle, A., 2004).

1.3.1.6.2 Somatic mutations

Tumour formation in FAP patients (and in the majority of sporadic CRC cases) requires somatic mutation of the wild-type *APC* allele either through loss of heterozygosity (LOH) or more frequently, protein truncating mutations. Most nonsense and frameshift mutations found in FAP tumours lie within the mutation cluster region (MCR; codons 1281-1556; Cheadle, J.P. *et al*, 2002), with hotspots at codons 1309 and 1450 (Figure 1.6, Thierry Soussi database, personal communication). Promoter hypermethylation of the wild-type *APC* allele that silences its expression is very rarely found in tumours from FAP patients (Schneikert, J. & Behrens, J., 2007). Tumourigenesis in FAP occurs through the chromosomal instability pathway (Kinzler, K.W. & Vogelstein, B., 1996)

Germline and somatic mutations of *APC* in tumours are not independent; the position of the germline mutation appears to determine the type and location of somatic mutation. Analysis of *APC* mutations in tumours from classical FAP patients revealed relationships between germline and somatic mutations and a general pattern emerged (Figure 1.7) (Lamlum, H. *et al*, 1999; Albuquerque, C. *et al*, 2002; Crabtree, M. *et al*, 2003). If the germline mutation results in a protein lacking all its 20-aa repeats, the somatic mutation is usually a truncating mutation that leads to a protein with one or two 20-aa repeats. If the germline mutation lies between the first and second 20-aa repeats, the somatic mutation is usually loss of heterozygosity as a result of mitotic recombination. If the germline mutation lies between the second and third, or third and fourth 20-aa repeats, the somatic mutation is usually a truncating mutation before the first 20-aa repeat (Albuquerque, C. *et al*, 2002; Crabtree, M. *et al*, 2003). It is thought that the reason for this selection lies in the ability of APC to regulate the canonical Wnt signalling pathway. If both *APC* alleles encoded proteins that lacked all 20-aa repeats, neither of these proteins would be able to bind and downregulate β-catenin in response to a Wnt signal. β-catenin would be free to accumulate to high levels in the nucleus and although this may seem

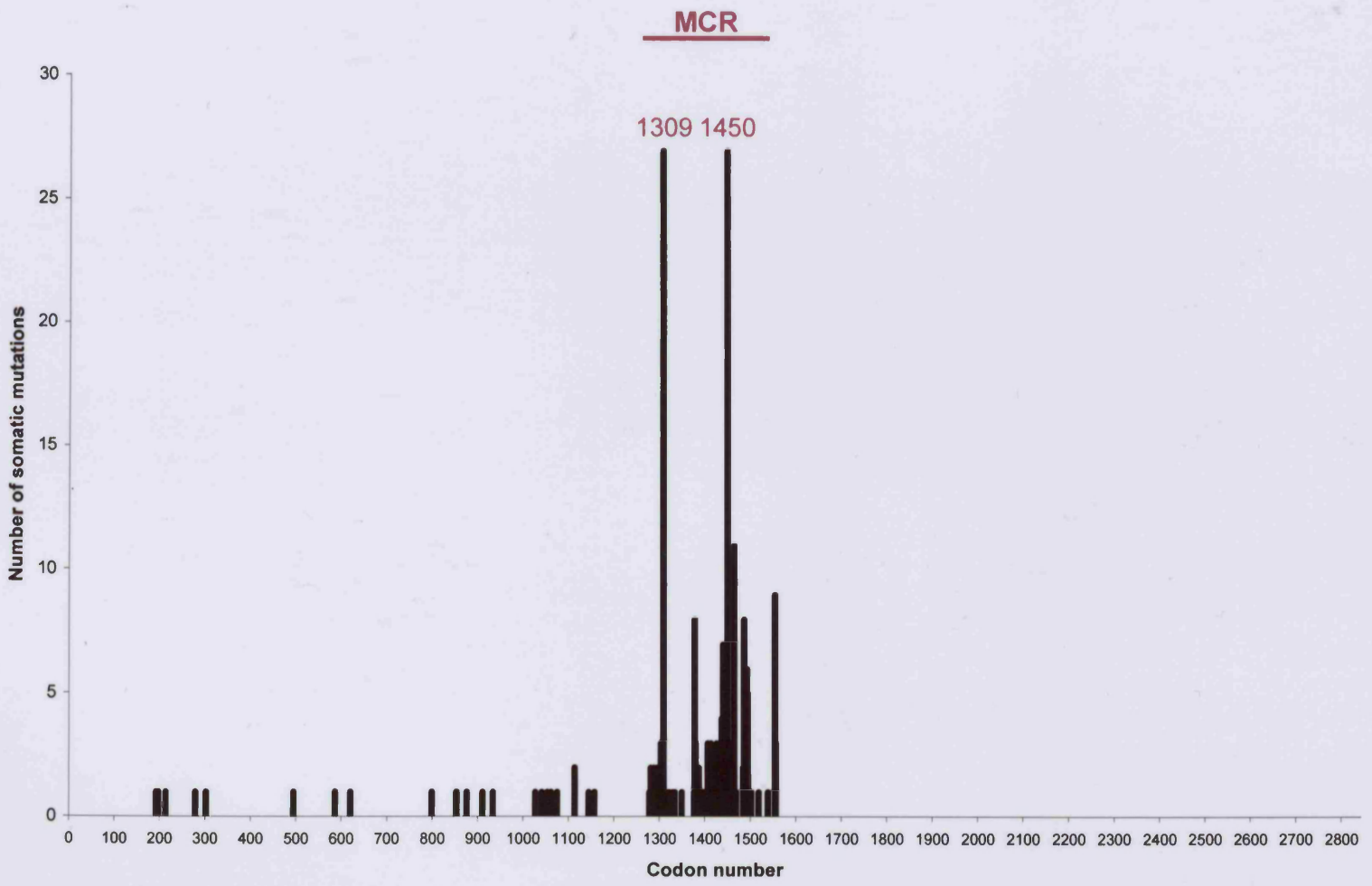


Figure 1.6 Distribution of somatic *APC* mutations in colorectal adenomas or CRC in FAP patients. Somatic mutations are mainly found within the mutation cluster region (MCR), codons 1281-1556 with hotspots at codons 1309 and 1450. Thierry Soussi database, personal communication.

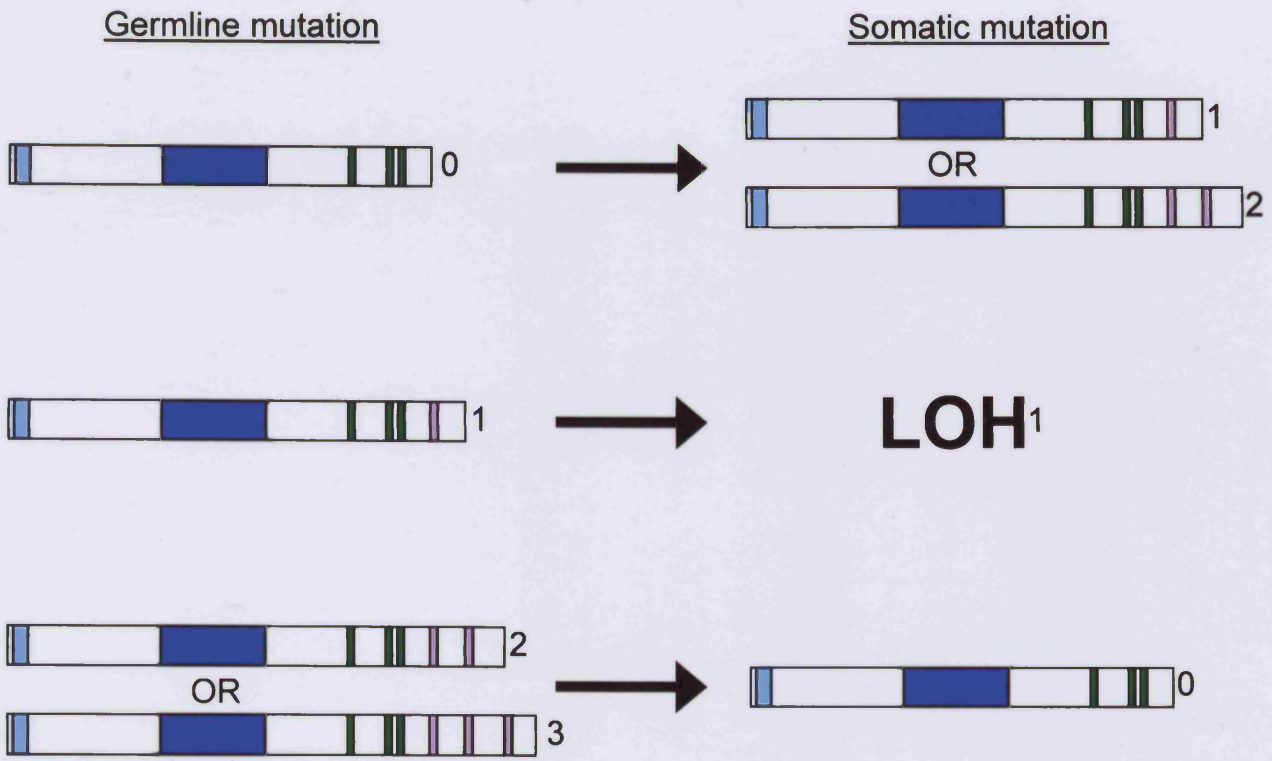


Figure 1.7 Interdependence of *APC* germline and somatic mutations in classical FAP. The location of the *APC* germline mutation determines the type and location of the somatic mutation in classical FAP tumours such that the total number of 20-aa repeats encoded by both alleles is most frequently two, but ranges from one and three. As LOH found in FAP tumours does not alter the gene dosage and is therefore likely to be a result of mitotic recombination not deletion, this allele still contributes a 20-aa repeat. The majority, but not all, of FAP tumours conform to the model shown. ▌,20-aa repeats, number shown at the end of each protein. Other colourings as in Figure1.3. Adapted from Albuquerque, C. *et al*, 2002.

advantageous for tumour formation, overexpression of β-catenin triggers apoptosis. Clearly there has to be some impairment of APC's ability to downregulate β-catenin in order to confer a selective growth advantage on the cell and truncated APC containing one, two or three 20-aa repeats have some residual β-catenin regulating activity. This 'just-right signalling' model suggests that some, but not too much, β-catenin signalling gives the greatest advantage for tumour formation and is why particular *APC* genotypes are selected (Albuquerque, C. *et al*, 2002; Crabtree, M. *et al*, 2003).

This model also explains why somatic truncating mutations cluster within the MCR. Patients with germline mutations between the first and second 20-aa repeats do not contribute much to the MCR data as they tend to have LOH as their somatic mutation. Classical FAP patients rarely have germline mutation after the second 20-aa repeat due to the relatively small size of this region (codons 1399-1580) so patients with germline mutations before the first 20-aa repeat contribute most of the MCR data. The somatic mutations found in these patients generally encode proteins with one or two 20-aa repeats and thus lie between codons 1285 and 1494, within the MCR (Crabtree, M. *et al*, 2003).

Unlike tumours from classical FAP patients, some AFAP tumours contain two somatic mutations, one of which affects the germline mutant *APC* allele (Spirio, L.N. *et al*, 1998; Su, L-K. *et al*, 2000; Sieber, O. *et al*, 2006). The highest frequency of these 'third hits' is found in tumours from patients with a germline mutation in the alternatively spliced region of exon 9. The third hit often leaves three 20-aa repeats intact on the germline mutant allele, in combination with either LOH or proximal somatic mutation of the wild type allele (Sieber, O. *et al*, 2006). Bearing in mind the alternative splicing of exon 9, these combinations of mutations may be selected to give an optimal level of β-catenin signalling for tumour formation. Somatic mutations in tumours from AFAP patients with a germline mutation at the 5' or 3' end of *APC* can follow the pattern expected in tumours from classical FAP patients with a germline mutation before the first 20-aa repeat; truncating mutations on the wild type allele resulting in a protein with one or two 20-aa repeats (Sieber, O. *et al*, 2006). AFAP mutations at the 5' or 3' end may produce a small quantity of functionally impaired or unstable

protein and such somatic mutations could be selected to give an optimal level of β-catenin signalling for tumour formation. Not all AFAP tumours require three hits in *APC* which could be a result of modifier alleles affecting levels of functional APC protein and its splicing efficiency, or the presence of a third hit at another locus (Sieber, O. *et al*, 2006).

1.3.1.6.3 Genotype-phenotype correlations

FAP patients do not all develop the same phenotype and this variation is partially accounted for by differences in the position of the germline *APC* mutation. Patients with severe polyposis tend to have germline mutations between codons 1250 and 1464 (Laurent-Puig, P. *et al*, 1998), with early onset severe polyposis linked to mutations at, or around codon 1309 (Friedl, W. *et al*, 2001). The reason for the association of a particularly severe form of FAP with germline mutations around codon 1309 is related to the type of somatic mutation selected for in these tumours, LOH. It is thought that germline mutations around 1309, in combination with somatic LOH of the wild type allele, result in a protein that gives an optimal level of β-catenin signalling, conferring a strong selective advantage on the cell for tumourigenesis. LOH occurs spontaneously at a higher frequency than truncating mutations close to codon 1300 so the frequency of pathogenic mutations in these patients is effectively higher, leading to severe polyposis (Lamlum, H. *et al*, 1999; Crabtree, M. *et al*, 2003).

AFAP generally results from germline mutations at the 5' or 3' ends of *APC* (codons <163 and 1596-2644) or between codons 329 and 338 in the alternatively spliced region of exon 9 (Galiatsatos, P. & Foulkes, W.D., 2006). AFAP patients with germline mutations at the 5' end tend to have a more severe phenotype than those AFAP patients with mutations in exon 9 or at the 3' end (Soravia, C. *et al*, 1998; Sieber, O. *et al*, 2006). Translation of 5' mutant alleles can start downstream of the mutation, at codon 184, producing a nearly full-length protein. Initiation at this site is probably inefficient so a limited amount of partially functional protein is predicted to be produced (Heppner Goss, K. *et al*, 2002). An alternatively spliced minor transcript of *APC* lacks exons 1-4 (Samowitz, W.S. *et al*, 1995) so a germline mutation within this region would not

affect this transcript and a small amount of protein with residual activity is predicted to be produced. Similarly, part of exon 9 is alternatively spliced in a minor *APC* transcript (Groden, J. *et al*, 1991) which would be unaffected by a mutation in this region so a limited amount of functional protein would also be predicted to be produced. Mutations at the 3' end may result in an unstable mutant APC protein (Knudsen, A.L. *et al*, 2003) but some partially functional protein is likely to be present since such mutations result in AFAP rather than classical FAP. AFAP mutations often seem to result in production of an APC protein with sufficient function that not only does the wild type allele need to be mutated, but a second mutation in the mutant allele is often also necessary for tumour initiation. AFAP patients may therefore have a milder phenotype than classical FAP patients because of the requirement for 'three mutational hits' rather than two (Sieber, O. *et al*, 2006).

Extracolonic manifestations are also associated with particular germline *APC* mutations, CHRPE is found in patients with germline mutations between codons 457 and 1444 (Fearnhead, N.S. *et al*, 2001) and desmoid tumours are associated with mutations at codons 1445-1580 (Friedl, W. *et al*, 2001). If the germline mutation lies between codons 976 and 1067, the patient appears to be at a higher risk of developing severe duodenal disease (Galiatsatos, P. & Foulkes, W.D., 2006). Papillary thyroid cancer has been found in patients with germline mutations ranging from codon 140 to 1309, an increased risk of osteomas is associated with mutations after codon 1444 (Galiatsatos, P. & Foulkes, W.D., 2006)

1.3.1.6.4 Phenotypic variability

Phenotypic variation also exists between FAP patients with the same *APC* germline mutation. Differences in the density of colorectal polyps have been found between and within FAP families with the same mutation at codon 1309, one of the recurrently mutated codons (Giardiello, F.M. *et al*, 1994). Similarly, variation in the number of colorectal polyps exists between families with the same, recurrent, codon 1061 mutation (Paul, P. *et al*, 1993). Extracolonic manifestations including; duodenal or gastric adenomas, desmoids, epidermoid cysts, osteomas, occult jaw lesions and extracolonic cancer also show inter-

and intrafamilial variability in families with the same mutation, either at codon 1309 or 1061. CHRPE, however, appears to be quite a consistent feature within and between these FAP families (Giardiello, F.M. *et al*, 1994; Paul, P. *et al*, 1993). Similar extracolonic variation was also seen within a family of 12 affected members, all with the same mutation affecting codon 1068 (Paul, P. *et al*, 1993).

Phenotypic differences are not limited to families with mutations predicted to result in classical FAP. Inter- and intrafamilial variation is also seen in families with a 3' mutation affecting codon 1979 which, on the basis of position, would be expected to result in AFAP. Most patients with this mutation do appear to have a phenotype characteristic of AFAP, with fewer than 100 colorectal polyps and a later age of onset than classical FAP. However, some patients with the same mutation had more than 100 colorectal polyps, traditionally classed as classical FAP. Extracolonic features were also variable within and between families with this mutation although lack of CHRPE was consistent (Brensinger, J.D. et al, 1998). A similar variation in phenotype was observed between members of the same family with a mutation affecting the alternatively spliced region of exon 9. The colonic features of these patients ranged from those typical of classical FAP to no clinical manifestation at all (Rozen, P. *et al*, 1999).

The inter- and intrafamilial variation in phenotype seen in FAP patients with the same germline mutation suggests that other genetic or environmental factors are contributing to the expression of germline *APC* mutations. Environmental factors such as diet are more likely to vary between families rather than within them, and although a high-fat low fibre diet can increase polyp number in a FAP mouse model (Hioki, K. *et al*, 1997), it is unlikely to have a profound effect on the phenotype of classical FAP patients due to the early age at onset of this disease (Crabtree, M.D. *et al*, 2002). First degree relatives have a more similar FAP phenotype than second degree relatives, which provides evidence for the role of unlinked genetic factors in phenotypic variation (Crabtree, M.D. *et al*, 2002). Consistent with this, the Apc^{Min} mouse model of FAP shows variation in polyp number dependent on the genetic background (Dietrich, W.F. *et al*, 1993) and a major modifier locus of this model, *Mom1* (modifier of *Min* 1), was

identified as the gene encoding secretary phospholipase A_2, *Pla2g2a* (Dietrich, W.F. *et al*, 1993; Cormier, R.T. *et al*, 1997). Three other Mom loci have been identified (Cormier, R.T. *et al*, 2000; Silverman, K.A. *et al*, 2002; Haines, J. *et al*, 2005) and a variant in an ATP synthase gene (*Atp5a1*) has recently been found at the *Mom2* locus (Baran, A.A. *et al*, 2007). The human orthologue of *Pla2g2a* does not appear to be a modifier of FAP phenotype in humans (Dobbie, Z. *et al*, 1996a) but variations at the *NAT1* and *NAT2* loci (which encode N-acetyl transferases) appear to influence the disease severity of FAP patients (Crabtree, M.D. *et al*, 2004).

1.3.2 MUTYH-Associated Polyposis (MAP)

1.3.2.1 Clinical presentation

MAP is a recently described colorectal adenoma and carcinoma predisposition syndrome (Al-Tassan, N. *et al*, 2002) typically characterized by the presence of tens to hundreds of colorectal adenomas by the fifth or sixth decade. Approximately half of MAP patients have already developed CRC by presentation (Sampson, J.R. *et al*, 2003; Sieber, O.M *et al*, 2003; Nielsen, M. *et al*, 2005; Aretz, S. *et al*, 2006) and CRC can develop in the apparent absence of colorectal polyposis (Croitoru, M.E. *et al*, 2004; Wang, L., *et al*, 2004). MAP and FAP (particularly AFAP) are difficult to differentiate clinically because of similarities in their colorectal phenotypes. In addition, duodenal adenomas that are frequently present in FAP also appear to be the most common extracolonic feature of MAP (Nielsen, M. *et al*, 2005; Aretz, S. *et al*, 2006). Other extracolonic manifestations have been reported in MAP patients, but not at significant frequencies. Like FAP, MAP is a rare disease thought to be responsible for only about 1% of all colorectal cancers (Fleischmann, C. *et al*, 2004).

1.3.2.2 Genetics

MAP is an autosomal recessive disorder caused by biallelic germline mutations in the *MUTYH* (also commonly, but incorrectly, referred to as *MYH*) gene (Jones, S. *et al*, 2002). The penetrance of this disease appears to be high as no unaffected older carriers of biallelic *MUTYH* mutations have been identified to date.

1.3.2.3 The MUTYH gene

Slupska *et al* (1996) cloned and sequenced the human *MUTYH* gene found on the short arm of chromosome 1, between p32.1 and p34.3. This 7.1kb gene is made up of 16 exons (Slupska, M.M. *et al*, 1996) and can produce three major mRNA transcripts (α, β and γ) which differ in their 5' sequence, indicating that *MUTYH* has three independent transcription initiation sites (Ohtsubo, T. *et al*, 2000). Each transcript can be alternatively spliced to result in at least ten different mature transcripts, as shown in Figure 1.8.

The α transcript encodes a mitochondrial targeting signal (MTS) within its 5' sequence and subtypes α1, α2 and α3 code for proteins which can localise to the mitochondria (Ohtsubo, T. *et al*, 2000). Consistent with this, the protein encoded by transcript α3 localises to mitochondria (Takao, M. *et al*, 1999) and a mitochondrial MUTYH polypeptide was identified by Ohtsubo *et al* (2000) and Boldogh *et al* (2001). By contrast, a MUTYH polypeptide corresponding to the predicted size of the α3-encoded protein (59kDa) was identified in nuclear extracts by Tsai-Wu *et al* (2000). This inconsistency may be the result of MUTYH shuttling between the nucleus and the mitochondria (Tsai-Wu, J-J. *et al*, 2000). Translation of subtype α4 from the same initiation codon as the other α transcripts would result in a stop codon so translation may start at a downstream initiation site in exon 4 resulting in a protein lacking the MTS (Ohtsubo, T. *et al*, 2000). The β and γ transcripts also skip exon 1 (which encodes the MTS) and translation is thought to start at a downstream initiation codon in exon 2 resulting in a protein that lacks the first 14 amino acids (Ohtsubo, T. *et al*, 2000; Takao, M. *et al*, 1999). The exception to this is γ4, whose translation is thought to start at the same initiation site as α4, in exon 4 (Ohtsubo, T. *et al*, 2000). A putative nuclear localisation signal (NLS) is found within exon 2 and is thought to be non-functional in the context of the MTS (i.e. in proteins encoded by α1,2 and 3) but may function in those proteins lacking the first 14 amino acids (Takao, M. *et al*, 1999). Another NLS is located within exon 16 and is common to all transcripts (Tsai-Wu, J-J. *et al*, 2000).

MUTYH mRNA is expressed in a wide range of human tissues including the small intestine, lung, spleen and mammary gland but differences in the level of

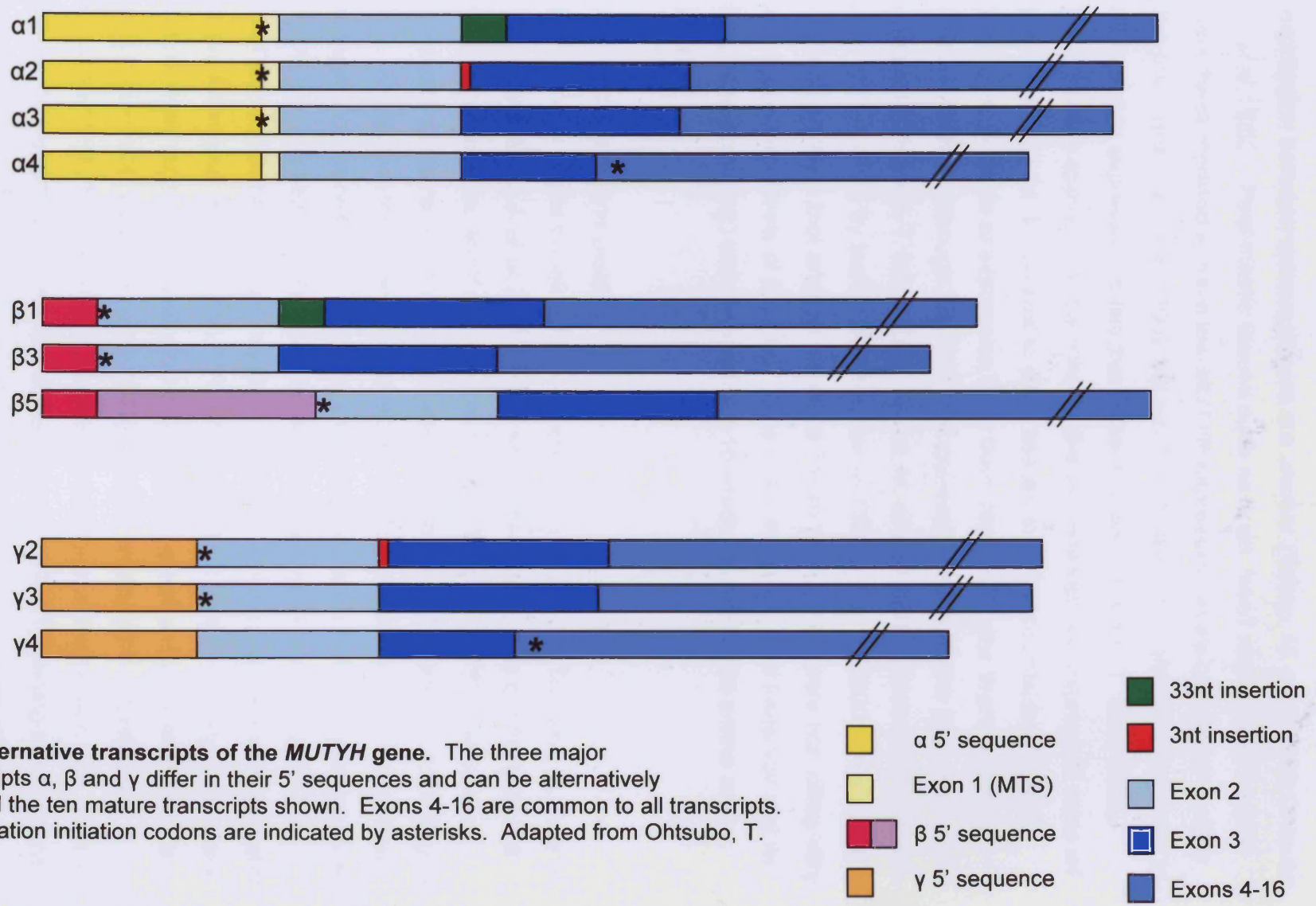


Figure 1.8 Alternative transcripts of the *MUTYH* gene. The three major mRNA transcripts α, β and γ differ in their 5' sequences and can be alternatively spliced to yield the ten mature transcripts shown. Exons 4-16 are common to all transcripts. Putative translation initiation codons are indicated by asterisks. Adapted from Ohtsubo, T. *et al*, 2000.

expression between various tissues are unclear (Takao, M. *et al*, 1999;Ohtsubo, T. *et al*, 2000). Post-mitotic tissues such as brain, heart and skeletal muscle have been reported to have low *MUTYH* expression levels compared to other tissues (Takao, M. *et al*, 1999), but adult brain tissue has also been reported to have higher expression levels than other tissues (Ohtsubo, T. *et al*, 2000). Brain tissues appeared to be richer in the γ4 transcript and expressed more α4 than α1 transcript, in contrast to other tissues examined (Ohtsubo, T. *et al*, 2000). High levels of expression have been reported in the thymus, testis, kidney and liver although high *MUTYH* expression levels in the liver have been disputed (Ohtsubo, T. *et al*, 2000; Takao, M. *et al*, 1999). These discrepancies may be explained by technical difficulties as Ohtsubo *et al* (2000) were able to perform Northern blot analysis whereas Takao *et al* (1999) were not, citing very low expression levels of *MUTYH*. Type α is a major *MUTYH* transcript and its longest subtype (α1) includes exons 1-16 which encode a 546 amino acid protein.

1.3.2.4 The MUTYH protein

MUTYH is a 59kDa protein when encoded by transcript α3 (Slupska, M.M. *et al*, 1999; Tsai-Wu, J-J. *et al*, 2000) which has 41% identity to its *Escherichia coli* (*E.coli*) homologue, MutY (Slupska, M.M. *et al*, 1996). MUTYH has several functional domains which allow it to interact with other proteins and DNA (Figure 1.9). A region at the N-terminus facilitates replication protein A (RPA) binding through a conserved motif which includes the sequence K/RXK/R (Parker, A. *et al*, 2001). The study of MutY functional domains and their homology to the MUTYH protein has allowed the identification of several highly conserved motifs in the N-terminal domain of MUTYH; the DNA minor groove reading motif, helix-hairpin-helix motif (HhH), pseudo HhH motif, iron-sulphur cluster and adenine recognition motif (Guan, Y. *et al*, 1998). MutS homologue 6 (MSH6) can interact with MUTYH through a 23 amino acid region of the protein (Gu, Y. *et al*, 2002). The apurinic/apyrimidinic endonuclease 1 (APE1) binding site contains the motif SGXXDV (Parker, A. *et al*, 2001) and partially overlaps with the recently identified Hus1 binding site (Shi, G. *et al*, 2006). MUTYH also interacts with proliferating cell nuclear antigen (PCNA) through a C-terminal binding site which contains a QXXLXXFF motif (Parker, A. *et al*, 2001). The N-terminus of

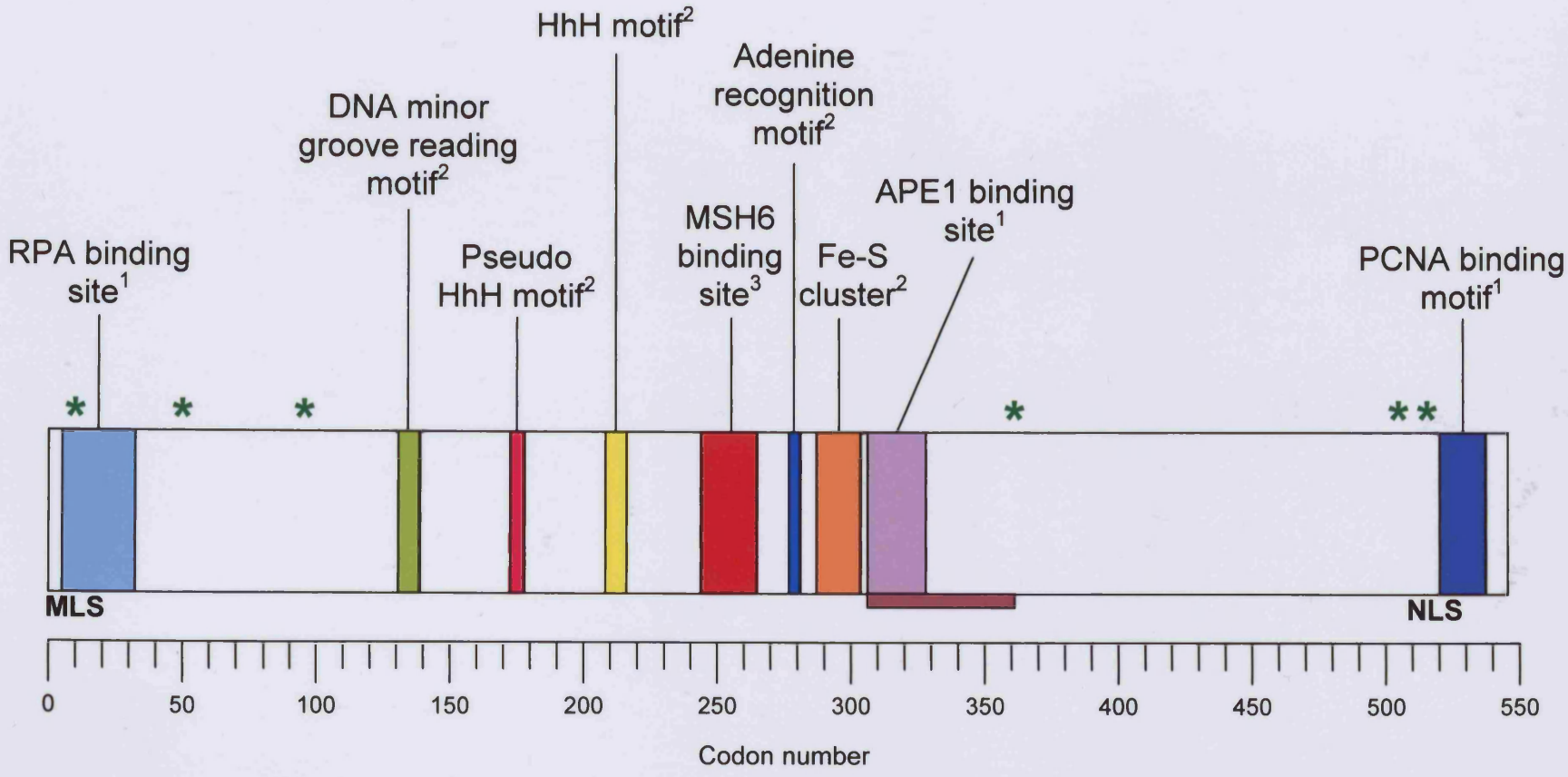


Figure 1.9 Functional domains of the MUTYH protein. Protein shown is encoded by *MUTYH* transcript α1. HhH; Helix-hairpin-Helix, Fe-S cluster; iron-sulphur cluster. [1]Parker, A. *et al*, 2003; [2]based on MutY, Guan, Y. *et al*, 1998; [3]Gu, Y. *et al*, 2002; [4]Shi, G. *et al*, 2006; [5]Ohtsubo, T. *et al*, 2000; [6]Takao, M. *et al*, 1999.

MUTYH contains a MTS, and a functional NLS is found at the C-terminus (Takao, M. *et al*, 1999). Six putative MUTYH serine phosphorylation sites have been identified, three in the N-terminal region (codons 9, 49 and 85) and three within the C-terminal half of the protein at codons 349, 494 and 504 (Parker, A.R. *et al*, 2003).

1.3.2.5 Function of MUTYH

MUTYH is DNA repair protein which helps to maintain genomic stability through its role in the base excision repair (BER) of oxidative DNA damage.

1.3.2.5.1 Oxidative DNA damage

In humans, an estimated 10^4 oxidative DNA damage lesions occur in each cell per day (Ames, B.N. & Gold, L.S., 1991). The main cause of this damage is reactive oxygen species (ROS) which include hydroxyl radicals and hydrogen peroxide. ROS are generated endogenously in the cell by normal metabolic pathways such as the mitochondrial electron transport chain and are also produced as a result of lipid peroxidation or upon exposure to ionizing radiation (King, R.J.B., 2000; Lu, A-L. *et al*, 2001). Commensal bacteria and dietary carcinogens may also contribute to high levels of oxidative stress in the colon (Huycke, M.M. & Gaskins, H.R, 2004; Ames, B.N. & Gold, L.S., 1991). The DNA damage produced by ROS can take the form of strand breaks or several types of individual base lesions. If the damage is extensive, this is sensed and can lead to cell death. If only limited damage results, it can be repaired and does not have a deleterious effect on the cell. However, if modified bases are not repaired before DNA replication, they can be misread and result in mutations (King, R.J.B., 2000; Lu, A-L. *et al*, 2001). Oxidative DNA damage is thought to contribute to aging, degenerative diseases and cancer (Ames, B.N. & Gold, L.S., 1991).

1.3.2.5.2 Base excision repair

BER is the main mechanism that protects against oxidative DNA damage (Hazra, T.K, *et al*, 2007). Lesions resulting from methylation, deamination and hydroxylation can also be repaired through this pathway (Hoeijmakers, J.H.J, 2001). As shown in Figure 1.10, the damaged base is recognised by a DNA

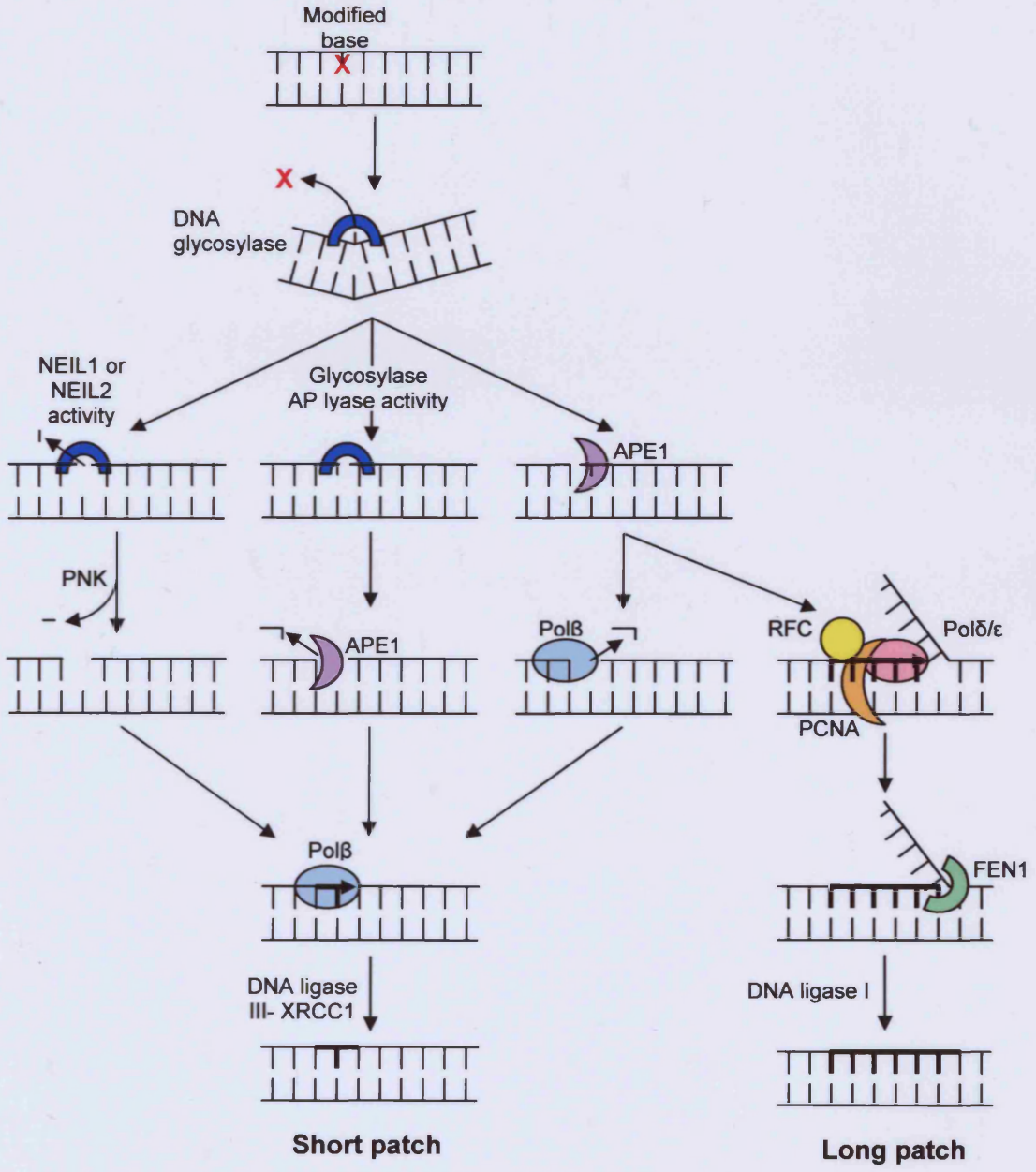


Figure 1.10 Base excision repair of a damaged DNA base. The use of short patch or long-patch repair is determined by the DNA glycosylase recruited, type of DNA damage and cell cycle phase. Polβ incorporates the first nucleotide but Pol δ/ε carries out elongation in the DNA synthesis step of long-patch repair. (Adapted from Wilson III, D.M. & Bohr, V.A., 2007; Fortini, P. & Dogliotti, E., 2007).

glycosylase which cleaves it from the sugar-phosphate backbone, resulting in an apurinic/apyrimidinic (AP) site. AP sites produced as a result of spontaneous hydrolysis can also be repaired through BER. Some glycosylases have an AP lyase activity and cleave the sugar-phosphate backbone 3' to the AP site. The APE1 endonuclease then processes the DNA in preparation for DNA synthesis. In the absence of a glycosylase lyase activity, APE1 prepares the DNA for synthesis by cleaving the sugar-phosphate backbone 5' to the AP site (Wilson III, D.M. & Bohr, V.A., 2007). Recently an APE1-independent BER pathway has been described (Wiederhold, L. *et al*, 2004) in which the NEIL1 or NEIL2 DNA glycosylase cleaves DNA at the AP site, leaving a 3' phosphate group. This is then removed by polynucleotide kinase (PNK) so DNA synthesis can occur (Wiederhold, L. *et al*, 2004). The repair pathway involving monofunctional glycosylases now diverges into either short-patch or long-patch BER (Wilson III, D.M. & Bohr, V.A., 2007). If long-patch repair is carried out, DNA polymerase β incorporates the first nucleotide but elongation is carried out by DNA polymerases δ/ε. A two or more nucleotide repair patch is synthesized from the damaged site, facilitated by PCNA and replication factor C (RFC). Flap endonuclease 1 (FEN1) then removes the displaced DNA flap and the nick is sealed by DNA ligase I. In short-patch repair, DNA polymerase β cleaves the sugar-phosphate backbone 3' to the AP site through its phosphodiesterase activity and fills the one-nucleotide gap. The remaining nick is sealed by the DNA ligase III/XRCC1 (X-ray repair cross complementing 1) heterodimer. DNA synthesis in repair pathways initiated by glycosylases with AP lyase activity or the NEILs is also short-patch and carried out by DNA polymerase β, with the nick sealed by the DNA ligase III/XRCC1 complex. The choice of short patch or long-patch repair is dependent on the DNA glycosylase recruited, type of DNA damage present and cell cycle phase. Direct single strand breaks (e.g those induced by ROS) can also be repaired by either short-patch or long-patch BER once their termini have been processed to contain a 3' hydroxyl group and a 5' phosphate group i.e. prepared for DNA synthesis (Fortini, P. & Dogliotti, E., 2007).

1.3.2.5.3 Repair of 7,8-dihydro-8-oxoguanine

One of the most stable products of oxidative damage is 7,8-dihydro-8-oxoguanine (8-oxoG) (Ames, B.N. & Gold, L.S., 1991) which can readily mispair with adenine during DNA replication (Shibutani, S. *et al*, 1991) and if left unrepaired, results in a G:C → T:A transversion following the next round of replication (Michaels, M.L. & Miller, J.H., 1992). The repair system which protects against 8-oxoG damage was first characterised in *E.coli* and the key enzymes involved were identified as MutT, MutM and MutY (Michaels, M.L. & Miller, J.H., 1992) (Figure 1.11). MutT is a nucleoside triphosphatase which hydrolyzes 8-oxodGTP in the nucleotide pool, preventing its incorporation into nascent DNA (Maki, H. & Sekiguchi, M., 1992). The MutM DNA glycosylase recognises and excises 8-oxoG paired opposite cytosine (Boiteux, S. *et al*, 1992) and MutY is a DNA glycosylase responsible for removing adenine residues misincorporated into the daughter strand opposite 8-oxoG following replication (Michaels, M.L. *et al*, 1992). Cytosine is then inserted in place of the excised adenine, resulting in a 8-oxoG:C pair which can be repaired by MutM. The MutS DNA glycosylase is part of the mismatch repair (MMR) system and can perform the same function as MutY (Zhao, J. & Winkler, M.E., 2000). Another DNA glycosylase, endonuclease VIII (or Nei protein) provides a backup pathway for repair of 8-oxoG as it can excise this damaged base from 8-oxoG:C pairs (Blaisdell, J.O. *et al*, 1999) and the nascent strand when misincorporated opposite adenine during replication (Hazra, T.K. *et al*, 2001). Human homologues of *mutT*, *mutM, mutY, mutS* and *nei* have been identified as *NUDT1* (Sakumi, K. *et al*, 1993), *OGG1* (Roldán-Arjona, T. *et al*, 1997), *MUTYH* (Slupska, M.M. *et al*, 1996), the *MSH*s (Fishel, R. *et al*, 1993; Acharya, S. *et al*, 1996) and the *NEIL*s (Bandaru, V. *et al*, 2002), respectively.

1.3.2.5.4 MutY catalytic mechanism

The bacterial MutY protein has been studied more extensively than its human homologue, MUTYH, and has provided the majority of the structural and biochemical data on this enzyme. MutY is a member of the helix-hairpin-helix (HhH) superfamily of proteins which all have a conserved catalytic domain containing an HhH motif followed by a glycine/proline-rich loop and a catalytic aspartate residue (HhH-GPD motif) (Fromme, J.C. *et al*, 2004). The N-terminal

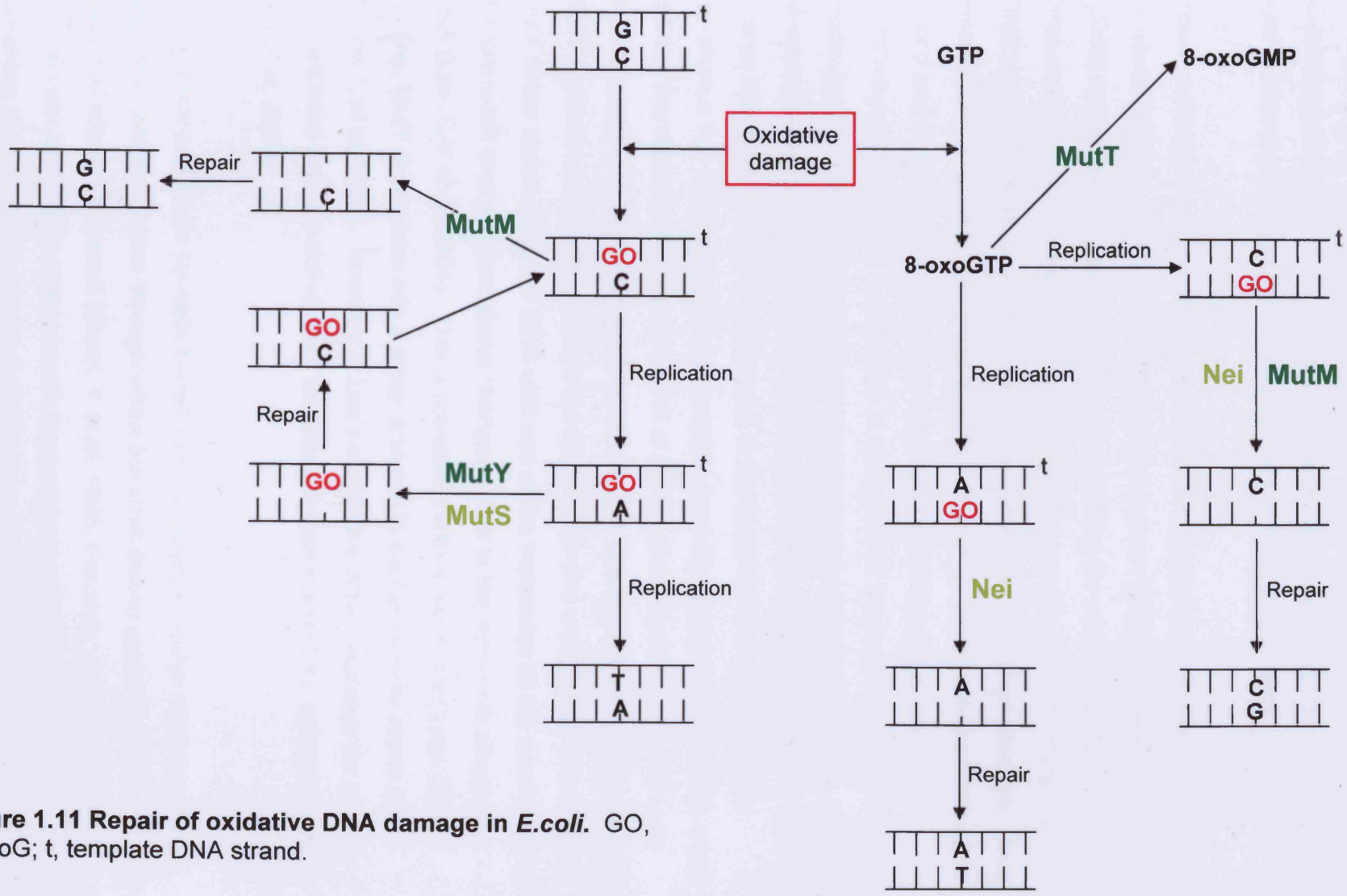


Figure 1.11 Repair of oxidative DNA damage in *E.coli*. GO, 8-oxoG; t, template DNA strand.

catalytic domain of MutY is made up of two modules; a six-helix barrel containing the HhH motif and an iron-sulphur cluster (Guan, Y *et al*, 1998). The C-terminal domain of MutY is unique among HhH proteins, but similar to MutT and is necessary for substrate specificity (Noll, D.M. *et al*, 1999).

The mechanism by which MutY searches for mismatches within the vast excess of normal DNA is not known, but models have been proposed. Boon *et al* (2003) suggested a model of long-range scanning for mismatches through DNA-mediated charge transport chemistry. Non-specific binding of MutY to DNA results in oxidation of its iron-sulphur cluster and loss of an electron. This electron is then transported along the DNA to another MutY molecule already bound and its iron-sulphur cluster is reduced, promoting dissociation of the protein from the DNA. If a mismatch is encountered between the two MutY molecules, the DNA-mediated charge transport cannot occur and the protein remains bound to the DNA and processively diffuses to the mismatch. The charge transport process is continuous in the absence of a mismatch and therefore a few MutY molecules can rapidly scan the genome (Boon, E.M. *et al*, 2003). The model proposed by Lee *et al* (2004) also requires more than one MutY molecule. A MutY monomer binds to DNA non-specifically then a second MutY molecule cooperatively binds the same DNA and a dimer is formed. The MutY dimer slides along the DNA until one of the molecules binds specifically to the mismatch then conformational changes occur in the active monomer and DNA (Lee, C-Y *et al*, 2004). This is consistent with a report that upon DNA binding, MutY assembles into a dimer and this is the functionally active state (Wong, I. *et al*, 2003). However in Lee's model the dimer searches for a mismatch but only the active MutY molecule remains bound for catalysis (Lee, C-Y *et al*, 2004).

The N-terminal domain six-helix barrel and iron-sulphur cluster modules form a positively-charged groove through which the DNA strand containing the mispaired adenine is bound (Guan, Y *et al*, 1998; Fromme, J.C. *et al*, 2004). The six-helix barrel also interacts with the backbone of the DNA strand containing the complementary 8-oxoG residue (Fromme, J.C. *et al*, 2004). The C-terminal domain of MutY confers specificity for 8-oxo-G:A mispairs by

recognising the 8-oxoG base (Noll, D.M. *et al*, 1999; Li, X. *et al*, 2000). It makes extensive contacts with the 8-oxo-G-strand of DNA and the oxidized base itself. 8-oxoG is hydrogen-bonded to both the C-terminal and catalytic domains, ensuring that adenines correctly paired opposite thymines are not excised (Fromme, J.C. *et al*, 2004). MutY mainly contacts the G-strand and interacts with DNA via at least five phosphates and purine bases on either side of the mismatch (Lu, A-L. *et al*, 1995). A double base flipping mechanism was proposed to facilitate recognition of the 8-oxo-G:A mispair and aid adenine extrusion into the active site (Bernards, A.S. *et al*, 2002) but crystal structures of MutY bound to DNA showed that although the DNA is bent and the mispaired adenine is extruded, 8-oxoG remains within the helix (Fromme, J.C. *et al*, 2004). The HhH and pseudo HhH motifs at either end of the DNA binding groove compress the DNA intrastrand phosphate distance either side of the mispaired adenine causing the DNA to bend and the target base to flip out of the DNA helix, into the active site pocket (Guan, Y *et al*, 1998). When MutY binds to an 8-oxoG:A mispair, the 8-oxoG nucleoside changes its glycosydic bond conformation from a *syn* to an *anti* orientation. The *anti* conformation of 8-oxoG would sterically clash with an adenine base opposite and so may also promote flipping of adenine into the active site (Fromme, J.C. *et al*, 2004). 8-oxoG binding to the C-terminal domain facilitates base-flipping and thus accelerates the glycosylase reaction (Noll, D.M. *et al*, 1999). The active site pocket is complementary to the extrahelical adenine residue, which is deeply sequestered and hydrogen-bonded by the MutY protein. The catalytic and C-terminal domains interact to allow the enzyme to surround the DNA duplex (Fromme, J.C. *et al*, 2004). The flipped-out adenine leaves a gap in the DNA base stack which may be stabilized by the minor groove reading motif. MutY then cleaves the N-glycosylic bond through a hydrolytic mechanism (Guan, Y et al, 1998). The adenine base is released quickly from the enzyme and the rate-limiting step is release of DNA containing an AP site as it is tightly bound by MutY (McCann, J.A.B. & Berti, P.J., 2003).

The DNA glycosylase endonuclease III also has AP lyase activity which is attributed to a lysine residue within its HhH motif. MutY lacks a lysine residue within the HhH motif and primarily functions as a pure DNA glycosylase, but

does have weak AP lyase activity (Guan, Y et al, 1998). Unlike endonuclease III, this additional activity is uncoupled to its glycosylase activity and is alternatively regulated by two lysine residues positioned near the active site (Manuel, R.C. *et al*, 2004).

1.3.2.5.5 Role of MUTYH in DNA repair

The human MutY homologue, MUTYH, demonstrates glycosylase activity with DNA substrates containing the following mispairs; A:G, 2-hydroxyadenine (2-OH-A):8-oxo-G, A:8-oxo-G, 2-OH-A:G and 2-OH-A:A but exhibits little AP lyase activity (Ohtsubo, T. *et al*, 2000). 2-OH-A is another product of oxidative DNA damage (Lu, A-L. *et al*, 2001). Under physiological salt conditions, MUTYH specifically catalyzes the glycosylase reaction on A:8-oxoG not A:G mispairs (Shinmura, K. *et al*, 2000). Mutant MUTYH is defective in binding and repair of A:8-oxoG and 8-oxoadenine (8-oxoA):G mispairs (Parker, A.R. *et al*, 2005). Phosphorylation of serine residues enhances MUTYH repair of A:8-oxoG mispairs and may regulate protein-protein interactions (Parker, A.R. *et al*, 2003). MUTYH interacts with APE1, PCNA and RPA1 suggesting that repair of DNA lesions by MUTYH occurs via the long-patch pathway of BER (Parker, A. *et al*, 2001). Indeed, short-patch BER initiated by murine MUTYH is futile and the repair has to proceed to the long-patch pathway (Hashimoto, K. *et al*, 2004). MUTYH can interact with Hus1 (part of the 9-1-1 DNA damage sensor complex) and may act as an adaptor for sensor checkpoint proteins following oxidative DNA damage (Shi, G. *et al*, 2006). Nuclear MUTYH co-localizes with PCNA at replication foci and levels of MUTYH are maximal during S phase indicating that MUTYH repair may be coupled to replication (Boldogh, I. *et al*, 2001). Consistent with this, DNA replication enhances murine MUTYH repair through its interaction with PCNA (Hayashi, H. *et al*, 2002). The coupling of replication and MUTYH repair allows MUTYH to be directed to the daughter strand, thereby preventing the A:T to C:G transversions that would arise from adenine glycosylase activity on the template strand. MUTYH also interacts with the MMR protein MSH6 which forms a heterodimer with MSH2. The mammalian MSH2/MSH6 complex can repair 8-oxoG incorporated into the daughter strand during replication (Colussi, C. *et al*, 2002) and A:8-oxoG mismatches where the adenine is on the template strand are mainly recognised by this complex (Gu,

Y. *et al*, 2002). However, it has been reported that MSH2/MSH6 repair of 8-oxoG incorporated during replication opposite adenine or cytosine is inefficient (Larson, E.D. *et al*, 2003). Recently, Macpherson *et al* (2005) have shown MSH2/MSH6 can efficiently recognise 8-oxoG opposite adenine or cytosine when these mispairs lie within repetitive sequences that have undergone a slippage event (Macpherson, P. *et al*, 2005). The MSH2/MSH6 complex binds 8-oxoG:G and 8-oxoG:T mispairs efficiently and although such mispairs are rare *in vivo*, when the oxidised base is on the template strand MMR can result in 8-oxoG:A mispairs which are recognised by MUTYH (Larson, E.D. *et al*, 2003). The MSH2/MSH6 dimer can enhance binding of MUTYH to A:8-oxoG mismatches and interactions between these repair proteins may help target MUTYH to the daughter strands (Gu, Y. *et al*, 2002). Wong *et al* (2003) proposed a mechanism by which dimeric MutY could determine the level of oxidative damage present and remove adenines only present on the nascent strand. A:8-oxoG mispairs resulting from incorporation of adenine into the nascent strand opposite 8-oxoG are found in isolation and repaired by MutY. However, A:8-oxoG mispairs formed as a result of incorporation of 8-oxo-GTP into the nascent strand opposite adenine would arise as patches of damage. Wong *et al* (2003) suggested that concurrent binding of two of these mispairs by dimeric MutY would inhibit its glycosylase activity, thereby preventing A:T to C:G transversions and allow recruitment of more appropriate repair enzymes such as those involved in MMR (Wong, I. *et al*, 2003). Mammalian MUTYH protects its DNA product from OGG1 as this DNA glycosylase could excise 8-oxoG opposite the AP site, resulting in a loss of informative bases on both strands and a double strand break. MUTYH may encircle and tightly bind the DNA duplex (like its bacterial homologue) thus physically preventing OGG1 from accessing the DNA (Tominaga, Y. *et al*, 2004)

1.3.2.6 Genetic pathway to CRC

The identification of MAP arose from investigating a family in which several siblings had an AFAP-like phenotype, but in whom a germline truncating mutation in *APC* was not identified. Analysis of somatic *APC* mutations in colorectal tumours from the affected siblings revealed that most of them were G:C to T:A transversions which almost exclusively occurred at GAA sequences,

resulting in stop codons. The pattern of *APC* somatic mutations was then linked to functionally compromising biallelic germline *MUTYH* mutations in the affected cases (Al-Tassan, N. *et al*, 2002) and this association was rapidly confirmed in further unrelated MAP cases (Jones, S. *et al*, 2002). A mouse model has also shown that homozygous *Mutyh* deficiency increases intestinal tumourigenesis in $Apc^{Min/+}$ mice and some polyps from these mice harboured G:C to T:A transversions resulting in termination codons (Sieber, O.M. *et al*, 2004). A specific somatic *K-RAS* mutation has been identified in some MAP adenomas and is associated with increased dysplasia. The G12C activating mutation is the result of a G:C to T:A transversion (Lipton, L. *et al*, 2003; Jones, S. *et al*, 2004). Although MAP tumours have somatic mutations in genes often mutated in the CIN pathway of tumourigenesis, they appear to be near-diploid yet they do not harbour somatic mutations in genes frequently mutated in MSI tumours such as *BRAF*, *SMAD4* and *TGFβIIR* (Lipton, L. *et al*, 2003).

Unlike other genes involved in inherited CRC predisposition, somatic mutations in *MUTYH* do not appear to be involved in sporadic colorectal tumourigenesis. No somatic *MUTYH* mutations were found in 75 sporadic CRCs and mRNA and protein were expressed in all of the 35 CRC cell lines examined, indicating that epigenetic silencing was not occurring (Halford, S.E.R. *et al*, 2003).

1.3.3 Hereditary Non-Polyposis Colorectal Cancer (HNPCC)

1.3.3.1 Clinical presentation

HNPCC is the most common Mendelian CRC syndrome, contributing to 2-5% of all new CRC cases (Lynch, H.T. *et al*, 2006). The population incidence of HNPCC is estimated to be between 1 in 2000 and 1 in 660 (de la Chapelle, A., 2005). The average age at CRC onset is 45 years and patients are at increased risk of synchronous and metachronous CRCs (Lynch, H.T. *et al*, 2006). A small number of colorectal adenomas (only exceptionally more than ten) may be present, but florid polyposis is not observed (Lucci-Cordisco, E. *et al*, 2003). Unlike FAP, where tumour initiation is accelerated but the progression rate is probably normal, HNPCC has accelerated tumour progression but probably a normal tumour initiation rate (Kinzler, K.W. & Vogelstein, B., 1996). HNPCC is also known as Lynch syndrome, in order to

reflect the extracolonic malignancies associated with this disease including endometrial cancer and cancers of the small bowel, urinary tract, stomach, hepatobiliary tract, ovaries and brain (Vasen, H.F.A, 2005; Watson, P. & Riley, B., 2005). Sebaceous skin tumours and internal malignancies are associated with the Muir-Torre variant (Ponti, G. & Ponz de Leon, M., 2005) and gliomas in addition to colorectal tumours are characteristics of Turcot syndrome (Lucci-Cordisco, E. *et al*, 2003).

1.3.3.2 Genetics

HNPCC is an autosomal dominant disorder caused by germline mutations in the *MSH2* (Fishel, R. *et al*, 1993), *MLH1* (Lindblom, A. *et al*, 1993), *MSH6* (Miyaki, M. *et al*, 1997) or *PMS2* (Nicolaides, N.C. *et al*, 1994) genes. Penetrance for CRC in males is approximately 80% (but is lower in females), up to 60% for endometrial cancer and less than 20% for other cancers (de la Chapelle, A., 2004).

1.3.3.3 Mismatch repair (MMR)

MSH2, *MLH1*, *MSH6* and *PMS2* all encode proteins which are involved in MMR. This pathway repairs a wide range of single base mismatches as well as insertion/deletion loops arising from DNA polymerase slippage during replication, often within repetitive sequences (Jascur, T. & Boland, C.R., 2006). MMR is specific to the daughter DNA strand and in *E. coli* the MutH protein recognises a particular sequence which is methylated in the template strand but transiently unmethylated in the daughter strand, thus targeting repair to the daughter strand (Kunkel, T.A. & Erie, D.A., 2005). No such methylation patterns occur in humans and no human MutH homologue has been identified. Communication between MMR proteins and enzymes at the replication fork through interactions with PCNA may allow strand discrimination to occur (Plotz, G. *et al*, 2006). As shown in Figure 1.12, the DNA damage is recognised by a MSH2 heterodimer, formed with either MSH6 or MSH3 depending on the type of DNA error. MSH2/MSH6 heterodimers recognise single base mismatches or small insertion/deletion loops whereas MSH2/MSH3 heterodimers recognise insertion/deletion loops. A MLH1/PMS2 heterodimer is then recruited and docks to the MSH2 heterodimer (Jascur, T. & Boland, C.R., 2006; Plotz, G. *et*

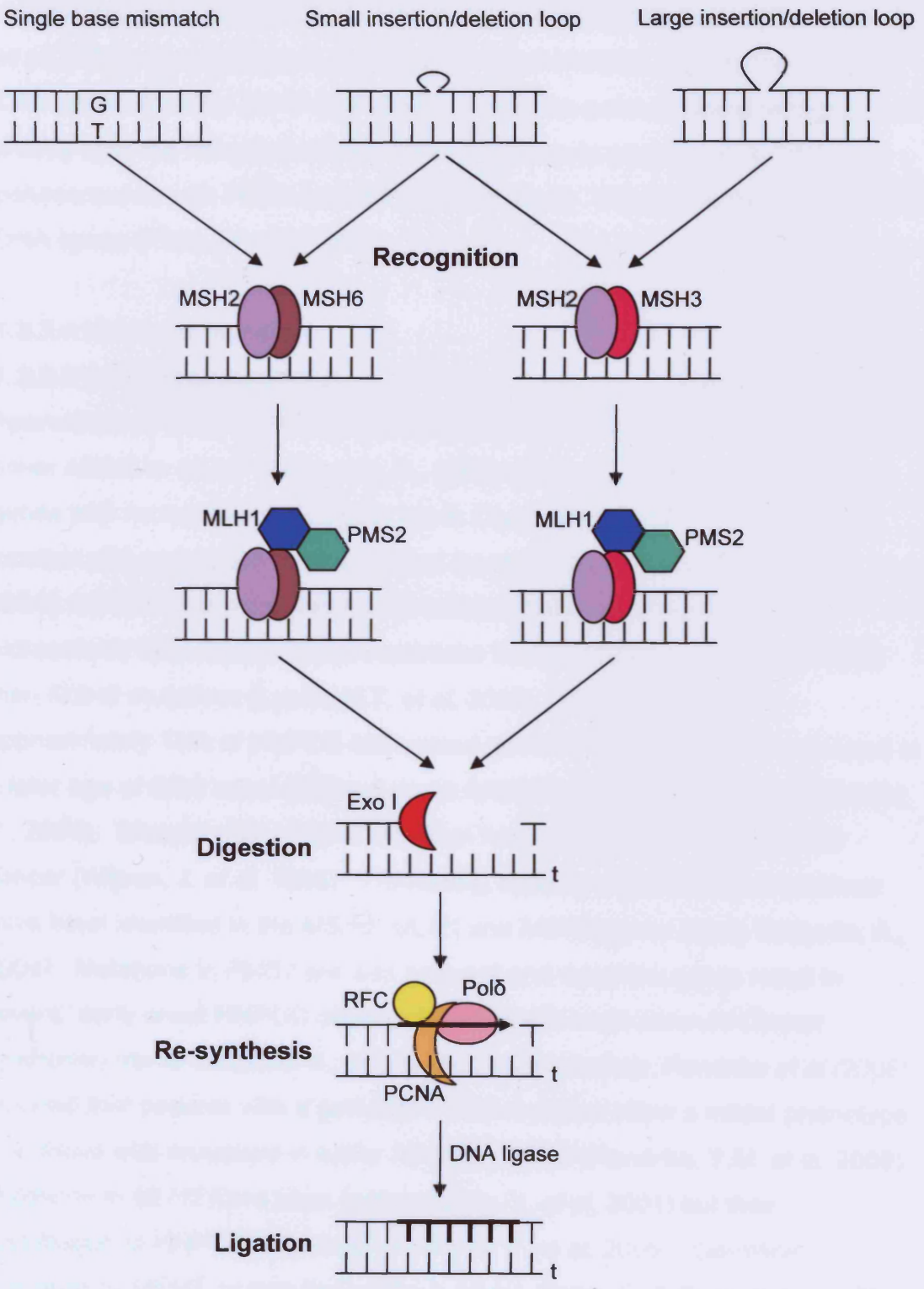


Figure 1.12 Human mismatch repair. Excision of the faulty daughter strand can take place in either the 3' or 5' direction. t, template strand

al, 2006). The MLH1/MLH3 heterodimer may function in repair of MSH2/MSH6 substrates, but is not thought to play a major role in MMR *in vivo* (Cannavo, E. *et al*, 2005). Exonuclease I, aided by RPA and HMGB1, digests the daughter DNA strand in either the 5' or 3' direction from the point of strand recognition to shortly after the mismatched site. DNA synthesis is carried out by DNA polymerase δ, with PCNA and RFC also required. The nick is then sealed by a DNA ligase (Plotz, G. *et al*, 2006).

1.3.3.4 Molecular genetics

1.3.3.4.1 Germline mutations

Approximately 90% of HNPCC-associated germline mutations are found in either *MSH2* or *MLH1* (Peltomäki, P., 2005) and are spread throughout these genes with no mutational hotspots (de la Chapelle, A. 2004). The intragenic location of the mutation does not affect the phenotype (de la Chapelle, A., 2004), but *MSH2* is more frequently mutated than *MLH1* in patients with extracolonic tumours and *MLH1* mutations have a higher expression of CRC than *MSH2* mutations (Lynch, H.T. *et al*, 2006). *MSH6* accounts for approximately 10% of HNPCC-associated germline mutations and may result in a later age of CRC onset compared with *MSH2* or *MLH1* mutations (Peltomäki, P., 2005). Women with a *MSH6* mutation have a high risk of endometrial cancer (Wijnen, J. *et al*, 1999). Truncating, splicing and missense mutations have been identified in the *MSH2*, *MLH1* and *MSH6* genes (de la Chapelle, A., 2004). Mutations in *PMS2* are less frequent and were thought to result in severe, early onset HNPCC often associated with brain tumours (Turcot syndrome) (de la Chapelle, A., 2004). Recently however, Hendriks *et al* (2006) reported that patients with a germline *PMS2* mutation show a milder phenotype than those with mutations in either *MLH1* or *MSH2* (Hendriks, Y.M. *et al*, 2006). Mutations in *MLH3* have been reported (Wu, Y. *et al*, 2001) but their contribution to HNPCC is unclear (Cannavo, E. *et al*, 2005). Germline mutations in *MSH2*, or less frequently in *MLH1* (Ponti, G. & Ponz de Leon, M., 2005) and *MSH6* (Arnold, A. *et al*, 2007), can cause Muir-Torre syndrome and mutations in *MSH2*, *MLH1* or *PMS2* are associated with Turcot syndrome (Lucci-Cordisco, E. *et al*, 2003).

Homozygotes for either *MSH2* (Whiteside, D. *et al*, 2002), *MLH1* (Wang, Q. *et al*, 1999) or *PMS2* (De Vos, M. *et al*, 2004) point mutations are rare and have a severe phenotype with haematological malignancies and some skin characteristics of neurofibromatosis type 1 (NF1). Compound heterozygotes for *PMS2* point mutations have a phenotype consistent with Turcot syndrome (De Rosa, M. *et al*, 2000). Recently, Will *et al* (2007) described a large (400kb) homozygous *PMS2* deletion in a patient with 10 CRCs, 35 colorectal adenomas and duodenal cancer in his twenties. The patient had no brain tumour but had mental retardation and café-au-lait spots. This case shows that gross homozygous mutations in a MMR gene can result in a severe phenotype uncharacteristic of HNPCC (Will, O. *et al*, 2007).

1.3.3.4.2 Somatic mutations

A somatic inactivating mutation in the inherited wild-type allele of the MMR gene results in defective MMR (Yuen, S.T. *et al*, 2002). Errors normally corrected by this pathway are left unrepaired, leading to a mutator phenotype which increases the chance of mutations occurring in growth regulatory genes (Ionov, Y. *et al*, 1993; Fishel, R. *et al*, 1993). Microsatellite instability (MSI) is a consequence of defective MMR (Thibodeau, S.N. *et al*, 1993) and colorectal tumourigenesis in HNPCC occurs via the MSI pathway, with genes such as *TGF-βRII* or *BAX* often somatically mutated (Söreide, K *et al*, 2006). Tumours deficient in MLH1, MSH2 or PMS2 show MSI at no less than 40% of loci and are termed MSI-high (Peltomäki, P., 2005; Umar, A. *et al*, 2004), whereas tumours from patients with *MSH6* mutations can have lower levels of instability (Peltomäki, P., 2005). Some sporadic tumours also display MSI but unlike HNPCC tumours, their mismatch repair deficiency is most often the result of hypermethylation of the *MLH1* promoter (Kuismanen, S.A. *et al*, 2000).

1.4 Hypotheses and aims

1.4.1 FAP and MAP as causes of adenomatous colorectal polyposis

Previous studies of unrelated colorectal polyposis patients have indicated that a large proportion (274/660, 42% (Aretz, S. *et al*, 2006); 37/60, 62% (Aceto, G. *et al*, 2005)) are not accounted for by germline *APC* or *MUTYH* mutations. Families recorded on the Wales Polyposis Register were screened for APC and MUTYH mutations in order to identify those in which novel polyposis genes could be sought. Novel candidate genes were assessed for the presence or absence of high penetrance polyposis alleles in *APC* and *MUTYH* mutation-negative cases from the Wales Polyposis Register and collaborating registers in the UK.

1.4.2 Characterization of MAP and the MUTYH mutational spectrum

MUTYH mutations in MAP cases from three European countries were characterized to better define the mutation spectrum and identify population-specific changes. Germline biallelic *MUTYH* mutations result in a BER defect which could lead to an increased risk of extracolonic manifestations in MAP. Genotypes and clinical features of MAP patients from three European centres were assessed to define the MAP phenotype and establish genotype-phenotype correlation.

1.4.3 Cancer risk in MUTYH heterozygotes

Previous studies have provided evidence for an increased CRC risk in monoallelic *MUTYH* mutation carriers (Farrington, S.M. *et al*, 2005; Peterlongo, P. *et al*, 2006) but other reports have contested this (Webb, E.M. *et al*, 2006; Balaguer, F. *et al*, 2007). To clarify whether *MUTYH* heterozygosity has implications for health, 350 monoallelic *MUTYH* mutation carriers identified through three European polyposis registers were included in a retrospective study of CRC, other cancer and mortality.

Chapter Two

Materials and Methods

2.1 Suppliers

Materials and equipment were purchased from the following companies: AB gene (Surrey, UK), Ambion (Cambridgeshire, UK), Amersham Biosciences (Buckinghamshire, UK), Anachem (Bedfordshire, UK), Applied Biosystems (Cheshire, UK), Beckman Coulter (Buckinghamshire, UK), Bibby Sterilin (Staffordshire, UK), Bioquote (Yorkshire, UK), Biorad (Hertfordshire, UK), Corning CoStar (The Netherlands), Difco Laboratories Ltd (Surrey, UK), DuPont Instruments (Hertfordshire, UK), Eppendorf, (Cambridgeshire, UK), Eurogentec (Hampshire, UK), Fisher Scientific (Leicestershire, UK), Fluka Biochemika (Dorset, UK), IKA (Staufen, Germany), Invitrogen Life Technologies (Strathclyde, UK), Jencons-PLS (Bedfordshire, UK), Kendro Laboratory Products (Hertfordshire, UK), Labtech International (East Sussex, UK), Leica (Wetzlar, Germany), Millipore (Hertfordshire, UK), MJ Research (Massachusetts, USA), MRC-Holland (Amsterdam, The Netherlands), MWG-Biotech (Buckinghamshire, UK), New England Biolabs (Hertfordshire, UK), Palm (Germany), PreAnalytiX (Hombrechtikon, Switzerland), Promega (Hampshire, UK), Qiagen (West Sussex, UK), Roche Biochemicals (East Sussex, UK), Sarstedt (Germany), Sartorius (Epsom, UK), Sigma Ltd (Dorset, UK), StarLab (Milton Keynes, UK) and Thermo Electron Corporation (Hampshire, UK).

2.2 Materials

2.2.1 Chemicals

General chemicals were supplied from Sigma Ltd unless otherwise stated.

2.2.2 Nucleic acid extraction and purification

QIAamp DNA Micro and QIAprep Miniprep kits were supplied by Qiagen. PAXgene Blood RNA tubes and PAXgene Blood RNA kit were from PreAnalytiX. TURBO DNA-*free* kit, RNase*Zap*, DNA*Zap* and DEPC-treated water were obtained from Ambion.

2.2.3 First strand synthesis

SuperScript III First-Strand Synthesis System for RT-PCR was supplied by Invitrogen Life Technologies.

2.2.4 PCR

AmpliTaq Gold DNA polymerase and GeneAmp PCR buffer were from Applied Biosystems. Individual deoxynucleotidetriphosphates (dNTPs) were purchased from Amersham Biosciences. HPSF purified oligonucleotide primers were supplied by MWG-Biotech or Eurogentec. Dimethyl Sulfoxide (DMSO) was obtained from Sigma Ltd.

2.2.5 PCR purification

Exonuclease I and Shrimp alkaline phosphatase were purchased from New England Biolabs and Amersham Biosciences respectively. QIAquick PCR purification kit was supplied by Qiagen

2.2.6 Multiplex ligation-dependent probe amplification (MLPA)

SALSA P043 APC MLPA kit was supplied by MRC-Holland.

2.2.7 Electrophoresis

Multipurpose agarose was obtained from Roche. New England Biolabs and Invitrogen Life Technologies supplied the 100bp DNA ladder and 1kb DNA ladder respectively. Ethidium bromide was supplied by Fluka Biochemika.

2.2.8 Sequencing and fluorescent product sizing

BigDye Terminator Cycle Sequencing Kit (Version 1.1 or 3.1), POP6 polymer, HiDi formamide and Genescan 500-ROX size standard were purchased from Applied Biosystems. Millipore supplied Montage SEQ_{96} sequencing reaction clean up kits and Sigma provided capillary electrophoresis running buffer.

2.2.9 Restriction enzymes

All restriction endonucleases were supplied, with the appropriate buffer and BSA, if required, by New England Biolabs.

2.2.10 Cloning

pGEM-T Easy vector system I was purchased from Promega and subcloning efficiency DH5α chemically competent *E.coli* cells were obtained from Invitrogen Life Technologies. Tryptone, yeast extract and agar were supplied by Difco Laboratories Ltd. Ampicillin, X-gal (5-bromo-4-chloro-3-indoyl-D-galactoside) and IPTG (isopropyl-β-D-thio-galactopyranoside) were obtained from Sigma Ltd.

2.2.11 Clinical Material

All tissue and blood samples were obtained with patient consent and ethical approval for research.

2.3 Equipment

2.3.1 Plastics and glassware

Sterile tips for Gilson pipettes were supplied by Bioquote or StarLab. RNase- and DNase-free sterile barrier tips were obtained from Promega. Distritips and sterile tips for multi-channel pipettes were bought from Anachem. Sterile 25ml stripettes were obtained from Corning CoStar. 0.65ml, 1.5ml and 2.0ml plastic eppendorfs were purchased from Bioquote and 1.5ml microcentrifuge tubes were supplied by Fisher Scientific. Thermo Electron Corporation provided 96 well Thermo-fast PCR reaction plates. Adhesive PCR sealing sheets, 0.2ml Thermo-strip tubes and 96 well Thermo-fast skirted detection plates were purchased from ABgene. Sterile universals and petri dishes were obtained from Bibby Sterilin and Sarstedt respectively. Glass flasks and beakers were provided by Jencons-PLS or Fisher Scientific.

2.3.2 Thermocycling

Thermocycling was carried out using a MJ Research DNA Engine Tetrad PTC-225 or an Applied Biosystems GeneAmp 9700.

2.3.3 Electrophoresis

Electrophoresis was carried out using 96 well gel apparatus from ABgene or the Horizon 11.14 gel tank from Invitrogen Life Technologies. Power packs were

purchased from BioRad. Visualisation of ethidium bromide stained gels was carried out using a Biorad GelDoc200 or Biorad GelDoc XR transluminator. Capillary electrophoresis of fluorescent sequencing or MLPA products was carried out using an ABI 3730 DNA Analyser or an ABI 3100 Genetic Analyser.

2.3.4 Other equipment

DuPont Instruments supplied the Sorvall RT6000B refrigerated centrifuge and the Sorvall LegendRT centrifuge was provided by Kendro Laboratory Products. The minishaker MS2 was purchased from IKA. The Biofuge pico centrifuge and the Vortex-Genie 2 were obtained from Jencons-PLS. Sartorius supplied the BP1200 top pan balance. The Biomek FX robot was provided by Beckman Coulter. Quantitation of DNA and RNA was carried out using the NanoDrop ND-1000 spectrophotometer (Labtech International).

2.4 Sequence data

Genbank (http://www.ncbi.nlm.nih.gov/Genbank/) accession numbers for genes analysed in this project are as follows:

	Genomic	mRNA
APC	NC_000005	NM_000038
MUTYH	NC_000001	NM_012222
MBD4	NC_000003	NM_003925
NUDT1	NC_000007	NM_198952
NUDT5	NC_000010	NM_014142
NTHL1	NC_000016	NM_002528
POLL	NC_000010	NM_013274
SMUG1	NC_000012	NM_014311
MPG	NC_000016	NM_002434
TDG	NC_000012	NM_003211
UNG	NC_000012	NM_080911, NM_003362
APE1	NC_000014	NM_001641

2.5 Bioinformatic tools

BLAST searches (http://www.ncbi.nlm.nih.gov/blast/) were carried out against DNA sequences from Genbank. Multiple sequence alignments were carried out

using Clustal W (http://www.ebi.ac.uk/clustalw/). Reported single nucleotide polymorphisms and pathogenic mutations were identified from the Human Gene Mutation Database (http://www.hgmd.cf.ac.uk/), dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/) and *APC* gene database (Soussi,T., personal communication). Splice site prediction was carried out using NetGene2 (http://www.cbs.dtu.dk/services/NetGene2) or SpliceSiteFinder (http://violin.genet.sickkids.on.ca/~ali/splicesitefinder.html) and ESEfinder v3.0 (http://rulai.cshl.edu/tools/ESE/)was used to predict putative exonic splicing enhancers (ESEs). Protein structures were obtained from the RCSB Protein Data Bank (http://www.rcsb.org/pdb/) and viewed using RasMol v2.6-β-2 (http://openrasmol.org/).

2.6 Methods

All solutions were made using MilliQ water and autoclaved at 15lb/sq.in at 121°C for 40 min where necessary.

2.6.1 General reagents

TAE buffer: 40mM Tris-acetate, 1mM EDTA pH8

2.6.2 Nucleic acid extraction

2.6.2.1 DNA extraction from paraffin-embedded tissue

DNA was extracted from paraffin-embedded tissue cores using the QIAamp DNA Micro kit, according to the manufacturer's instructions. To 2-5 tissue cores (diameter 6mm), 15µl SDS-containing lysis buffer (Buffer ATL; contents trade secret [CTS]) and 10µl proteinase K solution were added. The sample was mixed by vortexing then incubated at 56°C for 16 hours. To aid efficient lysis, the sample was occasionally vortexed during incubation. An additional 10µl proteinase K solution was added to the sample which was vortexed then incubated at 56°C for a further 3 hours. 25µl buffer ATL and 50µl buffer AL (CTS) were added to the sample which was then mixed by pulse-vortexing for 15 seconds to obtain a homogeneous solution. To this solution 50µl 100% ethanol was added, the sample was mixed by pulse-vortexing and then incubated at room temperature for 5 minutes. This addition of Buffer AL and

ethanol allows optimal adsorption of DNA to the silica gel membrane of the column. The solution was then applied to the QIAamp MinElute spin column and centrifuged at 6000 x g for 1min to facilitate DNA binding. The column was then transferred to a clean collection tube, and washed to remove any residual contaminants. 500 µl of the first wash buffer, AW1 (CTS) was added, and the column re-centrifuged for 1min at 6000 x g. The eluate was discarded before a second wash step was carried out by adding 500µl buffer AW2 (contains ethanol, CTS) then centrifuging for 1 min at 6000 x g. Any residual carryover of the ethanol-containing wash buffer would inhibit downstream reactions, so the column was centrifuged at 13,000 x g for 3 min to ensure its complete removal. DNA was eluted into a clean microcentrifuge tube by applying 30µl distilled water to the column, incubating at room temperature for 5 min and centrifuging at 13,000 x g for 1 min. Samples were stored at -20°C.

2.6.2.2 RNA extraction

RNA was extracted using the PAXgene Blood RNA system and TURBO DNA-*free* kit, according to the manufacturers' instructions. All equipment was treated with RNase*Zap* and DNA*Zap* then rinsed with DEPC-treated water. This removed any ribonucleases that would degrade the RNA and any DNA that would contaminate the sample. Blood (2.5ml) was collected from the patient into a PAXgene Blood RNA tube which already contained 6.9ml additive (CTS). The additive stabilises the *in vivo* gene transcription profile of a sample by reducing *in vitro* RNA degradation and minimising gene induction, allowing accurate detection and quantification of gene transcripts. The blood and additive were mixed by inverting the tube 8-10 times which was then kept upright at room temperature for 3.5 hours during transportation. The sample was then either frozen at -20°C for later use or incubated at room temperature for a further 17 hours before processing. If the sample had been frozen, it was left to thaw and incubate at room temperature for 17 hours before processing. After incubation the sample was centrifuged at 3000 x g for 10 minutes to pellet the nucleic acids. The supernatant was discarded and the pellet washed with the addition of 4ml RNase-free water. The sample was vortexed to dissipate the pellet then centrifuged again at 3000 x g for 10 minutes. The supernatant

was discarded and 350µl resuspension buffer (CTS) added to the pellet. The sample was vortexed until the pellet was visibly dissolved then transferred to a 1.5ml microcentrifuge tube. 300µl binding buffer (contains guanidine thiocyanate, CTS) and 40µl proteinase K were added to the sample which was vortexed for 5 seconds then incubated for 10 minutes at 55°C in a shaker-incubator at 450rpm. This resulted in digestion of the protein present. The solution was then applied to a PAXgene Shredder spin column and centrifuged for 3 minutes at 13,000 x g to homogenize the cell lysate and remove residual cell debris. The supernatant of the flow-through fraction was transferred to a clean 1.5ml microcentrifuge tube without disturbing the pellet in the processing tube. To this solution 350µl 100% ethanol was added to adjust the binding conditions and the sample was mixed by vortexing. 700µl of the solution was applied to a PAXgene RNA spin column and centrifuged for 1 minute at 13,000 x g to facilitate total RNA binding. The eluate was discarded and the column transferred to a clean collection tube. The remainder of the sample was applied to the column and centrifuged for 1 minute at 13,000 x g. The column was then transferred to a clean collection tube and washed to remove any residual contaminants. 350µl wash buffer 1 (contains ethanol and guanidine thiocyanate, CTS) was added and the column centrifuged for 1 minute at 13,000 x g. DNase I treatment of the column was carried out between wash steps to remove trace amounts of bound DNA. 10µl DNase I stock solution (contains bovine DNase I, 2.7 U/µl, CTS) was mixed with 70µl DNA digestion buffer (CTS) by gentle flicking as DNase I is sensitive to physical denaturation. This mixture was then applied to the column and incubated at room temperature for 15 minutes. The column was washed again with 350µl wash buffer 1 then centrifuged at 13,000 x g for 1 minute. The column was transferred to a clean collection tube and washed with a second buffer by adding 500µl wash buffer 2 (contains ethanol, CTS) then centrifuging for 1 minute at 13,000 x g. The column was transferred to a clean collection tube and washed again with wash buffer 2 (500µl) then centrifuged for 3 minutes at 13,000 x g. Any residual carryover of the ethanol-containing wash buffer would inhibit downstream reactions, so the column was transferred to a clean collection tube then centrifuged at 13,000 x g for 1 min to ensure its complete removal. RNA was

eluted into a clean 1.5ml microcentrifuge tube by applying 40µl elution buffer (CTS) to the column then centrifuging at 13,000 x g for 1 min. This elution step was repeated to obtain the maximum RNA yield. The RNA was then denatured ready for downstream applications by incubating at 65°C for 5 minutes then chilling immediately on ice.

Chai, V. *et al* found the on-column DNase I treatment insufficient to remove low levels of DNA contamination so in order to remove any DNA still remaining in the RNA sample, it was treated again using the TURBO DNA-*free* kit. 4µl 10X TURBO DNase buffer (contains Tris, NaCl, CTS) and 1µl TURBO DNase (2 U/µl, CTS) were added to 40µl RNA (1.2µg to 4.5µg). The sample was mixed gently then incubated at 37°C for 30 minutes. A further 1µl TURBO DNase was added and the sample incubated again for 30 minutes at 37°C. 8µl DNase inactivation reagent (CTS) was mixed with the sample and incubated for 2 minutes at room temperature with occasional vortexing to remove the DNase and divalent cations which can catalyze heat-mediated RNA cleavage. The sample was centrifuged at 10,000 x g for 1.5 minutes to pellet the DNase inactivation reagent. The supernatant (RNA) was transferred to a clean tube and stored at -80°C.

2.6.3 Quantification of nucleic acids

The concentration of DNA and RNA samples was determined by UV spectrophotometry at wavelengths of 260nm and 280nm to quantify the amount of DNA or RNA present and allow protein content to be established. An absorbance ratio of 1.8 at 260nm:280nm indicates high sample purity.

2.6.4 First strand synthesis

In order to amplify messenger RNA, which makes up between 1% and 5% of total RNA, a complementary DNA strand is required to provide a template for PCR. Random 6bp oligonucleotides bind along the RNA and prime the reverse transcription reaction. This first strand synthesis was carried out using the SuperScript III first strand synthesis system for RT-PCR, according to the manufacturer's instructions. A 10µl reaction containing 0.24µg to 0.58µg total

RNA, 1mM dNTPs and 50ng random hexamers was incubated at 65°C for 5 minutes and then chilled on ice. 10X RT buffer (200mM Tris-HCl, pH 8.4, 500mM KCl), 5mM $MgCl_2$, 2μl 0.1M dithiothreitol, 40U RNaseOUT recombinant ribonuclease inhibitor and 200U SuperScript III reverse transcriptase were then added. The reaction was incubated for 10 minutes at 25°C followed by 50 minutes at 50°C. The reaction was terminated by denaturation of the enzyme at 85°C for 5 minutes and then was chilled on ice. To increase the sensitivity of cDNA amplification by PCR, the RNA template was removed from the cDNA:RNA hybrid molecule by adding 2U *E.coli* RNase H and incubating at 37°C for 20 minutes. The cDNA reaction was stored at -20°C.

2.6.5 Primer design

Oligonucleotides were designed using Oligo Version 4 software (Molecular Biology Insights Inc., USA). The melting temperatures of the primer pair were within 2°C of each other and oligonucleotides were more stable at their 5' regions if possible. Primers were between 18 and 25 nucleotides in length, lacked repetitive motifs and had little predicted dimerization or secondary structure formation.

2.6.6 Polymerase chain reaction (PCR)

The polymerase chain reaction allows the specific *in vitro* amplification of defined DNA target sequences within a source of DNA. This occurs in an exponential manner. Double stranded DNA templates are heat denatured and oligonucleotides bind specifically to complementary target sites on each strand. Thermostable DNA polymerases extend the primers in the 5' to 3' direction by incorporating dNTPs to create a complementary DNA strand. This cycle is repeated 20-40 times to produce up to 5μg DNA (Mullis et al, 1986).

For standard PCR, 25ng template DNA, 0.25mM dNTPs, 25pmoles forward and reverse primers, 10X GeneAmp PCR buffer (100mM Tris-HCl, pH8.3, 500mM KCl, 15mM $MgCl_2$, 0.01% w/v gelatin), and 0.5U AmpliTaq Gold DNA polymerase were used in a total volume of 20μl. Cycling conditions were 95°C for 12 minutes, followed by 35 cycles of 95°C for 30 seconds, annealing

temperature (52°C to 62°C) for 30 seconds, 72°C for 30 seconds, and a final elongation step of 72°C for 10min.

A 20μl RT-PCR reaction contained 3μl cDNA, 0.25mM dNTPs, 25 pmoles forward and reverse primers, 10X GeneAmp PCR buffer (100mM Tris-HCl, pH8.3, 500mM KCl, 15mM $MgCl_2$, 0.01% w/v gelatin), and 0.5U AmpliTaq Gold DNA polymerase. Cycling conditions were 95°C for 12 minutes, followed by 35 cycles of 95°C for 30 seconds, 58°C for 30 seconds, 72°C for 30 seconds, and a final elongation step of 72°C for 10min.

2.6.7 Amplification refractory mutation system (ARMS)

ARMS is an allele-specific PCR technique that detects known single nucleotide point mutations. The normal and mutant alleles are amplified in separate PCR reactions. Allele-specific primers of 19-23bp were designed with the 3' base complementary to either the normal or mutant allele. A mismatched base was included at the third to last position of the 3' end to ensure that both primers had an equal binding affinity and a similar yield of PCR products was obtained from both reactions. Pyrimidine/pyrimidine and purine/purine mismatches are more destabilising than pyrimidine/purine mismatches. If the allele-specific base of the primer bonds to its complementary base with 3 hydrogen bonds, a strongly destabilising mismatch was inserted (the same base as in the target DNA sequence). If the allele-specific base of the primer bonds to its complementary base with 2 hydrogen bonds, a weaker destabilising mismatch was inserted (the opposite type of base as in the target DNA sequence and that is not the complementary one). The common primer was designed as described in section 2.6.5. All ARMS reactions also included a pair of control primers to prevent false negative results. The annealing temperature and relative concentrations of ARMS and control oligonucleotides were optimised to ensure specific amplification. Heterozygous samples produce a PCR product in both ARMS reactions, whereas samples that are homozygous generate products only during the PCR specific for that allele.

2.6.8 Multiplex ligation-dependent probe amplification (MLPA)

MLPA is a PCR-based method that can be used to detect copy number changes of single exons in one reaction. Instead of amplifying sample DNA, probe oligonucleotides are hybridized to the target sequences and a ligation reaction occurs resulting in a copy of each target sequence present in the sample. It is these ligated probes that are then amplified in a single multiplex PCR reaction (Figure 2.1). Each probe consists of two oligonucleotides which hybridize to adjacent target sequences. Only matching oligonucleotides are ligated together to produce a probe and one of the pair contains a non-hybridizing stuffer sequence. The length of this sequence is different in each different probe, allowing amplification products to be separated by capillary electrophoresis. Every probe ends with the same PCR primer sequences so only one pair of universal primers is needed to amplify all probes in a single reaction. Non-ligated probe oligonucleotides cannot be exponentially amplified as they do not contain both PCR primer sequences. Once the amplification products have been identified and quantified using capillary electrophoresis, they are compared to control samples to determine their copy number (Schouten, J.P. *et al*, 2002).

Known point mutations can also be detected by MLPA. For every mutation there are two possible probes that could be amplified. Each probe consists of two ligated oligonucleotides, one M13 derived oligonucleotide which is common to both probes and one allele specific synthetic oligonucleotide. The probes end with the same PCR primer sequences so are amplified in one reaction using a pair of universal primers. The allele specific oligonucleotides differ at their 3' ends in order to be complementary to either the mutant or wild type allele. Their specific target sequences are adjacent to the target sequence of the common oligonucleotide so when hybridized to their complementary sequence they can be ligated with the common oligonucleotide to form a probe. Mismatched oligonucleotides, for example an oligonucleotide complementary to the mutant allele mismatched to a wild type target, cannot be ligated and therefore exponentially amplified as they lack one of the PCR primer sequences. The allele specific oligonucleotides differ in length slightly so the probes can be separated and identified using capillary electrophoresis. If the

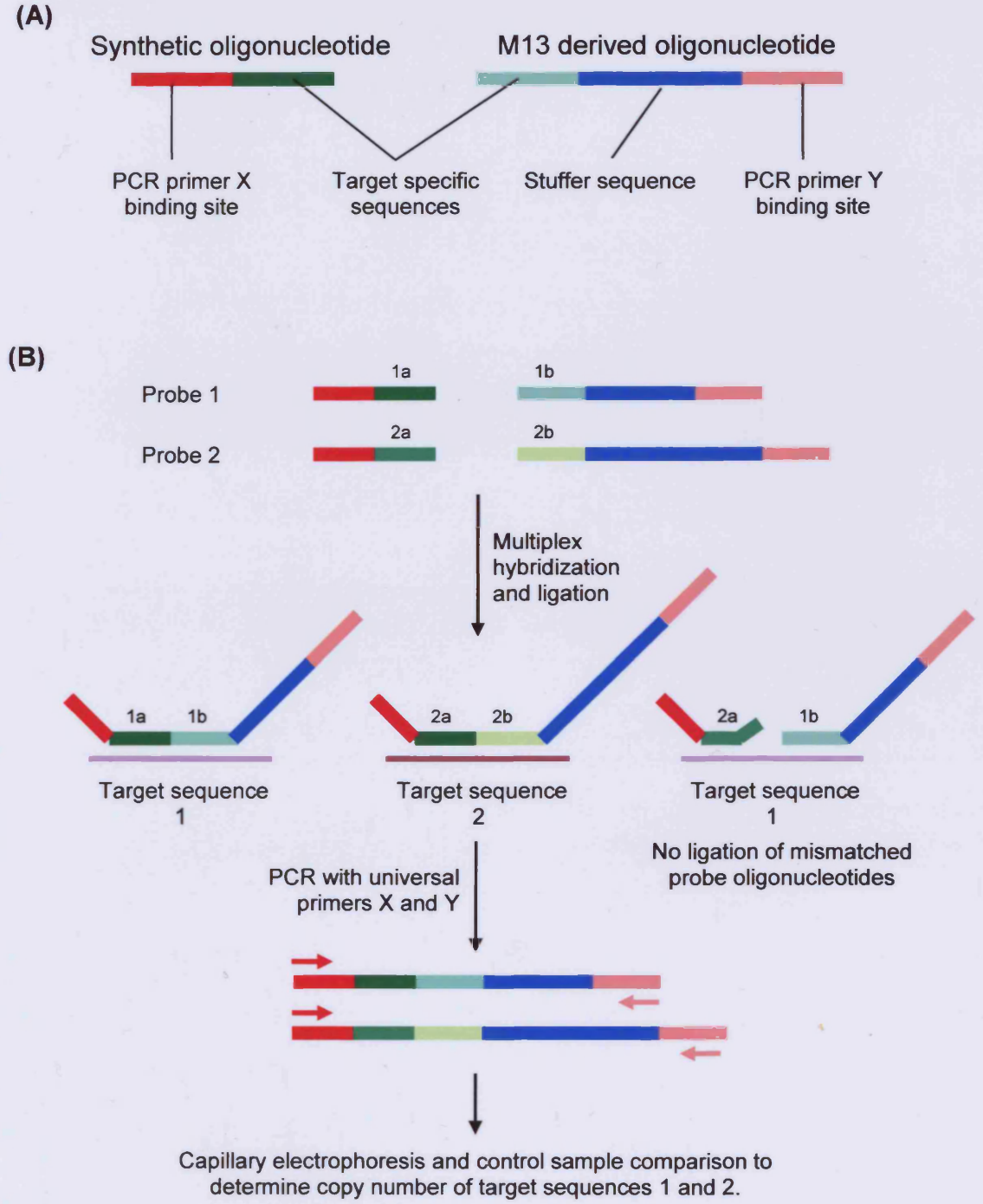


Figure 2.1. Multiplex ligation-dependent probe amplification (MLPA).
(A) Each probe consists of two oligonucleotides. A non-hybridizing stuffer sequence is incorporated into the M13 derived oligonucleotide. All probes have the same PCR primer sequences at their ends. (B) Multiple probe oligonucleotides are hybridized to sample DNA and bind to their complementary target sequences. Matching oligonucleotides are ligated together and the probes produced are amplified by PCR using the same pair of universal primers. The amplification products undergo capillary electrophoresis and control sample comparison to determine their copy number.

sample is heterozygous for the mutation two products will be seen. If the sample is either wild type or homozygous mutant only one product will be seen but the mutant and wild type products will differ in size and so can be identified.

MLPA was performed using the SALSA P043 APC MLPA kit, according to the manufacturer's instructions. 200 ng DNA was denatured by incubation at 98°C for 5 minutes. The sample was cooled to 25°C then 1.5µl SALSA probe mix (CTS) and 1.5µl MLPA buffer (1.5M NaCl, EDTA, CTS) were added. The mixture was incubated for 1 minute at 95°C then 16 hours at 60°C to allow hybridization of the probes to their target sequences. 3µl Ligase-65 buffer A (contains NAD), 3µl Ligase-65 buffer B (contains $MgCl_2$), 1µl Ligase-65 and 25µl sterile water were mixed together then chilled on ice. The hybridized sample was cooled to 54°C and the Ligase-65 mix was added. The reaction was incubated for 15 minutes at 54°C to allow ligation of probes then heated to 98°C for 5 minutes to inactivate the ligase. A 40µl PCR mix contained 10X SALSA PCR buffer (200mM Tris buffer pH 8.5; 500mM KCl; 16mM $MgCl_2$; non-ionic detergents), 2µl SALSA PCR primers (contained FAM labelled forward primer and dNTPs), 2µl SALSA enzyme dilution buffer (contains buffer, NaCl, non-ionic detergents) and 0.5µl SALSA Polymerase (a thermostable polymerase). 10µl ligation reaction was added to the PCR mix and the cycling conditions were 95°C for 10 seconds, followed by 33 cycles of 95°C for 30 seconds, 60°C for 30 seconds, 72°C for 1 minute, and a final elongation step of 72°C for 20min. The fluorescent PCR product was diluted 1 in 4 with sterile water and 2µl of this dilution was added to 10µl highly deionised formamide and 0.5µl ROX GS500 size standard and denatured at 95°C for 3 minutes. The reaction was cooled rapidly on ice to prevent the formation of heteroduplexes before running on the 3100 Genetic Analyzer. Products were visualised and quantified using GeneScan software. Genotyper software was then used to identify each peak as a particular probe amplification product based on the sizes of the expected MPLA products. The peak areas for these products were exported to an Excel file and normalised. Dosage quotients for each product were calculated by comparing the normalised test sample peak area to the average normalised peak area of 5 normal control samples. If the dosage

quotient is 1.0, there are two copies of the product in test sample i.e. normal. If the dosage quotient is 0.5, there is only one copy of the product in test sample i.e. deleted. If the dosage quotient is 1.5, there are three copies of the product in test sample i.e. duplicated. Internal quality control for each sample was assessed by examining the variation between control probe amplification products in that sample.

2.6.9 Agarose gel electrophoresis

Agarose gels were prepared with 1X TAE buffer to a concentration of 2.0% to 2.5% w/v. The higher the gel concentration, the better the separation of small DNA fragments because of the smaller pore size. 0.05μg/ml ethidium bromide was included in the gel as it intercalates DNA, enabling the DNA to be visualised under UV light at a wavelength of 300nm. 3μl loading dye (15% w/v ficol, 10mM Tris pH8, 1mM EDTA, 0.2% orange G) was added to each sample before loading and electrophoresis was performed in 1X TAE buffer at 100V. 1 kb or 100bp DNA ladders were used to allow fragment sizing. DNA was visualised on a UV transluminator and photographed using the GelDoc 200 or XR system.

2.6.10 Restriction endonuclease digestion

NEBcutter software (http://tools.neb.com/NEBcutter2/index.php) version 2.0 was used to identify restriction endonuclease recognition sites in DNA sequences. A total reaction volume of 20μl contained 17μl PCR product or 1μl plasmid DNA, 10X appropriate buffer, 0.1X bovine serum albumin (BSA) where required, and 2.5U restriction endonuclease. The reactions were incubated at 37°C, 50°C or 60°C for 6 hours and the products analysed using gel electrophoresis.

2.6.11 PCR purification

PCR products were purified using an ExoSap method or the QIAquick PCR purification kit. For the ExoSap method, 3μl of PCR product was purified by adding 10U exonuclease I and 0.5U shrimp alkaline phosphatase. The sample was incubated at 37°C for 15min to allow digestion of excess primers and

removal of phosphate groups from dNTPs before denaturation of the enzymes at 80°C for 15min.

Alternatively, the QIAquick PCR purification kit was used according to the manufacturer's instructions. 15µl of PCR product was mixed with 5 volumes of buffer PB (contains guanidine hydrochloride and isopropanol, CTS), applied to the silica gel membrane of the QIAquick column and centrifuged at 13,000 x g for 1 minute to facilitate DNA binding. Buffer PB provides the correct salt concentration and pH for adsorption of DNA to the membrane. The eluate was discarded and the column was washed with 750µl buffer PE (contains ethanol) and centrifuged for 1 minute at 13,000 x g to remove any residual contaminants. The flow through was discarded and the column was centrifuged again as any residual carryover of the ethanol-containing wash buffer would inhibit downstream reactions. The DNA was eluted into a clean microcentrifuge tube by applying 30µl buffer EB (10mM Tris.Cl, pH 8.5) to the column, incubating at room temperature for 5 min and centrifuging at 13,000 x g for 1 min.

2.6.12 Cycle sequencing

Sanger sequencing uses dideoxynucleotide triphosphates (ddNTPs) which lack the 3' hydroxyl group present in deoxyribose sugars. As a result of this, ddNTPs can be efficiently incorporated into a nascent strand by DNA polymerases but prevent further extension of the growing chain (Sanger et al, 1977). In automated sequencing the reaction can take place in a single tube because each ddNTP is labelled with a different fluorphore. The template DNA is denatured and bound by a single specific oligonucleotide. DNA polymerase extends this primer by incorporating either an unlabelled dNTP or a chain terminating ddNTP at each position. The relative concentrations of dNTPs and ddNTPs are such that the labelled products formed differ in size by one nucleotide. Capillary electrophoresis is used to separate the single stranded DNAs, with smaller fragments migrating fastest through the polymer and passing through the laser beam first. The emitted wavelength is detected and used to determine the ddNTP incorporated at a particular position. The order of nucleotides provides a sequence read of up to 500bp.

Sequencing reactions were performed using the BigDye Terminator Cycle Sequencing kit (Version 3.1 or 1.1) according to the manufacturer's instructions. A total reaction volume of 10μl contained 0.6-1.5μl purified PCR product (~5ng) or 1μl plasmid DNA (~250ng), 1μl Terminator Ready Reaction Mix (Labelled A, C, G and T dye terminators, dNTPs, AmpliTaq DNA polymerase FS, $MgCl_2$ and Tris-HCl buffer, pH 9.0), 1.6pmoles primer and 1.5μl BigDye terminator sequencing buffer (CTS). Cycle sequencing conditions were 25 cycles of 96°C for 10s, 50°C for 5s and 60°C for 3min 30s.

The Montage SEQ_{96} Sequencing Reaction Cleanup Kit was used to purify sequencing reactions. 20μl injection solution (CTS) was mixed with the reaction and transferred to a filter plate. A vacuum of 18 inches Hg was applied until the wells were empty, then a further 25μl injection solution was added and the vacuum applied again to ensure that all contaminating salts and unincorporated dye terminators were filtered out. Purified sequencing products were resuspended in 25μl injection solution by shaking for 12 min. Samples were run on an ABI 3100 or ABI 3730 Genetic Analyser.

2.6.13 Bacteriological methods

2.6.13.1 Bacteriological media and solutions

LB-Broth: 5g Bactotryptone, 2.5g yeast extract and 2.5g NaCl in 500ml dH_2O

LB agar: 5g Bactotryptone, 2.5g yeast extract, 2.5g NaCl and 8g Bactoagar in 500ml dH_2O

2.6.13.2 Ligation

Purified PCR products were ligated into the pGEM-T Easy vector which has thymidine overhangs to facilitate TA cloning. The reaction consisted of plasmid vector and insert at a molar ratio of 1:2, 2X rapid ligation buffer (60mM Tris-HCl, pH 7.8, 20mM $MgCl_2$, 20mM DTT, 2mM ATP, 10% PEG), and 3U T4 DNA ligase in a total volume of 10μl. The reaction was incubated at 4°C overnight.

2.6.13.3 Transformation

50µl *E. coli* subcloning efficiency DH5α competent cells were thawed on ice and 4µl ligation product was added. The cells were then incubated at 4°C for 20min before being subjected to a heat shock at 42°C for 45 seconds to semi-permeabilise the cell membrane and facilitate transfer of the plasmid into the cell. 950µl LB medium was added and the cells incubated at 37°C with agitation at 150rpm for 1.5 hr. 50µl bacteria were then diluted with an equal volume of LB medium and spread onto LB agar plates containing 50µg/ml ampicillin, 32µg/ml X-gal (5-bromo-4-chloro-3-indoyl-D-galactoside) and 32µg/ml IPTG (isopropyl-β-D-thio-galactopyranoside). Plates were incubated overnight at 37°C. White colonies are formed when the *lacZ* gene is insertionally inactivated and there is no β-galactosidase produced to convert the colourless X-gal substrate into its blue product. These colonies therefore contained recombinant plasmids and were streaked out on LB plates containing 50µg/ml ampicillin and incubated at 37°C overnight.

2.6.13.4 Plasmid preparations

The QIAprep miniprep kit was used to purify plasmids according to the manufacturer's instructions. 5ml of LB medium containing ampicillin (100µg/ml) were inoculated with single colonies and cultures grown for 16h at 37°C with shaking at 250rpm. Bacterial cells were pelleted by centrifugation at 10,000rpm for 3min and resuspended in 250µl Buffer P1 (100µg/ml RnaseA, 50mM Tris-HCl, 10mM EDTA). 250µl of the alkaline Buffer P2 (200mM NaOH, 1% SDS) was then added to lyse the cells. The sample was gently mixed to avoid shearing of chromosomal DNA and then incubated at room temperature for 5min. 350µl Buffer N3 (3M Potassium acetate) was added to neutralise the solution and provide the high salt concentration necessary to precipitate celllular debris, chromosomal DNA and proteins. The precipitate was pelleted by centrifugation at 13,000rpm for 10min and the supernatant was applied to the spin column. The column was then centrifuged at 13,000rpm for 1min to facilitate binding of plasmid DNA to the silica membrane. Any salts were removed by adding 750µl of the ethanol-based buffer PE (CTS) and centrifuging at 13,000rpm for 1min. The flow through was discarded and the column re-

centrifuged to ensure complete removal of the buffer before the DNA was eluted in 30µl buffer EB (10mM Tris.Cl; pH 8.5). This method typically yields 5-10µg DNA. Plasmids were digested using a restriction endonuclease, visualised using agarose electrophoresis and stored at -20°C.

Chapter Three

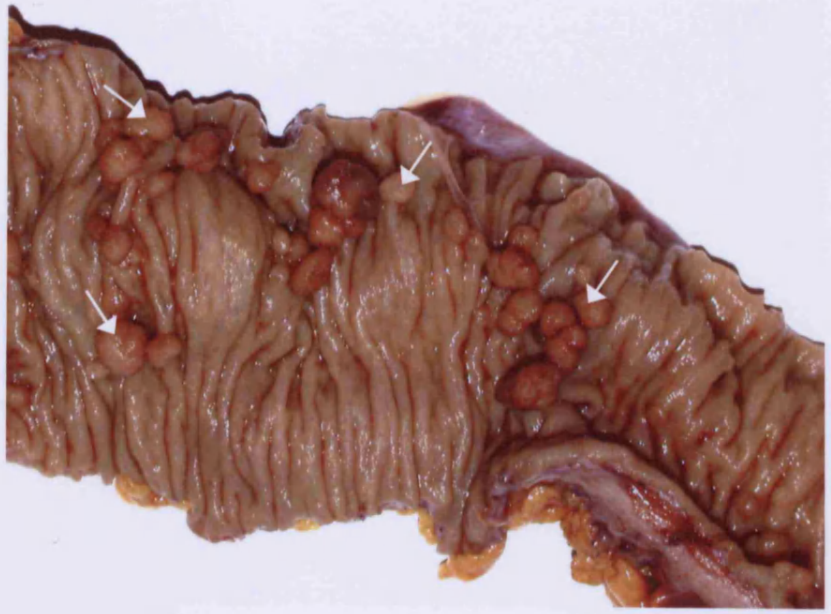
The contribution of germline *APC* and *MUTYH* mutations to colorectal polyposis

3.1 Introduction

Polyposis registers enable the coordination of genetic testing and clinical management in families affected by colorectal polyposis (Bülow, S. *et al*, 1995). Results from the Danish Polyposis Register show how such registration can help reduce CRC prevalence and improve the prognosis of FAP patients by effective use of prophylactic procedures (Bülow, S., 2003). Most UK regional polyposis registers cover geographically defined health regions, with the Wales Polyposis Register responsible for a population of approximately 3 million.

Cases with more than 100 macroscopic colorectal adenomas have traditionally been diagnosed with classical FAP and about a third of patients with 10-100 adenomas can be accounted for by AFAP (Nielsen, M. *et al*, 2007a; Friedl, W. *et al*, 2001). The MAP phenotype overlaps with that of FAP (especially AFAP, Figure 3.1) as cases typically have tens or hundreds of colorectal adenomas (Sampson, J.R. *et al*, 2003; Sieber, O.M *et al*, 2003; Nielsen, M. *et al*, 2005; Aretz, S. *et al*, 2006). The distinction between FAP and MAP would appear straightforward as the former is dominantly inherited whereas the latter is a recessively inherited disease. However, approximately 25% of cases with colorectal adenomatous polyposis occur sporadically or have affected siblings but unaffected parents (Bisgaard, M.L. *et al*, 1994). Such cases could result from *de novo APC* gene mutations or gonadal mosaicism in a clinically unaffected parent, but could also represent MAP. In addition, a small number of MAP cases have been reported with an apparently dominant family history of polyposis and/or CRC and thus could resemble a family affected by FAP (Nielsen, M. *et al*, 2007a; Aretz, S. *et al*, 2006; Jo, W-S. *et al*, 2005). One family was confirmed to have MAP cases in two generations as the spouse of an affected mother was a heterozygous *MUTYH* mutation carrier and their three children were all affected (Nielsen, M. *et al*, 2007a). An accurate diagnosis is important for appropriate genetic counselling and clinical management of

A)



B)

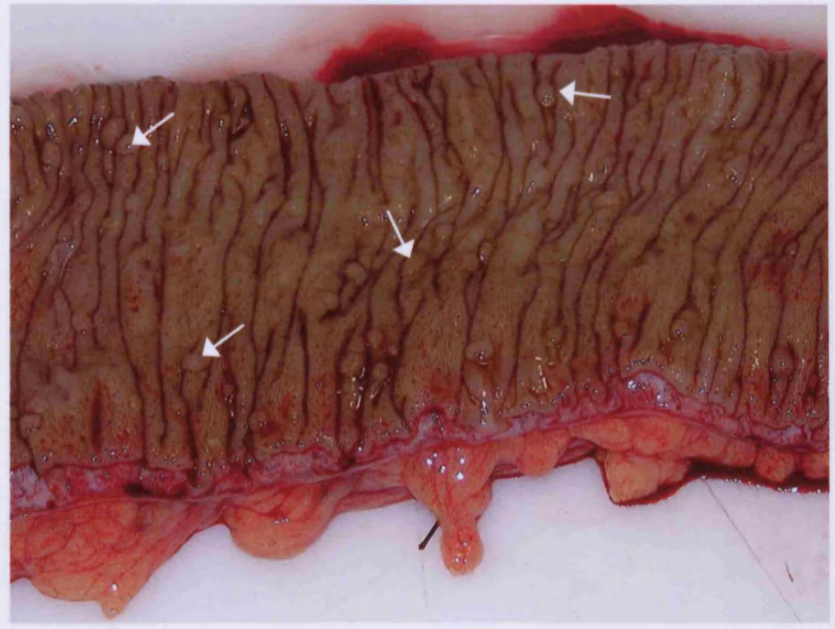


Figure 3.1 Colorectal adenomas in typical colectomy specimens from A) AFAP and B) MAP cases. The phenotypes of these diseases overlap making an accurate diagnosis on the basis of symptoms alone more difficult. Arrows indicate examples of adenomas.

affected families. An apparently sporadic case with FAP has a 50% risk to offspring and a low risk to siblings whereas MAP poses little risk to offspring but a 25% risk to siblings. Decisions regarding surveillance and preventative measures such as prophylactic colectomy are influenced by disease type. Diagnosis based on phenotype and family history alone can no longer be relied on and must be complemented by molecular genetic analysis.

3.2 Methods

3.2.1 Clinical samples

As of 1st January 2008, 107 apparently unrelated families affected by adenomatous colorectal polyposis were registered with the Wales Polyposis Register. The criterion for inclusion was one or more affected family members with at least ten colorectal adenomas at colonoscopy or colectomy. Blood DNA samples from affected members of 92 of the 107 families have been subject to a variety of forms of molecular analysis at the *APC* and *MUTYH* loci since the first identification of these genes in 1991 and 2002. Causative mutations had been identified in 57/107 families (Figure I), but in many of the remainder comprehensive analysis had not been completed since the progressive improvements in diagnostic specificity and sensitivity had been applied only sporadically in response to immediate diagnostic demands in specific families.

Prior to this project pathogenic germline *APC* mutations had been identified in 43 families, 36 of which were identified by the All Wales Medical Genetics Service (AWMGS) or by researchers of the Academic Department of Medical Genetics. The remaining seven *APC* mutations were identified in affected relatives living outside Wales. Biallelic germline *MUTYH* mutations had been identified in 14 families, with mutations in 11 of them identified by the AWMGS or by researchers of the Academic Department of Medical Genetics. The *MUTYH* mutations in three families were identified in affected relatives living outside of Wales. DNA samples were not available from eight families, four of whom had members clinically diagnosed with FAP. Three further families had members referred to the Wales Polyposis Register as they were at risk because of affected family members outside Wales (in whom no mutations had been identified) but were themselves found to be clinically unaffected. An index case from one family is currently being screened for mutations in *APC* by the AWMGS. Index cases from three families have been screened for mutations in *APC* by the AWMGS but both sequencing analysis and MLPA did not identify any pathogenic mutations so these samples are currently being screened for mutations in *MUTYH* ORF by this service. The initial step in this study was to complete comprehensive mutation analysis of *APC* and *MUTYH* in the remaining 35 families (shaded boxes in Figure I, appendix).

3.2.2 Screening for mutations in APC

Exons 1-15 of *APC* were PCR amplified as 35 fragments (Table 3.1) and screened for mutations using automated sequencing. Large deletions and duplications at the *APC* locus were screened for by MLPA. Two probes covered the promoter region, one probe detected changes within the 5' UTR and exons 1-14 were each covered by one probe (Table 3.2). Five probes identified changes exon 15, two of which covered the common small deletions at mutational hotspots codons 1061 and 1309. Thirteen control probes for other genes on different chromosomes were also included to assess the internal quality of the sample. If the standard deviation of the mean dosage quotients for all the control probes was less than 0.10 this was taken as an acceptable level of DNA quality. If the mean dosage quotient for an *APC* probe in a test sample was in the range 0.85-1.20, 0.35-0.65 or greater than 1.35 the region covered by the probe was considered to be normal, deleted or duplicated respectively. Analysis of MLPA data was performed using an Excel spreadsheet created by A. Wallace, National Genetics Reference Laboratory Manchester.

3.2.3 Screening for mutations in MUTYH

Cases in which a pathogenic *APC* mutation was not detected in the present study were screened for mutations in *MUTYH*. Exons 1-16 were PCR amplified and screened using automated sequencing as previously described (Al-Tassan, N. *et al*, 2002), with the exception of exon 3 which was PCR amplified using a primer set (Table 3.1) designed to cover exon 3 including the additional 33bp of coding sequence upstream of the originally described exon 3 (Ohtsubo, T. *et al*, 2000).

3.2.4 Assay for MUTYH sequence variant in controls

The putative splice site mutation c.1509+2 T>C was assayed by ARMS using normal (5'- TGGGAGAGGCCTAGGAGACCTA) or mutant (5'- TGGGAGAGGCCTAGGAGACATG) allele specific primers with a common forward primer (5'-TGAAGTTAAGGGCAGAACACCG). Primers were annealed at 58°C and generated a 120bp product. TSC2_Ex26F (5'-GAGCTTTGGCCCTTGGTGATA) and TSC2_Ex26R

Gene/Primer name	Primer sequence (5'→3')	Annealing temperature (°C)	Product size (bp)
APCx1F*	AGGTCCAAGGGTAGCCAAGG	55	197
APCx1R*	TAAAAATGGATAAACTACAATTAAAAG		
APCx2F*	AAATACAGAATCATGTCTTGAAGT	55	212
APCx2R*	ACACCTAAAGATGACAATTTGAG		
APCx3F	TGCTTAAAGCAATTGTTGTAT	55	286
APCx3R	GTACACAAGGCAATGTTTACT		
APCx4(2)F	TGCAGTCTTTATTAGCATTGTT	58	305
APCx4(2)R	CAGGCCTAAAGTTGGGTAA		
APCx5F	CTTTTTTGCTTTTACTGATTAACG	55	243
APCx5R	TGTAATTCATTTTATTCCTAATAGCTC		
APCx6F*	GGTAGCCATAGTATGATTATTTCT	55	204
APCx6R*	CTACCTATTTTTATACCCACAAAC		
APCx7F*	AAGAAAGCCTACACCATTTTTGC	57	238
APCx7R*	GATCATTCTTAGAACCATCTTGC		
APCx8F*	ACCTATAGTCTAAATTATACCATC	55	184
APCx8R*	GTCATGGCATTAGTGACCAG		
APCx9F	AGTCGTAATTTTGTTTCTAAACTC	55	457
APCx9R	CTTTGAAACATGCACTACGA		
APCx10F*	AAACATCATTGCTCTTCAAATAAC	55	216
APCx10R*	TACCATGATTTAAAAATCCACCAG		
APCx10AF	CCTGTATTCCAATGGATTGTAG	55	200
APCx10AR	CTATGTCCCCAGCAGTCAC		
APCx11F*	GATGATTGTCTTTTTCCTCTTGC	58	215
APCx11R*	CTGAGCTATCTTAAGAAATACATG		
APCx12F	AAGCTTGGCTTCAAGTTGTC	55	199
APCx12R	CAGAGTGAGACCCTGCCTC		
APCx13(2)F	TCCCAAAGTGATAGGATTACA	58	440
APCx13(2)R	AGGGAATCTCATGGCTAAA		
APCx14F*	TAGATGACCCATATTCTGTTTC	55	308
APCx14R*	CAATTAGGTCTTTTTGAGAGTA		
APCx15.1F	CAAAAGGAGATGTGGAATACT	55	420
APCx15.1R	AGCCAGGAGACATAATATTG		
APCx15.2F	CAGGAAGCATTATGGGACAT	55	422
APCx15.2R	GAGGAGCTGGGTAACACTGTAG		
APCx15.3F	AAGGCATCTCATCGTAGTAA	55	480
APCx15.3R	TTCCGACTTAGTGAAATTGTA		
APCx15.4F	CTGGGTCTACCACTGAATTAC	55	506
APCx15.4R	AGTTGTACTTTGATTCCTTGAT		
APCx15.5(1)F	CCTAGCCCATAAAATACATAGT	52	522
APCx15.5(1)R	ATAGGCTGATCCACATGAC		

APCx15.6F	AATCAAAATGTAAGCCAGTCT	55	486
APCx15.6R	AGCTGATGACAAAGATGATAA		
APCx15.7F	CTGCCACTTGCAAAGTTTC	52	502
APCx15.7R	ATTCCACTGCATGGTTCAC		
APCx15.8F	CAAAAGTGGTGCTCAGACAC	58	505
APCx15.8R	TTCCTGAACTGGAGGCATTA		
APCx15.9F	GACCTAAGCAAGCTGCAGTA	55	509
APCx15.9R	CCCCGGTGTAAAACTAACA		
APCx15.10F	CCCAGACTGCTTCAAAATTAC	55	444
APCx15.10R	CACGGAAAGGCTTGTGACT		
APCx15.11F	AAAACCTCATCTGTAACCATAC	55	445
APCx15.11R	TTCCTTCAATAGGCGTGTA		
APCx15.12F	AATTTAAATGCTGAGAGAGT	55	441
APCx15.12R	CCCTCTGTCTGGTATGTCT		
APCx15.13(1)F	TACCAGCCACACAGAACTAAC	58	527
APCx15.13(1)R	CACCCATATTTCTGGGACTAT		
APCx15.14(3)F	CACAGGGAGAACCAAGTAAAC	58	460
APCx15.14(3)R	AATGGTGATCCCAGAGAGAT		
APCx15.15F	GAAGGTGCAAATTCCATAGTA	55	486
APCx15.15R	GGCTGTTTGACCTTCACTAG		
APCx15.16F	AGGCAGGACAATGATTCATAT	55	487
APCx15.16R	GGCATTCTTGGATAAACCTG		
APCx15.17F	CCCTAGTACTGCTTCAACTAAG	55	517
APCx15.17R	TGGTCTTCCATCATTATACTCT		
APCx15.18F	AGACCAGCTTCTCCCACTAG	55	492
APCx15.18R	ACCTGAGGAAACGGTCTGA		
APCx15.19F	ACAAAGTAAAGAAAACCAAGTA	55	486
APCx15.19R	GGGGTACGTTCCACTATAG		
APCx15.20F	TCGCCTGAACTCCTTTATTC	55	512
APCx15.20R	TTCCAGAACAAAAACCCTCTA		
[illegible]			
MUTYHx3(1)F	GGCCAGAAACTTAGCCACAG	58	413
MUTYHx3(1)R	CAACCCCAGATGAGGAGTTAGG		

Table 3.1 Sequences of primers and PCR conditions used to screen *APC* and *MUTYH* exon 3. *Primers published by Groden, J. *et al*, 1991.

DQ control bands*	Not ligation dependent	64, 70, 76, 82
Synthetic control probe	2q14	94
Control probe	5q31	130
APC probe	Promoter	139
APC probe	Promoter	148
Control probe	14q13	157
APC probe	5' UTR	166
APC probe	Exon 1	175
Control probe	4q34	184
APC probe	Exon 2	193
APC probe	Exon 3	202
Control probe	9q21	211
APC probe	Exon 4	220
APC probe	Exon 5	229
Control probe	2q14.2	238
APC probe	Exon 6	247
APC probe	Exon 7	256
Control probe	2p12	265
APC probe	Exon 8	274
APC probe	Exon 9	283
Control probe	11p12	292
APC probe	Exon 10	305
APC probe	Exon 11	312
Control probe	8q24	319
APC probe	Exon 12	328
APC probe	Exon 13	337
Control probe	6q26	346
APC probe	Exon 14	355
APC probe	Start of exon 15	364
Control probe	14q32	373
APC probe	Middle of exon 15	382
APC probe	End of exon 15	391
Control probe	15q31	400
APC probe	Exon 10A	409
APC probe	Codon 1061 hotspot	418
APC probe	Codon 1309 hotspot	427
Control probe	16p13	436
Control probe	2p16	445

Table 3.2 *APC* MLPA probe positions and product sizes. *Indicate the amount of sample DNA; prominent peaks from these control bands represent a very small amount of sample DNA. Probes designed by A. Nygren at MRC-Holland.

(5'-CTCGCCCACAGGAGACCTAGA) primers were used as internal controls, producing a product of 388bp.

3.2.5 Diagnostic confirmation of mutations

All mutations identified in *APC* or *MUTYH* were confirmed by the AWMGS prior to application in clinical diagnosis or prediction.

3.2.6 Statistical analysis

Fisher's exact test was used to compare proportions of colorectal polyposis families with biallelic *MUTYH* mutations in Minitab 14.

3.3 Results

3.3.1 Contribution of APC mutations to colorectal polyposis in Wales

Of the 92 unrelated index cases screened for *APC* and/or *MUTYH* mutations (either previously or in the present study), pathogenic germline *APC* mutations were identified in 70 patients (76%). The mutations found in 27 of the 35 patients screened in the present study are summarized in Table 3.3. Overall, nonsense mutations contributed to 37% (26 cases) of *APC* mutations identified, 24 cases (34%) carried small deletions, 10% (7 cases) of mutations were accounted for by large deletions, small insertions were identified in 7 (10%) cases and splice site mutations were found in 3 (4%) cases. The precise identities of three (4%) *APC* mutations were unknown as they were characterised in affected relatives outside of Wales and full details were not provided (Figure 3.2).

3.3.1.1 Nonsense, splice site, small deletion and insertion mutations in APC

This category of mutations accounted for 90% of the pathogenic *APC* mutations for which details were available (60/67). All but three of these were located within the coding region and just over half were found within the 5' half of exon 15 (codons 653-1748) (Figure 3.3). The only pathogenic mutation located outside this region but still within exon 15 was c.6030delA which affects codon 2010. This was the most 3' pathogenic change identified and was found in one case with an AFAP phenotype. This patient presented aged 52 years with between 50 and 100 colorectal adenomas in the absence of CRC and also had multiple fundic gland polyps (FGPs). There was a dominant family history of CRC and colorectal polyps as her father was affected at 58 years. Two unrelated cases harboured a mutation at codon 1061 (c.3183_3187delACAAA) of which one was identified in the present study (1157; Table 3.3). Neither patient had a family history of CRC or polyposis and their phenotypes were very similar with both having over a thousand colorectal adenomas diagnosed aged 29 and 33 years, with the latter also diagnosed with CRC. Codon 1309 was mutated (c.3927_3931delAAAGA) in six cases, two of whom were screened in the present study (0900, 1208; Table 3.3). The age at diagnosis of patients carrying the codon 1309 mutation varied, with two diagnosed in their second decades, one in her twenties and two in their thirties. Clinical information was

Patient	Exon / Intron	Nucleotide change	AA change	Age at diagnosis (yrs)	Polyp count	CRC	Extracolonic features	Family history	Reference
1663	1-15	Promoter_Ex15del	/	26	Large numbers	N	None noted	N	De Rosa, M. *et al*, 1999
1222	1-15	Promoter_Ex15del	/	40	>100	N	DP & FGP	D	De Rosa, M. *et al*, 1999
1173	**1-2**	**Promoter_Ex2del**	**/**	**57**	**Multiple**	**Y(1)**	**None noted**	**D**	**Novel**
9712	4-5	Ex4_5del	/	24	20	N	None noted	D	Nielsen, M. *et al*, 2007b
2256	**9-10A**	**Ex9_10Adel**	**/**	**47**	**>1000**	**Y(4)**	**FGP**	**N**	**Novel**
1191	5	c.637 C>T	R213X	32	>1000	N	None noted	D	Miyoshi, Y. *et al*, 1992
4048	**6**	**c.667 C>T**	**Q223X**	**40**	**Numerous (>100)**	**Y(3)**	**FGP, desmoids & osteoma**	**D**	**Novel**
0966	6	c.694 C>T	R232X	50	100s	Y (1)	SC	N	Miyoshi, Y. *et al*, 1992
3318	7	c.739 C>T	Q247X	16	Numerous	N	DP & brain tumour	D	Moisio, A-L. *et al*, 2002
7680	7	c.832 C>T	Q278X	18	10	N	None noted	D	Wallis, Y.L. *et al*, 1999
1249	9	c.994 C>T	R332X	34	Multiple	N	SC & osteoma	D	Soravia, C. *et al*, 1998
1346	**9**	**c.1239_1240insA**	**R414fsX418**	**36**	**Polyposis**	**N**	**FGP**	**D**	**Novel**
2284	IVS9	c.1312+3 A>G	/	35	Polyposis	N	None noted	D	Olschwang, S. *et al*, 1993
1070	**IVS13**	**c.1744-4 C>G**	**/**	**30s**	**Polyposis**	**N**	**None noted**	**D**	**Novel**
5449	14	c.1779 G>A	W593X	56	100s	Y(1)	None noted	D	Nagase, H. *et al*, 1992a
1042	**IVS14**	**c.1958+1 G>C**	**/**	**18**	**~300**	**N**	**None noted**	**D**	**Novel**
1842	15	c.2805 C>A	Y935X	23	>100	N	None noted	D	Fodde, R. *et al*, 1992
2857	15	c.2805 C>A	Y935X	21	100s	N	None noted	N	Fodde, R. *et al*, 1992
1157	15	c.3183_3187delACAAA	K1061fsX1062	29	>1000	N	None noted	N	Miyoshi, Y. *et al*, 1992
0107	15	c.3202_3205delTCAA	S1068fsX1124	23	Multiple	N	None noted	A	Bisgaard, M.L. *et al*, 2004a
3383	15	c.3202_3205delTCAA	S1068fsX1124	40	100s	N	None noted	D	Bisgaard, M.L. *et al*, 2004a
1105	15	c.3366_3369delTCAA	N1122fsX1124	27	100s	N	Dermoid cyst	D	Nagase, H. *et al*, 1992a
0900	15	c.3927_3931delAAAGA	E1309fsX1312	13	100s	N	None noted	D	Miyoshi, Y. *et al*, 1992
1208	15	c.3927_3931delAAAGA	E1309fsX1312	35	Unknown	N	None noted	D	Miyoshi, Y. *et al*, 1992
1239	**15**	**c.4132 C>T**	**Q1378X**	**38**	**>1000**	**Y(1)**	**GP**	**N**	**Novel**
1098	15	c.4393_4394delAG	S1465fsX1467	22	Polyposis	N	SDT	D	Miyoshi, Y. *et al*, 1992
1078	15	c.4666delA	T1556fsX1564	18	Quite a few throughout colon	N	Desmoids, SC & dental abnormalities	N	Oka, T. *et al*, 2004

Table 3.3 Pathogenic germline *APC* mutations identified in 27 unrelated colorectal polyposis patients in the present study. Novel mutations are shown in bold and references provided for previously reported identical mutations. AA; amino acid, DP; duodenal polyps, FGP; fundic gland polyps, GP; gastric polyps, SC; sebaceous cysts, SDT; subcutaneous dermoid tumours, D; dominant, N; no family history, A; adopted.

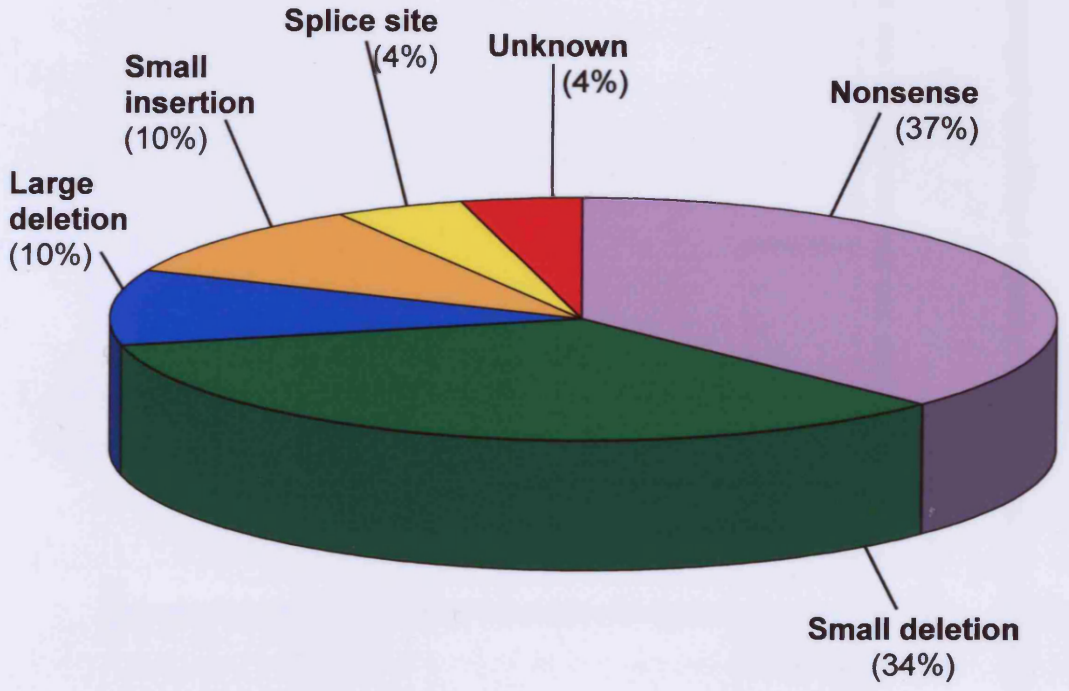


Figure 3.2 Class of pathogenic germline *APC* mutations identified in 70 apparently unrelated colorectal polyposis patients. The class of *APC* mutation in three families on the Wales Polyposis Register is unknown as these mutations were identified in affected relatives outside of Wales and details were not available.

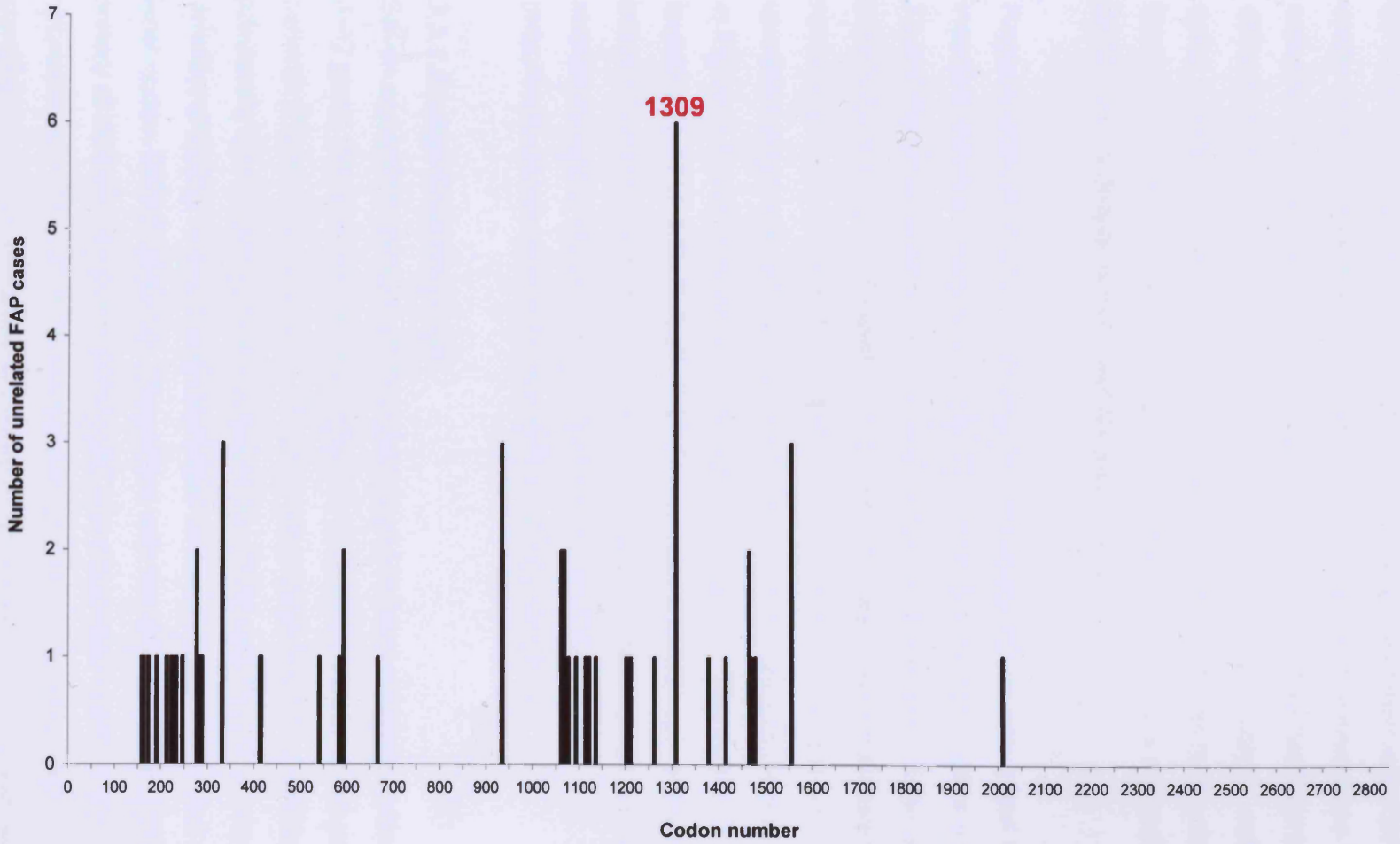


Figure 3.3 Distribution of coding region pathogenic germline *APC* mutations in 57 colorectal polyposis patients. The most frequently mutated codon is 1309.

not available for one patient. Two cases were described as having hundreds and a carpet of colorectal polyps, respectively. A polyp count was not known for the other patients but they had been clinically diagnosed with FAP. Four cases were not affected by CRC and no extracolonic manifestations had been noted in two patients. One case had a scalp cyst, another had learning difficulties and one case was diagnosed with fundic gland polyps, sebaceous cysts, a lipoma and an osteoma of the skull. Three patients had a dominant family history of CRC and/or colorectal polyps, one case was sporadic and family history details of two patients were unknown.

Regions other than exon 15 affected by nonsense, splice site, small deletion or insertion mutations were exons 4-9, 12 and 14, the acceptor splice site in intron 13 and the donor splice sites in introns 9 and 14. Seventeen of the twenty-nine base substitutions (59%) were C to T transversions. One small insertion, two nonsense and two splice site mutations identified in the present screen of 35 unrelated polyposis patients had not been previously unreported and are shown in Figure 3.4, with phenotypes described in Table 3.3. The intron 13 splice site mutation (c.1744-4 C>G) is predicted to activate a cryptic splice acceptor site in intron 13 and the intron 14 splice site mutation (c.1958+1 G>C) is predicted to abolish the splice donor site of intron 14 by SpliceSiteFinder (http://violin.genet.sickkids.on.ca/~ali/splicesitefinder.html).

3.3.1.2 Large deletions in APC

Seven apparently unrelated cases were found to have a large deletion in the *APC* gene, five of which were identified in the present screen of 35 polyposis patients (Figure 3.5, Table 3.3). The two other large deletions had been previously identified in patients on the Wales Polyposis Register; one was a deletion of exons 11 and 12 and the other covered exons 13, 14 and up to at least codon 1628 in exon 15. The patient with the former mutation had a family history of multiple colorectal polyps and was diagnosed aged 40 years with approximately 100 colorectal polyps without CRC. The latter mutation was identified in a case with a family history of colorectal polyposis and was diagnosed aged 29 years with colorectal and gastric polyps in the absence of CRC.

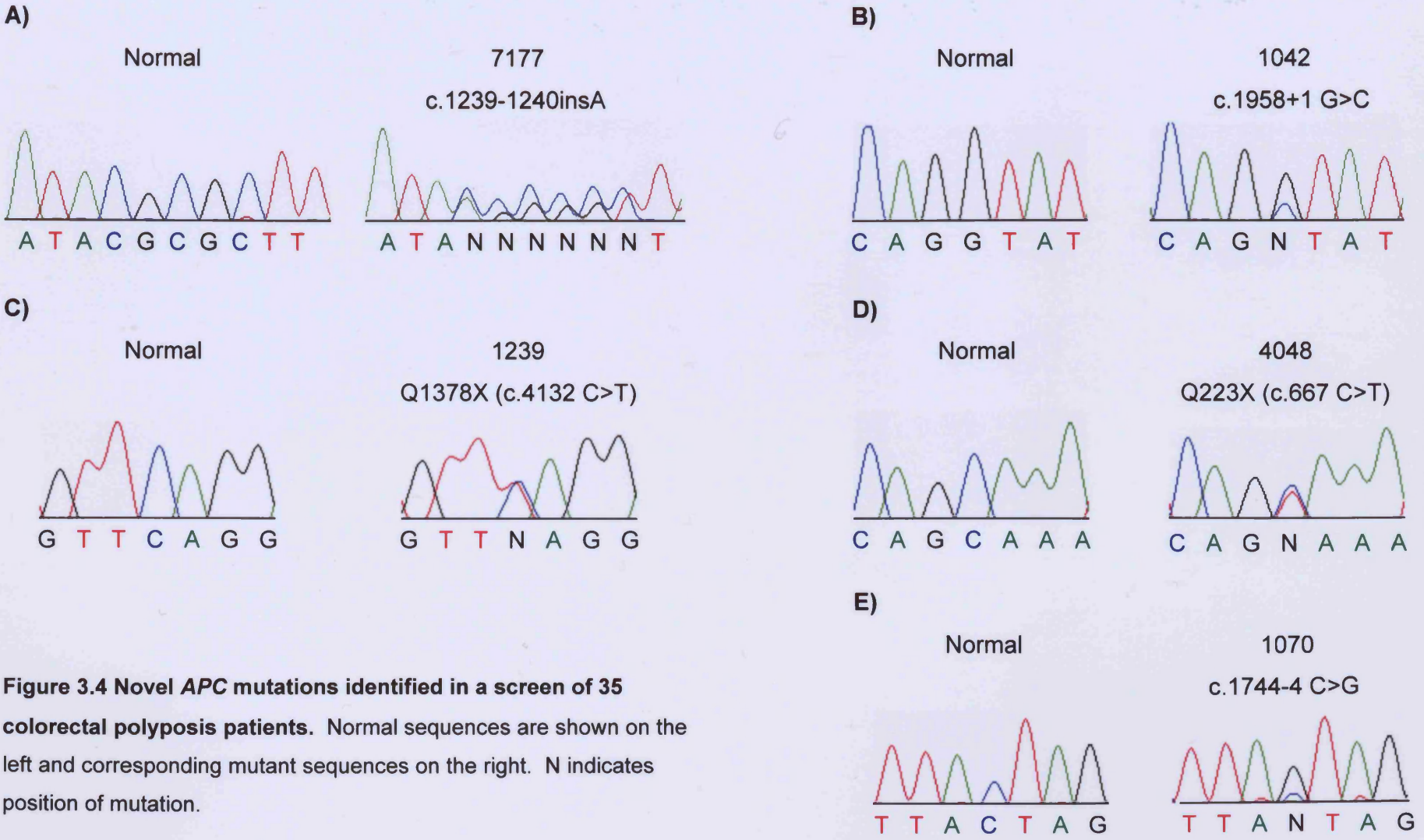


Figure 3.4 Novel *APC* mutations identified in a screen of 35 colorectal polyposis patients. Normal sequences are shown on the left and corresponding mutant sequences on the right. N indicates position of mutation.

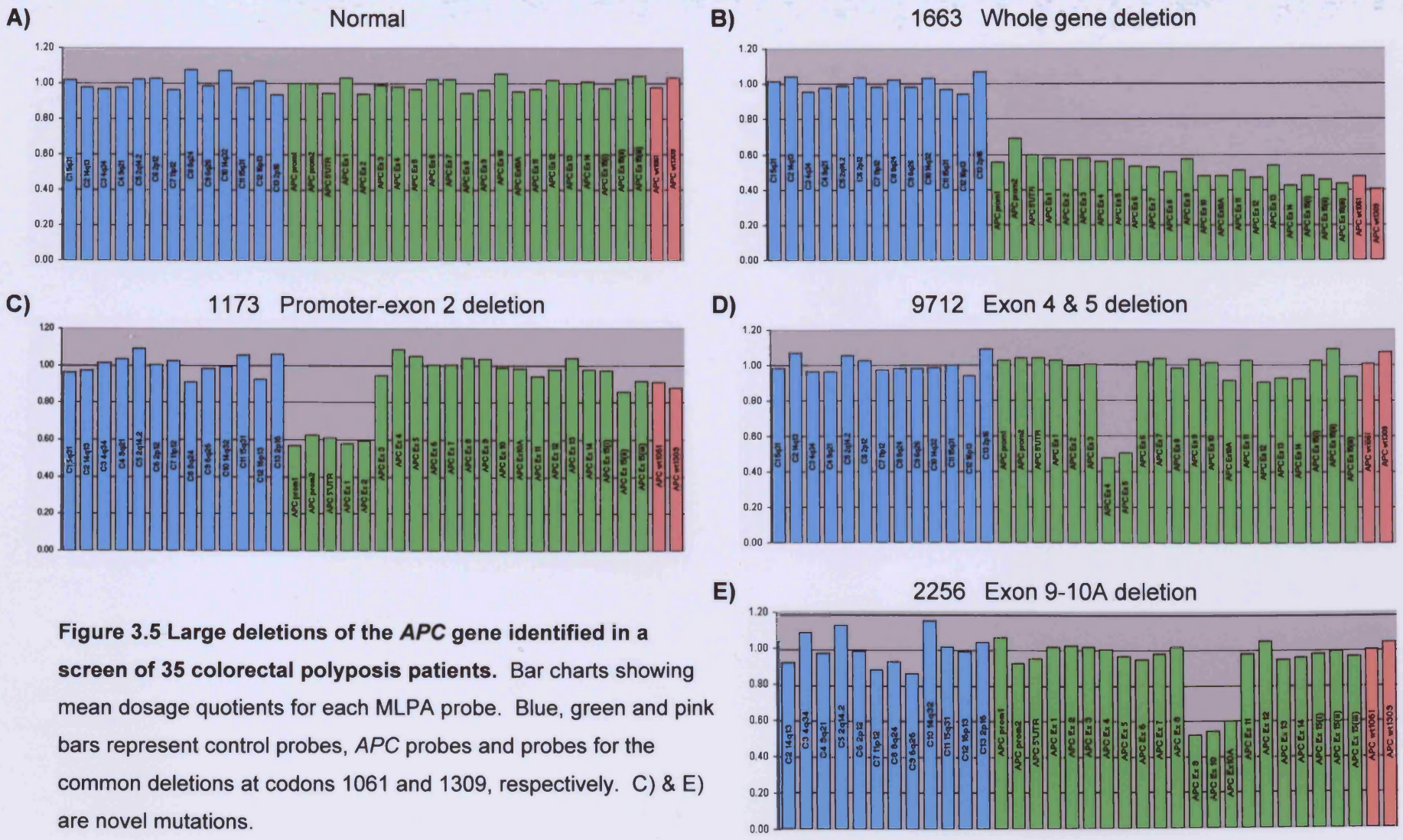


Figure 3.5 Large deletions of the *APC* gene identified in a screen of 35 colorectal polyposis patients. Bar charts showing mean dosage quotients for each MLPA probe. Blue, green and pink bars represent control probes, *APC* probes and probes for the common deletions at codons 1061 and 1309, respectively. C) & E) are novel mutations.

3.3.1.3 Missense and silent mutations in APC

Five *APC* missense mutations (R106H, R1450L, M2491T, G2502S and S2621C) were each identified as heterozygous changes in single cases in the present screen of 35 colorectal polyposis patients. All but one (S2621C c.7862 C→G) were detected in patients who also harboured a truncating *APC* mutation. The S2621C variant was identified in case 8163 and although it was originally reported as a pathogenic mutation (Miyoshi, Y. *et al*, 1992), subsequent studies have shown this change does not segregate with the FAP phenotype (Walon, C. *et al*, 1997; Ruiz-Ponte, C. *et al*, 2001) and it is not predicted to affect splicing (Sharp, A. *et al*, 2004). The G2502S (c.7504 G→A) missense mutation has been previously reported with no evidence that this variant affects splicing *in vivo* and it was not thought to affect protein function (rs2229995; Sharp, A. *et al*, 2004; Miyoshi, Y. *et al*, 1992). The three novel missense mutations R106H (c.317 G→A), R1450L (c.4349 G→T) and M2491T (c.7472 T→C) were not found in the Human Gene Mutation Database (http://www.hgmd.cf.ac.uk/) or the dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/) but were found in patients with a truncating *APC* mutation so are unlikely to be the disease-causing mutations. The previously reported missense variant E1317Q (c.3439 G→C; rs1801166, Frayling, I.M. *et al*, 1998) had been identified in four Wales Polyposis Register families by the AWMGS or by researchers of the Academic Department of Medical Genetics but only in conjunction with a co-existing *APC* truncating mutation or biallelic *MUTYH* mutations.

The common *APC* polymorphisms Y486Y (c.1458 C→T; rs2229992), T1493T (c.4479 A→G; rs41115), G1678G (c.5034 A→G; rs42427), S1756S (c.5268 G→T; rs866006), D1822V (c.5465 A→T; rs459552) and P1960P (c.5880 A→G; rs465899) were detected along with the previously reported silent variant S2307S (rs2229993).

3.3.2 Contribution of biallelic MUTYH mutations to colorectal polyposis in Wales

Of the 92 unrelated colorectal polyposis cases screened for mutations in *APC* and/or *MUTYH* (either previously or in the present study), biallelic germline *MUTYH* mutations were identified in 15 families (16%). One MAP case was

identified in the present screen. This patient was already known to be a Y176C (c.527 A>G) heterozygote but the current screen identified a second germline *MUTYH* mutation; a previously reported intron 15 splice site mutation (c.1509+2 T>C; Aretz, S. *et al*, 2006), which was found in the compound heterozygous state in this patient but was not found in over 728 control chromosomes. This variant is predicted to abolish the splice donor site by the splice site prediction program NetGene2 (http://www.cbs.dtu.dk/services/NetGene2). This case without a family history of CRC and/or polyps was diagnosed aged 41 years with approximately ten colorectal adenomas, metaplastic polyps and a CRC. She then had two additional colorectal carcinomas at 65 and 68 years and a breast carcinoma aged 71 years. The clinical and molecular genetic details of all Wales Polyposis Register patients with biallelic germline *MUTYH* mutations are described in chapter four. The common *MUTYH* polymorphisms V22M (c.64 G→A; rs3219484) and Q335H (c.1005 C→G; rs3219489) were detected in the present screen.

3.3.3 Colorectal polyposis cases without pathogenic APC or MUTYH mutations

In seven of the 92 unrelated polyposis cases screened for mutations in *APC* and/or *MUTYH* (either previously or in the present study), no pathogenic germline mutations were detected in *APC* or *MUTYH*. All seven patients were in the present screen and had the entire *APC* and *MUTYH* ORFs screened by sequencing. MLPA was performed to detect large deletions at the *APC* locus in all but one of these cases which failed due to poor quality DNA. One of the seven patients had an affected sibling but the remaining six were sporadic cases. The colorectal phenotype varied (Table 3.4) with polyp counts ranging from 14 to hundreds of polyps and the age at diagnosis was between 23 and 56 years. Five cases developed CRC of which one had two cancers.

Patient	Age at diagnosis	Polyp count	CRC	Family history
1187	23	50-100	Y (1)	Sporadic
0830	24	57	N	Sporadic
1137	28	14	N	Sporadic
4093	33	100s	Y (1)	Sporadic
2302	34	100-200	Y (1)	Sporadic
1149	42	16	Y (2)	Sister CRC aged 35yr
8163	56	~20	Y (1)	Sporadic

Table 3.4 Phenotypes of colorectal polyposis patients in whom no pathogenic germline *APC* or *MUTYH* mutations were identified. All cases had their entire *APC* and *MUTYH* ORFs screened and MLPA of the *APC* gene performed in the present study.

3.4 Discussion

3.4.1 Prevalence of germline APC mutations in polyposis patients

The Wales Polyposis Register includes the families of all patients with at least ten colorectal adenomas identified by the All Wales Medical Genetics Service. The 107 apparently unrelated families vary in size from one to eleven affected family members. Seventy-six percent of families on the Wales Polyposis Register that were screened for mutations in *APC* and/or *MUTYH* harboured a pathogenic germline mutation in the *APC* gene. This is a higher proportion than previously reported in several recent screens of polyposis registers which described pathogenic *APC* mutations in 50-69% of unrelated cases with a clinical diagnosis of FAP, AFAP or multiple adenomas (Bisgaard, M.L. *et al*, 2004a; Vandrovcova, J. *et al*, 2004; Aceto, G. *et al*, 2005; Truta, B. *et al*, 2005; Aretz, S. *et al*, 2006). Some of the variation could be due to differences in techniques used to screen for *APC* mutations, for example a number of previous studies would not have been able to detect single exon, multi-exon or whole gene deletions (Bisgaard, M.L. *et al*, 2004a; Vandrovcova, J. *et al*, 2004; Truta, B. *et al*, 2005) which were screened for in all families on the Wales Polyposis Register in whom no truncating *APC* or biallelic *MUTYH* mutations were identified. However, this cannot account for all differences as Aceto *et al* (2005) and Aretz *et al* (2006) undertook comprehensive *APC* screening in their polyposis patients but only identified a pathogenic mutation in 52% and 50% of unrelated cases, respectively. The variation is unlikely to be a result of screening families from different populations as previous studies of UK FAP families identified pathogenic *APC* mutations in just 45% of those in Birmingham (Wallis, Y.L *et al*, 1999) and 54% of families on the Northwest of England Regional Polyposis Register (Armstrong, J.G. *et al*, 1997) of cases. Cao *et al* (2006) recently detected pathogenic *APC* mutations in 94% (50/53) of families on the Singapore Polyposis Register but strict selection criteria were applied; only cases with more than a hundred colorectal adenomas or those diagnosed with CRC and at least twenty adenomas under the age of 50 years with a dominant family history were screened. *APC* analysis was also extensive, comprising PTT assay, MLPA and differential expression analysis (Cao, X. *et al*, 2006). Aretz *et al* (2006) observed that pathogenic *APC* mutations were detected at a higher frequency in unrelated patients with typical

FAP (83%) than those with an AFAP phenotype (22%) and that neither *APC* nor biallelic *MUTYH* mutations were identified in 17% and 64% of unrelated classical FAP and AFAP cases, respectively. Similarly, recent screens of unrelated patients with over a hundred colorectal adenomas identified pathogenic *APC* mutations in 71% (Kim, D-W. *et al*, 2005), 76% (De Rosa, M. *et al*, 2004) and 88% (De la Fuente, M.K. *et al*, 2007) of cases and Nielsen *et al* (2007a) described pathogenic *APC* mutations in 36% of AFAP families. The proportions of cases with these phenotypes will therefore affect the number of *APC* mutations detected. All affected members of families on the Wales Polyposis Register had at least ten colorectal adenomas but clinical details were not available for all cases so exact proportions of patients with a clinical diagnosis of classical FAP or AFAP could not be calculated. An additional 15 families on the Wales Polyposis Register have not been screened to date due to lack of DNA from an affected member or are in the process of being screened for mutations in *APC* and/or *MUTYH*. The proportion of families on the Wales Polyposis Register with pathogenic *APC* mutations may therefore alter once these additional families have been screened.

3.4.2 Spectrum and frequency of pathogenic APC mutations

3.4.2.1 APC nonsense, splice site, small deletion and insertion mutations

Just over half (52%) of the mutations in this category detected in Wales Polyposis Register families for which details were available were small deletions or insertions. Forty-three percent were nonsense mutations and 5% were predicted to affect splicing. Previous *APC* screens in colorectal polyposis cases reported slightly higher proportions of small deletions and insertions (59-70%) and lower proportions of nonsense mutations (27-39%) (van der Luijt, R.B. *et al*, 1997; Friedl, W. *et al*, 2001; Bisgaard, M.L. *et al*, 2004a; Vandrovcova, J. *et al*, 2004; Truta, B. et al, 2005; Cao, X. *et al*, 2006). However, Aceto *et al* (2005) and Stekrova *et al* (2007) described similar proportions of small deletions and insertions (53% and 50%, respectively) and nonsense mutations (40% and 43%, respectively) as in the present study. Splice site mutations have been reported to contribute 0-11% of mutations (van der Luijt, R.B. *et al*, 1997; Friedl, W. *et al*, 2001; Bisgaard, M.L. *et al*, 2004a; Truta, B. et al, 2005) in colorectal polyposis patients but no more than nine cases have been described in a single

study (Friedl, W. *et al*, 2001). Variation in the proportions of nonsense, splice site, small deletion and insertion mutations may be due to differences in the sensitivity of screening techniques (as discussed above) or sample sizes; Friedl *et al* (2001) reported pathogenic point mutations, small insertions or deletions in 322 colorectal polyposis cases whereas such mutations were detected in 59 families on the Wales Polyposis Register.

Consistent with previously reported coding region germline pathogenic *APC* mutations (Figure 1.5, Thierry Soussi APC mutation database, personal communication), the majority of such mutations identified in Wales Polyposis Register families were found in the 5' half of the gene (Figure 3.3). No mutations were identified before codon 159 but a small number have been described within this region (Stekrova, J. *et al*, 2007; Figure 1.5, Thierry Soussi APC mutation database, personal communication). Only one mutation was detected downstream of codon 1600 in accordance with the reported rarity of mutations 3' to this codon (Figure 1.5, Thierry Soussi APC mutation database, personal communication).

The most frequent *APC* mutation identified in unrelated cases on the Wales Polyposis Register was c.3927_3931delAAAGA, affecting codon 1309. This is by far the most commonly identified mutation (Figure 1.5, Thierry Soussi APC mutation database, personal communication) with very few reports from registries in which this change was not detected in any colorectal polyposis cases (Ruiz-Ponte, C. *et al*, 2001; Scott, R.J. *et al*, 2001). The high frequency of this mutation is not likely to be a result of a founder effect as it has been identified in patients from a wide range of ethnic backgrounds (Miyoshi, Y. *et al*, 1992; Cao, X. *et al*, 2006) and there is a high incidence of this deletion among *de novo* cases (Ripa, R. *et al*, 2002; Aretz, S. *et al*, 2004a). This suggests that the mutation is recurrent, supported by the fact that codon 1309 is part of a short repeat sequence which may increase its mutability (Segditsas, S. & Tomlinson, I., 2006). Aretz *et al* (2004a) found significantly fewer c.3927_3931delAAAGA mutations in familial FAP cases than in *de novo* cases but of the six unrelated index patients carrying this mutation on the Wales Polyposis Register, the family history of two cases was unknown (one through

lack of index case contact with family), three were familial cases and only one had no family history of CRC or polyps. The difference may be because the probands screened are offspring of *de novo* cases who have survived to reproduce due to successful treatment. This appears to be the case for two of the three unrelated familial cases as both their fathers were the first members of their families to be affected. The small numbers of unrelated cases with this mutation on the Wales Polyposis Register may also not reflect the frequency of the deletion in familial and *de novo* cases very accurately.

Codon 1061 has been reported as a hotspot for germline *APC* mutations (Figure 1.5, Thierry Soussi APC mutation database, personal communication; van der Luijt, R.B. *et al*, 1997; Friedl, W. et al, 2001) in colorectal polyposis patients but mutations at this codon only accounted for two of the mutations detected in Wales Polyposis Register families. Similarly, Bisgaard *et al* (2004a), Cao *et al* (2006) and Stekrova *et al* (2007) did not find mutations affecting this codon at higher frequencies than mutations at several other codons. Differences in the frequency of codon 1061 mutations do not appear to be the result of population differences as Wallis *et al* (1999) described pathogenic changes at this codon in 10% of cases with *APC* mutations from the Birmingham area. Both of the unrelated Wales Polyposis Register cases with a codon 1061 mutation did not have a family history of CRC or polyps and mutations affecting this codon have been previously reported in *de novo* FAP cases, although at a much lower frequency than 1309 mutations (Ripa, R. *et al*, 2002; Aretz, S. *et al*, 2004a). Codon 1061 is part of a short repeat sequence which may increase its mutability (Segditsas, S. & Tomlinson, I., 2006) and sample size may account for the differences between studies. Despite the relatively low frequency of mutations at codon 1061 in Wales Polyposis Register families, the region of *APC* between codons 1061 and 1136 (225 bp) harboured 17% of the nonsense, splice site, small deletion and insertion mutations detected in Wales Polyposis Register families for which details were available.

Mutations affecting codons 332, 934 and 1556 were each identified in three unrelated cases on the Wales Polyposis Register, with the remaining mutations detected in one or two families. The R332X mutation has been previously

described in a maximum of two unrelated cases in any one study (Thierry Soussi *APC* mutation database, personal communication; Hutter, P. *et al*, 2001) and so is not a frequently reported mutation. The c.2802-2805delTTAC mutation affecting codon 934 has been previously reported to account for a similar proportion of mutations as in the present study (5-6%) (Gavert, N. *et al*, 2002; Bisgaard, M.L. et al, 2004a) which may be due to its proximity to a short repeat sequence. The insertion or deletion of a single adenine at codon 1556 was identified in three unrelated cases on the Wales Polyposis Register and a small number of cases have been reported to carry the insertion mutation (Thierry Soussi APC mutation database, personal communication) which is likely to be because the poly$(A)_6$ tract in which codon 1556 lies increases its mutability. All unreported *APC* mutations were identified in single families on the Wales Polyposis Register.

Consistent with previously reported germline pathogenic *APC* point mutations (Thierry Soussi database, personal communication), the majority of base substitutions (59%) were C to T transversions which may be the result of spontaneous deamination of 5-methylcytosine, a product of intracellular metabolism (Olschwang, S. *et al*, 1993).

3.4.2.2 Large deletions in the APC gene

Of the pathogenic *APC* mutations detected in Wales Polyposis Register families for which details were available, 10% were large deletions (7/67). This is consistent with Michils *et al* (2005) and Cao *et al* (2006) who reported large deletions accounting for 12% (4/34) and 10% (5/50) of *APC* mutations in polyposis patients, respectively. Studies with larger numbers of patients described lower proportions of 8% (Nielsen, M. *et al*, 2007b) and 5% (Aretz, S. *et al*, 2005) of *APC* mutations accounted for by large deletions, which may be a more accurate reflection of the contribution of such mutations. As shown in Figure 3.6, whole gene and partial gene deletions affecting a range of exons have been identified. Unlike previous reports (Michils *et al*, 2005; Aretz, S. *et al*, 2005; Cao *et al*, 2006; Nielsen, M. *et al*, 2007b), no single exon deletions were found in Wales Polyposis Register families; five unrelated patients carried deletions of two or three exons and two harboured a whole gene deletion.

Figure 3.6 Distribution and frequency of large submicroscopic *APC* deletions in FAP patients. Deletions are shown as blue lines, novel mutations identified in the present study are shown in red and the green line represents a novel mutation identified outside of Wales reported as a deletion of exons 13, 14 & up to at least codon 1628 of exon 15. Dashed lines represent unknown 5' and 3' breakpoints. The most frequently reported deletion is that of the whole gene. Adapted from Nielsen *et al* (2007b) with additional information from Cao *et al* (2006), Stekrova *et al* (2007) and Cattaneo *et al* (2007).

Whole gene deletions are the most frequently reported large deletion (Figure 3.6; De Rosa, M. *et al*, 1999; Sieber, O.M. *et al*, 2002; Michils *et al*, 2005; Aretz, S. *et al*, 2005; Cao *et al*, 2006; Nielsen, M. *et al*, 2007b) and were the most common large deletion in the present study. One of the two whole-gene deletion cases was familial and the other was a *de novo* case. The extent of the whole gene deletions identified in the present study is not known; they may extend significantly outside of the *APC* gene as has been previously described in a few cases (De Rosa, M. *et al*, 1999; Sieber, O.M. *et al*, 2002). All novel deletions were identified in single families on the Wales Polyposis Register.

3.4.3 APC missense mutations

The previously reported missense mutations I1307K (Laken, S.J. *et al*, 1997) and E1317Q (Frayling, I.M. *et al*, 1998) have been associated with increased CRC susceptibility but only E1317Q was identified in families on the Wales Polyposis Register and only in conjunction with an *APC* truncating mutation (3 families) or biallelic *MUTYH* mutations (1 family).

3.4.4 Genotype-phenotype relationships in FAP

3.4.4.1 Severe colorectal polyposis

Mutations between codons 1250 and 1464 have been identified in cases with profuse polyposis (more than 5000 colorectal polyps; Nagase, H. *et al*, 1992b) with codon 1309 mutations associated with an early onset of severe polyposis (Caspari, R. *et al*, 1994; Ficari, F. *et al*, 2000; Friedl, W. *et al*, 2001). As expected, patients harbouring mutations within codons 1250-1464 (including codon 1309) develop CRC at an earlier age than cases carrying mutations outside this region (Caspari, R. *et al*, 1994; Friedl, W. *et al*, 2001). It is thought that the selection for somatic LOH in the context of a germline mutation around codon 1309 leads to severe polyposis as LOH occurs at a higher spontaneous frequency than truncating mutations close to codon 1300, (Lamlum, H. *et al*, 1999; Crabtree, M. *et al*, 2003). Within the six families on the Wales Polyposis Register carrying the same codon 1309 mutation, there were a total of twelve clinically affected patients but family history information was not available for two index patients. The mean age at diagnosis of nine cases who presented symptomatically was 25 years (range 6-41 years) and two of these

had CRC. Exact polyp counts were not available for many patients but they were described as having polyposis, hundreds of polyps or in one case a colon 'carpeted' with polyps. Only the youngest patient (6 years) may not have had more than a hundred polyps but no description was available. Three cases developed symptoms before the age of 20 years and another died from CRC aged 26 years, consistent with previous reports of an early disease onset associated with mutations at codon 1309 (Nagase, H. *et al*, 1992b; Caspari, R. *et al*, 1994; Ficari, F. *et al*, 2000; Friedl, W. *et al*, 2001). Inter- and intrafamilial differences in colorectal polyp density have been described in patients with the same codon 1309 mutation (Giardiello, F.M. *et al*, 1994) and these were also observed in families on the Wales Polyposis Register. This suggests that other genetic or environmental factors are contributing to the expression of the disease. Detailed clinical information on two Wales Polyposis Register patients carrying mutations in the region codons 1250 and 1464 (at 1262 and 1415) was not available.

3.4.4.2 Attenuated colorectal polyposis

AFAP patients generally carry germline *APC* mutations at the 5' or 3' ends of *APC* (codons <163 and 1596-2644) or between codons 329 and 338 in the alternatively spliced region of exon 9 (Galiatsatos, P. & Foulkes, W.D., 2006). Unlike classical FAP, mutations in these regions may allow residual tumour suppressor activity (Spirio, L.N. *et al*, 1998; Su, L-K. *et al*, 2000; Heppner Goss, K. *et al*, 2002) so two somatic *APC* mutations (one on the mutated allele and another on the wild-type allele) may be required for inactivation of the tumour suppressor activity. This delays tumour initiation so AFAP cases have fewer adenomas at a later age of onset than classical FAP patients. Six Wales Polyposis Register families harboured a germline mutation within, or at the boundary of, these AFAP-associated regions of *APC*. Three families carried the R332X mutation and had nine clinically affected patients between them. All but one had been diagnosed with CRC at a mean age of 50 years (range 44-54 years) consistent with AFAP, but no colorectal polyp details were available.

The case without CRC also had an AFAP phenotype; she was diagnosed with multiple polyps aged 34 years. A single family carried a mutation within the 3' region (at codon 2010) and the two affected members displayed AFAP rather than classical FAP. The index case was diagnosed with 50-100 polyps at 52 years and her father had CRC and polyps at 58 years. Consistent with previous reports (Soravia, C. *et al*, 1998; Sieber, O. *et al*, 2006), the two families with mutations at the 5' end of the gene (Y159X and Q163X) had a phenotype that was more severe than the other AFAP patients but remained milder than classical FAP. The Q163X mutation was identified in an apparently *de novo* case who presented aged 45 years with CRC in the absence of recognised polyposis then later re-presented with hundreds of colorectal polyps. Her daughter had approximately ten adenomas in her thirties. The family harbouring the Y159X mutation included 11 clinically affected members, six of whom died from CRC aged 33-55 years without a diagnosis of FAP having been recognised. The index case presented with hundreds of mainly right-sided polyps and CRC at 45 years but differences in colorectal phenotype existed between affected family members. Intrafamilial variability has been reported in several pedigrees with mutations affecting the 5', exon 9 and 3' regions associated with AFAP (Scott, R.J. *et al*, 1995; Brensinger, J.D. *et al*, 1998; Soravia, C. *et al*, 1998; Rozen, P. *et al*, 1999). No other Wales Polyposis Register families with mutations outside the regions above appeared to have AFAP but clinical information was not available for all cases.

3.4.4.3 Extracolonic manifestations

Detailed data on extracolonic manifestations have not been collected systematically on the Wales Polyposis Register. Previous studies have shown that CHRPE is associated with mutations between codons 311 and 1444 (Nieuwenhuis, M.H. & Vasen, H.F.A., 2007) and desmoids are mostly found in patients with mutations beyond codon 1444 (Caspari, R. *et al*, 1995; Friedl, W. *et al*, 2001). Mutations 3' to codon 1400 are also frequently associated with the presence of multiple extracolonic manifestations including osteomas (Dobbie, A. *et al*, 1996b). Desmoids were known to be present in four unrelated cases from Wales Polyposis Register families with mutations affecting codons 1465, 1468 and 1556. Three had other affected family members who were not reported to

have desmoids and one was a sporadic case. Consistent with Friedl *et al* (2001) who identified desmoids in some patients with a mutation 5' to codon 1444, a single case carrying the novel Q223X mutation was diagnosed with desmoids, but not all affected family members had desmoids. This case and her daughter had osteomas, as did five others each carrying a mutation at either codon 332, 1309, 1465, 1468 or 1477. Osteomas have been reported in patients with mutations at 1309 (Paul, P. *et al*, 1993; Giardiello, F.M. *et al*, 1994) and after 1400 (Dobbie, A. *et al*, 1996b) but were not identified in several patients with mutations at codon 332 (Soravia, C. *et al*, 1998; Hutter, P. *et al*, 2001; Cao, X. *et al*, 2006) which is consistent with the observation that extracolonic manifestations (apart from upper GI polyps) are rare in AFAP patients (Knudsen, A.L. *et al*, 2003). Osteomas are therefore not confined to cases with mutations at codons 767-1513 as has been suggested (Bisgaard, M.L. & Bülow, S., 2006). A brain tumour was diagnosed in a Q247X carrier, a patient with the Q223X mutation had papillary thyroid cancer and two hepatoblastoma cases had been identified, including one carrying the c.1239_1240insA mutation. Such tumours are only rarely associated with FAP (Nieuwenhuis, M.H. & Vasen, H.F.A., 2007) so it is not surprising that no other family members were affected by these tumours. Epidermoid cysts have been reported to be more frequent in cases with mutations at codons 1395-1493 (Wallis, Y.L. *et al*, 1999) but skin lesions were observed in Wales Polyposis Register patients with a wider range of mutations. Bertario *et al* (2003) reported an increased risk of duodenal adenomas in cases with mutations between codons 976 and 1067 but duodenal and gastric polyps were detected in Wales Polyposis Register patients with a range of mutations, in accordance with previous studies (Friedl, W. *et al*, 2001; Nieuwenhuis, M.H. & Vasen, H.F.A., 2007). Extracolonic manifestations in patients with the same mutation showed inter- and intrafamilial variability as previously described (Giardiello, F.M. *et al*, 1994; Paul, P. *et al*, 1993).

3.4.4.4 Phenotypic expression of novel nonsense, splice site and small deletion mutations

Five novel mutations were each identified once in the present screen of 35 unrelated patients from the Wales Polyposis Register. The nonsense mutation

Q223X was detected in a case with classical colorectal polyposis (well over a hundred polyps aged 40 years), as would be predicted by the location of this mutation. As for other mutations resulting in classical polyposis, the Q223X mutation is predicted to truncate the APC protein before any 20aa repeat sequences and somatic *APC* mutations resulting in a truncated protein with one or two 20aa repeats are predicted to be selected for tumour initiation (Albuquerque, C. *et al*, 2002). Unlike mutations resulting in AFAP, no residual tumour suppressor activity is predicted to be present in the protein encoded by this mutant allele as exon 6 is not alternatively spliced and codon 223 is beyond the alternative translational start site in exon 5. This case was diagnosed with three synchronous CRCs and Cao *et al* (2006) reported synchronous CRCs in 5% (6/122) of *APC* mutation positive cases. The presence of multiple CRCs in this Wales Polyposis Register case may be because of the late age at diagnosis as prophylactic surgery was not undertaken. Only tumour initiation is accelerated in FAP, not its progression, so the probability of three adenomas progressing to CRC at a similar time in a patient who was not described as having thousands of polyps would not be expected to be high. Additional environmental or genetic factors may be contributing to the expression of the disease and evidence for this is the intrafamilial variability; the patient's mother was diagnosed with CRC and polyps in her sixties whereas the patient's daughter had multiple polyps at 16 years. A mutation at codon 223 would not be predicted to result in a high risk of extracolonic manifestations (Dobbie, A. *et al*, 1996b) but this patient had desmoids, an osteoma and fundic gland polyps (FGPs). Her daughter also had an osteoma, FGPs, epidermoid cysts and a papillary thyroid cancer. Although desmoids, osteomas, FGPs and epidermoid cysts have been reported in FAP cases with mutations at a wide range of codons, Venesio *et al* (2003a) detected Q222X in a patient with classical FAP diagnosed aged 44 years without CRC or extracolonic lesions.

The novel exon 9 small insertion c.1239_1240insA is predicted to result in a stop codon at codon 418 and was identified in a case with classical colorectal polyposis (polyposis aged 36 years without CRC) and FGPs. This is as expected from the mutation location as it does not affect the AFAP-associated alternatively spliced region of exon 9. De Rosa *et al* (2007) recently described

alternative splicing of the entire of exon 9 resulting in a stop codon in exon 10 so despite being located in an alternatively spliced exon, the c.1239_1240insA mutation is predicted to result in a truncated protein without any residual tumour suppressor activity or 20aa repeat sequences. Therefore as for other mutations resulting in classical polyposis, a somatic *APC* mutation resulting in a truncated protein with one or two 20aa repeats is predicted to be necessary for tumour initiation (Albuquerque, C. *et al*, 2002). Similarly, other exon 9 mutations not located within the AFAP-associated alternatively spliced region have been described in classical FAP cases (Hutter, P. *et al*, 2001; Bisgaard, M.L. *et al*, 2004a). Six additional members of this family showed a classical FAP phenotype; three were found to have multiple colorectal polyps in their second decades, two were diagnosed with polyposis in their early thirties and one had CRC aged 42 years. FGPs and subcutaneous cysts were each identified in one of the six cases and another family member died aged 7 years from a hepatoblastoma. Olschwang *et al* (1993) reported a c.1238_1239insT mutation but no phenotypic details were provided.

The novel splice site mutation c.1744-4 C>G is predicted to activate a cryptic acceptor splice site which results in an in-frame 3bp insertion upstream of exon 14. The 3bp encode a stop codon, truncating the protein at codon 582. This mutation was identified in a case diagnosed in his thirties with classical colorectal polyposis without CRC. The truncated protein is not predicted to have any residual tumour suppressor activity or 20aa repeat sequences. A somatic *APC* mutation resulting in a truncated protein with one or two 20aa repeats is predicted to be necessary for tumour initiation (Albuquerque, C. *et al*, 2002), as is expected for other mutations resulting in classical polyposis. Seven other members of this family are thought to be affected; one was diagnosed with hundreds of colorectal polyps without CRC at 53 years and another was described as having symptoms of FAP and underwent colectomy at 24 years. A further five family members died from CRC at ages ranging from ~29 years to 58 years. One of these had two CRCs and multiple adenomas but polyp counts were not available for the other affected family members. Van der Luijt *et al* (1997) reported a c.1744-2 A>G mutation in a single index patient clinically described as having FAP but no further phenotypic details were provided.

The novel donor splice site mutation c.1958+1 G>C is predicted to skip exon 14 resulting in a stop codon at position 671 and was identified in a case with classical colorectal polyposis (~300 polyps aged 18 years) without CRC. Aretz *et al* (2004b) have previously reported G>T and G>A mutations at this position and transcript analysis revealed exon 14 skipping. This mutation is therefore predicted to result in a truncated protein without any residual tumour suppressor activity or 20aa repeat sequences. A somatic *APC* mutation resulting in a truncated protein with one or two 20aa repeats is predicted to be necessary for tumour initiation (Albuquerque, C. *et al*, 2002), as is expected for other mutations resulting in classical polyposis. Three other members of this family are thought to be affected, one underwent colectomy aged 14 years but no further clinical information about these cases was available. Aretz *et al* (2004b) described two patients with mutations at c.1958+1; one had hundreds of colorectal polyps at 25 years and duodenal adenomas at 32 years and the other had over a hundred colorectal polyps at 36 years. Neither case had any extracolonic manifestations. Bala *et al* (1997) reported a mutation at the exon 14 splice acceptor site resulting in skipping of this exon in a FAP family. Similar to the patient in the present study, both affected members had hundreds of colorectal adenomas diagnosed at an early age (16 and 27 years) with the latter also having CRC at diagnosis.

The novel nonsense mutation Q1378X was identified in an apparently *de novo* case diagnosed with over a thousand colorectal adenomas, CRC and gastric polyps at 38 years. The affected codon lies within the region associated with severe polyposis and Q1378X is predicted to result in a truncated protein with a single 20aa repeat. Somatic selection for LOH mutations would be predicted in tumours in this patient. Since this event occurs at a higher spontaneous frequency than do truncating mutations close to codon 1300, a severe phenotype would be predicted (Lamlum, H. *et al*, 1999; Crabtree, M. *et al*, 2003). The patient's son is also said to be affected but detailed clinical information was not available.

The phenotypes observed in the five index cases with previously unreported mutations are therefore consistent with previous studies and with the current

models of tumourigenesis in FAP (Lamlum, H. *et al*, 1999; Albuquerque, C. *et al*, 2002; Crabtree, M. *et al*, 2003).

3.4.4.5 Phenotypes associated with large deletions

Far fewer reports of colorectal polyposis patients with large submicroscopic deletions have been published than those with truncating point mutations, small insertions and deletions due in part to previous difficulties in detecting such gross mutations but some correlation between genotype and phenotype has been observed. Whole gene deletions tend to result in classical rather than severe polyposis with some cases also exhibiting extracolonic manifestations such as CHRPE, upper GI polyps, desmoids, osteomas and epidermoid cysts (Sieber, O.M. *et al*, 2002; Aretz, S. *et al*, 2005; Michils, G. *et al*, 2005; Nielsen, M. *et al*, 2007b; Stekrova, J. et al, 2007). Thyroid, duodenal, ovarian and nasopharyngeal cancers have also been reported in single whole-gene deletion cases (Michils, G. *et al*, 2005; Cao, X. *et al*, 2006). Only one case with a whole-gene deletion has been reported to have an AFAP phenotype with less than 70 polyps and CRC at 65 years (Venesio, T. *et al*, 2003b). Unlike some cases with cytogenetically detectable deletions involving *APC* (Raedle, J. *et al*, 2001), patients with submicroscopic deletions have not been reported to have any mental retardation (Mandl, M. *et al*, 1996; De Rosa, M. *et al*, 1999). The classical FAP phenotypes of the four affected members of two families with whole gene deletions identified in the present study are consistent with previous observations. As with other mutations resulting in classical polyposis, a whole gene deletion will not result in a protein with any tumour suppressor activity so a somatic *APC* mutation resulting in a truncated protein with one or two 20aa repeats is predicted to be necessary for tumour initiation (Albuquerque, C. *et al*, 2002).

Most cases with partial *APC* gene deletions also display a classical FAP phenotype with or without extracolonic features but these mutations appear to be associated with more inter- and intrafamilial variability, ranging from classical FAP to AFAP phenotypes (Aretz, S. *et al*, 2005; Nielsen, M. *et al*, 2007b). The exon 4-5 deletion identified in the present study was found in a case with only 20 macroscopic colorectal polyps aged 24 years. Three other family members

were affected; two died from CRC, one at 55 years and the other at 70 years, and another had colorectal polyps in the absence of CRC at 40 years. Nielsen *et al* (2007b) reported the same mutation in a classical FAP patient with over a hundred polyps and CRC aged 41 years. This deletion is out-of-frame so would not be predicted to produce any functional protein and therefore result in a classical FAP phenotype. Modifying factors affecting expression of truncating point mutations, small insertions and deletions are also likely to affect the expression of large deletions. The out-of-frame deletion of exons 11-12 was found in a Wales Polyposis Register family with three members affected by multiple colorectal polyps, but they were diagnosed at a later age (33-50 years) than a case with the same mutation reported by Aretz *et al* (2005) who had multiple colorectal polyps at 17 years. A deletion of exons 13-15 was initially identified outside of Wales and was described as a deletion of exons 13-15J. No exon 13-15 deletions have been reported to date but deletions of exons 14-15, 14-part of 15 and 11-15 have been described in classical FAP patients, consistent with the Wales Polyposis Register patient who also harboured this mutation. Detailed clinical information was not available for the other 14 affected members of this family.

The novel promoter-exon 2 deletion was identified in a single family on the Wales Polyposis Register with 15 affected members. Detailed clinical information was not available for seven cases but five of the remaining eight patients had CRC at a mean age of 47 years (range 25-57 years). Three cases were diagnosed with multiple colorectal polyps at 14, 26 and 57 years, one patient had numerous polyps in his twenties, another had hundreds aged 50 years and a single case aged 19 years had fewer than a hundred colorectal polyps. Two cases were only known to have died from CRC so no information about colorectal polyps was available. Sebaceous cysts, FGPs and duodenal polyps were each noted in single cases. This family shows the intrafamilial phenotypic variability associated with *APC* mutations but the predicted phenotype would be that of classical polyposis because without the promoter this mutation is predicted to be functionally equivalent to a whole gene deletion. In accordance with this, only classical FAP patients have been described with

deletions of promoter-exon 1 and promoter-exon 5 (Aretz, S. *et al*, 2005; Nielsen, M. *et al*, 2007b).

A novel deletion of exons 9-10A was detected in a single apparently *de novo* case with a severe phenotype of over a thousand colorectal polyps and four CRCs aged 47 years. This is the first mutation reported to affect the alternatively spliced exon 10A and although exon 9 can also be alternatively spliced (Groden, J. *et al*, 1991; De Rosa, M. *et al*, 2007), exon 10 is constitutively expressed thus all transcripts will be affected and no functional protein produced. The predicted phenotype would be that of classical FAP as a somatic *APC* mutation resulting in a truncated protein with one or two 20aa repeats is predicted to be necessary for tumour initiation (Albuquerque, C. *et al*, 2002). The severity of this patient's phenotype is likely to be due to his late age at diagnosis resulting from the lack of family history of CRC or polyps.

3.4.5 Prevalence of biallelic germline MUTYH mutations in polyposis patients

Sixteen percent of families on the Wales Polyposis Register that were screened for mutations in *APC* and/or *MUTYH* harboured biallelic germline mutations in the *MUTYH* gene. This is a significantly higher proportion than reported by Aretz *et al* (2006) in unselected, unrelated cases with a clinical diagnosis of FAP or AFAP (15/92 vs 55/660, p=0.016; Fisher's exact test). Aceto *et al* (2005) also reported biallelic *MUTYH* mutations in just 8% of cases clinically diagnosed with FAP or AFAP, but this was not statistically different from the present study (15/92 vs 5/60, p=0.12; Fisher's exact test). Possible explanations for variation in the prevalence of biallelic *MUTYH* mutations include differences in ascertainment, population differences in mutant allele frequencies or differences in the sensitivity of molecular testing. However, since the entire *MUTYH* ORF was screened in the previous studies (Aceto, G. *et al*, 2005; Aretz, S. *et al*, 2006) any differences are unlikely to be a result of undetected MAP cases in these cohorts. Notably, of the 92 unrelated Wales Polyposis Register cases screened for *APC* and/or *MUTYH*, only seven (8%) were mutation negative. By comparison, Aceto *et al* (2005) and Aretz *et al* (2006) reported an *APC* or *MUTYH* mutation in only 62% and 58% of cases, respectively. An additional fifteen families on the Wales Polyposis Register

have not been screened to date due to lack of DNA from an affected member or are in the process of being screened for mutations.

As described above, the proportions of patients with classical FAP and AFAP phenotypes can affect the *APC* mutation detection rate and the same is true of detecting *MUTYH* mutations. Aretz *et al* (2006) did not identify biallelic *MUTYH* mutations in patients with 'typical' FAP but did detect such mutations in 64% of cases with an AFAP phenotype. The proportions of cases with these phenotypes will therefore affect the number of MAP patients detected. All families on the Wales Polyposis Register had one or more family member with at least ten colorectal adenomas but polyp counts were not available for all cases so exact proportions of patients with a clinical diagnosis of classical FAP or AFAP could not be calculated.

3.4.6 Colorectal polyposis patients without APC or biallelic MUTYH mutations

Seven of the 92 unrelated Wales Polyposis Register cases screened for mutations in *APC* and/or *MUTYH* did not carry pathogenic mutations in either gene that could be detected by *APC* and *MUTYH* ORF sequence analysis or MLPA of *APC*. These cases may carry an *APC* mutation not detected by the methods used such as intronic mutations or promoter mutation or hypermethylation. However, germline hypermethylation of the *APC* promoter has not been identified (Hitchins, M. *et al*, 2005; Romero-Gimenez, J. *et al*, 2007) and is unlikely to be the cause of the colorectal polyposis in our cases. Constitutional reduction in expression from one *APC* allele has been described in several FAP patients (Laken, S. *et al*, 1999; Yan, H. *et al*, 2002; Renkonen, E.T. *et al*, 2005; Cao, X. *et al*, 2006) but mutations in the promoter regions or 3' UTR could not be identified in these cases (Yan, H. *et al*, 2002; Renkonen, E.T. *et al*, 2005) suggesting the mutations responsible are in other noncoding or regulation regions. The two unrelated cases with reduced *APC* expression identified by Renkonen *et al* (2005) were apparently *de novo* cases with over a hundred colorectal polyps in the absence of extracolonic manifestations. Recently, Venesio *et al* (2007) described increased constitutional expression of an *APC* mRNA isoform connecting exon 10 to exon 15 which encoded a stable truncated peptide in AFAP cases. This confirms the need to perform

constitutional *APC* transcription analysis in cases in which no *APC* or *MUTYH* mutations are identified by conventional techniques.

Another reason for failure to detect *APC* mutations in FAP patients is somatic mosaicism. Mutant alleles may be present at only low levels in the blood DNA of these patients, or even absent in the context of segmental disease. Aretz *et al* (2007) identified mosaicism in 8/75 (11%) of *de novo* FAP cases who carried pathogenic mutations within the region associated with typical or severe FAP. However, six of the eight mosaic patients had an AFAP or atypical phenotype. Hes, F.J. *et al* (2007) estimated that 20% of FAP *de novo* cases are accounted for by mosaic *APC* mutations and described somatic mosaic cases with phenotypes ranging from AFAP to florid polyposis with extracolonic manifestations. Six of the seven *APC* and *MUTYH*-negative cases in the present study are sporadic and could be accounted for by somatic mosaic *APC* mutations.

Alternatively, undetected *MUTYH* mutations such as large deletion/insertions or promoter mutation or hypermethylation could be present in the mutation-negative cases but no such mutations have been described to date. An *AXIN2* nonsense mutation has been identified in an *APC* mutation-negative family with dominantly-inherited severe tooth agenesis in association with between 1 and 68 colorectal polyps (Lammi, L. *et al*, 2004). However, subsequent *AXIN2* mutational analysis in a total of 30 *APC* mutation-negative colorectal polyposis patients has failed to detect any obviously pathogenic mutations (Renkonen, E.T. *et al*, 2005; Lejeune, S. *et al*, 2006) suggesting that mutations in *AXIN2* very rarely contribute to colorectal polyposis and that other genetic factors must account for a small proportion of colorectal polyposis.

The phenotype of polyposis patients in whom no *APC* mutation has been identified is controversial; Giarola *et al* (1999), Heinimann *et al* (2001) and Moisio *et al* (2002) have suggested a milder phenotype but Bisgaard *et al* (2004b) reported a more severe phenotype in mutation-negative families. The mean age at diagnosis of mutation-negative index patients in the present study was 34 years (range 23-56 years), five cases had CRC at a mean age of 38

years (range 23-56 years) and the polyp count varied from 14 to hundreds but none had severe polyposis.

3.4.7 Management of colorectal polyposis patients

Identification of the genetic cause of colorectal polyposis in families is crucial for correct clinical management and counselling. Whereas apparently sporadic *APC* mutation-positive cases have a 50% risk to offspring and a low risk to siblings, MAP poses little risk to offspring but a 25% risk to siblings. The correlations between *APC* mutation site and disease severity led Nieuwenhuis *et al* (2007) to suggest that mutational data might inform the type of prophylactic surgery needed to prevent CRC. The most difficult patients to diagnose on the basis of symptoms and family history are *de novo* colorectal polyposis patients. They could be the result of biallelic *MUTYH* mutations or an *APC* mutation. The present study highlights the need to screen the entire ORF of both genes and to look for large deletions in *APC*. If these methods fail to detect causative mutations other possibilities such as somatic mosaicism and reduced *APC* constitutional expression should be investigated.

Chapter Four

Clinical and molecular genetic characterization of *MUTYH*-associated polyposis (MAP)

4.1 Introduction

MAP was first identified in a single British family with three siblings affected by colorectal tumours. Two siblings had approximately fifty adenomas in their fifties and the third had CRC with adenomas but a complete assessment of the colorectum was not performed. Analysis of the *APC* gene in these tumours revealed a high frequency of somatic G:C to T:A transversions which lead to germline screening of the *MUTYH* gene through which biallelic mutations were identified in all three affected siblings (Al-Tassan, N. *et al*, 2002). Subsequent studies carried out *MUTYH* mutation analysis in multiple colorectal adenoma patients and confirmed that biallelic *MUTYH* mutations predispose to colorectal adenoma and carcinoma (Jones, S. *et al*, 2002; Sieber, O.M. *et al*, 2003; Sampson, J.R. *et al*, 2003).

Additional studies of MAP patients have now been described, with most cases reported to have tens or hundreds of colorectal adenomas typically by the fifth or sixth decade (Sieber, O.M. *et al*, 2003; Venesio, T. *et al*, 2004; Wang, L. *et al*, 2004; Nielsen, M. *et al*, 2005; Jo, W-S *et al*, 2005; Aretz, S. *et al*, 2006). The colorectal phenotype can be very variable however, and in a small number of patients polyp number may apparently be as low as zero in mid-life (Croitoru, M.E. *et al*, 2004; Wang, L., *et al*, 2004) or as high as over a thousand (Isidro, G., *et al*, 2004; Kanter-Smoler, G. *et al*, 2006). The risk of developing CRC is not clearly defined at present and although extracolonic manifestations have been reported in MAP patients (Nielsen, M. *et al*, 2005; Aretz, S. *et al*, 2006), many have not been described at significant frequencies making their association with MAP unclear.

A better understanding and characterization of the MAP phenotype may allow surveillance strategies to be improved with extracolonic screening implemented if necessary. Unlike FAP, no genotype-phenotype relationships have been

described to date as the phenotypes of only a relatively small number of MAP patients have been extensively characterised and the true spectrum of *MUTYH* mutations is only now becoming apparent. A collaborative study was therefore established to generate a large series of well characterized MAP cases and allow a more accurate assessment of the MAP phenotype with consequences for patient care and demonstration of any correlations between *MUTYH* genotype and the MAP phenotype.

4.2 Methods

4.2.1 Clinical samples

A collaborative project was established between The Institute of Medical Genetics (IMG), Cardiff, The Institute of Human Genetics, Bonn and The Centre of Human and Clinical Genetics, Leiden resulting in a cohort of 237 MAP patients with biallelic *MUTYH* mutations (182 index cases and 55 affected relatives).

Fifty-four MAP index cases were identified from 281 apparently unrelated index cases referred from various Regional Clinical Genetics Services, Regional Polyposis Registers and Gastroenterology departments to the All Wales Medical Genetics Service (AWMGS) for *MUTYH* germline mutation analysis. Adequate phenotypic details were not always provided with referrals but most of the patients for whom information was provided had multiple colorectal adenomas with or without CRC. The AWMGS identified 45 index cases with biallelic *MUTYH* mutations and 17 index cases with apparently monoallelic mutations (Figure I, appendix) by screening for the more frequent *MUTYH* mutations (Y176C and G393D in non-Asian cases and additionally Y101X and E477X in Asian cases). Biallelic *MUTYH* mutations were also detected in 16 affected relatives by the AWMGS (Figure I, appendix).

I screened the *MUTYH* ORF for germline mutations in 17 index cases found to carry one of the more frequent *MUTYH* mutations (Y176C or G393D) and in five index patients who were negative for the common *MUTYH* mutations and five further cases in whom no *MUTYH* analysis had been undertaken but who had characteristics similar to those of previously described MAP cases; less than a hundred colorectal adenomas with or without CRC or CRC at a young age, a family history of CRC or colorectal adenomas consistent with a recessive inheritance or previous negative mutation screening of the *APC* gene (shaded boxes in Figure I, appendix). All mutations identified in *MUTYH* were confirmed by the AWMGS prior to application in clinical diagnosis or prediction.

MAP cases identified through The Institute of Medical Genetics, Cardiff were recruited into the 'Investigating cancer risk in people with *MUTYH* gene

changes' study (MREC for Wales ref. 06/MRE09/19) or the 'Genes and Multiple Colorectal Adenomas- The Polyp Study' (MREC for Wales ref. 02/9/22). Medical history details were obtained from Regional Clinical Genetics Services and additional information was obtained, if required, from the patients directly via questionnaires or phone calls. Consent forms, information letters and questionnaires are shown in the appendix. Pathology reports were requested from hospital histopathology departments if necessary. An honorary NHS contract was obtained for me for the duration of this study to enable these clinical elements of the work.

The Institute of Human Genetics, Bonn identified 73 unrelated index MAP cases following sequencing of the *MUTYH* ORF in adenomatous polyposis patients referred for *APC* and/or *MUTYH* testing. Fourteen affected relatives were also identified as biallelic *MUTYH* mutation carriers and are included in the present study. Fifty-eight of these MAP cases (52 index cases and 6 affected relatives) have been previously reported (Aretz, S. *et al*, 2006). The Centre of Human and Clinical Genetics, Leiden identified 55 unrelated index MAP cases through a combination of denaturing gradient gel electrophoresis (DGGE) analysis and sequencing of the *MUTYH* ORF in colorectal adenoma patients either referred to their DNA Diagnostic Laboratory or identified through the Netherlands National Polyposis Register, Leiden. Forty of these index biallelic *MUTYH* mutation carriers have been previously reported (Nielsen, M. *et al*, 2005). Twenty-five affected relatives were also identified as biallelic *MUTYH* mutation carriers and are included in the present study. Informed consent was obtained and clinical and pathological data were obtained from patient records for the German and Dutch cases as for cases identified through the Institute of Medical Genetics, Cardiff.

4.2.2 Screening for mutations in MUTYH

Exons 1-16 of *MUTYH* were PCR amplified and screened using automated sequencing as described in section 3.2.3. All mutations were confirmed by sequencing an independent PCR product.

4.2.3 Assays for sequence variants in controls

ARMS assays were designed to determine the presence and frequency of rare missense or splice site variants in at least 716 control chromosomes (Table 4.1).

4.2.4 MUTYH transcript analysis

A 439bp region of cDNA from exon 6 to exon 10 was PCR amplified using primers MYHx6F_Q216Q (5'-GTGGCCTACACTGCAGGACC) and MYHx10R_Q216Q (5'-AGGGCACTGGCTGCACAG) at an annealing temperature of 58°C to characterize the *MUTYH* mutation Q227Q (c.681 G→A). The RT-PCR products were cloned then sequenced using the same primers (MYHx6F_Q216Q & MYHx10R_Q216Q).

4.2.5 Statistical analysis

The chi-squared test or Fisher's exact test were used to assess differences in proportions. The chi-squared test was used to assess trends in proportions. Univariate ANOVA (Analysis of Variance) was used to compare the mean ages at presentation and CRC diagnosis between different genotypes. The relationship between number of Y176C alleles and ages at presentation and CRC diagnosis was assessed using linear regression analysis. A probability (p) of <0.05 was considered statistically significant and all tests were carried out with either SPSS 12 or Minitab 14.

Variant	Primer name	Primer sequence (5'→3')	Anneal temp (°C)	Product size (bp)
G213E (c.638 G→A)	MUTYH_G202E_GF	CAGCTCCTGCCTGGCGTCGG	60	141
	MUTYH_G202E_AF	CAGCTCCTGCCTGGCGTAGA		
	MUTYH_G202E_R	AGAGGGGCCAAAGAGTTAGC		
Q227Q (c.681 G→A)	MUTYH_Q216Q_GF	GCCTCTATCGCCTTTGGCGAG	60	146
	MUTYH_Q216Q_AF	GCCTCTATCGCCTTTGGCTAA		
	MUTYH_Q216Q_R	ACAGCACCCGTGCTACGTTG		
N235S (c.704 A→G)	MUTYH_N224S_AF	ACCGGTGTGGTGGATGGTAA	60 (A) 58 (G)	149
	MUTYH_N224S_GF	ACCGGTGTGGTGGATGGGAG		
	MUTYH_N224S_R	GCAGAGCTCCTTTGCAGAC		
R242H (c.725 G→A)	MUTYH_R231H_GF	CGTAGCACGGGTGCTGTGGCG	58 (G) 60 (A)	149
	MUTYH_R231H_AF	CGTAGCACGGGTGCTGTGTCA		
	MUTYH_R231H_R	GAAGGGAACACTGCTGTGAAG		
L417M (c.1249 C→T)	MUTYH_L406M_(1)_CF	GCAGCTTCAGCGCAAGGGCC	58	141
	MUTYH_L406M_(1)_AF	GCAGCTTCAGCGCAAGGTCA		
	MUTYH_L406M_(1)_R	GCCTGTGGATATAGCCTCAA		
T474M (c.1421 C→T)	MUTYH_T463M_CF	CCAGGTGCTCGCTGGCTCAC	56	138
	MUTYH_T463M_TF	CCAGGTGCTCGCTGGCTAAT		
	MUTYH_T463M_R	TACACAGTAATATATTCATGTAG		

Table 4.1 ARMS primers and conditions used to assay for specific rare *MUTYH* variants in controls. Internal control primers were used to validate the assays; TSC2_Ex26F (5'-GAGCTTTGGCCCTTGGTGATA) and TSC2_Ex26R (5'-CTCGCCCACAGGAGACCTAGA), product size 388bp.

4.3 Results & Discussion

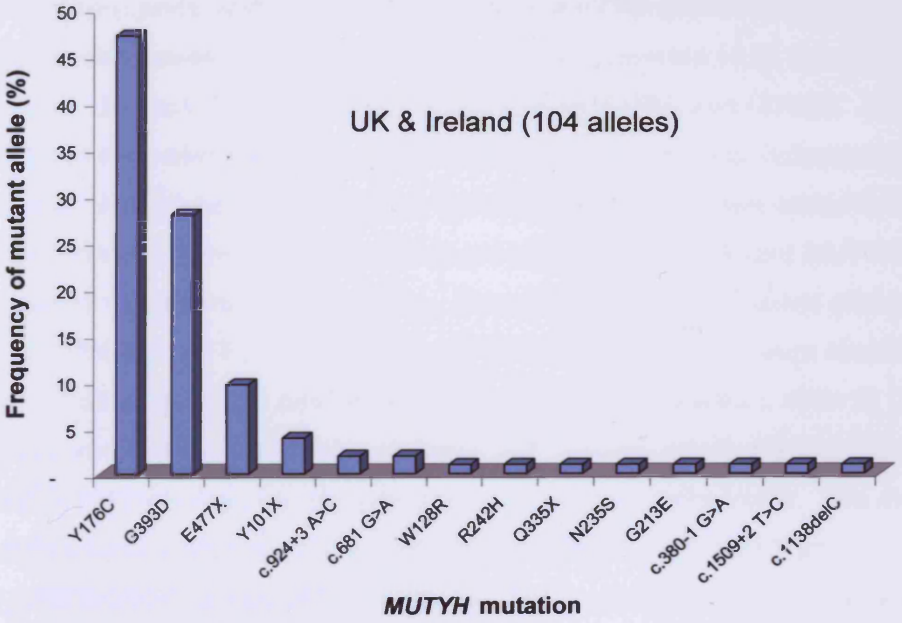
4.3.1 Spectrum of biallelic MUTYH mutations

A total of 36 different *MUTYH* mutations were detected in 182 unrelated MAP index cases referred for genetic testing in Cardiff, Bonn and Leiden. Twenty-six of these had been reported previously in MAP patients with biallelic *MUTYH* mutations. Two of the remaining ten mutations, R306C and R106W, were identified in single Dutch cases in combination with Y176C and G393D respectively, and had been previously reported but only as monoallelic *MUTYH* mutations (Sieber, O.M *et al*, 2003; Olschwang, S. *et al*, 2007). The novel *MUTYH* mutations R19X, V212M, R238Q and G247D were identified in Bonn and I identified four novel mutations among patients referred to IMG, Cardiff; G213E, N235S, c.681 G>A and c.1092delC.

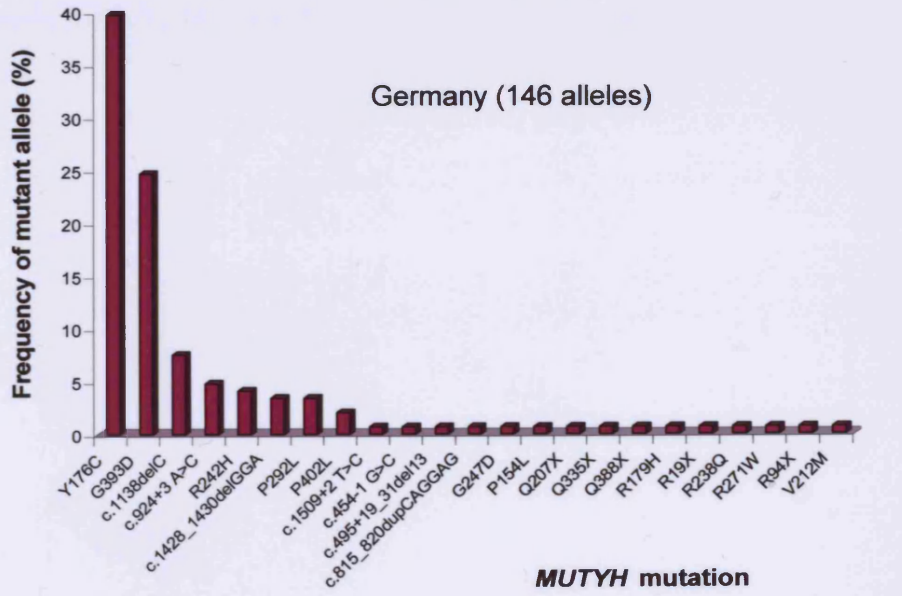
Although Y176C and G393D were the most frequent mutations in all three centres, there were some differences in the nature and frequencies of other mutations (Figure 4.1). As the strategy for mutation screening in most UK cases included a pre-screen for the common mutations, compound heterozygosity or homozygosity for rare mutations may have gone undetected in some cases.

To date, 54 different *MUTYH* mutations have been reported in the literature in patients with biallelic mutations. Together with the additional 10 mutations described above, their distribution along the *MUTYH* gene is shown in Figure 4.2. Mutations have been identified along the length of the gene in all exons except the extreme 5' (exon 1) and 3' (exon 16) ends and also in introns 4-6, 8, 10, 12, 13, 15. Nearly half are missense changes (30 mutations, 47%), nonsense mutations account for a fifth, nine (14%) are predicted to affect splicing, nine are small insertions or deletions and three in-frame insertions or deletions have been identified. More than one change has been detected at codons 242 (R242H, R242C, R242L), 238 (R238W, R238Q) and 179 (R179H, R179C), with nucleotide c.1092 affected by both a small insertion and deletion. Two splice site mutations have been described at the same nucleotide in intron 4 (c.380-1 G>A, c.380-1 G>C) but all other mutations affected different codons and nucleotides.

A)


Frequency of mutant allele (%)
50
45
40
35
30
25
20
15
10
5
-
UK & Ireland (104 alleles)
Y176C
G393D
E477X
Y101X
c.924+3 A>C
c.681 G>A
W128R
R242H
Q335X
N235S
G213E
c.380-1 G>A
c.1509+2 T>C
c.1138delC
MUTYH mutation

B)


Frequency of mutant allele (%)
40
35
30
25
20
15
10
5
0
Germany (146 alleles)
Y176C
G393D
c.1138delC
c.924+3 A>C
R242H
c.1428_1430delGGA
P292L
P402L
c.1509+2 T>C
c.454-1 G>C
c.495+19_31del13
c.815_820dupCAGGAG
G247D
P154L
Q207X
Q335X
Q388X
R179H
R19X
R238Q
R271W
R94X
V212M
MUTYH mutation

C)

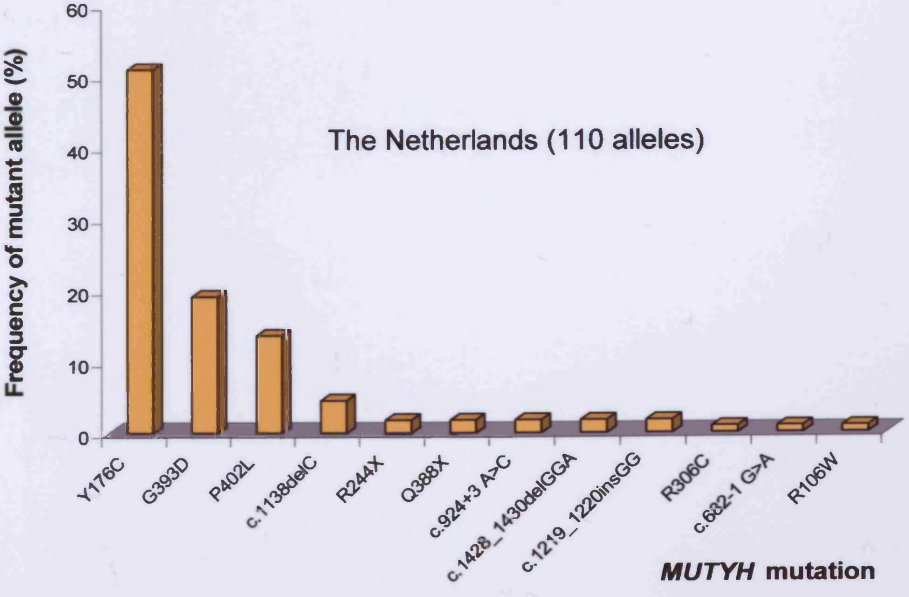

Frequency of mutant allele (%)
60
50
40
30
20
10
0
The Netherlands (110 alleles)
Y176C
G393D
P402L
c.1138delC
R244X
Q388X
c.924+3 A>C
c.1428_1430delGGA
c.1219_1220insGG
R306C
c.682-1 G>A
R106W
MUTYH mutation

Figure 4.1 Frequency and spectrum of mutant *MUTYH* alleles in 182 unrelated European index cases. A) Mutant *MUTYH* allele frequencies in 52 cases from the UK and Ireland. The most frequently mutated alleles are Y176C and G393D. All E477X and Y101X mutant alleles were identified in patients of Indian and Pakistani descent, respectively. Biallelic *MUTYH* mutations were also identified in two cases living outside the UK and Ireland so these mutant alleles are not shown. **B)** Mutant *MUTYH* allele frequencies in 73 patients from Germany. The most frequently mutated alleles are Y176C and G393D. All P292L, V212M and R179H mutant alleles were identified in patients of Turkish or Arabic descent and the single mutant alleles c.495+19_30del13 and Q335X were carried by patients of Greek and Russian origin, respectively. **C)** Mutant *MUTYH* allele frequencies in 55 patients from the Netherlands. The most frequently mutated alleles are Y176C, G393D and P402L. All Q388X and c.1428_1430delGGA mutant alleles were identified in patients of Turkish descent and only a case of Moroccan descent carried c.1219_1220insGG mutation.

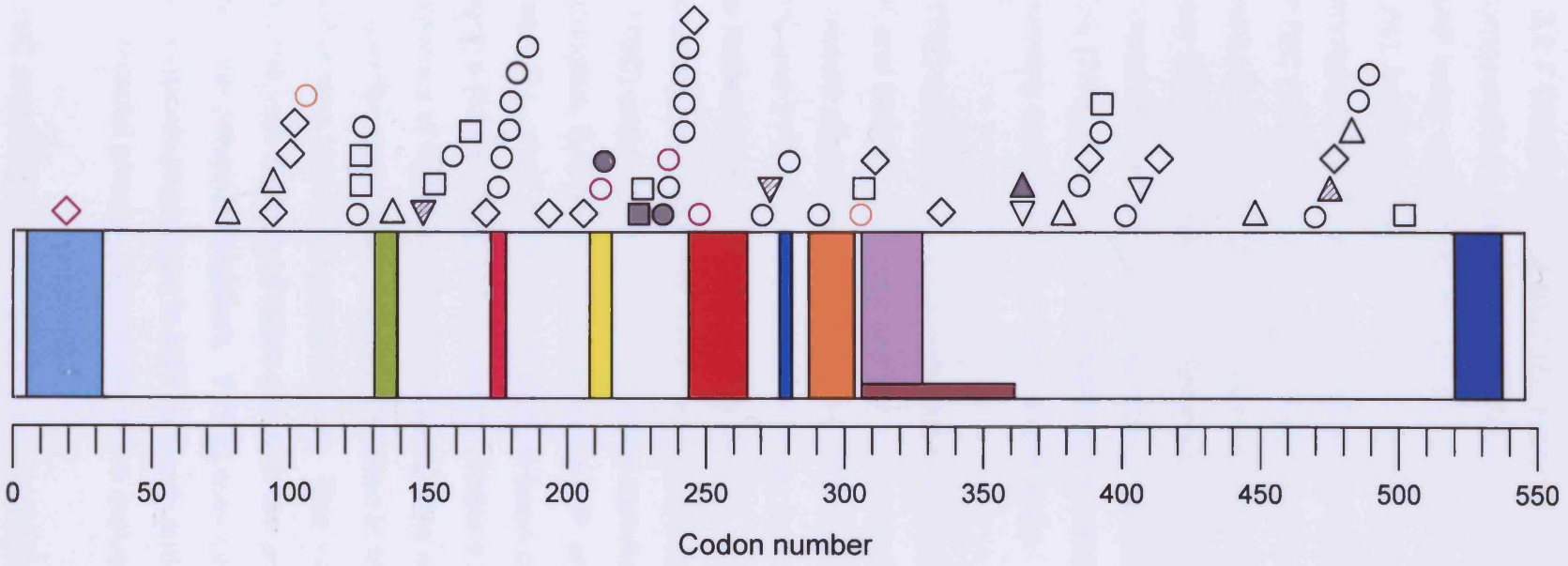


Figure 4.2 Spectrum and distribution of germline *MUTYH* mutations. Only mutations from patients with biallelic *MUTYH* mutations are shown. Novel mutations I identified are shown as grey-filled shapes and novel mutations identified by the German and Dutch centres are outlined in purple and orange, respectively. References for the previously reported mutations are shown in appendix Table I.

- Missense
- Nonsense
- Small deletion
- Small deletion (in-frame)
- Small insertion
- Small insertion (in-frame)
- Splice site

- RPA binding site
- DNA minor groove reading motif
- Pseudo HhH motif
- HhH motif
- MSH6 binding site
- Adenine recognition motif
- Fe-S cluster
- APE1 binding site
- Hus1 binding site
- PCNA binding motif

4.3.2 Mutation frequency in MUTYH

4.3.2.1 Comparison of MUTYH mutation frequencies between populations

Compound heterozygosity for Y176C/G393D was the most frequent genotype in MAP index cases from the UK and Ireland (16/52, 31%) and Germany (17/73, 23%), but not in cases from the Netherlands (8/55, 15%). The most common genotype in MAP index patients from the Netherlands was homozygosity for Y176C (18/55, 33%). This was the second most frequent genotype in the UK and Ireland (13/52, 25%) and Germany (12/73, 16%). G393D homozygotes were found at a lower frequency than Y176C homozygotes and Y176C/G393D compound heterozygotes in all three populations, accounting for 8% (4/52), 10% (7/73) and 7% (4/55) of MAP index cases from the UK and Ireland, Germany and the Netherlands, respectively.

Y176C was the most frequently mutated allele in MAP index patients from the UK and Ireland, Germany and the Netherlands, contributing 47%, 40% and 51% of mutant alleles, respectively, while the G393D mutation accounted for 28%, 25% and 21% of mutant alleles in cases from the UK and Ireland, Germany and the Netherlands, respectively (Figure 4.1). Current data suggests that these mutations occur at minor allele frequencies (MAFs) of approximately 0.3% (Y176C) and 0.7% (G393D) in healthy non-Asians (Croitoru, M.E. *et al*, 2004; Farrington, S.M. *et al*, 2005; Peterlongo, P. *et al*, 2006; Tenesa, A. *et al*, 2006; Webb, E.L. *et al*, 2006) and both have been commonly identified in MAP patients from a range of populations (Table 4.2, Figure 4.3). Given the higher frequency of G393D in healthy controls, this change would be expected to account for more mutant alleles identified in MAP cases than Y176C but the inverse was true in the present study. This may be due to an ascertainment bias; the vast majority of index cases in the present study were referred from colorectal polyposis registers. Y176C may be more likely than G393D to result in a polyposis phenotype in MAP patients and G393D mutations may account for colorectal phenotypes outside of the inclusion criteria for such registers.

Y176C and G393D mutations are not consistently associated with specific alleles at D1S2667 which suggests that there are likely to have been recurrent

MUTYH Mutation	Proportions of different mutant alleles in a particular population/ethnic group													
	Dutch	**Swedish**	**British & Irish**	**French**	**Swiss**	**German**	**Italian**	**Spanish**	**Portuguese**	**Australian**	**American**	**Canadian**	**Indian**	**Turkish & Arabic**
Y176C	57/106 (54%)	5/10 (50%)	68/158 (43%)	15/43 (35%)	6/12 (50%)	61/156 (39%)	33/100 (33%)	6/16 (38%)	20/46 (43%)	18/40 (45%)	56/112 (50%)	9/24 (38%)	/	/
G393D	22/106 (21%)	1/10 (10%)	69/158 (44%)	11/43 (26%)	4/12 (33%)	42/156 (27%)	17/100 (17%)	5/16 (31%)	11/46 (24%)	16/40 (40%)	40/112 (36%)	12/24 (50%)	/	/
1138delC	5/106 (5%)	/	2/158 (1%)	3/43 (7%)	/	12/156 (8%)	9/100 (9%)	1/16 (6%)	1/46 (2%)	/	2/112 (2%)	/	/	/
c.924+3 A>C	2/106 (2%)	1/10 (10%)	2/158 (1%)	1/43 (2%)	/	8/156 (5%)	2/100 (2%)	/	/	2/40 (5%)	3/112 (3%)	2/24 (8%)	/	/
P402L	**15/106 (14%)**	**3/10 (30%)**	/	1/43 (2%)	/	3/156 (2%)	/	/	/	/	/	/	/	/
E477X	/	/	/	/	/	/	/	/	/	/	2/112 (2%)	/	**10/10 (100%)**	/
Y101X	/	/	/	/	/	/	6/100 (6%)	/	/	/	/	1/24 (4%)	/	/
c.1428_1430delGGA	/	/	1/158 (0.6%)	/	/	7/156 (4%)	**15/100 (15%)**	/	/	3/40 (8%)	4/112 (4%)	/	/	**5/14 (36%)**
c.1219_1220insGG	/	/	/	**5/43 (12%)**	/	/	3/100 (3%)	**4/16 (25%)**	**8/46 (17%)**	/	/	/	/	/
P292L	/	/	/	1/43 (2%)	/	/	1/100 (1%)	/	/	/	/	/	/	**5/14 (36%)**

Table 4.2 Prevalence of particular *MUTYH* mutations in MAP cases from different populations. Only mutations identified in unrelated MAP patients with biallelic *MUTYH* mutations and either previously reported or part of the present study are included. Populations with at least 10 reported mutant alleles are shown. Mutations were selected for a frequency of 5% or more in at least one population which was due to more than one mutant allele. Mutations identified in patients who were reported to be descended from one population but living in another were counted as their population of descent. Frequencies highlighted in bold show a higher prevalence of the mutation in particular populations. References are shown in appendix Table I.

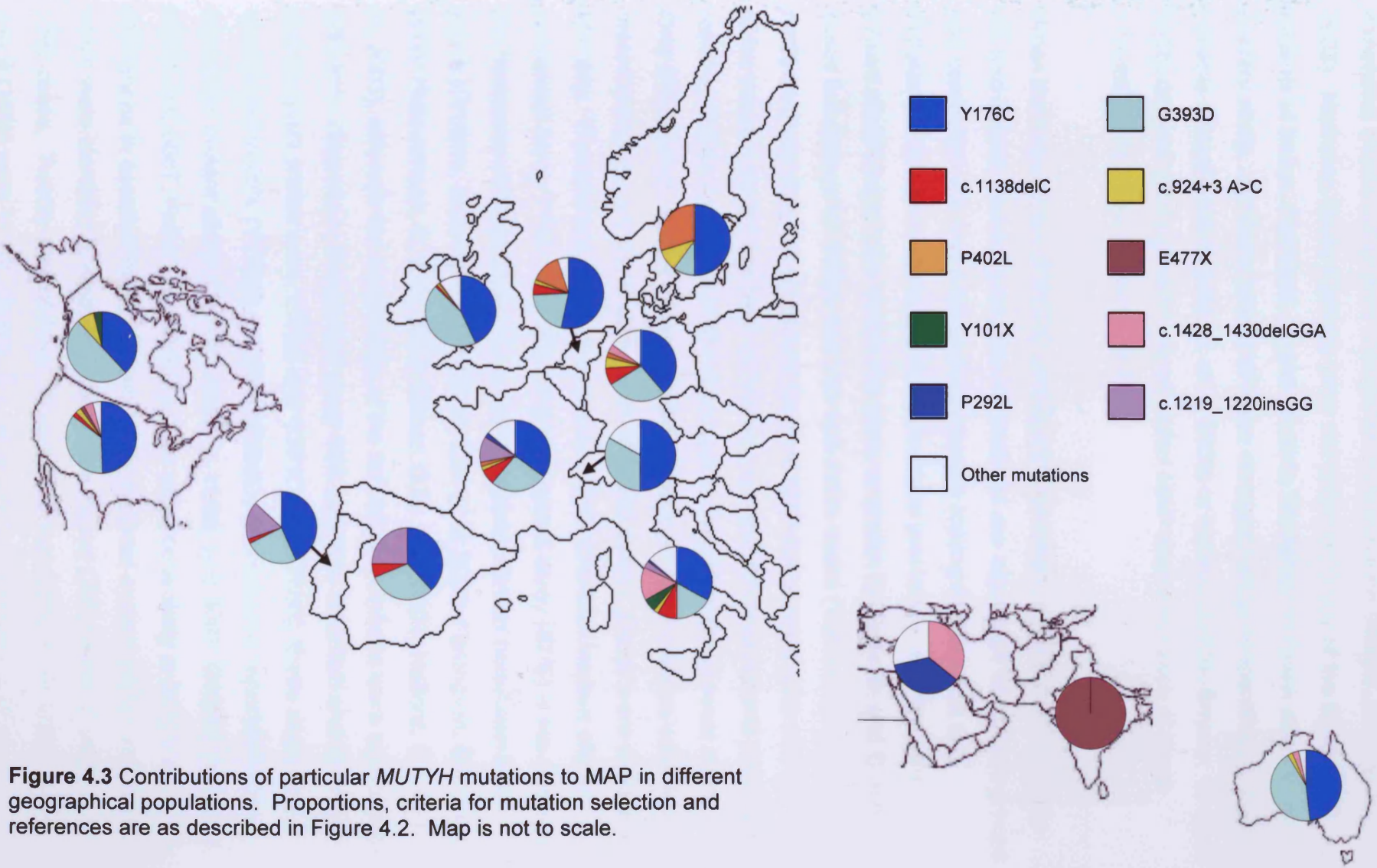


Figure 4.3 Contributions of particular *MUTYH* mutations to MAP in different geographical populations. Proportions, criteria for mutation selection and references are as described in Figure 4.2. Map is not to scale.

mutational events rather than a single ancestral change (Sieber, O.M. *et al*, 2003). However, these changes were not detected in any of the index MAP patients of Indian, Pakistani, Turkish, Arabic, Moroccan or Greek ethnicity in the present study. Similarly, neither of these changes has been identified in MAP patients of Japanese (Miyaki, M. *et al*, 2005) or Iranian (Kanter-Smoler, G. *et al*, 2006) descent although only one unrelated MAP case from each of these populations has been reported to date.

When the frequencies of Y176C and G393D identified in patients living in the UK and Ireland, Germany and the Netherlands are separated into ethnic groups and combined with data from previous studies (taking into account any mutations in the present study which have been previously reported), the mutant allele frequencies remain the same or similar for German and Dutch cases but change for British and Irish non-Asian cases (Table 4.2). This is partly because very few Dutch and German MAP index cases who are not part of the present study have been previously reported whereas several previously reported British non-Asian MAP patients were not part of the present study. Only 5% of Dutch mutant alleles and 6% of German mutant alleles were from recent immigrant groups whereas 13% of British mutant alleles were of Asian ethnicity. The total proportion of British and Irish non-Asian mutant alleles accounted for by Y176C and G393D in the present study (87%) is the same as the frequency of these alleles in previously reported British non-Asian MAP cases (Croitoru, M.E. *et al*, 2007; Webb, E.L. *et al*, 2006; Farrington, S.M. *et al*, 2005; Fleischmann, C. *et al*, 2004; Sieber, O.M. *et al*, 2003; Halford, S.E.R. *et al*, 2003), although the frequencies of the individual mutations were significantly different. Whereas in the present study 54% and 32% of British and Irish non-Asian mutant alleles were Y176C and G393D respectively, these mutations contributed to 28% (Y176C) and 59% (G393D) of previously reported British non-Asian mutant alleles (Y176C; 49/90 vs 19/68, p=0.0007, G393D; 29/90 vs 40/68, p=0.0007, Fisher's exact test). This variation is likely to be the result of differences in ascertainment; the previously reported mutant alleles included 38 which were identified in studies of population-based CRC cases or early onset CRC cases. Twenty-nine of these 38 mutant alleles (76%) were G393D and only 6 (16%) were Y176C (Webb, E.L. *et al*, 2006; Farrington, S.M. *et al*, 2005;

Fleischmann, C. *et al*, 2004; Halford, S.E.R. *et al*, 2003). The vast majority of index cases in the present study were referred from colorectal polyposis registers and included multiple colorectal adenoma patients with and without CRC. Y176C may therefore be more likely than G393D to result in a polyposis phenotype. Y176C is more frequently mutated than G393D in reported non-UK European, Australian and American MAP patients (Table 4.2, Figure 4.3).

Several mutations have been identified previously in more than one population but some appear to occur at higher frequencies in certain populations (Table 4.2, Figure 4.3).

4.3.2.2 Overall prevalence of MUTYH mutations

A total of 766 mutant *MUTYH* alleles have been reported in apparently unrelated MAP index cases to date and together with previously unreported mutant alleles in the present study, nearly nine hundred have been described. As shown in Figure 4.4, Y176C and G393D are ten and seven times more prevalent than any other single mutation, respectively. Together they account for 70% (631/896) of reported mutant alleles but as described above there has been a reporting bias towards these two mutations. Over twenty reports have been published to date in which biallelic *MUTYH* patients have been identified by analysis of the *MUTYH* ORF without pre-screening and as more of these unbiased screens are reported, more accurate frequencies of individual mutations among MAP patients will be obtained with implications for molecular diagnostic testing for MAP.

The remaining 62 reported *MUTYH* mutations each contribute less than 5% of mutant alleles described in unrelated MAP index cases. The most frequent of these less common mutations are c.1138delC and c.1428_1430delGGA which each account for 4% of mutant alleles. Thirty-seven *MUTYH* mutations have only been described in a single MAP index case, of which four were identified in the homozygous state (Figure 4.4).

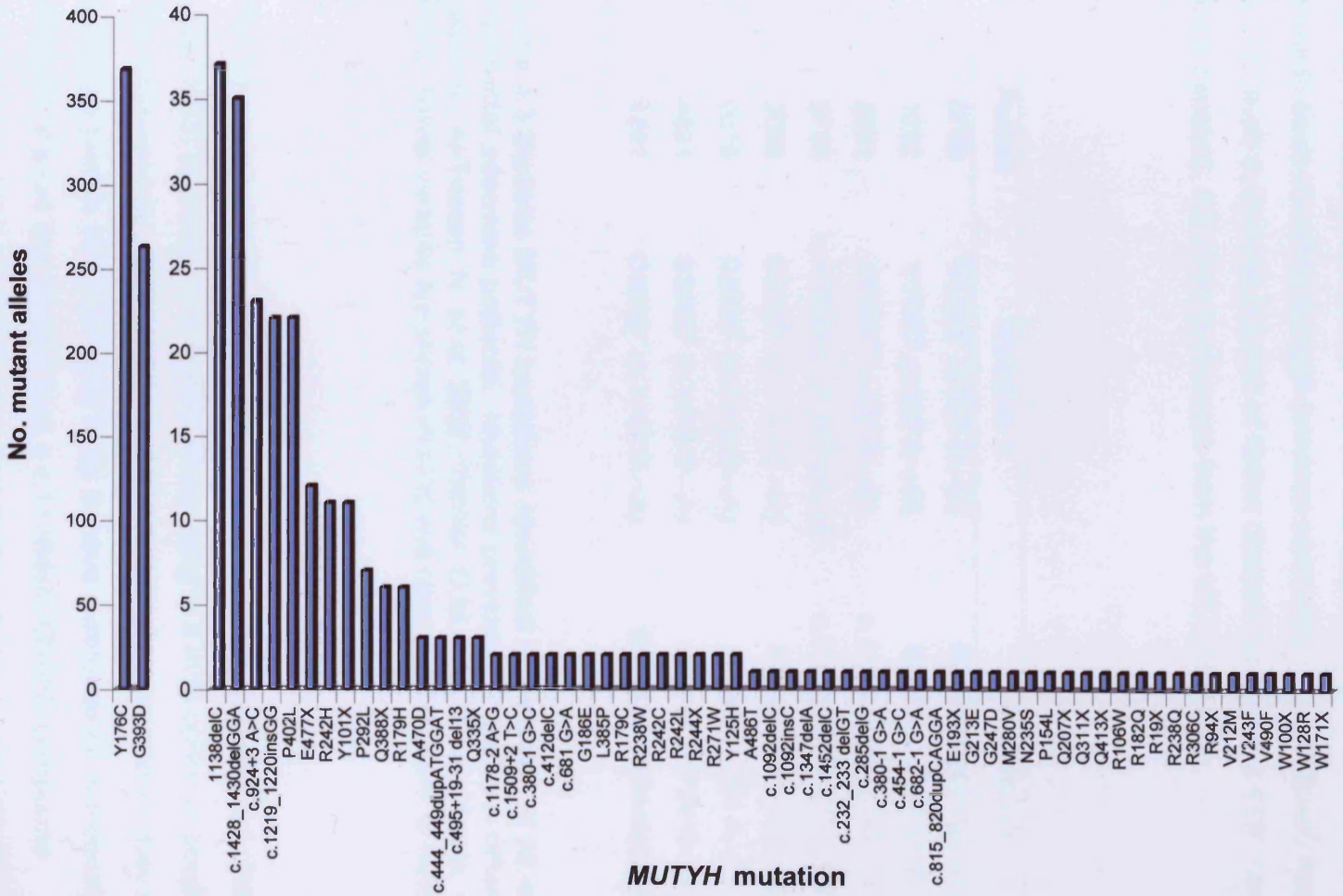


Figure 4.4 Frequency of *MUTYH* mutations previously reported or identified in MAP cases in the present study. Only mutations identified in unrelated MAP patients with biallelic *MUTYH* mutations are included. Y176C and G393D are ten and seven times more prevalent than any other single mutation, respectively. References are shown in appendix Table I.

4.3.3 Rare and novel pathogenic MUTYH mutations identified in Cardiff

Biallelic *MUTYH* mutations were identified in eight of the 26 unrelated index cases I screened (Table 4.3). Y176C and G393D mutations in these cases had already been identified through previous screening. A c.1138delC homozygote (3718) lived in Canada but was of Italian descent and patient 4891 came from New Zealand. All other cases were from the UK or Ireland.

Patient	Mutation 1	Mutation 2
0748	Y176C[1] (c.527 A→G)	**Q227Q (c.681 G→A)**
1232	Y176C[1] (c.527 A→G)	**Q227Q (c.681 G→A)**
2921	Y176C[1] (c.527 A→G)	*c.1138delC[2] (L370fsX404)*
3718	*c.1138delC[2] (L370fsX404)*	*c.1138delC[2] (L370fsX404)*
2058	G393D[1] (c.1178 G→A)	**G213E (c.638 G→A)**
0376	G393D[1] (c.1178 G→A)	**N235S (c.704 A→G)**
4021	G393D[1] (c.1178 G→A)	*R242H[3] (c.725 G→A)*
4891	G393D[1] (c.1178 G→A)	**1092delC (P364fsX404)**

Table 4.3 Biallelic *MUTYH* mutations identified in a screen of 26 multiple colorectal adenoma patients. Mutations previously reported in other MAP patients; [1]Al-Tassan, N. *et al*, 2002, [2]Sieber, O.M. *et al*, 2003, [3]Aceto, G et al, 2005. Novel variants are shown in bold and rare mutations are in italics.

The infrequent c.1138delC mutation has been previously reported (Sieber, O.M. *et al*, 2003) and leads to a frameshift resulting in a stop codon at position 404. The pathogenicity of frameshift mutations is usually assumed as they result in a truncated protein which is highly likely to have a detrimental functional effect. Analysis of a cell line derived from a c.1138delC/G393D compound heterozygote MAP patient showed that these cells have approximately half the wild-type level of MUTYH as a result of protein instability caused by the c.1138delC mutation (Parker, A.R. *et al*, 2005).

The previously reported R242H rare missense mutation (Aceto, G. *et al*, 2005) was not found in 726 control chromosomes and is predicted to abrogate putative exonic splicing enhancers (ESEs) responsive to SR proteins SF2/ASF and SC35 (ESEfinder version 3.0, http://rulai.cshl.edu/tools/ESE/, Cartegni, L. *et al*, 2003). Functional analysis of other *MUTYH* mutations that lie adjacent to, or within, the MSH6 binding domain of the protein (R238W, R242L and V243F) has been undertaken. None of the variants affect the physical interaction of MUTYH with MSH6 but these mutant MUTYH proteins exhibit defects in 8-oxoG:A binding and adenine glycosylase activities. In addition, none of these enzymes can complement MutY deficiency in *E. coli* (Bai, H. *et al*, 2005; Bai, H. *et al*, 2007) suggesting that R242H is likely to have an effect on the protein.

G213E is a novel missense mutation (Figure 4.5) and was found in combination with G393D in a single patient. This variant is a non-conservative amino acid change at an evolutionarily conserved site, is predicted to have a probably damaging effect on the protein (PolyPhen; http://genetics.bwh.harvard.edu/pph/) and was not detected in 724 control chromosomes. The glycine residue at codon 213 is conserved through evolution (Figure 4.6) and the equivalent amino acid in *E. coli* MutY (Gly118) forms part of a HhH motif, a highly conserved structural element found in the catalytic domain of several DNA glycosylases (Guan, Y. *et al*, 1998). Crystal structures of *E. coli* MutY have shown the protein forms a positively-charged DNA binding groove which enables the HhH motif to approach the DNA backbone. Gly118 is located within the hairpin turn of this HhH motif (Figure 4.6) and interacts with DNA phosphates. The pseudo HhH motif also has the ability to approach the DNA backbone and together both elements may compress the DNA intrastrand phosphate distance either side of the mispaired adenine causing the DNA to bend and the target base to flip out of the DNA helix, into the active site pocket (Guan, Y *et al*, 1998). The non-conservative substitution of G213E results in a negatively-charged amino acid which would repel DNA phosphates and may therefore affect the positioning of adenine into the active site.

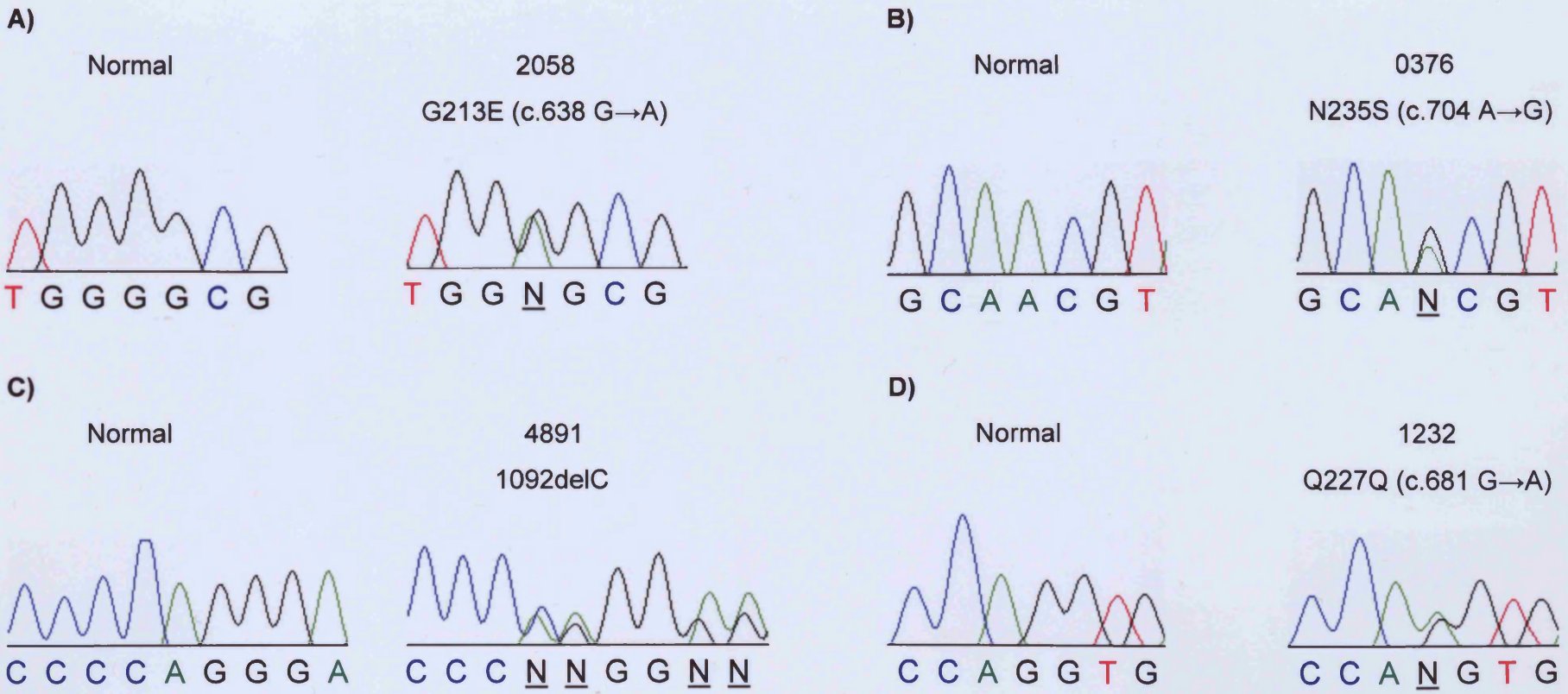


Figure 4.5 Novel *MUTYH* mutations identified in multiple colorectal adenoma cases. All mutations were identified in combination with either Y176C or G393D. Normal sequences are shown on the left and corresponding mutant sequences on the right. N indicates position of mutation.

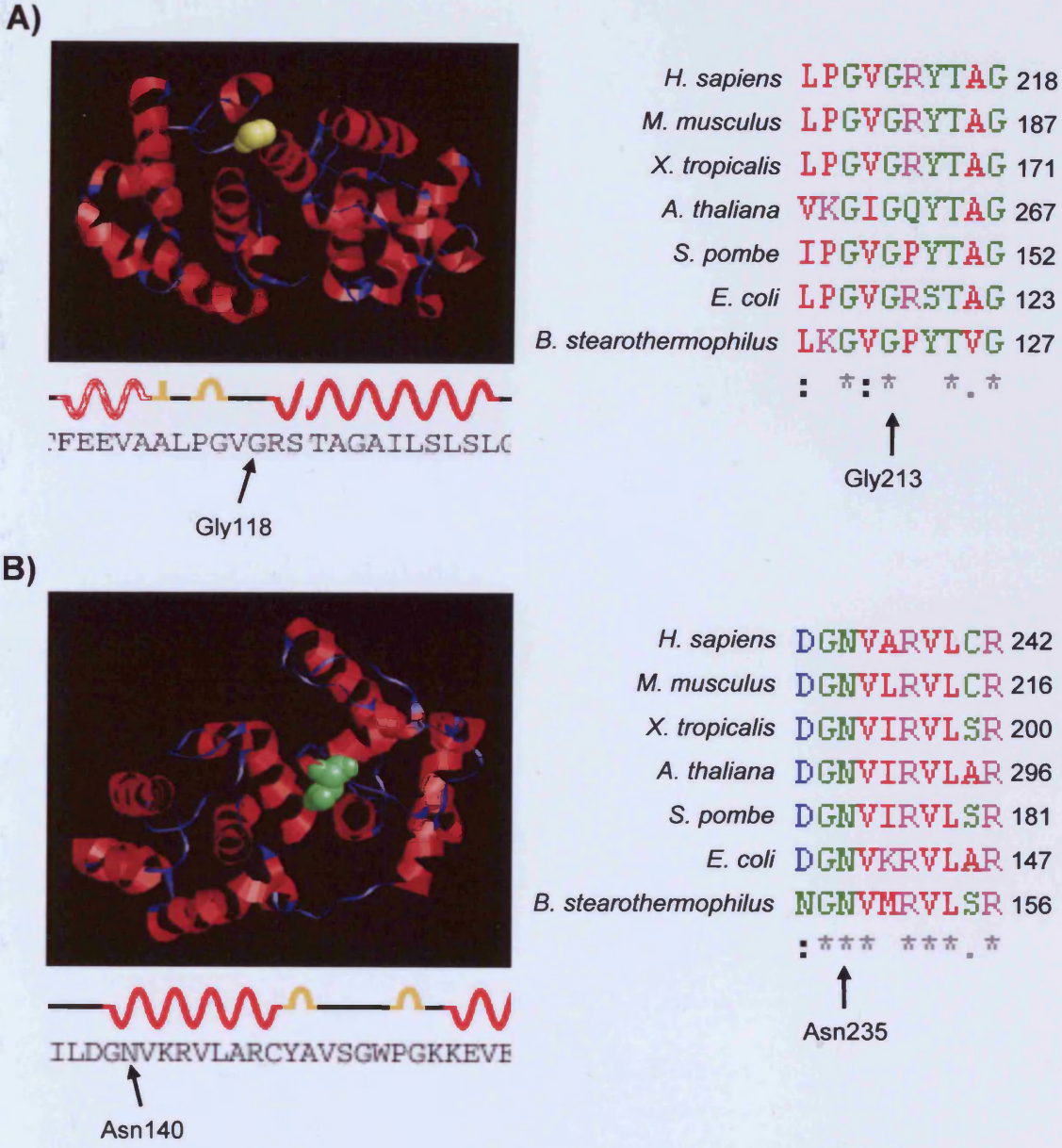


Figure 4.6 Crystal structure of the *E. coli* MutY catalytic domain and evolutionary conservation of MUTYH. A) Gly118 of *E. coli* MutY (left, shown as yellow spacefill) corresponds to Gly213 of human MUTYH and lies within the hairpin turn of a HhH motif, as indicated by the arrow (below) (RCSB Protein Data Bank www.rcsb.org/pdb/). The glycine at residue 213 of MUTYH is highly conserved through evolution, as indicated by the arrow (right). **B)** Asn140 of *E. coli* MutY (left, shown as green spacefill) corresponds to Asn235 of human MUTYH and forms part of an α-helix as indicated by the arrow (below) (RCSB Protein Data Bank www.rcsb.org/pdb/) which is thought to form part of the enzyme's active site. The asparagine at residue 235 of MUTYH is highly conserved through evolution, as indicated by the arrow, right. PDB file 1MUY was used to generate the figures using RasMol v2.6-β-2 (http://openrasmol.org/) and α-helices are shown in red. Alignment of human (NP_036354), mouse (NP_573513), *Xenopus* (NP_001072831), *Arabidopsis* (NP_193010), fission yeast (AAC36207), *E. coli* (NP_417436) and *B. stearothermophilus* (1RRQA) MUTYH homologues was carried out using ClustalW (right). Amino acids are coloured according to their biochemical properties.

The novel N235S mutation (Figure 4.5) is a semi-conservative amino acid change at a site conserved through evolution and is predicted to have a possibly damaging effect on the protein (PolyPhen; http://genetics.bwh.harvard.edu/pph/). This variant was detected in a single patient together with G393D but was not found in 722 control chromosomes. N235S occurs at a residue conserved through evolution (Figure 4.6) and the equivalent amino acid in *E. coli* MutY (Asn140) forms part of the enzyme's active site (Guan, Y. *et al*, 1998). Crystal structures of *E. coli* MutY suggest that the amide group of Asn140 hydrogen-bonds to the sugar residue of the target adenine nucleotide (Guan, Y. *et al*, 1998). The N235S substitution results in a change in the amino acid side chain from an amide to a hydroxyl group. The serine side chain is still capable of hydrogen bonding but spatial positioning of this group may be such that bonding is weaker or may even be disrupted. No other *MUTYH* mutations have been reported to affect this codon but a human R238W mutant MUTYH protein shows defective 8-oxoG:A binding and adenine glycosylase activities (Bai, H. *et al*, 2005).

The only novel small deletion identified, 1092delC (Figure 4.5), was found in a single patient in combination with G393D. This mutation results in a frameshift from codon 364 and creates a stop codon at position 404. The resulting truncated protein lacks the functionally important C-terminal domain needed for substrate recognition (Noll, D.M. *et al*, 1999; Fromme, J.C. *et al*, 2004), the PCNA binding motif (Parker, A. *et al*, 2001) and the NLS (Takao, M. *et al*, 1999). The 1138delC frameshift mutation also results in a stop codon at residue 404 so 1092delC is likely to result in protein instability and reduced cellular levels of MUTYH (Parker, A.R. *et al*, 2005). An insertion mutation of a single cytosine at nucleotide 1092 resulting in a stop codon near the C-terminus of the protein at residue 529 has been previously reported (Lejeune, S. *et al*, 2006). Both mutations occur at a polycytosine $(C)_6$ tract and are likely to be the result of replication slippage as repetitive sequences are more prone to this error.

The novel splice site mutation c.681 G→A (Figure 4.5) was found together with Y176C in two apparently unrelated cases and was not detected in 722 control

chromosomes. This mutation affects the last nucleotide of exon 8 so transcript analysis was undertaken to determine what effect the variant had on splicing. In addition to the predicted 439bp fragment, RT-PCR from exons 6 to 10 in patient 0748 revealed two smaller mRNA species, one of which appeared to be much less abundant (Figure 4.7). Sequencing of the cloned products showed that 46% of transcripts underwent normal splicing but harboured the Y176C mutation (located in exon 7), 46% of transcripts skipped exon 8 and did not carry the Y176C mutation and 8% of transcripts skipped exons 8 and 9 and did not carry the Y176C mutation (Figure 4.7). Exon 8 skipping results in an in-frame deletion of 38 amino acids which, based on crystal structures of *E. coli* MutY, include a region of the six helix barrel domain which contains the HhH motif. As described above, the HhH element interacts with the DNA backbone and works with the pseudo HhH motif to bend DNA and facilitate flipping of the target adenine base into the active site (Guan, Y *et al*, 1998). If both exons 8 and 9 are skipped, a frameshift results that leads to a stop codon 65 amino acids downstream. The truncated protein would lack most functional elements including the entire C-terminal domain which is necessary for substrate specificity (Noll, D.M. *et al*, 1999; Fromme, J.C. *et al*, 2004). Wooden *et al* suggested that a truncated human MUYTH protein (residues 1-135) may have a toxic effect on cells as *E. coli* expressing this mutant exhibited slowed growth (Wooden, S.H. *et al*, 2003). The 1138delC frameshift mutation leads to a stop codon further downstream than that caused by skipping of exons 8 and 9 yet has been shown to result in protein instability in human cells (Parker, A.R. *et al*, 2005) which may also be the case for c.681 G>A. A splice site mutation affecting the splice acceptor site of exon 8 has been reported, c.682-1 G>A, but further analysis was not undertaken (Nielsen, M. et al, 2005).

4.3.4 Functional analysis of pathogenic MUTYH mutations

Only a few *MUTYH* variants have undergone functional analysis, with the common Y176C and G393D mutations most comprehensively characterized. Both residues are highly conserved through evolution and are located within functionally important domains of the protein (Al-Tassan, N. *et al*, 2002). Crystal structures of homologous *E. coli* and *B. stearothermophilus* MutY proteins show the equivalent tyrosine residue is part of a pseudo HhH motif

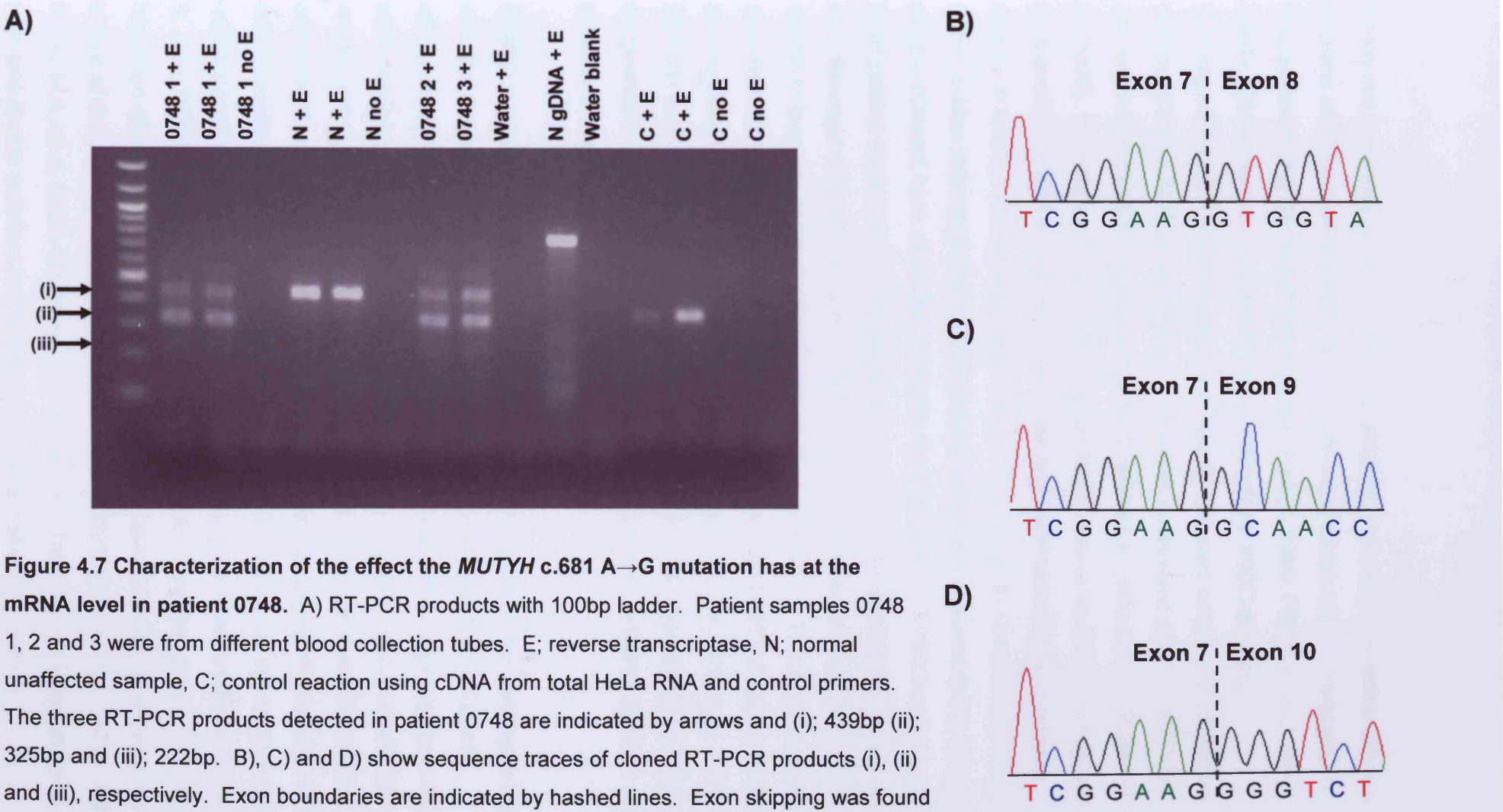


Figure 4.7 Characterization of the effect the *MUTYH* c.681 A→G mutation has at the mRNA level in patient 0748. A) RT-PCR products with 100bp ladder. Patient samples 0748 1, 2 and 3 were from different blood collection tubes. E; reverse transcriptase, N; normal unaffected sample, C; control reaction using cDNA from total HeLa RNA and control primers. The three RT-PCR products detected in patient 0748 are indicated by arrows and (i); 439bp (ii); 325bp and (iii); 222bp. B), C) and D) show sequence traces of cloned RT-PCR products (i), (ii) and (iii), respectively. Exon boundaries are indicated by hashed lines. Exon skipping was found exclusively on the allele which did not harbour the Y176C mutation.

which may promote flipping of the mispaired adenine into the active site and the tyrosine itself intercalates into DNA 5' to 8-oxoG which may be required for recognition of this damaged base (Guan, Y *et al*, 1998; Fromme, J.C. *et al*, 2004). The mutant *E. coli* MutY protein carrying a Y82C mutation (corresponding to Y176C) shows significantly reduced adenine glycosylase activity (Chmiel, N.H. *et al*, 2003), as does the equivalent mutant MUTYH mouse protein (Y150C), compared to their wild-type counterparts (Pope, M.A. *et al*, 2005). The human Y176C MUTYH protein displays no DNA binding activity on 8-oxoG:A or G:A containing substrates and lacks adenine glycosylase activity *in vitro* (Wooden, S.H. *et al*, 2003). The ability to complement *E. coli mutY*⁻ is also reduced in the Y176C mutant enzyme compared to the wild-type protein (Chmiel, N.H. *et al*, 2003). Cells derived from a Y176C homozygous MAP patient have low MUTYH levels and this enzyme exhibits minimal binding and cleavage of 8-oxoG:A and 8-oxoA:G mispairs, even when tested at levels similar to those in normal cells (Parker, A.R. *et al*, 2005). This mutation may also alter the effect other DNA repair proteins have on MUTYH; human APE1 almost completely inhibited the adenine glycosylase activity of mouse Y150C MUTYH on 8-oxoG:A mispairs without affecting the wild-type enzyme. This was suggested to be a result of competition for DNA binding between the two proteins (Pope, M.A. *et al*, 2005).

G393D is predicted to structurally disrupt a C-terminal turn that hydrogen bonds with the damaged DNA strand and is therefore thought to affect 8-oxoG recognition (Fromme, J.C. *et al*, 2004). An 85% reduction in glycosylase activity on 8-oxoG:A mispairs was observed with the equivalent *E. coli* mutant, G253D, *in vitro* and human MUTYH has a reduced ability to complement *E. coli mutY*⁻ compared to the wild-type protein (Chmiel, N.H. *et al*, 2003). Expression of G365D mutant mouse MUTYH in MUTYH-null ES cells could not suppress their mutator phenotype or produce detectable adenine glycosylase activity (Hirano, S. *et al*, 2003). In contrast, *E. coli*-expressed G365D mutant mouse MUTYH has been observed to excise adenine opposite 8-oxoG *in vitro* at levels similar to that of the wild-type protein (Hirano, S. *et al*, 2003; Tominaga, Y. *et al*, 2004; Pope, *M.A. et al*, 2005; Ushijima, Y. *et al*, 2005). This discrepancy between *in vitro* and *in vivo* activities is thought to be related to differences in

posttranslational modifications of MUTYH, in particular its phosphorylation which may be affected by the G365D mutation (Hirano, S. *et al*, 2003). The human G393D MUTYH protein displays a significantly reduced affinity for 8-oxoG:A mispairs and lacks adenine glycosylase activity *in vitro* (Wooden, S.H. *et al*, 2003). Cells derived from a G393D homozygous MAP patient have wild-type levels of MUTYH but the mutant enzyme exhibits significantly reduced binding and cleavage of 8-oxoG:A and 8-oxoA:G mispairs (Parker, A.R. *et al*, 2005). Mouse G365D MUTYH has a lower affinity for 8-oxoG:A pairs resulting in a reduced ability to prevent OGG1 from excising 8-oxoG opposite adenine or the generated AP site, leaving an uncorrected adenine to cause a G:C to T:A transversion or the loss of informative bases on both strands leading to a double strand break, respectively (Tominaga, Y. *et al*, 2004). Recombinant human G393D MUTYH protein shows only 1.5% of the wild-type glycosylase activity excising 2-hydroxyadenine opposite guanine which could also lead to G:C to T:A transversions (Ushijima, Y. *et al*, 2005).

Alhopuro *et al* recently analyzed the missense mutation A470D located in the MutT-like domain of MUTYH in human cell lines and found that this variant had significantly reduced adenine glycosylase activity on 8-oxoG:A mispairs compared to a wild-type cell line (Alhopuro, P. *et al*, 2005). Characterization of two splice site mutations through RNA analysis has been reported to date. The intron 12 variant c.1178-2 A>G identified in a patient in combination with G393D was confirmed to reside on the opposite allele to the common mutation and cDNA analysis using primers in exons 10 and 14 showed only mutant G393D transcript (Farrington, S.M. *et al*, 2005). The intron 10 splice site mutation c.924+3 A>C was detected in two patients in the present study and Kanter-Smoler *et al* recently reported that this variant causes skipping of exon 10 resulting in a frameshift and a stop codon at amino acid 267 (Kanter-Smoler, G. *et al*, 2006).

4.3.5 Other variants in MUTYH

The novel missense variant T474M (c.1401 C→T) is of uncertain significance as it was identified as a heterozygous change in a single patient in association with both G393D and 1092delC. It is a semi-conservative substitution at a

residue not conserved in prokaryotes. This variant was not detected in 724 control chromosomes and is predicted to have a possibly damaging effect on the protein (PolyPhen; http://genetics.bwh.harvard.edu/pph/). T474M is predicted to abolish one putative ESE responsive to the SR protein SF2/ASF and disrupt another so that it may promote exon inclusion more weakly (ESEfinder version 3.0, http://rulai.cshl.edu/tools/ESE/, Cartegni, L. *et al*, 2003). The previously reported rare missense variant L417M (c.1249 C→A, Peterlongo, P. *et al*, 2006) was detected in the heterozygous state in two cases, one was also a G393D heterozygote and the other patient carried both G393D and 1092delC pathogenic changes. This variant was not identified in 716 control chromosomes and is predicted to abolish a putative ESE responsive to the SR protein SC35 (ESEfinder version 3.0, http://rulai.cshl.edu/tools/ESE/, Cartegni, L. *et al*, 2003) but is a semi-conservative change at a site only partially conserved through evolution. The common *MUTYH* polymorphisms V22M (c.64 G→A; rs3219484), Q335H (c.1005 C→G; rs3219489) and S512F (c.1535 C→T; Al-Tassan, N. *et al*, 2003) were also detected, along with the previously reported silent variant S343S (c.1029 G→A) (Kairupan, C.F. *et al*, 2005). S343S is predicted to abrogate three putative ESEs responsive to the SR protein SF2/ASF but generates a novel enhancer responsive to the same protein within three nucleotides of these disrupted ESEs. Similarly, this variant also abolishes a putative ESE responsive to the SR protein SRp40 but results in a novel enhancer responsive to the same protein four nucleotides downstream (ESEfinder version 3.0, http://rulai.cshl.edu/tools/ESE/, Cartegni, L. *et al*, 2003).

4.3.6 Colorectal adenoma patients without biallelic MUTYH mutations

Biallelic *MUTYH* mutations were not identified in eighteen of the 26 unrelated index cases in whom I sequenced the entire *MUTYH* ORF. No potentially pathogenic *MUTYH* mutations were identified in eight patients and ten patients were heterozygous for either Y176C (4 cases) or G393D (6 cases) and did not appear to harbour a pathogenic mutation of the second allele (Table 4.4). The proportions of Y176C and G393D heterozygotes in unrelated index cases referred to Cardiff for genetic testing of *MUTYH* in whom biallelic mutations were not identified were not significantly different from those found in healthy UK non-Asians (Y176C heterozygotes; 34/5656 vs 4/227, p=0.057, G393D

Patient	Pathogenic mutation	Sex	Polyp count	CRC (age)	Family history	Previous *APC* screening
1598	Y176C	M	10s	2 (62)	None	PTT prom.& exons 3-part 15, MLPA
3196	Y176C	F	Polyposis	N	Unknown	Entire ORF sequenced[1]
1068	Y176C	M	Unknown	Unknown	Unknown	Unknown
3898	Y176C	F	Unknown	Unknown	Unknown	Unknown
4120	G393D	M	Multiple	N	Sister CRC + polyps, grandmother CRC	Unknown
3939	G393D	F	Polyposis aged 35	N	None	Entire ORF sequenced[1]
5161	G393D	M	Multiple aged 31	N	None	Entire ORF sequenced[1], MLPA
2088	G393D	F	>100	1 (58)	Unknown	Entire ORF sequenced[1], MLPA[1]
1997	G393D	M	None	1 (28)	Brother CRC + ~100 polyps. Consanguineous parents	Unknown
2125	G393D	M	Unknown	Unknown	Unknown	Unknown
4822	/	M	Multiple	1 (47)	Brother CRC aged 45. Consanguineous parents	Entire ORF sequenced[1], MLPA
3037*	/	F	22	N	1 of 4 sibs polyps no CRC	Unknown
3607	/	F	10-15	1 (35)	None. Consanguineous parents	Entire ORF sequenced[1]
3858	/	F	≥50	N	None	Entire ORF sequenced[1]
4147	/	M	~50 aged 40s	N	None	Entire ORF sequenced[1]
0652	/	M	30-40 aged 39	N	Unknown	Entire ORF sequenced
4234	/	M	<100	Unknown	Dominant family history	PTT prom.& exons 3-part 15, MLPA
1365	/	M	Unknown	Unknown	Unknown	APC –ve, no details

Table 4.4 Clinical features of patients referred to the All Wales Medical Genetics Service found to carry one or no pathogenic *MUTYH* mutation. Previous screening of the *APC* gene is shown in the last column. [1]Carried out by me as described in section 3.2.2. *This patient was also diagnosed with breast cancer aged 50.

heterozygotes; 75/5656 vs 6/227, p=0.092, Fisher's exact test) (Al-Tassan, N. *et al*, 2002; Sieber, O.M. *et al*, 2003; Fleischmann, C. *et al*, 2004; Farrington, S.M. *et al*, 2005; Tenesa, A. *et al*, 2006; Webb, E.L. *et al*, 2006). These cases may carry a *MUTYH* mutation not detected by ORF sequencing such as a large deletion/insertion or promoter mutation or hypermethylation. However, Nielsen *et al* (2005) carried out Southern blot analysis in 95 patients which did not identify any gross genomic rearrangements. Similarly, no additional *MUTYH* mutations have been found in patients following long-range PCR or fluorescent semi-quantitative multiplex PCR analysis (Eliason, K. *et al*, 2005; Peterlongo, P. *et al*, 2006). MLPA of exons 2 and 16 of the *MUTYH* gene also did not detect any large deletions or duplications (Kanter-Smoler, G. *et al*, 2006). AFAP or FAP could not be fully excluded in six of the 26 unrelated index cases in whom I sequenced the entire *MUTYH* ORF (Table 4.4).

4.3.7 MAP phenotype

All 237 MAP patients (182 index cases and 55 affected relatives) identified between the three centres were included in the study of MAP phenotype, regardless of ethnicity. Ninety-nine MAP cases were female (42%) and 58% (138 cases) were male. The mean age at presentation of 195 MAP patients who came to medical attention for reasons other than screening due to an affected relative was 45 years (median 45 years, range 12-68 years), consistent with previous reports (Wang, L. *et al*, 2004; Gismondi, V. *et al*, 2004; Kairupan, C.F. *et al*, 2005; Aceto, G. *et al*, 2005; Jo, W-S., *et al*, 2005; Bouguen, G. *et al*, 2006). Only a single patient presented symptomatically under the age of 20 years, with more than a hundred colorectal polyps but no CRC, and just 5% of cases presented before the age of 30 years. Three patients presented over the age of 65 years of whom two were diagnosed as a result of symptoms and one due to population screening. All three cases had more than a hundred colorectal polyps and two had CRC. These extremes were therefore very rare and the majority of patients (134 cases, 69%) presented symptomatically between the ages of 35 and 54 years.

4.3.7.1 Colorectal adenomas and CRC

Colorectal polyp counts were divided into categories based on number or terminology used. The colorectal phenotypes of MAP patients in the present study closely resembled AFAP (<100 adenomas) or classical FAP (100-1000 adenomas) (Table 4.5). Only 7% of MAP patients were described as having 'few' or less than ten colorectal polyps and unlike previous studies (Wang, L. *et al*, 2004; Croitoru, M.E. *et al*, 2004; Balaguer, F. *et al*, 2007), no cases with CRC in the absence of polyps were identified. This is likely to be because index cases in the present study were referred mainly from colorectal polyposis registers, clinical genetics or gastroenterology services whereas some previous studies screened population-based series of CRC cases (Croitoru, M.E. *et al*, 2004; Balaguer, F. *et al*, 2007) or early-onset CRC patients (Wang, L. *et al*, 2004). Very few MAP cases with CRC in the absence of polyps have been reported, suggesting that this phenotype is uncommon. More MAP cases had 10-99 than 100-1000 colorectal polyps (Table 4.5) which is consistent with the majority of previous reports (Wang, L. *et al*, 2004; Gismondi, V. *et al*, 2004;

Polyp count	No. of cases	Proportion of cases (%)
<10	14	6
10-99	119	50
100-1000	54	23
Few	3	1
Multiple	27	11
Numerous	6	3
Polyposis	11	5
Unknown	3	1

Table 4.5 Colorectal polyp count in 237 MAP patients. Cumulative number of colorectal polyps is shown grouped into categories based on number or terminology. Counts were derived from pathology and/or colonoscopy reports. A polyp count was not available for 3 patients.

Isidro, G. *et al*, 2004; Venesio, T. *et al*, 2004; Kairupan, C.F. *et al*, 2005; Jo, W-S. *et al*, 2005; Russell, A.M. *et al*, 2006; Lefevre, J.H. *et al*, 2006; Lejeune, S. *et al*, 2006; Croitoru, M.E. *et al*, 2007) and may reflect the mechanism of tumourigenesis. Only one somatic *APC* mutation is needed for adenoma initiation in FAP but both *APC* alleles have to be somatically inactivated in MAP. However, a significant proportion (23%) of MAP patients in the present study had 100-1000 colorectal polyps and this figure may be an underestimation as an additional 8% were described as having polyposis or numerous polyps. Two somatic *APC* mutations also appear to be required for tumourigenesis in AFAP but the BER defect in MAP is thought to increase the mutation rate in the *APC* gene making adenoma initiation more likely to occur. Two MAP cases with over a thousand colorectal polyps have been reported (Isidro, G. *et al*, 2004; Kanter-Smoler, G. *et al*, 2006) but none of the 237 patients in the present study had this number of polyps indicating the rarity of this phenotype in MAP.

The colorectal polyps in most MAP cases were adenomas but in 9% of patients a small number of hyperplastic polyps were identified in addition to adenomas. Colonic hyperplastic polyps have been previously reported in less than ten MAP patients (Sieber, O.M. *et al*, 2003; Ponti, G. *et al*, 2005; Croitoru, M. E. *et al*, 2007; Ponti, G. *et al*, 2007) and may be present in more cases but go unreported as adenomas have traditionally been seen as the polyps with malignant potential. A link between hyperplastic polyposis syndrome (HPS) and CRC through a serrated neoplasia pathway has been suggested more recently (Leggett, B.A. *et al*, 2001). Biallelic *MUTYH* mutations have been reported in a single HPS case with more than 40 adenomas, multiple hyperplastic polyps but no CRC (Chow, E. *et al*, 2006).

Fifty-eight percent (138/237) of MAP patients in the present study had been diagnosed with CRC. Previous studies of more than ten MAP patients reported CRC in 29% (Kairupan, C.F. *et al*, 2005), 44% (Wang, L. *et al*, 2004), 50% (Sieber, O.M. *et al*, 2003), 62% (Isidro, G. *et al*, 2004) and 75% (Gismondi, V. *et al*, 2004) of MAP cases. This variability may be due to differences in ascertainment, the age of MAP patients, the number of cases who had undergone prophylactic surgery or the small numbers of cases identified in

these studies (maximum 21 patients). CRC showed no preference for either sex (77/138 males, 61/99 females; p=0.370, chi squared test) and of the cases with CRC, 82% (113/138) had CRC at presentation whereas 25 patients developed CRC later. The mean age at CRC diagnosis was 48 years which is consistent with previous reports (Wang, L. *et al*, 2004; Gismondi, V. et al, 2004; Russell, A.M. *et al*, 2006; Bouguen, G *et al*, 2007; Croitoru, M.E. *et al*, 2007). A quarter of MAP patients with CRC were diagnosed at 45-49 years but the age at CRC diagnosis ranged from 21 to 70 years (Figure 4.8). The majority (80%) of MAP cases with CRC in the present study were diagnosed at an older age than classical FAP patients who usually develop CRC between the ages of 35 and 40 years (Galiatsatos, P. & Foulkes, W.D., 2006). HNPCC patients are diagnosed with CRC at a mean age of 45 years (Lynch, H.T. *et al*, 2006) and AFAP cases typically develop CRC at a mean age of 50-55 years (Galiatsatos, P. & Foulkes, W.D., 2006) although unlike FAP, not all patients affected by either disease develop CRC within their lifetimes (Strate, L.L. & Syngal, S., 2005). Classical FAP patients develop CRC earlier as adenoma initiation is quicker so many adenomas arise at an earlier age resulting in a high probability of progression to carcinoma, although the rate of progression is not accelerated in FAP. HNPCC has accelerated tumour progression but adenoma initiation occurs infrequently as this requires somatic mutation of a wild-type MMR allele followed by biallelic inactivation of a tumour suppressor. Adenoma initiation in MAP and AFAP is not as frequent or early as in classical FAP but it is possible that the BER defect in MAP may result in an increased risk of progression leading to an earlier onset of CRC than in typical AFAP.

Of the MAP patients diagnosed with CRC in the present study, 36% (49/138 cases) had at least one synchronous or metachronous CRC. Two, three or four synchronous CRCs were identified in 26, 3 and 4 patients, respectively. Sixteen cases had metachronous CRC, of whom three had two metachronous CRC. Several previous reports have also described MAP cases with more than one CRC (Lipton, L. *et al*, 2003; Gismondi, V. *et al*, 2004; Venesio, T. *et al*, 2004; Jo, W-S. *et al*, 2005; Lefevre, J.H. *et al*, 2006; Croitoru, M.E. *et al*, 2007). HNPCC patients are also at increased risk of synchronous and metachronous CRCs (Lynch, H.T. *et al*, 2006) and Vasen *et al* (2005) recently reported 18% of

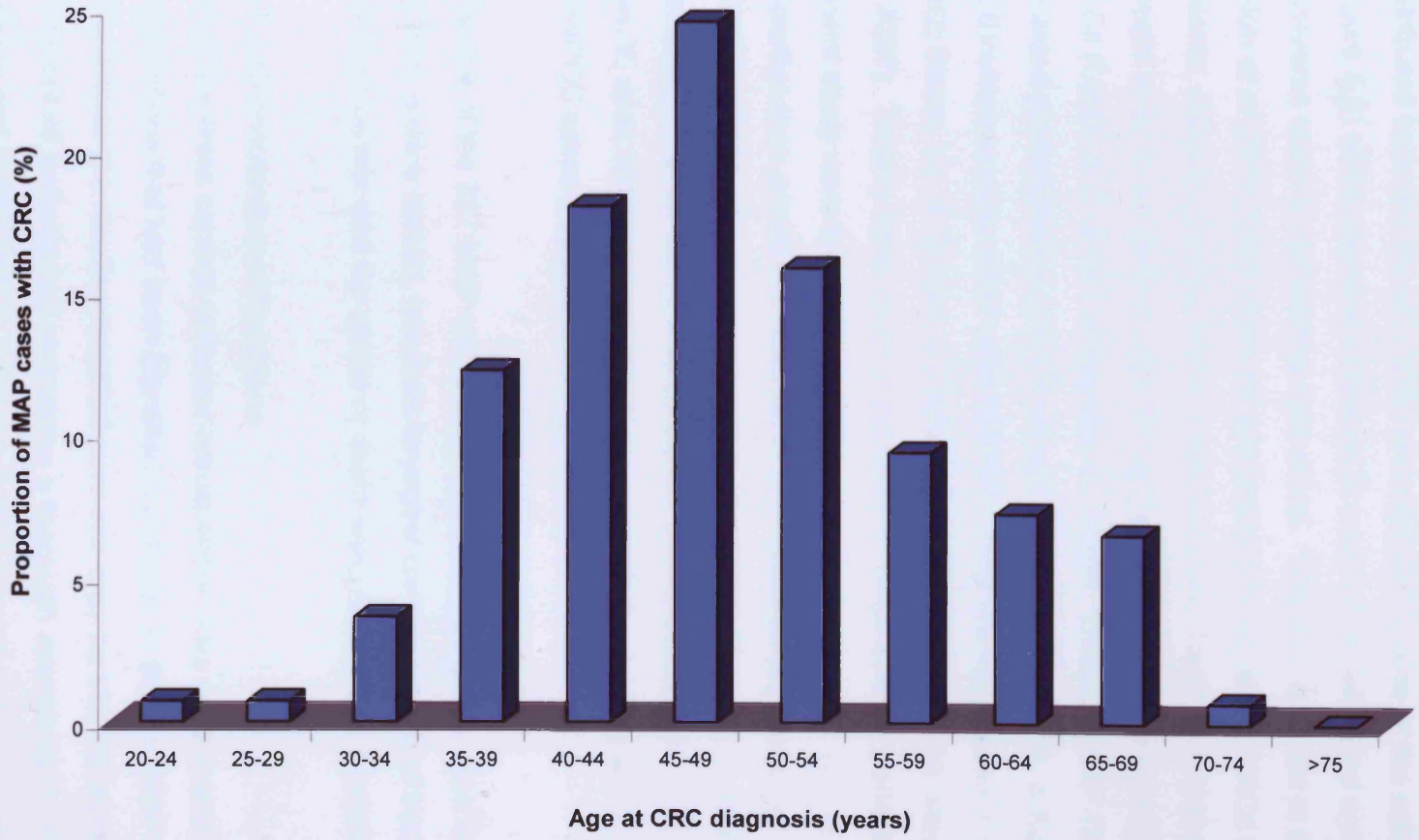


Figure 4.8 Age at CRC diagnosis of 138 MAP patients with CRC. A quarter of MAP cases with CRC were diagnosed between the ages of 45 and 49 years. Only 5% of cases were diagnosed before the age of 35 years, the age at diagnosis ranged from 21 to 70 years.

cases with CRC had more than one CRC.

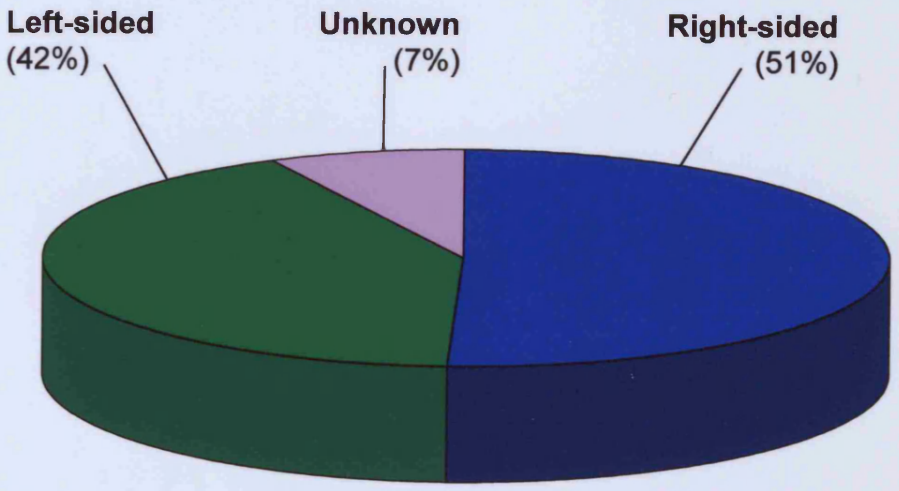
The 201 CRCs identified in 138 MAP patients in the present study were distributed between the left (42%) and right (51%) side of the colorectum (Figure 4.9) where the splenic flexure is counted as left-sided and the transverse colon is considered right-sided. This is in contrast to the report of Lipton *et al* (2003) who observed predominantly left-sided CRCs in MAP patients, although only 24 CRCs were analyzed. Distal (left-sided) CRCs are thought to be more likely to behave aggressively than proximal (right-sided) CRCs (Lynch, H.T. & de la Chapelle, A., 2003). Classical FAP CRCs show a left-sided predominance (Cao, Y. *et al*, 2002; Galiatsatos, P. & Foulkes, W.D., 2006) whereas more AFAP and HNPCC CRCs are right-sided (Cao, Y. *et al*, 2002; Strate, L.L. & Syngal, S., 2005; Lynch, H.T. *et al*, 2006; Mecklin, J-P. *et al*, 2007). Twenty-three percent (47/201) of CRCs in MAP patients in the present study were located in the rectum or rectosigmoid. This is a lower proportion than previously reported in MAP patients (Lipton, L. *et al*, 2003; Jo, W-S. *et al*, 2005; Kanter-Smoler, G. *et al*, 2006; Lefevre, J.H. *et al*, 2006; Croitoru, M.E. *et al*, 2007). Rectal cancer is also observed in FAP patients (Cao, Y. *et al*, 2002) but is rarely seen in AFAP (Lyster Knudsen, A. *et al*, 2003) or HNPCC cases (Mecklin, J-P. *et al*, 2007).

Forty-one of the 237 MAP cases were deceased; 26 (63%) died from CRC, two died from gastric cancer, one from laryngeal cancer, four died from causes other than cancer and the cause of death was unknown in 8 cases.

4.3.7.2 Extracolonic manifestations

Each of the three centres collected retrospective data on extracolonic manifestations that had been identified previously in the 237 MAP patients (Table 4.6) so the results presented here can only be viewed as preliminary data as not all patients had undergone a thorough assessment. As in FAP, duodenal and gastric polyps were the most frequently occurring extracolonic manifestations (Table 4.6) but had been previously reported in less than ten MAP patients (Sieber, O.M *et al*, 2003; Jo, W-S. *et al*, 2005; Kanter-Smoler, G. *et al*, 2006; Lejeune, S. *et al*, 2006; Bouguen, G. *et al*, 2007; Olschwang, S. *et*

A)



B)

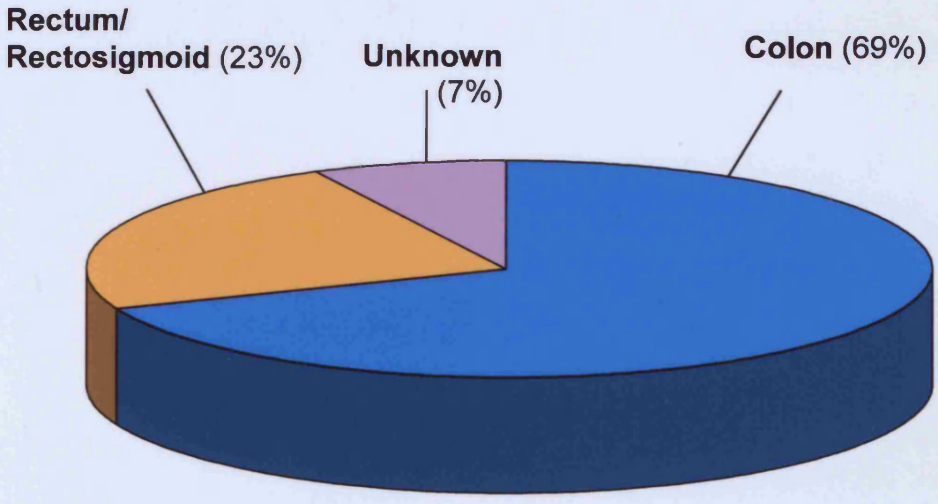


Figure 4.9 Location of 201 CRCs identified in 138 MAP cases. A) Proportion of left- and right-sided CRCs, B) Proportion of CRCs located in the colon or the rectum/rectosigmoid. CRC location data were not available for 15 of the 201 CRCs.

Extracolonic manifestation	No. of patients	Proportion of patients (%)	Mean age at diagnosis (yrs)	M:F ratio
GI tract				
Appendix carcinoid tumour	2	0.8	46	0:2
Duodenal polyps	25	11	47	10:15
Jejunal polyps	1	0.4	44	0:1
Duodenal cancer	2	0.8	61	2:0
Ileum cancer	2	0.8	53	2:0
Small intestine carcinoid tumour	2	0.8	70	1:1
All gastric polyps	15	6	50	9:6
Gastric adenomas	3	1	38	2:1
Fundic gland polyps	10	4	52	4:6
Gastric polyps with unknown histology	5	2	47	4:1
Gastric cancer	3	1	34	2:1
Skin lesions				
Basal cell carcinoma (M)	7	3	55	4:3
Melanoma (M)	2	0.8	31	1:1
Squamous cell carcinoma (M)	2	0.8	64	1:1
Sebaceous gland epithelioma (M)	1	0.4	51	1:0
Sebaceous gland hyperplasia (B)	4	2	41	3:1
Sebaceous gland adenomas (B)	4	2	40	3:1
Steatocystoma (B)	1	0.4	Unknown	1:0
Actinic keratoses (pre-malignant)	2	0.8	46	2:0
Epidermoid cyst (B)	5	2	50	2:3
Follicle cyst formation with branches (B)	1	0.4	38	1:0
Seborrhoeic wart (B)	1	0.4	53	1:0
Lipoma (B)	7	3	41	2:5
Pilar cyst (B)	1	0.4	50	1:0
Dermatofibroma (B)	1	0.4	54	0:1
Naevus naevocellularis pigmentosus (B)	2	0.8	36	1:1
Pigmented spots (B)	1	0.4	Unknown	1:0
Verruca vulgaris (B)	1	0.4	60	1:0
Skin lesions of unknown histology	8	3	40	6:2
Other cancers				
Breast cancer	9	4	57	1:8
Bladder cancer	4	2	58	3:1
Laryngeal cancer	1	0.4	57	1:0
Prostate cancer	1	0.4	62	1:0
Teratoma	2	0.8	31	1:1
Testicular cancer	1	0.4	41	1:0
Thyroid cancer	1	0.4	38	0:1
Cervical cancer	1	0.4	39	0:1
Gangliocytoma	1	0.4	52	0:1
Lymphoma	1	0.4	46	1:0
Ovarian cancer	1	0.4	56	0:1
Endometrial cancer	1	0.4	54	0:1
Oesophageal cancer	1	0.4	59	2:0
Soft palate cancer	1	0.4	46	1:0

Table 4.6 Retrospective identification of extracolonic manifestations in 237 MAP cases. The most frequent extracolonic manifestations are duodenal and gastric polyps. Several cancers other than CRC have been diagnosed in MAP patients. Additional benign lesions assumed to be of no clinical importance and patient identifiers are shown in appendix Table III. B; benign, M; malignant.

al, 2007). The small number of MAP patients reported to have upper GI polyps may reflect a lack of routine upper GI screening or the identification of relatively few MAP cases in previous studies. A single patient has been described as having jejunal polyps (Lejeune, S. *et al*, 2006) and one further case was identified in the present study. Small intestine carcinoma is rare in the general population, accounting for 0.4% of new cancer cases in the U.S in 2005 (Jemal, A. *et al*, 2005) but four cases (two duodenal and two ileal) were identified in the present study (Table 4.6) and two have been previously reported (Kumar, V.K.A. *et al*, 2007; Olschwang, S. *et al*, 2007). FAP and HNPCC are associated with small bowel cancer (Galiatsatos, P. & Foulkes, W.D., 2006; Vasen, H.F.A., 2005). Gastric carcinoma has been reported in a single case with biallelic *MUTYH* mutations (Olschwang, S. *et al*, 2007) and three patients in the present study were diagnosed at 48, 38 and 17 years (Table 4.6). However, the youngest case also harboured a germline mutation in *E-cadherin* and mutations in this gene have been identified in familial gastric cancer (Guilford, P. *et al*, 1998). Two MAP patients in the present study had a small intestine carcinoid tumour and two had a carcinoid tumour of the appendix (Table 4.6). Carcinoid tumours have not been described before in MAP cases and are rare in the general population but are most commonly found in these locations (Cancer Research UK). Larger studies are required to investigate the possibility of an aetiological association.

Epidermoid cysts, lipomas and fibromas are skin lesions associated with the FAP variant Gardner's syndrome (Galiatsatos, P. & Foulkes, W.D., 2006) but they are also common in the general population (Burns, T. *et al*, 2004) and were identified in 5, 7 and 1 MAP patients in the present study, respectively (Table 4.6). Dermoid and epidermoid cysts have been previously described in one and two MAP cases, respectively (Wang, L *et al*, 2004; Kairupan, C.F. *et al*, 2005; Gismondi, V. *et al*, 2004). Patients affected by the HNPCC variant Muir Torre syndrome have sebaceous adenomas or carcinomas (Lynch, H.T. *et al*, 2006). Four MAP cases with sebaceous adenomas and a single case with sebaceous epithelioma were identified in the present study (Table 4.6). Two MAP patients have been previously reported to have sebaceous adenomas (Ponti, G. *et al*, 2005; Kumar, V.K.A. *et al*, 2007) and a single case with sebaceous carcinoma

(Barnetson, R.A. *et al*, 2007). Both lesions are rare, with sebaceous carcinomas accounting for less than 1% of all skin malignancies (Burns, T. *et al*, 2004). Sebaceous gland hyperplasia was identified in four MAP cases in the present study (Table 4.6) but occurs in 1% of healthy individuals (Ponti G. *et al*, 2007) and is generally considered of no clinical importance (Burns, T. *et al*, 2004). However, Ponti *et al* (2007) recently detected *BRAF* mutations in sebaceous hyperplasias from MAP cases and mutations in this gene are linked to tumourigenesis in some skin tumours although the mutation detected was a T>A not G>T change. A single case in the present study had been diagnosed with steatocystoma which also affects the sebaceous glands. Further studies of sebaceous gland tumourigenesis are warranted in MAP. Malignant melanoma and squamous cell carcinoma were each identified in two MAP cases and basal cell carcinoma (BCC) was found in seven MAP patients in the present study (Table 4.6). Melanoma has been previously reported in a single MAP case (Enholm, S. *et al*, 2003) as has BCC (Olschwang, S. *et al*, 2007) but these and squamous cell carcinoma are common cancers and there is insufficient information at present to confirm an association with MAP. The only other skin lesions reported in MAP cases are pilomatricomas which were identified in a single kindred (Baglioni, S. *et al*, 2005). Two cases in the present study had pre-malignant actinic keratoses, seven patients had a range of benign skin lesions and eight cases had skin lesions of unknown histology (Table 4.6).

A germline defect in the critical BER repair pathway might be expected to give rise to tumours in many organs, as is seen in cases with defective MMR in HNPCC. Most extracolonic cancers were identified in only one or two MAP patients in the present study (Table 4.6) and overall 49 malignant lesions (other than CRC) were reported in 237 cases (21%) suggesting that these malignancies are not commonly observed in MAP and may just reflect the 1 in 3 population risk of cancer (Cancer Research UK). However, not all MAP patients in the present study have lived through their lifetime cancer risk and some may have died from CRC before other cancers had the time to develop. The present study of extracolonic manifestations in MAP patients is the largest to date as only case reports of extraintestinal cancers have been described

previously which cannot inform the clinical management of MAP. Single cases of cervical cancer (Alhopuro, P. *et al*, 2005), uterine cancer (Olschwang, S. *et al*, 2007), endometrial cancer (Barnetson, R.A. *et al*, 2007), thyroid cancer (Ponti G. *et al*, 2005), leukaemia (Croitoru, M.E. *et al*, 2007), sarcoma (Balaguer, F. *et al*, 2007) and carcinoma of the ampulla of Vater (Croitoru, M.E. *et al*, 2007) have been described in patients with biallelic *MUTYH* mutations. A central nervous system carcinoma was described by Olschwang *et al* (2007) and a schwannoma was reported in a single MAP case (Croitoru, M.E. *et al*, 2007). Bladder cancer has not been previously reported in MAP cases but three urothelial and one squamous cell bladder carcinoma were identified in four patients in the present study (Table 4.6). Two of the cases with urothelial cancer were siblings so additional genetic factors may have contributed to their development. Breast cancer was the most frequent extracolonic cancer in the present study, affecting 8% (8/99) of female MAP patients at a mean age of 57 years and a single male case aged 56 years. This cancer has also been previously reported in four MAP cases (Isidro, G. *et al*, 2004; Jo, W-S. *et al*, 2005; Balaguer, F. *et al*, 2007; Olschwang, S. *et al*, 2007). One in nine women (11%) will develop breast cancer during their lifetime and eight out of ten cases are diagnosed aged 50 years or over (Cancer Research UK). The earliest diagnosis of breast cancer in the present study was 45 years and 75% (6/8) of cases were diagnosed aged 50 years or over so there is no conclusive evidence that MAP patients are at increased risk of developing early breast cancer as is the case in Muir Torre syndrome (Strate, L.L. & Syngal, S., 2005). Unlike HNPCC and FAP, current data does not suggest that MAP confers an increased risk of cancer outside the GI tract.

The phenotypic similarities between MAP and FAP extend beyond the colorectum and upper GI tract in a few cases and may reflect the common inactivation of *APC* during tumourigenesis. A benign bone tumour was described in a single patient in the present study and osteomas have been reported in two other MAP patients (Gismondi, V. *et al*, 2004). Four cases in the present study had jaw-bone cysts at a mean age of 17 years and dental or jaw abnormalities have been described in three other MAP patients (Gismondi, V. *et al*, 2004; Lejeune, S. *et al*, 2006; Kairupan, C.F. *et al*, 2006). CHRPE was

not identified in the present study but has been previously reported in some MAP cases (Sieber, O.M. *et al*, 2003; Gismondi, V. *et al*, 2004; Jo, W-S. *et al*, 2005; Lefevre, J.H. *et al*, 2006).

4.3.8 Genotype-phenotype relationship

The most frequently mutated alleles in our 237 MAP cases were Y176C and G383D. To investigate if these mutations differed in their phenotypic effect, the colorectal phenotypes of Y176C homozygotes (55 cases), G393D homozygotes (20 cases) and Y176C/G393D compound heterozygotes (59 cases) were compared (Table 4.7). The mean age at presentation was inversely correlated with the number of Y176C alleles (p=0.003, linear regression) and although the proportion of patients with CRC was not significantly different between the three genotypes (p=0.087, chi squared test 2 degrees of freedom), the mean age at CRC diagnosis was also inversely correlated with the number of Y176C alleles (p<0.001, linear regression). Farrington *et al* (2005) reported CRC in all G393D homozygotes, but only by the age of 65 years and biallelic Y176C mutations have been previously suggested to have a greater effect on CRC risk than biallelic G393D mutations, although the difference was not statistically significant (Tenesa, A. *et al*, 2006). In the present study, the proportion of patients with less than ten polyps significantly increased as the number of Y176C alleles decreased (p=0.016, chi-squared Test for Trend in proportions, 1 degree of freedom), but there was little difference in the proportion of cases with 10-99 or 100-1000 colorectal polyps between the three genotypes and numeric counts were not available for all cases. CRC was distributed similarly between the left- and right-side of the colorectum in the three genotype groups and although no rectal cancer was seen in G393D homozygotes, only ten CRCs had been diagnosed in these cases and the same proportion of rectal cancer was seen in Y176C homozygotes as compound heterozygotes. Balaguer *et al* (2007) reported Y176C carriers (homozygotes and heterozygotes) were diagnosed with CRC at a younger age and had more adenomas and right-sided CRCs than CRC patients who did not carry this mutant allele. The differences observed between the three genotypes in the present study could be a result of the different effects these mutations have on the MUTYH protein. As described above, Y176C is thought to have a more detrimental effect on MUTYH

Genotype/ no. truncating alleles	No. MAP patients	M:F	Mean age at presentation (yrs) ± SD	Polyp count			Proportion of cases with CRC	Age at CRC diagnosis (yrs) mean ± SD (range)	Location CRC			
				<10	10-99	100-1000			Colon	Rectum	Left-sided	Right-sided
Y176C/Y176C	55	33:22	43 ± 8**	1/55* (2%)	27/55 (49%)	12/55 (22%)	38/55 (69%)	47 ± 6** (30-65)	43/60 (72%)	17/60 (28%)	25/60 (42%)	35/60 (58%)
Y176C/G393D	59	37:22	48 ± 11**	7/59* (12%)	29/59 (49%)	13/59 (22%)	31/59 (53%)	51 ± 10** (30-67)	31/43 (72%)	12/43 (28%)	21/43 (49%)	22/43 (51%)
G393D/G393D	20	12:8	50 ± 12**	4/20* (20%)	10/20 (50%)	3/20 (15%)	9/20 (45%)	58 ± 9** (37-70)	10/10 (100%)	0	3/10 (30%)	7/10 (70%)
0	172	102:70	45 ± 10	14/172 (8%)	88/172 (51%)	35/172 (20%)	102/172 (59%)	49 ± 9 (29-70)	106/142 (75%)	36/142 (25%)	64/142 (45%)	78/142 (55%)
1	44	22:22	43 ± 10	0	22/44 (50%)	9/44 (20%)	25/44 (57%)	46 ± 9 (21-68)	25/31 (81%)	6/31 (19%)	14/31 (45%)	17/31 (55%)
2	21	14:7	46 ± 9	0	9/21 (43%)	10/21* (48%)	11/21 (52%)	50 ± 11 (36-66)	8/13 (62%)	5/13 (38%)	6/13 (46%)	7/13 (54%)

Table 4.7 Colorectal phenotype of MAP cases according to genotype. Cases homozygous or compound heterozygous for the most frequently mutated alleles Y176C and G393D were analyzed for phenotypic differences (top of table). All 237 cases were divided into groups determined by the number of truncating alleles they carried; 0, 1 or 2 and were analyzed separately (bottom of table). SD, standard deviation. *$p<0.05$, chi squared test; **$p<0.005$, linear regression.

glycosylase activity than G393D and whereas mutant G393D MUTYH protein is found at wild-type levels in cells derived from a G393D homozygous patient, cells derived from a Y176C homozygous patient have lower levels of mutant MUTYH protein (Parker, A.R. *et al*, 2005). This more severe defect in BER resulting from Y176C mutations could explain the inverse correlation between age at onset of CRC and number of Y176C alleles as the mutation rate in cells would be expected to increase with number of Y176C alleles thus increasing the chance of mutations needed for tumourigenesis.

A milder disease in G393D homozygotes may explain why this genotype was identified much less frequently than homozygosity for Y176C in index cases from each of the three centres in the present study, despite the G393D mutant allele being approximately twice as frequent as Y176C in healthy individuals (Croitoru, M.E. *et al*, 2004; Farrington, S.M. *et al*, 2005; Peterlongo, P. *et al*, 2006; Tenesa, A. *et al*, 2006; Webb, E.L. *et al*, 2006). The vast majority of index cases in the present study were referred from colorectal polyposis registers and it could be that because G393D homozygotes appear to present later and may have fewer colorectal polyps, they are not being referred to such registers.

The colorectal phenotype of MAP patients was also analyzed according to how many of their biallelic germline *MUTYH* mutations were predicted to truncate the MUTYH protein; none, one or two. Truncating mutations included nonsense, frameshift and splice site mutations whereas non-truncating mutations comprised missense, in-frame insertion and in-frame deletion mutations. No significant differences were observed in mean age at presentation (p=0.485, ANOVA) or mean age at CRC diagnosis (p=0.547, ANOVA) between MAP case with none, one or two truncating alleles. Only patients with no truncating alleles were found to harbour less than ten colorectal polyps whereas similar proportions of MAP cases with 10-99 polyps were identified in all three groups. Twenty percent of patients carrying either one or no truncating alleles had 100-1000 colorectal polyps but a significantly higher proportion of patients with two truncating alleles (10/21 vs 44/216, p=0.008 Fisher's exact test) showed this more severe phenotype. However, 26% of MAP cases did not have a

numerical polyp count and the number of cases with no truncating alleles was eight times the number of patients with two truncating alleles. A similar proportion of MAP cases developed CRC in the three groups and the proportion of rectal cancer was not significantly different in cases with one, two or no truncating alleles (p=0.412, chi squared test 2 degrees of freedom). From the present study it appears that the number of truncating *MUTYH* alleles a patient carries cannot be correlated with disease severity.

MAP patients with identical *MUTYH* mutations showed phenotypic variability in the present study; one Y176C/G393D compound heterozygote had CRC and 100-1000 polyps aged 38 years yet another patient with the same genotype developed less than ten polyps without cancer aged 52 years. Intrafamilial variation was also seen in the present study and has been previously reported (Raoof, M. *et al*, 2007). This suggests that as for other colorectal cancer syndromes, additional genetic or environmental factors are modifying the MAP phenotype (Crabtree, M.D. *et al*, 2002).

4.3.9 Management of MAP patients

The present study confirmed that Y176C and G393D are the most frequently mutated *MUTYH* alleles in three European countries. Molecular genetic services specifically screen for these mutations but currently not all offer comprehensive screening for mutations. Seventeen percent of MAP cases identified in the present study did not carry either of these more common mutations and would not have been detected unless a full gene screen was undertaken. Similarly, 29% of MAP patients only carried Y176C or G393D in combination with a less frequent mutation so would have been classed as heterozygotes, with implications for surveillance of these patients. The very wide spectrum of mutations already identified in cases with MAP suggests that sequencing of the entire *MUTYH* ORF is justified. This will identify mutations of uncertain significance and functional analysis will be important in discriminating between pathogenic and non-pathogenic changes.

As for FAP, endoscopic surveillance in affected or at-risk individuals could begin in their second decade as only a single MAP case presented symptomatically

before the age of 20 years in the present study but should continue until late in life as a patient in the present study was diagnosed with CRC aged 70 years. Most CRC was diagnosed at presentation so earlier screening and prophylactic surgery should reduce the CRC mortality in MAP patients. Upper GI screening could be worthwhile in MAP cases as several of the patients who had undergone such screening in the present study were found to harbour either duodenal or gastric polyps and a small number of cases were diagnosed with carcinoma of the small intestine or stomach. Also of note were the four cases of intestinal carcinoid tumours but larger numbers of cases are needed before an association with MAP can be confirmed. MAP patients may also be at increased risk of sebaceous adenomas but a thorough assessment of skin lesions in more cases is needed to assess this. Although prospective studies are indicated, this retrospective study provides reassurance that systematic screening outside the GI tract is unlikely to be needed in MAP.

Chapter Five

Mortality and cancer risk in monoallelic *MUTYH* mutation carriers

5.1 Introduction

Since the identification of MAP by Al-Tassan *et al* in 2002, several studies have undertaken germline *MUTYH* mutation analysis in unrelated colorectal polyposis patients. Many have reported monoallelic *MUTYH* mutations in a small number of cases without pathogenic *APC* mutations (Sieber, O.M. *et al*, 2003; Aceto, G. *et al*, 2005; Nielsen, M. *et al*, 2005; Aretz, S. *et al*, 2006; Russell, A.M. *et al*, 2006; chapter four). The clinical significance of carrying a single *MUTYH* mutation is unclear at present as each of the two most common *MUTYH* mutations (Y176C and G393D) are also found as heterozygous changes in healthy individuals at a combined minor allele frequency of approximately 1% in non-Asians (Croitoru, M.E. *et al*, 2004; Farrington, S.M. *et al*, 2005; Peterlongo, P. *et al*, 2006; Tenesa, A. *et al*, 2006; Webb, E.L. *et al*, 2006).

Case-control studies in cohorts of CRC patients have described an over-representation of heterozygous carriers in CRC cases (Croitoru, M.E. *et al*, 2004; Fleischmann, C. *et al*, 2004; Farrington, S.M. *et al*, 2005; Zhou, X-L. *et al*, 2005; Colebatch, A. *et al*, 2006; Peterlongo, P. *et al*, 2006; Kury, S. *et al*, 2007) but independent statistical significance was reached in just two studies, and then only by considering G393D status in CRC cases and controls over the age of 55 years (Farrington, S.M. *et al*, 2005) or CRC cases from MMR gene mutation-negative HNPCC-like families (Peterlongo, P. *et al*, 2006). Results from case-control studies investigating *MUTYH* heterozygote CRC risk may not reach statistical significance because of the large samples size required. Webb *et al* (2006) estimated that 22000 cases and 22000 controls would be required to reliably detect a relative risk of 1.2 due to a low penetrance allele with a population frequency of less than 2%. In this study I used a kin-cohort approach to assess cancer risk and mortality in obligate *MUTYH* heterozygotes, the majority of whom had lived through their lifetime CRC risk. Clarification of whether or not monoallelic *MUTYH* mutation carriers have an increased risk of

CRC is important for the genetic counselling and clinical management of these individuals and their families.

5.2 Methods

As described in section 4.2.1, a collaborative project was established between The Institute of Medical Genetics (IMG), Cardiff, The Institute of Human Genetics, Bonn and The Centre of Human and Clinical Genetics, Leiden which resulted in a cohort of 182 unrelated MAP index cases. The present study aimed to gather all-cause mortality and cancer incidence and mortality data on the obligate *MUTYH* heterozygote parents of these MAP index cases. All of the index cases came to medical attention through symptomatic presentation of polyposis or CRC or (in 16 cases) through population screening programmes, and none because of their family history.

5.2.1 Study approval

As shown in Figure 5.1, several stages of application were necessary for approval of the 'Investigating cancer risk in people with *MUTYH* gene changes' study in the UK. An honorary NHS contract was obtained for the duration of the study as handling identifiable patient data was involved. Cardiff University agreed to act as sponsor (ref. SPON CU 237) and the MREC for Wales gave ethical approval (ref. 06/MRE09/19). As the study also involved other centres in the UK (Birmingham, Manchester and Southampton), once MREC approval had been received a site-specific assessment application was made to each of the LRECs with the help of the principal investigator (PI) at each centre (Prof. E. Maher, Prof. D.G.R. Evans, Prof. D. Eccles). This was followed by applications for study approval from the Research and Development (R&D) departments of NHS Trusts covering the centres involved with the help of PIs.

Protocols for obtaining informed consent in the Netherlands were approved by local ethics review boards and the ethics committee of the University of Bonn approved the German part of the study.

5.2.2 Clinical information and samples

Biallelic *MUTYH* mutations had been identified in 54 apparently unrelated index cases referred from Regional Clinical Genetics Services, Regional Polyposis Registers and Gastroenterology departments to AWMGS for *MUTYH* germline mutation analysis. These results had been reported back to clinical geneticists

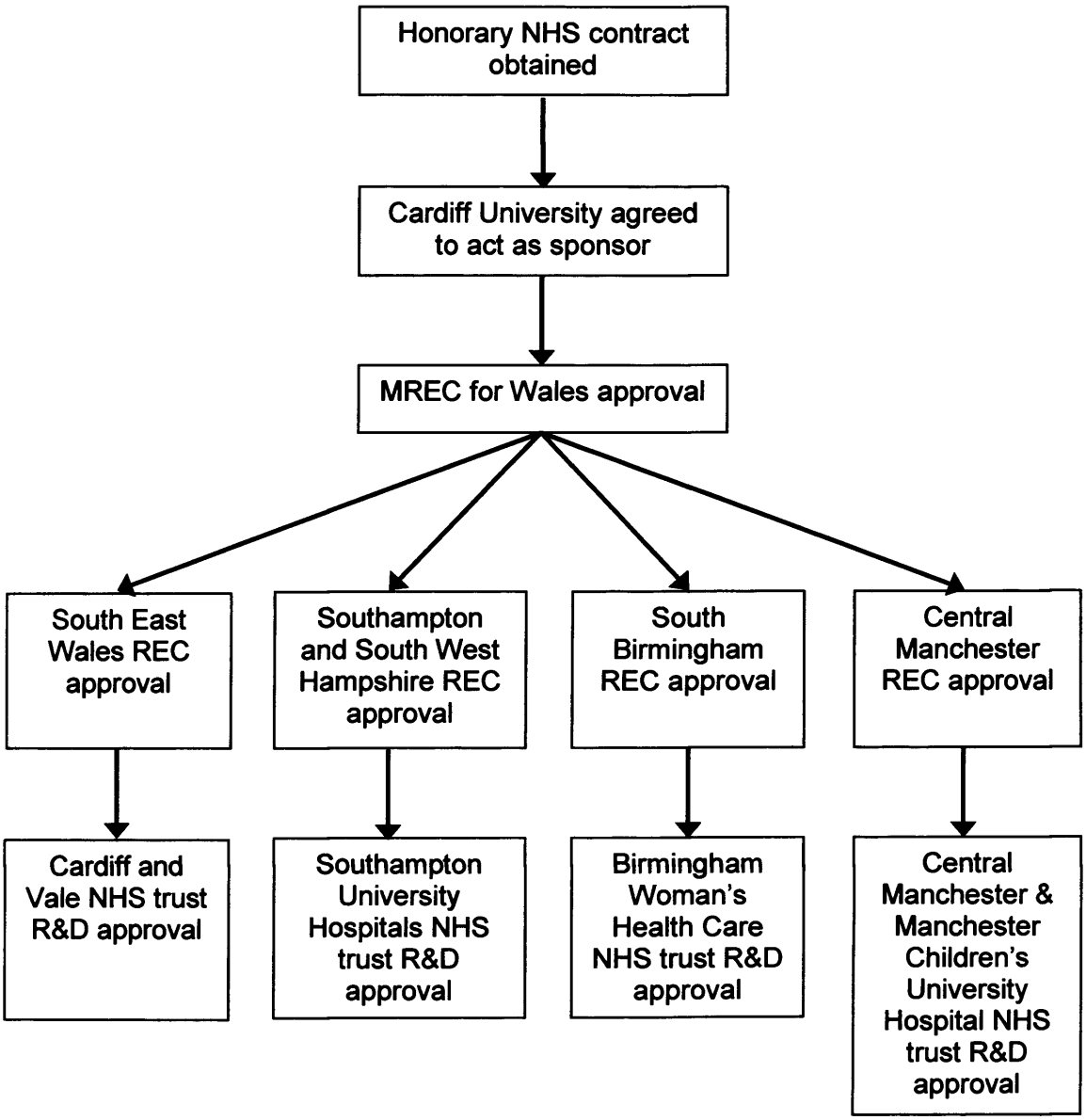


Figure 5.1 Approval of the 'Investigating cancer risk in people with *MUTYH* gene changes' study in the UK. Ethics committee and NHS Trust R&D approval was obtained for each research site with the aid of the PI at each centre.

at the Regional Genetics Services (RGSs) and discussed with the patients concerned. As MAP index cases usually present in the age range 40-70 years it was anticipated that the majority of their obligate *MUTYH* mutation heterozygote parents would be deceased. Information about these monoallelic carriers was obtained through RGSs, MAP index cases, or the siblings, spouses or adult offspring of the MAP index case as shown in Figure 5.2 (forms shown in appendix). Medical and family history details for 7 MAP index cases had already been obtained for a previous study (Sampson, J.R. *et al*, 2003) and no further details were sought. Informed consent for the release of medical and family history details from 41 patients or consenting relatives was obtained through recruitment into the 'Investigating cancer risk in people with *MUTYH* gene changes' study or the 'Genes and Multiple Colorectal Adenomas- The Polyp Study' (MREC for Wales ref. 02/9/22). Causes of death were confirmed with death certificates and cancer diagnoses were verified with Regional Cancer Registries when possible. If a parent had been diagnosed with CRC, pathology details along with the date and hospital where any surgery for this cancer took place were requested from the Regional Genetics Service. Again, this information was not frequently known to the Regional Genetics Service and was therefore requested from the MAP index case or consenting relative via a questionnaire. These details were then obtained from the relevant hospital records. Consent to obtain tissue blocks already stored was obtained from the MAP patient or consenting relative then a request was sent to the pathology department at the appropriate hospital for their release (Figure 5.2).

The Institute of Human Genetics, Bonn and The Centre of Human and Clinical Genetics, Leiden identified 73 and 55 unrelated MAP index cases, respectively. These centres obtained clinical information about the monoallelic mutation carrier parents of 71 and 55 MAP index cases, respectively, from relatives and/or hospital records. Causes of death were confirmed with death certificates and cancer diagnoses were verified with hospital records when possible. If a parent had been affected by CRC and comprehensive *MUTYH* mutation analysis had not been previously undertaken, a request was sent to the pathology department at the appropriate hospital to obtain tissue blocks from their CRC surgery. Division of work is shown in Figure I in the appendix.

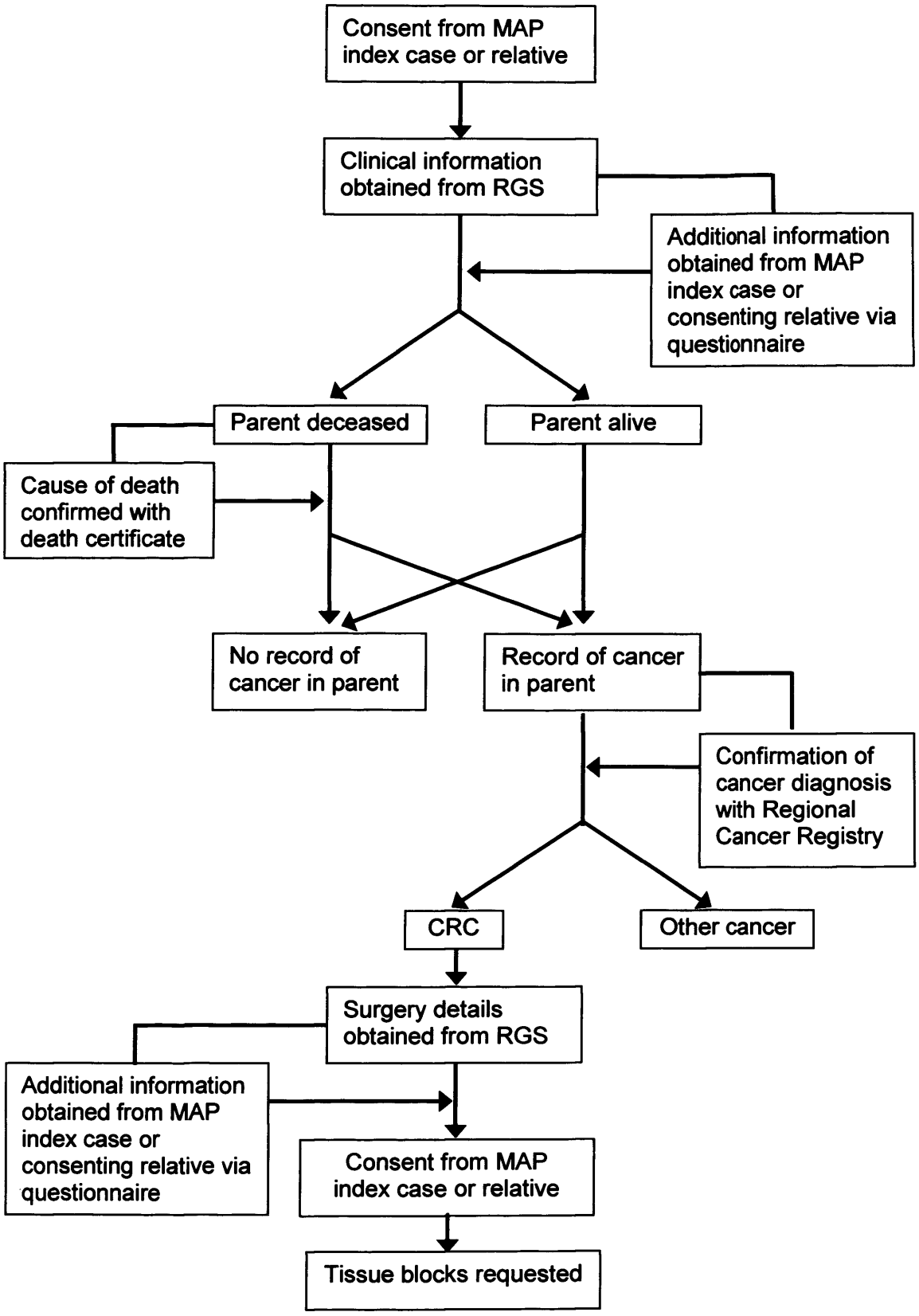


Figure 5.2 Protocol for obtaining clinical information about obligate *MUTYH* mutation heterozygote parents and tissue blocks from CRC surgery. RGS, Regional Genetics Service.

5.2.3 Screening for mutations in MUTYH

To exclude the possibility that any of the parents affected by CRC carried biallelic *MUTYH* mutations, I extracted DNA from cores of paraffin-embedded non-cancerous tissue taken at surgery for CRC as described in section 2.6.2.1. I PCR amplified exons 1-16 of *MUTYH* as 21 fragments (Table 5.1) and screened for mutations using automated sequencing (Figure I).

5.2.4 Population data and statistical analysis

Cancer incidence and mortality and overall mortality in monoallelic *MUTYH* mutation carriers were compared with nation, sex, age and date-specific population rates (5 year intervals). Table 5.2 shows the periods of time for which these population data were available for the different countries. The data for England and Wales were obtained by Dr. P. Wark from The Office of National Statistics, Dutch data were requested by Dr. F.J. Hes from the Eindhoven Cancer Registry and Dr. S. Vogt obtained German data from the Saarland Cancer Registry (Figure I). The ICD codes used to select appropriate data for CRC and all cancer incidence and mortality rates in the different countries are shown in Table 5.3.

Exact standardized mortality ratios (SMRs), exact standardized incidence ratios (SIRs) and their exact confidence intervals (CIs) were calculated using Stata version 9 (Stata Corporation, USA) and the smri function programmed by P. Sasieni. If dates of birth (DOBs), dates of death (DODs) or dates of cancer diagnoses of monoallelic *MUTYH* mutation carriers were incomplete, certain assumptions were made so that the data could be used for SMR and SIR calculations. If the month and year were available, the day was assumed to be the 15th and if only the year was given, the day and month were assumed to be the 30th June. If the date of an event was incomplete but the age of the carrier at this event was known, the mid-point of the possible time period was used. The chi-squared test was used to assess the significance of SIRs. Univariate ANOVA (Analysis of Variance) was used to compare the mean ages at CRC diagnosis between MAP patients (chapter 4) and *MUTYH* mutation heterozygotes. Fisher's exact test was used to assess the difference in proportions of monoallelic mutation carriers excluded for the CRC SIR analysis.

Primer name	Primer sequence (5'→3')	Annealing temperature (°C)	Product size (bp)
MYHx1F*	GAAGCTGCGGGAGCTGAAA	60	133
MYHx1R*	ATCCCCGACTGCCTGAACC		
MYHx2F*	CTGCATTTGGCTGGGTCTTT	55	172
MYHx2R	TGGGCCACAACCTAGTTCC		
MYHx3.1F	CCAAGACCCTGATGCACAG	58	155
MYHx3.1R	CCTCGGAAGGCTGTGACTT		
MYHx3.2F	AGCCGGAAGAGGTGGTATTG	58	216
MYHx3.2R	CAACCCCAGATGAGGAGTTAGG		
MYHx4F*	CTCATCTGGGGTTGCATTGA	58	167
MYHx4R*	GGGTTGGCATGAGGACACTG		
MYHx5F*	GGGCAGGTCAGCAGTGTC	58	189
MYHx5R*	TACACCCACCCCAAAGTAGA		
MYHx6F*	TACTTTGGGGTGGGTGTAGA	58	185
MYHx6R*	AAGAGATCACCCGTCAGTCC		
MYHx7(1)F	CACCCTAGGGTAGGGGAAAT	58	182
MYHx7(1)R	ACTCCTGGGGTTCCTACCCT		
MYHx8F	CCAGGAGTCTTGGGTGTCTTAT	58	200
MYHx8R	CTGGGCACGCACAAAGTG		
MYHx9F*	AACTCTTTGGCCCCTCTGTG	58	196
MYHx9R*	GAAGGGAACACTGCTGTGAAG		
MYHx10.1F	AAGGGTGCTTCAGGGGTGTC	58	156
MYHx10.1R	GCGCTGTGGGGTACACACTG		
MYHx10.2F	AGCCCGGCCAGGAGATTT	58	171
MYHx10.2R	AGGGCAGAGTCACTCCTTAGGAC		
MYHx11(1)F	GAAGGGGCAGTGAGAAGTCC	58	176
MYHx11(1)R	CTGACTGGGCCAGGAAGGGT		
MYHx12.1F	AGCCCTCTTGGCTTGAGTA	58	188
MYHx12.1R	CCAGGCTGTTCCAGAACAC		
MYHx12.2(1)F	GGGACCAGACCCTGGGAGTG	58	200
MYHx12.2(1)R	CTGCCGATTCCCTCCATTCTC		
MYHx13.1F	AGGGCAGTGGCATGAGTAACA	58	150
MYHx13.1R	TTGCGCTGAAGCTGCTCTG		
MYHx13.2F	TGGCAGGACTGTGGGAGTT	58	177
MYHx13.2R	ATAGCCTCAAAAGCCAACATC		
MYHx14.1F	TTGGCTTTTGAGGCTATATCC	58	147
MYHx14.1R	CTGGTGGTACGGTGGTCAC		
MYHx14.2F	CAGGTTGTCCACACCTTCTC	58	128
MYHx14.2R	CATGTAGGAAACACAAGGAAGTA		
MYHx15(1)F	CTCACCTCCCTGTCTTCTTG	58	180
MYHx15(1)R	CCAGTGAAGCCTGGAGTGGA		
MYHx16F	GGCCTCCCTCCTTCCATTT	58	182
MYHx16R	GGGAATGGGGGCTTTCAGA		

Table 5.1 Sequences of primers and PCR conditions used to screen *MUTYH*.

*Primers published by Al-Tassan, N. *et al*, 2002. Forty cycles instead of the standard 35 cycles of PCR were used.

Rate	England & Wales	The Netherlands	Germany
CRC incidence	1971-2003	1960-2006	1970-2005
All cancer incidence	1971-2003	1960-2006	1970-2005
CRC mortality	1940-2005	1970-2006	1970-2005
All cancer mortality	1940-2005	1970-2006	1970-2005
All cause mortality	1901-2005	1970-2006	1970-2005

Table 5.2 Time periods for which incidence and mortality rates were used in the present study. Mortality data were available before 1940 in England and Wales but were not used for CRC or all cancer mortality rates as the data were not classified clearly enough. Similarly, some earlier mortality and incidence data from Germany was provided but was not classified clearly enough and was not sufficiently divided into age groups to be used for determination of CRC or all cancer mortality or incidence rates in the present study. For all other rates, no earlier data could be provided.

Cancer	England & Wales	The Netherlands	Germany
CRC	ICD5: 46c, 46d ICD6: 153, 154 ICD7: 153.0-153.8, 154 ICD8: 153.0-153.8, 154 ICD9: 153, 154 ICD10: C18-C21	ICD8: 153, 154 ICD9: 153, 154 ICD10: C18-C20	ICD8: 153, 154 ICD9: 153, 154 ICD10: C18-C20
All cancer	ICD5: 45-55 ICD6: 140-190, 192-205 ICD7: 140-190, 192-207 ICD8: 140-172, 173.5, 174-207 ICD9: 140-172, 174-208 ICD10: C00-43, C45-C85, C88-C94.3, C94.7-C97.9, D32.0-D33.9, D38, D42.0-43.9	ICD8: 140-208 ICD9: 140-208 ICD10: all C codes	ICD8: 140-172, 174-203 ICD9: 140-172, 174-203 ICD10: all C codes

Table 5.3 ICD codes used to select appropriate data for CRC and all cancer incidence and mortality rates. ICD codes 5, 6, 7, 8, 9 & 10 cover time periods 1940-1949, 1950-1957, 1958-1967, 1968-1978, 1979-2000, 2001-present, respectively.

5.3 Results

5.3.1 Data collection

Mortality and cancer data for the parents of 48, 55 and 71 MAP index patients were collected by the Cardiff, Leiden and Bonn centres, respectively. Two families from the Netherlands and one family from Germany were already known to have two consecutive generations affected by MAP with characterized biallelic mutations in both generations and so data for an additional obligate *MUTYH* mutation heterozygote parent were obtained from each of these families (parents and spouses of the affected individual in the first generation). In total, data were collected for 351 apparent monoallelic *MUTYH* mutation carrier parents.

5.3.2 CRC risk in MUTYH mutation heterozygotes

5.3.2.1 Verification of MUTYH mutation status in CRC cases

Twenty-two parents of MAP index cases had been diagnosed with CRC, of whom one was already known to be a *MUTYH* mutation heterozygote through previous molecular genetic analysis. To assess whether other parents with CRC were biallelic *MUTYH* mutation carriers, attempts were made to obtain non-cancerous tissue blocks from their CRC surgery (if undertaken). Tissue blocks were obtained for all four of the UK CRC cases, four of the eight Dutch CRC cases of unknown mutation status and one of the nine German CRC patients of unknown mutation status.

MUTYH ORF sequencing of DNA extracted from non-cancerous tissue cores revealed biallelic mutations in the single German sample (i.e. 1 of 9 cases with tissue identified and 1 of 7 in which at least some molecular analysis was possible). This patient was a P154L homozygote and therefore excluded from the study, bringing the monoallelic *MUTYH* mutation carrier total to 350. A heterozygous Y176C mutation was identified in four CRC cases, two of which were screened for all exons and the other two were screened for all exons except exon 4 due to an insufficient quantity of DNA. Two CRC patients harboured a heterozygous G393D mutation, of which one was screened for all exons and the other was screened for all exons except exons 5, 12 and 14 due to an insufficient quantity of DNA. Amplification of DNA extracted from the

remaining two tissue blocks failed for all *MUTYH* exons. One of these blocks had a very small amount of tissue which was mainly fat cells and the other tissue block had been stained which could have interfered with the extraction or amplification process.

5.3.2.2 CRC incidence and mortality

Of the 350 obligate monoallelic *MUTYH* mutation carriers identifed by the three centres, 21 (7 male and 14 female) had been diagnosed with CRC (Table 5.4) at a mean age of 70 years (median 71 years, range 58-82 years). To assess whether or not more monoallelic *MUTYH* mutation carriers were affected by CRC than would have been expected in the relevant general populations, SIRs were calculated for the countries separately and together (Table 5.5). A total of 77 monoallelic *MUTYH* mutation carriers had to be excluded for this analysis as detailed in Table 5.4. Significantly more Dutch monoallelic *MUTYH* mutation carriers were diagnosed with CRC than would have been expected ($p<0.05$, chi-squared test) and the numbers of CRCs were also higher than expected in monoallelic mutation carriers from the UK and Ireland and from Germany (Table 5.5), although the individual data from these countries did not reach significance ($p>0.05$). Combining the data from all three centres revealed a significantly increased SIR for CRC of 2.12 (95% CIs 1.30-3.28, $p<0.01$, chi-squared test) among obligate monoallelic carrier parents.

Eight of the 350 monoallelic *MUTYH* mutation carriers (4 male and 4 female) died from CRC (Table 5.4). The cause of death for 23 monoallelic *MUTYH* mutation carriers could not be confirmed or excluded to be CRC. To investigate whether or not more monoallelic *MUTYH* mutation carriers died from CRC than would have been expected in the general population, SMRs were calculated for the countries separately and together (Table 5.5). A total of 84 monoallelic *MUTYH* mutation carriers had to be excluded for this analysis as detailed in Table 5.4. No difference in CRC mortality was observed between monoallelic mutation carriers and the general population, even when the data from all three centres were combined (Table 5.5).

	UK & Ireland	The Netherlands	Germany	Total
CRC	4	9	8	21
No CRC	86	103	126	315
CRC unknown	6	0	8	14
Exclusions for CRC SIR analysis				
Lived & died outside country	9	0	0	9
No DOB &/or DOD/last contact	10	22	7	39
Death/CRC diagnosis before population rate data recorded	5	4	13	22
Remaining unknown CRC status	2	0	5	7
Total excluded	26	26	25	77
Cause of death CRC	1	5	2	8
Cause of death not CRC	87	99	133	319
Cause of death unknown	8	8	7	23
Exclusions for CRC SMR analysis				
Lived & died outside country	9	0	0	9
No DOB &/or DOD/last contact	10	22	7	39
Death before population rate data recorded	0	10	13	23
Remaining unknown if CRC cause of death	2	7	4	13
Total excluded	21	39	24	84

Table 5.4 CRC incidence and mortality data for 350 monoallelic *MUTYH* mutation carriers. Exclusions for SIR and SMR analysis are shown. Start dates for population rate data are shown in Table 5.2.

Country	Observed	Expected	CRC SIR	95% Confidence Intervals Lower	Upper
UK & Ireland	4	1.77	2.26	0.62	5.79
The Netherlands	9	3.49	2.58*	1.18	4.90
Germany	7	4.18	1.68	0.67	3.45
All countries	20	9.43	2.12**	1.30	3.28
Country	**Observed**	**Expected**	**CRC SMR**	**95% Confidence Intervals** Lower	Upper
UK & Ireland	1	1.46	0.69	0.02	3.82
The Netherlands	4	3.22	1.24	0.34	3.18
Germany	2	2.20	0.91	0.11	3.28
All countries	7	6.88	1.02	0.41	2.10

Table 5.5 Standardized CRC incidence and mortality ratios. Exact SIRs, SMRs and confidence intervals were calculated. Observed refers to the number of *MUTYH* heterozygotes diagnosed with CRC (SIR) or died from CRC (SMR). Exclusions required for this analysis are shown in Table 5.4. *$p<0.05$, **$p<0.01$

5.3.3 All cancer risk in MUTYH mutation heterozygotes

Eighty-four of the 350 monoallelic *MUTYH* mutation carriers (40 male and 44 female) had been diagnosed with at least one cancer (Table 5.6). No data were available on the cancer status of 20 monoallelic mutation carriers. To assess whether or not more monoallelic *MUTYH* mutation carriers were diagnosed with cancer than would have been expected in the general population, SIRs were calculated for the countries separately and together (Table 5.7). A total of 83 monoallelic *MUTYH* mutation carriers had to be excluded for this analysis as detailed in Table 5.6. No significant difference in overall cancer incidence was observed between monoallelic mutation carriers and the relevant general populations (Table 5.7). The types of malignant cancer identified in monoallelic *MUTYH* mutation carriers are shown in Table 5.8. Other than CRC, lung cancer was the most frequently observed cancer, with 15 cases reported. Seven obligate monoallelic mutation carriers had been diagnosed with more than one type of cancer. Two cases had CRC and breast cancer aged 82 and 90 years and 75 and 55 years, respectively. A single monoallelic mutation carrier had CRC and lung cancer aged 63 and 74 years, respectively, and another was diagnosed with lymphoma (aged 67 years), lung cancer (aged 68 years) and CRC (aged 75 years). One case had laryngeal cancer and astrocytoma aged 57 and 64 years, respectively, and another case was diagnosed with prostate and bladder cancer aged 59 years. A single case had gastric cancer aged 59 years and uterine cancer at an unknown age.

Sixty-two of the 350 monoallelic *MUTYH* mutation carriers (32 male and 30 female) died from cancer (Table 5.6). The cause of death for 14 obligate monoallelic mutation carriers could not be confirmed or excluded to be cancer. To investigate whether or not more monoallelic *MUTYH* mutation carriers died from cancer more frequently than those in the general population, SMRs were calculated for the countries separately and together (Table 5.7). A total of 74 monoallelic *MUTYH* mutation carriers had to be excluded for this analysis as detailed in Table 5.6. No significant difference in cancer mortality was observed between monoallelic carriers and the general population (Table 5.7).

	UK & Ireland	The Netherlands	Germany	Total
Any cancer	20	29	35	84
No cancer	64	81	101	246
Unknown if had cancer	12	2	6	20
Exclusions for all cancer SIR analysis				
Lived & died outside country	9	0	0	9
No DOB &/or DOD/last contact	10	22	7	39
Death/cancer diagnosis before population rate data recorded	6	4	13	23
No date of diagnosis	0	1	5	6
Remaining unknown cancer status	1	0	5	6
Total excluded	26	27	30	83
Cause of death cancer	14	25	23	62
Cause of death not cancer	75	84	115	274
Cause of death unknown	7	3	4	14
Exclusions for all cancer SMR analysis				
Lived & died outside country	9	0	0	9
No DOB &/or DOD/last contact	10	22	7	39
Death before population rate data recorded	0	10	13	23
Remaining unknown if cancer cause of death	1	1	1	3
Total excluded	20	33	21	74

Table 5.6 All cancer incidence and mortality data for 350 monoallelic *MUTYH* mutation carriers. Exclusions for SIR and SMR analysis are shown. Start dates for population rate data are shown in Table 5.2.

Country	Observed	Expected	Cancer SIR	95% Confidence Intervals	
				Lower	Upper
UK & Ireland	14	12.02	1.17	0.64	1.95
The Netherlands	23	28.27	0.81	0.52	1.22
Germany	23	25.01	0.92	0.58	1.38
All countries	60	65.31	0.92	0.70	1.18
Country	**Observed**	**Expected**	**Cancer SMR**	**95% Confidence Intervals**	
				Lower	Upper
UK & Ireland	12	12.65	0.95	0.49	1.66
The Netherlands	18	15.73	1.14	0.68	1.81
Germany	20	16.18	1.24	0.76	1.91
All countries	50	44.56	1.12	0.83	1.48

Table 5.7 Standardized all cancer incidence and mortality ratios. Exact SIRs, SMRs and confidence intervals were calculated. Observed refers to the number of *MUTYH* heterozygotes diagnosed with any cancer (SIR) or died from cancer (SMR).

Cancer	No. of cases	Mean age at diagnosis (yrs) (range)
CRC	21	70 (58-82)
Lung	16	70 (57-85)
Breast	6	60 (45-90)
Gastric	6	67 (54-78)
Prostate	6	72 (59-80)
Leukaemia	5	64 (52-93)
Lymphoma	3	61 (55-67)
Bladder	3	66 (59-77)
Uterine	2	47
Laryngeal	2	57
Osteosarcoma	2	78 (72-83)
Ovarian	1	72
Cervical	1	>70
Testicular	1	Unknown
Duodenal	1	55
Pancreatic	1	67
Liver	1	59
Plasmocytoma	1	60
Fibrosarcoma	1	68
Glioblastoma	1	65
Astrocytoma	1	64
Thyroid	1	Unknown
Cancer, primary unknown	9	60 (33-79)

Table 5.8 Cancers identified in 84 monoallelic *MUTYH* mutation carriers. Other than CRC, lung cancer was the most frequently identified cancer. Seven monoallelic mutation carriers had more than one type of cancer.

5.3.4 Overall mortality in MUTYH mutation heterozygotes

Of the 350 obligate monoallelic *MUTYH* mutation carriers, 240 (134 male and 106 female) were deceased (Table 5.9). It was unknown whether four were alive or deceased. To investigate the mortality of monoallelic *MUTYH* mutation carriers, SMRs were calculated for the countries separately and together (Table 5.10). A total of 71 monoallelic *MUTYH* mutation carriers had to be excluded for this analysis as detailed in Table 5.9. No difference in overall mortality was observed between monoallelic carriers and the general population (Table 5.10). The mean age at death (among the 240 who had died) was 74 years (range 26-106 years) for females and 68 years (range 29-94) for males. The mean age at last contact for living monoallelic *MUTYH* mutation carriers was 72 years (range 47-92 years) for females and 73 years (range 54-98 years) for males.

	UK & Ireland	The Netherlands	Germany	Total
Deceased	62	83	95	240
Alive at last contact	32	27	47	106
Unknown if deceased	2	2	0	4
Exclusions for all cause SMR analysis				
Lived & died outside country	9	0	0	9
No DOB &/or DOD/last contact	10	22	7	39
Death before population rate data was recorded	0	10	13	23
Total excluded	19	32	20	71

Table 5.9 All cause mortality data for 350 monoallelic *MUTYH* mutation carriers. Exclusions for SMR analysis are shown. Start dates for population rate data are shown in Table 5.2.

Country	Observed	Expected	All cause SMR	95% Confidence Intervals	
				Lower	Upper
UK & Ireland	48	61.67	0.78	0.57	1.03
The Netherlands	57	63.13	0.90	0.68	1.17
Germany	75	67.37	1.11	0.88	1.40
All countries	180	192.17	0.94	0.80	1.08

Table 5.10 Standardized all cause mortality ratios. Exact SMRs and confidence intervals were calculated. Observed refers to the number of *MUTYH* heterozygotes deceased.

5.4 Discussion

5.4.1 Monoallelic MUTYH mutation carriers and CRC risk

The SIR analysis performed on 273 obligate monoallelic mutation carriers from UK and Ireland, the Netherlands and Germany indicated that these individuals had a two-fold increase in CRC risk. As expected, the mean age at CRC diagnosis in monoallelic mutation carriers was significantly higher than that in biallelic mutation carriers (70 years [range 58-82 years] vs. 48 years [range 21-70 years; chapter four], $p<0.001$; ANOVA).

5.4.1.1 Mechanism underlying increased CRC risk

One possible mechanism which may explain the observed increase in CRC risk in monoallelic *MUTYH* mutation carriers is analogous to that in HNPCC, requiring somatic mutation of the wild-type allele. Once both *MUTYH* alleles are mutated, the mutation rate in the compromised cell would be expected to increase leading to increased biallelic inactivation of *APC* and tumour initiation. The later age at CRC diagnosis in monoallelic *MUTYH* mutation carriers compared with MAP patients is consistent with the need for an additional somatic *MUTYH* mutation before *APC* mutations accumulate and tumourigenesis is initiated. Consistent with such a mechanism, LOH of the *MUTYH* locus on the short arm of chromosome 1 has been reported to be more frequent in CRCs from germline monoallelic mutation carriers than in CRCs from wild-type individuals but only small numbers of tumours were described (Croitoru, M.E. *et al*, 2004; Kambara, T. *et al*, 2004) and this loss occurs relatively frequently in sporadic CRC (Shih, I-M. *et al*, 2001; Goel, A. *et al*, 2003)

Another mechanism which could explain the increased CRC risk in monoallelic mutation carriers is that the mutant *MUTYH* allele may lead the heterozygous cell to have a slightly reduced repair activity so that the mutation rate would increase, eventually resulting in the biallelic inactivation of *APC* and tumour initiation. Each of the two somatic *APC* mutations would take longer to arise than in MAP, thus delaying the onset of tumourigenesis. However, any increase in mutation rate in monoallelic mutation carriers might be expected to

lead to earlier biallelic inactivation of *APC* (and thus tumour initiation) than in wild-type individuals.

Given the genotype-phenotype correlations observed in MAP cases in chapter four, certain *MUTYH* mutations may result in an earlier onset of CRC in monoallelic carriers. Five of the seven monoallelic mutation carriers with CRC screened for *MUTYH* mutations in the present study were Y176C heterozygotes and two were G393D heterozygotes. From the *MUTYH* genotypes of their MAP-affected offspring, an additional 12 monoallelic mutation carriers with CRC could potentially be Y176C heterozygotes and only two definitely do not carry this mutation. Y176C is thought to have a more detrimental effect on MUTYH glycosylase activity than G393D. Wild-type levels of mutant MUTYH protein are found in cells derived from a MAP patient with homozygous G393D mutations whereas cells derived from a MAP patient with homozygous Y176C mutations have lower levels of mutant MUTYH protein (Parker, A.R. *et al*, 2005). Therefore if a *MUTYH* heterozygote carries the Y176C mutation, not only is the mutant protein produced less active but there may also be less of it. An increase in mutation rate may be greater and biallelic inactivation of *APC* more likely in Y176C than G393D heterozygotes.

5.4.1.2 Previous kin-cohort studies

There are few kin-cohort studies of CRC risk in monoallelic mutation carriers reported in the literature to date. Jenkins *et al* (2006) recently reported a statistically significant (p=0.02) 3-fold increased CRC risk in monoallelic carriers and Webb *et al* (2006) described a non-significant CRC hazard ratio of 1.74 (95% CI: 0.62-3.60) for monoallelic mutation carriers. Croitoru *et al* (2004) showed that monoallelic mutation carrier CRC cases were significantly more likely than wild-type CRC cases to have a first or second-degree relative with CRC and suggested that this reflected transmission of monoallelic CRC risk as an autosomal dominant trait but the findings of Balaguer *et al* (2007) did not confirm this.

5.4.1.3 Previous case-control studies

Several previous case-control studies have reported a trend towards over-representation of *MUTYH* heterozygotes in CRC cases and suggested the possibility of an increased CRC risk in such individuals (Table 5.11). However, statistical significance was only achieved in two studies and under certain constraints; Farrington *et al* (2005) described a 1.68-fold (95% CI: 1.07-2.95) increase in CRC risk in monoallelic mutation carriers but only over the age of 55 years and Peterlongo *et al* (2006) reported an odds ratio (OR) of 2.79 (95% CI: 1.07-7.25; p=0.04) for monoallelic mutation carriers but significance was lost when this was adjusted for ethnic background, age and gender. The scarcity of statistically significant results is likely to reflect a lack of power to detect any increase in CRC risk due to the low penetrance and population frequency of monoallelic *MUTYH* mutations. The largest case-control study to date is that of Webb *et al* (2006) which did not find an increased CRC risk in monoallelic mutation carriers (Table 5.11). The approach used in the present study allowed analysis of a much larger number of monoallelic mutation carriers.

A few case-control studies did not find an increase in CRC risk in monoallelic carriers but as in the study of Webb *et al* (2006), their CIs for CRC risk included values over 1.0 (Table 5.11). Differences in CRC risk estimates between case-control studies may be influenced by the method of CRC case selection. Even meta-analyses of the ORs of CRC risk associated with monoallelic Y176C or G383D mutations from several case-control studies have not yielded statistically significant results as combined ORs of 1.26 (95% CI: 0.99-1.60) and 1.11 (95% CI: 0.90-1.37) have been reported (Webb, E.L. *et al*, 2006; Balaguer, F. *et al*, 2007). Tenesa *et al* (2006) used a different method of assessing risk for their meta-analysis and found a significant genotype relative risk of 1.27 (95% CI: 1.01-1.61) for monoallelic Y176C or G383D mutation carriers. The largest meta-analysis was that of Balaguer *et al* (2007), with 191/10582 CRC cases and 162/9728 controls carrying monoallelic Y176C or G383D mutations which is still fewer carriers than were involved in the present study.

Reference	CRC cases	Controls	Risk (95% CI)	Significant result
Enholm, S. *et al* (2003)	5/1042	0/424	NC	Yes (p<0.025)
Croitoru, M.E. *et al* (2004)	29/1238	21/1255	OR=1.4 (0.8-2.5)	No
Fleischmann, C. *et al*, (2004)	6/358	2/354	OR=2.9 (0.6-14.4)	No
Kambara, T. *et al*, (2004)	20/92	14/53	NC	No
Wang, L. *et al* (2004)	10/444	4/313	OR=1.78 (0.55-5.71)	No
Farrington, S.M. *et al*, (2005)	45/2239	28/1845	GRR=1.35 (0.92-2.07)	No
>55 years only	31/1345	18/1289	GRR=1.68 (1.07-2.95)	Yes
Peterlongo, P. *et al*, (2005)	4/555	7/918	OR=0.95 (0.20-3.75)	No
Zhou, X-L. *et al*, (2005)	6/438	3/469	NC	No
Colebatch, A. *et al*, (2006)	11/872	5/478	NC	No
Peterlongo, P. *et al*, (2006)	6/137	16/967	OR=2.79 (1.07-7.25)	Yes
Adjusted			OR=1.99 (0.70-5.69)	No
Tenesa, A. *et al*, (2006)	18/928	20/845	GRR=0.8 (0.5-1.4)	No
Webb, E.M. *et al*, (2006)	53/2561	57/2695	OR=0.98 (0.66-1.46)	No
Balaguer, F. *et al*, (2007)	19/1108	21/934	OR=0.76 (0.40-1.42)	No
Küry, S. *et al*, (2007)	24/1023	21/1121	OR=1.26 (0.70-2.27)	No

Table 5.11 Results of previous case-control studies investigating CRC risk in monoallelic *MUTYH* mutation carriers. Only two studies showed a significantly increased CRC risk in monoallelic mutation carriers but several others indicated an over-representation of monoallelic carriers in CRC cases. OR, odds ratio; GRR, genotype relative risk; CI, confidence interval.

5.4.1.4 Potential causes of inaccurate CRC risk estimation

5.4.1.4.1 Incorrect genotype assumptions

The present study attempted to verify that the parental CRC cases identified were monoallelic rather than biallelic mutation carriers by screening their non-cancerous tissue DNA for *MUTYH* mutations. Tissue blocks for nine of the 22 CRC cases were obtained, seven were successfully screened and in one further case blood DNA was available. One of these cases was found to carry pathogenic biallelic *MUTYH* mutations. It is possible that other parental cases with CRC for whom no blood or tissue was available may also have carried biallelic mutations. However, the parental MAP case had 11-20 colorectal adenomas at 74 years and a strong family history of CRC; her father died of CRC aged 80 years and her sister had CRC at 80 years. No information could be obtained about the presence or lack of consanguinity in this family. None of the other parental CRC cases for whom blood or tissue was unavailable were described as having a polyposis phenotype or strong family history of CRC.

A large proportion of the relatives included in the kin-cohort studies of Jenkins *et al* (2006) and Webb *et al* (2006) were not genotyped. Some of the assumed monoallelic mutation carriers with CRC could actually have been biallelic MAP cases and the CRC risk for monoallelic carriers therefore overestimated in these studies. Only one of the case-control studies reported to date screened the entire *MUTYH* ORF in CRC cases and controls (Fleischmann, C. *et al*, 2004). The other studies assayed for Y176C and G393D (and in some cases a small number of other specific changes) which was followed in most studies by an ORF screen of any monoallelic mutation carriers identified (Enholm, S. *et al*, 2003; Croitoru, M.E. *et al*, 2004; Kambara, T. *et al*, 2004; Wang, L. *et al*, 2004; Farrington, S.M. *et al*, 2005; Peterlongo, P. *et al*, 2005; Zhou, X-L. *et al*, 2005; Colebatch, A. *et al*, 2006; Peterlongo, P. *et al*, 2006; Tenesa, A. *et al*, 2006; Webb, E.M. *et al*, 2006; Balaguer, F. *et al*, 2007; Kury, S. *et al*, 2007). CRC cases carrying biallelic mutations other than those specifically screened for would have been incorrectly classified as wild-type cases and it could have contributed to an underestimation of risk. Kambara *et al* (2004) reported no significant difference in the frequency of monoallelic carriers in CRC cases and controls but their classification of a monoallelic carrier included individuals

harbouring variants such as V22M or S501F that appear to be non-pathogenic polymorphisms (Al-Tassan, N. *et al*, 2002). All monoallelic carriers in the present study carried pathogenic mutations that resulted in MAP in their offspring.

5.4.1.4.2 Ascertainment bias

To avoid a bias in favour of monoallelic mutation carriers with CRC, the present study included obligate heterozygote parents only of those MAP index cases who presented symptomatically or were identified as a result of population screening, but not those who presented because of a family history of CRC. The MAP index cases were identified through colorectal polyposis registers for which a family history of CRC was not an inclusion criterion.

SIR analysis is the method used by epidemiologists to investigate if there is a greater incidence (and therefore risk) of a disease in particular individuals than in the general population. In order to control for changing cancer incidence over time, for obligate heterozygote parents to be included in this analysis certain information (e.g. DOB and DOD) was required otherwise they had to be excluded. If a MAP index case had a parent affected by CRC, more effort may have been made by the clinical genetics services to obtain details these details than for an unaffected parent. This could have led to a bias in favour of inclusion of parents with CRC and an overestimation of CRC risk in monoallelic mutation carriers. Whereas 5% (1/21) of all the parents with CRC were excluded, 23% (76/329) of the remaining parents were excluded (p=0.034, Fisher's exact test). However, if these 76 parents were assumed to be unaffected by CRC and, along with the excluded CRC case, were included in the SIR analysis (through using randomly selected information from the existing cohort) a significantly increased CRC risk is still seen (SIR=1.87, 95% CIs 1.16-2.87; $p<0.05$, chi squared test).

Although many of the MAP patients from UK and Ireland underwent pre-screening for the two common mutations, 17% of MAP patients identified by the three centres (Cardiff, Bonn, Leiden) did not harbour a Y176C or G393D mutation and 29% carried Y176C or G393D in combination with a less frequent

mutation (see chapter 4). As noted in chapter four, the ratio of Y176C homozygotes to G393D homozygotes observed in our MAP patient cohort was greater than expected from the allele frequencies for these mutations in healthy individuals. In addition, G393D homozygotes presented and were diagnosed with CRC at an older age than Y176C homozygotes, suggesting that G393D homozygotes may be less likely to come to the attention of polyposis registers from which our index cases were selected for screening (chapter 4). The SIR we found may be an overestimation of CRC risk in monoallelic carriers as a whole in the general population. Balaguer *et al* (2007) observed a larger monoallelic effect for Y176C carriers than for G393D carriers but Tenesa *et al* (2006) found similar effects for both mutations. Neither of these observations was statistically significant and the 95% CI range for the analysis by Balaguer *et al* (2007) was large (OR=5.33; 95% CI: 0.70-38.40).

5.4.2 CRC mortality in monoallelic MUTYH mutation carriers

The SMR analysis performed on 266 monoallelic mutation carriers from UK and Ireland, the Netherlands and Germany indicated that no more of these individuals died from CRC than would have been expected in the general population. This is the first study to investigate CRC mortality in monoallelic mutation carriers and given the significantly increased CRC incidence observed above, it might have been expected that CRC mortality would also have been increased in these individuals. If the treatment and prognosis of CRCs in monoallelic mutation carriers was the same as for CRCs in the general population, a similar proportion of CRC cases would be expected to die in both groups. The numbers in the present study were small and the SMR 95% CI does not preclude an increase in CRC mortality for monoallelic mutation carriers (0.41-2.10). Of the 21 monoallelic mutation carrier CRC cases, only 8 were known to have died from CRC and one of these had to be excluded from the SMR analysis as they died before the appropriate population rate data was collected. Four parents with CRC were still alive at the end of the study, one died as a result of an overdose and the cause of death for another was unknown. The collection of data for the population CRC mortality rates for the Netherlands began ten years later than the CRC incidence rate data from this

country, leading to the additional exclusion of six monoallelic carriers from the study

5.4.3 All cancer incidence and mortality in monoallelic MUTYH mutation carriers

The SIR and SMR analyses performed on 267 and 276 monoallelic mutation carriers, respectively, from UK and Ireland, the Netherlands and Germany indicated that these individuals do not have a significantly increased risk of being diagnosed with, or dying from, cancer. However, neither analysis could exclude an increase as the CIs included values over one (SIR 95% CI: 0.70-1.18; SMR 95% CI: 0.83-1.48). This is the first large study to investigate all cancer incidence and mortality in monoallelic mutation carriers. Although CRC risk was found to be significantly increased in monoallelic mutation carriers, the all cancers risk was not. The all cancers SIR analysis only used information on the first cancer as subsequent cancers could have been the result of treatment for the first cancer or misdiagnosed metastases. Monoallelic carriers reported to be affected by cancer but for whom no date of diagnosis was known (or could be reasonably assumed) had to be excluded from the SIR analysis as the end of follow-up date could not be identified and not all monoallelic carriers were old enough to have lived through their cancer risk so an increase in risk in later life was not fully excluded.

Other than CRC, the most frequently identified cancer in monoallelic mutation carriers was lung cancer. Of the 330 carriers for whom cancer status was known, 16 (4.8%) had lung cancer which is as expected because this is the second most common cancer in the UK (Cancer Research UK). Al-Tassan *et al* (2004) reported only a single germline G393D heterozygote in a cohort of 276 lung cancer patients which is consistent with the frequency of this mutant allele in healthy individuals. The only GI tract cancer in addition to CRC identified in more than one monoallelic mutation carrier in the present study was gastric cancer (6 cases, 1.8%). Several Asian gastric cancer cases have been screened for mutations in *MUTYH*. Kim *et al* (2004) reported biallelic *MUTYH* mutations (one somatic missense mutation and LOH of the wild-type allele) in 2/95 Korean sporadic gastric cancers and 4/23 cases showed LOH at the *MUTYH* locus. Tao *et al* (2004) described a homozygous germline splice site

mutation in 1/148 Japanese gastric cancer patients but the heterozygous change was no more frequent in cases than in healthy controls. The frequency of a variant haplotype at two *MUTYH* polymorphisms was reported to be associated with an increased risk of familial gastric cancer in a Chinese study (Zhang, Y. *et al*, 2006). Gastric cancer is more prevalent in these Asian populations than in Europe and the frequency of this cancer in monoallelic mutation carriers in the present study in consistent with the UK population (3% of diagnosed cancers). A single Y176C heterozygote has been reported previously with prostate cancer (Shin, E.J. *et al*, 2007) and one Y176C heterozygote has been described with acute myeloid leukaemia (Akyerli, C.B. *et al*, 2003). The frequencies of prostate cancer, breast cancer and leukaemia identified in the present study are not greater than would be expected in the UK with 24% and 3% of all cancers accounted for by prostate cancer and leukaemia, respectively and the lifetime risk of breast cancer for women is 11% (Cancer Research UK). All other cancers identified in the present study were only diagnosed in less than 1% of monoallelic carriers. Monoallelic germline Y176C or G383D mutations have been identified in hepatocellular carcinoma and cholangiocarcinoma patients but at no higher frequency than in cancer-free controls (Baudhuin, L.M. *et al*, 2006) and no pathogenic germline mutations were detected in 29 sporadic squamous oral carcinoma patients (Görgens, H. *et al*, 2007). Given the finding that MAP patients do not appear to be at a significant risk of cancers outside the GI tract (chapter 4), it is not surprising that monoallelic carriers also do not appear to have a significantly increased risk of such cancers. The present study is underpowered to detect small increases in risk for rare cancers.

5.4.4 All cause mortality in monoallelic MUTYH mutation carriers

The SMR analysis performed on 279 monoallelic mutation carriers from UK and Ireland, the Netherlands and Germany showed no difference to the general population. This is the first study to investigate overall mortality in monoallelic *MUTYH* mutation carriers and the results will be helpful in reassuring relatives of MAP cases.

5.4.5 Clinical implications

The present study identified a statistically significant two-fold increase in CRC risk in monoallelic *MUTYH* mutation carriers which is a clinically relevant outcome. However, before genetic counselling and surveillance strategies for monoallelic carriers are changed in view of these findings, a study of whether or not all monoallelic mutations confer the same CRC risk is indicated. Evidence from MAP patients suggests that the Y176C mutant allele is associated with an earlier CRC onset than the G393D mutant allele (chapter 4) and it has been suggested (although no significant results have been reported) that Y176C monoallelic carriers may have a greater CRC risk than G393D monoallelic carriers. If this is the case, it may explain why some previous population based CRC case-control studies have failed to detect a significantly increased CRC risk in monoallelic carriers whereas the present study found a doubling of risk. A large collaborative study of genotyped heterozygotes would be needed to assess if different CRC risks are associated with different *MUTYH* mutations, but even then the CRC risk of rarer mutations will be difficult to determine. If different mutations are associated with different levels of CRC risk, this will have implications for genetic counselling and surveillance of monoallelic mutation carriers. For example, screening could be offered only to those with a higher risk mutation thus reducing costs. The present study does not provide evidence that monoallelic carriers are at a significantly increased risk of cancers other than CRC or of other potentially fatal diseases. Large prospective studies of monoallelic carriers are ideally needed to confirm the findings of the present study but will take a long time to complete.

Chapter Six

Mutational analysis of the base excision repair and DNA damage protection genes *TDG*, *MPG*, *SMUG1*, *MBD4*, *APE1*, *POLL* and *NUDT5* in patients with multiple colorectal adenomas

6.1 Introduction

6.1.1 Inherited CRC and BER

Known genetic predisposition syndromes caused by high penetrance mutations only account for a small proportion of hereditary CRC cases and incompletely penetrant mutations are thought to contribute to most of the remaining familial CRC cases (Kemp, Z. *et al*, 2004; de la Chapelle, A., 2004). The present study sought inherited high penetrance pathogenic mutations in candidate genes involved in DNA repair in fifty-eight multiple colorectal adenoma cases in whom no underlying genetic defect had been identified.

MAP is a recently described autosomal recessive disease resulting from biallelic germline mutations in the *MUTYH* gene (Al-Tassan, N. *et al*, 2002; Sieber, O.M. *et al*, 2003). MUTYH is a DNA glycosylase which functions in the BER pathway to remove adenine misincorporated opposite 8-oxoG as a result of oxidative DNA damage (Shinmura, K. *et al*, 2000). BER is the primary mechanism of protection against DNA damage caused by cellular metabolism including lesions resulting from ROS, deamination, methylation and hydroxylation (Hoeijmakers, J.H.J, 2001). Some base modifications caused by exogenous agents can also be repaired by BER (King, R.J.B., 2000). The identification of MAP has shown that compromised BER can influence CRC susceptibility and thus components of this pathway are good candidates for novel CRC predisposition genes.

DNA glycosylases initiate BER by recognising DNA damage and most can be grouped into families according to structural and/or functional similarities.

6.1.2 MutM/Nei glycosylases

Members of this DNA glycosylase family are characterised by a C-terminal zinc finger and a helix-two turn-helix motif which are involved in DNA binding. The human enzymes NEIL1, NEIL2 and NEIL3 are part of this family (Zharkov, D.O. *et al*, 2003).

6.1.2.1 NEIL1, NEIL2 & NEIL3

The *NEIL1* gene is found on chromosome 15q22, consists of nine exons and spans 6.1kb (Bandaru, V. *et al*, 2002; Morland, I. *et al*, 2002). The encoded protein is expressed in the nuclei of a range of tissues (Morland, I. *et al*, 2002) and is under S phase specific regulation (Hazra, T.K. *et al*, 2002a). NEIL1 lacks a zinc finger, but contains a 'zincless finger' instead (Doublié, S. *et al*, 2004). The *NEIL2* gene is made up of four exons and lies at chromosome 8p23 (Bandaru, V. *et al*, 2002). NEIL2 expression is cell cycle independent and found at high levels in muscle and testis (Hazra, T.K. *et al*, 2002b). Both NEIL1 and NEIL2 are glycosylases with AP lyase activity and unlike other glycosylases use a βδ-elimination mechanism for catalysis. The resulting 3' terminal phosphate is a poor substrate for mammalian APE1 but is removed by PNK in an APE1-independent BER mechanism (Hazra, T.K. *et al*, 2007). Both of these enzymes are active on substrates within DNA bubble structures suggesting that NEIL1 and NEIL2 repair may be associated with replication or coupled to transcription, respectively (Dou, H. *et al*, 2003). ROS-derived pyrimidine lesions are substrates for both glycosylases and NEIL1 preferentially excises thymine glycol, 5-hydroxyuracil as well as ring-opened purines-formamidopyrimidine (Fapy)-adenine and –guanine. The preferred substrates of NEIL2 include hydantoins (oxidation products of 8-oxoG), 5-hydroxyuracil and 5-hydroxycytosine (Hazra, T.K. *et al*, 2007). *NEIL3* is found at chromosome 4q34 and consists of ten exons spanning 55kb (Bandaru, V. *et al*, 2002; Morland, I. *et al*, 2002). The encoded protein has not been biochemically characterised to date.

6.1.3 Endonuclease III glycosylases

Members of this DNA glycosylase family are characterised by a conserved catalytic domain containing an HhH motif followed by a glycine/proline-rich loop

and a catalytic aspartate residue (HhH-GPD motif). Some HhH glycosylases also contain an iron-sulphur cluster (Nash, H.M. *et al*, 1996). The human enzymes MUTYH, OGG1, OGG2, NTHL1 and MBD4 are part of this family.

6.1.3.1 OGG1 & OGG2

The *OGG1* gene is located at chromosome 3p26.2 (Arai, K. *et al*, 1997) and consists of nine alternatively spliced exons which encode both nuclear and mitochondrial isoforms of the protein (Kohono, T. et al, 1998; Nishioka, K. *et al*, 1999). OGG1 is ubiquitously expressed (Arai, K. *et al*, 1997) and the predominant, nuclear isoform has both glycosylase and AP lyase activities (i.e is bifunctional) (Roldán-Arjona, T., et al, 1997). OGG1 substrates include 8-oxoG mispaired opposite cytosine, Fapy-G and 7,8-dihydro-8-oxoadenine (8-oxoA) (Klungland, A. & Bjelland, S., 2007). OGG2 also excises 8-oxoG, but prefers 8-oxoG:A and 8-oxoG:G mispairs. This bifunctional glycosylase was identified in human cell extracts and is likely to act on the daughter strand following misincorporation of 8-oxodGTP opposite adenine during DNA replication (Hazra, T.K. *et al*, 1998).

6.1.3.2 NTHL1

The *NTHL1* gene is found on chromosome 16p13 and contains six exons which span 8kb (Imai, K. *et al*, 1998). NTHL1 is the human homologue of *E.coli* endonuclease III and possesses both glycosylase and AP lyase activities (Aspinwall, R. *et al*, 1997). Multiple nuclear isoforms of this enzyme exist (Ikeda, S. *et al*, 2002) and NTHL1 expression is ubiquitous and increased during S-phase (Imai, K. *et al*, 1998; Luna, L. *et al*, 2000). NTHL1 excises a range of oxidized pyrimidine lesions including thymine glycol, dihydrouracil, 5-hydroxycytosine and urea, (Hazra, T.K. *et al*, 2007). This glycosylase also recognises AP sites and 8-oxoG when paired opposite guanine (Eide, L. *et al*, 2001; Matsumoto, Y. *et al*, 2001).

6.1.3.3 MBD4

MBD4 (also known as *MED1*) is found on the long arm of chromosome 3, at 3q21-q22 (Riccio, A. *et al*, 1999) and is made up of 8 exons (Hendrich, B. *et al*, 1999a). Alternatively spliced transcripts, including one which encodes a

truncated protein, have been identified. *MBD4* is expressed in a range of human tissues and the encoded protein localizes to the nucleus (Hendrich, B. & Bird, A, 1998). This enzyme is a member of the methyl-CpG-binding domain (MBD) family of proteins which mediate histone deacetylase-dependent transcriptional silencing at symmetrically methylated CpG islands (Ballestar, E. & Wolffe, A.P., 2001). CpG islands are GC-rich regions of DNA that mostly overlap the promoter regions of RNA polymerase II-transcribed genes (Bird, A., 2002). Aberrant methylation of CpG islands in the promoters of many cancer-related genes results in silencing of their expression (Baylin, S.B. *et al*, 2001). The MBD proteins target corepessor complexes which contain histone deacetylase(s) (HDACs) and other transcriptional repressors to methylated DNA, thereby silencing expression (Ballestar, E. & Wolffe, A.P., 2001). MBD4 can bind methylated DNA *in vivo* (Hendrich, B. & Bird, A, 1998) and binds to the hypermethylated promoters of human cancer-related genes *CDKN2A* and *MLH1*, repressing their transcription in a histone deacetylase-dependent manner (Kondo, E. *et al*, 2005).

Unlike the other MBD proteins, MBD4 is the only one to have an innate DNA glycosylase activity (Hendrich, B. *et al*, 1999b). MBD4 is a monofunctional glycosylase (lacks AP lyase activity) (Hendrich, B. *et al*, 1999b; Petronzelli, F. *et al*, 2000a) and binds tightly to AP sites which may protect them from non-specific processing (Petronzelli, F. *et al*, 2000a). MBD4 recognises the mismatched products of 5-methylcytosine (5meC) and cytosine deamination, removing the respective T or U mispaired opposite G (Hendrich, B. *et al*, 1999b; Petronzelli, F. *et al*, 2000a; Zhu, B. *et al*, 2000). G:T mispairs are often associated with CpG dinucleotides as 70-80% of all CpG dinucleotides (with the exception of CpG islands) are methylated in humans (Bird, A., 2002). G:T mismatches within CpG or 5meCpG sites are the preferred substrates of this enzyme (Petronzelli, F. *et al*, 2000a; Hendrich, B. *et al*, 1999b) and mice deficient in MBD4 show a 2-3 fold increase in the frequency of C to T transitions at CpG sites (Millar, C.B. *et al*, 2002; Wong, E. *et al*, 2002). On a cancer-susceptible $Apc^{Min/+}$ background, $Mbd4^{-/-}$ mice have increased numbers of intestinal tumours and accelerated tumour progression associated with an increase in somatic CpG to TpG mutations in the *Apc* gene (Millar, C.B. *et al*,

2002; Wong, E. *et al*, 2002). Other MBD4 substrates include 5meC, thymine glycol, 5-formyluracil, 5-fluorouracil or 3,N^4-ethenocytosine paired opposite guanine (Zhu, B. *et al*, 2000; Yoon, J-H., *et al*, 2003; Liu, P. *et al*, 2003; Petronzelli, F. *et al*, 2000a; Petronzelli, F. *et al*, 2000b). This glycosylase can also remove thymine from O^6-methylguanine (O^6-meG):T mismatches (Cortellino, S. *et al*, 2003).

MBD4 interacts with Fas-associated death domain protein (FADD), an adaptor protein implicated in multiple apoptotic responses (Screaton, R.A. *et al*, 2003) and deficiency of MBD4 reduces intestinal apoptosis in response to a variety of DNA damaging agents (Sansom, O.J. *et al*, 2003). *Mbd4*$^{-/-}$ MEFs do not undergo G_2-M cell cycle arrest and apoptosis when treated with antitumour DNA damaging agents so DNA damage accumulates (Cortellino, S. *et al*, 2003).

Frameshift mutations at a poly(A)$_{10}$ tract in *MBD4* have been identified in 24-43% of MSI colorectal cancers, including sporadic and hereditary cases (Riccio, A. *et al*, 1999; Bader, S. *et al*, 1999). Most commonly one *MBD4* allele is mutated at the poly(A)$_{10}$ tract and the other remains wild-type, but biallelic inactivation of *MBD4* through LOH or frameshift mutation of the second allele has been found in a small number of cases (Riccio, A. *et al*, 1999; Bader, S. *et al*, 2000).

6.1.4 Uracil DNA glycosylases

Members of the uracil DNA glycosylase (UDG) superfamily are characterized by the presence of two conserved active site motifs and an aromatic residue which forms part of the substrate binding pocket (Pearl, L.H., 2000). These enzymes are monofunctional (lack AP lyase activity) and share a common α/β-fold structure (Aravind, L. & Koonin, E.V., 2000). The superfamily can be divided into four families of glycosylases based on their specificity and mechanism of substrate recognition (Pearl, L.H., 2000). The human enzymes TDG, SMUG1 and UNG are members of different families within the UDG superfamily

6.1.4.1 UNG

The *UNG* gene is found at chromosome 12q23-q24.1 (Haug, T. *et al*, 1996) and is made up of seven exons (Nilsen, H. *et al*, 1997). Alternative splicing and the use of its two promoters result in nuclear (UNG2) and mitochondrial (UNG1) isoforms (Nilsen, H. *et al*, 1997). Detectable levels of UNG2 expression are found only in proliferating tissues which include the colon, small intestine, thymus, placenta and testis (Haug, T. *et al*, 1998). UNG2 expression is cell cycle dependent and peaks early in S phase, accumulating in replication foci then is degraded during late S phase (Otterlei, M. *et al*, 1999; Hardeland, U. *et al*, 2005; Kavli, B. *et al*, 2002; Fischer, J.A. *et al*, 2004). UNG2 repairs deaminated cytosine lesions (U:G) and is the enzyme mainly responsible for excision of uracil misincorporated into nascent DNA during replication (U:A) (An, Q. *et al*, 2005; Kavli, B. *et al*, 2002). Other UNG2 substrates include 5-hydroxyuracil, isodialuric acid and alloxan (Dizdaroglu, M. et al, 1996). This glycosylase can also remove uracil from single-stranded DNA (Kavli, B. *et al*, 2002) and is involved in immunoglobulin class-switch recombination and somatic hypermutation (Kavli, B. *et al*, 2005).

6.1.4.2 SMUG1

The *SMUG1* gene is located on the long arm of chromosome 12 (Haushalter, K.A. *et al*, 1999), spans approximately 7.5 kb and consists of two coding exons. The encoded protein is nuclear (Haushalter, K.A. *et al*, 1999) and accumulates in nucleoli but not in replication foci (Kavli, B. et al, 2002). SMUG1 excises uracil and its derivatives from double-stranded and single-stranded DNA (Haushalter, K.A. *et al*, 1999; Masaoka, A. *et al*, 2003). The modified uracil bases recognised by SMUG1 include 5-hydroxyuracil, 5-hydroxymethyluracil and 5-formyluracil (Masaoka, A. *et al*, 2003). SMUG1 binds tightly to AP sites and is important in the repair of deaminated cytosine (U:G) in non-replicating chromatin (Pettersen, H.S. *et al*, 2007). Two deaminated derivatives of oxidized cytosine, isodialuric acid and alloxan, are also SMUG1 substrates (An, Q. *et al*, 2005). Silencing of *SMUG1* in MEFs produces a mutator phenotype which moderately increases the frequency of C:G to T:A transitions (An, Q. *et al*, 2005).

6.1.4.3 TDG

The *TDG* gene is found at chromosome 12q24.1 (Sard, L. *et al*, 1997), spans approximately 23 kb and is made up of ten exons. The encoded protein is expressed at a similar level in a range of tissues (Sard, L. *et al*, 1997). TDG is a nuclear glycosylase under cell cycle regulation, it is eliminated by cells entering S-phase and deficiency is maintained until G2 (Hardeland, U. *et al*, 2005). TDG excises uracil and thymine when mispaired opposite guanine. These mismatches result from the spontaneous deamination of cytosine and 5-methylcytosine, respectively, and if left unrepaired lead to C to T transitions (Neddermann, P. & Jiricny, J., 1994). TDG shows a preference for G:T mismatches in a CpG sequence context (Neddermann, P. *et al*, 1996). A chicken homologue of TDG possesses 5-methylcytosine DNA glycosylase activity and its overexpression results in demethylation of the promoter of a hormone-regulated reporter gene (Zhu, B. *et al*, 2001). Other TDG substrates include 3,N^4-ethenocytosine, a highly mutagenic adduct, and derivatives of uracil such as 5-hydroxyuracil and 5-hydroxymethyluracil (Cortázar, D. *et al*, 2007). Most TDG substrates are mispaired opposite guanine which may reflect the enzyme's substrate recognition mechanism. It is likely that interactions with the nucleotide opposite the aberrant base are important for substrate recognition by TDG (Cortázar, D. *et al*, 2007). Similar to MUTYH, TDG can interact with the 9-1-1 DNA damage sensor complex, stimulating its glycosylase activity. TDG may act as an adaptor for sensor checkpoint proteins by recognizing DNA damage then recruiting the 9-1-1 complex to stimulate BER and initiate signal transduction pathways (Guan, X. *et al*, 2007). TDG binds with high affinity to AP sites but is displaced by APE1 (Waters, T.R. *et al*, 1999). TDG can be modified by SUMOs (Small Ubiquitin-like Modifiers) which induce a conformational change involving the N-terminal domain of TDG that aids release of the enzyme from AP sites due to the abrogation of non-specific DNA binding (Steinacher, R. & Schär, P, 2005). Thymine excision from G:T mispairs is virtually absent in SUMO-conjugated TDG but G:U mismatch processing occurs with enhanced efficiency (Hardeland, U. *et al*, 2002; Steinacher, R. & Schär, P, 2005). TDG may also have a role in the regulation of gene expression through interactions with transcription factors of the nuclear receptor family including RAR (retinoic acid receptor), RXR (retinoid X receptor) and

oestrogen receptor α (Um, S. *et al*, 1998; Chen, D. *et al*, 2003). Transcriptional coactivators CBP (CREB binding protein), p300 and SCR1 also interact with TDG (Tini, M. *et al*, 2002; Lucey, M.J. *et al*, 2005).

6.1.5 MPG

The *MPG* gene lies within the terminal region of chromosome 16p, consists of 5 exons and is expressed in a range of tissues (Vickers, M.A. *et al*, 1993). Two MPG isoforms with the same DNA glycosylase activities result from alternative splicing of this gene (Pendlebury, A. *et al*, 1994; O'Connor, T.R., 1993). MPG is a monofunctional glycosylase found mainly at the periphery of the nucleus and around nucleoli (Campalans, A. *et al*, 2005). Expression of *MPG* is cell cycle dependent, increasing early in G1 then remaining constant until after mitosis (Bouziane, M. *et al*, 2000). MPG excises a range of damaged bases including adenine and guanine alkylation adducts, hypoxanthine, 1,N^6-ethenoadenine, 8-oxoguanine and its oxidation product, cyanuric acid (Chakravarti, D. *et al*, 1991; O'Connor, T.R., 1993; Saparbaev, M. & Laval, J., 1994; Dosanjh, M.K. *et al*, 1994; Bessho, T. *et al*, 1993; Dherin, C. *et al*, 2004). MBD1 and MPG have been shown to interact, linking transcriptional repression with DNA repair in chromatin (Watanabe, S. *et al*, 2003).

6.1.6 Additional enzymes involved in BER and DNA damage protection

Several enzymes function downstream of DNA glycosylases in the BER pathway, including APE1 and pol λ. Other enzymes such as NUDT1 and NUDT5 act to prevent the incorporation of aberrant nucleotides into nascent DNA.

6.1.6.1 Apurinic/apyrimidinic endonuclease 1 (APE1)

The *APE1* gene is found at chromosome 14q11.2-q12 and spans 2.64 kb. It consists of five exons (the first of which is untranslated) and is ubiquitously expressed (Akiyama, K. *et al*, 1994; Harrison, L. *et al*, 1992). APE1 is a nuclear protein which possesses AP endonuclease, 3'→5' exonuclease, 3'-repair diesterase and 3' phosphatase activities (Demple, B. *et al*, 1991; Seki, S. *et al*, 1991; Chou, K-M. & Cheng, Y-C., 2002). This is the major human AP endonuclease (Chen, D.S. *et al*, 1991) which hydrolytically cleaves the DNA

phosphodiester backbone 5' to an AP site, generating a terminal 3' hydroxyl group ready for repair synthesis (Levin, J.D. & Demple, B., 1990). DNA glycosylases produce AP sites which are recognised by APE1 and some, such as OGG1, UNG, TDG and MPG, are specifically stimulated by the endonuclease (Vidal, A.E. *et al*, 2001; Parikh, S.S. *et al*, 1998; Waters, T.R. *et al*, 1999; Xia, L. *et al*, 2005). APE1 can also stimulate the activities of enzymes downstream in the BER pathway; it facilitates binding of Pol β to DNA and stimulates the dRP lyase activity of this polymerase (Bennett, R.A. *et al*, 1997). The enzymatic activities of FEN1 and DNA ligase I (which are involved in long-patch BER) are stimulated by this endonuclease and APE1 blocks re-ligation of a damaged site and prevents a futile cleavage and ligation cycle (Ranalli, T.A. *et al*, 2002). APE1 has DNA exonuclease activity on mismatched nucleotides at 3' termini of nicked or gapped DNA and may increase the fidelity of BER (Chou, K-M. & Cheng, Y-C., 2002). Suppression of APE1 in colon cancer cells causes inhibition of cell proliferation and activation of apoptosis. The steady-state level of endogenous AP DNA damage is up to 10 times greater in APE1-suppressed cells than in control cells (Fung, H. *et al*, 2005). APE1 also has a second function, mediated by its Ref1 domain, regulating the DNA binding activity of proteins such as Fos, Jun and p53 through posttranslational reduction (Xanthoudakis, S. *et al*, 1992; Jayaraman, L. *et al*, 1997). Another AP endonuclease is present in humans, APE2. This enzyme has only weak AP endonuclease activity but shows strong 3'→5' exonuclease and 3' phosphodiesterase activities and may have a role in processing 3'-blocking termini or 3' mismatched nucleotides in nascent DNA (Burkovics, P. *et al*, 2006).

6.1.6.2 DNA polymerase lambda (pol λ)

The *POLL* gene spans approximately 9kb on chromosome 10q24.3-q25.1 and consists of 8 exons. The major 2.7kb transcript is most highly expressed in testis, ovary and foetal liver but is also weakly expressed in many other tissues. Minor splice variants have been identified but are considered to be non-functional (Aoufouchi, S. *et al*, 2000; Nagasawa, K-I. *et al*, 2000).

Pol λ is part of the X family of DNA polymerases which also includes Pol β and all members contain a pol-X core domain which mediates the addition of dNTPs to the 3' end of the growing DNA chain (Ramadan, K. *et al*, 2004a). Pol λ is a DNA repair, not replicative, polymerase and exhibits both template-dependent (DNA polymerase) (García-Díaz, M. *et al*, 2002) and template independent (terminal transferase) (Ramadan, K. *et al*, 2003) activities, but lacks a 3'→5' exonuclease proofreading activity (García-Díaz, M. *et al*, 2002). Short gaps containing a 5' phosphate group are the preferred substrates of this enzyme in human cells and it may be able to function under low cellular levels of dNTPs (García-Díaz, M. *et al*, 2002). Similar to pol β, pol λ possesses dRPlyase activity, a function required in short-patch BER, and pol λ can coordinate the gap-filling DNA synthesis and dRP excision steps in a reconstituted BER reaction (García-Díaz, M. *et al*, 2001). This polymerase has been shown to contribute to mammalian BER as pol λ null MEF extracts show a deficiency in BER compared to the wild-type cells and lack of pol λ in Pol β-deficient cell extracts reduces *in vitro* BER (Braithwaite, E.K. *et al*, 2005a). Pol λ mediates repair of 5-hydroxymethyluracil *in vitro* and Pol λ deficient MEFs are hypersensitive to the incorporation of this lesion into genomic DNA. Pol λ localises to sites repairing oxidative DNA lesions and can interact with SMUG1 (Braithwaite, E.K. *et al*, 2005b). Pol λ also has the ability to perform strand displacement synthesis on gapped DNA which would be necessary for its involvement in long-patch BER (García-Díaz, M. *et al*, 2001). RPA can stimulate the polymerase activity of pol λ while suppressing its terminal transferase activity (Maga, G. *et al*, 2005) and can selectively prevent the generation of mismatches by pol λ *in vitro*, suggesting that pol λ could function in long-patch BER (Maga, G. *et al*, 2006). This polymerase can insert dCTP or dATP opposite 8-oxo-G (Brown, J.A. *et al*, 2007; Picher, A.J. & Blanco, L., 2007), but in the presence of PCNA and RPA it correctly incorporates dCTP more than a thousand times more efficiently than dATP (Maga, G. et al, 2007). Pol λ strongly discriminates against the extension of the mutagenic 8-oxo-G:A pair and readily extends the non-mutagenic 8-oxo-G:C pair (Picher, A.J. & Blanco, L., 2007). Pol λ may function in translesion synthesis (Maga, G. *et al*, 2002; Ramadan, K. *et al*, 2003; Crespan, E. *et al*, 2007) and non-homologous end joining (Bebenek, K. *et al*, 2003).

6.1.6.3 NUDT1 & NUDT5

NUDT1 (also known as *MTH1*) is located at chromosome 7p22, spans approximately 9 kb and is made up of five exons, the first of which is non-coding (Furuichi, M. *et al*, 1994; Oda, H. *et al*, 1997). The *NUDT1* gene is alternatively spliced to produce seven mRNA transcripts and is expressed in a range of tissues, with highest expression in the testis and thymus (Oda, H. *et al*, 1997). The *NUDT5* gene is found on chromosome 10p13-p14, consists of nine exons and spans approximately 18.5 kb (Gasmi, L. *et al*, 1999; Yang, H. *et al*, 2000). The major mRNA transcript is 1.4kb and expressed in a wide range of tissues, with highest expression in liver, kidney, pituitary, placenta and thymus. Minor transcripts of 5.0kb and 1.1kb have also been identified (Gasmi, L. *et al*, 1999).

NUDT1 and NUDT5 are members of the Nudix (Nucleoside diphosphate-linked moiety X) hydrolase family of pyrophosphohydrolases whose substrates often take the form of a nucleoside diphosphate linked to another moiety, X, resulting in a nucleoside monophosphate and X-P products. Examples of X include a sugar such as ribose or another phosphate group, as in (d)NTPs and their oxidized derivatives (McLennan, A.G., 2006). *NUDT1* is the mammalian homologue of *Escherichia coli mutT* and preferentially hydrolyzes 8-oxodGTP to 8-oxodGMP accordingly (Sakumi, K. *et al*, 1993). NUDT1-deficient mouse cells have a mutation frequency approximately double that of the wild-type cells and NUDT1-deficient mice develop a greater number of lung, stomach and liver tumours (Tsuzuki, T. *et al*, 2001). Human NUDT1 expression in *E. coli mutT* reduces the spontaneous A:T to C:G mutation frequency (Sakumi, K. *et al*, 1993). NUDT5 provides human cells with an additional mechanism for preventing incorporation of 8-oxoG into DNA from the nucleotide pool. NUDT5 specifically hydrolyzes 8-oxodGDP and can also cleave 8-oxodGTP, but only at very low levels (Ishibashi, T. *et al*, 2003). Direct oxidation of dGDP and cleavage of 8-oxodGTP by nucleoside triphosphatase (Mo, J.Y. *et al*, 1992) both result in 8-oxodGDP which is readily phosphorylated by nucleoside diphosphate kinase to generate 8-oxodGTP (Hayakawa, H. *et al*, 1995). It is therefore important for 8-oxodGDP to be hydrolyzed to the monophosphate as this is not a substrate for cellular guanylate kinase and therefore cannot be used for DNA synthesis (Hayakawa, H. *et al*, 1995). Expression of NUDT5 in *E.*

coli MutT deficient cells reduced the mutation rate to the wild-type level which may be due to the action of nucleoside triphosphatase followed by NUDT5 (Ishibashi, T. *et al*, 2003). NUDT5 also promotes the NUDT1 reaction by removing its inhibitor, 8-oxodGDP (Bialkowski, K. & Kasprzak, K.S., 1998).

Similar to 8-oxodGTP, 8-oxoGTP can be generated either by direct oxidation of GTP or phosphorylation of 8-oxoGDP by nucleoside diphosphate kinase and can be incorporated into mammalian cells by RNA polymerase II (Hayakawa, H. *et al*, 1999) but in contrast to DNA, once formed in the RNA it cannot be removed (Ishibashi, T. *et al*, 2005). NUDT1 and NUDT5 function to prevent transcriptional errors caused by misincorporation of 8-oxoG-containing ribonucleotides by specifically hydrolyzing 8-oxoGDP to 8-oxoGMP (Ishibashi, T. *et al*, 2005). The monophosphate is not a substrate for guanylate kinase and therefore cannot be reutilized (Hayakawa, H. *et al*, 1999). NUDT1 can also selectively cleave 8-oxoGTP to 8-oxoGMP. Expression of either enzyme in *E. coli mutT* deficient cells reduced the increase in erroneous protein production from 28-fold more than the wild-type, to less than 1.5-fold (Ishibashi, T. *et al*, 2005).

NUDT5 was originally identified as an ADP-sugar pyrophosphatase as it can hydrolyze ADP-ribose (and ADP-mannose to a lesser extent) (Gasmi, L. *et al*, 1999; Yang, H. *et al*, 2000). However, NUDT5 hydrolyzes 8-oxodGDP more efficiently than ADP-sugars (Ishibashi, T. *et al*, 2003).

6.1.7 Aims

Our group (Jones, S., Dallosso, A., Dolwani, S., personal communication) have previously screened *OGG1*, *NEIL1*, *NEIL2*, *NEIL3*, *NTHL1* and *NUDT1* in unrelated multiple colorectal adenoma cases and found no evidence to suggest that any of these genes possess high penetrance alleles at a significant frequency for this phenotype. The present work extended the study to other genes involved in BER or DNA damage protection; *TDG*, *MPG*, *MBD4*, *SMUG1*, *NUDT5*, *APE1* and *POLL*.

6.2 Methods

6.2.1 Clinical samples

Blood DNA samples from one case from USA and fifty-seven unrelated index cases from regional polyposis registers in Cardiff, Manchester, Birmingham, Liverpool, Southampton and Cambridge were analysed. All patients had between 10 and 100 adenomas recorded at colonoscopy or colectomy, except for two who had hundreds of adenomas at colectomy. Eleven cases were comprehensively sequenced for germline mutations in the *APC* and *MUTYH* ORFs with normal results and six of these also underwent *APC* MLPA with normal results. There was no evidence of a germline *APC* mutation or biallelic *MUTYH* mutations in the remaining forty-seven cases following testing in NHS molecular diagnostic laboratories in the referring centres. *OGG1*, *NEIL1*, *NEIL2*, *NEIL3*, *NTHL1* and *NUDT1* had been screened in all fifty-eight cases and no clearly pathogenic mutations had been identified. In general, only blood DNA samples from the index cases were available and progression to family studies and somatic mutation analysis in tumours would only have been undertaken were there very strong leads that high penetrance alleles had been identified.

6.2.2 Screening for mutations in TDG, MPG, MBD4, SMUG1, NUDT5, APE1 and *POLL*

Primers were designed to PCR amplify the coding regions of *TDG*, *MPG*, *MBD4* and *SMUG1* as 9, 5, 10 and 3 fragments, respectively (Jones, S., Table 6.1). A. Dallosso designed PCR primers to amplify the coding region of *APE1* as 4 fragments (Table 6.1). I designed primers to PCR amplify the coding regions of *NUDT5* and *POLL* as 8 and 10 fragments, respectively (Table 6.1). I carried out PCR optimization for all genes except *APE1* which was undertaken by A. Dallosso (Table 6.1). The coding regions of all genes were PCR amplified and screened for mutations in the 58 cases using automated sequencing. A. Dallosso undertook screening of *TDG*, *MPG* and *APE1*, the Wales Gene Park Genomic Facility screened *SMUG1* and I screened *MBD4*, *NUDT5* and *POLL* in the 58 cases. Division of work is shown in Figure I, appendix, with my contribution highlighted in shaded boxes.

Gene/Primer name	Primer sequence (5'→3')	Annealing temperature (°C)	Product size (bp)
TDG			
TDG_Ex1F_NEW	CAGCCACTGTCTGGGTACTG	62	358
TDG_Ex1R_NEW	AGAGCAGCCCCGACCTC		
TDG_Ex2F	CTCTCCTCTGTAATCCACTCTA	59	329
TDG_Ex2R	ATCCGATGTTGAACTTTCTAA		
TDG_Ex3F	AGCTGCTAAAGTTTCTAAGTTAA	59	372
TDG_Ex3R	CAAGGACAACTGTTAAGTAAAG		
TDG_Ex4F	TCCACCACTCCTCCATAGAA	59	360
TDG_Ex4R	ACATCCCTCCATTCTCATAGAC		
TDG_Ex5F	GATCGTGCCACTACACTCTA	62	317
TDG_Ex5R	AGCTCAGCTTGAACTAGATACA		
TDG_Ex6_7F	GCTGTCTGAATTTAGCATATTATA	56	409
TDG_Ex6_7R	TCACAATGGATAGGACAAATAA		
TDG_Ex8F	ACAAATATTCTAATCTCAATGAGT	56	293
TDG_Ex8R	TATACACACACAAAATGAATAAA		
TDG_Ex9F	CGGTTTTTACAGTTCTTATG	53	406
TDG_Ex9R	ATTCCCATTCTTCAATAATTT		
TDG_Ex10F	CTGCAAAGAGCTGTGATCAT	53	343
TDG_Ex10R	AGCAAACTGAGGTTCTACTTGT		
MPG			
MPG_Ex1F	TCGAGTGTGTCAGGGTGTT	58	194
MPG_Ex1R	CGTCGGCAAAACTGTAATG		
MPG_Ex2F	CCTATTCGGATGCTTATTTA	55*	378
MPG_Ex2R	GGGTTCAGGGACAACTG		
MPG_Ex3F	GGGCACTGTTAGGGTGAG	60	357
MPG_Ex3R	CCACCTCAGTCCTCCTAG		
MPG_Ex4F1new	GCTCCACTTCCAAACTGTC	55	242
MPG_Ex4R1new	CCAGCCATACAGCTTCATC		
MPG_Ex4F2_NEW	CCAGCCGTGTCCTCAAG	60	311
MPG_Ex4R2_NEW	AAAATCTTGTCTGGGCAGG		
SMUG1			
SMUG_Ex1F	TGGATCCCTCCTACTCTG	56	420
SMUG_Ex1R	CCAAGCATCCACCTAGAA		
SMUG_Ex2F1_NEW1	GGCCTCAGGTCTCCAGTT	56	314
SMUG_Ex2R1_NEW1	GGCAGCTCAGCAGGAGTA		
SMUG_Ex2F2	GAGGTCTTCTTCCATCACTGT	58	400
SMUG_Ex2R2	CTTCGAGGTCTTGAATGTGTC		
APE1			
FEX1_2s.APE	GAGGCTAAGCGTCTCCGTCAC	58	720
REX1_2s.APE	GGTTTGTGAAGAAGTCGCAG		
FEX3s.APE	GTGTGAAGAAGTCGCAGGAAC	58	383

REX3a.APE	TTCTCCGTTTAGCCTTCAGG		
FEX4s.APE	GAATATTGTGCTGCTTGACTC	58	420
REX4a.APE	CACCTCTTGATTGCTTTCCC		
FEX5s.APE	TTGCTAATTCTCTATCTCTG	58	673
REX5a.APE	CCATTCCTTCTTTAAACACTC		
FEX3Rseq	AGAAACCTTCACGGGGTGGAG		
[illegible]	[illegible]	[illegible]	[illegible]
MBD4_Ex1F	CTTTCGCAACATTCAGACCTC	58	297
MBD4_Ex1R	ACTGTCCACTCTCCCGATACC		
MBD4_Ex2F	TGAGTAGGCAGTGGAAGATAA	58	374
MBD4_Ex2R	AAGCTAAGATTCCTGCTATGC		
MBD4_Ex3F1new	AATGTGGTCCAGTTCTTTTAA	58	399
MBD4_Ex3R1new	ATCAACACCCTCATCTTCTTT		
MBD4_Ex3F2	TTCAAACTGGAACCTCAGGAC	58	361
MBD4_Ex3R2	AGCATCAGAAATGCAGACAGT		
MBD4_Ex3F3	GATGCTGAAAGTGAACCTGTT	58	496
MBD4_Ex3R3	TCTTGGCTCTATTTTCACATC		
MBD4_Ex4F	ATTATTTGCATCCCTCAATAT	55	214
MBD4_Ex4R	ATAGTGCATCAGAATTGAAAA		
MBD4_Ex5F	AATCAGAACAGCAAATTCTAA	58	298
MBD4_Ex5R	TGACACACTCAAAATGGACT		
MBD4_Ex6F	CCACCTGGAGTCTTGTAATCA	58	225
MBD4_Ex6R	TATGTTTTTCCTTTGGGTGTA		
MBD4_Ex7F	ATTTTGGGAGGGTGTCTTTAG	60*	205
MBD4_Ex7R	CAGAGACCAAATGTGCTGAAT		
MBD4_Ex8F	CGTCTCTGCCTCTGTATCTTA	58	262
MBD4_Ex8R	CATTGGCACACACATTAAGA		
[illegible]	[illegible]	[illegible]	[illegible]
NUDT5x1F	GCGGCAGTTCCCTCACATAC	58	389
NUDT5x1R	CGAGCCCAAAAAAGGGAGTA		
NUDT5x2F	AAACGTGCCCACAGATTATT	58	389
NUDT5x2R	CCCAATTTTATTCAGCAACTAT		
NUDT5x3F	GGCAAGAGTTTTCGTTGTTA	58	527
NUDT5x3R	AGGTGTGGCTTCAAACTATAT		
NUDT5x4(1)F	CTGTGGAAGGTCAGTGTAGTC	58	557
NUDT5x4(1)R	CTCGCTACGAAATGGTTTAG		
NUDT5x5F	GATTTCCCGCTCCCATCAC	58	313
NUDT5x5R	AGGCCCTTCTGGCTCCAG		
NUDT5x6F	CTACCAAATGTGAGAACTAGT	58	500
NUDT5x6R	GTGGGAATACACTTCATATAG		
NUDT5x7/8F	GCTCGAGTTTGACAATGTAT	58	465
NUDT5x7/8R	CTAGGCATTTGACTTTAGTGA		
NUDT5x9F	AGCATAGGAAGTGACATATA	58	319

NUDT5x9R	AGCTAATGGCAAATCTAC		
POLL			
POLLx1F	CTACCCCCAAAGCCTGGTCAG	58	365
POLLx1R	AGGCCCTGGACAGGCAGAGT		
POLLx2(1)F	TTTCAGGGTAGGGGACTGT	61	567
POLLx2(1)R	CTCCCACCTAAACCTCTTATAG		
POLLx3F	GCCTTGATCGTACCACTGTAC	58	492
POLLx3R	AGAGCTAAATGGCTTCACAATA		
POLLx4.1F	GGCCTCACACCCAAGGAGA	58	445
POLLx4.1R	GCATTGATGGCCTTGGCATAG		
POLLx4.2F	CCAGAAGGCGACCAATCAC	58	329
POLLx4.2R	CTCCCAGCTTCAACAACTATCAA		
POLLx5F	TTACCCAGCCCTCATTCTATC	58	471
POLLx5R	CCCATCAGAGCACAGCATAG		
POLLx6F	TTCCCAAGTCCTGCTGAGTAC	58	456
POLLx6R	GGCCTGGAGCTTCAGTCTTA		
POLLx7F	GCTTGCCTCCTGCACAGT	58	489
POLLx7R	TGCCTCAGGACTGGAACTTC		
POLLx8.1F	AAGGGCCCAGAGAGGGTAGT	58	474
POLLx8.1R	AGGCCTAAGAGCCTGAAGACAT		
POLLx8.2F	GCCAAAACCAAGGGCATGAGTCT	58	305
POLLx8.2R	GGCCCTGCTCGCTGAGGAA		

Table 6.1 Sequences of primers and PCR conditions used to screen *TDG*, *MPG*, *MBD4*, *SMUG1*, *NUDT5*, *APE1* and *POLL*. * 4% DMSO added to PCR reaction.

6.2.3 Assays for sequence variants in controls

Specific assays were designed to determine the presence and frequency of previously unreported rare missense variants (identified in only one or two cases in the present study) in at least 700 control chromosomes (Table 6.2).

Gene/Variant	Primer name	Primer sequence (5'→3')	Anneal. temp (°C)	Product size (bp)	Assay
TDG					
Q391K (c.1171 C→A)	TDG_Q391K_CF	GACCCAGTCATTTACAGTCC	58 (C) 54 (A)	156	ARMS
	TDG_Q391K_AF	GACCCAGTCATTTACAGGCA			
	TDG_Q391K_R	AGCAAACTGAGGTTCTACTTGT			
I134V (c.400 A→G)	TDG_Ex3F	AGCTGCTAAAGTTTCTAAGTTAA	59	372	*HpyCH4IV* digest Mut allele digested (291bp + 81bp)
	TDG_Ex3R	CAAGGACAACTGTTAAGTAAAG			
MPG					
E240D (c.240 A→C)	MPG_E240D_F	CCAGCCGTGTCCTCAAGGA	58	126	ARMS
	MPG_E240D_TR	CACGCTCCAGCCATACAACT			
	MPG_E240D_GR	CACGCTCCAGCCATACACCG			
MBD4					
N467S (c.1400 A→G)	MBD4_N467S_F	CCACGACGTAAAGCCTTTAAG	58	133	ARMS
	MBD4_N467S_AR	TCCCCAAACCTGAGGTCCAAT			
	MBD4_N467S_GR	TCCCCAAACCTGAGGTCCCAC			
NUDT5					
G165A (c.910 G→C)	NUDT5_G165A_F	GCTCGAGTTTGACAATGTAT	54 (G) 56 (C)	177	ARMS
	NUDT5_G165A_GR	GGAGTTGCAGAACATACCTC			
	NUDT5_G165A_CR	GGAGTTGCAGAACATACCTG			
POLL					
P303S (c.907 C→T)	POLLx5F	TTACCCAGCCCTCATTCTATC	58	471	*BsaJI* digest Mut allele digested (372bp + 99bp) WT allele digested (206bp + 165bp + 99bp + 1bp)
	POLLx5R	CCCATCAGAGCACAGCATAG			

Table 6.2 Assays for specific missense variants in *TDG*, *MPG*, *MBD4*, *NUDT5* and *POLL*. Internal control primers were used to validate the ARMS assays; TSC2_Ex26F (5'-GAGCTTTGGCCCTTGGTGATA) and TSC2_Ex26R (5'-CTCGCCCACAGGAGACCTAGA), product size 388bp.

6.3 Results

6.3.1 Coding variants identified in TDG

Four missense variants were identified in the *TDG* gene; I134V, G199S, V367A and Q391K (Table 6.3). No cases were found to have two variants in the *TDG* gene. V367M was found in a heterozygous state in nine cases and in a homozygous state in one case, all with between 11 and 100 colorectal adenomas. This variant is a semi-conservative amino acid substitution and valine is not found at this residue in other species.

The three rare missense variants, G199S, I134V and Q391K, were identified in the heterozygous state only. G199S was found in one patient (CB1) with between 11 and 100 colorectal adenomas and is a conservative amino acid change at a site partially conserved through evolution. I134V was detected in one case (1137) who presented in the third decade, had 14 colorectal adenomas without CRC and no family history of CRC. This case had previously been screened (and was negative) for germline mutations in the ORFs of *APC* and *MUTYH* (see chapter 3). This variant is a conservative amino acid substitution at a site partially conserved through evolution but was not found in over 700 control chromosomes. Tumour samples were not available to screen for somatic inactivation of the wild-type *TDG* allele and DNA from the previous generation was not available to determine whether I134V was a *de novo* substitution.

Q391K was identified in a single case (M2) with at least 30 colorectal adenomas without CRC. This case has a family history of CRC; two of five siblings were diagnosed aged 48 and 39 years and their mother died from CRC aged 82. This variant is an uncharged to a charged amino acid substitution, although both amino acids are polar. This site is partially conserved through evolution and the change was not detected in over 700 control chromosomes. DNA samples were not available to determine if this variant segregated with the disease phenotype. Tumour samples were also not available to screen for somatic inactivation of the wild-type *TDG* allele.

Gene / Exon	AA change	Nucleotide Change	Conservation		PolyPhen[2]	MAF in cases[3]	Case no.	NCBI dbSNP (MAF)[4]	MAF in controls[5]
			AA	Evolution[1]					
TDG									
3	I134V	c.400 A→T	C	Partial	Benign	1/116 (0.9%)	1137	/	0/724
5	G199S	c.595 G→A	C	Partial	Possibly damaging	1/116 (0.9%)	CB1	Rs 4135113 (18.5%)	ND
10	V367M	c.1099 G→A	SC	No	Benign	11/116 (9.5%)	B1, B2, *C1*, C2, CB2, M1, SO1, SO2, SY1, 4822	Rs 2888805 (22.1%)	ND
10	Q391K	c.1171 C→A	SC	Partial	Possibly damaging	1/116 (0.9%)	M2	/	0/722
MPG									
4	E240D	c.720 A→C	C	No	Benign	1/116 (0.9%)	B3	/	0/716
APE1									
3	Q51H	c.153 G→C	SC	Partial	Benign	2/116 (1.7%)	1137, 1187	Rs 1048945 (3.3%)	ND
5	D148E	c.444 T→G	C	No	Benign	56/116 (48.3%)	*B1*, B2, *B3*, B5, B6, *B7*, C1, C2, C3, C4, CB1, CB3, CB5, CB6, CB7, B4, *L1*, M1, M2, *M3*, M4, M5, M6, M7, M8, *M9*, SO1, SO2, SO3, SO6, *S07*, *SO8*, SO9, SO10, SY1, SY2, 0322, *1187*, 3432, *3607*, 3858, 4093, *4822*, 8163	Rs 1130409 (48.6%)	ND
MBD4									
3	A273T	c.817 G→A	SC	Partial	Benign	12/116 (10.3%)	C4, CB1, M1, *M4*, S04, SO5, SY1, 2302 4093, 4822, 8163	Rs 10342 (22.9%)	ND

3	I358T	c.1073 T→C	SC	Partial	Benign	1/116 (0.9%)	CB3	Rs 2307298 (0.6%)	ND
5	N467S	c.1400 A→G	SC	Partial	Benign	2/116 (1.7%), 1/218 (0.5%)*	M3, SO3	/	0/712, 1/502 (0.2%)*
8	D568H	c.1702 G→C	SC	Partial	Benign	1/116 (0.9%)	C3	Rs 2307293 (0.6%)	ND
NUDT5									
7	G165A	c.910 G→C	C	Partial	Possibly damaging	1/116 (0.9%)	M3	/	0/706
POLL									
4	T221P	c.661 A→C	SC	Partial	Benign	12/116 (10.3%)	B2, B4, C2, C5, M3, M5, M6, SO1, SO6, SY1, 1137, 4093	Rs 3730463 (11.0%)	ND
5	P303S	c.907 C→T	SC	Yes	Probably damaging	1/116 (0.9%)	5161	/	0/722
7	R438W	c.1312 C→T	SC	Partial	Benign	24/116 (20.7%)	B5, *C6*, *CB2*, *CB4*, CB5, M2, M3, M7, SO1, SO2, SO7, *SO8*, 0322, 1149, 2302, *3432*, 3607, 4147, 5161	Rs 3730477 (17.1%)	ND

Table 6.3 Coding variants identified in *TDG*, *MPG*, *APE1*, *MBD4*, *NUDT5* and *POLL*. [1]Alignments of human (NP_003202; NP_002425; NP_001632; NP_003916; NP_054861; NP_037406), mouse (NP_766140; NP_034952; NP_033817; NP_034904; NP_058614; NP_064416), *Xenopus* (NP_001084290; NP_001017178; NP_001086779; NP_001037916; NP_001016613; NP_001093716), *Drosophila* (NP_651925; NP_476841, TDG and APE1 only), *Arabidopsis* (NP_181677; NP_191862; NP_178524; NP_172522, not TDG or MPG), *S. cerevisiae* (EDN64598; NP_009669, APE1 and NUDT5 only), *S. pombe* (NP_588515; NP_592977, TDG and Pol λ only), *C. noyvi* (YP_877811, MPG only), *M. thermautotrophicus* (P29588, MBD4 only), *P. stutzeri* (YP_001174562, MBD4 only), *P. putida* (NP_746917, MPG only), *E. coli*

(NP_417540; NP_754044; NP_417856, TDG, APE1 and NUDT5 only) homologues of TDG, MPG, APE1, MBD4, NUDT5 and Pol λ were carried out using ClustalW. Residues were determined to be conserved, partially conserved or non-conserved through evolution based on the number of species with similar amino acids. [2]PolyPhen (http://genetics.bwh.harvard.edu/pph/) prediction of effect variant may have on protein. [3]The minor allele frequency (MAF) of each variant in multiple colorectal adenoma patients is shown. Case numbers in normal and italic font indicate heterozygotes and homozygotes, respectively. [4]NCBI dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/) reference numbers and MAF are indicated. [5]Rare variants (MAF<2% in cases) with no entry in dbSNP were also typed in control samples. *Dutch multiple adenoma cases and controls, AA; amino acid, C; conservative substitution, SC; semi-conservative substitution, NC; non-conservative substitution, ND; not determined.*Dutch cases and controls.

6.3.2 Coding variants identified in MPG

One missense variant was identified in the MPG gene, E240D (Table 6.3). This rare variant was identified in a heterozygous state in a single case (B3) with 28 colorectal adenomas and CRC aged 48 but no family history of bowel problems. E240D is a conservative amino acid substitution at a site which is not conserved through evolution but was not detected in more than 700 control chromosomes. No second germline variant in the *MPG* gene was identified in this case. Tumour samples were not available to screen for somatic inactivation of the wild-type *MPG* allele and DNA from the previous generation was not available to determine whether E240D was a *de novo* substitution.

6.3.3 Coding variants identified in APE1

Two missense variants were identified in *APE1*, Q51H and D148E (Table 6.3). The MAF of the D148E variant was nearly 50%, with 33 heterozygotes and 4 homozygotes identified. The phenotype of heterozygous cases ranged from between 11 and 100 to hundreds of colorectal adenomas and all homozygotes had between 11 and 100 adenomas. This variant is a conservative amino acid substitution at a site which is not evolutionarily conserved. Q51H is an uncharged to a charged amino acid substitution, although both amino acids are polar, and this site is partially conserved through evolution. Q51H was detected in the heterozygous state in two cases; one with 50 to 100 colorectal adenomas and CRC (1187) and the other had 14 colorectal adenomas in the absence of CRC (1137). Neither of these cases had a family history of CRC and both had been comprehensively sequenced for germline mutations in the *APC* and *MUTYH* ORFs with normal results (chapter 3). *APC* MLPA failed for case 1137 but a normal result was obtained for patient 1187 (chapter 3). No cases were found to have two rare variants in the *APE1* gene.

6.3.4 Coding variants identified in MBD4

Four missense variants were identified in *MBD4*; A273T, I358T, N467S and D568H (Table 6.3). A273T was found in the heterozygous state in ten cases, eight of whom had between 11 and 100 colorectal adenomas and two had hundreds of colorectal adenomas and CRC. One homozygote case (M4) with 50 colorectal adenomas was identified. This variant is a semi-conservative

amino acid substitution at a residue which is partially conserved through evolution. One case heterozygous for I358T was identified (CB3) and had between 11 and 100 colorectal adenomas. This variant is a non-polar to a polar amino acid substitution at a site which is partially conserved through evolution. D568H was also detected in the heterozygous state in a single case (C3) with between 11 and 100 colorectal adenomas. This variant is a negatively-charged to a positively-charged amino acid substitution, although both amino acids are polar, at a residue that is partially conserved though evolution.

Two N467S heterozygotes were identified, one (M3) had at least 70 colorectal adenomas with CRC and had previously been screened (and found to be negative) for germline mutations in *APC* exons 3, 5-15 by PTT and SSCP. This case had a family history of CRC, with all three siblings diagnosed in their forties or fifties and their father diagnosed aged 72. There was no evidence that the father had colorectal adenomas but at least one of the siblings had a colorectal polyp in addition to CRC. The other patient (SO3) had between 11 and 100 colorectal adenomas. Both cases were screened for germline mutations in *APC* exons 4 and 9 and the *MUTYH* ORF by sequencing. No mutations were detected. N467S is a semi-conservative amino acid substitution at a site which is partially conserved through evolution. This variant was not identified in over 700 control chromosomes so a cohort of Dutch multiple colorectal adenoma patients and Dutch controls (courtesy of Dr. F.J. Hes) was screened (Table 6.3). One out of 109 cases and one of the 251 control samples were found to be heterozygous for this variant. No cases were found to have two rare variants in the *MBD4* gene and tumour samples were not available to screen for somatic inactivation of the wild-type *MBD4* allele.

6.3.5 Coding variants identified in NUDT5

One missense variant was identified in the *NUDT5* gene, G165A (Table 6.3, Fig 6.1). This variant was identified in a heterozygous state in patient M3 (see above for phenotype). G165A is a conservative amino acid substitution at a site which is partially conserved through evolution but was not found in over 700 control chromosomes. No second germline variant in the *NUDT5* gene was identified in this case and tumour samples were not available to screen for

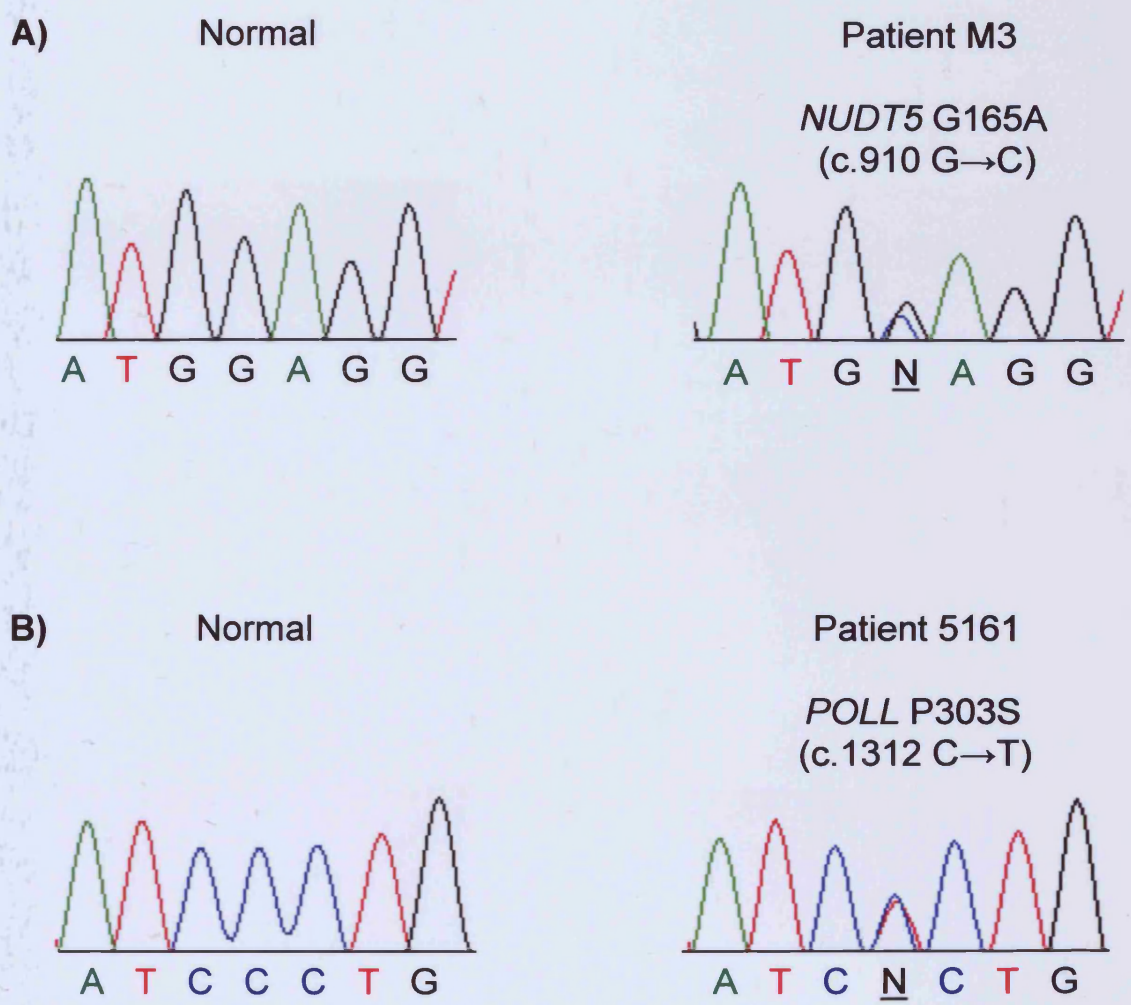


Figure 6.1 Rare missense variants in the *NUDT5* and *POLL* genes in patients with colorectal adenomas. Patient M3 had at least 70 colorectal adenomas and patient 5161 had multiple colorectal adenomas. Both variants are novel changes that were not found in controls.

somatic inactivation of the wild-type allele.

6.3.6 Coding variants identified in POLL

Three missense variants were identified in *POLL*; T221P, P303S and R438W (Table 6.3). Twelve cases were found to be heterozygous for T221P and all but one had between 11 and 100 colorectal adenomas, with a single case having hundreds of colorectal adenomas. This variant is a polar to a non-polar, cyclic amino acid substitution at a residue partially conserved though evolution. R438W was found in the heterozygous state in fourteen cases, of which thirteen had between 11 and 100 colorectal adenomas and one had hundreds of colorectal adenomas with CRC. Five cases with between 11 and 100 colorectal adenomas were identified as R428W homozygotes. This variant is a semi-conservative amino acid substitution at a residue which is partially conserved through evolution.

P303S was identified in a single heterozygous case (5161) (Fig 6.1) with multiple polyps aged 31 years without a family history of CRC or colorectal polyps and both parents had normal colonoscopies in their sixties. This patient had been previously screened (and found to be negative) for germline mutations in *APC* by PTT, SSCP, MLPA and sequencing of the entire *APC* coding region. Comprehensive sequencing of the *MUTYH* ORF was undertaken (see chapter 4) and patient 5161 was found to carry only a monoallelic G393D mutation. The *POLL* P303S variant is a non-polar, cyclic to a polar amino acid substitution at a residue which is conserved through evolution. It was not detected in over 700 control chromosomes. No second rare germline variant in the *POLL* gene was identified in this case and the P303S variant was found to be inherited from his unaffected father. Tumour samples were not available to screen for somatic inactivation of the wild-type *POLL* allele.

6.4 Discussion

6.4.1 Mutations in the TDG gene

The missense variants V367M and G199S have been previously reported as polymorphisms (rs2888805; rs4135113) but were identified at a lower frequency in the present study than in reported normal population samples. This may be due to variation in allele frequencies between different populations. V367M is a common variant that lies outside of the glycosylase domain (Gallinari, P. & Jiricny, J., 1996) within the less conserved C-terminal region of TDG. It is not predicted to have a deleterious effect on the protein, has a MAF of approximately 20% in dbSNP and has recently been reported in familial CRC cases by Broderick *et al* who identified more than ten heterozygotes in a cohort of 94 cases and concluded that V367M is a polymorphism that does not contribute significantly to CRC susceptibility (Broderick, P. *et al*, 2006).

G199S is located in the catalytic core domain of TDG (Baba, D. *et al*, 2005) but is not one of the critical residues of the active site or necessary for substrate recognition (Hardeland, U. et al, 2000). It is predicted to have a possibly damaging effect on the protein but the substitution of serine for glycine is conservative as both are uncharged and small. Despite being classed as non-polar, the tiny glycine side chain does not really contribute to hydrophobic interactions (Nelson, D.L. & Cox, M.M., 2000) so this substitution is unlikely to disrupt crucial interactions. G199S has been reported in unaffected individuals at a MAF twenty times greater than that observed in cases in the present study (18.5% vs 0.9%). Such a large difference in MAF suggests that this allele could have a protective effect, decreasing CRC predisposition as has been described for homozygous carriers of the *MTHFR* variant A677V (Houlston, R.S. & Tomlinson, I.P.M., 2001). A serine side chain at codon 199 of TDG could contribute more to hydrophobic interactions within the catalytic core domain than the glycine side chain which could possibly improve substrate recognition and/or catalysis.

Two novel coding variants were identified in *TDG*, Q391K and I134V. I134V is located in the catalytic core domain of TDG (Baba, D. *et al*, 2005) but is not itself one of the residues crucial for catalysis or substrate recognition

(Hardeland, U. et al, 2000). It forms part of a β-sheet structure in the core domain (Baba, D. *et al*, 2005) which is unlikely to be affected by the substitution as both amino acids are non-polar and aliphatic. Despite being part of the sequence motifs that mediate SUMO binding (V/I-X-V/I-V/I) (Mohan, R.D. *et al*, 2007) and oestrogen receptor α interaction (LDIVI) (Chen, D. *et al*, 2003), the amino acid substitution leaves them functionally unaffected, consistent with the prediction of a benign effect on the protein. Although I134V was not identified in control samples, there is little evidence that this change is pathogenic.

Q391K is found in a region of the C-terminal domain not required for glycosylase activity but lies within an area (residues 273-421) needed for interaction with the CH3 domain of transcriptional coactivator CBP (Tini, M. *et al*, 2002). This variant was predicted to be possibly damaging and was not detected in control chromosomes. Both amino acids are hydrophilic but whereas glutamine is uncharged, lysine is positively-charged which could affect electrostatic interactions. Further work is required to determine whether this mutation has a pathogenic effect and any influence on colorectal adenoma predisposition.

6.4.2 Mutations in the MPG gene

The only missense change identified in *MPG* was the novel variant E240D but there is little evidence this is pathogenic. E240D is not one of those residues critical for glycosylase activity (Lau, A.Y. *et al*, 1998) but is found within a region (residues 233-294) required for interaction with the transcriptional repressor MBD1 (Watanabe, S. *et al*, 2003). E240D was not detected in control samples but both amino acids are negatively-charged so the substitution is predicted to have little effect on the protein.

6.4.3 Mutations in the APE1 gene

The two non-synonymous variants detected in the *APE1* gene, Q51H and D148E, have both been identified previously in normal populations at frequencies similar to those reported in this study (Hadi, M.Z. et al, 2000; rs1048945; rs1130409). D148E is a common polymorphism located within the nuclease domain but has no impact on endonuclease or DNA binding activities

of the protein (Hadi, M.Z. et al, 2000), as predicted. Q51H is found outside of the nuclease domain, within the Ref domain of APE1 (Hadi, M.Z. et al, 2000) and is predicted to have little effect on the protein. There is therefore no evidence to suggest that either of these variants represent pathogenic mutations.

6.4.4 Mutations in the MBD4 gene

The *MBD4* missense variant A273T has been previously reported as a polymorphism found at similar (Bader, S. *et al*, 1999) or higher frequencies (rs10342) than in the present study. This discrepancy may be due to variation in allele frequencies between different populations. There is no evidence to suggest that this variant represents a pathogenic mutation as it is predicted to have little effect on the structure or function of the protein.

Three rare variants were detected in the *MBD4* gene in the present study. I358T and D568H have been identified previously (rs2307298; rs2307293) and are predicted to have little effect on the MBD4 protein. The I358T residue is not required for glycosylase activity and lies outside the methyl-binding domain (Wu, P. *et al*, 2003). Despite being located within the glycosylase domain in a region that forms the top of the cleft, D568H is predicted to have a benign effect on the protein as both amino acids are hydrophilic and it is the hydrophobic residues within this region that are important for cleft formation (Wu, P. *et al*, 2003). N467S is a novel rare variant that lies within a loop in the glycosylase domain that is important for DNA-binding and base flipping (Wu, P. *et al*, 2003). It was not identified in the initial set of control chromosomes but after analyzing a further set of Dutch multiple colorectal adenoma cases and controls, it was identified in a single control sample. Both amino acids are polar and uncharged so this variant is predicted to have little effect on MBD4. It is therefore unlikely that these three variants contribute significantly to multiple colorectal adenoma predisposition.

6.4.5 Mutations in the NUDT5 gene

The only missense change identified in *NUDT5* was the novel variant G165A. This residue is located within the catalytic Nudix domain but is not found in the

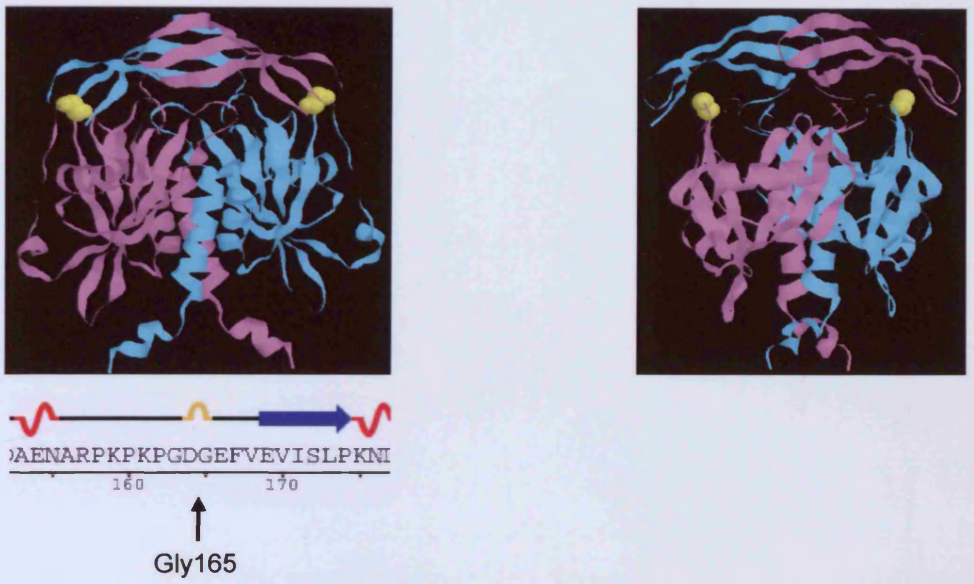
Nudix motif itself (Zha, M. *et al*, 2006). G165A is predicted to be possibly damaging, despite both amino acids being non-polar and aliphatic. NUDT5 forms homodimers and this variant forms part of a loop which is involved in stabilizing interactions between the two subunits and is important for dimerization (Fig 6.2) (Zha, M. *et al*, 2006). G165A was not detected in control samples so further work is required to determine whether this mutation has a pathogenic effect and any influence on colorectal adenoma predisposition.

6.4.6 Mutations in the POLL gene

The non-synonymous variants T221P and R438W found in *POLL* have previously been reported as polymorphisms (rs3730463; rs3730477). The R438W variant was identified at a higher frequency in the present study which may to be due to variation in allele frequencies between different populations. T221P is located in the serine-proline-rich domain which is dispensable for enzymatic activity (Delarue, M. *et al*, 2002). This variant is not predicted to have an effect on the protein as both amino acids are polar and uncharged and the substituted proline lies within a proline-rich domain. The common variant R438W is found in the palm subdomain of the Pol λ polymerization domain but is not one of the critical catalytic residues (Garcia-Diaz, M. *et al*, 2004) and is not predicted to have a deleterious effect on the protein. Therefore there is little evidence that T221P or R438W are pathogenic changes.

The novel, rare *POLL* missense change P303S is located in the 8kDa domain of Pol λ and forms part of a helix-hairpin-helix (HhH) motif which binds downstream primer DNA and contains the catalytic residue for dRP lyase activity (Fig 6.2) (Garcia-Diaz, M. *et al*, 2004). Proline reduces the structural flexibility of polypeptide regions containing this amino acid (Nelson, D.L. & Cox, M.M., 2000) and is therefore important in maintaining the HhH motif structure. Substitution of serine for this proline is likely to disrupt the protein structure and is predicted to have a deleterious effect on Pol λ. P303S was not detected in control samples but was identified in the patient's unaffected father so further work is required to determine whether this mutation has an incompletely penetrant pathogenic effect and any influence on colorectal adenoma predisposition. This patient also harbours a monoallelic germline *MUTYH*

A)



B)

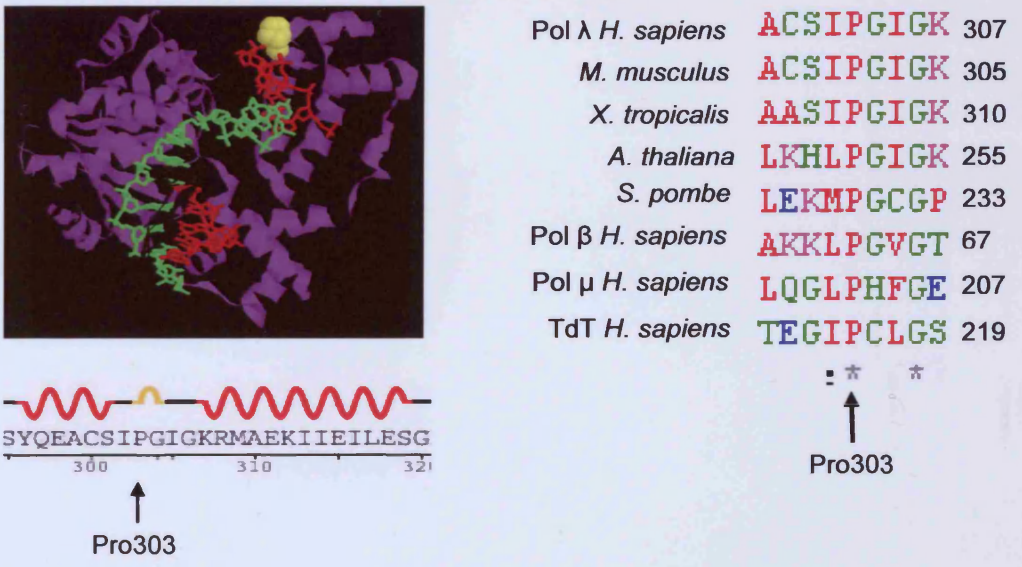


Figure 6.2 (A) Crystal structure and structural features of the human NUDT5 protein. Two NUDT5 monomers (blue and pink ribbons) form a homodimer that is stabilized by interactions between the subunits at an interface which includes residue 165 (shown as yellow spacefill). This residue lies within a hydrogen-bonded turn (yellow loop) as indicated by the arrow (below) (RCSB Protein Data Bank www.rcsb.org/pdb/). **(B) Crystal structure, structural features and evolutionary conservation of the human Pol λ protein.** The Pol λ 8kDa domain (purple ribbon) bound to a two nucleotide gap. The template strand and primers are shown in green and red sticks respectively. This domain interacts with the downstream primer in a

region which includes residue 303 (shown as yellow spacefill). This residue lies within a hydrogen-bonded turn (yellow loop) of a HhH motif, as indicated by the arrow (below) (RCSB Protein Data Bank www.rcsb.org/pdb/). Alignment of human (NP_037406), mouse (NP_064416), *Xenopus* (NP_001093716), *Arabidopsis* (NP_172522) and yeast (NP_592977) Pol λ homologues, along with human Polβ, Polμ and TdT (terminal deoxynucleotidyltransferase), was carried out using ClustalW (right). Amino acids are coloured according to their biochemical properties. The highly conserved proline at residue 303 is indicated by the arrow. Chains A and B of PDB file 2dsb (NUDT5) and chain A of PDB file 1rzt (Pol λ) were used to generate the figures using RasMol v2.6-β-2 (http://openrasmol.org/).

mutation (G393D) which was inherited from his unaffected mother. Neither variant may significantly predispose to colorectal adenomas alone (chapter 5), but in combination may affect BER sufficiently to produce a phenotypic consequence. Rare germline missense variants in *NUDT1* (R31Q) and *OGG1* (R197W) have been found individually in combination with a heterozygous *MUTYH* mutation (either Y176C or G393D) in multiple polyposis cases (Farrington, S.M. *et al*, 2005). Alternatively, the presence of a heterozygous G393D mutation in combination with P303S may just be coincidental as this *MUTYH* mutant allele is found at a MAF of 1% (G393D) in healthy Caucasians (Croitoru, M.E. *et al*, 2004; Farrington, S.M. *et al*, 2005; Webb, E.L. *et al*, 2006).

6.4.7 No evidence for significant contribution of highly penetrant germline mutations in the TDG, SMUG1, MPG, APE1, MBD4, NUDT5 or POLL genes to colorectal adenoma predisposition

In contrast to *MUTYH*, no clearly pathogenic mutations were identified in the BER and DNA damage protection genes analyzed in the present study. Similar negative results have been reported after screening *NTHL1*, *NEIL1*, *NEIL2*, *NEIL3*, *NUDT1*, *OGG1*, *UNG*, *MPG*, *TDG* and *SMUG1* in patients with multiple colorectal adenomas (Al-Tassan, N. *et al*, 2002; Sieber, O.M. *et al*, 2003; Broderick, P. *et al*, 2006; S. Jones, A. Dallosso and S. Dolwani, personal communication). There are several possibilities which may explain these results.

6.4.7.1 Functional redundancy

Protection against DNA damage is crucial for maintaining genomic integrity so organisms have evolved a system with a high degree of functional redundancy. DNA glycosylases have partially overlapping substrate specificities and can, to a certain extent, functionally compensate for each other *in vivo*. Mice deficient in OGG1, UNG or MBD4 do not display phenotypic abnormalities and exhibit only a small (1.4- to 3-fold) increase in mutation rate (Minowa, O. *et al*, 2000; Nilsen, H. *et al*, 2000; Miller, C.B. *et al*, 2002; Wong, E. *et al*, 2002). *Nthl1* knock-out mice retain glycosylase activity and are phenotypically normal (Takao, M. *et al*, 2002), as are $Mpg^{-/-}$ mice which show limited glycosylase activity (Engelward, B.P. *et al*, 1997). BER proteins APE1 and Pol λ have

catalytic activities which are also exhibited by APE2 and Pol β respectively (Burkovics, P. *et al*, 2006; Braithwaite, E.K. *et al*, 2005). Chicken DT40 cells lacking Pol λ and Pol β showed hypersensitivity to oxidative DNA damage and lower BER capacity, but lack of either polymerase alone resulted in only a modest effect on sensitivity to oxidative DNA damage and no effect on BER capacity compared with wild-type cells (Tano, K. *et al*, 2007). NUDT1 and NUDT5 overlap functionally in preventing incorporation of 8-oxoG into nascent DNA or RNA. NUDT1 deficient cells show only a 2-fold increase in mutation frequency and both enzymes can reduce the mutation rate in *E. coli mutT* cells (Tsuzuki, T. *et al*, 2001; Sakumi, K. *et al*, 1993; Ishibashi, T. *et al*, 2003).

6.4.7.2 Embryonic lethality

Some of the core components of BER are essential for embryonic development as demonstrated by the embryonic lethality of mice lacking Ape1, Pol β, DNA ligase I or XRCC1 (Xanthoudakis, S. *et al*, 1996; Friedberg, E.C. & Meira, L.B., 2003). $Ape^{+/-}$ mice are phenotypically normal but are haploinsufficient and the observed changes in BER activity may be related to the redox regulatory activity of APE1 (Raffoul, J.J. *et al*, 2004). Unlike other DNA glycosylases, TDG deficiency is embryonic lethal in mice so this enzyme must have a non-redundant function necessary for development which may relate to BER but is not simply the excision of damaged bases (Cortázar, D. *et al*, 2007). Biallelic germline mutations in these genes which result in a lack of functional protein may therefore not be viable in humans but mutations that produce either functionally compromised protein or a reduced amount of functional protein may be sufficient for survival and could result in defective BER.

6.4.7.3 Tissue specificity

Inactivation of genes involved in protection against DNA damage other than *MUTYH* may not predispose to colorectal adenomas, but have a different phenotypic effect. NUDT1-deficient mice develop more lung, liver and stomach tumours than their wild-type counterparts (Tsuzuki, Y. *et al*, 2001) and there is variable expression of some enzymes including Pol λ and NUDT5 (Aoufouchi, S. *et al*, 2000; Nagasawa, K-I. *et al*, 2000; Gasmi, L. *et al*, 1999) between tissues. Somatic mutations in the *MBD4* gene have been reported in sporadic

and hereditary MSI colorectal tumours (Riccio, A. *et al*, 1999; Bader, S. *et al*, 1999). Similar to *Mutyh*$^{-/-}$ mice on a *Apc*$^{Min/+}$ background (Sieber, O.M. *et al*, 2004), *Mbd4*$^{-/-}$ mice on the same background have increased numbers of intestinal tumours and accelerated tumour progression (Millar, C.B. *et al*, 2002; Wong, E. *et al*, 2002) suggesting that MBD4 function may be important in the colorectum. However, *MBD4* mutations have only been found at hypermutable polyadenine tracts, are not detected in MSS colorectal tumours (Riccio, A. *et al*, 1999; Bader, S. *et al*, 1999) and MBD4 deficiency or heterozygosity provides no selective advantage on MMR-deficient tumourigenesis in mice (Sansom, O.J. *et al*, 2004). These data suggest that mutations in *MBD4* may just be a consequence of MMR deficiency and this glycosylase may not have a critical role in the colorectum. Consistent with this, no pathogenic germline *MBD4* mutations were identified in patients with hyperplastic polyposis syndrome (Chow, E. *et al*, 2006). The *MBD4* polymorphism E346K may be associated with oesophageal squamous cell carcinoma susceptibility (Hao, B. *et al*, 2004) and there is conflicting evidence as to whether this polymorphism is associated with lung cancer risk (Shin, M.C. *et al*, 2006; Sakiyama, T. *et al*, 2005). Different glycosylases preferentially act at certain DNA sequences, cell cycle stages or when a particular base is paired opposite the damaged site (Petronzelli, F. *et al*, 2000a; Hendrich, B. *et al*, 1999b; Hardeland, U. *et al*, 2005; Neddermann, P. *et al*, 1996; Bouziane, M. *et al*, 2000; Cortázar, D. *et al*, 2007). Different tissues are exposed to different carcinogens and have distinct gatekeeper genes which are susceptible to particular types of mutation. Defects in different DNA glycosylases may therefore result in different, tissue-specific phenotypes. APE1 also exhibits tissue-specific variation in activity as shown by the haploinsufficient *Ape*$^{+/-}$ mice. These mice show different alterations in Pol β-dependent BER activity in different tissues. Activity was decreased in the liver, increased in the testes and remained the same in the brain (Raffoul, J.J. *et al*, 2004).

6.4.7.4 Low frequency of mutations

Due to the small sample size of the present study, mutations in *TDG*, *MPG*, *SMUG1*, *MBD4*, *NUDT5*, *APE1* and *POLL* could predispose to colorectal adenomas but remain undetected if they are found at a lower frequency than

MUTYH mutations. This situation occurs in HNPCC, with only a very small proportion of cases accounted for by mutations in the *PMS2* gene (Peltomäki, P., 2005). The patient set analyzed in the present study consisted predominantly of cases with between 11 and 100 adenomas without evidence of vertical transmission of polyposis and included few cases with more than 100 adenomas and none with early onset CRC in the absence of polyposis. Mutations resulting in a slightly different phenotype from that of MAP would therefore not have been detected. Patients in whom one potentially pathogenic germline variant was identified could also possess a germline mutation on the other allele that remained undetected such as a large deletion, a promoter mutation or an epimutation, as recently reported in the *MSH2* gene in a family affected by HNPCC (Chan, T.L. *et al*, 2006).

6.4.7.5 Reduced penetrance and multifactorial inherited susceptibility

A substantial proportion of inherited CRC susceptibility is thought to be the result of incompletely penetrant, relatively rare variants in different genes. These might individually confer only a very modest increase in relative risk of CRC but act in an additive or multiplicative manner (Fearnhead, N.S. *et al*, 2005). A significant association between the combined effects of missense variants in the *APC*, *AXIN*, *CTNNB1*, *MLH1* and *MSH2* genes and colorectal adenomas has been reported (Fearnhead, N.S *et al*, 2004). Rare variants identified in the present study may be good candidates for multifactorial inherited CRC susceptibility as defects in repair could increase genetic instability which plays a key role in tumourigenesis. In addition to *MLH1* and *MSH2* variants, a rare variant in the *BLM* gene may also be involved in CRC predisposition (Fearnhead, N.S *et al*, 2004; Gruber, S.B. *et al*, 2002). High frequency, low penetrance alleles may also be associated with CRC risk but this study did not have the power to detect an association between such alleles and the multiple adenoma phenotype. There is good evidence of an increased CRC risk associated with polymorphisms in the *HRAS1*, *MTHFR* and *NAT2* genes (Kemp, Z. *et al*, 2004). Large case-control studies will be needed to determine whether variants identified in the present study contribute to the hereditary CRC predisposition.

Thirlwell *et al* recently analyzed the frequencies and spectra of somatic mutations in the *APC*, *K-ras* and *BRAF* genes in patients with 5-100 colorectal adenomas without evidence of vertical transmission in whom no germline mutation in known predisposition genes had been identified. Unlike MAP tumours, no mutational signature was detected and the genetic pathway of tumourigenesis appeared very similar to the classical pathway in sporadic adenomas. It was suggested that germline variation acting after tumour initiation to cause accelerated progression from microadenoma to macroadenoma may result in the multiple colorectal adenoma phenotype and that this variation was unlikely to be a in gatekeeper or caretaker gene, but in one which could alter the intra- or extracellular microenvironment of cells (Thirlwell, C. *et al*, 2007).

Chapter Seven

General Discussion

7.1 Summary of results of the present study

Patients and families are included on the Wales Polyposis Register (WPR) when at least ten colorectal adenomas with or without CRC are present. A high proportion (85/92, 92%) of families on the WPR were confirmed by molecular analysis to have either FAP (70/92, 76%) or MAP (15/92, 16%). The seven index patients in whom no pathogenic *APC* or *MUTYH* mutations were identified did not have a dominant family history and had tens to hundreds of colorectal adenomas with or without CRC. The proportion of patients whose polyposis could not be accounted for by known genes was substantially smaller than reported from other polyposis registers (Aceto, G. *et al*, 2005; Aretz, S. *et al*, 2006).

Mutational analysis of 182 unrelated MAP index cases identified through a European collaborative project extended the *MUTYH* mutation spectrum by identifying eight novel mutations and highlighted the need for ORF sequencing as a significant proportion (31/182, 17%) of cases did not carry either of the two common non-Asian mutations (Y176C and G393D). Of 138 cases with CRC, 49 (36%) had more than one colorectal cancer. MAP cases were not found to be at increased risk of cancers outside the GI tract. Genotype-phenotype analysis of MAP cases with the common non-Asian mutations revealed that the mean age at presentation and at CRC diagnosis were inversely correlated with the number of Y176C alleles carried.

Analysis of the obligate heterozygote parents of European MAP index patients identified a two-fold increased risk of CRC compared with the general population. No increase in mortality or other cancer risk was observed.

No clearly pathogenic mutations were identified in the candidate BER or DNA damage protection genes (*TDG*, *MPG*, *SMUG1*, *MBD4*, *APE1*, *POLL* and

NUDT5) that were screened in 58 unrelated multiple colorectal adenoma patients in whom *APC* and *MUTYH* mutations had not been found.

7.2 Highly penetrant colorectal adenoma and CRC predisposition alleles

Pathogenic mutations in *APC*, *MUTYH* and MMR genes account for some, but not all, familial CRC predisposition. Other high penetrance alleles, if they exist, remain unidentified. Linkage analysis could be used to map such alleles but requires a large family or many smaller families to be successful. Large CRC families are rare and obtaining samples from a large number of affected members is not always possible. Even highly penetrant alleles may not achieve complete penetrance even by old age and interpretation of results can be complicated by the presence of unaffected carriers or individuals who appear to be affected but do not carry the disease-associated allele (phenocopies). In addition, linkage analysis is difficult if there is locus heterogeneity within the group of families studied or if more than one locus is contributing to the disease in a family. Consanguineous families are extremely useful for the identification of recessive high penetrance alleles. Even if a locus is found to be linked to CRC susceptibility through genetic linkage analysis, the disease-associated region often remains very large and usually contains many candidate genes. Most of the multiple adenoma cases in whom no known genetic defect could be identified in the present study occurred as sporadic cases and so were not suitable for linkage analysis.

Alternatively, high penetrance alleles could be identified through screening of functional candidate genes in early-onset CRC and/or colorectal adenoma patients. Such alleles will be rare in the general population but should be enriched in the patient panel so large cohorts may not be essential. One problem with this approach is candidate gene selection. Highly penetrant CRC predisposition genes may act in pathways known to be involved in colorectal tumourigenesis but could function in pathways not currently known to be related to CRC. Recent analysis of somatic mutations in non-familial CRC identified changes in genes not previously known to be associated with tumourigenesis (Sjöblom, T. *et al*, 2006; Wood, L.D., *et al*, 2007), providing new potential candidates for germline screening. A second and major problem is the

interpretation of rare and possibly pathogenic mutations in patients since proving their pathogenicity is difficult. Bioinformatic tools can only predict the effects of these changes on the encoded protein, with functional work necessary to confirm such predictions. Identification of mutations in the same gene in several patients provides more evidence of an association with CRC susceptibility. In this study rare or unique missense variants were identified in plausible novel candidates but functional work was not able to be carried out and with the exception of a single case, samples were not available from other family members to assess the segregation of these changes with the disease. Mouse models can sometimes give an indication of whether or not mutation of a particular gene is likely to have a phenotypic effect. If a mutation of uncertain significance is identified in a candidate gene, a combination of functional, family and population studies may be required to clarify its pathogenicity and the nature of any association with the disease.

7.3 Low penetrance alleles predisposing to colorectal adenomas and CRC

The genetic basis of most inherited CRC susceptibility is thought to be due to low penetrance alleles and these are now being identified. Genes in which highly penetrant CRC susceptibility alleles have been described may also harbour incompletely penetrant changes. The *APC* missense variant I1307K is found in approximately 6-8% of healthy Ashkenazi individuals and confers a significantly increased risk of colorectal adenomas and CRC in this population, probably through the creation of a hypermutable tract which predisposes to somatic mutations (Laken, S.J. *et al*, 1997; Frayling, I.M. *et al*, 1998). E1317Q is a rare *APC* variant found in less than 1% of UK controls and has been suggested to predispose to colorectal adenomas and CRC with a low and variable penetrance (Frayling, I.M. *et al*, 1998; Lamlum, H. *et al*, 2000). Others have not observed this association (Popat, S. *et al*, 2000; Gismondi, V. *et al*, 2002) and its pathogenicity remains unclear.

Our group, in collaboration with Myriad Genetic Laboratories and Genzyme Genetics, recently reported that multiple rare (MAFs <2%) inherited APC missense mutations may predispose to colorectal adenomas (Azzopardi, D. *et al*, 2008). A significantly higher proportion of patients without truncating *APC*

mutations or biallelic *MUTYH* mutations (non-FAP non-MAP) carried rare *APC* missense mutations than did FAP or MAP cases. The highest over-representation of these rare missense mutations was in non-AFAP non-MAP patients with 11-99 colorectal adenomas. Compared to healthy controls, significantly more non-FAP non-MAP cases carried rare missense variants in the β-catenin down-regulating domain of APC and again the highest over-representation was in non-FAP non-MAP patients with 11-99 colorectal adenomas. Although I1307K and E1317Q contributed to the excess of rare missense changes in non-FAP non-MAP cases, when only variants with a MAF of <0.5% were considered (thereby excluding I1307K, E1317Q and G2502S), there was still an excess of rare *APC* missense variants in non-FAP non-MAP cases. Rare silent variants and common polymorphisms were detected at similar frequencies in non-FAP non-MAP patients, FAP or MAP cases and healthy controls (all of whom were of similar ethnic backgrounds) so the findings are unlikely to be the result of population stratification. Not all missense variants would be expected to have a detrimental effect on APC function but *in silico* analysis predicted 39-46% of the variants were likely to be damaging. Functional analysis of 16 missense variants found in the β-catenin down-regulating domain showed that seven altered β-catenin-regulated transcription *in vitro*. A novel *APC* missense mutation (N1026S) has recently been reported to be associated with AFAP in a single family. The variant co-segregated with the disease in the reported family and functional analysis showed that it altered β-catenin-regulated transcription (Menéndez, M. *et al*, 2008).

The *MLH1* missense variant D132H has been reported to confer susceptibility to CRC in the Israeli population through reduced (but not abolished) MMR activity. This variant was found at a significantly higher frequency in CRC cases than unaffected controls. CRC developed at a later age in D132H carriers than in HNPCC patients and the tumours were mostly MSS. Unlike HNPCC cases, multiple primary tumours were rare and no metachronous CRC was found in the D132H carriers (Lipkin, S.M. *et al*, 2004).

Hereditary mixed polyposis syndrome (HMPS) reported in Ashkenazi families is associated with the highly penetrant *CRAC1* locus and is characterized by

multiple colorectal polyps (which are not classical adenomas) and CRC. An association study of UK CRC cases (selected for family history and/or early onset) and unaffected controls found that common SNPs within the *CRAC1* locus were strongly associated with an increased CRC risk suggesting that common low penetrance alleles at this locus can contribute to CRC predisposition in the general population (Jaeger, E. *et al*, 2008).

To identify low penetrance CRC predisposition alleles at loci not previously associated with an inherited CRC syndrome or colorectal tumourigenesis, genome-wide association studies can be used. Such studies have recently identified a CRC susceptibility locus at 8q24.21 (Tomlinson, I. *et al*, 2007; Zanke, B.W. *et al*, 2007). Two common SNPs (rs6983267 & rs10505477) in linkage disequilibrium were strongly associated with colorectal neoplasia (Tomlinson, I. *et al*, 2007; Zanke, B.W. *et al*, 2007) and mapped to a haplotype block which had been previously associated with prostate cancer risk (Haiman, C.A. *et al*, 2007; Yeager, M. *et al*, 2007). This locus does not contain any known genes but does have two ORFs. One is an uncharacterized gene with multiple alternatively spliced transcripts and harbours rs10505477 within an intron. The other is a processed pseudogene of *POU5F1* (which encodes a transcription factor) and is found 20kb telomeric to rs6983267 (Tomlinson, I. *et al*, 2007; Zanke, B.W. *et al*, 2007). The *MYC* oncogene lies 116kb telomeric to rs6983267 and outside the haplotype block but its function could be influenced by the risk variant through genomic instability or long-range regulation of expression. The CRC susceptibility locus at 8q24.21 influences the risk of colorectal adenomas so may be affecting tumour initiation rather than progression (Tomlinson, I. *et al*, 2007). Three common *SMAD7* SNPs which influence CRC risk have also recently been identified through a genome-wide association study. This gene is involved in TGF-β and Wnt signalling and all three variants map to the same block of linkage disequilibrium located in intron 3 of this gene. Similar to the 8q24.21 locus, the *SMAD7* SNPs significantly influence the risk of colorectal adenomas so may be affecting tumour initiation rather than progression (Broderick, P. *et al*, 2007).
Individually, common low penetrance alleles may confer only a modest effect on CRC risk but could have a substantial effect by acting in an additive or

multiplicative manner. As more of these alleles are identified, different combinations may be associated with varying degrees of risk and some could generate a CRC risk which justifies increased tumour surveillance.

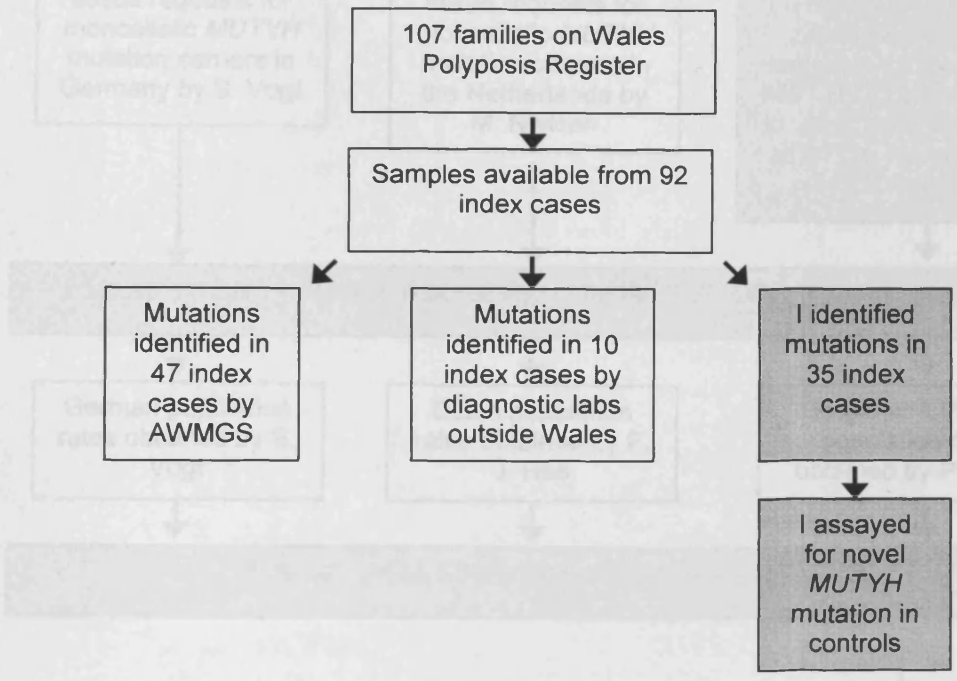
7.4 Application of CRC predisposition alleles in clinical practice

High penetrance CRC predisposition alleles can be readily exploited by medical genetics services to improve the counselling and surveillance of families at high risk. If a patient carries a pathogenic mutation(s), they can undergo colonic surveillance and prophylactic surgery if necessary whereas family members who do not carry the mutant allele(s) do not require additional surveillance. However, genetic counselling and management of patients who carry variants of unknown pathogenicity such as missense changes is difficult, with a requirement for unequivocal genetic and functional data before the variant can be used with confidence in the setting of pre-symptomatic testing.

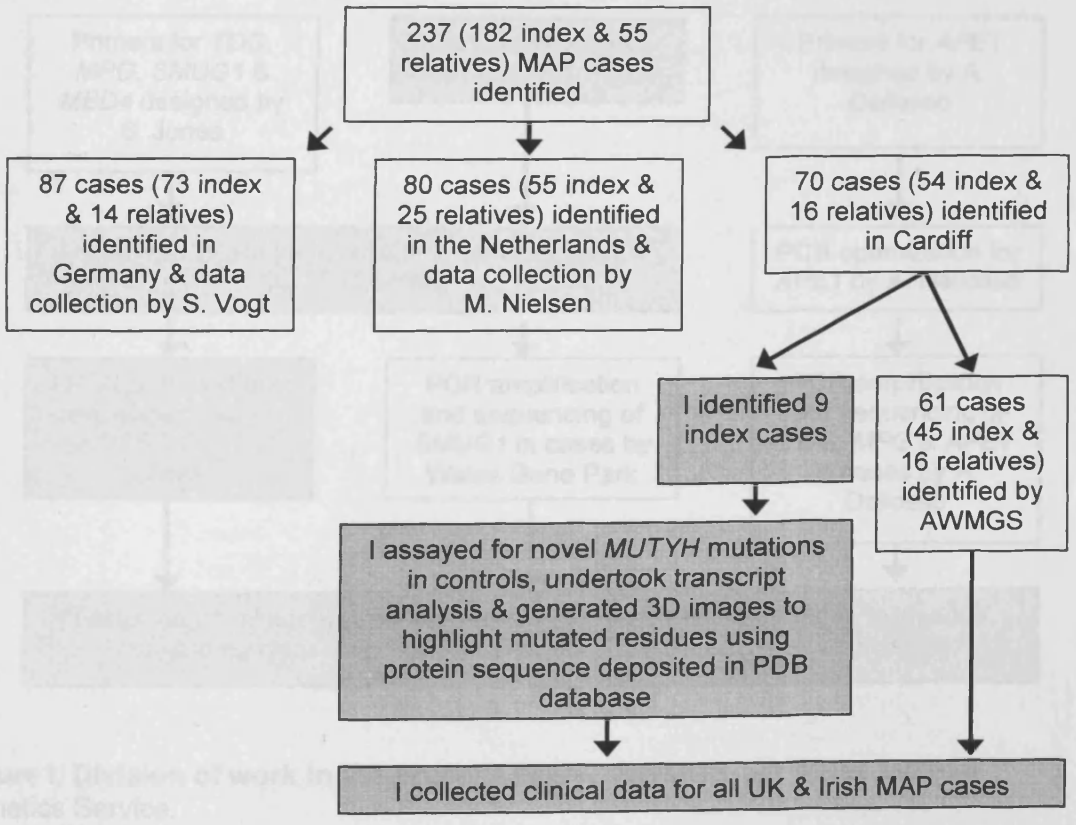
Testing for low penetrance alleles should not be applied in a clinical setting at present but this situation is likely to change as the risks associated with such alleles are clarified. In practice, family history is currently used as an indicator of risk in families in which no high penetrance genetic defect can be identified. In the future common SNPs associated with increased CRC risk may well be used to identify individuals in the general population for whom additional CRC screening is appropriate.

Appendix

Chapter 3



Chapter 4



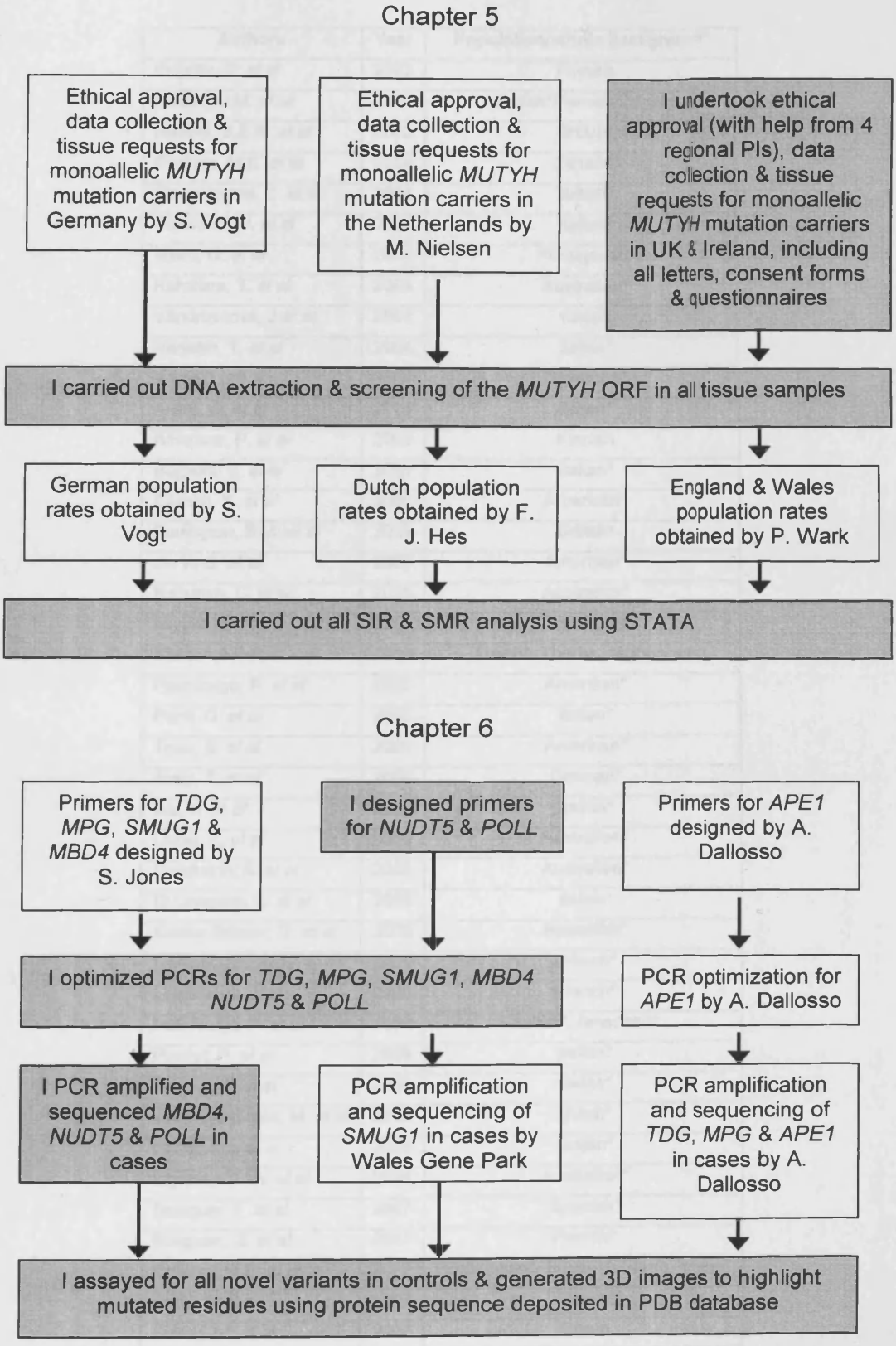


Figure I. Division of work in the present study. AWMGS, All Wales Medical Genetics Service.

Authors	Year	Population/ethnic background*
Enholm, S. *et al*	2003	Finnish
Sieber, O.M. *et al*	2003	British*/Finnish/Danish
Halford, S.E.R. *et al*	2003	British*
Croitoru, M.E. *et al*	2004	Canada*
Fleischmann, C. *et al*	2004	British*
Gismondi, V. *et al*	2004	Italian*
Isidro, G. *et al*	2004	Portugeuse*
Kambara, T. *et al*	2004	Australian*
Vandrovcová, J *et al*	2004	Czech
Venesio, T. *et al*	2004	Italian*
Wang, L. *et al*	2004	American*
Aceto, G. *et al*	2005	Italian*
Alhopuro, P. *et al*	2005	Finnish
Baglioni, S. *et al*	2005	Italian*
Eliason, K. *et al*	2005	American*
Farrington, S.M. *et al*	2005	British*
Jo, W-S. *et al*	2005	American*
Kairupan, C. *et al*	2005	Australian*
Miyaki, M. *et al*	2005	Japanese
Nielsen, M. *et al*	2005	Dutch*, Turkish, Morroccan
Peterlongo, P. *et al*	2005	American*
Ponti, G. *et al*	2005	Italian*
Truta, B. *et al*	2005	American*
Aretz, S. *et al*	2006	German*
Bai, H. *et al*	2006	Italian*
Chow, E. *et al*	2006	Australian*
Colebatch, A. *et al*	2006	Australian*
Di Gregorio, C. *et al*	2006	Italian*
Kanter-Smoler, G. *et al*	2006	Sweedish*
Lefevre, J.H. *et al*	2006	French*
Lejeune, S. *et al*	2006	French*
Peterlongo, P. *et al*	2006	Italian*, American*
Piccioli, P. *et al*	2006	Italian*
Russell, A. *et al*	2006	Swiss*
van Puijenbroek, M. *et al*	2006	Dutch*
Webb, E.L. *et al*	2006	British*
Wynter, C.V.A. *et al*	2006	Australian*
Balaguer, F. *et al*	2007	Spanish*
Bouguen, G. *et al*	2007	French*
Croitoru, M.E. *et al*	2007	Portuguese*, English*, Polish, Italian*
Ponti, G. *et al*	2007	Italian*
Šulová, M. *et al*	2007	Czech
Kury, S. *et al*	2007	French*

Table I. References used to create Figures 4.2, 4.3, 4.4 and Table 4.2.
*Populations with at least 10 reported mutant alleles and are therefore included in Table 4.2 and Figure 4.3.

Family no.	Patient	Sex	Index/ Sib	Ethnicity	Mutation 1	Mutation 2	Presentation	Age at diagnosis	Polyp count	CRC	Age at CRC	Location	Extracolonic (age)
11	1083	M	Index	Pakistani	Y101X	Y101X	Symptomatic	51	432	Y (1)	51	Rectum	N
20	4773	M	Index	Pakistani	Y101X	Y101X	Symptomatic	62	>100	Y (1)	66	Caecum	N
7	0718	M	Index	British	Y176C	Y176C	Symptomatic	41	156	N	/	/	N
13	1610	M	Index	British	Y176C	Y176C	Symptomatic	45	10+ small no.	N	/	/	N
41	2134	M	Index	British	Y176C	Y176C	Symptomatic	33	50-100	N	/	/	N
5	2120	F	Index	British	Y176C	Y176C	Symptomatic	36	10s	Y (1)	51	Caecum	N
17	1120	F	Index	British	Y176C	Y176C	Symptomatic	38	18+ multiple adenomas	Y (3)	38 (2) 41 (1)	Rectum (2) Perineum (1)	N
21	27864	F	Index	British	Y176C	Y176C	Symptomatic	35	8+ several small	Y (1)	45	Rectum	N
27	2131	F	Index	British	Y176C	Y176C	Symptomatic	45	numerous	Y (1)	45	Caecum	N
40	2133	M	Index	British	Y176C	Y176C	Symptomatic	50	50	Y (1)	50	Unknown	N
42	1675	M	Index	British	Y176C	Y176C	Symptomatic	30	>25 + multiple	Y (1)	30	Ascending colon	>10 duodenal adenomas (30-49) 1 gastric adenoma (30-49)
49	2512	F	Index	British	Y176C	Y176C	Symptomatic	43	100s	Y (2)	43	Transverse colon Transverse/ Descending colon	N
50	2139	F	Index	Irish	Y176C	Y176C	Symptomatic	31	88	Y (1)	57	Rectum	N
51	4312	F	Index	Irish	Y176C	Y176C	Symptomatic	33	13	Y (1)	40	Caecum	N
52	0366	F	Index	Irish	Y176C	Y176C	Symptomatic	48	Multiple+ 25 further removed in 2005	Y (1)	49	Caecum	Carcinoid appendix (49)
9	2124	F	Index	British	Y176C	c.380-1 G>A	Symptomatic	56	14	N	/	/	N
44	2135	M	Index	British	Y176C	W117R	Symptomatic	30	152	N	/	N	
48	1232	M	Index	British	Y176C	c.681 G>A	Symptomatic	42	~120	N	/	/	Salivary adenoma (47) Sebaceous gland hyperplasia (45)
46	0748	F	Index	British	Y176C	c.681 G>A	Symptomatic	47	~200	Y(4)	48	Caecum (3) Descending colon (1)	N
45	2136	M	Index	British	Y176C	924+3 A>C	Screening- due to anaemia	46	119	Y (1)	47	Caecum	Basal cell carcinoma (45)
39	2921	M	Index	British	Y176C	c.1138delC	Symptomatic	53	11-30	Y (1)	53	Rectum	Benign prostatic hyperplasia (56)
18	1949	F	Index	British	Y176C	c.1509+3 C>T	Symptomatic	41	Multiple, >10	Y (3)	41 (1) 65 (1) 68 (1)	Transverse colon (1) Sigmoid (1) Caecum (1)	Breast carcinoma (71)
6	0444	M	Index	British	Y176C	G393D	Symptomatic	47	16+ 'numerous tiny polyps'	N	/	/	N

Family no.	Patient	Sex	Index/ Sib	Ethnicity	Mutation 1	Mutation 2	Presentation	Age at diagnosis	Polyp count	CRC	Age at CRC	Location	Extracolonic (age)
10	0964	M	Index	British	Y176C	G393D	Symptomatic	48	<100	N	/	/	N
15	27459	M	Index	British	Y176C	G393D	Screening due to affected sib	49	50-100	N	/	/	N
26	4481	M	Index	British	Y176C	G393D	Screening due to affected sib	46	50-75	N	/	/	N
28	4292	M	Index	British	Y176C	G393D	Symptomatic	52	20-30	N	/	/	Small intestine carcinoid tumour (62)
29	3771	M	Index	British	Y176C	G393D	Symptomatic	44	>20	N	/	/	Soft palate cancer (46) Multiple actinic porokeratoses (28) Sebaceous adenomas (28)
31	2127	M	Index	British	Y176C	G393D	Symptomatic	12	>100	N	/	/	Gastric cancer (17)
43	2132	M	Index	British	Y176C	G393D	Symptomatic	49	Multiple	N	/	/	N
1	0662	M	Index	British	Y176C	G393D	Symptomatic	31	2	Y (1)	37	Rectum	Sebaceous hyperplasia (33)
4	3890	M	Index	British	Y176C	G393D	Symptomatic	47	25-30	Y (2)	47	Sigmoid (2)	N
14	1016	F	Index	British	Y176C	G393D	Symptomatic	43	3	Y (1)	43	Rectum	N
22	1064	F	Index	British	Y176C	G393D	Symptomatic	65	>100	Y (4)	65	Caecum (1) Splenic flexure (1) Rectosigmoid junction (1) Rectum (1)	N
25	2835	M	Index	British	Y176C	G393D	Symptomatic	65	several/numerous	Y (3)	65	Rectum (1) Transverse colon (2)	N
32	2180	F	Index	British	Y176C	G393D	Symptomatic	58	3 + multiple/numerous	Y (4)	58	Rectum (1) Caecum (1) Transverse colon (1) Unknown (1)	N
36	3215	F	Index	Irish	Y176C	G393D	Screening due to affected sib	43	19	Y (1)	43	Descending colon	N
47	0034	F	Index	British	Y176C	G393D	Symptomatic	53	15	Y (1)	53	Rectum (in situ)	N
38	4690	F	Index	Irish	G393D	G393D	Symptomatic	50	123	N	/	/	N
8	2123	M	Index	British	G393D	G393D	Symptomatic	62	~70	Y (2)	62	Unknown	N
30	1107	M	Index	British	G393D	G393D	Symptomatic	36	4	Y (1)	37	Caecum	N
37	3776	F	Index	British	G393D	G393D	Symptomatic	57	10	Y (2)	57	Sigmoid (2)	N
34	2058	M	Index	British	G393D	G213E	Symptomatic	65	22	Y (1)	68	Appendix	N
24	0376	M	Index	British	G393D	N235S	Symptomatic	50	5+ multiple small sessile polyps	Y (1)	50	Sigmoid	Pilar cyst under right mandible (50)

Family no.	Patient	Sex	Index/ Sib	Ethnicity	Mutation 1	Mutation 2	Presentation	Age at diagnosis	Polyp count	CRC	Age at CRC	Location	Extracolonic (age)
54	4021	F	Index	British	G393D	R242H	Symptomatic	49	22	N	/	/	Carcinoid appendix (42)
2	2121	F	Index	Irish	G393D	c.924+3 A>C	Screening due to affected sib	34	6+ polyps throughout the colon	N	/	/	Benign breast lump (22)
16	2126	M	Index	British	G393D	Q335X	Screening due to affected sib	49	multiple+25 +>20	N	/	/	N
35	4891	M	Index	New Zealand	G393D	1092delC	Symptomatic	36	Multiple	Y (1)	36	Unknown	3 gastric adenomas (43-45)
53	3718	M	Index	Italian	1138delC	1138delC	Screening due to age/programme	51	30-40	Y (1)	51	Recto-sigmoid	N
3	0958	F	Index	Indian	E477X	E477X	Symptomatic	51	40-50	N	/	/	N
19	18831	M	Index	Indian	E477X	E477X	Symptomatic	49	~200	N	/	/	Skin lesions (55)
12	19173	M	Index	Indian	E477X	E477X	Symptomatic	37	>25	Y (1)	37	Rectum	N
23	17191	M	Index	Indian	E477X	E477X	Symptomatic	65	>100+ 74	Y (3)	65 (2) 69 (1)	Transverse colon (1) Caecum (1) Unknown (1)	N
33	2130	M	Index	Indian	E477X	E477X	Symptomatic	55	40-50	Y (1)	55	Caecum	N
20	brother1	M	Sib	Pakistani	Y101X	Y101X	Symptomatic	60	multiple in resection, at least 12	Y (1)	60	Caecum	N
20	Daughter 1	F	Offspring	Pakistani	Y101X	Y101X	Screening to due affected parent	38	>100	N	/	/	N
20	Daughter 2	F	Offspring	Pakistani	Y101X	Y101X	Screening to due affected parent	40	No assessment yet	N	/	/	N
20	nephew1	M	Offspring of sib	Pakistani	Y101X	Y101X	Screening to due affected parent	46	>100	N	/	/	N
42	brother1	M	Sib	British	Y176C	Y176C	Screening due to affected sib	29	Multiple	N	/	/	N
42	brother2	M	Sib	British	Y176C	Y176C	Screening due to affected sib	28	>100	N	/	/	N
15	sister1	F	Sib	British	Y176C	G393D	screening due to affected sib	51	77+multiple	N	/	/	N
26	sister1	F	Sib	British	Y176C	G393D	unknown	45	>100	N	/	/	N
31	brother1	M	Sib	British	Y176C	G393D	screening due to affected sib	14	>100	N	/	/	N
14	brother3	M	Sib	British	Y176C	G393D	Symptomatic	53	>100	Y (1)	61	Rectum	N
15	brother1	M	Sib	British	Y176C	G393D	Symptomatic	46	2 adenomas in resection	Y (1)	46	Rectum	N
26	brother3	M	Sib	British	Y176C	G393D	symptomatic	44	76	Y (1)	44	Caecum	N
2	0675	F	Sib	Irish	G393D	c.924+3 A>C	Symptomatic	57	Unknown	Y (1)	57	Unknown	N
53	3717	M	Sib	Italian	1138delC	1138delC	Screening due to affected sib	45	>100	Y (1)	45	Caecum	N

Family no.	Patient	Sex	Index/ Sib	Ethnicity	Mutation 1	Mutation 2	Presentation	Age at diagnosis	Polyp count	CRC	Age at CRC	Location	Extracolonic (age)
12	sister2	F	Sib	Indian	E477X	E477X	Screening due to affected sib	44	143	N	/	/	N
12	sister1	F	Sib	Indian	E477X	E477X	Screening due to affected sib	36	120	Y (1)	36	Transverse colon	N

Table II. Genotype and phenotype details of 70 MAP cases referred to the All Wales Medical Genetics Service for *MUTYH* germline mutation analysis.

Extracolonic manifestation	Patient (age at diagnosis in years)
GI tract	
Appendix carcinoid tumour	0366 (49), 4021 (42)
Duodenal polyps	1675 (30-49), D2031 (68), D2137 (50), D2124 (56), D2121 (67), D123 (57), D203 (52), D16 (43), D194 (65), D128 (36), G26-2 (36-48), G370-1 (41-47), G398-1 (44-60), G489-1 (49), G526-1 (41-51), G526-2 (25-43), G719-1 (47-54), G787-1 (56-58), G818-2 (40), G848-1 (60-65), G925-1 (48), G1125-1 (51-60), G1260-1 (48), G1334-1 (30-31), G872-1 (37-42)
Jejunal polyps	G526-2 (44)
Duodenal cancer	D2124 (56), D194 (65)
Ileum cancer	D2140 (65), D2159 (41)
Small intestine carcinoid tumour	4292 (62), D193 (77)
All gastric polyps	1675 (30-49), 4891 (43-45), 17191 (65-66), D48 (46), G660-1 (46-54), G719-1 (47-54), G1175-2 (40-41), G1468-1 (50-78), D2111 (50), D202 (50), D21 (59), D2137 (53), D2121 (67), D196 (65), G1111-1 (44-52)
Gastric adenomas	1675 (30-49), 4891 (43-45), G1175-2 (40-41)
Fundic gland polyps	D2111 (50), D202 (50), D48 (46), D21 (59), D2137 (53), D2121 (67), D196 (65), G1111-1 (44-52), G719-1 (47-54), G1175-2 (40-41)
Gastric polyps with unknown histology	1675 (30-49), 17191 (65-66), D48 (46), G660-1 (46-54), G1468-1 (50-78)
Gastric cancer	2127 (17), G548-1 (38), G1293-1 (48)
Skin lesions	
Melanoma (M)	G885-1 (32), D2124 (30)
Basal cell carcinoma (M)	2136 (45), D6 (41), D54 (44), D123 (58), D2031 (71), D193 (63), G858-1 (62)
Squamous cell carcinoma (M)	G395-1 (60), G561-1 (68)
Sebaceous gland epithelioma (M)	G1180-1 (51)
Sebaceous gland hyperplasia (B)	0662 (33), 1232 (45), D48 (46), D176 (unknown)
Sebaceous gland adenomas (B)	3771 (28), G925-1 (15), G1222-1 (47), D101 (68, 74)
Steatocystoma (B)	D176 (unknown)
Actinic keratoses (pre-malignant)	3771 (28), D180 (63)
Epidermoid cyst (B)	G1323-1 (unknown), G489-1 (unknown), D2118 (50), G561-1 (unknown), D176 (unknown)
Follicle cyst formation with branches (B)	D2113 (38)
Seborrhoeic wart (B)	D2111 (53)
Lipoma (B)	G26-1 (unknown), G26-2 (30), G848-1 (65), G1309-1 (33), G1412-1 (unknown), G489-2 (37), G489-1 (unknown)
Pilar cyst (B)	0376 (50)
Dermatofibroma (B)	D2329 (54)
Naevus naevocellularis pigmentosus	G1175-2 (36), D2113 (36)
Pigmented spots	D84 (unknown)
Verruca vulgaris	D123 (60)
Skin lesions of unknown histology	18831 (55), G1421-1 (25), G1065-1 (36), G774-1 (58), G395-1 (50), D2140 (unknown), D128 (unknown), G370-1 (15)
Other cancers	
Breast cancer	1949 (71), G787-1 (56), G1358-1 (49), G561-1 (60, 68), G1293-1 (45), D39 (50), D193 (76, 78), D190 (55), D112 (50)
Bladder cancer	G526-1 (45), G858-1 (62), D176 (unknown), D178 (67)
Larynx cancer	D178 (57)
Prostate cancer	D180 (62)
Teratoma	G1451-1 (28), G26-2 (33)
Testicular cancer	G914-1 (41)

Thyroid cancer	G1293-1 (38)
Cervical cancer	G1434-1 (39)
Gangliocytoma	1114-1 (52)
Lymphoma	G1087-1 (46)
Ovarian cancer	G489-1 (56)
Endometrial cancer	G1323-1 (54)
Oesophageal cancer	D80 (59), G676-1 (unknown)
Soft palate cancer	4292 (46)
Other benign abnormalities	
Benign breast lump	2121 (22), D196 (unknown) G1125-1 (59), G641-1 (31)
Mammary cyst	G641-1 (32)
Endometrial polyp	1451-2 (48)
Endometrial hyperplasia/ tubulaire metaplasia	D52 (unknown)
Cervical plaveisel	D122 (27)
Uterine myomas	D54 (unknown)
Uterine polyps/cysts	G1065-1 (48, 50), G641-1 (32)
Ovarian cysts	G826-1 (unknown), G1371-1 (unknown)
Uterine leiomyoma	G641-1 (32), G826-1 (unknown), G370-1 (unknown)
Prostatic hyperplasia	2921 (56), D101 (unknown), D178 (67)
Fibrous histiocytoma	D178 (60)
Renal cyst	G1229-1 (64)
Hepatic cysts	G719-1 (47), G1077-2 (70), G1229-1 (64), G489-2 (48)
Salivary adenoma	1232 (47)
Jaw-bone cysts	G526-2 (10), G757-1 (14), G370-1 (21), G641-1 (23)
Bone tumour	G1222-1 (48)
Capillary hemangioma (unknown)	D176 (unknown)
Leg exostosis	G26-2 (33)
Carpall tunnel syndrome	D11 (unknown)
Graves disease	D52 (unknown)
Basedows disease	G826-1 (unknown)

Table III. Extracolonic manifestations of 237 MAP patients. D or G in patient identifier represents cases from the Dutch or German centres, respectively. All other patients were referred to The All Wales Medical Genetics Service.

Covering letter for MAP patients version 2 06/06/06 *To be put on RGS-headed paper*

Date Tel no.

Address of MAP patient.....
.....
.....
.....

Dear ... *MAP patient*.....,

You are invited to participate in a research project as you have been affected by bowel polyps and/or bowel cancer. Recently developed gene tests requested by your doctor have shown that in your case the bowel problem is caused by changes in the *MUTYH* gene (also called the *MYH* gene). This causes an inherited predisposition to bowel polyps and bowel cancer that is now called MAP (MUTYH-associated polyposis).

Prof. Julian Sampson's group at the Institute of Medical Genetics, Cardiff was the first to define the disease MUTYH-associated polyposis (MAP), the condition that has affected you. Prof. Sampson's team are continuing to undertake research into this disease and have asked if you would be willing to take part in their study of families with MAP entitled "INVESTIGATING CANCER RISK IN PEOPLE WITH *MUTYH* GENE CHANGES" as there is still much that is not known about the disorder. This would involve relevant details on your medical and family history being passed from us to the researchers at Cardiff. If either of your parents were affected by bowel cancer in the past, the researchers would also like permission for any tissues stored at the hospital to be made available for analysis.

Please read the enclosed information sheet and complete and return the enclosed consent form. A stamped addressed envelope is also enclosed.

Please do not hestitate to contact me on the above number if you have any questions or if I can be of any further help. If you have any questions for Prof. Julian Sampson you can contact him on 02920 746412.

Yours sincerely

Information sheet for MAP patients version 2 (06/06/06) - *to be put on IMG-headed paper.*

Date Direct line: 02920 746412

Dear ...*Name of MAP patient*.....,

I am part of a team undertaking research into *MUTYH*-associated polyposis (MAP), a disease that causes an inherited predisposition to bowel polyps and bowel cancer.

You are being invited to take part in a research study. Before you decide whether to participate it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with friends, relatives and your GP if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Consumers for Ethics in Research (CERES) publish a leaflet entitled 'Medical Research and You'. This leaflet gives more information about medical research and looks at some questions you may want to ask. A copy may be obtained from CERES, PO Box 1365, London N16 0BW.

Thank you for reading this.

WHY HAVE YOU BEEN CHOSEN?

You are invited to participate in this research project as you have been affected by bowel polyps and/or bowel cancer. Recently developed gene tests requested by your doctor have shown that in your case the bowel problem is caused by changes in the *MUTYH* gene (also called the *MYH* gene). This causes an inherited predisposition to bowel polyps and bowel cancer that is now called MAP (MUTYH-Associated Polyposis).

We are conducting a study of families with MAP entitled "INVESTIGATING CANCER RISK IN PEOPLE WITH *MUTYH* GENE CHANGES". The genetic cause of MAP was only identified in 2002 and there is still much that is not known about the disorder. We can only improve our knowledge of the disorder and how best to manage affected families with the help of those who are affected. I am therefore writing to ask if you would be willing to take part in our study.

WHAT WOULD THIS RESEARCH INVOLVE FOR YOU?

We would like your permission for your Regional Genetics Service to pass on details of your medical history and family history to us and we may ask you for further details if required. Current data indicates that in a very small proportion of families a parent of the patient (or patients) with MAP has been affected by bowel cancer. If you agree to the study and

one (or both) of your parents had bowel cancer but is no longer alive, we would also like to perform genetic analysis on any bowel cancer tissue that was removed at surgery and stored in the hospital. We may wish to confirm details about this surgery with you if this information is not already known to your Regional Genetics Service. If required we will post a questionnaire to you about this along with a stamped, addressed envelope to be returned to us after completion. **All this information will remain strictly confidential.**

DOES THIS AFFECT YOUR FUTURE MEDICAL CARE?

You are under no obligation either to enter this study in the first instance or to continue with it if you change your mind at any point. This study will not in any way affect your on-going treatment or care or your legal rights. Participation in this study is unlikely to have direct benefits for you but we hope it will be of benefit to others in the future.

ARE THERE ANY IMPLICATIONS FOR INSURANCE?

We will not disclose any clinical or genetic information about you to any third party such as insurance companies and strict confidentiality about all information pertaining to you will be maintained at all times.

WHAT WILL HAPPEN TO THE RESULTS OF THE RESEARCH?

The results of the study will be published in a medical journal. You will not be identified in any reports or publications. If you would like a copy of the published results you could contact us and we will supply you with these.

If you would like to help with the study please complete the enclosed consent form and return it in the pre-paid envelope.

If you do not wish to take part in the study I would be grateful if you could indicate this on the enclosed form and return it in the pre-paid envelope. Once the form has been received stating that you do not wish to take part you will not be contacted again.

If you have any questions or would like further information before agreeing to help, please do not hesitate to contact me at the above address or telephone number.

Many thanks indeed for your help.

Yours sincerely,

Prof. J.R. Sampson
Professor of Medical Genetics
THE NEXT STEP IS:-

PLEASE SIGN THE ENCLOSED CONSENT FORM AND POST IT BACK TO US IN THE ENVELOPE PROVIDED.

THANK YOU

INVESTIGATING CANCER RISK IN PEOPLE WITH *MUTYH* GENE CHANGES.

CONSENT FOR PARTICIPATION IN STUDY (Consent form 1) VERSION 2 (06.06.2006)

RE: *Name of MAP patient*

I have read and understood the nature of the above clinical study as explained in the information sent to me by Professor Sampson (letter version 2 06/06/06). I confirm that I have had the opportunity to ask questions.

I understand that as part of the study: **(Please put your initials in the boxes, circle either Yes or No & sign at the bottom of the letter where indicated)**

- I agree to take part in this study and hereby give consent for my medical history and family history to be reviewed by the researchers involved in the study for the specific purposes of this research.

  Yes ☐ No ☐

- I agree to provide additional information on my medical and family history if required.

  Yes ☐ No ☐

- I understand that any information about me and my family history will remain confidential and that at any point I am free to withdraw from the study without giving any reason and without this affecting my future treatment or standard of care or my legal rights.

Yes ☐ No ☐

SIGNED----------(Patient)

DATE OF BIRTH:

NAME:

DATE:

ADDRESS:

SIGNED----------(Researcher)

NAME: DATE:

Please post this back to us in the envelope provided.

INVESTIGATING CANCER RISK IN PEOPLE WITH *MUTYH* GENE CHANGES.

CONSENT FOR OBTAINING PREVIOUSLY STORED TISSUE BLOCKS (Consent form 2) VERSION 2 (06.6.2006)

RE: *Name of affected parent*

I have read and understood the nature of the above clinical study as explained in the information sent to me by Professor Sampson (letter version 2 06/06/06). I confirm that I have had the opportunity to ask questions.

I understand that as part of the study: **(Please put your initials in the boxes, circle either Yes or No & sign at the bottom of the letter where indicated)**

- I consent to research investigations being carried out using tumour tissues that have previously been removed from my late parent, ..., at surgery.

Yes ☐ No ☐

- I am happy to be contacted again to provide further information on my late parent and their bowel problem if required for the research .

Yes ☐ No ☐

- I understand that any information about me or my family will remain confidential and that at any point I am free to withdraw from the study without giving any reason and without this affecting my future treatment or standard of care or my legal rights.

Yes ☐ No ☐

SIGNED----------

DATE OF BIRTH:

NAME:

DATE:

ADDRESS:

SIGNED----------(Researcher)

NAME: DATE:

Please post this back to us in the envelope provided.

INVESTIGATING CANCER RISK IN PEOPLE WITH *MUTYH* GENE CHANGES.

Questionnaire 1 version 2 (8/5/06)

Name of MAP index case: *To be completed by IMG* Your ref no. .....

Address: ...*Usually Completed by IMG*..... D.O.B *Completed by IMG*
.....
.....
.....
.....
.....
.....

1. Is the MAP index case deceased □ (please tick)
alive □
unknown ? □

If deceased, please give the D.O.D.....

2. Please provide the following information about the MAP index case;

a. Date of presentation.....

b. Mode of presentation: symptomatic presentation □ (please tick)
screening when asymptomatic for (i) family history □
(ii) other □
(please specify)
.....
.....
unknown □
other □

If other, please specify.....

- Colorectal cancer at presentation? Yes/No/Unknown (delete as appropriate)
- Subsequent colorectal cancer? Yes/No/Unknown (delete as appropriate)

If yes, please give the date of diagnosis for this cancer.....

- Please give details (diagnosis, age at diagnosis) of any other cancer.

.....
.....
.....
.....
.....

3. Please give the following information about the parents of the MAP index case.

Mother: Name ..... D.O.B .....
D.O.D .....
Or age if alive .....
Or age at last Regional Genetics Service contact with family .....
Or age at death if DOB and DOD not known .....

Father: Name ..... D.O.B .....
D.O.D .....
Or age if alive .....
Or age at last Regional Genetics Service contact with family .....
Or age at death if DOB and DOD not known .....

4. If deceased, what was the cause of death and how was this information obtained?

	Mother deceased	Father deceased
Cause of death:	.....	.....
Info obtained from:	relatives □	relatives □

death certificate □
hospital records □
other □

death certificate □
hospital records □
other □

If other, please specify.....

5. Is there a record of colorectal cancer in mother? Yes/No/Unknown
or father? Yes/No/Unknown

6. Please give details of any other cancer in either parent.

Mother	Father
.....	.....
.....	.....
.....	.....

INVESTIGATING CANCER RISK IN PEOPLE WITH *MUTYH* GENE CHANGES.

Questionnaire 2 version 1 (09/03/06)

Name of MAP index case: *To be completed by IMG* Your ref no. .....

Address: ...*Usually Completed by IMG*..... D.O.B...*Completed by IMG*
.....
.....
.....
.....
.....
.....

Name of parent affected by colorectal cancer (CRC) .....*Completed by IMG*.....

D.O.B .....

Please provide the following clinical information about this parent;

- Date of CRC operation .....

- Hospital(s) where operation(s) took place
.....
.....
.....
.....
.....
.....

- A copy of the pathology report.

INVESTIGATING CANCER RISK IN PEOPLE WITH *MUTYH* GENE CHANGES.

Questionnaire 3 for MAP patients version 1 (09/03/06)

Your parent,.....*Completed by IMG*....., was affected by bowel cancer.

Would you please complete the following questions about .....*Completed by IMG*..... to help us to carry out research investigations using any tumour tissues that were removed from your late parent at surgery and stored at the hospital.

1. Affected parent's date of birth .....

2. Date of surgery for bowel cancer .....
(approximate if not known precisely)

3. Hospital(s) where surgery(s) took place

.....
.....
.....
.....
.....

4. Date of death .....

If we have difficulty tracking down the records from the hospitals where your affected parent was treated we may want to contact you for further details. Please could you provide us with your telephone number and times that would be convenient to call you.

Tel no:.....

Please call between ..... and .....

Thank you for your help.

INVESTIGATING CANCER RISK IN PEOPLE WITH *MUTYH* GENE CHANGES.

Questionnaire 4 version 1 (01/12/06)

Name of MAP index case: *To be completed by IMG* Your ref no. *To be completed*
D.O.B *Completed by IMG*

1. If the MAP index case is deceased, please give the cause of death:

.....

- Info obtained from: relatives □
  death certificate □
  hospital records □
  other □

If other, please specify.....

Colorectal adenomas

- Age at first colonoscopy .....
- No. of polyps at first colonoscopy .....
- Histologic type of colorectal polyps (adenomas, hyperplastic etc.)

.....

- Age at last colonoscopy .....
- Cumulative no. of polyps (approx.) .....

Colorectal Cancer

- Colorectal cancer? Yes/No/Unknown (delete as appropriate)

If yes, date of presentation .....

- In which part of the colon was the cancer located? .....

(rectum, sigmoid, descending colon, splenic flexure, transverse colon, c.ascendens, caecum, hepatic flexure, if not exactly known; right-sided, left-sided or unknown)

- Duke's stage of colorectal cancer .....

(A, B1-3, C1-3, D) Carcinoma in situ will not be considered as carcinoma and not included in survival analysis.

- Treatment of colorectal carcinoma (surgery, radiotherapy, chemotherapy)

.....

- Synchronous colorectal cancer? Yes/No/Unknown (delete as appropriate)

If yes, please give location(s).....

- Subsequent (metachronous) colorectal cancer? Yes/No/Unknown (delete as appropriate)

If yes, please give; date of diagnosis for this cancer.....

location of this cancer .....

Duke's stage of this subsequent colorectal cancer.....

treatment of this colorectal carcinoma (if known) (surgery, radiotherapy, chemotherapy)
.....

Gastroduodenal Tumours

- Gastroduodenoscopy? Yes/No/Unknown (delete as appropriate)
- Symptomatic? Yes/No (delete as appropriate)
- Age/date at gastroduodenoscopy .....
- Cumulative no. of duodenal polyps.....
- Histologic type of duodenal polyps (adenomas, hyperplastic etc.)

.....

- Cumulative no. of gastric polyps.....
- Histologic type of gastric polyps (adenomas, hyperplastic etc.)

.....

- Duodenal or gastric cancer? Yes/No/Unknown (delete as appropriate)

If yes, please give details below.
.....
.....

Extraintestinal Tumours

Please give details of any other tumour/cancer in this patient

Type	Diagnosis was verified by the following examination	Histological result	Location	Age at diagnosis	Mode of diagnois (symptomatic, screening, by chance)

- Please give details of any other significant health problem.
.....
.....
.....

- Does the MAP index patient have other family members affected by MAP?

  Yes/No/Unknown (delete as appropriate)

Publications resulting from this work

N. Jones, M. Nielsen, M.C. Joerink - van de Beld, S. Vogt, C.M. Tops, H.F.A. Vasen, F.J. Hes, S. Aretz, J. R. Sampson. Hereditary Colorectal Cancer, Chapter V.19 *MUTYH*-associated polyposis. *Awaiting publication*

M. Nielsen *, M.C. Joerink - van de Beld *, N. Jones, S. Vogt, C.M. Tops, H.F.A. Vasen, J. R. Sampson, S. Aretz, F.J. Hes. Genotype and phenotype heterogeneity in *MUTYH* Associated Polyposis. *Manuscript in preparation.* *co-first authors

Dallosso, A.R.*, Dolwani, S.*, Jones, N.*, Jones, S.*, Colley, J., Maynard, J., Idziaszczyk, S., Humphreys, V., Arnold, J., Donaldson, A., Eccles,D., Ellis, A., Evans, D.G., Frayling, I.M., Hes, F.J., Houlston, R.S., Maher, E.R., Nielsen, M., Parry, S., Tyler, E., Moskvina, V., Cheadle, J.P., Sampson, J.R. Inherited predisposition to colorectal adenomas caused by multiple rare alleles of MUTYH but not *OGG1*, *NUDT1*, *NTH1* or *NEIL 1*, *2* or *3*. *Gut, in press*. *co-first authors listed in alphabetical order

Azzopardi, D., Dallosso, A.R., Eliason, K., Hendrickson, B.C., Jones, N., Rawstorne, E., Colley, J., Moskvina, V., Frye, C., Sampson, J.R., Wenstrup, R., Scholl, T., Cheadle, J.P. (2008). Multiple rare nonsynonymous variants in the *Adenomatous Polyposis Coli* gene predispose to colorectal adenomas. *Cancer Res.* **68(2)**:358-363

References

Aceto, G., Curia, M.C., Veschi, S., De Lellis, L., Mammarella, S., Catalano, T., Stuppia, L., Palka, G., Valanzano, R., Tonelli, F., Casale, V., Stigliano, V., Cetta, F., Battista, P., Mariani-Costantini, R., Cama, A. (2005) Mutations of *APC* and *MYH* in unrelated Italian patients with Adenomatous Polyposis Coli. *Hum. Mutat.* **26(4)**:394

Acharya, S., Wilson, T., Gradia, S., Kane, M.F., Guerrette, S., Marsischky, G.T., Kolodner, R., Fishel, R. (1996). hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. *Proc. Natl. Acad. Sci. U.S.A.* **93**:13629-13634

Akiyama, K., Seki, S., Oshida, T., Yoshida, M.C. (1994). Structure, promoter analysis and chromosomal assignment of the human *APEX* gene. *Biochim. Biophys. Acta.* **1219**:15-25

Akyerli, C.B., Özbek, U., Aydin-Sayitoğlu, M., Sirma, S., Özçelik, T. (2003) Analysis of *MYH* Tyr165Cys and Gly382Asp variants in childhood leukemias. *J. Cancer Res. Clin. Oncol.* **129**:604-605

Albuquerque, C., Breukel, C., van der Luijt, R., Fidalgo, P., Lage, P., Slors, F.J.M., Leitão, Fodde, R., Smits, R. (2002). The 'just-right' signalling model: *APC* somatic mutations are selected based on a specific level of activation of the β-catenin signalling cascade. *Hum. Mol. Genet.***11(13)**:1549-1560

Alhopuro, P., Parker, A.R., Lehtonen, R., Enholm, S., Järvinen, H.J., Mecklin, J-P., Karhu, A., Eshleman, J.R., Aaltonen, L.A. (2005). A novel functionally deficient *MYH* variant in individuals with colorectal adenomatous polyposis. *Hum. Mutat.* **26(4)**:393

Al-Tassan, N., Chmiel, N.H., Maynard, J., Fleming, N., Livingston, A.L., Williams, G.T., Hodges, A.K., Davies, D.R., David, S.S., Sampson, J.R., Cheadle, J.P. (2002). Inherited variants of *MYH* associated with somatic G:C→T:A mutations in colorectal tumours. *Nat. Genet.* **30**:227-232

Al-Tassan, N., Eisen, T., Maynard, J., Bridle, H., Shah, B., Fleischmann, C., Sampson, J.R., Cheadle, J.P., Houlston, R.S. (2004) Inherited variants in *MYH* are unlikely to contribute to the risk of lung carcinoma. *Hum. Genet.* **114**:207-210

Ames, B.N. & Gold, L.S. (1991). Endogenous mutagens and the causes of aging and cancer. *Mutat. Res.* **250**:3-16

An, Q., Robins, P., Lindahl, T., Barnes, D.E. (2005). C→T mutagenesis and γ-radiation sensitivity due to deficiency in the Smug1 and Ung DNA glycosylases. *EMBO J.* **24**:2205-2213

Aoufouchi, S., Flatter, E., Dahan, A., Faili, A., Bertocci, B., Storck, S., Delbos, F., Cocea, L., Gupta, N., Weill, J-C., Reynaud, C-A. (2000). Two

novel human and mouse DNA polymerases of the polX family. *Nucleic Acids Res.* **28(18)**:3684-3693

Arai, K., Morishita, K., Shinmura, K., Kohno, T., Kim, S-R., Nohmi, T., Taniwaki, M., Ohwada, S., Yokota, J. (1997). Cloning of a human homolog of the yeast *OGG1* gene that is involved in the repair of oxidative DNA damage. *Oncogene* **14**:2857-2861

Aravind, L. & Koonin, E.V. (2000). The α/β fold uracil DNA glycosylases: a common origin with diverse fates. *Genome Biol.* **1(4)**:research0007.1-0007.8

Aretz, S., Ulhaas, S., Caspari, R., Mangold, E., Pagenstecher, C., Propping, P., Friedl, W. (2004a). Frequency and parental origin of *de novo APC* mutations in familial adenomatous polyposis. *Eur. J. Hum. Genet.* **12**:52-58

Aretz, S., Ulhaas, S., Sun, Y., Pagenstecher, C., Mangold, E., Caspari, R., Möslein, G., Schulmann, K., Propping, P., Friedl, W. (2004b). Familial Adenomatous Polyposis: aberrant splicing due to missense or silent mutations in the *APC* gene. *Hum. Mutat* **24**:370-380

Aretz, S., Stienen, D., Uhlhaas, S., Pagenstecher, C., Mangold, E., Caspari, R., Propping, P., Friedl, W. (2005). Large submicroscopic genomic *APC* deletions are a common cause of typical Familial Adenomatous Polyposis. *J. Med. Genet.* **42**:185-192

Aretz, S., Ulhaas, S., Goergens, H., Siberg, K., Vogel, M., Pagenstecher, C., Mangold, E., Caspari, R., Propping, P., Friedl, W. (2006). *MUTYH*-associated polyposis: 70 out of 71 patients with biallelic mutations present with an attenuated or atypical phenotype. *Int. J. Cancer* **119**:807-814

Aretz, S., Stienen, D., Friedrichs, N., Stemmler, S., Ulhaas, S., Rahner, N., Propping, P., Friedl, W. (2007). Somatic APC mosaicism: a frequent cause of Familial Adenomatous Polyposis (FAP). *Hum. Mutat* **28(10)**:985-992

Armstrong, J.G., Davies, D.R., Guy, S.P., Frayling, I.M., Evans, D.G.R. (1997). *APC* mutations in Familial Adenomatous Polyposis families in the Northwest of England. *Hum. Mutat* **10**:376-380

Arnold, A., Payne, S., Fisher, S., Fricker, D., Soloway, J., White, S.M., Novelli, M., MacDonald, K., Mackay, J., Groves, R., Canham, N. (2007). An individual with Muir-Torre syndrome found to have a pathogenic *MSH6* gene mutation. *Fam. Cancer* **6(3)**:317-321

Aspinwall, R., Rothwell, D.G., Roldan-Arjona, T., Anselmino, C., Ward, C.J., Cheadle, J.P., Sampson, J.R., Lindahl, T., Harris, P.C., Hickson, I.D. (1997). Cloning and characterization of a functional human homolog of *Escherichia coli* endonuclease III. *Proc. Natl. Acad. Sci. U.S.A.* **94**:109-114

Azzopardi, D., Dallosso, A.R., Eliason, K., Hendrickson, B.C., Jones, N., Rawstorne, E., Colley, J., Moskvina, V., Frye, C., Sampson, J.R., Wenstrup,

R., Scholl, T., Cheadle, J.P. (2008). Multiple rare nonsynonymous variants in the *Adenomatous Polyposis Coli* gene predispose to colorectal adenomas. *Cancer Res.* **68(2)**:358-363

Baba, D., Maita, N., Jee, J-G., Uchimura, Y., Saitoh, H., Sugasawa, K., Hanaoka, F., Tochio, H., Hiroaki, H., Shirakawa, M. (2005). Crystal structure of thymine DNA glycosylase conjugated to SUMO-1. *Nature* **435**:979-982

Bader, S., Walker, M., Hendrich, B., Bird, A., Bird, C., Hooper, M., Wyllie, A. (1999). Somatic frameshift mutations in the *MBD4* gene of sporadic colon cancers with mismatch repair deficiency. *Oncogene* **18**:8044-8047

Bader, S., Walker, M., Harrison, D. (2000). Most microsatellite unstable sporadic colorectal carcinomas carry *MBD4* mutations. *Br. J. Cancer* **83(12)**:1646-1649

Bader, S., Walker, M., Harrison, D. (2007). A human cancer-associated truncation of MBD4 causes dominant negative impairment of DNA repair in colon cancer cells. *Br. J. Cancer* **96**:660-666.

Baglioni, S., Melean, G., Gensini, F., Santucci, M., Scatizzi, M., Papi, L., Genuardi, M. (2005). A kindred with MYH-associated polyposis and pilomatricomas. *Am. J. Med. Genet. A* **134(2)**:212-214

Bai, H., Grist, S., Gardner, J., Suthers, G., Wilson, T.M., Lu, A-L. (2007). Functional characterization of human MutY homolog (hMYH) missense mutation (R231L) that is linked with hMYH-assoicated polyposis. *Cancer Lett.* **250**:74-81

Bai, H., Jones, S., Guan, X., Wilson, T.M., Sampson, J.R., Cheadle, J.P., Lu, A-L. (2005). Functional characterization of two human MutY homolog (hMYH) missense mutations (R227W and V232F) that lie within the putative hMSH6 binding domain and are associated with hMYH polyposis. *Nucleic Acids Res.* **33(2)**:597-604

Bala, S., Sulekova, Z., Ballhausen, W.G. (1997). Constitutive *APC* exon 14 skipping in early-onset Familial Adenomatous Polyposis reveals a dramatic quantitative distortion of *APC* gene-specific isoforms. *Hum. Mutat* **10**:201-206

Balaguer, F., Castellvi-Bel, S., Castells, A., Andreu, M., Munoz, J., Gisbert, J.P., Ilor, X., JOver, R., de Cid, R., Gonzalo, V., Bessa, X., Xicola, R.M., Pons, E., Alenda, C., Paya, A., Pique, J.M., Gastrointestinal Oncology Group of the Spanish Gastroenterological Association. (2007). Identification of *MYH* mutation carriers in colorectal cancer: a multicenter, case-control, population-based study. *Clin. Gastroenterol. Hepatol.* **5**:379-387

Ballestar, E. & Wolffe, A.P. (2001). Methyl-CpG-binding proteins: targeting specific gene repression. *Eur. J. Biochem.* **268**:1-6

Bandaru, V., Sunkara, S., Wallace, S.S., Bond, J.P. (2002). A novel DNA glycosylase that removes oxidative DNA damage and is homologous to *Escherichia coli* endonuclease VIII. *DNA Repair (Amst.)* **1**:517-529

Baran, A.A., Silverman, K.A., Zeskand, J., Koratkar, R., Palmer, A., McCullen, K., Curran Jr, W.J., Bocker Edmonston, T., Siracusa, L.D., Buchberg, A.M. (2007). The modifier of Min 2 (Mom2) locus: embryonic lethality of a mutation in the Atp5a1 gene suggests a novel mechanism of polyp suppression. *Genome Res.* **17(5)**:566-576

Barnetson, R.A., Devlin, L., Miller, J., Farrington, S.M., Slater, S., Drake, A.C., Campbell, H., Dunlop, M.G., Porteous, M.E. (2007) Germline mutation prevalence in the base excision repair gene, MYH, in patients with endometrial cancer. *Clin. Genet.* **72**:551-555

Baudhuin, L.M., Roberts, L.R., Enders, F.T.B., Swanson, R.L., Mettler, T.A., Aderca, I., Stadheim, L.M., Highsmith, W.E. (2006) MYH Y165C and G382D mutations in hepatocellular carcinoma and cholangiocarcinoma patients. *J. Cancer Res. Clin. Oncol.* **132**:159-162

Baylin, S.B., Esteller, M., Rountree, M.R., Bachman, K.E., Schuebel, K., Herman, J.G. (2001). Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Hum. Mol. Genet.* **10(7)**:687-692

Bebenek, K., Garcia-Diaz, M., Blanco, L., Kunkel, T.A. (2003). The frameshift infidelity of human DNA polymerase λ. *J. Biol. Chem.* **278(36)**:34685-34690

Behrens, J., Jerchow, B-A., Wurtele, M., Grimm, J., Asbrand, C., Wirtz, R., Kühl, M., Wedlich, D., Birchmeier. (1998). Functional interaction of an axin homolog, conductin, with β-catenin, APC and GSK3β. *Science* **280**:596-599

Bellacosa, A., Cicchillitti, L., Schepis, F., Riccio, A., Yeung, A.T., Matsumoto, Y., Golemis, E.A., Genuardi, M., Neri, G. (1999). MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. *Proc. Natl. Acad. Sci. U.S.A.* **96**:3969-3974

Bennett, R.A.O., Wilson III, D.M., Wong, D., Demple, B. (1997). Interaction of human apurinic endonuclease and DNA polymerase β in the base excision repair pathway. *Proc. Natl. Acad. Sci. U.S.A.* **94**:7166-7169

Bernards, A.S., Miller, J.K., Bao, K.K., Wong, I. (2002). Flipping duplex DNA inside out: a double base-flipping reaction mechanism by *Escherichia coli* MutY adenine glycosylase. *J. Biol. Chem.* **277(23)**:20960-20964

Bertario, L., Russo, A., Sala, P., Varesco, L., Mondini, P., Pierotti, M., Spinelli, P., Radice, P. (2003). Multiple approach to the exploration of genotype-phenotype correlations in Familial Adenomatous Polyposis. *J. Clin. Oncol.* **21(9)**:1698-1707

Bialkowski, K. & Kasprzak, K.S. (1998). A novel assay of 8-oxo-2'-deoxyguanosine 5'-triphosphate pyrophosphohydrolases (8-oxo-dGTPasae) activity in cultured cells and its use for evaluation of cadmium(II) inhibition of this activity. *Nucleic Acids Res.* **26(13)**:3194-3201

Bienz, M. & Hamada, F. (2004). Adenomatous polyposis coli proteins and cell adhesion. *Curr. Opin. Cell Biol.* **16**:528-535

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**:6-21

Bisgaard, M.L., Fenger, K., Bülow, S., Niebuhr, E., Mohr, J. (1994). Familial Adenomatous Polyposis (FAP): frequency, penetrance and mutation rate. *Hum. Mutat.* **3**:121-125

Bisgaard, M.L., Ripa, R.S., Bülow, S. (2004a) Mutation analysis of the adenomatous polyposis coli (*APC*) gene in Danish patients with familial adenomatous polyposis (FAP). *Hum Mut.* **23(5)**:522

Bisgaard, M.L., Ripa, R.S., Kudsen, A.L., Bülow, S. (2004b). Familial adenomatous polyposis patients without an identified *APC* germline mutation have a severe phenotype. *Gut* **53**:266-270

Bisgaard, M.L. & Bülow, S. (2006). Familial Adenomatous Polyposis (FAP): genotype correlation to FAP phenotype with osteomas and sebaceous cysts. *Am. J. Med. Genet. A* **140(3)**:200-204

Blaisdell, J.O., Hatahet, Z., Wallace, S.S. (1999). A novel role for *Escherichia coli* endonuclease VIII in prevention of spontaneous G→T transversions. *J. Bacteriol.* **181(20)**:6396-6402

Boiteux, S., Gajewski, E., Laval, J., Dizdaroglu, M. (1992). Substrate specificity of the *Escherichia coli* Fpg protein (formaminopyrimidine-DNA glycosylase): excision of purine lesions in DNA produced by ionizing radiation or photosensitization. *Biochemistry* **31**:106-110

Boldogh, I., Milligan, D., Soog Lee, M., Bassett, H., Lloyd, R.S., McCullough, A.K. (2001). hMYH cell cycle-dependent expression, subcellular localization and association with replication foci: evidence suggesting replication-coupled repair of adenine:8-oxoguanine mispairs. *Nucleic Acids Res.* **29(13)**:2802-2809

Boon, E.M., Livingston, A.L., Chimel, N.H., David, S.S., Barton, J.K. (2003). DNA-mediated charge transport for DNA repair. *Proc. Natl. Acad. Sci. U.S.A.* **100(22)**:12543-12547

Bouguen, G., Manfredi, S., Blayau, M., Dugast, C., Buecher, B., Bonneau, D., Siproudhis, L., David, V., Bretagne, J-F. (2007). Colorectal adenomatous polyposis associated with *MYH* mutations: genotype and phenotype characteristics. *Dis. Colon Rectum* **50(10)**:1612-1617

Braithwaite, E.K., Prasad, R., Shock, D.D., Hou, E.W., Beard, W.A., Wilson, S.L. (2005a). DNA polymerase λ mediates a back-up base excision repair activity in extracts of mouse embryonic fibroblasts. *J. Biol. Chem.* **280(18)**:18469-18475

Braithwaite, E.K., Kedar, P.S., Lan, L., Polosina, Y.Y., Asagoshi, K., Poltoratsky, V.P., Horton, J.K., Miller, H., Teebor, G.W., Yasui, A., Wilson, S.H. (2005b). DNA polymerase λ protects mouse fibroblasts against oxidative DNA damage and is recruited to sites of DNA damage/repair. *J. Biol. Chem.* **280(36)**:31641-31647

Brensinger, J.D., Laken, S.J., Luce, M.C., Powell, S.M., Vance, G.H., Ahnen, D.J., Petersen, G.M., Hamilton, S.R., Giardiello, F.M. (1998). Variable phenotype of familial adenomatous polyposis in pedigrees with 3' mutation in the *APC* gene. *Gut* 1998 **43**:548-552

Broderick, P., Bagratuni, T., Vijayakrishnan, J., Lubbe, S., Chandler, I., Houlston, R.S. (2006). Evaluation of *NTHL1, NEIL1, NEIL2, MPG, TDG, UNG* and *SMUG1* genes in familial colorectal cancer predisposition. *BMC Cancer* **6**:243

Broderick, P., Carvajal-Carmona, L., Pittman, A.M., Webb, E., Howarth, K., Rowan, A., Lubbe, S., Spain, S., Sullivan, K., Fielding, S., Jaeger, E., Vijayakrishmann, J., Kemp, Z., Gorman, M., Chandler, I., Papaemmanuil, E., Penegar, S., Wood, W., Sellick, G., Qureshi, M., Teixeira, A., Domingo, E., Barclay, E., Martin, L., Sieber, O., members of the CORGI Consortium, Kerr, D., Gray, R., Peto, J., Cazier, J-B., Tomlinson, I., Houlston, R. (2007). A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat. Genet.* **39(11)**:1315-1317

Brown, J.A., Duym, W.W., Fowler, J.D., Suo, Z. (2007). Single-turnover kinetic analysis of the mutagenic potential of 8-oxo-7,8-dihydro-2'-deoxyguanosine during gap-filling synthesis catalyzed by human DNA polymerases λ and β. *J. Mol. Biol.* **367**:1258-1269

Bülow, S., Bülow, C., Faurschou Nielsen, T., Karlsen, L., Moesgaard, F. (1995). Centralized registration, prophylactic examination and treatment results in improved prognosis in Familial Adenomatous Polyposis. *Scand. J. Gastroenterol.* **30**:989-993

Bülow, S. (2003) Results of national registration of familial adenomatous polyposis. *Gut* **52**:742-746

Burkovics, P., Szukacsov, V., Unk, I., Haracska, L. (2006). Human Ape2 has a 3'-5' exonuclease activity that acts preferentially on mismatched base pairs. *Nucleic Acids Res.* **34(9)**:2508-2515

Burns, T., Breathnach, S., Cox, N., Griffiths, C. (2004) Rook's textbook of dermatology, 7th edition. Chapter 37 Tumours of the skin appendages & Chapter 43 Disorders of the sebaceous glands. Blackwell Science (USA)

Butterworth, A.S., Higgins, J.P.T., Pharoah, P. (2006) Relative and absolute risk of colorectal cancer for individuals with a family history: A meta-analysis. *Eur. J. Cancer* **42**:216-227

Cahill, D.P., Lengauer, C., Yu, J., Riggins, G., Willson, J.K.V., Markowitz, S.D., Kinzler, K.W., Vogelstein, B. (1998). Mutations of mitotic checkpoint genes in human cancers. *Nature* **392**:300-303

Cancer Research UK at http://www.cancerresearch.org/

Cannavo, E., Marra, G., Sabates-Bellver, J., Menigatti, M., Lipkin, S.M., Fischer, F., Cejka, P., Jiricny, J. (2005). Expression of the MutL homologue hMLH3 in human cells and its role in DNA mismatch repair. *Cancer Res.* **65(23)**:10759-10766

Cao, X., Hong, Y., Eu, K.W., Loi, C., Cheah, P.Y. (2006). Singapore Familial Adenomatous Polyposis (FAP) patients with classical adenomatous polyposis but undetectable *APC* mutations have accelerated cancer progression. *Am. J. Gastroenterol.* **101**:1-8

Cao, Y., Pieretti, M., Marshall, J., Khattar, N.H., Chen, B., Kam-Morgan, L., Lynch, H. (2002). Challenge in the differentiation between attenuated Familial Adenomatous Polyposis and Hereditary Nonpolyposis Colorectal Cancer: case report with review of the literature. *Am. J. Gastroenterol.* **97(7)**:1822-1827

Capp, J-P., Boudsocq, F., Bertrand, P., Laroche-Clary, A., Pourquier, P., Lopez, B.S., Cazaux, C., Hoffmann, J-S., Canitrot, Y. (2006). *Nucleic Acids Res.* **34(10)**:2998-3007

Carson, D.J., Santoro, I.M., Groden, J. (2004). Isoforms of the APC tumour suppressor and their ability to inhibit cell growth and tumorigenicity. *Oncogene* **23**:7144-7148

Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q., Krainer, A.R. (2003). ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* **31(13)**:3568-3571

Caspari, R., Friedl, W., Mandl, M., Möslein, G., Kadmon, M., Knapp, M., Jacobasch, K.H., Ecker, K-W., Kreißler-Haag, D., Timmermanns, G., Propping, P. (1994). Familial adenomatous polyposis: mutation at codon 1309 and early onset of colon cancer. *Lancet* **343**:629-632

Caspari, R., Olschwang, S., Friedl, W., Mandl, M., Boisson, C., Böker, T., Augustin, A., Kadmon, M., Möslein, G., Thomas, G., Propping, P. (1995). Familial Adenomatous Polyposis: desmoid tumours and lack of ophthalmic

lesions (CHRPE) associated with *APC* mutations beyond codon 1444. *Hum. Mol. Genet.* **4(3)**:337-340

Cattaneo, F., Molatore, S., Mihalatos, M., Apessos, A., Venesio, T., Bione, S., Grignani, P., Nasioulas, G., Ranzani, G.N. (2007). Heterogeneous molecular mechanisms underlie attenuated familial adenomatous polyposis. *Genet. Med.* **9(12)**:836-841

Chai, V., Vassilakos, A., Lee, Y., Wright, J.A., Young, A.H. (2005). Optimization of the PAXgene™ blood RNA extraction system for gene expression analysis of clinical samples. *J. Clin. Lab. Anal.* **19**:182-188

Chan, T.L., Yuen, S.T., Kong, C.K., Chan, Y.W., Chan, A.SY., Ng, W.F., Tsui, W.Y., Lo, M.WS., Tam, W.Y., Li, V.SW., Leung, S.Y. (2006). Heritable germline epimutation of *MSH2* in a family with hereditary nonpolyposis colorectal cancer. *Nat. Genet.* **38(10)**:1178-1183

Cheadle, J.P., Krawczak, M., Thomas, M.W., Hodges, A.K., Al-Tassan, N., Fleming, N., Sampson, J.R. (2002). Different combinations of biallelic *APC* mutations confer different growth advantages in colorectal tumours. *Cancer Res.* **62**:363-366

Chen, D., Lucey, M.J., Phoenix, F., Lopez-Garcia, J., Hart, S.M., Losson, R., Buluwela, L., Coombes, R.C., Chambon, P., Schär, P., Ali, S. (2003). T:G mismatch-specific thymine-DNA glycosylase potentiates transcription of estrogen-regulated genes through direct interaction with estrogen receptor α. *J. Biol. Chem.* **278(40)**:38586-38592

Chen, D.S., Herman, T., Demple, B. (1991). Two distinct human DNA diesterases that hydrolyze 3'-blocking deoxyribose fragments from oxidized DNA. *Nucleic Acids Res.* **19(21)**:5907-5914

Chmiel, N.H., Livingston, A.L., David, S.S. (2003). Insight into the functional consequences of inherited variants of the hMYH adenine glycosylase associated with colorectal cancer: complementation assays with hMYH variants and pre-steady-state kinetics of the corresponding mutated *E. coli* enzymes. *J. Mol. Biol.* **327(2)**:431-443

Chou, K-M. & Chen, Y-C. (2002). An exonucleolytic activity of human apurinic/apyrimidinic endonuclease on 3' mispaired DNA. *Nature* **415**:655-659

Chow, E. & Macrae, F. (2005) Review of Juvenile polyposis syndrome. *J. Gastroenterol. Hepatol.* **20**:1634-1640

Chow, E., Lipton, L., Lynch, E., D'Souza, R., Aragona, C., Hodgkin, L., Brown, G., Winship, I., Barker, M., Buchanan, D., Cowie, S., Nasioulas, S., du Sart, D., Young, J., Leggett, B., Jass, J., Macrae, F. (2006). Hyperplastic polyposis syndrome: phenotypic presentations and the role of *MBD4* and *MYH*. *Gastroenterology* **131**:30-39

Colebatch, A., Hitchins, M., Williams, R., Meagher, A., Hawkins, N.J., Ward, R.L. (2006). The role of *MYH* and microsatellite instability in the development of sporadic colorectal cancer. *Br. J. Cancer* **95**:1239-1243

Colussi, C., Parlanti, E., Degan, P., Aquilina, G., Barnes, D., Macpherson, P., Karran, P., Crescenzi, M., Dogliotti, E., Bignami, M. (2002). The mammalian mismatch repair pathway removes DNA 8-oxodGMP incorporated from the oxidized dNTP pool. *Curr. Biol.* **12**:912-918

Cormier, R.T., Hong, K.H., Halberg, R.B., Hawkins, T.L., Richardson, P., Mulherkar, R., Dove, W.F., Lander, E.S. (1997). Secretory phospholipase Pla2g2a confers resistance to intestinal tumourigenesis. *Nat. Genet.* **17**:88-91

Cormier, R.T., Bilger, A., Lillich, A.J., Halberg, R.B., Hong, K.H., Gould, K.A., Borenstein, N., Lander, E.S., Dove, W.F. (2000). The *Mom1*^{AKR} intestinal tumor resistance region consists of Pla2g2a and a locus distal to *D4Mit64*. *Oncogene* **19**:3182-3192

Cortázar, D., Kunz, C., Saito, Y., Steinacher, R., Schär, P. (2007). The enigmatic thymine DNA glycosylase. *DNA Repair (Amst.)* **6**:489-504

Cortellino, S., Turner, D., Masciullo, V., Schepis, F., Albino, D., Daniel, R., Skalka, A.M., Meropol, N.J., Alberti, C., Larue, L., Bellacosa, A. (2003). The base excision repair enzyme MED1 mediates DNA damage response to antitumor drugs and is associated with mismatch repair system integrity. *Proc. Natl. Acad. Sci. U.S.A.* **100(25)**:15071-15076

Crabtree, M.D., Tomlinson, I.P.M., Hodgson, S.V., Neale, K., Phillips, R.K.S., Houlston, R.S. (2002). Explaining variation in familial adenomatous polyposis: relationship between genotype and phenotype and evidence for modifier genes. *Gut* **51**:420-423

Crabtree, M., Sieber, O.M., Lipton, L., Hodgson, S.V., Lamlum, H., Thomas, H.J.W., Neale, K., Phillips, R.K.S., Heinimann, K., Tomlinson, I.P.M. (2003). Refining the relation between 'first hits' and 'second hits' at the *APC* locus: the 'loose fit' model and evidence for differences in somatic mutation spectra among patients. *Oncogene* **22**:4257-4265

Crabtree, M.D., Fletcher, C., Churchman, M., Hodgson, S.V., Neale, K., Phillips, R.K.S., Tomlinson, I.P.M. (2004). Analysis of candidate modifier loci for the severity of colonic familial adenomatous polyposis, with evidence for the importance of the N-acetyl transferases. *Gut* **53**:271-276

Crespan, E., Hübscher, U., Maga, G. (2007). Error-free bypass of 2-hydroxyadenine by human DNA polymerase λ with Proliferating Cell Nuclear Antigen and Replication Protein A in different sequence contexts. *Nucleic Acids Res.* **35(15)**:5173-5181

Croitoru, M.E., Cleary, S.P., Berk, T., Di Nicola, N., Kopolovic, I., Bapat, B., Gallinger, S. (2007). Germline MYH mutations in a clinic-based series of Canadian multiple colorectal adenoma patients. *J Surg Oncol* **95(6)**:499-506.

Croitoru, M.E., Cleary, S.P., Di Nicola, N., Manno, M., Selander, T., Aronson, M., Redston, M., Cotterchio, M., Knight, J., Gryfe, R., Gallinger, S. (2004). Association between biallelic and monoallelic germline *MYH* gene mutations and colorectal cancer risk. *J. Natl. Cancer Inst.* **96(21)**:1631-1634

Davies, H., Bignell, G.R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M.J., Bottomley, W., Davis, N., Dicks, E., Ewing, R., Floyd, Y., Gray, K., Hall, S., Hawes, R., Hughes, J., Kosmidou, V., Menzies, A., Mould, C., Parker, A., Stevens, C., Watt, S., Hooper, S., Wilson, R., Jayatilake, H., Gusterson, B.A., Cooper, C.., Shipley, J., Hargrave, D., Pritchard-Jones, K., Maitland, N., Chenevix-Trench, G., Riggins, G.J., Bigner, D.D., Palmieri, G., Cossu, A., Flanagan, A., Nicholson, A., Ho, J.W.C., Leung, S.Y., Yuen, S.T., Weber, B.L., Seigler, H.F., Darrow, T.L., Paterson, H., Marais, R., Marshall, C.J., Wooster, R., Stratton, M.R., Futreal, P.A. (2002). Mutations of the *BRAF* gene in human cancer. *Nature* **417**:949-954

Deka, J., Herter, P., Sprenger-Haußels, M., Koosch, S., Franz, D., Müller, K-M., Kuhnen, C., Hoffmann, I., Müller, O. (1999). The APC protein binds to A/T rich DNA sequences. *Oncogene* **18**:5654-5661

de la Chapelle, A. (2004). Genetic predisposition to colorectal cancer. *Nat. Rev. Cancer* **4**:769-780

de la Chapelle, A. (2005). The incidence of Lynch syndrome. *Fam. Cancer* **4**:233-237

de la Fuente, M.K., Alvarez, K.P., Letelier, A.J., Bellolio, F., Acuña, M.L., León, F.S., Pinto, E., Carvallo, P., López-Köstner, F. (2007). Mutational screening of the *APC* gene in Chilean families with Familial Adenomatous Polyposis : nine novel truncating mutations. *Dis. Colon Rectum* **50(12)**:2142-2148

Delarue, M., Boule, J.B., Lescar, J., Expert-Bezancon, N., Jourdan, N., Sukumar, N., Rougeon, F., Papanicolaou, C. (2002). Crystal structures of a template-independent DNA polymerase: murine terminal deoxynucleotidyltransferase. *EMBO J.* **21**:427-439

Demple, B., Herman, T., Chen, D.S. (1991). Cloning and expression of APE, the cDNA encoding the major human apurinic endonuclease: definition of a family of DNA repair enzymes. *Proc. Natl. Acad. Sci. U.S.A.* **88**:11450-11454

De Rosa, M., Scarano, M.I., Panariello, L., Carlomagno, N., Rossi, G.B., Tempesta, A., Borgheresi, P., Renda, A., Izzo, P. (1999) Three

submicroscopic deletions at the APC locus and their rapid detection by quantitative-PCR analysis. *Eur. J. Hum. Genet.* **7**:695-703

De Rosa, M., Fasano, C., Panariello, L., Scarano, M.I., Belli, G., Iannelli, A., Ciciliano, F., Izzo, P. (2000). Evidence for a recessive inheritance of Turcot's syndrome caused by compound heterozygous mutations within the PMS2 gene. *Oncogene* **19**:1719-1723

De Rosa, M., Dourisboure, R.J., Morelli, G., Graziano, A., Gutiérrez, A., Thibodeau, S., Halling, K., Avila, K.C., Duraturo, F., Podesta, E.J., Izzo, P., Solano, A.R. (2004). First genotype characterization of Argentinean FAP patients: identification of 14 novel *APC* mutations. *Hum. Mutat.* **23(5)**:523-524.

De Rosa, M., Morelli, G., Cesaro, E., Duraturo, F., Turano, M., Rossi, G.B., Delrio, P., Izzo, P. (2007). Alternative splicing and nonsense-mediated mRNA decay in the regulation of a new adenomatous polyposis coli transcript. *Gene* **395**:8-14

De Vos, M., Hayward, B.E., Picton, S., Sheridan, E., Bonthron, D.T. (2004). Novel *PMS2* pseudogenes can conceal recessive mutations causing a distinctive childhood cancer syndrome. *Am. J. Hum. Genet.* **74**:954-964

Dherin, C., Gasparutto, D., O'Connor, T.R., Cadet, J., Boiteux, S. (2004). Excision by the human methylpurine DNA *N*-glycosylase of cyanuric acid, a stable and mutagenic oxidation product of 8-oxo-7,8-dihydroguanine. *Int. J. Radiat. Biol.* **80(1)**:21-27

Dietrich, W.F., Lander, E.S., Smith, J.S., Moser, A.R., Gould, K.A., Luongo, C., Borenstein, N., Dove, W. (1993). Genetic identification of *Mom-1*, a major modifier locus affecting *Min*-induced intestinal neoplasia in the mouse. *Cell* **75**:631-639

Di Gregorio, C., Frattini, M., Maffei, S., Ponti, G., Losi, L., Pedroni, M., Venesio, T., Bertario, L., Varesco, L., Risio, M., Ponz de Leon, M. (2006). Immunohistochemical expression of MYH protein can be used to identify patients with MYH-associated polyposis. *Gastroenterology* **131**:439-444

Dizdaroglu, M., Karakaya, A., Jaruga, P., Slupphaug, G., Krokan, H.E. (1996). Novel activities of human uracil DNA *N*-glycosylase for cytosine-derived products of oxidative DNA damage. *Nucleic Acids Res.* **24(3)**:418-422

Dobbie, Z., Müller, H., Scott, R.J. (1996a). Secretory phospholipase A_2 does not appear to be associated with phenotypic variation in familial adenomatous polyposis. *Hum. Genet.* **98**:386-390

Dobbie, Z., Spycher, M., Mary, J-L., Häner, M., Guldenschuh, I., Hürliman, R., Amman, R., Roth, J., Müller, H., Scott, R.J. (1996b). Correlation between

the development of extracolonic manifestations in FAP patients with mutations beyond codon 1403 in the *APC* gene. *J. Med. Genet.* **33**:274-280

Dolwani, S., Williams, G.T., West, K.P., Newman, J., Stock, D., Griffiths, A.P., Best, J., Cheale, J.P., Sampson, J.R. (2007). Analysis of inherited MYH/(MutYH) mutations in British Asian patients with colorectal cancer. *Gut* **56**:593

Dosanjh, M.K., Roy, R., Mitra, S., Singer, B. (1994). 1,N^6-ethenoadenine is preferred over 3-methyladenine as substrate by a cloned human *N*-methylpurine-DNA glycosylase (3-methyladenine-DNA glycosylase). *Biochemistry* **33**:1624-1628

Dou, H., Mitra, S., Hazra, T.K. (2003). Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2. *J. Biol. Chem.* **278(50)**:49679-49684

Doublié, S., Bandaru, V., Bond, J.P., Wallace, S.S. (2004). The crystal structure of human endonuclease VIII-like 1 (NEIL1) reveals a zincless finger motif required for glycosylase activity. *Proc. Natl. Acad. Sci. U.S.A.* **101(28)**:10284-10289

Eide, L., Luna, L., Gustad, E.C., Henderson, P.T., Essigmann, J.M., Demple, B., Seeberg, E. (2001). Human endonuclease III acts preferentially on DNA damage opposite guanine residues in DNA. *Biochemistry* **40**:6653-6659

Eliason, K., Hendrickson, B.C., Judkins, T., Norton, M., Leclair, B., Lyon, E., Ward, B., Noll, W., Scholl, T. (2005). The potential for increased clinical sensitivity in genetic testing for polyposis colorectal cancer through analysis of MYH mutations in North American patients. *J. Med. Genet.* **42**:95-96

Engelward, B.P., Weeda, G., Wyatt, M.D., Broekhof, J.L., de Wit, J., Donker, I., Allan, J.M., Gold, B., Hoeijmakers, J.H.J., Samson, L.D. (1997). Base excision repair deficient mice lacking the Aag alkyladenine DNA glycosylase. *Proc. Natl. Acad. Sci. U.S.A.* **94**:13087-13092

Enholm, S., Hienonen, T., Suomalainen, A., Lipton, L., Tomlinson, I., Karja, V., Eskelinen, M., Mecklin, J-P., Karhu, A., Jarvinen, H.J., Aaltonen, L.A. (2003). Proportion and phenotype of *MYH*-associated colorectal neoplasia in a population-based series of Finnish colorectal cancer patients. *Am. J. Path.* **163(3)**: 827-832

Erdmann, K.S., Kuhlmann, J., Lessmann, V., Herrmann, L., Eulenburg, V., Müller, O., Heumann, R. (2000). The Adenomatous Polyposis Coli protein (APC) interacts with the protein tyrosine phosphatase PTP-BL via an alternatively spliced PDZ domain. *Oncogene* **19**:3894-3901

Fan, W & Wu, X. (2004). DNA polymerase λ can elongate on DNA substrates mimicking non-homologous end joining and interact with XRCC4-ligase IV complex. *Biochem. Biophys. Res. Commun.* **323**:1328-1333

Farrington, S.M., Tenesa, A., Barnetson, R., Wiltshire, A., Prendergast, J., Porteous, H., Campbell, H., Dunlop, M.G. (2005). Germline susceptibility to colorectal cancer due to base-excision repair gene defects. *Am. J. Hum. Genet.* **77**:112-119

Fearnhead, N.S., Britton, M.P., Bodmer, W.F. (2001). The ABC of APC. *Hum. Mol. Genet.* **10(7)**:721-733

Fearnhead, N.S., Wilding, J.L., Winney, B., Tonks, S., Bartlett, S., Bicknell, D.C., Tomlinson, I.P.M., Mortensen, N.J.McC., Bodmer, W.F. (2004). Multiple rare variants in different genes account for multifactorial inherited susceptibility to colorectal adenomas. *Proc. Natl. Acad. Sci. U.S.A.***101(45)**: 15992-15997

Fearnhead, N.S., Winney, B., Bodmer, W.F. (2005). Rare variant hypothesis for multifactorial inheritance. *Cell Cycle* **4(4)**:521-525

Fiala, K.A., Duym, W.W., Zhang, J., Suo, Z. (2006). Up-regulation of the fidelity of human DNA polymerase λ by its non-enzymatic proline-rich domain. *J. Biol. Chem.* **281(28)**:19038-19044

Ficari, F., Cama, A., Valanzano, R., Curia, M.C., Palmirotta, R., Aceto, G., Esposito, D.L., Crognale, S., Lombardi, A., Messerini, L., Mariani-Costantini, R., Tonelli, F., Battista, P. (2000). *APC* gene mutations and colorectal adenomatosis in Familial Adenomatous Polyposis. *Br. J. Cancer* **82(2)**:348-353

Fischer, J.A., Muller-Weeks, S., Caradonna, S. (2004). Proteolytic degradation of the nuclear isoform of uracil-DNA glycosylase occurs during the S phase of the cell cycle. *DNA Repair (Amst.)* **3**:505-513

Fishel, R., Lescoe, M.K., Rao, M.R.S., Copeland, N.G., Jenkins, N.A., Garber, J., Kane, M., Kolodner, R. (1993). The human mutator gene homolog *MSH2* and its association with Hereditary nonpolyposis colon cancer. *Cell* **75**:1027-1038

Fishel, M.L., He, J., Smith, M.L., Kelley, M.R. (2007). Manipulation of base excision repair to sensitize ovarian cancer cells to alkylating agent temozolomide. *Clin. Cancer Res.* **13(1)**:260-267

Fleischmann, C., Peto, J., Cheadle, J., Shah, B., Sampson, J., Houlston, R.S. (2004). Comprehensive analysis of the contribution of germline *MYH* variation to early-onset colorectal cancer. *Int. J. Cancer* **109**:554-558

Fodde, R., van der Luijt, R., Tops, C., van der Klift, H., van Leeuwen-Cornelisse, I., Griffioen, G., Vasen, H., Khan, P.M. (1992). Eight novel inactivating germ line mutations at the *APC* gene identified by denaturing gradient gel electrophoresis. *Genomics* **13**:1162-1168

Fodde, R., Kuipers, J., Rosenberg, C., Smits, R., Kielman, M., Gaspar, C., van Es, J.H., Breukel, C., Wiegant, J., Giles, R.H., Clevers, H. (2001a). Mutations in the *APC* tumour suppressor gene cause chromosomal instability. *Nat. Cell Biol.* **3**:433-438

Fodde, R., Smits, R., Clevers, H. (2001b). *APC*, signal transduction and genetic instability in colorectal cancer. *Nat. Rev. Cancer* **1**:55-67

Fodde, R. & Brabletz, T. (2007). Wnt/β-catenin signalling in cancer stemness and malignant behaviour. *Curr. Opin. Cell Biol.* **19**:150-158

Fortini, P. & Dogliotti, E. (2007). Base damage and single-strand break repair: mechanisms and functional significance of short- and long-patch repair subpathways. *DNA Repair (Amst.)* **6**:398-409

Frayling, I.M., Beck, N.E., Ilyas, M., Dove-Edwin, I., Goodman, P., Pack, K., Bell, J.A., Williams, C.B., Hodgson, S.V., Thomas, H.J.W., Talbot, I.C., Bodmer, W.F., Tomlinson, I.P.M. (1998). The *APC* variants I1307K and E1317Q are associated with colorectal tumors, but not always with a family history. *Proc. Natl. Acad. Sci. U.S.A.* **95**:10722-10727

Frieaberg, E.C. & Meira, L.B. (2003). Database of mouse strains carrying targeted mutations in genes affecting biological responses to DNA damage version 5. *DNA Repair (Amst.)* **2**:501-530

Friedl, W., Caspari, R., Sengteller, M., Uhlhaas, S., Lamberti, C., Jungck, M., Kadmon, M., Wolf, M., Fahnenstich, J., Gebert, J., Möslein, G., Mangold, E., Propping, P. (2001). Can *APC* mutation analysis contribute to therapeutic decisions in familial adenomatous polyposis? Experience from 680 FAP families. *Gut* **48**:515-521

Fromme, J.C., Banerjee, A., Huang, S.J., Verdine, G.L. (2004). Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature* **427**:652-654

Fukushige, S., Kondo, E., Gu, Z., Suzuki, H., Horii, A. (2006). RET finger protein enhances MBD2- and MBD4-dependent transcriptional repression. *Biochem. Biophys. Res. Commun.* **351**:85-92

Fung, H. & Demple, B. (2005). A vital role for Ape1/Ref1 protein in repairing spontaneous DNA damage in human cells. *Mol. Cell* **17**:463-470

Furuichi, M., Yoshida, M.C., Oda, H., Tajiri, T., Nakabeppu, Y., Tsuzuki, T., Sekiguchi, M. (1994). Genomic structure and chromosome location of the human *mutT* homologue gene *MTH1* encoding 8-oxodGTPase for prevention of A:T to C:G transversion. *Genomics* **24**:485-490

Galiatsatos, P. & Foulkes, W.D. (2006). Familial Adenomatous Polyposis. *Am. J. Gastroenterol.* **101**:385-398

Gallinari, P & Jiricny, J. (1996). A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase. *Nature* **383**:735-738

García-Díaz, M., Domínguez, O., López-Fernández, L.A., de Lera, L.T., Saníger, M.L., Ruiz, .J.F., Párraga, M., García-Ortiz, M.J., Kirchhoff, T., del Mazo, J., Bernad, A., Blanco, L. (2000). DNA polymerase lambda (Pol λ), a novel eukaryotic DNA polymerase with a potential role in meiosis. *J. Mol. Biol.* **301**:851-867

García-Díaz, M., Bebenek, K., Kunkel, T.A., Blanco, L. (2001). Identification of an intrinsic 5'-deoxyribose-5-phosphate lyase activity in human DNA polymerase λ. *J. Biol. Chem.* **276(37)**:34659-34663

García-Díaz, M., Bebenek, Sabariegos, R., Domínguez, O., Rodríguez, J., Kirchhoff, T., García-Palomero, E., Picher, A.J., Juárez, R., Ruiz, J.F., Kunkel, T.A., Blanco, L. (2002). DNA polymerase λ, a novel DNA repair enzyme in human cells. *J. Biol. Chem.* **277(15)**:13184-13191

García-Díaz, M., Bebenek, K., Krahn, J.M., Blanco, L., Kunkel, T.A., Pedersen, L.C. (2004). A structural solution for the DNA polymerase λ-dependent repair of DNA gaps with minimal homology. *Mol. Cell* **13**:561-572

Gasmi, L., Cartwright, J.L., McLennan, A.G. (1999). Cloning, expression and characterization of YSA1H, a human adenosine 5'-diphosphosuagr pyrophosphatase possessing a MutT motif. *Biochem. J.* **344**:331-337

Gavert, N., Yaron, Y., Naiman, T., Bercovich, D., Rozen, P., Shomrat, R., Legum, C., Orr-Urtreger, A. (2002). Molecular analysis of the *APC* gene in 71 Israeli families: 17 novel mutations. *Hum. Mutat.* **19(6)**:664.

Giardiello, F.M., Krush, A.J., Petersen, G.M., Booker, S.V., Kerr, M., Tong, L.L., Hamilton, S.R. (1994). Phenotypic variability of familial adenomatous polyposis in 11 unrelated families with identical *APC* gene mutation. *Gastroenterology* **106**:1542-1547

Giarola, M., Stagi, L., Presciuttini, S., Mondini, P., Radice, M.T., Sala, P., Pierotti, M.A., Bertario, L., Radice, P. (1999). Screening for mutations of the *APC* gene in 66 Italian Familial Adenomatous Polyposis patients: evidence for phenotypic differences in cases with and without identified mutation. *Hum. Mutat.* **13**:116-123

Gismondi, V., Bonelli, L., Sciallero, S., Margiocco, P., Viel, A., Radice, P., Mondini, P., Sala, P., Montera, M.P., Mareni, C., Quaia, M., Fornasarig, M., Gentile, M., Pietro, G., Rossini, P., Arrigoni, A., Meucci, G.M., Bruzzi, P., Varesco, L. (2002). Prevalence of the E1317Q variant of the *APC* gene in Italian patients with colorectal adenomas. *Genet. Test.* **6(4)**:313-317

Gismondi, V., Meta, M., Bonelli, L., Radice, P., Sala, P., Bertario, L., Viel, A., Fornasarig, M., Arrigoni, A., Gentile, M., Ponz de Leon, M., Anselmi, L., Mareni, C., Bruzzi, P., Varesco, L. (2004). Prevalence of the Y165C, G382D

and 1395delGGA germline mutations of the *MYH* gene in Italian patients with adenomatous polyposis coli and colorectal adenomas. *Int. J. Cancer* **109**:680-684

Goel, A., Arnold, C.N., Niedzwiecki, D., Chang, D.K., Ricciardiello, L., Carethers, J.M., Dowell, J.M., Wasserman, L., Compton, C., Mayer, R.J., Bertagnolli, M.M., Boland, C.R. (2003). Characterization of sporadic colon cancer by patterns of genomic instability. *Cancer Res.* **63**:1608-1614

Görgens, H., Müller, A., Krüger, S., Kuhlisch, E., König, I.R., Ziegler, A., Schackert, H.K., Eckelt, U. (2007) Analysis of the base excision repair genes *MTH1*, *OGG1* and *MUTYH* in patients with squamous oral carcinomas. *Oral Oncol.* **43**:791-795

Grady, W.M. & Markowitz, S.D. (2002). Genetic and epigenetic alterations in colon cancer. *Annu. Rev. Genomics Hum. Genet.* **3**:101-128

Green, R.A. & Kaplan, K.B. (2003). Chromosome instability in colorectal tumour cells is associated with defects in microtubule plus-end attachments caused by a dominant mutation in *APC*. *J. Cell Biol.* **163(5)**:949-961

Groden, J., Thliveris, A., Samowitz, W., Carison, M., Gelbert, L., Albertsen, H., Joslyn, G., Stevens, J., Spirio, L., Robertson, M., Sargeant, L., Krapcho, K., Wolff, E., Burt, R., Hughes, J.P., Warrington, J., McPherson, J., Wasmuth, J., Le Paslier, D., Abderrahim, H., Cohen, D., Leppert, M., White, R. (1991). Identification and characterization of the Familial Adenomatous Polyposis Coli gene. *Cell* **66**:589-600

Gu, Y., Parker, A., Wilson, T.M., Bai, H., Chang, D-Y., Lu, A-L. (2002). Human MutY homolog, a DNA glycosylase involved in base excision repair, physically and functionally interacts with mismatch repair proteins human MutS homolog 2/human MutS homolog 6. *J. Biol. Chem.* **277(13)**:11135-11142

Guan, X., Madabushi, A., Chang, D-Y., Fitzgerald, M.E., Shi, G., Drohat, A.C., Lu, A-L. (2007). The human checkpoint sensor Rad9-Rad1-Hus1 interacts with and stimulates DNA repair enzyme TDG glycosylase. *Nucleic Acids Res.* **35(18)**:6207-6218

Guan, Y., Manuel, R.C., Arvai, A.S., Parikh, S.S., Mol, C.D., Miller, J.H., Lloyd, R.S., Tainer, J.A. (1998). MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily. *Nat. Struct. Biol.* **5(12)**:1058-1064

Guilford, P., Hopkins, J., Harraway, J., MeLeod, M., McLeod, N., Harawira, P., Taite, H., Scoular, R., Miller, A., Reeve, A.E. (1998) E-cadherin germline mutations in familial gastric cancer. *Nature* **392**:402-405

Hadi, M.Z., Coleman, M.A., Fidelis, K., Mohrenweiser, H.W., Wilson III, D.M. (2000). Functional characterization of Ape1 variants identified in the human population. *Nucleic Acids Res.* **28(20)**:3871-3879

Haiman, C.A., Le Marchand, L., Yamamato, J., O Stram, D., Sheng, X., Kolonel, L.N., Wu, A.H., Reich, D., Henderson, B.E. (2007). A common genetic risk factor for colorectal and prostate cancer. *Nat. Genet.* **39(8)**:954-956

Haines, J., Johnson, V., Pack, K., Suraweera, N., Slijepcevic, P., Cabuy, E., Coster, M., Ilyas, M., Wilding, J., Sieber, O., Bodmer, W., Tomlinson, I., Silver, A. (2005). Genetic basis of variation in adenoma multiplicity in $Apc^{Min/+}$ $Mom1^{s}$ mice. *Proc. Natl. Acad. Sci. U.S.A.* **102(8)**:2868-2873

Halford, S.E.R., Rowan, A.J., Lipton, L., Sieber, O.M., Pack, K., Thomas, H.J.W., Hodgson, S.V., Bodmer, W.F., Tomlinson, I.P.M. (2003). Germline mutations but not somatic changes at the *MYH* locus contribute to the pathogenesis of unselected colorectal cancers. *Am. J. Pathol.* **162(5)**:1545-1548

Hamada, F. & Bienz, M. (2004). The APC tumour suppressor binds to C-terminal binding protein to divert nuclear β-catenin from TCF. *Dev. Cell* **7**:677-685.

Hanson, C.A. & Miller, J.R. (2005). Non-traditional roles for the Adenomatous Polyposis Coli (APC) tumour suppressor protein. *Gene* **361**:1-12

Hao, B., Wang, H., Zhou, K., Li, Y., Chen, X., Zhou, G., Zhu, Y., Miao, X., Tan, W., Wei, Q., Lin, D., He, F. (2004). Identification of genetic variants in base excision repair pathway and their associations with risk of esophageal squamous cell carcinoma. *Cancer Res.* 64: 4378-4384

Hardeland, U., Bentele, M., Jiricny, J., Schär, P. (2000). Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *J. Biol. Chem.* **275(43)**:33449-33456

Hardeland, U., Steinacher, R., Jiricny, J., Schär, P. (2002). Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover. *EMBO J.* **21(6)**:1456-1464

Hardeland, U., Kunz, C., Focke, F., Szadkowski, M., Schär, P. (2007). Cell cycle regulation as a mechanism for functional separation of the apparently redundant uracil DNA glycosylases TDG and UNG2. *Nucleic Acids Res.* **35(11)**:3859-3867

Harrison, L., Ascione, G., Menninger, J.C., Ward, D.C., Demple, B. (1992). Human apurinic endonuclease gene (*APE*): structure and genomic mapping (chromosome 14q11.2-12). *Hum. Mol. Genet.* **1(9)**:677-680

Hart, M.J., de los Santos, R., Albert, I., Rubinfeld, B., Polakis, P. (1998). Downregulation of β-catenin by human axin and its association with the APC tumor suppressor, β-catenin and GSK3β. *Curr. Biol.* **8**:573-581

Hashimoto, K., Tominaga, Y., Nakabeppu, Y., Moriya, M. (2004). Futile short-patch DNA base excision repair of adenine:8-oxoguanine mispair. *Nucleic Acids Res.* **32(19)**:5928-5934

Haug, T., Skorphen, F., Kvaløy, K., Eftedal, I., Lund, H., Korkan, H.E. (1996). Human uracil-DNA glycosylase gene: sequence organization, methylation pattern, and mapping to chromosome 12q23-q24.1. *Genomics* **36**:408-416

Haug, T., Skorpen, F., Aas, P.A., Malm, V., Skjelbred, C., Krokan, H.E. (1998). Regulation of expression of nuclear and mitochondrial forms of human uracil-DNA glycosylase. *Nucleic Acids Res.* **26(6)**:1449-1457

Haushalter, K.A., Stukenberg, P.T., Kirschner, M.W., Verdine, G.L. (1999). Identification of a new uracil-DNA glycosylase family by expression cloning using synthetic inhibitors. *Curr. Biol.* **9**:174-185

Hayakawa, H., Taketomi, A., Sakumi, K., Kuwano, M., Sekiguchi, M. (1995). Generation and elimination of 8-oxo-7,8-dihydro-2'-deoxyguanosine 5'-triphosphate, a mutagenic substrate for DNA synthesis, in human cells. *Biochemistry* **34**:89-95

Hayakawa, H., Hofer, A., Thelander, L., Kitajima, S., Cai, Y., Oshiro, S., Yakushji, H., Nakabeppu, Y., Kuwano, M., Sekiguchi, M. (1999). Metabolic fate of oxidized guanine ribonucleotides in mammalian cells. *Biochemistry* **38**:3610-3614

Hayashi, H., Tominaga, Y., Hirano, S., McKenna, A.E., Nakabeppu, Y., Matsumoto, Y. (2002). Replication-associated repair of adenine:8-oxoguanine mispairs by MYH. *Curr. Biol* **12**:335-339

Hazra, T.K., Izumi, T., Maidt, L., Floyd, R.A., Mitra, S. (1998). The presence of two distinct 8-oxoguanine repair enzymes in human cells: their potential complementary roles in preventing mutation. *Nucleic Acids Res.* **26(22)**:5116-5122

Hazra, T.K., Hill, J.W., Izumi, T., Mitra, S. (2001). Multiple DNA glycosylases for repair of 8-oxoguanine and their potential *in vivo* functions. *Prog. Nucleic Acid Res. Mol. Biol.* **68**:193-205

Hazra, T.K., Izumi, T., Bolodogh, I., Imhoff, B., Kow, Y.W., Jaruga, P., Dizdaroglu, M., Mitra, S. (2002a). Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA. *Proc. Natl. Acad. Sci. U.S.A.* **99(6)**:3523-3528

Hazra, T.K., Kow, Y.W., Hatahet, Z., Imhoff, B., Boldogh, I., Mokkapati, S.K., Mitra, S., Izumi, T. (2002b). Identification and characterization of a novel human DNA glycosylase for repair of cytosine-derived lesions. *J. Biol. Chem.* **277(34)**:30417-30420

Hazra, T.K., Das, A., Das, S., Choudhury, S., Kow, Y.W., Roy, R. (2007). Oxidative DNA damage repair in mammalian cells: a new perspective. *DNA Repair (Amst.)* **6**:470-480

Heinimann, K., Thompson, A., Locher, A., Furlanetto, T., Bader, E., Wolf, A., Meier, R., Walter, K., Bauerfeind, P., Marra, G., Müller, H., Foernzler, D., Dobbie, Z. (2001). Nontruncating *APC* germline mutations and mismatch repair deficiency play a minor role in *APC* mutation-negative polyposis. *Cancer Res.* **61**:7616-7622

Heppner Goss, K., Trzepacz, C., Tuohy, T.M.F., Groden, J. (2002). Attenuated APC alleles produce functional protein from internal translation initiation. *Proc. Natl. Acad. Sci. U.S.A.* **99(12)**:8161-8166

Hendrich, B. & Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* **18(11)**:6538-6547

Hendrich, B., Abbott, C., McQueen, H., Chambers, D., Cross, S., Bird, A. (1999a). Genomic structure and chromosomal mapping of the murine and human *Mbd1*, *Mbd2*, *Mbd3* and *Mbd4* genes. *Mamm. Genome* **10**:906-912

Hendrich, B., Hardeland, U., Ng, H-H., Jiricny, J., Bird, A. (1999b). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* **401**:301-304

Hendriks, Y.M.C., Jagmohan-Changur, S., Van der Klift, H.M., Morreau, H., Van Puijenbroek, M., Tops, C., Van Os, T., Wagner, A., Ausems, M.G.F.M., Gomez, E., Breuning, M.H., Bröcker-Vriends, A.H.J.T., Vasen, H.F.A., Wijnen, J.TH. (2006). Heterozygous mutations in PMS2 cause Hereditary nonpolyposis colorectal carcinoma (Lynch syndrome). *Gastroenterology* **130**:312-322

Herman, J.G., Merlo, A., Mao, L., Lapidus, R.G., Issa, J-P.J., Davidson, N.E., Sidransky, D., Baylin, S.B. (1995). Inactivation of the *CDKN2/p16/MTS1* gene is frequently associated with aberrant DNA methylation in all common human cancers. *Cancer Res.* **55**:4525-4530

Herman, J.G., Umar, A., Polyak, K., Graff, J.R., Ahuja, N., Issa, J-P.J., Markowitz, S., Willson, J.K.V., Hamilton, S.R., Kinzler, K.W., Kane, M.F., Kolodner, R.D., Vogelstein, B., Kunkel, T.A., Baylin, S.B. (1998). Incidence and functional consequences of *hMLH1* promoter hypermethylation in colorectal carcinoma. *Proc. Natl. Acad. Sci. U.S.A.* **95**:6870-6875

Hes, F.J., Nielsen, M., Bik, E.C., Konvalinka, D., Wijnen, J.T., Bakker, E., Vasen, H.F.A., Breuning, M.H., Tops, C.M.J. (2008). Somatic APC mosaicism: an underestimated cause of polyposis coli. *Gut* **57(1)**:71-76.

Hioki, K., Shivapurkar, N., Oshima, H., Alabaster, O., Oshima, M., Taketo, M.M. (1997). Suppression of intestinal polyp development by low-fat and high-fibre diet in $Apc^{\Delta 716}$ knockout mice. *Carcinogenesis* **18(10)**:1863-1865

Hirano, S., Tominaga, Y., Ichinoe, A., Ushijima, Y., Tsuchimoto, D., Honda-Ohnishi, Y., Ohtsubo, T., Sakumi, K., Nakabeppu, Y. (2003). Mutator phenotype of MUTYH-null mouse embryonic stem cells. *J. Biol. Chem.* **278(40)**:38121-38124

Hitchins, M., Suter, C., Wong, J., Cheong, K., Hawkins, N., Leggett, B., Scott, R., Spigelman, A., Tomlinson, I., Martin, D., Ward, R. (2005). Germline epimutations of *APC* are not associated with inherited colorectal polyposis. *Gut* **122(6)**:1422-1425

Hoeijmakers, J.H.J. (2001). Genome maintenance mechanisms for preventing cancer. *Nature* **411**:366-374

Hollstein, M., Sidransky, D., Vogelstein, B., Harris, C.C. (1991). p53 mutations in human cancers. *Science* **253**:49-53

Horii, A., Nakatsuru, S., Ichii, S., Nagase, H., Nakamura, Y. (1993). Multiple forms of the *APC* gene transcripts and their tissue-specific expression. *Hum. Mol. Genet.* **2(3)**:283-287

Houlston, R.S. & Tomlinson, I.P.M. (2001). Polymorphisms and colorectal tumor risk. *Gastroenterology* **121**:282-301

Hutter, P., Rey-Berthod, C., Chappuis, P.O., Couturier, A., Membrez, V., Murphy, A., Joris, F., Schorderet, D.F., Delozier-Blanchet, C., Soravia, C. (2001). Molecular and clinical characteristics in 32 families affected with Familial Adenomatous Polyposis. *Hum. Mutat.* **18(6)**:550

Huycke, M.M. & Gaskins, H.R. (2004). Commensal bacteria, redox stress, and colorectal cancer: mechanisms and models. *Exp. Biol. Med. (Maywood)* **229**:586-597

Ikeda, S., Kishida, S., Yamamoto, H., Murai, H., Koyama, S., Kikuchi, A. (1998a). Axin, a negative regulator of the Wnt signalling pathway, forms a complex with GSK3β and β-catenin and promotes GSK-3β-dependent phosphorylation of β-catenin. *EMBO J.* **17(5)**:1371-1384

Ikeda, S., Biswas, T., Roy, R., Izumi, T., Boldogh, I., Kurosky, A., Sarker, A.H., Seki, S., Mitra, S. (1998b). Purification and characterization of human NTH1, a homolog of *Escherichia coli* endonuclease III. *J. Biol. Chem.* **273(34)**:21585-21593

Ikeda, S., Kishida, M., Matsuura, Y., Usui, H., Kikuchi, A. (2000). GSK-3β-dependent phosphorylation of adenomatous polyposis coli gene product can be modified by β-catenin and protein phosphatase 2A complexed with Axin. *Oncogene* **19**:537-545

Ikeda, S., Kohmoto, T., Tabata, R., Seki, Y. (2002). Differential intracellular localization of the human and mouse endonuclease III homologs and analysis of the sorting signals. *DNA Repair (Amst.)* **1**:847-854

Imai, K., Sarker, A.H., Akiyama, K., Ikeda, S., Yao, M., Tsutsui, K., Shohmori, T., Seki, S. (1998). Genomic structure of a human homologue (NTHL1/NTH1) of *Escherichia coli* endonuclease III with those of the adjacent parts of *TSC2* and *SLC9A3R2* genes. *Gene* **222**:287-295

Ionov, Y., Peinado, M.A., Malkhosyan, S., Shibata, D., Perucho, M. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* **363**:558-561

Ishibashi, T., Hayakawa, H., Sekiguchi, M. (2003). A novel mechanism for preventing mutations caused by oxidation of guanine nucleotides. *EMBO Rep.* **4(5)**:479-483

Ishibashi, T., Hayakawa, H., Ito, R., Miyazawa, M., Yamagata, Y., Sekiguchi, M. (2005). Mammalian enzymes for preventing transcriptional errors caused by oxidative damage. *Nucleic Acids Res.* **33(12)**:3779-3784

Isidro, G., Laranjeira, F., Pires, A., Leite, J., Regateiro, F., Castro e Sousa, F., Soares, J., Castro, C., Giria, J., Brito, M.J., Medeira, A., Teixeira, R., Morna, H., Gaspar, I., Marinho, C., Jorge, R., Brehm, A., Silva Ramos, J., Guida Boavida, M. (2004). Germline *MUTYH* (*MYH*) mutations in Portuguese individuals with multiple colorectal adenomas. *Hum. Mutat.* **24(4)**:353-354

Ito, R., Hayakawa, H., Sekiguchi, M., Ishibashi, T. (2005). *Biochemistry* **44**:6670-6674

Jaeger, E., Webb, E., Howarth, K., Carvajal-Carmona, L., Rowan, A., Broderick, P., Walther, A., Spain, S., Pittman, A., Kemp, Z., Sullivan, K., Heinimann, K., Lubbe, S., Domingo, E., Barclay, E., Martin, L., Gorman, M., Chandler, I., Vijayakrishnan, J., Wood, W., Papaemmanuil, E., Penegar, S., Qureshi, M., members of the CORGI Consortium, Farrington, S., Tenesa, A., Cazier, J-B., Kerr, D., Gray, R., Peto, J., Dunlop, M., Campbell, H., Thomas, H., Houlston, R., Tomlinson, I. (2008). Common genetics variants at the *CRAC1* (*HMPS*) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat. Genet.* **40(1)**:26-28

Jascur, T. & Boland, C.R. (2006). Structure and function of the components of the human DNA mismatch repair system. *Int. J. Cancer* **119**:2030-2035

Jayaraman, L., Murthy, K.G.K., Zhu, C., Curran, T., Xanthoudakis, S., Prives, C. (1997). Identification of redox/repair protein Ref-1 as a potent activator of p53. *Genes Dev.* **11**:558-570

Jemal, A., Murray, T., Ward, E., Samuels, A., Tiwari, R.C., Ghafoor, A., Feuer, E.J., Thun, M.J. (2005) Cancer statistics, 2005. *CA Cancer J. Clin.* **55(1)**:10-30

Jenkins, M.A., Croitoru, M.E., Monga, N., Cleary, S.P., Cotterchio, M., Hopper, J.L., Gallinger, S. (2006). Risk of colorectal cancer in monoallelic and biallelic carriers of *MYH* mutations: a population-based case-family study. *Cancer Epidemiol. Biomarkers Prev.* **15(2)**:312-314

Jimbo, T., Kawasaki, Y., Koyama, R., Sato, R., Takada, S., Haraguchi, K., Akiyama, T. (2002). Identification of a link between the tumour suppressor APC and the kinesin superfamily. *Nat. Cell Biol.* **4**:323-327

Jo, W-S., Bandipalliam, P., Shannon, K.M., Niendorf, K.B., Chan-Smutko, G., Hur, C., Syngal, S., Chung, D.C. (2005) Correlation of polyp number and family history of colon cancer with germline *MYH* mutations. *Clin. Gastroenterol. Hepatol.* **3**:1022-1028

Jones, S., Emmerson, P., Maynard, J., Best, J.M., Jordan, S., Williams, G.T., Sampson, J.R., Cheadle, J.P. (2002). Biallelic germline mutations in *MYH* predispose to multiple colorectal adenoma and somatic G:C→T:A mutations. *Hum. Mol. Genet.* **11(23)**:2961-2967

Jones, S., Lambert, S., Williams, G.T., Best, J.M., Sampson, J.R., Cheadle, J.P. (2004). Increased frequency of the k-*ras* G12C mutation in MYH polyposis colorectal adenomas. *Br. J. Cancer* **90**:1591-1593

Juwana, J-P., Henderikx, P., Mischo, A., Wadle, A., Fadle, N., Gerlach, K., Arends, J.W., Hoogenboom, H., Pfreundschuh, M., Renner, C. (1999). *EB/RP* gene famaily encodes tubulin binding proteins. *Int. J. Cancer* **81**:275-284

Kambara, T., Whitehall, V.L.J., Spring, K.J., Barker, M.A., Arnold, S., Wynter, C.V.A., Matsubara, N., Tanaka, N., Young, J.P., Leggett, B.A., Jass, J.R. (2004). Role of inherited defects of *MYH* in the development of sporadic colorectal cancer. *Genes Chromosomes Cancer* **40**:1-9

Kairupan, C.F., Meldrum, C.J., Crooks, R., Milward, E.A., Spigelman, A.D., Burgess, B., Groombridge, C., Kirk, J., Tucker, K., Ward, R., Williams, R., Scott, R.J. (2005). Mutation analysis of the *MYH* gene in an Australian series of colorectal polyposis patients with or without germline APC mutations. *Int. J. Cancer* **116**:73-77

Kanter-Smoler, G., Björk, J., Fritzell, K., Engwall, Y., Hallberg, B., Karlsson, G., Grönberg, H., Karlsson, P., Wallgren, A., Wahlström, J., Hultcrantz, R., Nordling, M. (2006). Novel findings in Swedish patients with *MYH*-

Associated Polyposis: mutation detection and clinical characterization. *Clin. Gastroenterol. Hepatol.* **4**:499-506

Kavli, B., Sundheim, O., Akbari, M., Otterlei, M., Nilsen, H., Skorpen, F., Aas, P.A., Hagen, L., Krokan, H.E., Slupphaug, G. (2002). hUNG2 is the major repair enzyme for removal of uracil from U:A mismatches, U:G mismatches and U in single-stranded DNA, with hSMUG1 as a broad specificity backup. *J. Biol. Chem.* **277(42)**:39926-39936

Kavli, B., Andersen, S., Otterlei, M., Liabakk, N.B., Imai, K., Fischer, A., Durandy, A., Krokan, H.E., Slupphaug, G. (2005). B cells from hyper-IgM patients carrying *UNG* mutations lack ability to remove uracil from ssDNA and have elevated genomic uracil. *J. Exp. Med.* **201(12)**:2011-2021

Kawasaki, Y., Sato, R., Akiyama, T. (2003). Mutated APC and Asef are involved in the migration of colorectal tumour cells. *Nat. Cell Biol.* **5**:211-215

Kemp, Z., Thirlwell, C., Sieber, O., Silver, A., Tomlinson, I. (2004). An update on the genetics of colorectal cancer. *Hum. Mol. Genet.* **13 Spec No 2**:R177-R185

Kim, C.J., Cho, Y.G., Park, C.H., Kim, S.Y., Nam, S.W., Lee, S.H., Yoo, N.J., Lee, J.Y., Park, W.S. (2004) Genetic alterations of the *MYH* gene in gastric cancer. *Oncogene* **23**:6820-6822

King, R.J.B. (2000). Cancer Biology. Chapter 7 Chemical and radiation carcinogenesis. Pearson Education Limited (Essex, UK)

Kinzler, K.W., Nilbert, M.C., Su, L-K., Vogelstein, B., Bryan, T.M., Levy, D.B., Smith, K.J., Preisinger, A.C., Hedge, P., McKechnie, D., Finniear, R., Markham, A., Groffen, J., Boguski, M.S., Altschul, S.F., Horii, A., Ando, H., Miyoshi, Y., Miki, Y., Nishisho, I., Nakamura, Y. (1991). Identification of FAP locus genes from chromosome 5q21. *Science* **253**:661-665

Kinzler, K.W. & Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. *Cell* **87**:159-170

Klungland, A. & Bjelland, S. (2007). Oxidative damage to purines in DNA: role of mammalian Ogg1. *DNA Repair (Amst.)* **6**:481-488

Kim, D-W., Kim, I-J., Kang, H.C., Park, H-W., Shin, Y., Park, J-H, Jang, S-G., Yoo, B.C., Lee, M.R., Hong, C.W., Park, K-J., Oh, N-G., Kim, N.K., Sung, M.K., Lee, B.W., Kim, Y.J., Lee, H., Park, J-G. (2005). Mutation spectrum of the *APC* gene in 83 Korean FAP families. *Hum. Mutat.* **26(3)**:281.

Kim, I-J., Ku, J-L., Kang, H.C., Park, J-H., Yoon, K-A., Shin, Y., Park, H-W., Jang, S.G., Lim, S-K., Han, S.Y., Shin, Y-K., Lee, M.R., Jeong, S-Y., Shin, H-R., Lee, J.S., Kim, W-H., Park, J-G. (2004). Mutational analysis of *OGG1*, *MYH*, *MTH1* in FAP, HNPCC and sporadic colorectal cancer patients:

R154H *OGG1* polymorphism is associated with sporadic colorectal cancer patients. *Hum. Genet.* **115**:498-503

Knudsen, A.L., Bisgaard, M.L., Bülow, S. (2003). Attenuated familial adenomatous polyposis (AFAP). A review of the literature. *Fam. Cancer* **2**:43-55

Knudson, A.G. (1996). Hereditary cancer: two hits revisited. *J. Cancer Res. Clin. Oncol.* **122**:135-140

Kohno, T., Shinmura, K., Tosaka, M., Tani, M., Kim, S-R., Sugimura, H., Nohmi, T., Kasai, H., Yokota, J. (1998). Genetic polymorphisms and alternative splicing of the *hOGG1* gene, that is involved in the repair of 8-hydroxyguanine in damaged DNA. *Oncogene* **16**:3219-3225

Kolonel, L.N., Altshuler, D., Henderson, B.E. (2004). The multiethnic cohort study: exploring genes, lifestyle and cancer risk. *Nat. Rev. Cancer* **4**:1-9

Kondo, E., Gu, Z., Horii, A., Fukushige, S. (2005). The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated *p16^{INK4}a* and *hMLH1* genes. *Mol. Cell. Biol.* **25(11)**:4388-4396

Kuismanen, S.A., Holmberg, M.T., Salovaara, R., de la Chapelle, A., Peltomäki, P. (2000) Genetic and epigenetic modification of *MLH1* accounts for a major share of microsatellite-unstable colorectal cancers. *Am. J. Path.* **156(5)**:1773-1779

Kumar, V.K.A., Gold, J.A., Mallon, E., Thomas, S., Hodgson, S.V. (2007). Sebaceous adenomas in an MYH associated polyposis patient of Indian (Gujarati) origin. *Fam. Cancer* DOI 10.1007/s10689-007-9161-9

Kunkel, T.A. & Erie, D.A. (2005). DNA mismatch repair. *Annu. Rev. Biochem.* **74**:681-710

Küry, S., Buecher, B., Robiou-du-Pont, S., Scoul, C., Coleman, H., Lelièvre, B., Olschwang, S., Le Houérou, C., Le Neel, T., Faroux, R., Ollivry, J., Lafraise, B., Chupin, L-D., Bézieau, S. (2007) The thorough screening of the MUTYH gene in a large French cohort of sporadic colorectal cancers. *Genet. Test.* **11(4)**:373-379

Laken, S.J., Petersen, G.M., Gruber, S.B., Oddoux, C., Ostrer, H., Giardiello, F.M., Hamilton, S.R., Hampel, H., Markowitz, A., Klimstra, D., Jhanwar, S., Winawer, S., Offit, K., Luce, M.C., Kinzler, K.W., Vogelstein, B. (1997). Familial colorectal cancer in Ashkenazim due to a hypermutable tract in *APC*. *Nat. Genet.* **17**:79-83

Laken, S.J., Papadopoulos, N., Petersen, G.M., Gruber, S.B., Hamilton, S.R., Giardiello, F.M., Brensinger, J.D., Vogelstein, B., Kinzler, K.W. (1999). Analysis of masked mutations in familial adenomatous polyposis. *Proc. Natl. Acad. Sci. U.S.A.* **96**:2322-2326

Lamlum, H., Ilyas, M., Rowan, A., Clark, S., Johnson, V., Bell, J., Frayling, I., Efstathiou, J., Pack, K., Payne, S., Roylance, R., Gorman, P., Sheer, D., Neale, K., Phillips, R., Talbot, I., Bodmer, W., Tomlinson, I. (1999). The type of somatic mutation at *APC* in familial adenomatous polyposis is determined by the site of the germline mutation: a new facet to Knudson's 'two-hit' hypothesis. *Nat. Med.* **5(9)**:1071-1075

Lamlum, H., Al-Tassan, N., Jaeger, E., Frayling, I., Sieber, O., Bin Reza, F., Eckert, M., Rowan, A., Barclay, E., Atkin, W., Williams, C., Gilbert, J., Cheadle, J., Bell, J., Houlston, R., Bodmer, W., Sampson, J., Tomlinson, I. (2000). Germline APC variants in patients with multiple colorectal adenomas, with evidence for the particular importance of E1317Q. *Hum. Mol. Genet.* **9(15)**:2215-2221

Lammi, L., Arte, S., Somer, M., Järvinen, H., Lahermo, P., Thesleff, I., Pirinen, S., Nieminen, P. (2004). Mutations in *AXIN2* cause familial tooth agenesis and predispose to colorectal cancer. *Am. J. Hum. Genet.* **74**:1043-1050

Larson, E.D., Iams, K., Drummond, J.T. (2003). Strand-specific processing of 8-oxoguanine by the human mismatch repair pathway: inefficient removal of 8-oxoguanine paired with adenine or cytosine. *DNA Repair (Amst.)* **2**:1199-1210

Lau, A.Y., Schärer, O.D., Samson, L., Verdine, G.L., Ellenberger, T. (1998). Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell* **95**:249-258

Laurent-Puig, P., Béroud, C., Soussi, T. (1998). *APC* gene: database of germline and somatic mutations in human tumours and cell lines. *Nucleic Acids Res.* **26(1)**:269-270

Lee, C-Y., Bai, H., Houle, R., Wilson, G.M., Lu, A-L. (2004). An Escherichia coli MutY mutant without the six-helix barrel domain is a dimer in solution and assembles cooperatively into multisubunit complexes with DNA. *J. Biol. Chem.* **279(50)**:52653-52663

Lee, J.W., Blanco, L., Zhou, T., Garcia-Diaz, M., Bebenek, K., Kunkel, T.A., Wang, Z., Oovirk, L.F. (2004). Implications of DNA polymerase λ in alignment-based gap filling for nonhomologous DNA end joining in human nuclear extracts. *J. Biol. Chem.* **279(1)**:805-811

Lefevre, J.H., Rodrigue, C.M., Mourra, N., Bennis, M., Flejou, J-F., Parc, R., Tiret, E., Gespach, C., Parc, Y.R. (2006). Implication of MYH in colorectal polyposis. *Ann. Surg.* **244(6)**:874-880

Leggett, B.A., Devereaux, B., Biden, K., Searle, J., Young, J., Jass, J. (2001) Hyperplastic polyposis. *Am. J. Surg. Path.* **25(2)**:177-184

Lejeune, S., Guillemot, F., Triboulet, J-P., Cattan, S., Mouton, C., PANFORD Group, Porchet, N., Manouvrier, S., Buisine, M-P. (2006). Low frequency of *AXIN2* mutations and high frequency of *MUTYH* mutations in patients with multiple polyposis. *Hum. Mutat.* **27(10)**:1064

Levin, J.D. & Demple, B. (1990). Analysis of class II (hydrolytic) and class I (β-lyase) apurinic/apyrimidinic endonucleases with a synthetic DNA substrate. *Nucleic Acids Res.***18(17)**:5069-5075

Li, X., Wright, P.M., Lu, A-L. (2000). The C-terminal domain of MutY glycosylase determines the 7,8-dihydro-8-oxo-guanine specificity and is crucial for mutation avoidance. *J. Biol. Chem.* **275(12)**:8448-8455

Li, Z. & Näthke, I.S. (2005). Tumor-associated NH_2-terminal fragments are the most stable part of the Adenomatous Polyposis Coli protein and can be regulated by interactions with COOH-terminal domains. *Cancer Res.* **65(12)**:5195-5204

Lichtenstein, P., Holm, N.V., Verkasalo, P.K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A., Hemminki, K. (2000). Environmental and heritable factors in the causation of cancer. *N. Engl. J. Med.* **343(2)**:78-85

Lieber, M.R., Ma, Y., Pannicke, U., Schwarz, K. (2003). Mechanisms and regulation of human non-homologous DNA end-joining. *Nat. Rev. Mol. Cell Biol.* **4**:712-720

Lindblom, A., Tannergård, P., Werelius, B., Nordenskjöld, M. (1993). Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. *Nat. Genet.* **5**:279-282

Lipkin, S.M., Rozek, L.S., Rennert, G., Yang, W., Chen, P-C., Hacia, J., Hunt, N., Shin, B., Fodor, S., Kokoris, M., Greenson, J.K., Fearon, E., Lynch, H., Collins, F., Gruber, S.B. (2004). The MLH1 D132H variant is associated with susceptibility to sporadic colorectal cancer. *Nat. Genet.* **36(7)**:694-699

Lipton, L., Halford, S.E., Johnson, V., Novelli, M.R., Jones, A., Cummings, C., Barclay, E., Sieber, O., Sadat, A., Bisgaard, M-L., Hodgson, S.V., Aaltonen, L.A., Thomas, H.J.W., Tomlinson, I.P.M. (2003). Carcinogenesis in MYH-associated polyposis follows a distinct genetic pathway. *Cancer Res.* **63**:7595-7599

Liu, P., Burdzy, A., Sowers, L.C. (2003). Repair of the mutagenic DNA oxidation product, 5-formyluracil. *DNA Repair (Amst.)* **2**:199-210

Lu, A-L., Li, X., Gu, Y., Wright, P.M., Chang, D-Y. (2001). Repair of oxidative DNA damage. *Cell Biochem. Biophys.* **35**:141-170

Lu, A-L., Tsai-Wu, J-J., Cillo, J. (1995). DNA determinants and substrate specificities of *Escherichia coli* MutY. *J. Biol. Chem.* 270(40): 23582-23588

Lucci-Cordisco, E., Zito, I., Gensini, F., Genuardi, M. (2003). Hereditary Nonpolyposis Colorectal Cancer and related conditions. *Am. J. Med. Genet.* **122A**:325-334

Lucey, M.J., Chen, D., Lopez-Garcia, J., Hart, S.M., Phoenix, F., Al-Jehani, R., Alao, J.P., White, R., Kindle, K.B., Losson, R., Chambon, P., Parker, M.G., Schär, P., Heery, D.M., Buluwela, L., Ali, S. (2005). T:G mismatch-specific thymine-DNA glycosylase (TDG) as a coregulator of transcription interacts with SRC1 family members through a novel tyrosine repeat motif. *Nucleic Acids Res.* **33(19)**:6393-6404

Luna, L., Bjørås, M., Hoff, E., Rognes, T., Seeberg, E. (2000). Cell cycle regulation, intracellular sorting and induced overexpression of the human NTH1 DNA glycosylase involved in removal of formamidopyrimidine residues from DNA. *Mutat. Res.* **460**:95-104

Lynch, H.T. & de la Chapelle, A. (2003). Hereditary colorectal cancer. *N. Engl. J. Med.* **348**:919-932

Lynch, H.T., Boland, C.R., Gong, G., Shaw, T.G., Lynch, P.M., Fodde, R., Lynch, J.F., de la Chapelle, A. (2006). Phenotypic and genotypic heterogeneity in the Lynch syndrome: diagnostic, surveillance and management implications. *Eur. J. Hum. Genet.* **14**:390-402

Macpherson, P., Barone, F., Maga, G., Mazzei, F., Karran, P., Bignami, M. (2005). 8-oxoguanine incorporation into DNA repeats *in vitro* and mismatch recognition by MutSα. *Nucleic Acids Res.* **33(16)**:5094-5105

Maga, G., Villani, G., Ramadan, K., Shevelev, I., Tanguy le Gac, N., Blanco, L., Blanca, G., Spadari, S., Hübscher, U. (2002). Human DNA polymerase λ functionally and physically interacts with proliferating cell nuclear antigen in normal and translesion synthesis. *J. Biol. Chem.* **277(50)**:48434-48440

Maga, G., Ramadan, K., Locatelli, G.A., Shevelev, I., Spadari, S., Hübscher, U. (2005). DNA elongation by the human DNA polymerase λ polymerase and terminal transferase activities are differentially coordinated by proliferating cell nuclear antigen and replication protein A. *J. Biol. Chem.* **280(3)**:1971-1981

Maga, G., Shevelev, I., Villani, G., Spadari, S., Hübscher, U. (2006). Human replication protein A can suppress the intrinsic *in vitro* mutator phenotype of human DNA polymerase λ. *Nucleic Acids Res.* **34(5)**:1405-1415

Maga, G., Villani, G., Crespan, E., Wimmer, U., Ferrari, E., Bertocci, B., Hübscher, U. (2007). 8-oxo-guanine bypass by human DNA polymerases in the presence of auxiliary proteins. *Nature* **447**:606-608

Magdinier, F. & Wolffe, A.P. (2001). Selective association of the methyl-CpG binding protein MBD2 with the silent *p14/p16* locus in human neoplasia. *Proc. Natl. Acad. Sci. U.S.A.* **98(9)**:4990-4995

Mahmoud, N.N., Boolbol, S.K., Bilinski, R.T., Martucci, C., Chadburn, A., Bertagnolli, M.M. (1997). *APC* gene mutation is associated with a dominant-negative effect upon intestinal cell migration. *Cancer Res.* **57**:5045-5050

Majumder, S., Kutay, H., Datta, J., Summers, D., Jacob, S.T., Ghoshal, K. (2006). Epigenetic regulation of metallothionein-I gene expression: differential regulation of methylated and unmethylated promoters by DNA methyltransferases and methyl CpG binding proteins. *J. Cell. Biochem.* **97**:1300-1316

Maki, H. & Sekiguchi, M. (1992). MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. *Nature* **355**:273-275

Mandl, M., Caspari, R., Jauch, A., Böker, T., Raschke, H., Sengteller, M., Propping, P., Friedl, W. (1996). Familial Adenomatous Polyposis: a submicroscopic deletion at the *APC* locus in a family with mentally normal patients. *Hum. Genet.* **97**:204-208

Manuel, R.C., Hitomi, K., Arvai, A.S., House, P.G., Kurtz, A.J., Dodson, M.L., McCullough, A.K., Tainer, J.A., Lloyd, R.S. (2004). Reaction intermediates in the catalytic mechanism of *Escherichia coli* MutY DNA glycosylase. *J. Biol. Chem.* **279(45)**:46930-46939

Markowitz, S., Wang, J., Myeroff, L., Parsons, R., Sun, L., Lutterbaugh, J., Fan, R.S., Zborowska, E., Kinzler, K., Vogelstein, B., Brattain, M., Willson, J.K.V. (1995). Inactivation of the type II TGF-β receptor in colon cancer cells with microsatellite instability. *Science* **268**:1336-1338

Masaoka, A., Matsbara, M., Tanaka, T., Terato, H., Phyama, Y., Kubo, K., Ide, H. (2003). Repair roles of hSMUG1 assessed by damage specificity and cellular activity. *Nucleic Acids Res. Suppl.* **3**:263-264

Matsumoto, Y., Zhang, Q-M., Takao, M., Yasui, A., Yonei, S. (2001). *Escherichia coli* Nth and human hNTH1 DNA glycosylases are involved in removal of 8-oxoguanine from 8-oxoguanine/guanine mispairs in DNA. *Nucleic Acids Res.* **29(9)**:1975-1981

McCann, J.A.B. & Berti, P.J. (2003). Adenine release is fast in MutY-catalyzed hydrolysis of G:A and 8-oxo-G:A DNA mismatches. *J. Biol. Chem.* **278(32)**:29587-29592

McLennan, A.G. (2006). The Nudix hydrolase superfamily. *Cell. Mol. Life Sci.* **63**:123-143

Mecklin, J-P., Aarino, M., Läärä, E., Kairaluoma, M.V., Pylvänäinen, K., Peltomäki, P., Aaltonen, L.A., Järvinen, H.J. (2007). Development of

colorectal tumours in colonoscopic surveillance in Lynch Syndrome. *Gastroenterology* **133**:1093-1098

Menéndez, M., González, S., Obrador-Hevia, A., Domínguez, A., Jesús Pujol, M., Valls, J., Canela, N., Blanco, I., Torres, A., Pineda-Lucena, A., Moreno, V., Bachs, O., Capellá, G. (2008). Functional characterization of the novel *APC* N1026S variant associated with Attenuated Familial Adenomatous Polyposis. *Gastroenterology* **134**:56-64

Menko, F.H. (1993). Genetics of colorectal cancer for clinical practice. Chapter 2 General aspects of colorectal cancer. Kluwer Academic Publishers (The Netherlands)

Michaels, M.L., Cruz, C., Grollman, A.P., Miller, J.H. (1992). Evidence that MutY and MutM combine to prevent mutations by an oxidatively damaged form of guanine in DNA. *Proc. Natl. Acad. Sci. U.S.A.* **89**:7022-7025

Michaels, M.L. & Miller, J.H. (1992). The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). *J. Bacteriol.* **174(20)**: 6321-6325

Michils, G., Tejpar, S., Thoelen, R., van Custem, E., Vermeesch, J.R., Fryns, J-P., Legius, E., Matthijs, G. (2005). Large deletions of the *APC* gene in 15% of mutation-negative patients with classical polyposis (FAP): a Belgian study. *Hum. Mutat.* **25**: 125-134

Millar, C.B., Guy, J., Sansom, O.J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P.D., Bishop, S.M., Clarke, A.R., Bird, A. (2002). Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* **297**: 403-405

Minowa, O., Arai, T., Hirano, M., Monden, Y., Nakai, S., Fukuda, M., Itoh, M., Takano, H., Hippou, Y., Aburatani, H., Masumura, K-I., Nohmi, T., Nishimura, S., Noda, T. (2000). *Mmh/Ogg1* gene inactivation results in accumulation of 8-hydroxyguanine in mice. *Proc. Natl. Acad. Sci. U.S.A.* **97(8)**: 4156-4161

Miyaki, M., Konishi, M., Tanaka, K., Kikuchi-Yanoshita, R., Muraoka, M., Yasuno, M., Igari, T., Koike, M., Chiba, M., Mori, T. (1997). Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nat. Genet.* **17**: 271-272

Miyaki, M., Iijima, T., Yamaguchi, T., Hishima, T., Tamura, K., Utsunomiya, J., Mori, T. (2005). Germline mutations of the *MYH* gene in Japanese patients with multiple colorectal adenomas. *Mutat. Res.* **578**:430-433

Miyoshi, Y., Ando, H., Nagase, H., Nishisho, I., Horii, A., Miki, Y., Mori, T., Utsunomiya, J., Baba, S., Petersen, G., Hamilton, S.R., Kinzler, K.W., Vogelstein, B., Nakamura, Y. (1992) Germ-line mutations of the *APC* gene

in 53 familial adenomatous polyposis patients. *Proc. Natl. Acad. Sci. U.S.A.* **89**: 4452-4456

Mo, J-Y., Maki, H., Sekiguchi, M. (1992). Hydrolytic elimination of a mutagenic nucleotide, 8-oxodGTP, by human 18-kilodalton protein: sanitization of nucleotide pool. *Proc. Natl. Acad. Sci. U.S.A.* **89**: 11021-11025

Mohan, R.D., Rao, A., Gagliardi, J., Tini, M. (2007). SUMO-1-dependent allosteric regulation of thymine DNA glycosylase alters subnuclear localization and CBP/p300 recruitment. *Mol. Cell. Biol.* **27(1)**: 229-243

Moisio, A-L., Järvinen, H., Peltomäki, P. (2002) Genetic and clinical characterisation of familial adenomatous polyposis: a population base study. *Gut* **50**: 845-850

Morin, P.J., Vogelstein, B., Kinzler, K.W. (1996). Apoptosis and APC in colorectal tumourigenesis. *Proc. Natl. Acad. Sci. U.S.A.* **93**: 7950-7954

Morin, P.J., Sparks, A.B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B., Kinzler, K.W. (1997). Activation of β-catenin-Tcf signalling in colon cancer by mutations in β-catenin or APC. *Science* **275**: 1787-1790

Morland, I., Rolseth, V., Luna, L., Rognes, T., Bjørås, M., Seeberg, E. (2002). Human DNA glycosylases of the bacterial Fpg/MutM superfamily: an alternative pathway for the repair of 8-oxoguanine and other oxidation products in DNA. *Nucleic Acids Res.* **30(22)**: 4926-4936

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. (1992) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. 1986. *Biotechnology,* 24, 17-27.

Nagasawa, K-I., Kitamura, K., Yasui, A., Nimura, Y., Ikeda, K., Hirai, M., Matsukage, A., Nakanishi, M. (2000). Identification and characterization of human DNA polymerase β2, a DNA polymerase β-related enzyme. *J. Biol. Chem.* **275(40)**: 31233-31238

Nagase, H., Miyoshi, Y., Horii, A., Aoki, T., Petersen, G.M., Vogelstein, B., Maher, E., Ogawa, M., Maruyama, M., Utsunomiya, J., Baba, S., Nakamura, Y. (1992a) Screening for germ-line mutations in familial adenomatous polyposis patients: 61 new patients and a summary of 150 unrelated patients. *Hum. Mut.* **1**: 467-473

Nagase, H., Miyoshi, Y., Horii, A., Aoki, T., Ogawa, M., Utsunomiya, J., Baba, S., Sasazuki, T., Nakamura, Y. (1992b). Correlation between the location of germ-line mutations in the *APC* gene and the number of colorectal polyps in Familial Adenomatous Polyposis patients. *Cancer Res.* **52**: 4055-4057

Nagy, R., Sweet, K., Eng, C. (2004). Highly penetrant hereditary cancer syndromes. *Oncogene* **23**: 6445-6470

Nash, H.M., Bruner, S.D., Schärer, O.D., Kawate, T., Addona, T.A., Spooner, E., Lane, W.S., Verdine, G.L. (1996). Cloning of a yeast 8-oxoguanine DNA glycosylase reveals the existence of a base-excision DNA –repair superfamily. *Curr. Biol.* **6(8)**: 968-980

Nassif, N.T., Lobo, G.P., Wu, X., Henderson, C.J.A., Morrison, C.D., Eng, C., Jalaludin, B., Segelov, E. (2004). *PTEN* mutations are common in sporadic microsatellite stable colorectal cancer. *Oncogene* **23**:617-628

Näthke, I.S. (2004). The Adenomatous Polyposis Coli protein: the Achilles heel of the gut epithelium. *Annu. Rev. Cell Dev. Biol.* **20**: 337-366

Neddermann, P. & Jiricny, J. (1994). Efficient removal of uracil from G·U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells. *Proc. Natl. Acad. Sci. U.S.A.* **91**: 1642-1646

Neddermann, P., Gallinari, P., Lettieri, T., Schmid, D., Truong, O., Hsuan, J.J., Wiebauer, K., Jiricny, J. (1996). Cloning and expression of human G/T mismatch-specific thymine DNA glycosylase. *J. Biol. Chem.* **271(22)**: 12767-12774

Nelson, D.L. & Cox, M.M. (2000). Lehninger Principles of Biochemistry, 3rd edition. Chapter 5 Amino acids, peptides and proteins. Worth Publishers (USA).

Nicolaides, N.C., Papadopoulos, N., Liu, B., Wei, Y.F., Carter, K.C., Ruben, S.M., Rosen, C.A., Haseltine, W.A., Fleischmann, R.D., Fraser, C.M., Adams, M.D., Venter, J.C., Dunlop, M.G., Hamilton, S.R., Petersen, G.M., de la Chapelle, A., Vogelstein, B., Kinzler, K. (1994). Mutations of two *PMS* homologues in hereditary nonpolyposis colon cancer. *Nature* **371**:75-80

Niederreither, K., Harbers, M., Chambon, P., Dollé (1998). Expression of T:G mismatch-specific thymidine-DNA glycosylase and DNA methyl transferase genes during development and tumorigenesis. *Oncogene* **17**: 1577-1585

Nielsen, M., Franken, P.F., Reinards, T.H.C.M., Weiss, M.M., Wagner, A., van der Klift, H., Kloosterman, S., Houwing-Duistermaat, J.J., Aalfs, C.M., Ausems, M.G.E.M., Bröcker-Vriends, A.H.J.T., Gomez Garcia, E.B., Hoogerbrugge, N., Menko, F.H., Sijmons, R.H., Verhoef, S., Kuipers, E.J., Morreau, H., Breuning, M.H., Tops, C.M.J., Wijnen, J.T., Vasen, H.F.A., Fodde, R., Hes, F.J. (2005). Multiplicity in polyp count and extracolonic manifestations in 40 Dutch patients with *MYH* associated polyposis coli (MAP) *J. Med. Genet.* **42**:e54

Nielsen, M., Hes, F.J., Nagengast, F.M., Weiss, M.M., Mathus-Vliegen, E.M., Morreau, H., Breuning, M.H., Wijnen, J.T., Tops, C.M.J., Vasen,

H.F.A. (2007a) Germline mutations in *APC* and *MUTYH* are responsible for the majority of families with attenuated familial adenomatous polyposis. *Clin. Genet.* **71**: 427-433

Nielsen, M., Bik, E., Hes, F.J., Breuning, M.H., Vasen, H.F.A., Bakker, E., Tops, C.M.J., Weiss, M.M. (2007b) Genotype-phenotype correlations in 19 Dutch cases with APC gene deletions and a literature review. *Eur. J. Hum. Genet.* **15(10)**:1034-1042.

Nieuwenhuis, M.H., Mathus-Vliegen, L.M., Slors, F.J., Griffioen, G., Nagengast, F.M., Schouten, W.R., Kleibeuker, J.H., Vasen, H.F.A. (2007). Genotype-phenotype correlations as a guide in the management of Familial Adenomatous Polyposis. *Clin. Gastroenterol. Hepatol.* **5**:374-378

Nieuwenhuis, M.H. & Vasen, H.F.A. (2007). Correlations between mutation site in APC and phenotype of Familial Adenomatous Polyposis (FAP): a review of the literature. *Crit. Rev. Oncol. Hematol.* **61**: 153-161

Nilsen, H., Otterlei, M., Haug, T., Solum, K., Nagelhus, T.A., Skorpen, F., Krokan, H.E. (1997). Nuclear and mitochondrial uracil-DNA glycosylases are generated by alternative splicing and transcription from different positions in the *UNG* gene. *Nucleic Acids Res.* **25(4)**: 750-755

Nilsen, H., Rosewell, I., Robins, P., Skjelbred, C.F., Andersen, S., Slupphaug, G., Daly, G., Krokan, H.E., Lindahl, T., Barnes, D.E. (2000). Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication. *Mol. Cell* **5**: 1059-1065

Nishioka, K., Ohtsubo, T., Oda, H., Fujiwara, T., Kang, D., Sugimachi, K., Nakabeppu, Y. (1999). Expression and differential intracellular localization of two major forms of human 8-oxoguanine DNA glycosylase encoded by alternatively spliced OGG1 mRNAs. *Mol. Biol. Cell* **10**: 1636-1652

Noll, D.M., Gogos, A., Granek, J.A., Clarke, N.D. (1999). The C-terminal domain of the adenine-DNA glycosylase MutY confers specificity for 8-oxoguanine·adenine mispairs and may have evolved from MutT, an 8-oxo-dGTPase. *Biochemistry* **38**: 6374-6379

Oda, H., Nakabeppu, Y., Furuichi, M., Sekiguchi, M. (1997). Regulation of expression of the human *MTH1* gene encoding 8-oxo-dGTPase. *J. Biol. Chem.* **272(28)**: 17843-17850

Ohtsubo, T., Nishioka, K., Imaiso, Y., Iwai, S., Shimokawa, H., Oda, H., Fujiwara, T., Nakabeppu, Y. (2000). Identification of human MutY homolog (hMYH) as a repair enzyme for 2-hydroxyadenine in DNA and detection of multiple forms of hMYH located in nuclei and mitochondria. *Nucleic Acids Res.* **28(6)**: 1355-1364

Oka, T., Takayama, T., Sato, Y., Sato, Y., Takada, K., Hayashi, T., Takahashi, M., Kuroda, M., Kato, J., Niitsu, Y. (2004). A case of Gardner

syndrome with a mutation at codon 1556 of *APC*: a suggested case of genotype-phenotype correlation in dental abnormality. *Eur J Gastroenterol Hepatol* **16**: 101-105

Olschwang, S., Laurent-Puig, P., Groden, J., White, R., Thomas, G. (1993). Germ-line mutations in the first 14 exons of the Adenomatous Polyposis Coli (*APC*) gene. *Am. J. Hum. Genet.* **52**: 273-279

Olschwang, S., Blanché, H., de Moncuit, C., Thomas, G. (2007). Similar colorectal cancer risk in patients with monoallelic and biallelic mutations in the *MYH* gene identified in a population with adenomatous polyposis. *Genet. Test.* **11(3)**: 315-320

Oshima, M., Sugiyama, H., Kitagawa, K., Taketo, M. (1993). *APC* messenger RNA: novel isoforms that lack exon 7. *Cancer Res.* **53**: 5589-5591

Otterlei, M., Warbrick, E., Nagelhus, T.A., Haug, T., Slupphaug, G., Akbari, M., Aas, P.A., Steinsbekk, K., Bakke, O., Krokan, H.E. (1999). Post-replicative base excision repair in replication foci. *EMBO J.* **18(13)**: 3834-3844

Parikh, S.S., Mol, C.D., Slupphaug, G., Bharati, S., Krokan, H.E., Tainer, J.A. (1998). Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J.* **17(17)**: 5214-5226

Parker, A., Gu, Y., Mahoney, W., Lee, S-H., Singh, K.K., Lu, A-L. (2001). Human homolog of the MutY repair protein (hMYH) physically interacts with proteins involved in long patch DNA base excision repair. *J. Biol. Chem.* **276(8)**: 5547-5555.

Parker, A.R., O'Meally, R.N., Sahin, F., Su, G.H., Racke, F.K., Nelson, W.G., DeWeese, T.L., Eshleman, J.R. (2003). Defective human MutY phosphorylation exists in colorectal cancer cell lines with wild-type MutY alleles. *J. Biol. Chem.* **278(48)**: 47937-47945

Parker, A.R., Sieber, O.M., Shi, C., Hua, L., Takao, M., Tomlinson, I.P., Eshleman, J.R. (2005). Cells with pathogenic biallelic mutations in the human *MUTYH* gene are defective in DNA damage binding and repair. *Carcinogenesis* **26(11)**: 2010-2018

Paul, P., Letteboer, T., Gelbert, L., Groden, J., White, R., Coppes, M.J. (1993). Identical *APC* exon 15 mutations result in a variable phenotype in familial adenomatous polyposis. *Hum. Mol. Genet.* **2(7)**: 925-931

Pearl, L.H. (2000). Structure and function in the uracil-DNA glycosylase superfamily. *Mutat. Res.* **460**: 165-181

Peltomäki, P. (2005). Lynch syndrome genes. *Fam. Cancer* 4: 227-232

Pendlebury, A., Frayling, I.M., Santibanez Koref, M.F., Margison, G.P., Rafferty, J.A. (1994). Evidence for the simultaneous expression of alternatively spliced alkylpurine *N*-glycosylase transcripts in human tissues and cells. *Carcinogenesis* **15(12)**: 2957-2960

Peterlongo, P., Mitra, N., Chuai, S., Kirchhoff, T., Palmer, C., Huang, H., Nafa, K., Offit, K., Ellis, N.A. (2005). Colorectal cancer risk in individuals with biallelic or monoallelic mutations of *MYH*. *Int. J. Cancer* **114**:505-507

Peterlongo, P., Mitra, N., Sanchez de Abajo, A., de la Hoya, M., Bassi, C., Bertario, L., Radice, P., Glogowski, E., Nafa, K., Caldes, T., Offit, K., Ellis, N.A. (2006). Increased frequency of disease-causing *MYH* mutations in colon cancer families. *Carcinogenesis* **27(11)**: 2243-2249

Petronzelli, F., Riccio, A., Markham, G.D., Seeholzer, S.H., Stoerker, J., Genuardi, M., Yeung, A.T., Matsumoto, Y., Bellacosa, A. (2000a). Biphasic kinetics of the human DNA repair protein MED1 (MBD4), a mismatch-specific DNA *N*-glycosylase. *J. Biol. Chem.* **275(42)**: 32422-32429

Petronzelli, F., Riccio, A., Markham, G.D., Seeholzer, S.H., Genuardi, M., Karbowski, M., Yeung, A.T., Matsumoto, Y., Bellacosa, A. (2000b). Investigation of the substrate spectrum of the human mismatch-specific DNA *N*-glycosylase MED1 (MBD4): fundamental role of the catalytic domain. *J. Cell. Physiol.* **185**: 473-480

Pettersen, H.S., Sundheim, O., Gilljam, K.M., Slupphaug, G., Krokan, H.E., Kavli, B. (2007). Uracil-DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms. *Nucleic Acids Res.* **35(12)**:3879-3892

Piccioli, P., Serra, M., Gismondi, V., Pedemonte, S., Loiacono, F., Lastraioli, S., Bertario, L., De Angioletti, M., Varesco, L., Notaro, R. (2006). Multiplex tetra-primer amplification refractory mutation system PCR to detect 6 common germline mutations of the *MUTYH* gene associated with polyposis and colorectal cancer. *Clin. Chem.* **52(4)**:739-743

Picher, A.J., García-Díaz, M., Bebenek, K., Pedersen, L.C., Kunkel, T.A., Blanco, L. (2006). Promiscuous mismatch extension by human DNA polymerase lambda. *Nucleic Acids Res.* **34(11)**: 3259-3266

Picher, A.J. & Blanco, L. (2007). Human DNA polymerase lambda is a proficient extender of primer ends paired to 7,8-dihydro-8-oxoguanine. *DNA Repair (Amst.)* **6**: 1749-1756

Plotz, G., Zeuzem, S., Raedle, J. (2006). DNA mismatch repair and Lynch syndrome. *J. Mol. Histol.* **37**: 271-283

Polakis, P. (2002). Casein kinase 1: a Wnt'er of disconnect. *Curr. Biol.* **12**: R499-R501

Ponti, G. & Ponz de Leon, M. (2005). Muir-Torre syndrome. *Lancet Oncol.* 6 :980-987

Ponti, G., Ponz de Leon, M., Maffei, S., Pedroni, M., Losi, L., Di Gregorio, C., Gismondi, V., Scarselli, A., Benatti, P., Roncari, B., Seidenari, S., Pellacani, G., Varotti, C., Prete, E., Varesco, L., Roncucci, L. (2005). Attenuated familial adenomatous polyposis and Muir-Torre syndrome linked to compound biallelic constitutional *MYH* gene mutations. *Clin. Genet.* **68**:442-447

Ponti, G., Venesio, T., Losi, L., Pellacani, G., Bertario, L., Sala, P., Pedroni, M., Petti, C., Maffei, S., Varesco, L., Lerch, E., Baggio, A., Bassoli, S., Longo, C., Seidenari, S. (2007). BRAF mutations in multiple sebaceous hyperplasias of patients belonging to *MYH*-associated polyposis pedigrees. *J. Invest. Dermatol.* **127(6)**:1387-1391

Popat, S., Stone, J., Coleman, G., Marshall, G., Peto, J., Frayling, I., Houlston, R. (2000). Prevalence of *APC* E1317Q variant in colorectal cancer patients. *Cancer Lett.* **149**:203-206

Pope, M.A., Chmiel, N.H., David, S.S. (2005). Insight into the functional consequences of hMYH variants associated with colorectal cancer: distinct differences in the adenine glycosylase activity and the response to AP endonucleases of Y150C and G365D murine MYH. *DNA Repair (Amst.)* **4**:315-325

Raedle, J., Friedl, W., Engels, H., Koenig, R., Trojan, J., Zeuzem, S. (2001). A *de novo* deletion of chromosome 5q causing Familial Adenomatous Polyposis, dysmorphic features and mild mental retardation. *Am. J. Gastroenterol.* **96(10)**: 3016-3020

Raffoul, J.J., Cabelof, D.C., Nakamura, J., Meira, L.B., Friedlberg, E.C., Heydari, A.R. (2004). Apurinic/apyrimidinic endonuclease (APE/REF-1) haploinsufficient mice display tissue-specific differences in DNA polymerase β-dependent base excision repair. *J. Biol. Chem.* **279(18)**: 18425-18433

Rajagopalan, H., Nowak, M.A., Vogelstein, B., Lengauer, C. (2003). The significance of unstable chromosomes in colorectal cancer. *Nat. Rev. Cancer* **3**: 695-701

Rajagopalan, H., Jallepalli, P.V., Rago, C., Velculescu, V.E., Kinzler, K.W., Vogelstein, B., Lengauer, C. (2004). Inactivation of *hCDC4* can cause chromosomal instability. *Nature* **428**: 77-81

Ramadan, K., Shevelev, I., Maga, G., Hübscher, U. (2002). DNA polymerase λ from calf thymus preferentially replicates damaged DNA. *J. Biol. Chem.* **277(21)**: 18454-18458

Ramadan, K., Maga, G., Shevelev, I.V., Villani, G., Blanco, L., Hübscher, U. (2003). Human DNA polymerase λ possesses terminal deoxyribonucleotidyl transferase activity and can elongate RNA primers: implications for novel functions. *J. Mol. Biol.* **328**: 63-72
Ramadan, K., Shevelev, I., Hübscher, U. (2004a). The DNA-polymerase-X family: controllers of DNA quality? *Nat. Rev. Mol. Cell Biol.* **5**: 1038-1043

Ramadan, K., Shevelev, I.V., Maga, G., Hübscher, U. (2004b). *De novo* DNA synthesis by human DNA polymerase λ, DNA polymerase μ and terminal deoxyribonucleotidyl transferase. *J. Mol. Biol.* **339**: 395-404

Rampino, N., Yamamoto, H., Ionov, Y., Li, Y., Sawai, H., Reed, J.C., Perucho, M. (1997) Somatic frameshift mutations in the *BAX* gene in colon cancers of the microsatellite mutator phenotype. *Science* **275**: 967-969

Ranalli, T.A., Tom, S., Bambara, R.A. (2002). AP endonuclease 1 coordinates flap endonuclease 1 and DNA ligase I activity in long patch base excision repair. *J. Biol. Chem.* **277(44)**: 41715-41724

Raoof, M., Canter, R.J., Paty, P.B. (2007). Variable phenotypic expression of identical *MYH* germline mutations in siblings with Attenuated Familial Adenomatous Polyposis. *Am Surg* **73(12)**: 1250-1253

Renkonen, E.T., Nieminen, P., Abdel-Rahman, W.M., Moisio, A-L., Järvelä, I., Arte, S., Järvinen, H.J., Peltomäki, P. (2005). Adenomatous polyposis families that screen *APC* mutation-negative by conventional methods are genetically heterogeneous. *J. Clin. Oncol.* **23**: 5651-5659

Reya, T. & Clevers, H. (2005). Wnt signalling in stem cells and cancer. *Nature* **434**: 843-850

Riccio, A., Aaltonen, L.A., Godwin, A.K., Loukola, A., Percesepe, A., Salovaara, R., Masciullo, V., Genuardi, M., Paravatou-Petsotas, M., Bassi, D.E., Ruggeri, B.A., Klein-Szanto, A.J.P., Testa, J.R., Neri, G., Bellacosa, A. (1999) *Nat. Genet.* **23**: 266-268

Ripa, R., Bisgaard, M.L., Bülow, S., Nielsen, F.C. (2002). *De novo* mutations in familial adenomatous polyposis (FAP). *Eur. J. Hum. Genet.* **10**: 631-637

Roldán-Arjona, T., Wei, Y-F., Carter, K.C., Klungland, A., Anselmino, C., Wang, R-P., Augustus, M., Lindahl, T. (1997). Molecular cloning and functional expression of a human cDNA encoding the antimutator enzyme 8-hydroxyguanine-DNA glycosylase. *Proc. Natl. Acad. Sci. U.S.A.* **94**: 8016-8020

Romero-Giménez, J., Dopeso, H., Blanco, I., Guerra-Moreno, A., Gonzalez, S., Vogt, S., Aretz, S., Schwartz Jr, S., Capella, G., Arango, D. (2008). Germline hypermethylation of the APC promoter is not a frequent cause of Familial Adenomatous Polyposis in *APC/MUTYH* mutation negative families. *Int. J. Cancer* **122(6)**:1422-5

Rozen, P., Samuel, Z., Shomrat, R., Legum, C. (1999). Notable intrafamilial phenotypic variability in a kindred with familial adenomatous polyposis and an *APC* mutation in exon 9. *Gut* **45**: 829-833

Rubinfeld, B., Souza, B., Albert, I., Müller, O., Chamberlain, S.H., Masiarz, F.R., Munemitsu, S., Polakis, P. (1993). Association of the *APC* gene product with β-catenin. *Science* **262**: 1731-1733

Rubinfeld, B., Albert, I., Porfiri, E., Fiol, C., Munemitsu, S., Polakis, P. (1996). Binding of GSK3β to the APC-β-catenin complex and regulation of complex assembly. *Science* **272**: 1023-1026

Rubinfeld, B., Albert, I., Porfiri, E., Munemitsu, S., Polakis, P. (1997). Loss of β-catenin regulation by the APC tumour suppressor protein correlates with loss of structure due to common somatic mutations of the gene. *Cancer Res.* **57**: 4624-4630

Rubinfeld, B., Tice, D.A., Polakis, P. (2001). Axin-dependent phosphorylation of the Adenomatous Polyposis Coli protein mediated by casein kinase 1ε. *J. Biol. Chem.* **276(42)**: 39037-39045

Ruiz-Ponte, C., Vega, A., Carracedo, A., Barros, F. (2001). Mutation analysis of the Adenomatous Polyposis Coli (*APC*) gene in Northwest Spanish patients with Familial Adenomatous Polyposis (FAP) and sporadic colorectal cancer. *Hum. Mutat.* **18(4)**:355

Russell, A.M., Zhang, J., Luz, J., Hutter, P., Chappuis, P.O., Rey Berthod, C., Maillet, P., Mueller, H., Heinimann, K. (2006). Prevalence of *MYH* germline mutations in Swiss *APC* mutation-negative polyposis patients. *Int. J. Cancer* **118**:1937-1940

Sakiyama, T., Kohno, T., Mimaki, S., Ohta, T., Yanagitani, N., Sobue, T., Kunitoh, H., Saito, R., Shimizu, K., Hirama, C., Kimura, J., Maeno, G., Hirose, H., Eguchi, T., Saito, D., Ohki, M., Yokota, J. (2005). Association of amino acid substitution polymorphisms in DNA repair genes *TP53*, *POLI*, *REV1* and *LIG4* with lung cancer. *Int. J. Cancer* **114**:730-737

Sakumi, K., Furuichi, M., Tsuzuki, T., Kakuma, T., Kawabata, S-I., Maki, H., Sekiguchi, M. (1993). Cloning and expression of cDNA for a human enzyme that hydrolyzes 8-oxo-dGTP, a mutagenic substrate for DNA synthesis. *J. Biol. Chem.* **268(31)**: 23524-23530

Samowitz, W.S., Thliveris, A., Spirro, L.N., White, R. (1995). Alternatively spliced Adenomatous Polyposis Coli (*APC*) gene transcripts that delete exons mutated in attenuated *APC*. *Cancer Res.* **55**: 3732-3734

Samowitz, W.S., Albertsen, H., Herrick, J., Levin, T.R., Sweeney, C., Murtaugh, M.A., Wolff, R.K., Slattery, M.L. (2005). Evaluation of a large,

population-based sample supports a CpG island methylator phenotype in colon cancer. *Gastroenterology* **129**: 837-845

Sampson, J.R., Dolwani, S., Jones, S., Eccles, D., Ellis, A., Evans, D.E., Frayling, I., Jordan, S., Maher, E., Mak, T., Maynard, J., Pigatto, F., Shaw, J., Cheadle J.P. (2003). Autosomal recessive colorectal adenomatous polyposis due to inherited mutations of *MYH*. *Lancet* **362**: 39-41

Sampson, J.R., Jones, S., Dolwani, S., Cheadle, J.P. (2005). *MutYH* (*MYH*) and colorectal cancer. *Biochem. Soc. Trans.* **33(4)**: 679-683

Samuels, Y., Diaz, L.A., Schmidt-Kittler, O., Cummins, J.M., DeLong, L., Cheong, I., Rago, C., Huso, D.L., Lengauer, C., Kinzler, K.W., Vogelstein, B., Velculescu, V.E. (2005). Mutant PIK3CA promotes cell growth and invasion of human cancer cells. *Cancer Cell* **7**: 561-573

Sansom, O.J., Zabkiewicz, J., Bishop, S.M., Guy, J., Bird, A., Clarke, A.R. (2003). MBD4 deficiency reduces the apoptotic response to DNA-damaging agents in the murine small intestine. *Oncogene* **22**: 7130-7136

Sansom, O.J., Bishop, S.M., Bird, A., Clarke, A.R. (2004). MBD4 deficiency does not increase mutation or accelerate tumorigenesis in mice lacking MMR. *Oncogene* **23**: 5693-5696

Santoro, I.M. & Groden, J. (1997). Alternative splicing of the *APC* gene and its association with terminal differentiation. *Cancer Res.* **57**: 488-494

Saparbaev, M., & Laval, J. (1994). Excision of hypoxanthine from DNA containing dIMP residues by the *Escherichia coli*, yeast, rat and human alkylpurine DNA glycosylases. *Proc. Natl. Acad. Sci. U.S.A.* **91**: 5873-5877

Sard, L., Tornielli, S., Gallinari, P., Minoletti, F., Jiricny, J., Lettieri, T., Pierotti, M.A., Sozzi, G., Radice, P. (1997). Chromosomal localizations and molecular analysis of *TDG* gene-related sequences. *Genomics* 44: 222-226

Schneikert, J. & Behrens, J. (2007). The canonical Wnt signalling pathway and its APC partner in colon cancer development. *Gut* **56**: 417-425

Schouten, J.P., McElgunn, C.J., Waaijer, R., Zwijnenburg, D., Diepvens, F., Pals, G. (2002). Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* **30 (12)**:e57

Scott, R.J., van der Luijt, R., Spycher, M., Mary, J-L., Müller, A., Hoppeler, T.H., Haner, M., Müller, H.J., Martinoli, S., Brazzola, Khan, P.M. (1995). Novel germline *APC* gene mutation in a large kindred displaying variable phenotypes. *Gut* **36**: 731-736

Scott, R.J., Meldrum, C., Crooks, R., Spigelman, A.D., Kirk, J., Tucker, K., Koorey, D., the Hunter Family Cancer Service (2001). Familial Adenomatous

Polyposis: more evidence for disease diversity and genetic heterogeneity. *Gut* **48**: 508-514

Screaton, R.A., Kiessling, S., Sansom, O.J., Millar, C.B., Maddison, K., Bird, A., Clarke, A.R., Frisch, S.M. (2003). Fas-associated death domain protein interacts with methyl-CpG binding domain protein 4: a potential link between genome surveillance and apoptosis. *Proc. Natl. Acad. Sci. U.S.A.* **100(9)**: 5211-5216

Segditsas, S., & Tomlinson, I. (2006). Colorectal cancer and genetic alterations in the Wnt pathway. *Oncogene* **25**: 7531-7537

Seki, S., Hatsushika, M., Watanabe, S., Akiyama, K., Nagao, K., Tsutsui, K. (1992). cDNA cloning, sequencing, expression and possible domain structure of human APEX nuclease homologous to *Escherichia coli* exonuclease III. *Biochim. Biophys. Acta* **1131**: 287-299

Sharp, A., Pichert, G., Lucassen, A., Eccles, D. (2004). RNA analysis reveals splicing mutations and loss of expression defects in *MLH1* and *BRCA1*. *Hum. Mutat.* **24(3)**:272

Shi, G., Chang, D-Y., Cheng, C-C., Guan, X., Venclovas, C., Lu, A-L. (2006). Physical and functional interactions between MutY glycosylase homologue (MYH) and checkpoint proteins Rad9-Rad1-Hus1. *Biochem. J.* **400**: 53-62

Shibutani, S., Takeshita, M., Grollman, A.P. (1991). Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature* **349**: 431-434

Shih, I-M., Zhou, W., Goodman, S.N., Lengauer, C., Kinzler, K., Vogelstein, B. (2001). Evidence that genetic instability occurs at an early stage of colorectal tumourigenesis. *Cancer Res.* **61**: 818-822

Shimazaki, N., Yazaki, T., Kubota, T., Sato, A., Nakamura, A., Kurei, S., Toji, S., Tamai, K., Koiwai, O. (2005). DNA polymerase lambda directly binds to proliferating cell nuclear antigen through its confined C-terminal region. *Genes Cells* **10**: 705-715

Shin, E.J., Chappell, E., Pethe, V., Hersey, K., van der Kwast, T., Fleshner, N., Bapat, B. (2007) MYH mutations are rare in prostate cancer. J. *Cancer Res. Clin. Oncol.* **133(6)**:373-378

Shin, M.C., Lee, S.J., Choi, J.E., Cha, S.I., Kim, C.H., Lee, W.K., Kam, S., Kang, Y.M., Jung, T.H., Park, J.Y. (2006). Glu346Lys polymorphism in the *methyl-CpG binding domain* 4 gene and the risk of primary lung cancer. *Jpn. J. Clin. Oncol.* **36(8)**: 483-488

Shinmura, K., Yamaguchi, S., Saitoh, T., Takeuchi-Sasaki, M., Kim, S-R., Nohmi, T., Yokota, J. (2000). Adenine excisional repair function of MYH

protein on the adenine:8-hydroxyguanine base pair in double-stranded DNA. *Nucleic Acids Res.* **28(24)**: 4912-4918

Sieber, O.M., Lamlum, H., Crabtree, M.D., Rowan, A.J., Barclay, E., Lipton, L., Hodgson, S., Thomas, H.J.W., Neale, K., Phillips, R.K.S., Farrington, S.M., Dunlop, M.G., Mueller, H.J., Bisgaard, M.L., Bulow, S., Fidalgo, P., Albuquerque, C., Scarano, M.I., Bodmer, W., Tomlinson, I.P.M., Heinimann, K. (2002). Whole-gene *APC* deletions cause classical familial adenomatous polyposis, but not attenuated polyposis or "multiple" colorectal adenomas. *Proc. Natl. Acad. Sci. U.S.A.* **99(5)**: 2954-2958

Sieber, O.M., Lipton, L., Crabtree, M., Heinimann, K., Fidalgo, P., Phillips, R.K.S., Bisgaard, M-L., Orntoft, T.F., Aaltonen, L.A., Hodgson, S.V., Thomas, H.J.W., Tomlinson, I.P.M. (2003). Multiple colorectal adenomas, classic adenomatous polyposis, and germline mutations in *MYH*. *N. Engl. J. Med.* **348(9)**: 791-799

Sieber, O.M., Howarth, K.M., Thirlwell, C., Rowan, A., Mandir, N., Goodlad, R.A., Gilkar, A., Spencer-Dene, B., Stamp, G., Johnson, V., Silver, A., Yang, H., Miller, J.H., Ilyas, M., Tomlinson, I.P.M. (2004). *Myh* deficiency enhances intestinal tumorigenesis in multiple intestinal neoplasia ($Apc^{Min/+}$) mice. *Cancer Res.* **64**: 8876-8881

Sieber, O.M., Segditsas, S., Knudsen, A.L., Zhang, J., Luz, J., Rowan, A.J., Spain, S.L., Thirlwell, C., Howarth, K.M., Jaeger, E.E.M., Robinson, J., Volikos, E., Silver, A., Kelly, G., Aretz, S., Frayling, I., Hutter, P., Dunlop, M., Guenther, T., Neale, K., Phillips, R., Heinimann, K., Tomlinson, I.P.M. (2006). Disease severity and genetic pathways in attenuated familial adenomatous polyposis vary greatly but depend on the site of the germline mutation. *Gut* **55**: 1440-1448

Silverman, K.A., Koratkar, R., Siracusa, L.D., Buchberg, A.M. (2002). Identification of the modifier of *Min* 2 (*Mom2*) locus, a new mutation that influences *Apc*-induced intestinal neoplasia. *Genome Res.* **12**: 88-97

Sjöblom, T., Jones, S., Wood, L., Parsons, D.W., Lin, J., Barber, T., Mandelker, D., Leary, R.L., Ptak, J., Silliman, N., Szabo, S., Buckhaults, P., Farrell, C., Meeh, P., Markowitz, S.D., Willis, J., Dawson, D., Willson, J.K.V., Gazdar, A.F., Hartigan, J., Wu, L., Liu, C., Parmigiani, G., Park, B.H., Bachman, K.E., Papadopoulos, N., Vogelstein, B., Kinzler, K.W., Velculescu, V.E. (2006). The consensus coding sequences of human breast and colorectal cancers. *Science* **314**: 268-274

Slupska, M.M., Baikalov, C., Luther, W.M., Chiang, J-H., Wei, Y-F., Miller, J.H. (1996). Cloning and sequencing a human homolog (*hMYH*) of the *Escherichia coli mutY* gene whose function is required for the repair of oxidative DNA damage. *J. Bacteriol.* **178(13)**: 3885-3892

Slupska, M.M., Luther, W.M., Chiang, J-H., Yang, H., Miller, J.H. (1999). Functional expression of hMYH, a human homolog of the *Escherichia coli* MutY protein. *J. Bacteriol.* **181(19)**: 6210-6213

Soravia, C., Berk, T., Madlensky, L., Mitri, A., Cheng, H., Gallinger, S., Cohen, Z., Bapat, B. (1998). Genotype-phenotype correlations in attenuated adenomatous polyposis coli. *Am. J. Hum. Genet.* **62**: 1290-1301

Söreide, K., Janssen, E.A.M., Söiland, H., Körner, H., Baak, J.P.A. (2006). Microsatellite instability in colorectal cancer. *Br. J. Surg.* **93**: 395-406

Spirio, L.N., Samowitz, W., Robertson, J., Robertson, M., Burt, R.W., Leppert, M., White, R. (1998). Alleles of *APC* modulate the frequency and classes of mutations that lead to colon polyps. *Nat. Genet.* **20**: 385-388

Steinacher,R. & Schär, P. (2005). Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation. *Curr. Biol.* **15**: 616-623

Stekrova, J., Sulova, M., Kebrdlova, V., Zidkova, K., Kotlas, J., Ilencikova, D., Vesela, K., Kohoutova, M. (2007). Novel *APC* mutations in Czech and Slovak FAP families: clinical and genetic aspects. *BMC Med. Genet.* **8**:16

Strate, L.L. & Syngal, S. (2005). Hereditary colorectal cancer syndromes. *Cancer Causes Control* **16**: 201-213

Su, L-K., Johnson, K.A., Smith, K.J., Hill, D.E., Vogelstein, B., Kinzler, K. (1993a). Association between wild-type and mutant *APC* gene products. *Cancer Res.* **53**: 2728-2731

Su, L-K., Vogelstein, B., Kinzler, K. (1993b). Association of the APC tumour suppressor proteins with catenins. *Science* **262**: 1734-1737

Su, L-K., Barnes, C.J., Yao, W., Qi, Y., Lynch, P.M., Steinbach, G. (2000). Inactivation of germline mutant *APC* alleles by attenuated somatic mutations: a molecular genetic mechanism for attenuated familial adenomatous polyposis. *Am. J. Hum. Genet.* **67**: 582-590

Sulekova, Z., Reina-Sanchez, J., Ballhausen, W.G. (1995). Multiple *APC* messenger RNA isoforms encoding exon 15 short open reading frames are expressed in the context of a novel exon 10A-derived sequence. *Int. J. Cancer* **63**: 435-441

Šulová, M., Zídková, K., Kleibl, Z., Štekrová, J., Kebrdlová, V., Bortlík, M., Lukáš, M., Kohoutová, M. (2007). Mutation analysis of the *MYH* gene in unrelated Czech *APC* mutation-negative polyposis patients. *Eur J Cancer* **43**:1617-1621

Sun, T., Gao, Y., Tan, W., Ma, S., Shi, Y., Yao, J., Guo, Y., Yang, M., Zhang, X., Zhang, Q., Zeng, C., Lin, D. (2007). A six-nucleotide insertion-

deletion polymorphism in the *CASP8* promoter is associated with susceptibility to multiple cancers. *Nat. Genet.* **39(5)**: 605-613

Taddei, F., Hayakawa, H., Bouton, M-F., Cirinesi, A-M., Matic, I., Sekiguchi, M., Radman, M. (1997). Counteraction by MutT protein of transcriptional errors caused by oxidative damage. *Science* **278**: 128-130

Tajiri, T., Maki, H., Sekiguchi, M. (1995). Functional cooperation of MutT, MutM, and MutY proteins in preventing mutations caused by spontaneous oxidation of guanine nucleotide in *Escherichia coli*. *Mutat. Res.* **336**: 257-267

Takao, M., Kanno, S-I., Shiromoto, T., Hasegawa, R., Ide, H., Ikeda, S., Sarker, A.H., Seki, S., Xing, J.Z., Le, X.C., Weinfeld, M., Kobayashi, K., Miyaazaki, J-I., Muijtjens, M., Hoeijmakers, J.H.J., van der Horst, G., Yasui, A. (2002). Novel nuclear and mitochondrial glycosylases revealed by disruption of the mouse *Nth1* gene encoding an endonuclease III homolog for repair of thymine glycols. *EMBO J.* **21(13)**: 3486-3493

Takao, M., Zhang, Q-M., Yonei, S., Yasui, A. (1999). Differential subcellular localization of human MutY homolog (hMYH) and the functional activity of adenine:8-oxoguanine DNA glycosylase. *Nucleic Acids Res.* **27(18)**: 3638-3644

Tano, K., Nakamura, J., Asagoshi, K., Arakawa, H., Sonoda, E., Braithwaite, E.K., Prasad, R., Buerstedde, J-M., Takeda, S., Watanabe, M., Wilson, S.H. (2007). Interplay between DNA polymerases β and λ in repair of oxidation DNA damage in chicken DT40 cells. *DNA Repair (Amst.)* **6**: 869-875

Tao, H., Shinmura, K., Hanaoka, T., Natsukawa, S., Shaura, K., Koizumi, Y., Kasuga, Y., Ozawa, T., Tsujinaka, T., Li, Z., Yamaguchi, S., Yokota, J., Sugimura, H., Tsugane, S. (2004) A novel splice-site variant of the base excision repair gene *MYH* is associated with production of an aberrant mRNA transcript encoding a truncated MYH protein not localized in the nucleus. *Carcinogenesis* **25(10)**:1859-1866

Tenesa, A., Campbell, H., Barnetson, R., Porteous, M., Dunlop, M., Farrington, S.M. (2006). Association of *MUTYH* and colorectal cancer. *Br. J. Cancer* **95(2)**:239-242

Terdiman, J.P. (2000). Genomic events in the adenoma to carcinoma sequence. *Semin. Gastrointest. Dis.* **11(4)**: 194-206

Thibodeau, S.N., Bren, G., Schaid, D. (1993). Microsatellite instability in cancer of the proximal colon. *Science* **260**:816-819

Thirlwell, C., Howarth, K.M., Segdisas, S., Guerra, G., Thomas, H.J.W., Phillips, R.K.S., Talbot, I.C., Gorman, M., Novelli, M.R., Sieber, O.M., Tomlinson, I.P.M. (2007). Investigation of pathogenic mechanisms in multiple colorectal adenoma patients without germline *APC* or *MYH/MUTYH* mutations. *Br. J. Cancer* **96**: 1729-1734

Thliveris, A., Samowitz, W., Matsunami, N., Groden, J., White, R. (1994). Demonstration of promoter activity and alternative splicing in the region 5' to exon 1 of the *APC* gene. *Cancer Res.* **54**: 2991-2995

Tini, M., Benecke, A., Um, S-J., Torchia, J., Evans, R.M., Chambon, P. (2002). Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. *Mol. Cell* **9**: 265-277

Tominaga, Y., Ushijima, Y., Tsuchimoto, D., Mishima, M., Shirakawa, M., Hirano, S., Sakumi, K., Nakabeppu, Y. (2004). MUTYH prevents OGG1 or APEX1 from inappropriately processing its substrate or reaction product with its C-terminal domain. *Nucleic Acids Res.* **32(10)**: 3198-3211

Tomlinson, I., Webb, E., Carvajal-Carmona, L., Broderick, P., Kemp, Z., Spain, S., Penegar, S., Chandler, I., Gorman, M., Wood, W., Barclay, E., Lubbe, S., Martin, L., Sellick, G., Jaeger, E., Hubner, R., Wild, R., Rowan, A., Fielding, S., Howarth, K., the CORGI Consortium, Silver, A., Atkin, W., Muir, K., Logan, R., Kerr, D., Johnstone, E., Sieber, O., Gray, R., Thomas, H., Peto, J., Cazier, J-B., Houlston, R. (2007). A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat. Genet.* **39(8)**: 984-988

Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J.G., Baylin, S.B., Issa, J-P J. (1999) CpG island methylator phenotype in colorectal cancer. *Proc. Natl. Acad. Sci. U.S.A.* **96**: 8681-8686

Truta, B., Allen, B.A., Conrad, P.G., Weinberg, V., Miller, G.A., Pomponio, R., Lipton, L.R., Guerra, G., Tomlinson, I.P.M., Sleisenger, M.H., Kim, Y.S., Terdiman, J.P. (2005). A comparison of the phenotype and genotype in adenomatous polyposis patients with and without a family history. *Fam. Cancer* **4**:127-133

Tsai-Wu, J-J., Su, H-T., Wu, Y-L., Hsu, S-M., Wu, C.H.H. (2000). Nuclear localization of the human *mutY* homologue *hMYH*. *J. Cell. Biochem.* **77**: 666-677

Tsuzuki, T., Egashira, A., Igarashi, H., Iwakuma, T., Nakatsuru, Y., Tominaga, Y., Kawate, H., Nakao, K., Nakamura, K., Ide, F., Kura, S., Nakabeppu, Y., Katsuki, M., Ishikawa, T., Sekiguchi, M. (2001). Spontaneous tumorigenesis in mice defective in the *MTH1* gene encoding 8-oxo-dGTPase. *Proc. Natl. Acad. Sci. U.S.A.* **98(20)**: 11456-11461

Um, S., Harbers, M., Benecke, A., Pierrat, B., Losson, R., Chambon, P. (1998). Retinoic acid receptors interact physically and functionally with T:G mismatch-specific thymine-DNA glycosylase. *J. Biol. Chem.* **273(33)**: 20728-20736

Umar, A., Boland, C.R., Terdiman, J.P., Syngal, S., de la Chapelle, A., Rüschoff, J., Fishel, R., Lindor, N.M., Burgart, L.J., Hamelin, R., Hamilton,

S.R., Hiatt, R.A., Jass, J., Lindblom, A., Lynch, H.T., Peltomäki, P., Ramsey, S.D., Rodriguez-Bigas, M.A., Vasen, H.F.A., Hawk, E.T., Barrett, J.C., Freedman, A.N., Srivastava. (2004). Revised Bethesda guidelines for Hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J. Natl. Cancer Inst.* **96(4)**: 261-268

Ushijima, Y., Tominaga, Y., Miura, T., Tsuchimoto, D., Sakumi, K., Nakabeppu, Y. (2005). A functional analysis of the DNA glycosylase activity of mouse MUTYH protein excising 2-hydroxyadenine opposite guanine in DNA. *Nucleic Acids Res.* **33(2)**:672-682

van der Luijt, R.B., Khan, P.M., Vasen, H.F.A., Tops, C.M.J., van Leeuwen-Cornelisse, I.S.J., Wijnen, J.T., van der Klift, H.M., Plug, R.J., Griffioen, G., Fodde, R. (1997). Molecular analysis of the *APC* gene in 105 Dutch kindreds with Familial Adenomatous Polyposis: 67 germline mutations identified by DGGE, PTT and Southern analysis. *Hum. Mutat.* **9**: 7-16

Vandrovcová, J., Štekrová, J., Kebrdlová, V., Kohoutová, M. (2004). Molecular analysis of the *APC* and *MYH* genes in Czech families affected by FAP or multiple adenomas: 13 novel mutations. *Hum. Mutat.* **23(4)**:397

van Puijenbroek, M., Nielsen, M., Reinards, T.H.C.M., Weiss, M.M., Wagner, A., Hendriks, Y.M.C., Vasen, H.F.A., Tops, C.M.J., Wijnen, J., van Wezel, T., Hes, F.J., Morreau, H. (2007). The natural history of a combined defect in MSH6 and MUTYH in a HNPCC family. *Fam. Cancer* **6(1)**:43-51

Vasen, H.F.A (2005). Clinical description of the Lynch syndrome [hereditary nonpolyposis colorectal cancer (HNPCC)]. *Fam. Cancer* 4: 219-225

Venesio, T., Balsamo, A., Scordamaglia, A., Bertolaso, M., Arrigoni, A., Sprujevnik, T., Rossini, F.P., Risio, M. (2003a). Germline *APC* mutation on the β-catenin binding site is associated with a decreased apoptotic level in colorectal adenomas. *Mod. Pathol.* **16(1)**: 57-65

Venesio, T., Balsamo, A., Rondo-Spaudo, M., Varesco, L., Risio, M., Ranzani, G.N. (2003b). *APC* haploinsufficiency, but not CTNNB1 or CDH1 gene mutations, accounts for a fraction of familial adenomatous polyposis patients without *APC* truncating mutations. *Lab. Invest.* **83(12)**: 1859-2003.

Venesio, T., Molatore, S., Cattaneo, F., Arrigoni, A., Risio, M., Ranzani, G.N. (2004). High frequency of *MYH* gene mutations in a subset of patients with familial adenomatous polyposis. *Gastroenterology* **126**:1681-1685

Venesio, T., Balsamo, A., Sfiligoi, C., Fuso, L., Molatore, S., Ranzani, G.N., Risio, M. (2007). Constitutional high expression of an *APC* mRNA isoform in a subset of attenuated familial adenomatous polyposis patients. *J Mol Med* **85**: 305-312

Vickers, M.A., Vyas, P., Harris, P.C., Simmons, D.L., Higgs, D.R. (1993). Structure of the human 3-methyladenine DNA glycosylase gene and

localization close to the 16p telomere. *Proc. Natl. Acad. Sci. U.S.A.* **90**: 3437-3441

Vidal, A.E., Hickson, I.D., Boiteux, S., Radicella, J.P. (2001). Mechanism of stimulation of the DNA glycosylase activity of hOGG1 by the major human AP endonuclease: bypass of the AP lyase activity step. *Nucleic Acids Res.* **29(6)**: 1285-1292

Vogelstein, B. & Kinzler, K. (2004). Cancer genes and the pathways they control. *Nat. Med.* **10(8)**: 789-799

Wallis, Y.L., Morton, D.G., McKeown, C.M., Macdonald, F. (1999). Molecular analysis of the *APC* gene in 205 families: extended genotype-phenotype correlations in FAP and evidence for the role of APC amino acid changes in colorectal cancer predisposition. *J Med Genet* **36**: 14-20

Walon, C., Kartheuser, A., Michils, G., Smaers, M., Lannoy, N., Ngounou, P., Mertens, G., Verellen-Dumoulin, C. (1997). Novel germline mutations in the *APC* gene and their phenotypic spectrum in familial adenomatous polyposis kindreds. *Hum. Genet.* **100**: 601-605

Wang, L., Baudhuin, L.M., Boardman, L.A., Steenblock, K.J., Petersen, G.M., Halling, K.C., French, A.J., Johnson, R.A., Burgart, L.J., Rabe, K., Lindor, N.M., Thibodeau, S.N. (2004). *MYH* mutations in patients with attenuated and classic polyposis and with young-onset colorectal cancer without polyps. *Gastroenterology* **127**: 9-16

Wang, Q., Lasset, C., Desseigne, F., Frappaz, D., Bergeron, C., Navarro, C., Ruano, E., Puisieux, A. (1999). Neurofibromatosis and early onset of cancers in h*MLH1*-deficient children. *Cancer Res.* **59**: 294-297

Watanabe, S., Ichimura, T., Fujita, N., Tsuruzoe, S., Ohki, I., Shirakawa, M., Kawasuji, M., Nakao, M. (2003). Methylated DNA-binding domain 1 and methylpurine-DNA glycosylase link transcriptional repression and DNA repair in chromatin. *Proc. Natl. Acad. Sci. U.S.A.***100(22)**: 12859-12864

Waters, T.R., Gallinari, P., Jiricny, J., Swann, P.F. (1999). Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1. *J. Biol. Chem.* **274(1)**: 67-74

Watson, P. & Riley, B. (2005). The tumour spectrum in the Lynch syndrome. *Fam. Cancer* **4**: 245-248

Webb, E.L., Rudd, M.F., Houlston, R.S. (2006). Colorectal cancer risk in monoallelic carriers of *MYH* variants. *Am. J. Hum. Genet.* **79**:768-771

Weitz, J., Koch, M., Debus, J., Höhler, T., Galle, P.R., Büchler, M. (2005). Colorectal cancer. *Lancet* **365**: 153-165

Whiteside, D., McLeod, R., Graham, G., Steckley, J.L., Booth, K., Somerville, M.J., Andrew, S.E. (2002). A homozygous germ-line mutation in the human *MHS2* gene predisposes to haematological malignancy and multiple café-au-lait spots. *Cancer Res.* **62**: 359-362

Wiederhold, L., Leppard, J.B., Kedar, P., Karimi-Busheri, F., Rasouli-Nia, A., Weinfeld, M., Tomkinson, A.E., Izumi, T., Prasad, R., Wilson, S.H., Mitra, S., Hazra, T.K. (2004). AP endonuclease-independent DNA base excision repair in human cells. *Mol. Cell* **15**: 209-220

Wijnen, J., de Leeuw, W., Vasen, H., van der Klift, H., Møller, P., Stormorken, A., Meijers-Heijboer, H., Lindhout, D., Menko, F., Vossen, S., Möslein, G., Tops, C., Bröcker-Vriends, A., Wu, Y., Hofstra, R., Sijmons, R., Cornelisse, C., Morreau, H., Fodde, R. (1999) Familial endometrial cancer in female carriers of *MSH6* germline mutations. *Nat. Genet.* **23(2)**:142-144

Will, O., Carvajal-Carmona, L.G., Gorman, P., Howarth, K.M., Jones, A.M., Polanco-Echeverry, G.M., Chinaleong, J.A., Günther, T., Silver, A., Clark, S.K., Tomlinson, I. (2007). Homozygous PMS2 deletion causes a severe colorectal cancer and multiple adenoma phenotype without extraintestinal cancer. *Gastroenterology* **132(2)**:527-530

Wilson III, D.M. & Bohr, V.A. (2007). The mechanics of base excision repair, and its relationship to aging and disease. *DNA Repair (Amst.)* **6**: 544-559

Wong, E., Yang, K., Kuraguchi, M., Werling, U., Avdievich, E., Fan, K., Fazzari, M., Jin, B., Brown, A.M.C., Lipkin, M., Edelmann, W. (2002). Mbd4 inactivation increases C→T transition mutations and promotes gastrointestinal tumor formation. *Proc. Natl. Acad. Sci. U.S.A.* **99(23)**: 14937-14942

Wong, I., Bernards, A.S., Miller, J.K., Wirz, J.A. (2003). A dimeric mechanism for contextual target recognition by MutY glycosylase. *J. Biol. Chem.* **278(4)**: 2411-2418

Wood, L., Parsons, D.W., Jones, S., Lin, J., Sjöblom, T., Leary, R.L., Shen, D., Boca, S.M., Barber, T., Ptak, J., Silliman, N., Szabo, S., Dezso, Z., Ustyanksky, V., Nikolskaya, T., Nikolsky, Y., Karchin, R., Wilson, P.A., Kaminker, J.S., Zhang, Z., Croshaw, R., Willis, J., Dawson, D., Shipitsin, M., Willson, J.K.V., Sukumar, S., Polyak, K., Park, B.H., Pethiyagoda, C.L., Pant, P.V.K., Ballinger, D.G., Sparks, A.B., Hartigan, J., Smith, D.R., Suh, E., Papadopoulos, N., Buckhaults, P., Markowitz, S.D., Parmigiani, G., Kinzler, K.W., Velculescu, V.E., Vogelstein, B. (2007). The genomic landscapes of human breast and colorectal cancers. *Science* **318**:1108-1113

Wooden, S.H., Bassett, H.M., Wood, T.G., McCullough, A.K. (2003). Identification of critical residues required for the mutation avoidance function of human MutY (hMYH) and implications in colorectal cancer. *Cancer Lett.* **205**:89-95

Woodford-Richens, K.L., Rowan, A.J., Gorman, P., Halford, S., Bicknell, D.C., Wasan, H.S., Roylance, R.R., Bodmer, W.F., Tomlinson, I.P.M. (2001). *SMAD4* mutations in colorectal cancer probably occur before chromosomal instability, but after divergence of the microsatellite pathway. *Proc. Natl. Acad. Sci. U.S.A.* **98(17)**: 9719-9723

Wu, G., Wu, W., Hegde, M., Fawkner, M., Chong, B., Love, D., Su, L-K., Lynch, P., Snow, K., Richards, C.S. (2001). Detection of sequence variations in the Adenomatous Polyposis Coli (APC) gene using denaturing high-performance liquid chromatography. *Genet. Test.* **5(4)**: 281-290

Wu, P., Qiu, C., Sohail, A., Zhang, X., Bhagwat, A.S., Cheng, X. (2003). Mismatch repair in methylated DNA. *J. Biol. Chem.* **278(7)**: 5285-5291

Wu, Y., Berends, M.J.W., Sijmons, R.H., Mensink, R.G.J., Verlind, E., Kooi, K.A., van der Sluis, T., Kempinga, C., van der Zee, A.G.J., Hollema, H., Buys, C.H.C.M., Kleibeuker, J.H., Hofstra, R.M.W. (2001). A role for MLH3 in hereditary nonpolyposis colorectal cancer. *Nat. Genet.* **29**: 137-138

Wynter, C.V.A., Kambara, T., Walsh, M.D., Leggett, B.A., Young, J., Jass, J.R. (2006). DNA methylation patterns in adenomas from FAP, multiple adenoma and sporadic colorectal carcinoma patients. *Int. J. Cancer* **118**:907-915

Xanthoudakis, S. & Curran, T. (1992). Identification and characterization of Ref-1, a nuclear protein that facilitates AP-1 DNA binding activity. *EMBO J.* **11(2)**: 653-665

Xanthoudakis, S., Smeyne, R.J., Wallace, J.D., Curran, T. (1996). The redox/DNA repair protein, Ref-1, is essential for early embryonic development in mice. *Proc. Natl. Acad. Sci. U.S.A.* **93**: 8919-8923

Xia, L., St. Denis, K.A., Bapat, B. (1995). Evidence for a novel exon in the coding region of the Adenomatous Polyposis Coli (*APC*) gene. *Genomics* **28**: 589-591

Xia, L., Zheng, L., Lee, H-W., Bates, S.E., Federico, L., Shen, B., O'Connor, T.R. (2005). Human 3-methyladenine-DNA glycosylase: effect of sequence context on excision, association with PCNA, and stimulation by AP endonuclease. *J. Mol. Biol.* **346**: 1259-1274

Yan, H., Dobbie, Z., Gruber, S.B., Markowitz, S., Romans, K., Giardiello, F.M., Kinzler, K.W., Vogelstein, B. (2002). Small changes in expression affect predisposition to tumourigenesis. *Nat. Genet.* **30**: 25-26

Yang, H., Slupska, M.M., Wei, Y-F., Tai, J.H., Luther, W.M., Xia, Y-R., Shih, D.M., Chiang, J-H., Baikalov, C., Fitz-Gibbon, S., Phan, I.T., Conrad, A., Miller, J.H. (2000). Cloning and characterization of a new member of the

nudix hydrolases from human and mouse. *J. Biol. Chem.* **275(12)**: 8844-8853

Yeager, M., Orr, N., Hayes, R.B., Jacobs, K.B., Kraft, P., Wacholder, S., Minichiello, M.J., Fearnhead, P., Yu, K., Chatterjee, N., Wang, Z., Welch, R., Staats, B.J., Calle, E.E., Spencer Feigelson, H., Thun, M.J., Rodriguez, C., Albanes, D., Virtamo, J., Weinstein, S., Schumacher, F.R., Giovannucci, E., Willett, W.C., Cancel-Tassin, G., Cussenot, O., Valeri, A., Andriole, G.L., Gelmann, E.P., Tucker, M., Gerhard, D.S., Fraumeni Jr, J.F., Hoover, R., Hunter, D.J., Chanock, S.J., Thomas, G. (2007). Genome-wide association study of prostate cancer identified a second risk locus at 8q24. *Nat. Genet.* **39(5)**: 645-649

Yoon, J-H., Iwai, S., O'Conner, T.R., Pfeifer, G.P. (2003). Human thymine DNA glycosylase (TDG) and methyl-CpG-binding protein 4 (MBD4) excise thymine glycol (Tg) from a Tg:G mispair. *Nucleic Acids Res.* **31(18)**: 5399-5404

Yuen, S.T., Chan, T.L., Ho, J.W.C, Chan, A.S.Y., Chung, L.P., Lam, P.W.Y., Tse, C.W., Wyllie, A.H., Leung, S.Y. (2002). Germline, somatic and epigenetic events underlying mismatch repair deficiency in colorectal and HNPCC-related cancers. *Oncogene* **21**: 7585-7592.

Zanke, B.W., Greenwood, C.M.T., Rangrej, J., Kustra, R., Tenesa, A., Farrington, S.M., Prendergast, J., Olschwang, S., Chiang, T., Crowdy, E., Ferretti, V., Laflamme, P., Sundararajan, S., Roumy, S., Olivier, J-F., Robidoux, F., Sladek, R., Montpetit, A., Campbell, P., Bezieau, S., O'Shea, A.M., Zogopoulos, G., Cotterchio. M., Newcomb, P., McLaughlin, J., Younghusband, B., Green, R., Gree, J., Porteous, M.E.M., Campbell, H., Blanche, H., Sahbatou, M., Tubacher, E., Bonaiti-Pellié, C., Buecher, B., Riboli, E., Kury, S., Chanock, S.J., Potter, J., Thomas, G., Gallinger, S., Hudson, T.J., Dunlop, M.G. (2007). Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat. Genet.* **39(8)**:989-994

Zha, M., Zhong, C., Peng, Y., Hu, H., Ding, J. (2006). Crystal structures of human NUDT5 reveal insights into the structural basis of the substrate specificity. *J. Mol. Biol.* **364**: 1021-1033

Zhang, Y., Liu, X., Fan, Y., Ding, J., Xu, A., Zhou, X., Hu, X., Zhu, M., Zhang, X., Li, S., Wu, J., Cao, H., Li, J., Wang, Y (2006) Germline mutations and polymorphic variants in *MMR*, *E-cadherin* and *MYH* genes associated with familial gastric cancer in Jiangsu of China. *Int. J. Cancer* **119**:2592-2596

Zhao, J. & Winkler, M.E. (2000). Reduction of GC→TA transversion mutation by overexpression of MutS in *Escherichia coli* K-12. *J. Bacteriol.* **182(17)**: 5025-5028.

Zharkov, D.O., Shoham, G., Grollman, A.P. (2003). Structural characterization of the Fpg family of DNA glycosylases. *DNA Repair (Amst.)* **2** 839-862

Zhou, X-L., Djureinovic, T., Werelius, B., Lindmark, G., Sun, X-F., Lindblom, A., the Swedish Low-Risk Colorectal Cancer Group (2005) Germline mutations in the *MYH* gene in Swedish familial and sporadic colorectal cancer. *Genet. Test.* **9(2)**:147-151

Zhu, B., Zheng, Y., Angliker, H., Schwarz, S., Thiry, S., Siegmann, M., Jost, J-P. (2000). 5-methylcytosine DNA glycosylase activity is also present in the human MBD4 (G/T mismatch glycosylase) and in a related avian sequence. *Nucleic Acids Res.* **28(21)**: 4157-4165

Zhu, B., Benjamin, D., Zheng, Y., Angliker, H., Thiry, S., Siegmann, M., Jost, J-P. (2001). Overexpression of 5-methylcytosine DNA glycosylase in human embryonic kidney cells EcR293 demethylates the promoter of a hormone-regulated reporter gene. *Proc. Natl. Acad. Sci. U.S.A.* **98(9)**: 5031-5036