Swansea University E-Theses

Reprimo G>C and P53R2 C>G 4696 single nucleotide polymorphisms and colorectal cancer: A case-control disease-association study.

Beasley, William D

How to cite:

Use policy:

REPRIMO G>C 824 AND *P53R2* C>G 4696 SINGLE NUCLEOTIDE POLYMORPHISMS AND COLORECTAL CANCER – A CASE-CONTROL DISEASE - ASSOCIATION STUDY.

William D Beasley MBBS MRCS.

Submitted to the University of Wales in fulfilment of the requirements for the degree of Master of Philosophy.

Swansea University

2008

ProQuest Number: 10807459

ProQuest 10807459

SUMMARY

This thesis presents a case – control disease association research project examining association between single nucleotide polymorphisms (SNPs) of the *P53R2* and *Reprimo* genes and colorectal cancer (CRC).

Identifying the genetic risk factors may help to reduce the incidence of CRC by allowing identification of individuals at increased CRC risk through genetic screening. Individuals identified as having increased CRC risk could then have tailored management aiming to reduce their risk and detect CRC at an early, potentially curable stage.

P53R2 and *Reprimo* are both involved in cell cycle control and DNA repair. *P53R2* supplies deoxyribonucleotides for repair of damaged DNA and *Reprimo* is involved in cell cycle arrest in response to genotoxic stress, allowing DNA repair to take place or for apoptosis. Both genes are induced by the tumour suppressor gene *TP53*. Dysfunction of *P53R2* or *Reprimo* may lead to errors in DNA repair or cell cycle arrest, mechanisms that are recognised in carcinogenesis. SNPs are known to be able to alter gene function and it is thought that SNPs may account for some of the molecular pathology involved in common complex diseases such as cancer.

Two SNPs were studied, *P53R2* C>G 4696 and *Reprimo* G>C 842, using a case control method. Cases with histologically proven CRC were compared with two control populations; young, healthy individuals and individuals with diverticular disease (DD).

DNA was obtained from buccal cell biopsies and the populations genotyped using an allele specific polymerase chain reaction (PCR) or polymerase chain reaction/restriction fragment length polymorphism (PCR/RFLP) methods. Tests for Hardy-Weinberg equilibrium (HWE) and allelic- and genotype-disease association were performed using an on line SNP-disease association calculator.

An association between the *Reprimo* G>C 842 C allele and CRC was detected on heterozygous and allele positivity testing when using the young, healthy population as a control. However, this association is likely to be a false positive result confounding factors and was not replicated when using the matched DD population. The DD population was found to deviate from HWE with respect to *Reprimo* G>C 824 and therefore a disease association may exist. Otherwise no associations between the study SNPs and CRC were detected.

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed .... (candidate)

Date .... 30th June 2008

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed .... (candidate)

Date .... 30th June 2008

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed .... (candidate)

Date .... 30th June 2008.

ACKNOWLEDGMENTS

ABBREVIATIONS

5-FU – 5 - fluorouracil
AD – autosomal dominant
AKAP - A kinase anchor protein
ALDH2 – aldehyde dehydrogenase 2 family
APC – adenomatous polyposis coli
APOE – apolipoprotein E
AR –autosomal recessive
ATM - ataxia telangiectasia mutated
BAX – BCL2-associated X protein
BCL – B cell lymphoma
BLM – Bloom syndrome
BMI – body mass index
BMPR – bone morphogenetic protein receptor
BRCA – breast cancer gene
CD – cluster of differentiation
Cdc – cell division cycle
Cdk – cyclin-dependent kinase
CEA – carcinoembryonic antigen
CHRPE – congenital hypertrophy of retinal pigment epithelium
CI – confidence interval
CRC – colorectal cancer
CRT - chemoradiotherapy
CT – computed tomography
CXCR – chemokine, CXC motif, receptor
CYP – cytochrome p450 polypeptide
DCC – deleted in colon cancer
DD – diverticular disease
DMH – 1,2-dimethylhydrazine
DNA – deoxyribonucleic acid
Drg – developmentally regulated GTP-binding protein
dNTP - deoxyribonucleotide
ECM – extracellular matrix
E2F – E2F transcription factor
FAP – familial adenomatous polyposis
FOB – faecal occult blood
GSK – glycogen synthase kinase
GSTM – glutathione-S-transferase, Mu
GSTP – glutathione-S-transferase psuedogene
GSTT - glutathione-S-transferase, theta
hDLG – homologue of Drosophila large discs
HNPCC – hereditary non-polyposis colon cancer
HRAS – Harvey Rat viral oncogenes homologue
HRT – hormone replacement therapy
HWE – Hardy-Weinberg equilibrium
IGF – insulin-like growth factor
JPS – juvenile polyposis syndrome
K-Ras – Kirsten rat sarcoma viral oncogenes homologue
Met – MET proto-oncogene

MLH – MutL homologue
MMP – matrix metalloproteinase
MMR – mismatch repair
mRNA – messenger RNA
MSH – MutS homologue
MSI – microsatellite instability
MTHFR – methylenetetrahydrofolate reductase
MYH – MutY homologue
NAT – N-acetyl transferase
NHS – National Health Service
OR – odds ratio
P53R2 - p53-inducible ribonucleotide reductase small subunit 2 homologue
PCR – polymerase chain reaction
PCR-SSCP – polymerase chain reaction – single-strand conformation polymorphism
PD-ECGF – platelet derived endothelial cell growth factor
PIA – polymorphism interaction analysis
PJS – Peutz-Jegher's syndrome
PLA2G2A – phospholipase A2, group 2A
PMS – post meiotic segregation increased
PTEN – phosphatase and tensin homologue
RFLP – restriction fragment length polymorphism
RNA – ribonucleic acid
RNR – ribonucleoside reductase
SMAD – mothers against decapentaplegic
SNP – single nucleotide polymorphism
STK – serine/threonine protein kinase
TGF – transforming growth factor
TGFBR – transforming growth factor beta receptor
TIMP – tissue inhibitor of metalloproteinase
TNF – tumour necrosis factor
TNM – tumour, nodes, metastases
UICC – Union Internacional Contra la Cancrum
UTR – untranslated region
uPAR – urokinase type plasminogen activator receptor
USS – ultra sound scan
VEGF – vascular endothelial growth factor
VNTR – variable number tandem repeat

TABLE OF CONTENTS

List Of Tables.

List Of Figures.

PREFACE

### i.	Problem statement

Colorectal cancer (CRC) is a fatal disease that represents a significant health problem in terms of numbers of people affected and burden on health care resources. Although advances in surgical technique and medical science have resulted in improved diagnostic technology, increasingly complex surgery and improved anaesthetic and intensive care resulting in lower operative mortality, overall mortality due to CRC over the past 40 years has not been significantly reduced (Little and Faivre, 1999, 1998, Grogan et al., 1996, Cunningham and Dunlop, 1996). The probable reason for the unchanging mortality is that the diagnosis of CRC is made at an often advanced and incurable stage, and that this situation has not changed with advances in surgical and medical science. It is likely that mortality from CRC will remain high until it becomes possible to diagnose the disease at an earlier, curable stage. This would, in part, involve being able to identify and treat individuals with a greater risk of developing colorectal cancer at a pre-symptomatic stage.

Screening of targeted individuals is, at present, limited to those with symptoms suggestive of CRC or those with strong family histories and usually involves faecal occult blood detection, endoscopic or x-ray examination to identify suspicious colonic lesions. Genetic screening is limited to those individuals with family histories suggestive of the recognized familial syndromes leading to the development of colorectal cancer.

The molecular biology of CRC has started to be unravelled and the various genetic steps necessary for cancer development have started to be understood. Analysis of the genotype of tumours can yield more accurate prognostic information than

traditional stage and grade information and can guide treatment strategies, particularly adjuvant chemoradiotherapy.

Hereditary forms of CRC, although rare, have long been recognized and recently the molecular biology of some of these syndromes has been determined, allowing genetic screening and guided therapy. As knowledge of CRC genetics advances, more genes are implicated in its development and the possibility of genetic screening to identify individuals at greater risk of developing the disease becomes more feasible.

Paralleling advances in cancer genetics have been advances in genetic technology allowing easier and cheaper methods of gene identification and functional analysis, allowing more efficient investigation of candidate disease genes and genetic screening.

Reprimo and *p53R2* are both genes involved in DNA repair and failure of DNA repair mechanisms are fundamental in carcinogenesis. Variations in genes such as single nucleotide polymorphisms may lead to functional variations. Although SNPs of *Reprimo* and *p53R2* have been discovered it is not known whether these variations have any effect on their functions and so potential to modulate DNA repair. If so it is conceivable that SNPs of *Reprimo* and *p53R2* could have variable DNA repair effectiveness that could lead to DNA replication errors, a fundamental step in carcinogenesis. As such *Reprimo* and *p53R2* present themselves as candidate genes for investigation into association with cancer.

ii. Aim

With these ideas in mind the aim of this research is to investigate possible association between the *Reprimo* G>C 824 and *p53R2* C>G 4696 SNPs and CRC using a case-control study method.

iii. Research approach

To achieve this aim Chapter 1 provides the background to the research. The current understanding of the molecular biology of CRC is reviewed. Single Nucleotide Polymorphisms are defined, their effects on gene function explained and their role in carcinogenesis reviewed. The two study genes *Reprimo* and *p53R2* are introduced and a detailed description of their genetics is given. Details of the discovery of their SNPs are also given. The control disease diverticulosis is reviewed in terms of its epidemiology, pathology and genetics.

The Materials and Methods are described in Chapter 2 where details of the laboratory techniques used to genotype the study populations and the statistical methods used to test the study genes for association with CRC are given.

The preliminary work involved in designing and developing the research methods and the problems encountered in designing the methods are given in Chapter 3.

Chapter 4 presents the results of the research. Population demographics and comparisons, DNA collection and isolation results and results of genotyping are given. Examples of typical electrophoresis gels that display genotypes are shown and the results of the statistical analyses of the genotyping data are presented.

Chapter 5 discusses the research in the context of the current understanding of the topic. The research method is discussed in terms of its strengths and weaknesses. Further work stimulated by this study is also discussed.

CHAPTER 1

BACKGROUND

1.1 COLORECTAL CANCER

1.1.1 Definition

CRCs are malignant tumours of the large bowel. The majority arise from the glandular mucosa and are termed adenocarcinomas, and this thesis is concerned only with adenocarcinomas.

1.1.2 Epidemiology.

CRC is the second most common cause of death from malignant disease accounting for nearly 400,000 deaths per year worldwide (2001). In England and Wales CRC accounts for 13% of all cancers. Only lung cancer and prostate cancer in men and lung cancer in women are more frequent and lung cancer more lethal in both sexes.

1.1.3 Incidence and Mortality.

The incidence of CRC between 2002 – 2004 was 34,395 cases in the United Kingdom and 1,957 in Wales. The mortality from CRC between 2002-2004 was 15,908 in the United kingdom and 905 in Wales (Statistics, 2007)

The survival from CRC is moderate, with a five-year survival rate of around 40%. Colorectal cancer occurs more frequently in men with an aged-standardised ratio of 1.5:1. Incidence rises steadily with age, and the disease is most common in the seventh decade of life.

1.1.4 Geography

CRC is more common in developed countries. Thus incidence rates are higher in Europe, North America and Australia, but lower in the developing regions of Asia and Africa. However migrants to areas of higher incidence acquire the local risk, suggesting a significant environmental risk factor in the aetiology of CRC.

Despite similar incidence rates survival from CRC is higher in Australia and North America than in Europe. England and Wales has poorer survival rates than Europe. This variation in survival may reflect earlier detection of CRC and better management in those countries with improved survival.

Within England and Wales the incidence of CRC is highest in Wales and lowest in North Thames. Mortality is also highest in Wales and lowest in North Thames. This may well reflect variations in health service provision.

1.1.5 Diagnosis.

The majority of cases of CRC are diagnosed following individuals presenting with the symptoms of the disease (as discussed below). Unfortunately symptomatic CRC usually represents relatively advanced disease, and so less likelihood of cure. A proportion of CRC is diagnosed by screening high risk individuals on the basis of family history. Some CRC is diagnosed incidentally during investigation of other problems.

Diagnosis depends on biopsying suspicious lesions found on physical examination, endoscopy and/or radiological investigations. Guidelines have been developed to direct physicians in the management of CRC and particularly in referral of patients with high risk symptoms or family history (2004, Ireland, 2001). The aim of such

guidelines is to minimise diagnostic delay from the onset of symptoms and ensure patients receive optimal treatment based on the best available evidence.

1.1.6   Pathology.

Colorectal adenocarcinomas arise from the glandular epithelium of the colon and rectum. At variable times in their development they metastasise to the local lymph nodes via lymphatic channels and to distant organs, commonly the liver, via the blood stream (haematogenous spread). The majority of colorectal cancers develop in pre-existing adenomatous polyps and there is a well defined progression from normal mucosa to adenomatous polyps to malignant carcinoma. This progression is mirrored by increasing dysplasia within the tissue.

As the tumours grow with in the lumen of the bowel they may be manifest symptomatically by blood loss per rectum due to tumour bleeding, by an alteration in the bowel habit due to obstruction of the passage of stools through the bowel or due to excessive mucus production. The tumours may grow through the bowel wall and invade adjacent tissues and organs. Tumours that have spread to the liver can cause jaundice and liver failure, and this is a frequent cause of death in CRC. Bleeding tumours may present with anaemia from slow bleeding or catastrophic haemorrhage when significant blood vessels are invaded. Obstructing tumours may present with chronic obstruction or acute obstruction with bowel perforation.

Although CRC may present with a variety of different symptoms and clinical signs the common denominator is that for the tumours to be clinically manifested they are usually at an advanced stage of development when metastatic spread is more likely and cure, which depends on complete tumour excision, is less likely.

1.1.7 Tumour Stage.

Tumour stage refers to the stage of development or growth of cancers and is used to make prognostic assumptions and therefore plan treatment. Indeed survival following diagnosis of cancer is related to tumour stage and when cancers are diagnosed it is important to determine their stage. Although most cancers evolved their own specific staging systems (for example Duke's stage for rectal cancer) there has been consensus on standardising staging using the TNM system, developed by the UICC (Union Internacional Contra la Cancrum). TNM stands for Tumour, (lymph) Nodes, Metastases and cancers are staged according to the size of the primary tumour, presence of lymph nodes and presence of distant metastases. Having established the TNM stage for a cancer a stage group can be assigned. Once a cancer has been staged assumptions regarding curability can be made according to prior determined stage related prognosis and decisions made regarding the most suitable treatment strategies. Clearly in order to allow this, staging needs to be a pre-treatment/operative process utilising various diagnostic modalities principally computed tomography (CT) and ultra sound scanning (USS), although as technology advances more sophisticated scans are becoming available, such as magnetic resonance imaging, endorectal ultrasound and positron emission tomography. Definitive staging for resected tumours is dependent on pathological examination of the resection specimens and this information is used to plan post-operative adjuvant chemo- or radiotherapy.

The Tables below show how CRC is staged (Tables 1.1-1.3) and how stage relates to survival (Table 1.4).

DUKE'S STAGE

A	Tumour confined to bowel wall, no positive lymph nodes.
B	Tumour growth beyond muscularis propria, no positive lymph nodes.
C1	Tumour growth beyond muscularis propria, positive lymph nodes but apical node negative.
C2	Tumour growth beyond muscularis propria, positive apical lymph node.

Table 1.1: Dukes' stage for rectal cancer (the apical lymph node is the node closest to the point at which the blood vessels are ligated in a resected tumour specimen).

Stage	T	N	M
x	Incomplete information		
is	Tumour "in-situ", confined to mucosa.		
0	No tumour found	No lymph nodes involved	No metastases
1	Tumour involves submucosa.	1-3 local lymph nodes involved	Evidence of metastases
2	Tumour involves muscularis propria.	> 4 local lymph nodes involved	
3	Tumour breeches muscularis propria.		
4	Tumour involves adjacent organs/tissues or has perforated.		

Table 1.2: TNM stages for colorectal cancer (N stage is only ever 0, 1 or 2 and M stage only ever 0 or 1).

STAGE GROUP	T	N	M	DUKE'S
0	Carcinoma In-situ	0	0	
I	1 – 2	0	0	A
IIA	3	0	0	B
IIB	4	0	0	B
IIIA	1-2	1	0	C
IIIB	3-4	1	0	C
IIIC	Any	2	0	C
IV	Any	any	1	

Table 1.3: Stage grouping for colorectal cancer.

STAGE GROUP	PERCENTAGE 5 YEAR SURVIVAL
I	93
IIA	85
IIB	72
IIIA	83
IIIB	64
IIIC	44
IV	8

Table 1.4: Five year survival according to stage for colorectal cancer (O'Connell et al., 2004) .

As Table 4 shows lower tumour stage is associated with better survival, and so it is desirable to diagnose CRC at an early stage if possible. However, as mentioned previously, CRC tends not to produce symptoms at the early stages, so is not clinically manifest and therefore diagnosed so frequently at the early, better prognostic stages. Indeed almost a third of CRC presents with metastatic spread with very little likelihood of cure (Table 1.5). Therefore in order to improve

survival from CRC it is necessary to develop methods to detect the disease at a pre-symptomatic, more favourable stage.

DUKE'S STAGE	FREQUENCY AT DIAGNOSIS
A	11%
B	35%
C	26%
METASTATIC DISEASE	29%

Table 1.5: Stage frequency at diagnosis of CRC.

1.1.8   Tumour grade.

Tumour grade refers to how the histological appearance of the tumour cells relates to tumour behaviour. Despite all adenocarcinomas being malignant, that is they have the potential to metastasize, there is a spectrum of 'aggressiveness' from low-grade tumours that histologically more closely resemble the tissue from which they derive (well differentiated) and tend to grow more slowly and metastasize later, and, to high grade tumours that grow rapidly, invade local tissues and metastasize early, and histologically are more distinct from the original tissue (poorly differentiated). Knowledge of a tumour's grade adds to prognostic information with higher grade, more poorly differentiated tumours tending to have a worse prognosis. Tumour grade influences treatment strategies, particularly adjuvant therapies.

1.1.9   Aetiology of CRC (Figure 1.1).

80% of CRC cases arise spontaneously with no known pre-existing inherited predisposition (and it is with this group that this thesis is concerned). Five percent

of cases are inherited due to autosomal dominantly inherited genetic conditions including familial adenomatous polyposis (FAP) and related conditions (*MYH*-associated polyposis and the hamartomatous syndromes), and hereditary non-polyposis colon cancer (HNPCC). Around 15% cases are familial and are identified by clustering of the disease within families and early diagnosis.
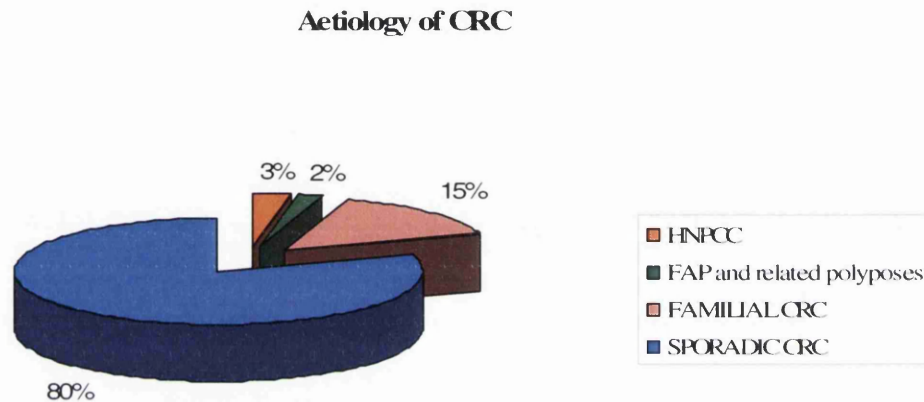
Aetiology of CRC

Figure 1.1 Aetiology of CRC

1.1.10 Environmental Aetiological Factors.

1.1.10.1 Diet.

Diet has been considered one of the most important environmental aetiological factors for the development of CRC for a long time. Increased risk of CRC has been shown to be associated with increased faecal pH, increased concentration of faecal bile acids and decreased faecal bulk (Thorton, 1981, Jenson et al., 1982, Reddy and Wynder, 1977). As dietary fibre increases faecal bulk and lowers faecal pH it was thought that it may protect against CRC (Eastwood, 1987). Furthermore the dilutional effect of increased faecal bulk from fibre was postulated to reduce the

exposure to carcinogens and it was speculated that specific anti-carcinogenic agents may be present in fruit and vegetables (Lampe et al., 1992). Red meat and processed meats were also implicated as an environmental risk factor for CRC (Willett et al., 1990, Goldbohm et al., 1994). Proposed mechanisms for the increased CRC risk from red meat included: increased bile acid secretion; mitogenic effects of diacylglyceride products of fat digestion; effects of increased saturated fatty acids or decreased polyunsaturated fatty acids on insulin receptors; carcinogenic effects of increased colonic ammonia from breakdown of dietary protein; effects of increased iron; enhanced production of carcinogenic N-nitroso compounds; formation of carcinogenic heterocyclic amines and polycyclic aromatic hydrocarbons by high temperature cooking and naked flames (Norat et al., 2002). Therefore the role of diet in CRC has been extensively studied (Glade, 1999, Michels et al., 2000). However the only consistent association that has been demonstrated from case control studies is that of an inverse association between vegetable intake and CRC.

Dietary fibre has not been shown to have a clear association with CRC (1997, Byers, 2000, Kim, 2000) and indeed no protective effects of wheat bran have been demonstrated in randomised controlled trials (Byers, 2000).

No statistically significant association has been demonstrated between meat consumption and CRC in meta-analyses (Sandhu et al., 2001).

1.1.10.2 Obesity.

Obesity has been shown to be a risk factor for developing CRC. A BMI of greater than $25kg/m^2$ or $30kg/m^2$ have been shown to increase an individuals risk by 15% and 33% respectively (Bergström et al., 2001).

1.1.10.3 Exercise.

The risk of Colon cancer but not rectal cancer has been shown to be reduced by physical exercise. A 50% risk reduction can be achieved with high levels of exercise, with the effect being strongest for men (Colditz et al., 1997).

1.1.10.4 Smoking.

Long term tobacco smokers are at increased risk of developing CRC with relative risks of 1.5 – 3. 20% of CRC cases in the USA are probably as a result of smoking (Giovannucci, 2001).

1.1.10.5 Alcohol.

There is no clear evidence for the relationship between alcohol consumption and CRC. Meta-analysis of the available studies shows significant heterogeneity (Corrao et al., 1999).

1.1.10.6 Hormone Therapy.

Postmenopausal oestrogen replacement therapy has been shown to reduce the risk of colon cancer (Crandall, 1999). Colocyte oestrogen receptor methylation has been proposed as an early event in colorectal carcinogenesis and demethylation from exogenous oestrogen may account for the reduced risk (Issa et al., 1994). However the reduction in cancer risk is offset by an increased risk of cardiovascular events, pulmonary embolism and invasive breast cancer (Rossouw et al., 2002). The protective effects of HRT do not endure after stopping the treatment and there does

not seem to be any effect on rectal cancer risk (Goldstein et al., 1999, Nanda et al., 1999).

There may be some risk reduction from oral contraceptives (Franceschi and Vecchia, 1998).

1.1.11 Colorectal Cancer Genetics.

Of all the cancers the genetics of CRC is perhaps the best understood and most extensively studied. This is due to it being a common disease and having a clearly understood pre-malignant phenotype; the adenomatous polyp. Extensive studies of the genetic changes that occur during the progression from normal mucosa to carcinoma have led to the development of a genetic paradigm; the "adenoma-carcinoma sequence"(Vogelstein et al., 1988, Fearon and Volgestein, 1990). This paradigm shows how the step-wise accumulation of genetic mutations in key regulatory genes leads to increasing cellular instability, uncontrolled growth and metastatic spread (Figure 1.2).

| | | *K-* | | | DCC | P53 |
| *APC* | | *Ras* | | | | |

| Normal Epithelium | ∂ | Early Adenoma | ∂ | Intermediate Adenoma | ∂ | Late Adenoma | ∂ | Invasive Cancer |

Mismatch repair genes; MSH2, MLH1, PMS2 etc.
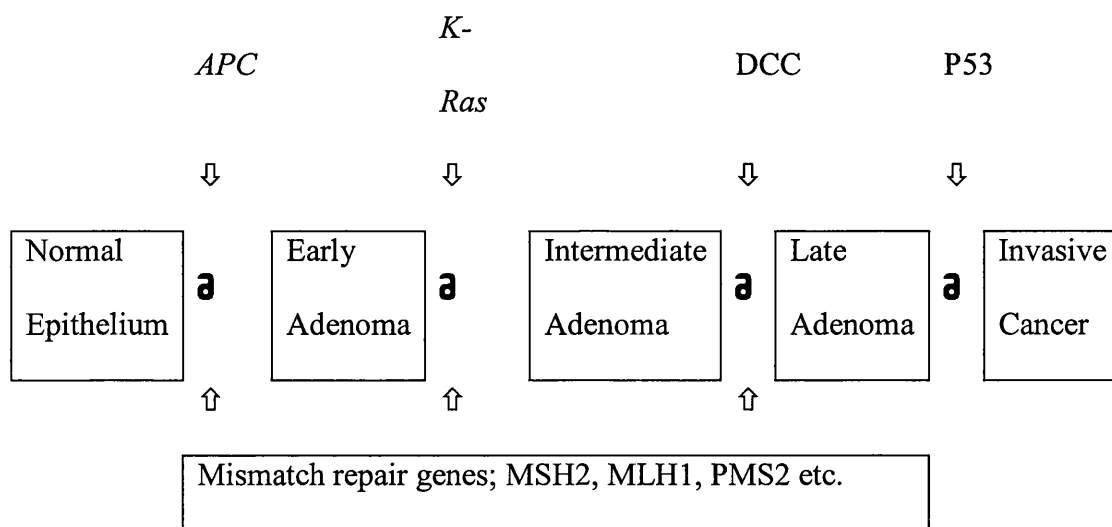
Figure 1.2: The adenoma-carcinoma sequence.

Three categories of genes are involved in colorectal carcinogenesis: tumour-suppressor genes (e.g. *APC*, p53 and p16), oncogenes (e.g. *K-Ras*) and DNA repair genes (e.g. MSH2, MLH1, PMS2)(Jo and Chung, 2005). Tumour suppressor genes inhibit cellular growth stimulatory pathways and classically require both alleles to be mutated before gene function is compromised. Some tumour suppressor genes exhibit a dominant negative effect where a single mutated allele leads to abnormal gene function. P53 is an example. The dominant negative effect arises because the functional p53 protein is tetrameric and a single mutant allele will give rise to abnormal proteins that are incorporated into the protein complex potentially impairing function. Oncogenes promote cellular growth and proliferation pathways and gene function is compromised by mutation in a single allele. DNA repair genes serve to maintain the integrity of the genome. DNA damage due to various genotoxic mechanisms and replication errors can lead to genetic instability and a "mutator phenotype" manifest by an increased mutation rate and mutation accumulation.

Clustering of CRC within families has long been recognised and syndromes with a very high risk of developing CRC have been identified. The conditions familial adenomatous polyposis (FAP) and hereditary non-polyposis colon cancer (HNPCC), in which the development of CRC is virtually inevitable, were identified as autosomal dominant conditions on the basis of family studies before the molecular pathology was elucidated.

Although hereditary CRC accounts for a small proportion of the total number of CRC cases, identification of the causative gene defects has lead to a greater understanding and provides a useful model of the molecular pathology of sporadic CRC.

1.1.11.1 Familial CRC.

Familial CRC can be divided into those tumours that arise as a result of high penetrance susceptibility genes and those that are due to low penetrance susceptibility genes. The former include HNPCC and FAP and are considered the hereditary CRCs. In the latter the precise genetics have not been elucidated but are characterised by multiple members of the same family being affected by CRC with an increased risk of developing CRC in unaffected members (Abdel-Rahman and Peltomaki, 2004, Kwak and Chung, 2007).

Hereditary CRC arises as a result of germ line mutations in key regulatory tumour suppressor or mismatch repair (MMR) genes. The majority are autosomal dominant conditions and therefore offspring of carriers have a 50% chance of inheriting the risk genes. Carriers have a germ line mutation in one allele and require mutation of the second allele to acquire the pathological phenotype (Knudson's "two-hit hypothesis"(Knudson, 1971)). HNPCC and FAP are the commonest forms of inherited CRC and carry a very high risk. Other inherited conditions are also recognised that have an increased risk of developing CRC (Table 1.6). All of the inherited forms of CRC also carry an increased risk of extra-colonic malignancies, demonstrating that all the cells in the body have the germ line mutations and so are vulnerable. There is a predilection for the colon, however.

Syndrome	Affected Genes	Risk of CRC	Inheritance	Age	Extra-colonic tumours
FAP	*APC*	100%	AD	Late 30s	Duodenal,desmoid, osteomas, thyroid, brain
Attenuated FAP	*APC*	>80%	AD	Mid 50s	Duodenal
HNPCC	*MLH1, MSH2, MSH6*	80%	AD	Mid 40s	Uterus, ovary, stomach, kidney, urinary tract, biliary, small intestine, skin
MYH polyposis	*MYH*	~100%	AR	Early 50s	
PJS	*STK 11 (LKB1)*	40%	AD	Mid 40s	Uterus, breast, lungs, pancreas, gallbladder
JPC	*SMAD4, BMPR1A*	10-40%	AD	Mid 30s	Stomach, duodenum

Table 1.6: Hereditary CRC syndromes. (FAP - familial adenomatous polyposis, HNPCC - hereditary non-polyposis colon cancer, PJS - Peutz-Jegher's syndrome, JPC - juvenile polyposis coli, HPS - hyperplastic polyposis syndrome).

Syndrome	Affected Genes	Risk of CRC	Inheritance	Age	Extra-colonic tumours
Bloom's syndrome	*BLM*	8%	AR	Early 30s	Leukaemia, lymphoma, ENT tumours, ovarian, breast, upper GIT
I1307K *APC* polymorphism	*APC*	8-11%	AD	>60	

Table 1.6 (continued): Hereditary CRC syndromes. (FAP - familial adenomatous polyposis, HNPCC - hereditary non-polyposis colon cancer, PJS - Peutz-Jegher's syndrome, JPC - juvenile polyposis coli, HPS - hyperplastic polyposis syndrome).

1.1.11.2 Familial Adenomatous Polyposis (FAP).

FAP is manifested by the development of hundreds to thousands of colonic adenomatous polyps, starting in adolescence, with inevitable development of adenocarcinoma by the age of forty. FAP is a highly penetrant autosomal dominant disease, caused by germ line mutations of the adenomatous polyposis coli (*APC*) gene located on chromosome 5q21 (Merg et al., 2005a, Lipton and Tomlinson, 2006, Maple and Boardman, 2006, Doxey et al., 2005). The condition is associated with various extra colonic tumours (Table 1.7) and individuals with FAP require screening for these conditions as part of their management. Indeed, as modern treatment has greatly reduced the mortality from CRC in FAP, these other tumours, particularly the upper gastrointestinal and desmoid tumours, are becoming more prevalent in FAP (Merg et al., 2005a, Galiatsatos and Foulkes, 2006).

Site	Lifetime risk (%)
Desmoid	15
Duodenum	3-5
Thyroid	2
Brain	2
Ampullary	1.7
Pancreas	1.7
Hepatoblastoma	1.6
Gastric	0.6

Table 1.7: Extra colonic tumours associated with FAP.

FAP is diagnosed by the presence of a hundred or more adenomatous colorectal polyps at endoscopy (Jarvinen, 2004). Diagnosis should be followed by screening for associated tumours, establishing any family history of hereditary CRC and genotyping for *APC* mutations. Offspring of known *APC* mutation carriers should be genotyped as half will inherit the mutation, and those that are found to be carriers should be endoscopically screened from early adolescence (Jarvinen, 2004, Galiatsatos and Foulkes, 2006, Maple and Boardman, 2006). Colectomy is the treatment of choice, either due to the presence of polyposis, symptomatic polyposis or prophylactically at an appropriate time in screened *APC* mutation carriers (Maple and Boardman, 2006, Galiatsatos and Foulkes, 2006, Jarvinen, 2004).

Many different mutations of the *APC* gene have been demonstrated in *APC*, but the majority give rise to a truncated, dysfunctional protein product (Doxey et al., 2005). The *APC* gene primary function is in cellular adhesion and migration (Galiatsatos and Foulkes, 2006). It has a fundamental tumour suppressor role in antagonising the Wingless/Wnt cellular signalling pathway via its interaction with β-catenin (Merg et al., 2005a, Lipton and Tomlinson, 2006, Doxey et al., 2005). *APC* mutations lead to impaired β-catenin degradation which thus accumulates within the cytoplasm and

activates the Wnt signalling pathway, leading to increased cell proliferation (Doxey et al., 2005) . *APC* is also involved with several other cellular proteins including α-catenin, GSK3β, axin, (Wnt signalling proteins) and conductin, tubulin, EB1 (microtubule and cell migration proteins) and hDLG (cell adhesion and motility) and has roles in cell cycle control by regulating $G_0/G_1 > S$ phase progression, apoptosis and neuronal differentiation (Hanson and Miller, 2005, Galiatsatos and Foulkes, 2006).

Therefore *APC* mutations can lead to increased cellular proliferation, decreased apoptosis, chromosomal instability and impaired cell cycle control, all important factors in carcinogenesis.

Disease severity in FAP has been shown to be linked to the position of mutations in the *APC* gene, so-called genotype-phenotype correlations, and demonstrates how the effect of mutation on gene function can be dependent on where in the genome the mutation occurs (Galiatsatos and Foulkes, 2006, Merg et al., 2005a). Codon 1309 *APC* mutations lead to an accelerated form of FAP where polyposis occurs at an earlier age, is more florid and malignant change is more rapid. Mutations between codons 976-1067 have a high incidence of duodenal tumours and mutations between codons 543-1309 are associated with congenital hypertrophy of retinal pigment epithelium (CHRPE), a condition leading to pigmented retinal lesions that occasionally can give rise to malignant transformation(Sheilds et al., 2001) and as they are congenital they are an important, early sign in *APC* mutation carriers. Mutations that occur at the extreme ends of the *APC* gene lead to "attenuated FAP", manifest by fewer polyps, later age of onset and lower risk of CRC.

APC gene function has also been found to be impaired by single nucleotide polymorphisms (SNP). A thymine to adenine transversion at nucleotide 3920 (3920 T>A) leads to isoleucine being substituted with lysine at codon 1307 (I1307K)

(Laken et al., 1997). This missense mutation leads to an increased risk of developing adenomas and CRC, but not the polyposis as seen in classical FAP. The SNP is carried by 6% of Ashkenazi Jews and in around 28% of Ashkenazim with a family history of CRC.

1.1.11.3 Hereditary Non-polyposis Colorectal Cancer (HNPCC).

HNPCC (also called Lynch syndrome) is manifest by the development of proximally sited CRC at a young age and is due to autosomal dominantly inherited mutations in MMR genes. The tumours develop from solitary polyps, in contrast to FAP (Jarvinen, 2004). The life time risk of developing CRC is in the order of 75-80%. The condition is associated with other, extra-colonic tumours and multiple family members are affected with either CRC or the other related tumours (Table1.8).

Site	Frequency (%)
Colon	80
Endometrium	40
Stomach	15
Ovary	12
Urothelium	5
Other (small bowel, pancreas, brain)	<5

Table 1.8: Cancers associated with HNPCC.

Clinical diagnosis is achieved when cases fulfil the Amsterdam criteria (Table 1.9) and those cases should be referred for genetic testing.

1. Three (or more) family members with HNPCC related cancer; one a first degree relative of the other two.

2. HNPCC cancer in at least two generations.

3. At least one individual diagnosed before 50 years of age.

4. FAP excluded.

Table 1.9: Amsterdam criteria for clinical definition of HNPCC(Vasen et al., 1999).

Those familial cancers that do not fulfil the Amsterdam criteria are evaluated according to the revised Bethesda guidelines (Table 1.10) and undergo genetic testing accordingly.

1. CRC diagnosed at less than 50 years of age.

2. Presence of synchronous or metachronous HNPCC related tumours regardless of age.

3. Microsatellite instability "high" CRC tumours in individuals less than 60 years of age.

4. CRC diagnosed in one or more 1st degree relatives with an HNPCC- related tumour, with one of the tumours diagnosed under the age of 50 years.

5. CRC diagnosed in two or more 1st or 2nd degree relatives with HNPCC-related tumours, regardless of age.

Table 1.10: Revised Bethesda guidelines.

Individuals who satisfy either the Amsterdam criteria or the Bethesda guidelines should be genetically tested to determine whether they carry germ line mutations in any of the DNA mismatch-repair (MMR) genes (Table 1.11).

Gene	Locus	Percentage of HNPCC
MHS2	2p22-p21	60%
MLH1	3p21.3	30%
MSH6	2p16	5%
PMS2	7p22	minimal
PMS1	2q31-q33	minimal
TFGBR2	3p22	minimal
MLH3	14q24.3	Minimal

Table 1.11: MMR genes mutated in HNPCC.

MMR genes function to correct errors of base-base mismatch and insertion-deletion loops that form during DNA replication. Base-base mismatches lead to single base substitutions and insertion-deletion loops involve gains or losses of short repetitive DNA nucleotide units (microsatellites) a phenomenon known as microsatellite instability (MSI). Although most microsatellite sequences are found within non-coding genomic regions certain growth regulatory genes have DNA microsatellites within their coding regions. Such genes include the *transforming growth factor-β receptor II* and *insulin-like growth factor II receptor* genes, the *E2F transcription factor 4* gene, the *p53 regulated proapoptotic BAX (BCL2-associated X protein)* gene and the MMR genes *hMSH3* and *hMSH6* (Jo and Chung, 2005). Loss of MMR function leads to accumulation of mutations in these genes and hence loss of cellular growth regulation, the hall-mark of cancer development. Defective MMR function leads to 100 or 1000 fold mutation rates in affected cells and so individuals who possess a germ-line mutation of an MMR gene are at great risk of accumulating mutations and thus developing cancer at a younger age than the general population (Abdel-Rahman and Peltomaki, 2004).

HNPCC is managed by aggressive screening with resection of identified high grade dysplastic lesions or carcinomas.

1.1.11.4 *MYH* Polyposis.

Individuals with FAP phenotype have been discovered who do not harbour an *APC* mutation. Genetic analysis has shown that mutations of the *E. coli mutY homologue* gene (*MYH*) at locus 1p34.3-p32.1 are responsible. Unlike the other inherited CRC genes the *MYH* gene has a recessive mode of inheritance, however mono- and bi-allelic carriers have an increased risk of CRC (Doxey et al., 2005). *MYH* is a base excision repair gene involved in repairing oxidative DNA damage due to reactive oxygen species (Doxey et al., 2005, Lipton and Tomlinson, 2006, Galiatsatos and Foulkes, 2006). Oxidation of guanine (G) nucleotides leads to the formation of 8-oxo-7, 8-dihydro2'deoxyguanosine (8-oxoG) which mispairs with adenine (A) rather than cytosine (C). This then leads to G:C to T:A transversion mutations in the daughter strand during DNA replication. *MYH* scans the daughter strands for such mutations and excises them. Losses of *MYH* function due to mutations leads to preservation of G:C to T:A transversions. When such mutations occur in *APC*, for example, polyps and carcinoma result.

85% cases of *MYH*-associated polyposis are due to 2 mutations; tyrosine 165 to cystine (Y165C) and glycine 382 to aspartine (G382D).

1.1.11.5 Bloom's Syndrome.

Bloom's syndrome is a very rare condition associated with multiple site cancers (Jo and Chung, 2005). There is a characteristic phenotype of short stature, photosensitivity, variable skin pigmentation, diabetes and infertility.

The syndrome is due to germ line mutations in the *RecQ helicase* gene family member, *BLM*, located at chromosome 15q26.1. BLM helicase functions to re start DNA replication at arrested or collapsed replication forks (Kaneko et al., 2004).

BLM mutations lead to chromosomal instability due to destabilization of enzymes involved in DNA replication and repair (Ellis et al., 1995) and hence tumorigenesis. Like *MYH* polyposis, BLM is recessively inherited.

1.1.11.6 Hamartomatous Polyposis Syndromes.

CRC arising from hamartomatous polyps is seen in the autosomal dominant Peutz-Jeghers (PJS) and Juvenile polyposis (JPS) syndromes (Jo and Chung, 2005, Lipton and Tomlinson, 2006, Jarvinen, 2004).

PJS is due to mutations of the tumour suppressor gene *STK 11* (previously named *LKB 1*) located at chromosome 19p13.3. STK 11 is a serine-threonine kinase with many functions, including p53-dependent apoptosis regulation (Karuman et al., 2001), cell cycle arrest (Tianen et al., 2002) and maintaining cellular polarity (Martin and Johnston, 2003).

JPS is caused by mutations in the mothers against *decapentaplegic, Drosophila, homolog of, 4 (SMAD4)* gene on chromosome 18q21.1 or *bone morphogenetic protein receptor type 1A (BMPR1A)* gene on chromosome 10q22.3. SMAD4 is involved in the intracellular transforming growth factor β (TGF- β) signalling pathway (Zhou et al., 1998). BMPR1A is a member of the TGF- β super family and is involved in a SMAD4 dependent intracellular signalling and transcription regulation bone morphogenetic protein (BMP) pathway (Merg et al., 2005b). Mutations of *SMAD4* and *BMPR1A* lead to increased cell proliferation and dysplasia, and hence CRC.

1.1.12 Sporadic CRC.

The majority of CRC arise spontaneously, that is, in individuals with no obvious genetic or familial risk factors. The hereditary CRCs, although accounting for only

a small proportion of CRC cases, illustrate some of the genetic mechanisms involved in sporadic colorectal carcinogenesis, for example the *APC* and the MMR genes have been shown to be involved in sporadic CRC.

As previously stated, CRC (and indeed all cancers) arise due to accumulation of mutations in key regulatory genes involved in cell cycle control and growth, apoptosis, DNA surveillance and repair, and genes that are important for cancer invasion and metastasis such as those involved in cell adhesion, migration, proteolysis and angiogenesis. Also involved are epigenetic mechanisms of DNA methylation and histone modification. Analysis of the molecular pathology of CRC demonstrates that carcinogenesis follows various genetic pathways (Takayama et al., 2006).

1.1.12.1 Chromosomal Instability Pathway.

85% of sporadic CRC tumour cells show aneuploidy (abnormal quantities of DNA) as a result of chromosomal losses and gains (Terdiman, 2000). Progression along the adenoma-carcinoma sequence is mirrored by chromosomal losses on 5q (*APC*), 17p (*TP53*) and 18q (*DCC/SMAD*) (Vogelstein et al., 1988).

APC (already discussed above) mutations occur early in adenoma development and may be the initiating event in colorectal carcinogenesis (Heinen et al., 2002).

K-Ras mutations also occur during the adenomatous stage and are found with increasing frequency with increasing adenoma size and dysplasia (Howe and Guillem, 1997). *K-Ras* controls various downstream genes involved in normal cellular proliferation and differentiation (Leslie et al., 2002). *K-Ras* mutation leads to activation and confers a growth advantage to the cell.

The transition of adenomas to carcinomas is thought to be due to mutations in *TP53* (Dietel, 1998). *TP53* mutations are found in most CRC tumours and, indeed, in

many other cancers. It has been termed "the guardian of the genome" (Navaratnam et al., 1999) as it has a wide range of functions involved in cell cycle control, DNA repair and apoptosis via its control of a large number of genes. CRCs displaying TP53 mutations have been shown to carry a worse prognosis than those where TP53 function is preserved (Pricolo et al., 1997).

Most advanced and invasive tumours show chromosome 18q losses. *DCC (Deleted in Colorectal Cancer)* was originally thought to be the candidate for these loses, but this has not been supported by subsequent research. (Leslie et al., 2002, Takayama et al., 2006). *SMAD* genes are also found at this locus and are involved in the TGF-β cell signalling pathway which itself is involved in cell growth, differentiation, matrix production and apoptosis. *SMAD2* and *SMAD4* mutations lead to inactivation of the TGF-β cell signalling pathway (Fukushima and Takenoshita, 2001, Terdiman, 2000).

Table 1.12 summarises the genes involved in the chromosomal instability pathway.

Gene	Chromosome	Function
Tumour suppressor genes		
APC	5q21	Inhibition of cell growth.
TP53	17p12	G1 cell cycle arrest/apoptosis.
SMAD	18q21	P15/p21 induced growth arrest.
DCC	18q21	Cell-cell interaction/apoptosis
Oncogenes		
K-Ras	12p12	Growth promotion
β-catenin	31q21	Cell adhesion/ proliferative signalling

Table 1.12: Genes involved in the chromosomal instability pathway.

1.1.12.2 Microsatellite Instability Pathway.

MSI has already been discussed in the context of HNPCC above, however MSI has also been recognised in sporadic CRC where up to 30% cases are MSI positive (Grogan et al., 1996). MSI positive CRC tumours can be classified as low- or high-frequency MSI (MSI-L or MSI-H). MSI-L tumours display only one to two altered loci and are associated with *APC*, *TP53* and *K-Ras* mutations whereas MSI-H tumours have three or more altered loci and have fewer *APC*, *TP53* etc mutations, indicating that MSI can lead to the malignant phenotype without chromosomal instability and that in some tumours both pathways contribute (Baba, 1997). Table 1.13 summarises the genes involved in the microsatellite instability pathway.

Gene	Chromosome	Function
TGF-β Receptor II	3p22	Cell growth inhibition
BAX	19q13.3-q13.4	Apoptosis induction
IGF-II receptor	6q26	Growth promotion
hMSH6	2p16	MMR
hMSH3	5q11-q12	MMR
PTEN	10q23.32	Cell growth inhibition
E2F-4	16q22.1	G1 cell cycle arrest

Table 1.13: Genes involved in the microsatellite instability pathway.

1.1.12.3 TGF-β /SMAD Signalling Pathway.

The TGF-β/SMAD pathway has been described previously in the context of the hereditary hamartomatous polyposis syndromes. TGF-βRII complexes with TGF-βRI as a result of TGF-β binding. This complex phosphorylates SMAD2 which

forms a complex with SMAD4 that induces p15 and p21 within the cell nucleus. P15 and p21 are Cdk inhibitors and function to arrest cell growth (Takayama et al., 2006). Up to 25% of sporadic CRC tumours have been shown to have inactivating *SMAD4* mutations, and around 6% have *SMAD2* mutations. *TGF-βRII* mutations are common in MSI positive tumours, and in those MSI positive tumours without *TGF-βRII* mutations, *IGF-II* mutations are frequently found. *TGF-βRII* and *SMAD* mutations are associated with adenoma-carcinoma transition. The TGF-β signalling pathway also interacts with the Wnt pathway to promote CRC carcinogenesis, illustrating how the various molecular pathways interact with each other to produce the malignant phenotype.

1.1.12.4 Genetics of Invasion and Metastasis.

Invasion and the ability to metastasise is the hallmark of malignancy. Many processes are involved including digestion of the extra-cellular matrix (ECM), cell-cell adhesion, angiogenesis, migration and evading the immune system that allow for tumour cells to invade tissues adjacent to their tissue of origin and to metastasise to distant sites. CRC tumour cells have been shown to express genes that confer these characteristic upon them (Table 1.14) (Takayama et al., 2006, Kountouras et al., 2000).

Gene	Chromosome	Function
	Proteolysis genes	
MMP-1,-8,-13	11q21-q23	Collagen I, II, III, IV, VI, IX, X and XI digestion
MMP-2, -9	16q13, 20q11.2-q13.1	Gelatins and collagen IV digestion
MMP-3	11q23	Fibronectin and laminin digestion
MMP-7	11q21-q22	Fibronectin, laminin, collagen IV and proteoglycans digestion
TIMP-1	Xp11.3-p11.23	MMP inhibition
uPAR	19q13	Plasminogen activation regulation
	Adhesion genes	
Integrins		Extra-cellular matrix component binding
Cadherins		Cell-cell adhesion
CD44	11pter-p13	Matrix adhesion lymphocyte activation
CEA	19q13.2	Kupffer cell binding
	Angiogenesis genes	
VEGF	6p12	Mitogen for vascular endothelial cells
PD-ECGF	22q12.32-qter	Angiogenesis
	Miscellaneous genes	
Tumour Necrosis Factor Receptor Family		Apoptosis regulation
CXCR4	2q21	CRC cell migration/metastasis
Drg-1		Cell differentiation
c-Met		Hepatocyte growth factor receptor, enhances metastatic growth

Table 1.14: Genes involved in CRC invasion and metastasis.

1.1.12.5 Epigenetics of CRC.

Epigenetics refers to heritable processes distinct from DNA sequence modification that determine gene expression such as histone modification, DNA methylation, and RNA modifications (Wong et al., 2007). Histones form the core of chromosomes around which DNA is wrapped and are subject to reversible modifications that have a role in regulating gene expression. DNA methylation is involved in regulating differential gene expression such as X chromosome silencing, age related and tissue specific gene expression. Abnormalities in epigenetic mechanisms have been shown to have a role in carcinogenesis and have been demonstrated in CRC (Kondo and Issa, 2004). Both hypo- and hypermethylation of DNA have been implicated in carcinogenesis. Hypomethylation is thought to increase chromosomal instability. Hypermethylation of promoter region of various tumour suppressor, MMR and cell-cycle regulatory genes leads to their silencing (Wong et al., 2007). Epigenetic abnormalities tend to occur early in CRC development, and are even seen in normal epithelium, and so may serve as a marker for CRC risk.

The development of the understanding of the genetic mechanisms involved in CRC has provided a great insight into the molecular pathology of cancer globally. It has also led to the understanding that tumour biology and behaviour is related to tumour genotype. Genetic staging of tumours may provide more accurate prognostic information than traditional pathological staging, and may also allow better prediction of therapeutic response to adjuvant therapies.

1.1.12.6 Genetic Predisposition to Sporadic CRC.

The hereditary CRCs show how highly penetrant genetic mutations greatly increase the risk of developing CRC and the genetic pathways described above shows how

the accumulation of mutations drives tumour development. However exactly how sporadic CRC is initiated is not well understood. The environmental risks have been discussed above and it is of considerable interest whether there are heritable genetic factors that increase the risk of developing sporadic CRC. Knowledge of such genes would be of benefit in allowing more accurate risk stratification than is currently possible on the basis of family history. It is clear that similar environmental exposure in different individuals does not necessarily result in cancer in all those individuals, therefore there must be intrinsic, and most probably genetic, factors modulating environmental risk.

There has been considerable research into the role of low penetrance genetic variants on CRC risk and although there are controversies and difficulties with such studies there has been some progress in identifying such risk factors (Kemp et al., 2004). Single nucleotide polymorphisms (SNPs) are thought to be the most likely candidates for influencing risk and have been subject to intense study. Disease-gene association studies have identified *APC*-I1307K, *HRAS1-VNTR*, and *MTHFR* polymorphisms as influencing the increased risk of CRC (Houlston and Tomlinson, 2001) however the effect is modest. Studies have also looked at the effect of gene-gene and gene-environment interactions on CRC risk (Goodman et al., 2006, Webb et al., 2006). The role of SNPs in CRC is discussed in greater detail later.

1.1.13 Treatment.

Treating CRC has one of two aims depending on the stage of disease at presentation. Those tumours that are at an early stage, without evidence of local invasion or distant metastases, are potentially curable with radical surgical excision. Those tumours that have invaded local tissues or that have metastasised to distant organs are considered incurable and their treatment is palliative. Curative treatment

depends of completely removing all malignant tissue with a wide margin of normal tissue by surgical excision. Patients who have had excised tumours that are shown on histological examination to have lymph node metastases are offered chemotherapy as this treatment has been shown to prolong survival. There is currently increasing interest and enthusiasm for excising solitary hepatic metastases, both those found at presentation and those that develop post curative surgery. For rectal cancers radiotherapy has a role, both before surgery where it may help to reduce the size of tumours in order to render them excisable and post operatively if there are found to be involved lymph nodes or margins.

Palliative treatment involves controlling the symptoms of CRC. This may involve surgery to bypass obstructing tumours or chemotherapy to try to slow the progression of metastatic disease.

At present the only hope of cure from CRC rests on excising early stage tumours. In order to reduce the mortality from CRC it is necessary to increase the proportion of cases that are diagnosed at stage I or II. However these stages tend to be relatively asymptomatic, therefore screening may be the only way to detect these tumours and so achieve lower mortality from CRC.

1.1.14 Screening.

In 1968 the World Health Organization published screening guidelines and specified the criteria for screening programs (Wilson and Junger, 1968):

1. The condition should be an important health problem.

2. There should be a treatment for the condition.

3. Facilities for diagnosis and treatment should be available.

4. There should be a latent stage of the disease.

5. There should be a test or examination for the condition.

6. The test should be acceptable to the population.

7. The natural history of the disease should be adequately understood.

8. There should be an agreed policy on who to treat.

9. The total cost of finding a case should be economically balanced in relation to medical expenditure as a whole.

10. Case-finding should be a continuous process, not just a "once and for all" project.

In 2003 the UK National Screening Committee developed similar criteria for screening. Although screening programmes are far from controversial, screening for breast and cervical cancers have led to a reduction in disease related mortality and disease incidence respectively (Blanks et al., 2000, Sasieni et al., 2003). CRC would seem for screening and indeed screening has been shown to reduce the risk of dying from CRC (Hewitson et al., 2007).

CRC screening should be considered in terms of high risk groups screening and population screening. High risk groups include the hereditary and familial CRCs; patients who have had a CRC or high risk adenoma resected and patients with long standing inflammatory bowel disease. Guidelines have been published with regard to these groups in order to aid clinicians (Dunlop, 2002a, Dunlop, 2002b, Scholefield and Steele, 2002, Atkin and Saunders, 2002, Eaden and Mayberry, 2002).

Population screening involves inviting asymptomatic individuals from the general population to under go a screening test. In terms of CRC screening can be performed using either radiological or endoscopic imaging or by faecal occult blood (FOB) detection, tests which have been shown to be adequately sensitive and specific. Currently the National Health Service is rolling out a bowel cancer

screening programme in England using FOB detection testing (www.cancerscreening.nhs.uk/bowel). All men and women aged between 60 and 69 will be invited to be screened every two years. It is anticipated that two percent of those screened will have an abnormal test and these individuals will then be referred for further investigation.

1.1.14.1 Genetic Screening.

Genetic screening for sporadic CRC depends on a comprehensive knowledge of the genetic factors that influence the risk of individuals developing CRC. This knowledge would allow an individual's genetic risk to be established and so inform whether targeted screening and behaviour modification would be appropriate (Burn et al., 2001). To this end there has been a huge drive to both identify susceptibility genes and develop the technology to allow rapid and efficient genotyping (Collins et al., 1998b, Stremmel et al., 2002). However, in contrast to the "Mendelian CRCs", FAP and HNPCC, the identification and disease risk determination of the genes is far more complex and has necessarily led to the development of new epidemiological methodologies (2007, Goodman et al., 2006, Harkin, 2006). Although several genes have been shown to be associated with CRC there is as yet insufficient information to be able to determine the overall effect on the risk of developing CRC and so determining individual CRC risk profiles. Therefore there are currently no genetic screening programmes for CRC, other than those for the Mendelian CRC syndromes.

1.1.15 Summary.

CRC is an important public health issue leading to many deaths each year. Current management depends on diagnosing symptomatic individuals or screening high risk individuals with inherited CRC syndromes. This approach results in a high proportion of sporadic CRC cases being diagnosed at an advanced stage and as a consequence mortality rates remain high. Population screening for CRC would lead to earlier diagnosis and should help to reduce mortality, as has been seen with breast cancer. Screening programmes being developed are based on occult bleeding tests or endoscopic surveillance.

Although the genetics of the inherited CRC syndromes and the adenoma-carcinoma sequence are well understood, low penetrance genetic susceptibility factors are less well understood and so population genetic screening has yet to be realised. However as knowledge of this area increases and technologies for rapid genotyping improve genetic screening for CRC susceptibility will become a reality.

1.2 SINGLE NUCLEOTIDE POLYMORPHISMS.

1.2.1 Introduction.

Single nucleotide polymorphisms (SNPs) are the most abundant form of genetic variation accounting for about 90% of sequence differences in the human genome (Wang et al., 1998, Wang and Moult, 2001). 85% of SNPs are common to all human populations and arose within the human genome after speciation but before the emergence of different populations, and so the majority of human genetic variance occurs within rather than between populations. SNPs are found scattered throughout the genome and account for several million single base differences between individuals. Coding SNPs result in around 100 000 amino acid differences in our proteomes.

SNPs may hold the key to understanding the genetics of common complex diseases such as diabetes, heart disease and cancer (Collins et al., 1998a). Common complex diseases are thought to arise due to both gene-environment and gene-gene interactions where variations in multiple interacting polymorphic genes modulate the risk of developing disease states (Kirk et al., 2002, Rueff et al., 2002). The role of SNPs as research tools has been shown to be important in the investigation of the molecular biology of common complex human diseases and a major part of the Human Genome Project is the cataloguing of SNPs (Risch and Merikangas, 1996, Collins et al., 1997, Wang et al., 1998). The power of SNPs lies in their use as genetic markers for identifying disease genes by association studies of candidate genes (Collins et al., 1998b) and in their suitability for automated gene identification techniques using micro array technology, where multiple candidate SNPs can be screened simultaneously (Friend and Stoughton, 2002). Knowledge of

an individual's SNP profile may allow disease susceptibility assessment and direct tailored screening and prevention strategies for at risk diseases (Brown et al., 2003).

1.2.2 Definition.

SNPs can be defined as "Single base pair positions in genomic DNA at which different sequence alternatives (alleles) exist in normal individuals in some population(s), wherein the least frequent allele has an abundance of 1% or greater" (Brookes, 1999).

1.2.3 Classification.

SNPs are classified as transitions, where a pyrimidine (Cytosine or Thymine) is substituted by a pyrimidine or a purine (Adenine or Guanine) is substituted by a purine, or transversions, substitution of a pyrimidine by a purine or vice versa.

Two thirds of SNPs are represented by the Cytosine to Thymine transition a disproportionately high frequency due the high frequency of 5-methlycytosine deamination reactions that occur at CpG dinucleotides (Brookes, 1999).

1.2.4 SNP genomic location.

In predicting the effects of SNPs on gene function knowledge of the location of the SNP in the genome is important and whether the SNP is transcribed and translated. SNPs in coding regions (cSNPs) and regulatory regions are most likely to affect gene function as these have the greatest potential to lead to altered protein structure (and hence function) or to variations in gene expression (Wang and Moult, 2001).

1.2.4.1 Intronic SNPs.

Introns are non-coding sections of DNA that are not translated into the protein product. Introns are transcribed with exons but are removed by post-transcriptional mRNA modification by a process know as splicing. Splicing takes place at 'splice sites' at the exon-intron boundaries. Correct splicing is essential for gene expression. Introns may also be involved in regulating gene expression.

SNPs that fall within introns may activate cryptic splice sites or may eliminate splice sites, leading to alternative genomic splicing and gene transcription, or may interfere with gene expression. Either mechanism may lead to gene dysfunction and disease.

Examples of diseases caused by intronic mutations include some forms of congenital adrenal hyperplasia (Higashi et al., 1988, Steinberg and Adams, 1982), most of the beta thalassaemias (Steinberg and Adams, 1982) and an autosomal dominant form of isolated growth hormone deficiency (McCarthy and Phillips, 1998).

1.2.4.2 Exonic SNPs.

Exons are coding or non-coding (untranslated) DNA sequences that occur between introns. The translation of the coding sequences gives rise to the protein product of genes.

1.2.4.2.1 Coding SNPs.

SNPs that occur with in the coding DNA can give rise to various mutations:

1.2.4.2.2 Synonymous (silent) mutations.

These are the most frequent SNP mutations in coding DNA. The mutation is neutral and involves nucleotide base substitution that creates a codon that codes for the same amino acid as the codon containing the wild type polymorphism. They tend to occur at the 3rd base position of the codon and can be referred to as "3rd base wobble".

1.2.4.2.3 Nonsense mutations.

These non-synonymous nucleotide substitutions create a stop codon from an amino acid codon leading to premature ending of translation, and therefore can profoundly interfere with gene function. Because of this they are under great selection pressure and are rare.

1.2.4.2.4 Missense mutations.

These are also non-synonymous nucleotide substitutions that create codons for different amino acids than the wild type. They can be conservative, where the new amino acid is chemically similar to the wild type thus having little functional effect on the protein product, or non-conservative where the substituted amino acid is chemically distinct and can alter the protein product morphology thus interfering with its function. Missense SNPs account for half of all pathogenic genetic mutations (Wang and Moult, 2001).

1.2.4.3   Untranslated region SNPs.

Untranslated regions (UTRs) flank genomic exons at the 5' and 3' ends. They are transcribed to mRNA and assist in binding and stabilising ribosomes during translation.

The 3' UTR has a key role in regulating, stabilising and localizing translation (Wickens et al., 1997, Day and Tuite, 1998, Siomi and Dreyfuss, 1997). SNPs in these regions could therefore interfere with the binding of ribosomes and other RNA binding proteins, thus interfering with gene function by modulating mRNA translation.

UTR SNPs may enhance or reduce gene function. It has been shown that a 3' UTR SNP of the human dihydrofolate reductase gene is associated with enhanced gene expression (Goto et al., 2001).

1.2.5   Use of SNPs.

SNPs have three primary uses; candidate polymorphism testing, diagnostics and risk profiling and prediction of response to xenobiotics (Schork et al., 2000). Candidate polymorphism testing involves selecting polymorphisms that are potentially associated with diseases and performing case-control studies to test that association. Selecting polymorphisms requires a reasonable knowledge of a diseases molecular pathology in order to target the studies. For example it would be reasonable to study polymorphisms of cell cycle control genes in cancer studies, as loss of cell cycle control is fundamental to carcinogenesis. In these types of study the polymorphisms could be causally associated, where they lead to altered gene function and disease phenotype, or they may be in linkage disequilibrium with disease genes. There has been a surge in these types of studies in the past few years

and in order to facilitate this interest in candidate polymorphism disease association there have been determined efforts to catalogue human SNPs (Collins et al., 1997, Wang et al., 1998, Collins et al., 1998b).

If SNPs are found to be associated with diseases then the risk of developing the disease can be attributed to either individual SNPs, SNP haplotypes or to an individual's SNP profile. Genetic profiling could be used for targeted disease screening, to ameliorate disease risk by adjusting environmental and life-style risk factors and to tailor disease management.

One of the potential phenotypic effects of SNPs is variation in drug metabolism leading to variation in response to the same drug between individuals. Knowledge of the variation of response of individuals to xenobiotics due to SNPs can be applied to drug design and knowledge of an individual's SNP profile could lead to individualised pharmacotherapy. This may be particularly applicable to variations in response of tumours to chemotherapy.

1.2.6   SNPs and carcinogenesis.

Whilst some cancers, for example familial adenomatous polyposis and familial retinoblastoma, are known to arise due to high-penetrance dominant genes, the majority of common cancers, such as lung, breast, colorectal and skin cancer, arise sporadically. These cancers, like other common complex diseases, are thought to be caused by the interaction of multiple low-penetrance genes both with themselves and with the environment to which the individual is exposed (Shields and Harris, 2000). SNPs are likely candidates for low-penetrance disease risk genes as they are so abundant and widespread throughout the genome. The risk that individual SNPs confer to an individual depends on the attributable risk of the SNP itself, the presence of other genes influencing disease and the environmental risks to which

the person is exposed. Therefore determining genetic risk is a complex process that must take into account all these factors. However, SNPs have been extensively studied with regard to cancer association, both due to the biological plausibility of their being the reason for inter-individual variation in cancer risk and due to their accessibility as they are abundant and easy to detect (Shields and Harris, 2000, Imyanitov et al., 2004, Houlston and Peto, 2004).

Research efforts at establishing SNP-cancer association have concentrated on the classes of genes known to be involved in carcinogenesis; xenobiotic metabolizers, hormone metabolizers, DNA repair genes, cell cycle control genes, angiogenesis genes, as well as with the oncogenes and tumour suppressor genes known to be involved in mutagenesis (Imyanitov et al., 2004).

Xenobiotic metabolizers are those enzymes involved in the detoxification of environmental carcinogens. Examples of extensively studied polymorphic xenobiotic metabolizers considered likely to influence carcinogenesis include the glutathione-S-transferases (GSTM1, GSTT1, GSTP1, GSTM3), N-acetyltransferases (NAT1 and NAT2), and cytochromes (CYP2D6, CYP2E1 and CYP1A1) (Benhamou et al., 2002, Lower et al., 1979, Ayesh et al., 1984, Seidegard et al., 1986, Vineis et al., 2003, Vineis et al., 2001) . The results of such studies have, however, yielded modest odds ratios and not led to the genes being utilised for clinical predictive testing (Imyanitov et al., 2004).

Breast and prostate cancers are known to be endocrine related, meaning that circulating hormones (specifically steroid hormones in these cases) influence their growth. Therefore attention has been focused on the influence of polymorphic variations of steroid metabolizer genes on the development of these cancers (Mitrunen and Hirvonen, 2003, Nelson and Witte, 2002). Again the results of these studies have been disappointing.

Failure of DNA damage repair is known to be fundamental to the development of many cancers and many of the highly penetrant genes (*TP53, BRCA1, MSH1, PMS1* etc) predisposing to cancer are involved in DNA damage surveillance and repair. These genes have therefore been targeted for studying the effects of low penetrance polymorphic variations on cancer susceptibility (Berwick and Vineis, 2000).

1.2.7 SNPs and Colorectal cancer.

CRC has and continues to be an important subject for SNP studies. This is due to it being the third most common cancer, accounting for 30,000 new cases and 17,000 deaths every year in the United Kingdom, and due to the fact that three quarters of those cases arise sporadically. Although the major environmental and genetic risk factors have been identified, it is clear that there is inter individual variation in disease susceptibility and that this variation is likely to be due to low penetrance genes, particularly SNPs (Kemp et al., 2004, Houlston and Tomlinson, 2001). Combined with the development in SNP technology allowing rapid identification of SNPs this has led to a large number of SNP-CRC association studies being performed with the aim of defining the risk of low penetrance genes in CRC and ultimately to be able to define an individual's risk according to their genotype (Harkin, 2006).

Most of the studies so far have concentrated on SNPs in candidate genes that are likely, or known, to be involved in colorectal carcinogenesis; carcinogen metabolizers (*CYP1A1, NAT1, NAT2, GSTM1, GSTT1, GSTP1, mEPHX*), methylation enzymes (*MTHFR, MTR*), DNA repair genes (MMR genes), micro environmental modifiers (*APOE, PLA2G2A*), oncogenes and tumour suppressors (*HRAS1, L-myc,TP53*) (Mort et al., 2003, Jong et al., 2002, Houlston and Tomlinson, 2001). Pooled analysis of studies of 30 polymorphisms demonstrated an

increased risk of CRC associated with *GSTT1, NAT2, HRAS1* and *ALDH2* polymorphisms; *MTHFR, TP53* polymorphisms showed a decreased risk; TNF-α displayed both an increased and decreased risk with different alleles (Jong et al., 2002). However the results of individual studies are inconsistent. A recent large study of 1467 SNPs demonstrated CRC risk associated with A kinase anchor protein 9 (*AKAP9*) M463I, ataxia telangiectasia mutated (*ATM*) F858L and P1054R, and methylenetetrahydrofolate reductase (*MTHFR*) A222V (Webb et al., 2006). Polymorphism interaction analysis (PIA) has identified the *GSTT1-02* SNP as a strong predictor of CRC and that the presence of p53 Arg72Pro or CASP8 Asp302His SNPs in combination with the *GSTT1-02* SNP modifies the CRC association (Goodman et al., 2006).

Although a great deal of work has been done looking at the relationships between SNPs and CRC, there is still insufficient data available to be able to use SNPs as a screening tool for CRC. Only with further research and a greater understanding of the role of SNPs and their influence on CRC risk will SNP screening become a reality. Integral to this work is to identify candidate CRC association SNPs.

Reprimo and *p53R2* are genes involved in DNA repair and cell cycle control. SNPs in these genes may lead to defective function and so defective DNA repair and cell cycling, events necessary for carcinogenesis. Neither gene has been extensively studied for CRC association so they would make ideal candidate genes for such studies and are the subjects of this thesis. They are described in more detail below.

1.2.8   Summary

Thus it can be seen that SNPs are potentially a powerful genetic tool. They can be used in population genetics studies as markers of disease genes, association analysis of patients and controls, and loss-of-heterozygosity in tumours. In considering their relevance to the development of disease, the fact that they constitute the most frequent genetic variations and therefore potential impact on clinical phenotypes, leads to the hypothesis that the risks of common, complex diseases, including cancer, are influenced by SNPs in key susceptibility genes. Although individual SNPs are unlikely to cause disease it is possible that combinations of SNPs can modify the risk of developing diseases. Developing an understanding of the relationship between SNPs and disease risk would potentially allow screening for diseases as, by definition, SNPs are inherited. Knowledge of an individual's SNP genotype could be used to make risk assessments of their potential clinical phenotype and disease susceptibility and therefore guide therapy.

.

1.3   *P53R2.*

1.3.1   Alternative names.

p53-inducible ribonucleotide reductase small subunit 2 homologue.

1.3.2   Symbol.

P53R2.

1.3.3   Genebank accession number.

AB036532

1.3.4   OMIM accession number.

604712

1.3.5   Locus and size.

P53R2 is a 6213 base gene located at chromosome 8q23.1. It contains a coding region from nucleotide position 245 to 1300.

1.3.6   Function.

p53R2 has been shown to be a direct target for the tumour suppressor gene *TP53* and contains a 351 amino acid protein containing a 20-nucleotide p53 binding site

at intron 1 (Tanaka et al., 2000). It has also been shown to be induced by p73, one of the *TP53* gene family (Nakano et al., 2000). *TP53* is known to have an important function in regulating cell cycle arrest or apoptosis in response to genotoxic stress and DNA damage thus preventing proliferation of damaged cells, an important step in carcinogenesis (Nakamura, 2004). P53 exerts its function via control of a large family of target genes involved in cell cycle arrest, apoptosis, DNA repair, inhibition of angiogenesis, oxidative stress and chemotaxis.

P53R2 has an 80% homology to human ribonucleotide reductase small subunit (RNR R2), a cytoplasmic protein involved in catalysis of ribonucleoside diphosphate to deoxyribonucleotide (dNTP) conversion in DNA synthesis (Guittet et al., 2001). RNR R2 is S phase specific and together with a 90 KDa subunit, R1, forms a ribonucleotide reductase complex (RNR) to provide deoxynucleotide precursors for DNA synthesis in the proliferating cell. R2 expression is cell-cycle specific and its synthesis and degradation control the activity of RNR (Yamaguchi et al., 2001). R2 is not involved in the supply of dNTPs for DNA repair in resting cells and cells in G1, instead p53 induced *p53R2* expression due to various genotoxic stresses (including γ-irradiation, UV-irradiation, the cytotoxic antibiotics, actinomycin-D and adriamycin) leads to the *p53R2* protein accumulating in the nucleus (Nakano et al., 2000, Yamaguchi et al., 2001, Yoshida et al., 2006), here the *p53R2* protein forms an active tetrameric RNR complex with R1 supplying dNTP precursors for DNA repair (Guittet et al., 2001, Xue et al., 2003, Nakamura, 2004).

It has been shown that p52R2 transcription and binding to R1 are impaired in p53 mutant cell lines, however where the p53 – *p53R2* DNA repair mechanism is impaired R2 is able to take over the role of *p53R2* and supply the necessary dNTPs (Zhou et al., 2003, Lin et al., 2004).

The role of *p53R2* in DNA repair has been studied *in vivo* demonstrating the pivotal role of the gene in maintaining dNTP levels in resting cells (Kimura et al., 2003).

1.3.7   *P53R2* Single Nucleotide Polymorphisms (Table 1.15).

A C to A transversion at nucleotide position 88 in the 5' untranslated region (UTR) has been identified by a PCR-SSCP (polymerase chain reaction – single-strand conformation polymorphism) detection method in 52 individuals (Smeds et al., 2001). The SNP was found to have a population frequency of 18%. The effect of the SNP on *p53R2* function remains obscure.

Three SNPs in the *p53R2* gene have been identified by expressed sequence tag database analysis and confirmed in a Caucasian population by allele specific polymerase chain reaction and polymerase chain reaction/restriction fragment length polymorphism methods (Ye and Parry, 2002a). The polymorphisms are represented by an A to C transversion at nucleotide position 2752, an A to G transition at nucleotide position 2759 and a G to C transversion at nucleotide position 4696. The rare alleles at these positions were found at frequencies of 6%, 6% and 15% respectively. All three polymorphisms fall with in the 3' UTR and their effects on *p53R2* expression and function is not known.

Three SNPs have been identified in the first intron of *p53R2* by cold SSCP analysis of 210 individuals (Deng et al., 2005); a C>G transversion at nucleotide position 789 occurring at a frequency of 8%, a G>A transition at nucleotide position 928 occurring at a frequency of 10% and a T>C transition at nucleotide position 933 occurring at a frequency of 1%. The frequencies of these three SNPs were not found to be significantly different in 163 sporadic colorectal adenomas. The functional significance of the SNPs is unknown.

SNP	Location	Rare Allele frequency	Reference
88 C>A	5' UTR	18%	(Smeds et al., 2001)
2752 A>C	3' UTR	6%	(Ye and Parry, 2002a)
2759 A>G	3' UTR	6%	(Ye and Parry, 2002a)
G>C 4696	3' UTR	15%	(Ye and Parry, 2002a)
789 C>G	Intron 1	8%	(Deng et al., 2005)
928 A>G	Intron 1	10%	(Deng et al., 2005)
933 T>C	Intron 1	1%	(Deng et al., 2005)

Table 1.15: *p53R2* SNPs

1.3.8 Role in carcinogenesis.

It is well understood that DNA damage is pivotal in carcinogenesis and the accumulation of genetic defects leads to loss of cell regulation and uncontrolled proliferation (Gibbs, 2003). Therefore cells have a complex mechanism for DNA damage surveillance and repair. The *TP53* tumour suppressor gene is central to this mechanism and acts in response to DNA damage from a wide variety of cellular stresses by activating various pathways that can lead to cell cycle arrest, DNA repair and apoptosis (Nakamura, 2004). *TP53* mutations are the most common genetic defects found in human malignancies. *P53R2* is one of the many genes induced by p53 and is involved in DNA repair. Should the p53 – *p53R2* DNA repair mechanism fail then it can be imagined that genetic mutations may arise. *In vitro* studies of the effects of organic acids on ulcerative colitis – cancer derived cell lines have demonstrated that disruption of the p53 – *p53R2* DNA repair mechanism is likely to play an important role in carcinogenesis (Yoshida et al., 2006). *In vivo* studies of *Rrm2b*-null mice (*Rrm2b* encodes *p53R2*) have suggested that

impairment of *p53R2* results in the accumulation of genetic alterations and increase in the rate of spontaneous mutations (Kimura et al., 2003).

Expression and mutation analyses of *p53R2* in gastric cancer, squamous cell carcinoma of the oesophagus and urinary tract transitional cell carcinoma did not find any association between the gene and diseases (Byun et al., 2002, Smeds et al., 2002, Hayashi et al., 2004). A mutation of *p53R2* has been discovered in a single urinary tract transitional cell carcinoma (Hayashi et al., 2004) leading to a Glu136 → Asp amino acid substitution, but its significance is unclear.

P53R2 expression has been found to be increased in oral squamous cell carcinoma and to be significantly associated with tumour stage and grade (Yanamoto et al., 2003). This finding prompted the authors to speculate that *p53R2* expression may be associated with oral carcinogenesis and may be a predictive factor in tumour development and chemoradiotherapy response.

A case-control study comparing the frequencies of three *p53R2* SNPs in 163 individuals with colorectal adenomas and 210 controls found no significant differences between the two populations (Deng et al., 2005).

1.3.9 Role in cancer chemoradiotherapy.

Differences in the responses of cancers to chemoradiotherapy (CRT) may, in part, be due to the genetic makeup of the tumour cells. Lack of *p53R2* expression has been found to be associated with a favourable response of oesophageal squamous cell carcinoma to CRT (Okumura et al., 2005). Such knowledge could lead to the ability to better select candidates for CRT based on tumour expression analysis. Similarly enhanced sensitivity to 5-fluoruracil (5-FU) has been demonstrated in *p53R2* silenced cancer cell lines (Yanamoto et al., 2005). 5-FU is a chemotherapeutic agent used to treat a wide range of cancers. That silencing of

p53R2 leads to increased sensitivity to 5-FU supports targeting *p53R2* for gene therapy in order to enhance adjuvant therapy of cancer.

The role of iron in the cell cycle and in neoplastic proliferation has lead to the possibility of using iron chelators as anti-tumour agents (Le and Richardson, 2002). Doxorubicin has been shown to be a more effective chemotherapy when combined with the iron chelator deferoxamine (Blatt and Huntley, 1989). Both R2 and *p53R2* proteins have been shown to bind Fe through highly conserved residues (Le and Richardson, 2002, Kauppi et al., 1996). The enzymatic activity of the RR molecule is related to the molecule's iron centre and iron chelation has been shown to impair the activity of RR, with *p53R2* being more susceptible to iron chelators than R2 (Le and Richardson, 2002, Blatt and Huntley, 1989, Shao et al., 2004). This knowledge could lead to the development of new chemotherapeutic agents based on iron chelation.

1.3.10 Summary.

P53R2 is a p53 inducible DNA repair gene that has a crucial role in supplying dNTPs for DNA repair. Several SNPs have been discovered but their functional effects remain obscure. There is little evidence that *p53R2* has a significant role in carcinogenesis, however further studies are required to fully elucidate *p53R2* function and role in carcinogenesis.

1.4   *REPRIMO.*

1.4.1   Symbol.

RPRM.

1.4.2   Genebank accession number.

AB043585

1.4.3   OMIM accession number.

-

1.4.4   Locus and size.

Reprimo is a 1510 base gene located at chromosome 2q23.3. The coding sequence is located between bases 244 to 573.

1.4.5   Function.

Reprimo acts down stream of the tumour suppressor gene p53 and is involved in cell cycle arrest at the G2 stage of the cell cycle in response to X-irradiation (Suzuki et al., 2006, Wong et al., 2005). *Reprimo* expression produces a glycosylated cytoplasmic protein and G2/M arrest of the cell cycle by inhibition of Cdc2 activity and nuclear translocation of cyclin B1. Arrest of damaged cells at G2 allows for DNA repair or apoptosis and thus protection from accumulation of mutations and

tumorigenesis. *Reprimo* dysfunction may lead to loss of cell cycle regulation and therefore contribute to tumorigenesis.

1.4.6 *Reprimo* Single Nucleotide Polymorphisms.

Two SNPs of the *Reprimo* gene have been identified using the human expressed sequence tags (ESTs) database (Ye and Parry, 2002b). The polymorphisms, a G to C transversion at nucleotide position 824 and a C to G transversion at position 839, were confirmed by allele specific polymerase chain reaction in a healthy Caucasian population. The allele frequencies of G and C alleles at position 824 were found to be 38.4% and 61.6% respectively and the frequencies of G and C alleles at position 839 to be 3.7% and 96.3%. Both polymorphisms fall within the 3' untranslated region of the gene.

There are no published *Reprimo* SNP – cancer association studies to date.

1.4.7 Role in carcinogenesis.

Allelic imbalance of the *Reprimo* chromosomal locus, 2q23, is associated with human cancers, including lung, colon and breast, suggesting a tumour suppressor role (Zhang et al., 2002).

Lowered expression of *Reprimo* has been observed in mice with p53 germ line mutations exposed to the colon carcinogen 1,2-Dimthylhydrazine (DMH) (Takahashi et al., 2005). Exposure of these mice to DMH led to an increased incidence of uterine sarcomas, colon carcinomas, lung adenomas and hepatomas. Lowered *Reprimo* expression was presumed to lead to increased cell proliferation.

Reprimo has been found to be aberrantly methylated in a number of cancers (Suzuki et al., 2005, Komazawa et al., 2004, Sato et al., 2003, Zhang et al., 2002, Taylor and

Stark, 2001, Ohki et al., 2000). Aberrant methylation of proto-oncogenes and tumour suppressor genes is an epigenetic event frequently observed in carcinogenesis. The effects of promoter region methylation on DNA binding and altered chromatin structure down regulate transcriptional activity leading to loss of gene function. Aberrant methylation of *Reprimo* has also been shown to be independent of p53 status supporting the hypothesis that *Reprimo* has a tumour suppressor gene function.

Reprimo has been shown to be up regulated by S100A2, one of the S100 protein family thought to have a role in cancer progression (Matsubara et al., 2005).

1.4.8 Summary.

Reprimo is a p53 dependent gene involved in cell cycle regulation. Its chromosomal locus is recognised as being associated with human cancers. It is a putative tumour suppressor gene whose dysfunction due to altered regulation or aberrant methylation may be involved in carcinogenesis. Although SNPs have been discovered their effect on gene function or association with disease has not been studied.

1.5   DIVERTICULAR DISEASE.

1.5.1   Introduction.

Diverticular disease is an acquired condition of the colon caused by mucosal and sub mucosal herniations through the circular muscle layer (Mimura et al., 2002, Jun and Stollman, 2002). It is very common and represents an important disease in terms of health care costs in the Western world (Petruzziello et al., 2006). The disease covers a clinical spectrum from asymptomatic disease to uncomplicated symptomatic disease manifest by chronic abdominal pain and to complicated disease manifest by acute inflammation (diverticulitis), abscess formation, fistulation or stricture formation and bowel obstruction, and haemorrhage (Stollman and Raskin, 2004). At the severe end of this clinical spectrum diverticular disease is potentially fatal.

1.5.2   Prevalence.

The true prevalence of diverticular disease is difficult to establish as in most cases it is asymptomatic. However prevalence increases with age with rates of 5-10% under the age of 40, 30% over the age of 50, 50% over the age of 70 and 66% over the age of 80 in the Western world being generally accepted (Stollman and Raskin, 2004, Delvaux, 2003, Jun and Stollman, 2002). The prevalence of diverticular disease appears to be increasing, this may be due to an increasingly elderly population, increasing bowel investigation due or due to a true prevalence rise (Delvaux, 2003, Jun and Stollman, 2002).

Bleeding and infective complications occur in around 5-15% and 15-20% of cases respectively (Jun and Stollman, 2002, Delvaux, 2003).

Diverticular disease occurs with equal frequency in both men and women, however more men under the age of 65 and more women over the age of 65 are admitted to hospital with diverticulitis (Jun and Stollman, 2002).

1.5.3   Geography.

Diverticular disease has been termed a 'disease of Western civilisation' due to the wide variation in geographical incidence. The prevalence of diverticular disease is highest in the industrialised nations (USA, Europe and Australia) and lowest in developing nations (rural Africa and Asia) (Kang et al., 2004, Stollman and Raskin, 2004, Nakaji et al., 2002). Ethnic differences also exist within countries, for example a lower prevalence of diverticular disease in Indian-subcontinent Asians as compared with other ethnic groups living in Britain has been shown from a study of colonoscopy reports (Kang et al., 2004). The prevalence of diverticular disease within ethnic groups has been shown to rise as they are Westernized, demonstrated by the increase in prevalence from 3.8% to 12% and 0.7% to 5.4% in Shepardic Jews and Arabs, respectively, in Israel over ten years (Jun and Stollman, 2002). These findings suggest both environmental and genetic effects on the aetiology of diverticular disease.

The site of diverticula formation in the colon also has geographic variation with the left side of the colon being affected in Western populations and the right side in Asian and Oriental populations (Jun and Stollman, 2002, Nakaji et al., 2002, Simpson et al., 2002, Kang et al., 2004). Although the prevalence of diverticular disease may increase with Westernisation the anatomical site preponderance persists (Jun and Stollman, 2002) again suggesting a genetic component to the disease.

1.5.4 Aetiology.

Diverticular disease is an acquired disease of the colon and is thought to arise as a result of a prolonged low fibre diet interacting with changes in the bowel wall to cause increased pressure within the lumen of the bowel (Mimura et al., 2002). This increased pressure causes herniations of the mucosal through the muscle wall at the points where blood vessels penetrate.

The role of fibre is well established in the development of diverticular disease. Vegetarians have been shown to have a lower incidence of diverticular disease than their carnivorous counterparts (Gear et al., 1979). It is thought that a diet low in fibre leads to a lower amount of stool within the colon (fibre is not absorbed and therefore serves to provide bulk to the stool). With fewer residues within the bowel as the muscles contract a higher intraluminal pressure is generated (Mimura et al., 2002, Stollman and Raskin, 2004). This effect is compounded by stiffening of the bowel wall due to changes in the extracellular matrix. Electron microscopic studies of colonic walls have demonstrated elastosis (Whiteway and Morson, 1985) and structural and compositional changes in collagen have been demonstrated (Wess et al., 1995, Wess et al., 1996). However the relationship between the bowel wall changes and pathogenesis of diverticular disease remains unclear. One theory put forward involves altered activity of metalloproteinases, enzymes involved in the degradation of extracellular matrix components including elastin and collagen, leading to altered colonic wall structure (Simpson et al., 2003).

1.5.5 Genetic aspects of Diverticular Disease.

Diverticular disease is not generally considered to be a genetic disease, however there has been little research into its molecular pathology or genetic epidemiology,

and is generally held to be an environmental disease. Certain aspects of diverticular disease suggest at least a contribution from molecular mechanisms to its pathogenesis.

The ethnic differences in both the frequency of diverticular disease and the predilection for affecting the right colon in Eastern races may reflect a genetic influence on the aetiology of the disease (Kang et al., 2004, Simpson et al., 2002, Mimura et al., 2002, Nakaji et al., 2002). Although it has been suggested the difference in site predilection may be due to morphological differences in the colon between races (Nakaji et al., 2002).

There are also associations of diverticular disease with genetic connective tissue disorders. Marfan's, William's and Ehler's-Danlos syndromes are diseases manifest by disorders of connective tissue. Marfan's syndrome is due to a mutation in the fibrillin-1 gene, located on chromosome 15q21.1, and results in a number of abnormalities of collagen and elastin (Gray and Davies, 1996). The Ehler's-Danlos syndromes are a group of inherited connective tissue disorders due to a defect in the type III collagen gene (2q31) (Yeowell and Pinnell, 1993). In type IV Ehler's-Danlos, the most severe type, the bowel is affected. William's syndrome is, again, a genetic connective tissue disease due to a deletion in the elastin gene, located at 7q11.2. All three of these syndromes have been reported associated with premature diverticular disease (Beighton et al., 1998, Eliashar et al., 1998, Deshpande et al., 2005). This would suggest that genetic factors can be responsible for the development of diverticular disease and leads to the supposition that acquired defects in genes involved in the extra-cellular matrix may be important in its development. Changes in type III collagen and reduced matrix metalloproteinase 1 expression have been demonstrated in diverticular disease further supporting molecular pathological mechanisms in its development (Stumpf et al., 2001).

Further work is clearly needed in this area in order to unravel the complex nature of this common disease.

1.5.6 Diverticular disease and Colorectal Cancer.

Both diverticular disease and colorectal cancer share similar aetiological and symptomological features. Both are diseases of old age, of the developed world and are associated with a diet poor in vegetables. Both diseases present with altered bowel habit and rectal bleeding and can cause bowel obstruction and perforation. However whether there is a causal relationship between the two diseases, and especially whether diverticular disease can increase the risk of developing colorectal cancer remains controversial. It is well established that the inflammatory bowel diseases Crohn's colitis and ulcerative colitis increase the risk of developing colorectal cancer and this stimulates the debate as to whether diverticular disease can cause colorectal cancer. However the evidence base is lacking. Some studies have shown that colorectal cancer is relatively common amongst patients with diverticular disease (Stefansson et al., 1995) and that there is an increased risk of left sided colon cancer (Stefansson et al., 1993). Other studies have shown an increased incidence of adenomas (precursors of cancer) in diverticular disease but that this does not necessarily indicate a causal association (Morini et al., 1988, Morini et al., 2002). Other studies still have shown that there is no significant association between the two diseases and that they represent heterogeneous groups (McCallum et al., 1987, Krones et al., 2006) or indeed that there is a lower incidence of colorectal cancer in patients with diverticular disease (Loffeld and Putten, 2001). The generally accepted view is that being diagnosed with diverticular disease does not increase an individual's risk of developing colorectal cancer,

however any individual who presents with symptoms that suggest diverticular disease must be assumed to have colorectal cancer until proven otherwise.

1.5.7 Diverticular disease as a control disease in genetic studies of colorectal cancer.

People diagnosed with diverticular disease provide a reasonable and practical control group for colorectal cancer genetic studies. Both diseases have similar environmental and geographical aetiologies and affect a similar age group but there is not thought to be a causal relationship between the two diseases. Therefore diverticular disease provides well-matched controls for colorectal cancer studies. Diverticular disease is also a common disease and so potential control subjects are plentiful.

People who have been diagnosed as having diverticular disease will have had investigations that would detect colorectal cancer and so it can be ensured that individuals do not have colorectal cancer when used as control subjects.

1.5.8 Summary.

Diverticular disease is a common disease affecting the colon. It is a disease of advancing age and of the developed world. Lack of dietary fibre is an important aetiological risk factor. Although it is not known to be a genetic disease, ethnic differences in disease prevalence and pattern and its association with genetic connective tissue diseases suggest a likely underlying molecular pathology.

Diverticular disease is not generally accepted to be causally related to colorectal cancer but does share demographic and aetiological factors and therefore provides a control disease for genetic studies of colorectal cancer.

Chapter 2

MATERIALS AND METHODS

2.1   INTRODUCTION

The following chapter gives the details of the methods followed in order to test the *Reprimo* G>C 824 and *p53R2* C>G 4696 SNPs for association with CRC. The first step was to establish the frequencies of the SNPs in the study populations by genotyping each individual in the populations. DNA was extracted from buccal cells and analysed using polymerase chain reaction techniques to determine individual's genotype. The population genotype frequency data was then analysed using an on-line statistical calculator specifically designed for SNP disease association studies.

2.2   METHOD DESIGN

A case-control study was used to test for genetic association between the *Reprimo* G>C 824 and *p53R2* C>G 4696 SNPs.

2.3   ETHICAL CONSIDERATIONS

The ethical principles governing research as laid down by the General Medical Council (Council) were followed in designing and conducting the work presented in this thesis.

The Iechyd Morgannwg Health authority local research ethics committee granted approval for the research (reference: 2002.033).

Each participant was given an information sheet outlining the purpose of the research and informed consent was obtained from individuals within the study populations prior to obtaining biopsy samples (see appendices A and B). Confidentiality was guaranteed by removing identification from extracted DNA so that it was not possible to match a given genotype to an individual.

2.4 POPULATION SELECTION

Three populations were studied:

2.4.1 Study population.

A Colorectal cancer population derived from patients treated by the Department of Colorectal Surgery, Singleton Hospital, Swansea NHS Trust, Wales. Only patients who had had a histologically proven diagnosis of colorectal cancer were included in the study. Patients with coexistent inflammatory bowel disease (ulcerative colitis, Crohn's disease etc.) or autosomal dominantly inherited CRC (FAP or HNPCC) were excluded. Patients with a family history of CRC were included.

2.4.2 Unmatched control population

A random population derived from students and staff of the Centre for Genetics and Toxicology, University of Wales Swansea. All students and staff who volunteered were included in the study. There were no exclusion criteria.

2.4.3 Matched control population.

A diverticular disease population derived from patients treated by the Department of Colorectal Surgery, Singleton Hospital, Swansea NHS Trust, Wales. Only

individuals with diverticular disease proven by radiological, endoscopic or histological means were included. Patients with coexistent inflammatory or malignant colorectal disease were excluded.

2.5   MATCHING CRITERIA.

2.5.1   Ethnic group.

Both the CRC and DD populations were derived from the same geographical area of South West Wales – an area of relative ethnic homogeneity, with 98% of the population being Caucasian (www.statistics.gov.uk). The student population was taken from a more diverse ethnic mix.

2.5.2   Age.

Both CRC and DD are age related disorders becoming increasingly common with advancing age.

2.5.3   Environmental factors.

The matched CRC and DD populations were assumed to have similar environmental backgrounds on the basis of their similar geographical locations, ethnicity and ages. The unmatched student population was more diverse and so may have been exposed to different environmental factors.

2.6 METHODS TO OBTAIN DNA.

2.6.1 Sample Collection.

DNA samples were collected by buccal scraping biopsy using cytology brushes (Cytosoft Cytology Brush, Medical Packaging Corporation, Panorama City, CA 91402, USA). The brush was rubbed over the buccal mucosa of participants, replaced in its packaging, identified with a patient label to allow data retrieval and the population from which the individual came from indicated on the packaging. The sample was then stored at -20°C until the DNA was extracted. Samples could be stored frozen indefinitely.

2.6.2 DNA Isolation Technique.

DNA isolation from collected buccal cell biopsies was performed using the Puregene™ DNA Purification System (Gentra Systems, Minnesota, USA). The system comprises the following five steps:

2.6.2.1 Cell Lysis.

Each biopsy brush was washed in 300µl of cell lysis solution (supplied in kit) and 1.5µl proteinase K solution (20mg/ml) added. After thorough mixing the lysate was incubated at 55°C for 1-3 hours.

2.6.2.2   RNase Treatment.

1.5µl RNase A solution was added to each lysate then thoroughly mixed and incubated at 37°C for 30 minutes.

2.6.2.3   Protein Precipitation.

The samples were cooled to room temperature and 100µl of protein precipitation solution (supplied in kit) was added to each sample and thoroughly mixed by vortexing. The samples were then cooled in an ice bath for five minutes and then centrifuged at 13000rpm for 15 minutes at 4°C to form the precipitated protein into a pellet, visible at the bottom of the tube. If no pellet was visible, the sample was re-vortexed and centrifuged.

2.6.2.4   DNA Precipitation.

The supernatant containing the DNA was separated from the precipitated protein pellet, mixed with 300µl of 100% isopropanol and 0.5µl of glycogen solution (20mg/ml) and incubated at room temperature for five minutes. This solution was then centrifuged at 13000rpm for ten minutes at 4°C. (The DNA was visualized at the bottom of the tube as a small white pellet if the yield was large enough). The supernatant was removed and 300µl of 70% ethanol was added to wash the DNA. This solution was then centrifuged at 13000 rpm for five minutes at 4°C. The ethanol was carefully removed and the tube containing the DNA allowed to dry.

2.6.2.5 DNA Hydration.

The DNA was rehydrated by adding 20μl of purified water and incubating at 65˚C for one hour or overnight at room temperature. The DNA samples were then stored at -20˚C.

2.6.3 DNA Concentration Analysis and Standardisation.

As the yield of DNA from each biopsy varied it was necessary to establish the concentration of each isolated solution so that the concentrations could be standardized (30ng μl^{-1}) for PCR and PCR/RFLP.

Each sample was thawed and vortexed to redisperse the DNA. 4μl of each sample was mixed with 96μl of purified water and the DNA concentration analysed in a spectrophotometer.

Samples with a concentration greater than 30ng μl^{-1} were diluted. The volume to be added (V) was calculated by the formula V = (Y x 16) − 16 where Y = (concentration of DNA/ 30) ng μl^{-1}.

2.7 METHODS TO DETECT *P53R2* C>G 4696 AND *REPRIMO* G>C 824 SINGLE NUCLEOTIDE POLYMORPHISMS.

NB: The p53R2 4696 SNP was described by Ye and Parry as a G>C transversion, with the G allele being rare (frequency 15%), as the G SNP appears in Genebank. However for this study the SNP is referred to as a C>G transversion as this nomenclature more accurately describes the SNP.

2.7.1 *Reprimo* G>C 824 - Allele specific polymerase chain reaction (PCR).

Allele specific PCR was used to detect a G to C transition at nucleotide position 824 in the *Reprimo* gene using one downstream primer (REP-3) and two upstream primers (REP-1 and REP-2), differing in the terminal base. Primer sequences are those used by Ye and Parry and shown below (Table 2.1). Primers were obtained from MWG Biotech (Ebersberg, Germany). Each sample was tested in parallel reactions, with the same downstream primer and one of the upstream primers. The reaction mixture used for the PCR is given below. Amplification only took place if the tube contained an exact matching upstream primer, i.e. in both tubes for heterozygotes or in a single tube for homozygotes. The reactions were performed in a PTC-225 Peltier Thermal Cycler (MJ Research,Waltham, MA, USA). Conditions for the reactions are shown below. Each reaction also included negative controls, where no DNA was included in the reaction mix, in order to detect any contamination and positive controls using standard DNA control(1:10 dilution of Promega Human Genomic DNA, 185µl/ml) of known genotype, in order to check reaction was successful. PCR products were visualized on a 2.5% agarose gel with ethidium bromide staining (2g agarose melted in 100ml TrisBorateEDTA with 10µl ethidium bromide added).

2.7.1.1 Primers used to detect *Reprimo* G>C 824.

Reprimo G>C 824 (allele specific)

REP-1 (Upstream)	5′-AGAGGGCGATTAGGGCGCAG-3′
REP-2 (Upstream)	5′-AGAGGGCGATTAGGGCGCAC-3′
REP-3 (Downstream)	5′-AGGAGAAGAGTGGGAGCGC-3′

Table 2.1: Primers used to identify *Reprimo* G>C 824.

2.7.1.2 Reaction Mix for PCR.

1.5 mM $MgCl_2$

10 mM Tris-HCl Buffer, pH 8.8

100 µM each dNTP

100 pmols Upstream Primer

100 pmols Downstream Primer

2.5 U *Taq* DNA polymerase

2 µl template DNA solution (30ng/ µl).

In a total volume of 50µl.

2.7.1.3 Reaction Conditions.

5-minute denaturation at 94°C.

32 cycles at 94°C for 45 seconds, 62°C for 30 seconds, 72°C for 45 seconds.

5-minute final extension at 72°C.

2.7.2 *p53R2* C>G 4696 – Polymerase chain reaction / restriction fragment length polymorphism (PCR/RFLP).

The polymerase chain reaction/restriction fragment length polymorphism method was used to detect a C to G transition at nucleotide position 4696 of the *p53R2* gene. This transversion creates a recognition site for the *Hind*III restriction enzyme. A fragment containing the polymorphism was isolated and amplified by PCR using primers R2-6 and R2-7 (Table 2.2) using the same reaction mix as shown previously. 10μl PCR product was digested at 37°C in a mixture containing 2μl *Hind*III restriction enzyme, 6μl purified water and 2μl buffer to give a total volume of 20 μl. The digested fragments were visualized on 2.5% agarose gel.

2.7.2.1 Primers used to detect *p53R2* C>G 4696.

P53R2 C>G 4696 (PCR/RFLP)
R2-6   5′-AGATGGAAGGCTGGGAGAAT-3′
R2-7   5′-TTCTGTACATGCAGGCTTGG-3′

Table 2.2.: Primers used to identify *p53R2* C>G 4696.

2.7.2.2 Reaction conditions.

5 minute denaturation at 94°C.

32 cycles of 94°C for 10 seconds, 57°C for 20 seconds and 72°C for 45 seconds.

Final extension at 72°C for 5 minutes.

2.8 VALIDATION OF GENOTYPING TECHNIQUES.

To ensure the genotyping was accurate and to exclude genotyping errors confounding interpretation of results measures were taken to verify the fidelity of the techniques used.

2.8.1 Repeat genotyping.

As the most likely genotyping error in allele specific PCR results is false positive heterozygotes all *Reprimo* G>C 824 heterozygotes were checked by repeating the PCR reaction and gel electrophoresis.

2.8.2 Independent sequencing.

A total of nine randomly selected PCR products of hetero- and homozygotes of both *Reprimo* G>C 824 and *p53R2* C>G 4696 from all three populations were purified using the Qiagen QIAquick® PCR Purification system (Hilden, Germany) and sent to MWG Biotech for sequencing in order to verify fidelity of methods to identify SNPs.

Sequencing was only performed with down stream primers therefore chromograms received from MWG showed complimentary and reversed sequences.

2.9 STATISTICAL ANALYSIS.

2.9.1 Population demographic data.

Gender distribution in the study populations was compared using a χ^2 test. Age distribution in the CRC and DD populations was compared using a t-test for means.

2.9.2 Population genotype and allele frequency analysis.

In order to test the SNPs for association with CRC they were tested for Hardy-Weinberg equilibrium and genotypic and allelic disease association on-line using the Finetti programme (http://ihg.gsf.de/cgi-bin/hw/hwa1.pl) accessed via the Genestat Statistical Genetics web site (http://www.genestat.org). The programme generates expected genotype frequencies, allele frequencies and inbreeding coefficient (F) from observed genotype frequencies. Deviation from Hardy-Weinberg equilibrium is tested by Pearson's goodness-of-fit, Log likelihood ratio and exact chi-square tests (with 1 degree of freedom). Tests for association are performed using chi square and odds ratio for allele-, heterozygote- and homozygote- frequency differences, allele positivity and Armitage's trend test. The tests were carried out twice for each SNP using both control populations. P values of less than 0.05 were significant.

2.9.2.1 Hardy – Weinberg Equilibrium (HWE).

The Hardy – Weinberg law states that allele frequencies remain in equilibrium in large, randomly mating populations and is expressed by the formula $p^2 + 2pq + q^2 = 1$ where p and q are the frequencies of the alleles A and a. From this formula the

allele and expected genotype frequencies can be derived from the observed genotype frequencies according to the following formulae; where the observed frequencies of the genotypes AA, Aa and aa are x, y and z in a population size of n then the frequencies of the alleles A and a are (2x+y)/n and (2z+y)/n respectively. The expected genotype frequencies in the population are thus $(A^2)n$, $(2Aa)n$ and $(a^2)n$.

2.9.2.1.1 Deviation from HWE.

Deviation from HWE is tested using Pearson's goodness-of-fit, log likelihood ratio and exact χ^2 tests to compare the observed and expected allele frequencies, as calculated using the Hardy-Weinberg equation. Three tests are used as the goodness of fit test ($\chi^2 = \Sigma \, (O - E)^2/E$) may be inaccurate when the expected frequencies are low or the sample sizes are small. In these cases, respectively, the log likelihood ratio and exact tests are more appropriate.

2.9.2.2 Inbreeding coefficient.

The inbreeding coefficient is defined as the probability that two alleles in an individual are identical by descent (autozygous) and is a measure of genotypic frequency deviation from panmictic frequencies in terms of heterozygous deficiency or excess. Calculated by the formula $F = (H_{obs}/H_{exp})$ where H_{obs} and H_{exp} are the observed and expected heterozygosity. Negative and positive values of F indicate heterozygous excess (outbreeding) and deficiency (inbreeding) respectively. Although deviation from HWE can be due to causes other than

inbreeding, the inbreeding coefficient gives a good indication of the degree of HWE deviation.

2.9.3 Tests for disease – SNP association.

Association between the study SNPs and CRC were tested using odds-ratio and Pearson's goodness-of-fit χ^2 tests. A range of tests are performed; allele frequency difference, heterozygous, homozygous, allele positivity and Armitage's trend test in order to account for different models of penetrance i.e. dominant, recessive, additive, multiplicative or no specific model. Each genetic model generates a different contingency table to which the χ^2 test is applied. The tests are performed with respect to each SNP allele, as it is not known which allele potentially confers disease risk.

For an SNP with alleles A and B the full genotype contingency table will be;

	AA	AB	BB
Cases	a	b	c
Controls	d	e	f

where a, b, c, d, e and f are observed genotype counts.

The χ^2 statistic tests for differences between the observed and expected genotype numbers in each cell of the table. If the observed (O) AA genotype number in cases is a, the expected (E) number of AA cases is $n_{AA}n_{case}/N$ i.e. (a+b)×(a+b+c)/(a+b+c+d+e+f). The χ^2 statistic is thus $\chi^2 = \Sigma (O - E)^2/E$.

The odds ratio (OR) indicates the increase in disease risk for carriers as compared with non-carriers and is calculated as the ratio of candidate allele carriers to non-carriers in the disease population compared with that in the control population, i.e.:

OR = [(b+c)d]/[a(e+f)] for the above contingency table.

2.9.3.1 Allele frequency difference test.

This test assumes a multiplicative penetrance model where the genotype AB increases risk by r and the genotype BB increases risk by r^2. The contingency table for this model is:

	A	B
Cases	2a+b	b+2c
Controls	2d+e	e+2f

2.9.3.2 Heterozygous and homozygous tests.

These tests assume no specified relationship with the genotypes. The contingency tables generated are:

	AB	AA
Cases	b	a
Controls	e	d

for the heterozygous test with respect to the A allele,

	AA	BB
Cases	a	c
Controls	d	f

for the homozygous test.

2.9.3.3 Allele positivity test.

This test assumes a dominant or recessive penetrance model. The contingency tables generated are:

	BB	AA+AB
Cases	c	a+b
Controls	f	d+e

for a dominant model with respect to A (single copy of A increases risk) or

	AB+BB	AA
Cases	b+c	a
Controls	e+f	d

for a recessive model with respect to A (two copies of A increases risk).

2.9.3.4 Armitage's trend test

Assumes an additive penetrance model where the genotype AB increases risk by r and the genotype BB increases risk by $2r$ with respect to risk allele B. The contingency table tested will be:

	AA	AB	BB
Cases	a	b	c
Controls	d	e	f

2.10 SUMMARY.

Ethical approval was granted for the study. Three populations (healthy, diverticular disease and CRC) were genotyped with respect to the *Reprimo* G>C 824 and *p53R2* C>G 4696 SNPs using allele specific PCR and PCR/RFLP techniques. Repeat genotyping and independent sequencing of PCR products tested the validity of the genotyping technique. The populations were tested for Hardy-Weinberg equilibrium and CRC association with respect of the candidate SNPs and differing modes of penetrance.

Chapter 3

PRELIMINARY WORK AND OPTIMISATION OF

GENOTYPING TECHNIQUES

3.1 INTRODUCTION.

This chapter will describe how the methodology described in the "materials and methods" chapter was developed. It will describe how the research method was designed in order to investigate the hypothesis that *Reprimo* and *p53R2* SNPs are associated with CRC. The original intention was to investigate three previously described *p53R2* SNPs (at nucleotide positions 2752, 2759 and 4696) and two *Reprimo* SNPs (at nucleotide positions 824 and 839) (Ye and Parry, 2002b, Ye and Parry, 2002a) however only two SNPs (*p53R2* C>G 4696 and *Reprimo* G>C 824) were studied for disease association with CRC. This was because either the SNPs were not detected in the study populations using the methods previously described, or the published methods failed or despite optimisation experiments it was not possible to detect the SNPs. The problems encountered and how they were overcome is detailed below.

It s important to emphasise that the work described below took up to around a third of the total time spent in the laboratory during this study.

3.2 POPULATION SELECTION.

Three populations were studied in this work; a young healthy population, a population of patients known to have CRC and a population of patients known to have diverticular disease. The young population had had tissue samples pre-collected prior to the start of the study and the DNA extracted was used to optimise

the conditions and techniques for the study SNP detection. It was also used to establish the frequency of the study SNPs in the general population. However it was realised that in case-control study a better matched population was required in order to eliminate any bias or confounding factors in the results that may be due to age. Therefore a population with diverticular disease was chosen as it is a common disease affecting a similar age group as CRC, but the two diseases have not been shown to be causally related.

3.3   POPULATION SIZE.

The size of the populations used was based on sizes of populations used in similar studies published prior to or at the time of the beginning of the study (2002) (Jong et al., 2002, Houlston and Tomlinson, 2001). Many published SNP disease association studies had used sample sizes of around one hundred individuals or less and it was felt that population sizes of one hundred would be sufficiently powerful to answer the question posed and more importantly that it would be feasible to collect and analyse the DNA within the time limits of the research period.

3.4   PROBLEMS WITH VISUALISING PCR PRODUCTS.

In Ye and Parry's papers all PCR and PCR/RFLP products were visualised on 2.5% agarose gels. It was felt that as the products were all around 250 base pairs (bp) in length, visualisation on acrylamide gels might be more appropriate. Therefore experiments to test this were performed using Promega human genomic DNA (Promega Corporation, Madison, WI 53711, USA), a standardised source of DNA, and samples collected from a population of students and laboratory staff (Control population 1). Allele specific PCR was used to detect an A>C transversion at

nucleotide position 2752 of the *p53R2* gene, using conditions published by Ye and Parry.

These experiments resulted in gels with multiple bands, the strongest band appearing at around 400bp, fainter bands at 600bp and 900bp and, importantly, no bands at 249bp, the length of the expected PCR product using these conditions (Figure 3.1).
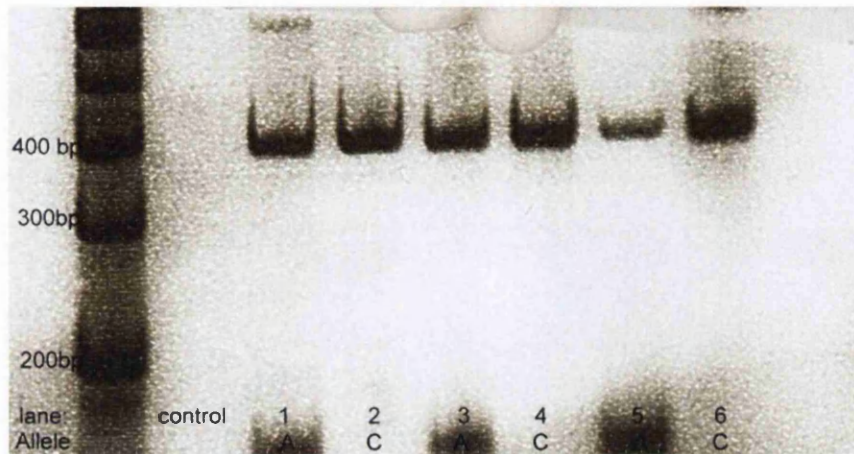
Figure 3.1: Typical acrylamide gel appearance following *p53R2* A>C 2752 allele specific PCR. The control lane contained no DNA in reaction to control for contamination (negative control) (bp = base pairs).

Therefore it was decided to use 2.5% agarose gels and indeed expected band sizes were visualised.

3.5   PROBLEMS IDENTIFYING *P53R2* SNPS.

3.5.1   SNP 1: A>C transversion at nucleotide position 2752.

Ye and Parry used an allele specific PCR technique to identify the *p53R2* A>C 2752 SNP with the following conditions and primers:

Upstream primer (R2-1)	5'-GGATGAAGAAGTCTTAATACTGAA-3'
Upstream primer (R2-2)	5'-GGATGAAGAAGTCTTAATACTGAC-3'
Downstream primer (R2-3)	5'-AAGATCTCAAGTTTCTGACAGTG-3'

Table 3.1: Primers used to detect *p53R2* A>C 2752

The reaction mix used was a 50μl solution containing 1.5mM $MgCl_2$, 10mM Tris-HCL, pH 8.8, 100μM each dNTP, 10 pmol each primer, 2.5 U *Taq* polymerase and 0.5μg template DNA. Amplification conditions were; a 5-minute initial denaturation at 94° C, followed by 32 cycles of 94° C for 45s, 57° C for 30s, 72° C for 45s with a final extension at 72° C for 5 minutes.

These conditions were followed using Promega DNA and collected samples and the products visualised on 2.5% agarose gel.

Negative controls were included, where no DNA was included in the reaction mix, in order to detect any contamination that might confound results.

However this method invariably detected the rare A allele in all samples, an unexpected result given that Ye and Parry found the frequency of the rare A allele 6% (Figure 3.2).
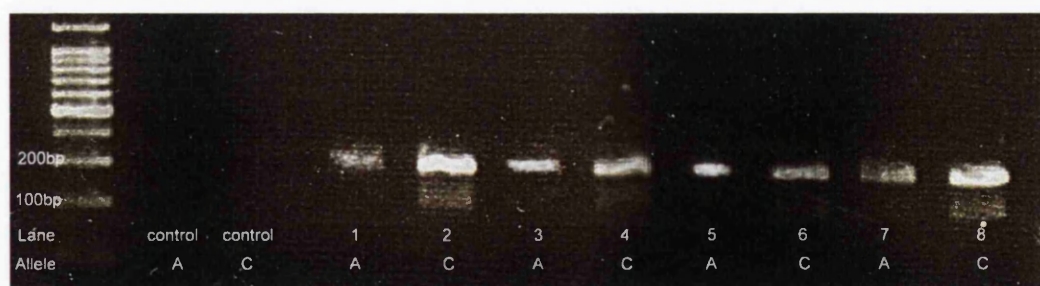
Figure 3.2: Typical agarose gel for *p53R2* A>C 2752 using Ye and Parry conditions. The rare A allele is detected in all samples. The two control lanes were negative controls and here indicate no contamination took place.

Therefore a series of experiments was conducted with various different PCR conditions in order to optimise the conditions. These involved performing the PCR with a re-annealing temperature gradient from 55°C to 65°C (Figure 3.3), varying

the DNA concentration from 1:10 to 1:100 dilutions (Figure 3.4) and varying the number of reaction cycles from 30 to 35.
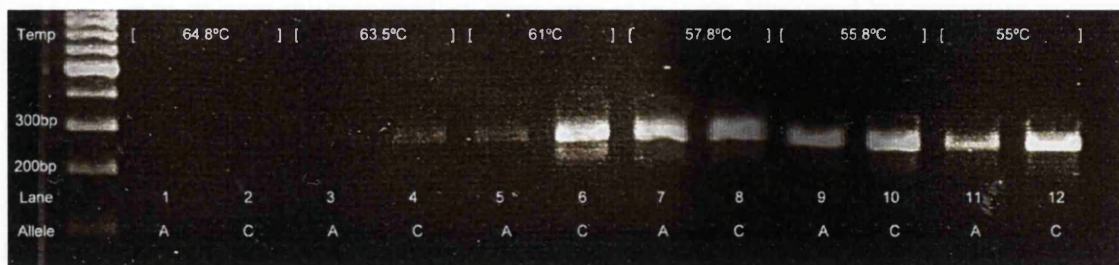
Figure 3.3: Typical agarose gel for *p53R2* A>C 2752 PCR re-annealling temperature gradient experiments. Re-annealling temperature is shown across top of Figure.
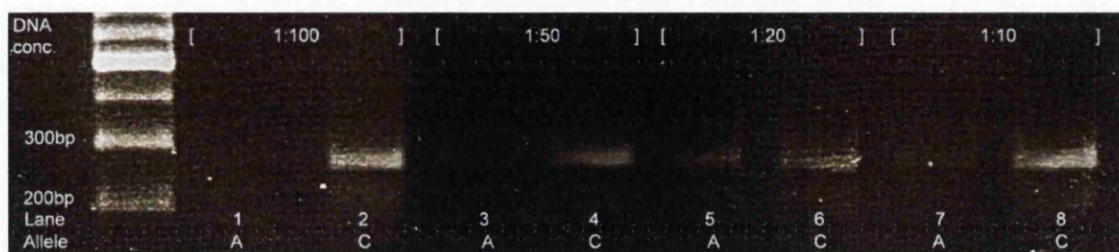
Figure 3.4: Typical agarose gel for *p53R2* A>C 2752 PCR DNA concentration gradient experiments.

However none of these experiments yielded accurate results that would allow confident identification of the alleles. Time constraints forced these experiments to be abandoned to allow the project to progress.

3.5.2   SNP 2: A>G transition at nucleotide position 2759.

A PCR/RFLP technique was used by Ye and Parry to identify the *p53R2* A>G 2759 polymorphism. The A>G transition eliminates a *Tsp*509I restriction enzyme recognition site. The reaction mix was the same as for the A>C 2752 SNP. The primers used to detect the *p53R2* A>G 2759 SNP are shown in table 3.2. Amplification conditions were: a 5-minute initial denaturation at 94° C, followed by 32 cycles of 94° C for 25s, 57° C for 30s, 72° C for 45s with a final extension at 72°

C for 5 minutes. The PCR product was then digested overnight with the *Tsp*509I restriction enzyme at 65° C. The digested products were visualised on 2.5% agarose gel.

| R2-4 | 5'-GATAGCCAATGGTGAGGGAA-3' |
| R2-5 | 5'-ACAGTTTGTTGGGAATGGGA-3' |

Table 3.2: Primers used to detect *p53R2* A>G 2759.

These conditions were used to successfully genotype the whole of control population 1. However the G allele was not identified in any sample (Figure 3.5). Therefore it was not possible to confirm the existence of the A>G transition at nucleotide position 2759 and this SNP was not studied further in either the CRC or diverticular disease populations.
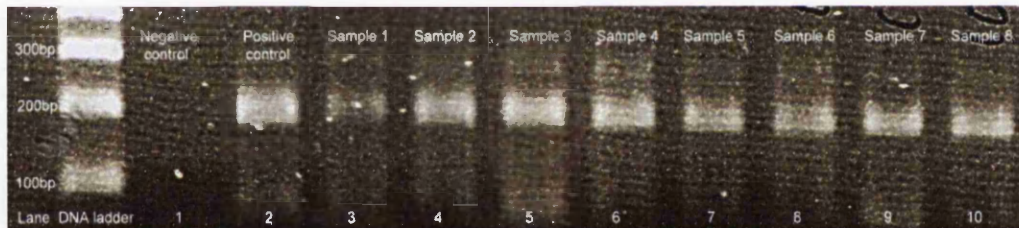
Figure 3.5: Typical agarose gel for *p53R2* A>G 2759 PCR/RFLP. Each lane represents a single DNA sample. The single bands indicate that the PCR products have not cut with the restriction enzyme indicating that the G allele was not present in any sample. Samples are arbitrarily numbered for sake of figure.

3.5.3 SNP 3: C>G transversion at nucleotide position 4696.

A PCR/RFLP technique was used by Ye and Parry to identify the *p53R2* C>G 4696 polymorphism. The G>C transition eliminates a *Hind*III restriction enzyme recognition site. The reaction mix was the same as for the A>C 2752 SNP. Amplification conditions were: a 5-minute initial denaturation at 94° C, followed by 32 cycles of 94° C for 10s, 57° C for 20s, 72° C for 45s with a final extension at 72° C for 5 minutes. The PCR product was then digested overnight with the *Hind*III

restriction enzyme at 37° C. The digested products were visualised on 2.5% agarose gel.

R2-6	5'- AGATGGAAGGCTGGGAGAAT-3'
R2-7	5'- TTCTGTACATGCAGGCTTGG-3'

Table 3.3: Primers used to detect *p53R2* C>G 4696.

This technique successfully identified the polymorphism and was used in the study presented in this thesis (see figure 4.2, page 93 for example of electrophoresis gel).

3.6   PROBLEMS IDENTIFYING *REPRIMO* SNPS.

3.6.1   SNP 1: G>C transversion at nucleotide position 824.

Ye and Parry used an allele specific PCR technique to identify the *Reprimo* C>G 824 SNP with the following conditions and primers:

Upstream primer (REP-1)	5'-AGAGGGCGATTAGGGCGCAG-3'
Upstream primer (REP-2)	5'-AGAGGGCGATTAGGGCGCAC-3'
Downstream primer (REP-3)	5'-AGGAGAAGAGTGGGAGCGC-3'

Table 3.5: Primers used to detect *Reprimo* C>G 824

The reaction mix used was the same as for the *p53R2* SNPs. Amplification conditions were; a 5-minute initial denaturation at 94° C, followed by 32 cycles of 94° C for 45s, 60° C for 30s, 72° C for 45s with a final extension at 72° C for 5 minutes. The products were visualised on 2.5% agarose gel.

These conditions were followed using Promega DNA and samples from the young, student population. However, although these methods yielded products of appropriate size (249 bp) multiple fainter bands were also visualised (Figure 3.6).
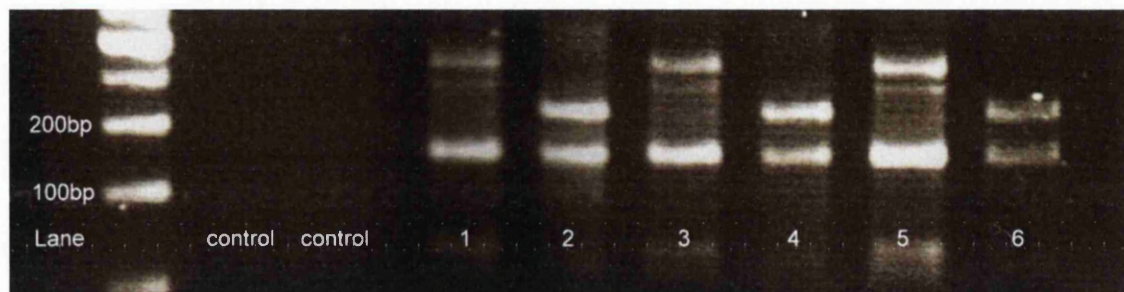
Figure 3.6: Typical agarose gel for *Reprimo* G>C 824 allele specific PCR using Ye and Parry's conditions.

Therefore performing the process with a re-annealing temperature gradient optimised the re-annealing temperature. This gave an optimum re-annealing temperature of 62°C (compared with 60°C) which reliably produced correctly sized products (see figure 4.1, page 92 for example of electrophoresis gel). The altered conditions were adopted for the subsequent work.

3.6.2 SNP 2: C>G transversion at nucleotide position 839.

Ye and Parry used an allele specific PCR technique to identify the *Reprimo* C>G 839 SNP with the following conditions and primers:

Upstream primer (REP-6)	5'-GCGCAGAACTTTGGAAGCTGC-3'
Upstream primer (REP-7)	5'-GCGCAGAACTTTGGAAGCTGG-3'
Downstream primer (REP-3)	5'-AGGAGAAGAGTGGGAGCGC-3'

Table 3.6: Primers used to detect *Reprimo* C>G 839

The reaction mix and amplification conditions used were the same as for the *p53R2* SNPs. These conditions were followed using Promega DNA and collected samples and the products visualised on 2.5% agarose gel. This method did not reliably yield the expected products. Optimisation experiments with re-annealing temperature gradients did not produce reliable conditions (Figures 3.7 and 3.8).
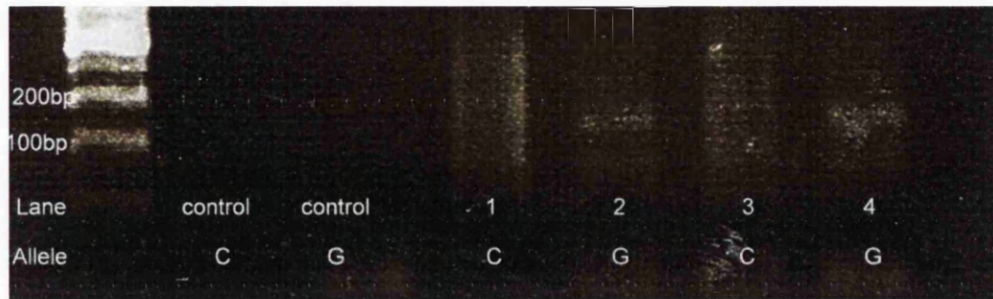
Figure 3.7: *Reprimo* C>G 839 allele specific PCR re-annealling at 58°C. The control lanes are both negative controls. No distinct bands are seen indicating failure of the PCR to detect the SNP.
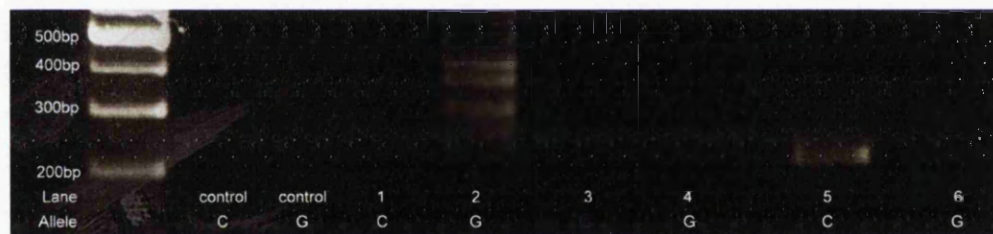
Figure 3.8: *Reprimo* C>G 839 allele specific PCR re-annealling at 59.5°C. Control lanes are negative controls. Lack of distinct bands indicates PCR failure.

Therefore, as the G polymorphism creates a recognition site for the *Rsa*I restriction enzyme, a PCR/RFLP technique was used to detect this polymorphism. The PCR products generated using the primers to identify the *Reprimo* C>G 824 SNP contained the C>G 839 polymorphism site, therefore these primers were used in the PCR/RFLP technique to identify the C>G 839 polymorphism. The PCR products were then digested overnight at 37°C with the *Rsa*I restriction enzyme. However the rare G allele was not identified (Figure 3.9). Due to time constraints, further work to detect the *Reprimo* 839 C>G SNP was not carried out.
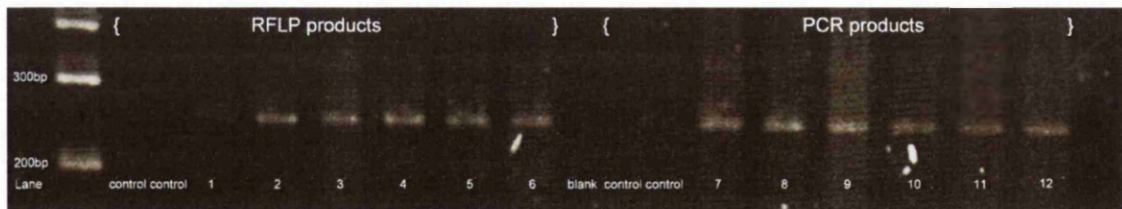
Figure 3.9: Agarose gel demonstrating *Reprimo* C>G 839 PCR/RFLP experiment. The RFLP product lanes show that the PCR products did not cut with the *Rsa* I restriction enzyme indicating that the G allele was not present.

3.7   STATISTICAL METHODS.

The genotypic and allelic data obtained in this study is categorical. In order to examine the phenotypic-genotypic relationship (i.e. disease-association) χ^2 tests can be used to compare the observed with the expected genotype frequencies. The expected allele frequencies can be determined using the Hardy-Weinberg equation, $p^2+2pq+q^2=1$ (where p and q are the allele frequencies). This was the approach taken early in the research. However as the research progressed literature searches revealed that this approach is too simplistic for SNP disease association studies as it does not take into account sample size, departures from HWE or the mode of penetrance, factors that can influence interpretation of results (Hattersley and McCarthy, 2005b, Hosking et al., 2004b, Gordon et al., 2002b, Lewis, 2002b, Ohashi et al., 2001, Slager and Schaid, 2001a, Wittke-Thompson et al., 2005b). It also became apparent that statistical calculators are available on-line specifically designed for SNP studies that take into account these important factors. Therefore all the statistical tests for this study were performed using the Finetti program (accessed at http://ihg.gsf.de/cgi-bin/hw/hwa1.pl). This program tests for HWE (using exact, goodness of fit and log likelihood χ^2 tests), the inbreeding coefficient and tests for allelic and genotypic disease association. The tests are described in greater detail in Section 2.9 of the Methods chapter.

3.8 SUMMARY.

Three populations were used to avoid age related bias; diverticular disease was used as a suitable, easily obtainable matched control disease. Population sizes were similar to previously reported SNP association studies and feasible within the time period of the research.

Three of the five previously reported SNPs were not detected in preliminary experiments and so only the *p53R2* C>G 4696 and *Reprimo* C>G 824 polymorphisms were studied with respect to their association with CRC by a case-control method. Statistical analysis was performed using an on-line SNP disease-association calculator.

Chapter 4

RESULTS

4.1 INTRODUCTION

The following chapter details the results obtained from the study methods as described in Chapter 2.

Tables are used to show results of statistical analysis where appropriate and results are given to two decimal places. P values less than 0.05 are taken as statistically significant and are shown in red.

4.2 POPULATION DATA

4.2.1 Colorectal cancer population

96 volunteers (32 female, 64 male, gender ratio 1:2).

Mean age 68 years (standard deviation = 9.9 years).

100% Caucasian.

4.2.2 Student population

107 volunteers (69 female, 38 male, gender ratio 1:0.6).

Mean age 19.3 years (standard deviation = 1.9 years).

92% Caucasian, 3% Middle Eastern, 2% Asian, 2% Oriental, 1% South American.

4.2.3 Diverticular disease population

52 volunteers (37 female, 15 male, 1:0.4).

Mean age 67.7 years (standard deviation 11.0 years).

100% Caucasian.

4.3 COMPARISON OF GENDER DISTRIBUTION IN MATCHED CRC AND DD POPULATIONS (χ^2 TEST).

There was a significant difference in gender distribution between the CRC and diverticular disease populations (Table 4.1).

Gender	CRC Population	DD Population	χ^2	P value
Male	62	15		
Female	34	37	17.90	< 0.01

Table 4.1: comparison of gender distribution in matched populations CRC and DD (χ^2 test).

4.4 COMPARISON OF MEAN AGES IN MATCHED CRC AND DD POPULATIONS (t-TEST).

There were no significant differences in the mean ages of the CRC and diverticular disease populations (Table 4.2).

	CRC Population	DD Population	Difference	t value (95% confidence interval)	P value
Population size	96	52			
Mean age (years)	68	67.7	0.3	0.17 (-3.2 to 3.8)	0.87
Standard deviation	9.9	11.0			

Table 4.2: Comparison of mean ages in matched populations CRC and DD (t-test).

4.5 DNA COLLECTION

4.5.1 Acceptability.

The methods for canvassing volunteers and collecting tissue samples for DNA analysis were well accepted by patient volunteers, with only very small numbers refusing to take part in the study. Most volunteers were enthusiastic and often extremely interested in the study. No adverse events were encountered in performing buccal scrape biopsies.

4.5.2 Efficiency.

The methods for canvassing and collecting tissue samples for DNA analysis were very time consuming and labour intensive resulting in a smaller diverticular disease population than intended. Indeed, tissue sample collection was an important rate limiting step in this study. Ethical approval was only granted some three months

into the study, from when sample collection could begin. This left 21 months to collect samples. However in this time it was only possible to collect 96 CRC and 52 diverticular disease samples, giving a collection rate of seven samples per month.

4.6 DNA ISOLATION.

4.6.1 DNA Yield.

The average DNA concentration obtained from the buccal scrapings was 2.94ng/ml (range 0.29 – 16.51ng/ml). This gave an average volume of DNA solution from each subject of 40.31μl (range 16 – 220μl).

4.7 GENOTYPING

4.7.1 Error Rate.

4% of all samples were not successfully genotyped for both SNPs studied.

4.7.1.1 *Reprimo* G>C 824.

Student population – 6% (6/107) samples were not successfully genotyped.

Diverticular disease population – 2% (1/52) samples were not successfully genotyped.

Colorectal cancer population – 2% (2/96) samples were not successfully genotyped.

4.7.1.2 *P53R2* **C>G 4696.**

Student population – 4% (4/107) samples were not successfully genotyped.

Diverticular disease population – 4% (2/52) samples were not successfully genotyped.

Colorectal cancer population – 3% (3/96) samples were not successfully genotyped.

4.8 ELECTROPHORESIS GELS

4.8.1 *Reprimo* G>C 824 (Allele specific PCR)

Two electrophoresis columns represent each individual's genotype, one for each allele. Amplification only took place in one of the reactions if the individual was a homozygote or in both reactions if the individual was a heterozygote, visualised as a single or double band pattern on electrophoresis as shown below (Figure 4.1).
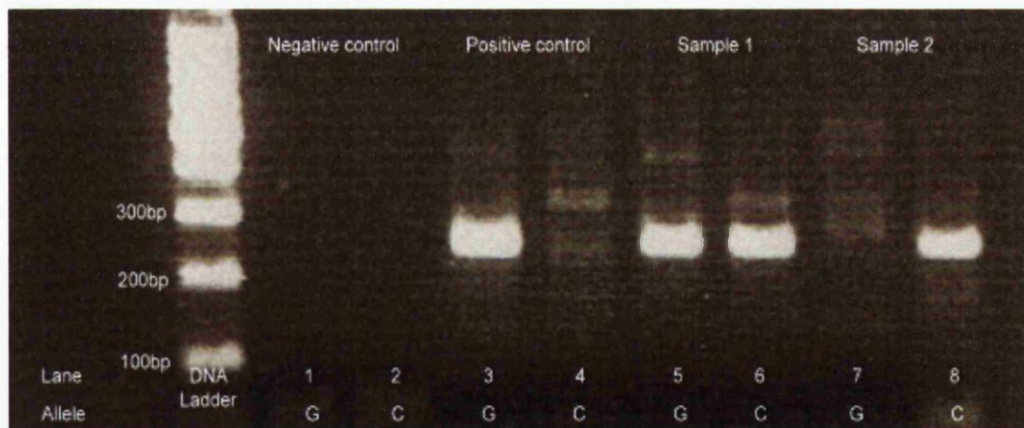
Figure 4.1 Electrophoresis gel showing band patterns for *Reprimo* G>C 824 SNP genotypes. Each sample requires two reactions to detect each allele and is therefore represented in two lanes, a band being visualised at 249bp if the allele is present. The positive control (standard human DNA) genotype is GG, sample 1 GC and sample 2 CC. Samples are arbitrarily numbered for sake of figure.

4.8.2   *P53R2* C>G 4696(PCR/RFLP)

Each individual's genotype is represented in a single column. GG homozygotes have no *Hind III* digestion site and so PCR products do not cut yielding a single 228 base pair band. Each allele of CC homozygotes contains the digestion site and so all PCR products are cut, yielding two bands of 149 and 78 base pairs. Heterozygous individuals show a mixture of these patterns, yielding three bands of 228, 149 and 78 base pairs (Figure 4.2).
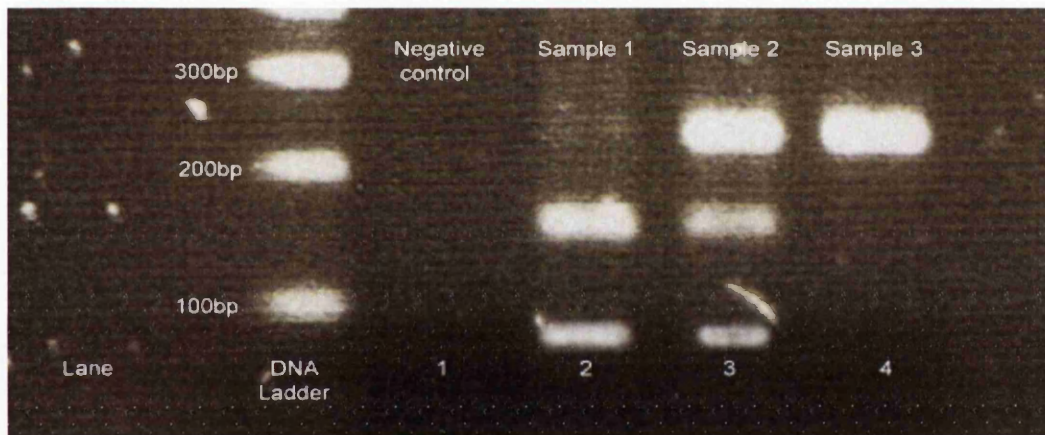
Figure 4.2 Electrophoresis gel showing band pattern for *P53R2* C>G 4696SNP genotypes. Each sample is represented in a single lane. The G allele is represented by a single 228bp band and, as the C allele creates the restriction site, presence of the C allele is shown by two bands at 149bp and 78bp. Positive control was omitted from this reaction. Samples are arbitrarily numbered for sake of figure.

4.8.3   Multiple sample genotyping gel for *Reprimo* G>C 824 (Figure 4.3).

Figure 4.3: Typical electrophoresis gel obtained from allele specific PCR *Reprimo* G>C 824 SNP genotyping of multiple samples. Lanes 1 and 2 are contamination (negative) control lanes. Lanes 3 and 4 are standard DNA (positive) control lanes, showing the standard DNA to be a GG homozygote. Sample 1 is a GC heterozygote and samples 2, 3 and 4 are CC homozygotes. Samples are arbitrarily numbered for sake of figure.

4.8.4   Multiple sample genotyping gel for *p53R2* C>G 4696 (Figure 4.4).
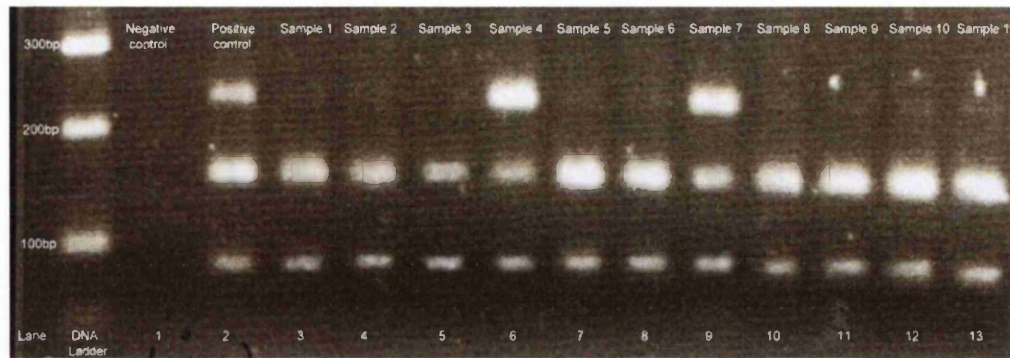
Figure 4.4: Typical electrophoresis gel obtained from PCR/RFLP *P53R2* C>G 4696SNP genotyping of multiple samples. Triple band patterns shown in samples 4 and 7 and positive control (standard Promega human genomic DNA) represent GC heterozygotes. Double band pattern represents CC homozygotes. The GG homozygote is not shown, but would be represented by a single band. Samples are arbitrarily numbered for sake of figure.

4.9   VERIFICATION OF GENOTYPING FIDELITY.

The results of genotyping initially showed an excess of *Reprimo* G>C 824 heterozygotes in the diverticular disease and CRC populations, raising the possibility of genotyping errors. Therefore all diverticular disease and CRC samples were re-genotyped for *Reprimo* G>C 824. 7/145 (5%) genotyping errors were identified amongst the heterozygotes and one genotyping error amongst the homozygotes.

In order to verify that the genotyping techniques were identifying the correct allele by isolating the correct nucleotide sequence, three representative homozygous and heterozygous samples for both study SNP and from each of the three populations were sent for sequencing for both study SNPs at an independent laboratory. All samples sent for sequencing had genotypes as expected from the PCR or PCR/RFLP genotyping. Examples of chromograms demonstrating the nucleotide sequences containing the alleles under study are shown below (Figures 4.5 – 4.10).

4.9.1   *Reprimo* G>C 824 chromograms.

SNPs are indicated by arrows.

⇓

T T A G G G C G C A G A A C T T T G G A
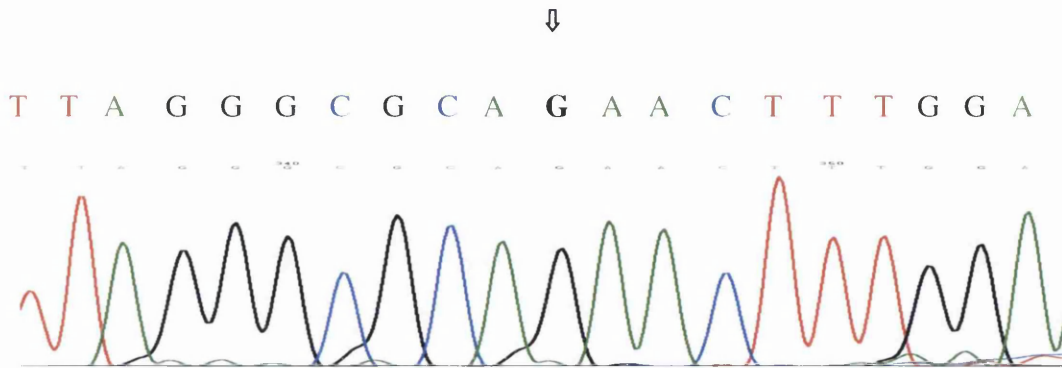
Figure 4.5: Chromogram demonstrating *Reprimo* G>C 824 G homozygote (nucleotide 824 arrowed).

⇓

T T A G G G C G C A G A A C T T T G G A

C
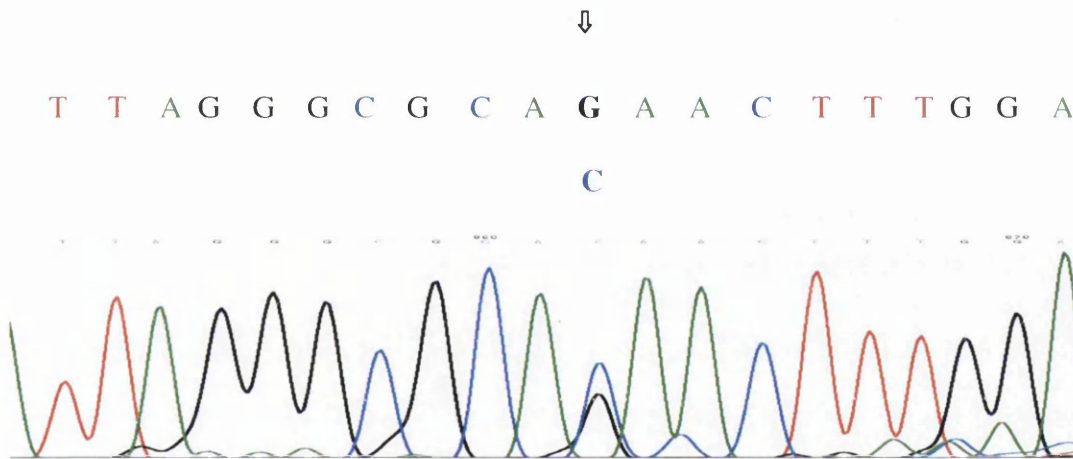
Figure 4.6: Chromogram demonstrating *Reprimo* G>C 824 GC heterozygote (nucleotide 824 arrowed).

⇓

T T A G G G C G C A C A A C T T T G G A
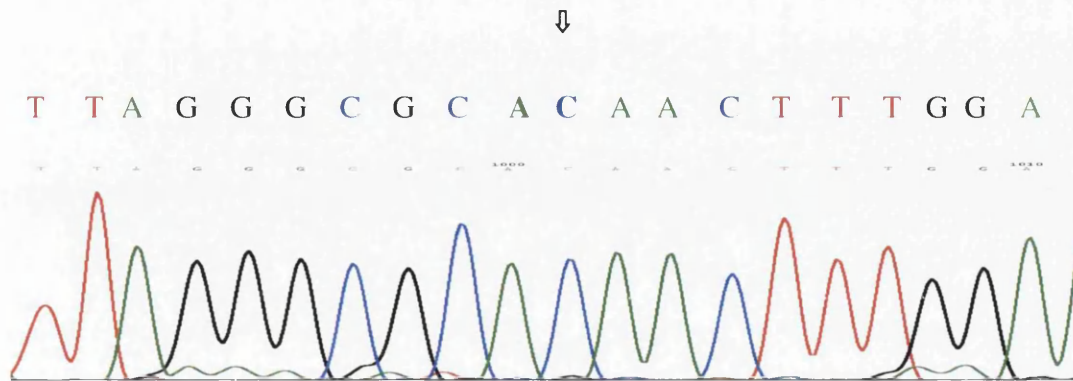
Figure 4.7: Chromogram demonstrating *Reprimo* G>C 824 CC homozygote (nucleotide 824 arrowed).

4.9.2   *P53R2* C>G 4696 chromograms.

SNPs are indicated by arrows.
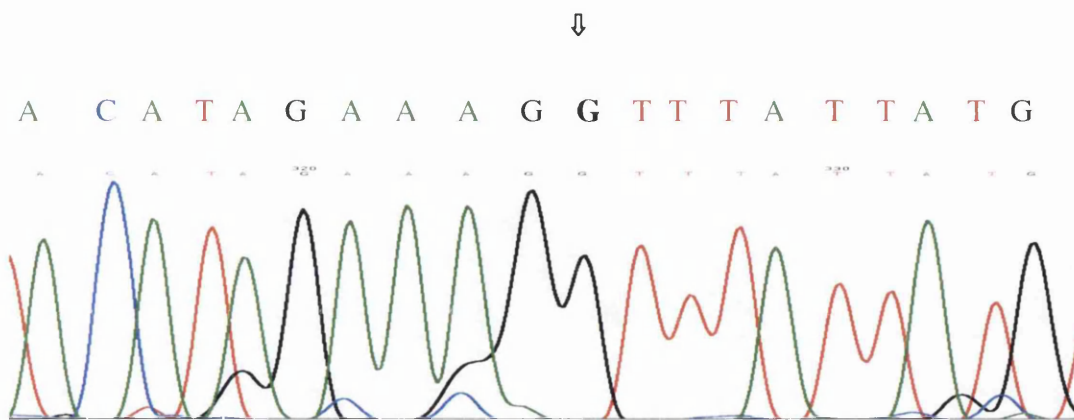
Figure 4.8: Chromogram demonstrating p52R2 C>G 4696 GG homozygote (nucleotide 4696 arrowed).
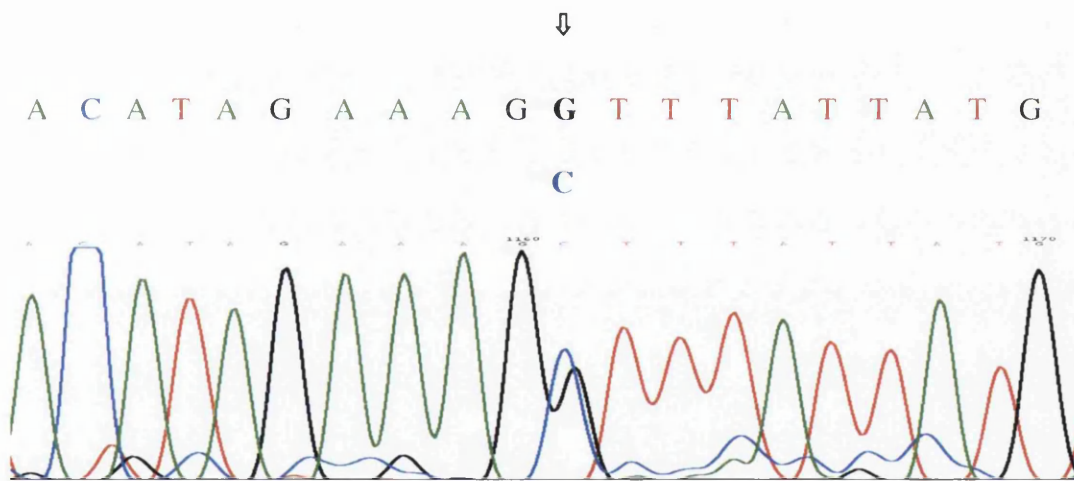
Figure 4.9: Chromogram demonstrating p52R2 C>G 4696 GC heterozygote (nucleotide 4696 arrowed).
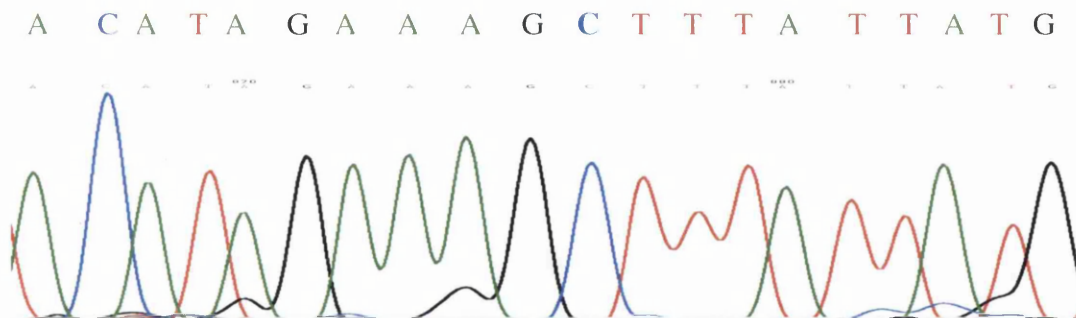
Figure 4.10: Chromogram demonstrating p52R2 C>G 4696 CC homozygote (nucleotide 4696 arrowed).

4.9.3   GENOTYPE FREQUENCIES.

4.9.3.1   *Reprimo* G>C 824.

The observed and expected *Reprimo* G>C 824 genotype frequencies are shown in Table 4.3.

Genotype	STUDENT POPULATION		DIVERTICULAR DISEASE POPULATION		COLORECTAL CANCER POPULATION	
	Observed	Expected	Observed	Expected	Observed	Expected
GG	28 (28%)	27	10 (19%)	14	13 (14%)	17
GC	48 (47%)	50	34 (67%)	25	55 (58%)	46
CC	25 (25%)	24	7 (14%)	11	26 (28%)	30

Table 4.3: Genotype frequencies, *Reprimo* G>C 824.

4.9.3.2   *P53R2* C>G 4696.

The observed and expected *p53R2* C>G 4696 genotype frequencies are shown in Table 4.4.

Genotype	STUDENT POPULATION		DIVERTICULAR DISEASE POPULATION		COLORECTAL CANCER POPULATION	
	Observed	Expected	Observed	Expected	Observed	Expected
GG	2 (2%)	1	1 (2%)	0	1 (1%)	1
GC	15 (16%)	17	2 (4%)	4	12 (13%)	13
CC	88 (82%)	87	46 (94%)	45	82 (86%)	82

Table 4.4: Genotype frequencies, *p53R2* C>G 4696.

4.9.4 ALLELE FREQUENCIES.

4.9.4.1 *Reprimo* G>C 824.

In both the student and diverticular control populations the C allele was rarer than the G allele, however in the CRC population the G allele was rarer (Table 4.5).

Allele	STUDENT POPULATION		DIVERTICULAR DISEASE POPULATION		COLORECTAL CANCER POPULATION	
	Number	Frequency	Number	Frequency	Number	Frequency
G	104	51 %	54	53%	81	43%
C	98	49%	48	47%	107	57%

Table 4.5: Allele frequencies, *Reprimo* G>C 824.

4.9.4.2 *P53R2* C>G 4696.

In all three populations the G allele was the rare polymorphism (Table 4.6).

Allele	STUDENT POPULATION		DIVERTICULAR DISEASE POPULATION		COLORECTAL CANCER POPULATION	
	Number	Frequency	Number	Frequency	Number	Frequency
G	19	9%	4	4%	14	7%
C	191	91%	94	96%	178	93%

Table 4.6: Allele frequencies, *p53R2* C>G 4696.

4.10 STATISTICAL ANALYSIS.

4.10.1 TESTS FOR HWE.

4.10.1.1 *Reprimo* G >C 824.

	Student population	Diverticular disease population	Colorectal cancer population
Pearson's goodness of fit χ^2 (df = 1).	P =0. 63	P = 0.02	P = 0.06
Log likelihood ratio χ^2 (df = 1).	P = 0.63	P = 0.02	P = 0.06
Exact test.	P = 0.69	P = 0.03	P = 0.09

Table 4.7: HWE tests, *Reprimo* G>C 824.

4.10.1.2 *P53R2* C>G 4696.

	Student population	Diverticular disease population	Colorectal cancer population
Pearson's goodness of fit χ^2 (df = 1).	P = 0.18	P = 0.01	P = 0.47
Log likelihood ratio χ^2 (df = 1).	P = 0.23	P < 0.01	P = 0.51
Exact test.	P = 0.19	P = 0.06	P = 0.40

Table 4.8: HWE tests, *p53R2* C>G 4696.

All three tests for deviation from HWE show that the Diverticular disease population deviates from HWE with respect to *Reprimo* G>C 824. Inspection of the observed and expected genotype frequencies reveals that there is an excess of GC heterozygotes in the Diverticular disease population. Both Pearson's goodness of fit and log likelihood χ^2 tests show that the Diverticular disease population deviates from HWE with respect to *p53R2* C>G 4696 although the exact test does not replicate this finding.

The student control population and CRC study population were both in HWE with respect to both study SNPs.

4.10.2 INBREEDING COEFFICIENTS.

4.10.2.1 *Reprimo* G>C 824.

Population	F value
Student	0.04
Diverticular disease	-0.34
Colorectal cancer	-0.19

Table 4.9: Inbreeding coefficients, *Reprimo* G>C 824.

The positive F value seen in the student population suggests heterozygote deficiency the negative values in the DD and CRC populations suggest heterozygote excess. This result is supported by the observed genotype frequencies and those expected according to the Hardy-Weinberg law.

4.10.2.2 *P53R2* C>G 4696.

Population	F value
Student	0.13
Diverticular disease	0.48
Colorectal cancer	0.08

Table 4.10: Inbreeding coefficients, *p53R2* C>G 4696.

All three populations show positive inbreeding coefficients and therefore heterozygote deficiency, again shown by the observed and expected genotype frequencies.

4.10.3 TESTS FOR DISEASE – SNP ASSOCIATION.

4.10.3.1 *Reprimo* G>C 824.

The results of the *Reprimo* G>C 824 – CRC association tests using the DD population as control do not suggest any association between the disease and the SNP (Table 4.11).

TESTS FOR *REPRIMO* G>C 824 - COLORECTAL CANCER ASSOCIATION (Diverticular disease control population)				
Allele frequency difference	Heterozygous	Homozygous	Allele positivity	Armitage's Trend test
Risk allele – C				
O.R.=1.50 C.I.=0.92-2.41 χ^2=2.58 p=0.11 (P)	O.R=1.24 C.I.=0.49-3.15 χ^2=0.21 p=0.64	O.R=2.90 C.I.=0.88-9.23 χ^2=3.18 p=0.08	O.R=1.52 C.I.=0.61-3.76 χ^2=0.83 p=0.36	O.R.=1.70 χ^2=3.37 p=0.07
Risk allele – G				
O.R. = 0.67 C.I.= 0.42-1.092 χ^2=2.58 p=0.11 (P)	O.R. = 0.44 C.I.= 0.17-1.11 χ^2=3.11 p=0.08	O.R.= 0.35 C.I.=0.11-1.13 χ^2=3.18 p=0.08	O.R =0.42 C.I.=0.17-1.04 χ^2=3.65 p=0.06	O.R. = 0.59 χ^2=3.37 p=0.07

Table 4.11: *Reprimo* G>C 824 - CRC association tests using diverticular disease control population (O.R., odds ratio; C.I., confidence interval).

TESTS FOR *REPRIMO* G>C 824 - COLORECTAL CANCER ASSOCIATION (Student control population)				
Allele frequency difference	Heterozygous	Homozygous	Allele positivity	Armitage's Trend test
Risk allele – C				
O.R. =1.40 C.I.=0.94-2.09 χ^2=2.76 p=0.10 (P)	O.R. =2.47 C.I.=1.15-5.29 χ^2=5.54 p=0.02	O.R. =2.24 C.I.=0.95-5.28 χ^2=3.46 p=0.06	O.R. =2.39 C.I.=1.15-4.96 χ^2=5.66 p=0.02	O.R. =1.47 χ^2=2.93 p=0.09
Risk allele – G				
O.R. =0.71 C.I.=[0.48-1.06] χ^2=2.76 p=0.10 (P)	O.R. =1.10 C.I.=[0.56-2.16] χ^2=0.08 p=0.78	O.R. =0.45 C.I.=[0.19-1.05] χ^2=3.46 p=0.06	O.R. =0.86 C.I.=[0.45-1.63] χ^2=0.21 p=0.64	O.R. =0.69 χ^2=2.93 p=0.09

Table 4.12: *Reprimo* G>C 824 - CRC association tests using student control population (O.R., odds ratio; C.I., confidence interval).

The heterozygous and allele positivity tests suggest association with CRC with respect to the C allele. Otherwise no other disease association between *Reprimo* G>C 824 and CRC is detected when the student population is used as the control (Table 4.11).

P53R2 C>G 4696.

TESTS FOR *P53R2* C>G 4696- COLORECTAL CANCER ASSOCIATION (Diverticular disease control population)				
Allele frequency difference	Heterozygous	Homozygous	Allele positivity	Armitage's Trend test
Risk allele – C				
O.R. = 0.54 C.I.=0.17-1.67 χ^2=1.19 p=0.28 (P)	O.R. = 6.00 C.I.=0.26-140.05 χ^2=1.47 p=0.23	O.R. = 1.78 C.I.=0.11-29.18 χ^2=0.17 p=0.68	O.R. = 1.96 C.I.=0.12-32.00 χ^2=0.23 p=0.63	O.R. = 0.75 χ^2=1.02 p=0.31
Risk allele – G				
O.R. =1.87 C.I.=0.60-5.84 χ^2=1.19 p=0.28 (P)	O.R. =3.37 C.I.=0.72-15.70 χ^2=2.64 p=0.10	O.R. =0.56 C.I.=0.03-9.18 χ^2=0.17 p=0.68	O.R. =2.43 C.I.=0.66-8.98 χ^2=1.87 p=0.17	O.R. =1.43 χ^2=1.02 p=0.31

Table 4.13: *p53R2* C>G 4696 - CRC association tests using diverticular disease control population (O.R., odds ratio; C.I., confidence interval).

No disease association is detected when using the diverticular disease population as control (Table 4.12).

TESTS FOR *P53R2* C>G 4696- COLORECTAL CANCER ASSOCIATION (Student control population)				
Allele frequency difference	Heterozygous	Homozygous	Allele positivity	Armitage's Trend test
Risk allele – C				
O.R. =1.25 C.I.=0.61-2.57 χ^2=0.37 p=0.54(P)	O.R. =1.60 C.I.=0.13-19.84 χ^2=0.14 p=0.71	O.R. =1.86 C.I.=0.17-20.94 χ^2=0.26 p=0.61	O.R. =1.83 C.I.=0.16-20.46 χ^2=0.25 p=0.62	O.R. =1.25 χ^2=0.34 p=0.56
Risk allele – G				
O.R. =0.80 C.I.=0.39-1.64 χ^2=0.37 p=0.54(P)	O.R. =0.86 C.I.=0.38-1.94 χ^2=0.13 p=0.71	O.R. =0.54 C.I.=0.05-6.03 χ^2=0.26 p=0.61	O.R. =0.82 C.I.=0.38-1.79 χ^2=0.25 p=0.62	O.R. =0.80 χ^2=0.34 p=0.56

Table 4.14: *p53R2* C>G 4696 - CRC association tests using student control population (O.R., odds ratio; C.I., confidence interval).

No disease association is detected when using the student population as control (Table 4.13).

4.11 SUMMARY OF RESULTS.

A total of 255 individuals were genotyped for the *Reprimo* G>C 824 and *p53R2* C>G 4696 SNPs. The individuals came from three populations; a colorectal cancer (CRC) study population (n=96) and two control populations; a diverticular disease (DD) age matched population (n=52) and an unmatched student (ST) population (n=107).

The DNA collection and isolation method was effective and yielded sufficient DNA to perform genotyping of 96% of individuals in the study. The genotyping technique was validated by replication and by independent laboratory sequencing of representative samples of all SNP genotypes of both genes.

The *Reprimo* G>C 824 genotype frequencies in the CRC population were 14%GG, 58%GC and 28% CC. In the DD population – 19% GG, 67% GC and 14% CC and in the ST population 28% GG, 47% GC and 25% CC. The *Reprimo* G>C 824 allele frequencies in the CRC population were G – 43% and C – 57%, in the DD population G – 53% and C – 47% and in the ST population G – 51% and C – 49%.

The *p53R2* C>G 4696 genotype frequencies in the CRC population were 1%GG, 13%GC and 86% CC. In the DD population– 2% GG, 4% GC and 94% CC and in the ST population 2% GG, 16% GC and 82% CC. The *p53R2* C>G 4696 allele frequencies in the CRC population were G – 7% and C – 93%, in the DD population G – 4% and C – 96% and in the ST population G – 9% and C – 91%.

All populations were in Hardy-Weinberg equilibrium with respect to both study SNPs except for the diverticular population with respect to *Reprimo* G>C 824, where an excess of heterozygotes was observed.

Heterozygous and allele positivity association tests detected a statistically significant association between the C allele of *Reprimo* G>C 824 and CRC when using the unmatched student population as a control. However the other association

tests did not detect any disease association with either the DD or ST population controls.

No association was detected between the *p53R2* C>G 4696 SNP and CRC with any association test or with either the DD or ST control populations.

The significance of these results will be discussed in detail in the next chapter.

Chapter 5

DISCUSSION

This chapter will discuss the results presented in chapter 5 in the context of what can be concluded about any association between the *Reprimo* G>C 824 and *p53R2* C>G 4696 SNPs with CRC and how the results compare with existing knowledge regarding the SNPs. The work as a whole will be discussed in terms of the study design and methodology in the context of similar studies and guidelines for such studies, paying attention to the strengths and weaknesses of the study. Finally there will be a discussion regarding further work that could follow on from this study.

In the past ten years there has been an explosion of SNP-disease association studies stimulated by the Human Genome Project and technological developments allowing increasingly rapid identification of SNPs. Paralleling these advances has been the refining of genetic association studies in terms of their methodology, analysis and reporting. This has lead to the publication of guidance and check-lists with the aim of increasing the reliability and veracity of such studies (Becker et al., 2003, Campbell and Rudan, 2002, Hattersley and McCarthy, 2005a, Silverman and Palmer, 2000, Cooper et al., 2002, Lewis, 2002a, Daly, 2003, Cordell and Clayton, 2005). These publications also allow readers to be able to critically appraise SNP association studies.

5.1   GENERAL DISCUSSION OF RESULTS.

The aim of the study was to investigate disease association between two SNPs (*Reprimo* G>C 824 and *p53R2* C>G 4696) and CRC by means of a case-control study. The methodology reliably identified the two SNPs and three populations

(CRC study population, diverticular disease control population and young, healthy control population) were genotyped with respect to the two SNPs. The data collected from the study were analysed using an on-line statistical calculator designed for SNP studies. This analysis involved testing each population for HWE (HWE) and allelic and genotypic association, taking into account the possible different models of gene penetrance. The results showed that all populations were in HWE with respect to both SNPs except for the diverticular disease population with respect to the *Reprimo* 824 C>G SNP, where an excess of heterozygotes was demonstrated. Deviation from HWE can be due to disease association and therefore this result may be interpreted as a positive association between heterozygotes for *Reprimo* G>C 824 and diverticular disease. Disease association tests did not show any significant association between *p53R2* C>G 4696 and CRC and so it may be concluded from these results that the *p53R2* C>G 4696 SNP has not been found to be associated with CRC in this study. An association between the C allele of the *Reprimo* G>C 824 SNP and CRC was demonstrated when heterozygous and allele positivity association tests were performed using the un-matched, young, healthy population as a control. On face value it would appear that there is an association of the C allele with CRC; however the interpretation of this result is not straight forward, as the populations are not matched and there was no similar association shown when testing with a matched control. This will be discussed further below.

5.1.2 Deviation from HWE with respect to *Reprimo* G>C 824.

The Diverticular disease population was shown to be in deviation from HWE with respect to the *Reprimo* G>C 824 SNP, with an excess of heterozygotes and lack of both G and C homozygotes. The usual causes of deviation from HWE, inbreeding, genetic drift, small population size and selection pressure (Connor and Ferguson-

Smith, 1993) cannot explain this finding in a large, randomly mating, static and late onset disease population. Deviation from HWE may also be due to genotyping errors, population stratification or chance (Silverman and Palmer, 2000, Wittke-Thompson et al., 2005a), however genotyping was shown to be accurate by independent laboratory sequencing of DNA and the sample was ethnically homogenous. The relatively small sample size does increase likelihood that HWE deviation is due to chance (Campbell and Rudan, 2002) and this must be considered in this case.

However if none of these factors are causing the deviation from HWE seen in the Diverticular disease population it may be due to genetic association (Balding et al., 2001). The nature of this association is obscure, however. It is also unusual in that the heterozygote genotype appears to be associated with the diverticular disease phenotype. Heterozygotes have been shown to alter gene function (Richmond and Powell, 1970) and survival advantage due to increased heterozygote fitness is termed heterosis, however diverticular disease does not affect biological fitness. The dominant negative effect seen in p53, where mutant gene products derived from the mutant allele form dysfunctional protein complexes with wild type proteins, is well recognised. However this study suggests both the GG and CC homozygotes function normally and it is the combination of the G and C alleles in the heterozygote that is associated with the diverticular disease phenotype.

This association would support the possibility that SNPs occurring in non-coding regions can modulate gene function, a phenomenon that has been demonstrated by 3'UTR TAF1 gene polymorphisms (Henry et al., 2001). In terms of possible biological mechanisms resulting in this association the *Reprimo* gene may have an as yet undiscovered pleiotropic function other than cell cycle control or *Reprimo* 824G>C heterozygotes may be less able to repair DNA damage from reactive oxygen species produced in inflammation in diverticulitis. Clearly more work is

required to investigate this finding, both from a population genetics and molecular pathological perspective.

5.1.3 Association between the *Reprimo* G>C 824 C allele and CRC.

When the student population was used as a control, an association was detected between the *Reprimo* G>C 824 C allele and CRC on heterozygous and allele positivity tests. These tests assume no specific genotype relationships or a dominant or recessive penetrance model. It has been demonstrated that SNP-disease associations may only be detected by χ^2 tests using different contingency Tables to account for different models of genetic penetration (Ohashi et al., 2001), and so the results may be considered significant. However, the same result was not found when association testing was performed using the diverticular disease control population, which was matched for age and ethnicity. Indeed when ethnic groups other than Caucasian are excluded from the student population and the association tests performed, no association is detected. Therefore association of the C allele of the *Reprimo* G>C 824 with CRC should be considered as a type I error, likely due to ethnic bias. Ethnicity matching in genetic case-control studies will be discussed below.

5.2 DISCUSSION OF THIS WORK IN THE CONTEXT OF SIMILAR STUDIES.

5.2.1 *Reprimo* G>C 824 and *p53R2* C>G 4696 studies.

The SNPs studied in this work were originally described by Ye and Parry (Ye and Parry, 2002b, Ye and Parry, 2002a). They found the genotype frequencies of the

Reprimo G>C 824 SNP to be CC 41.5%, GC 40.2% and GG 18.3% and of the *p53R2* G>C4696 SNP to be CC 73%, GC 23% and GG 4% in 82 healthy individuals (average age 43.5, 47 women and 35 men). The populations were in HWE. The ethnicity of the sample is not given, nor was it indicated whether the genotype frequencies were checked. These frequencies are not significantly different from the genotype frequencies found in the student population in this study, but are significantly different from the diverticular disease and CRC populations with respect to *Reprimo* G>C 824, (χ^2 tests: diverticular disease population; p<0.01, CRC population; p< 0.01) and from the diverticular population with respect to *p53R2* C>G 4696 (p=0.01). These differences may be due to population stratification due ethnicity or age differences (discussed below) or due to genotyping errors in Ye and Parry's samples. If Ye and Parry's samples are used as controls to the CRC cases in this study then disease-gene associations are detected between both the *Reprimo* G>C 824 and *p53R2* C>G 4696 SNPs. Similarly if Ye and Parry's samples are pooled with the student population used in this study (with which the genotype frequencies are similar) association is detected between *Reprimo* G>C 824 and CRC. However, these findings should be interpreted with caution due to potential population stratification in unmatched populations.

That significant differences in the genotype frequencies between Ye and Parry's studies and that three of the five SNPs reported in Ye and Parry's studies were not detected in this study (see preliminary work) demonstrates the difficulty of reproducing results of SNP studies (Ioannidis et al., 2001). The reasons for non-replication of the findings of gene-disease studies are varied, but include variations in study design, variations in relative risk of the candidate genes in different populations, the genetic complexity of complex diseases and selection of polymorphisms which are not causal (Tabor et al., 2002).

There are no published studies of either *Reprimo* G>C 824 or *p53R2* G>C association studies with either CRC, diverticular disease or any other diseases.

5.2.2   Other *Reprimo* and *p53R2* SNP studies.

Other than the papers by Ye and Parry, there are no other published papers describing *Reprimo* SNPs. In their paper Ye and Parry describe two *Reprimo* SNPs, the G>C 824 and C>G 839. As described previously it was not possible to demonstrate the *Reprimo* C>G 839 SNP despite altering the conditions as described by Ye and Parry. Similarly two other *p53R2* SNPs as described by Ye and Parry were not detected.

Including the SNPs described by Ye and Parry, seven *p53R2* SNPs have been described (Smeds et al., 2001, Ye and Parry, 2002a, Deng et al., 2005). Of these three have been tested for disease association with the colorectal adenomas, CRC precursor lesions (Deng et al., 2005). In that study no association was found between colorectal adenomas and a C>G 789 transversion, a G>A 928 transition or a T>C 933 transition.

Therefore although they are reasonable candidates for CRC association studies, there is, as yet, no evidence to support *Reprimo* and *p53R2* SNPs influencing the risk of developing CRC.

5.3   DISCUSSION OF THE STUDY DESIGN AND METHODOLOGY.

5.3.1   Population selection and matching.

5.3.1.1   Cases.

The cases used in this study were sampled from patients known to have CRC. The patients were recruited into the study at varying times from their diagnosis, i.e. they were prevalent cases. Ideally cases should be incident, that is recruited at diagnosis, to avoid selection bias (Vineis and McMichael, 1998). Also the cases recruited were mostly those with potentially curable stage disease at diagnosis and so represented a selected sample, missing those with advanced disease at diagnosis and those with very rapid progressive disease. In order to have had only incident cases recruitment would have needed to be more targeted and therefore logistically more complex. This would have made collecting sufficient samples within the time frame of the study period difficult. Also there are ethical issues with recruiting patients at the time of their diagnosis of cancer.

5.3.1.2   Control populations.

Two control populations were used in this study; a young, healthy student population and a diverticular disease population. The young, healthy student population was not matched for age or ethnicity with the cases and as such introduces the possibility of confounding due to population stratification and selection bias due to age differences (Vineis and McMichael, 1998, Campbell and Rudan, 2002). However the population was convenient and it was possible to collect tissue samples for genotyping rapidly.

The diverticular disease population was matched for age and ethnicity, but the gender distribution was different from the cases, there being a higher proportion of women in the diverticular population. Men are known to be at greater risk of CRC than women and so the gender difference is a possible source of selection bias. Further selection bias could have been introduced as all the diverticular disease patients were recruited from a population of hospital patients and as such they would have had symptomatic disease (Vineis and McMichael, 1998),(Potter, 2003). Most people with diverticulosis, i.e. the asymptomatic phenotype, do not require hospital treatment and so recruiting only hospital patients selects those with more severe forms of the disease. As with the CRC population, avoiding these biases would have made collecting the samples more difficult, and indeed it was only possible to collect a relatively small sample within the time frame of the study period.

5.3.2   Sample sizes.

Perhaps the greatest weakness of this study is the small size of the population samples used leading to a lack of statistical power. Statistical power refers to the ability of a study to eliminate type I, false positive, and type II, false negative, errors. Conventionally the ideal type I, α, level should be 0.05 and the type II, β, level 0.1. No formal power calculations were performed in the initial study design in order to determine sample sizes, rather previously similar studies were examined and it was decided to use similar sized population samples. It was also felt that it would only be possible to collect and analyse samples of around 100 individuals within the time frame of the study.

In order to detect susceptibility genes with low relative risks large sample sizes are required (Hattersley and McCarthy, 2005a, Becker et al., 2003, Dahlman et al.,

2002). The sample sizes used in this study are only likely to be able to identify significant associations for susceptibility alleles with odds ratios greater than four and frequencies of greater than 10% with α and β levels of 0.05 and 0.1 respectively.

5.3.3 DNA collection.

The DNA used for genotyping the populations in this study was isolated from buccal cells obtained by brush biopsy. This proved an efficient, reliable and well tolerated method of obtaining adequate DNA for analysis and would be recommended for further studies.

5.3.4 DNA isolation.

The method used to isolate DNA from the tissue samples proved efficient and yielded more than adequate quantities of DNA for genotyping both SNPs, repeat genotyping where necessary and to obtain independent laboratory genotyping.

5.3.5 Genotyping.

The preliminary work performed for this study indicates the difficulty in reproducing genotyping methods. Only two of the five SNPs described by Ye and Parry, *Reprimo* G>C 824 and *p53R2* C>G 4696, were successfully identified, one with altered PCR conditions. For one of the other SNPs, *Reprimo* G>C 839, it was not possible to optimize the genotyping conditions to give reliable results. For *p53R2* A>C 2752 the entire student population was genotyped as AA, and the C allele was not detected. Ye and Parry reported an AA frequency of 1% and a CC frequency of 89%. Optimisation experiments were not able to detect the C allele.

Similarly it was not possible to detect the G allele of *p53R2* A>G 2759. Genotyping errors are a common source of errors in genetic disease association studies and may lead to spurious associations (Hattersley and McCarthy, 2005a, Hosking et al., 2004a). Clues to genotyping errors may be given by unexpected results, lack of reproducibility or deviation from HWE and should prompt attempts to exclude genotyping errors, and bias, from the data (Ioannidis et al., 2001). In this study the diverticular population was found to be deviating from HWE and genotyping error was suspected. Therefore many of the samples from all three populations were re-genotyped for *Reprimo* G>C 824 and were also genotyped for both *Reprimo* G>C 824 and *p53R2* C>G 4696 by an independent laboratory. The results indicated that the genotyping that there was possibly a 5% error rate for the *Reprimo* G>C 824 allele specific PCR. Only 18 samples were sent for independent analysis, due to cost restraints, and so although the samples sent were correctly genotyped, errors in the rest of the samples could not be discounted. Therefore in interpreting the results of this study genotyping errors need to be considered.

5.3.6 Statistical methods.

The statistical methods employed in this study were straight forward and facilitated by the use of an on-line calculator specifically designed for SNP-disease association analysis. The calculator tests the population samples for HWE and disease-association based on various genetic models. As discussed above HWE testing is important to detect genotyping errors and may be used as an association test in itself (Balding, 2006). HWE was tested using three statistical tests in order to avoid confounding due to small sample sizes and small SNP frequencies. The effect of small SNP frequency on the results of the HWE tests can be seen in the results for *p53R2* C>G 4696 where there are differences in the p values of the three tests. In

this case the log likelihood ratio test is the most accurate as the *p53R2 C>G* 4696 rare allele was found to occur at 4%-9%. The extent of deviation from HWE, if present, was indicated by the inbreeding coefficient test and is used to support the tests for HWE (Salanti et al., 2005). Allele and genotype χ^2 and odds ratio association tests are important in order not to miss associations with differing genetic models of penetrance, when this is not known for the study SNP (Lewis, 2002a, Ohashi et al., 2001). The Armitage's test for trend had particular utility in this study as it is the preferred test when populations deviate from HWE, as was the case with the diverticular disease population (Slager and Schaid, 2001b, Balding, 2006).

Therefore in terms of the nature of this study the statistical methods employed are appropriate and robust enough to detect disease association in the event of various different genetic situations. However the interpretation of those results needs caution due the lack of statistical power and possible sources of bias and confounding as discussed previously.

5.4 IMPROVEMENTS THAT COULD BE MADE TO THIS STUDY.

5.4.1 Statistical power.

The fundamental weakness of this study is the small sample sizes leading to low statistical power and the increased likelihood of type I error. Therefore the most important improvement would be to repeat the study with larger sample sizes based on power calculations. Using the Power for Association With Error (P.A.W.E) on-line calculator (http://linkage.rockerfeller.edu/pawe/pawe.cgi) the minimum sample sizes necessary would be around 2000 for *p53R2 C>G* 4696 and 300 for *Reprimo G>C* 824 with a significance level of 0.05 and a power of 0.8. This programme

takes into account genotyping errors and adjusts the sample sizes necessary in order to maintain statistical power (Gordon et al., 2002a). However, collecting sample populations of this size would have been beyond the scope of this study.

5.4.2 Avoiding selection bias and confounding.

Selection bias and confounding could be reduced if only incident CRC cases were recruited and by recruiting unselected controls, this would be difficult to achieve using diverticular disease as a control population as most people with diverticulosis are asymptomatic. Therefore the control population should ideally be age and gender matched healthy individuals randomly sampled from the community.

5.5 FURTHER WORK STIMULATED BY THIS STUDY.

Although this study lacks statistical power and is prone to selection bias and confounding, questions are stimulated and further work prompted.

5.5.1 Repeating the study with greater power and avoiding bias.

The SNPs studied are reasonable candidate genes for CRC association studies and it would be useful to repeat the study with the improvements as outlined above, although this would be more difficult for a single handed investigator with limited time. However robust studies investigating the candidate disease SNPs are needed in order to understand the genetic susceptibility to CRC (Houlston and Tomlinson, 2001).

5.5.2 Investigating the association between *Reprimo* G>C 824 and diverticular disease.

The deviation from HWE seen in the diverticular disease population should be investigated firstly by genotyping a larger sample. If HWE deviation is seen in an adequately sized sample to give good statistical power then a well designed and conducted case-control disease association study should be performed. If diverticular disease is shown to have a molecular pathology it may become possible to identify genotypes that confer a greater risk developing complications, and so allowing targeted management strategies for different individuals.

5.5.3 Searching for further *Reprimo* and *p53R2* SNPs.

The SNPs studied in this study both occur in the 3'-UTR, a region thought less likely to affect gene function through polymorphisms (Tabor et al., 2002). Coding sequence SNPs are much more likely to have effects on gene function, therefore a search of the coding regions of *Reprimo* and *p53R2* should be prioritized in any search for further SNPs.

5.5.4 *Reprimo* and *p53R2* SNP interaction analysis.

As the aetiology of complex diseases such as CRC is thought to involve gene-gene, as well as gene-environment, interactions, investigating disease associations should extend beyond single candidate gene association to multiple polymorphisms (Goodman et al., 2006). Therefore further investigation into *Reprimo* and 53R2 association should involve multiple candidate genes and environmental factors.

5.5.5 Functional analysis of *Reprimo* G>C 824 and *p53R2* C>G 4696 SNPs.

Investigating the effects of DNA damaging agents on cells of known genotype would give valuable information as to any functional polymorphic variation. Such experiments would involve exposing cell cultures from individuals of known genotypes to various genotoxic agents and observing for differences in DNA damage, chromosomal damage and apoptosis.

5.6 SUMMARY.

Reprimo G>C 824 deviation from HWE seen in the diverticular disease population may be due to chance, which is more likely with the small sample size, or disease association. Disease association is difficult to explain, both from a mechanistic point of view in the light of the current understanding of the pathogenesis of diverticular disease, and from a genetic point of view as the disease association is with the heterozygote.

The CRC-*p53R2* C>G 4696 association seen on heterozygous and allele positivity testing, using the student population as a control, is likely due to different population sample ages and ethnic mix confounding and small sample sizes leading to weak statistical power.

The strengths of this study lie in the methods for collecting tissue samples using buccal brush biopsy, the efficient and high yielding DNA isolation technique, the effective genotyping method with a reasonable error rate and the use of on-line statistical software specifically designed for SNP studies.

The weaknesses are the potential for selection bias in the case samples due recruiting prevalent cases, confounding due age, gender and ethnic heterogeneity in

the control and case populations and small sample sizes leading to a statistically weak study.

The study does prompt further work: repeating the study with a more robust design to eliminate bias, confounding and increase statistical power; investigating the *Reprimo* G>C 824 – diverticular disease association; investigating other *Reprimo* and *p53R2* SNPs; functional studies of the polymorphisms and *Reprimo* and *p53R2* SNP interaction analysis.

REFERENCES

(1997) *Food, nutrition and the prevention of cancer: a global perspective,* London, World Cancer Research Fund.

(1998) *ABC of Colorectal Diseases,* BMJ books.

(2001) *ABC of Colorectal Cancer,* BMJ Books.

(2004) Improving Outcomes in Colorectal Cancers. IN SERVICE, N. H. (Ed.), National Institute for Clinical Excellence.

(2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature,* 447, 661-78.

ABDEL-RAHMAN, W. M. & PELTOMAKI, P. (2004) Molecular basis and diagnostics of hereditary colorectal cancers. *Ann Med,* 36, 379-388.

ATKIN, W. S. & SAUNDERS, B. P. (2002) Surveillance guidelines after removal of colorectal adenomatous polyps. *Gut,* 51, v6-v9.

AYESH, R., IDLE, J. R., RITCHIE, J. C., CROTHERS, M. J. & HETZEL, M. R. (1984) Metabolic oxidation phenotypes as markers for susceptibility to lung cancer. *Nature,* 312, 169-70.

BABA, S. (1997) Recent Advances in Molecular Genetics of Colorectal Cancer. *World Journal of Surgery,* 21, 678-687.

BALDING, D. J. (2006) A tutorial on statistical methods for population association studies. *Nat Rev Genet,* 7, 781-91.

BALDING, D. J., BISHOP, M. & CANNINGS, C. (2001) Chapter 19. *Handbook of Statistical Genetics.*

BECKER, N., NIETERS, A. & RITTGEN, W. (2003) Single nucleotide polymorphism - disease relationships: statistical issues for the performance of association studies. *Mutation Research,* 525, 11-18.

BEIGHTON, P., DE PAEPE, A., STEINMANN, B., TSIPOURAS, P. & WENSTRUP, R. J. (1998) Ehlers-Danlos syndromes: revised nosology, Villefranche, 1997. Ehlers-Danlos National Foundation (USA) and Ehlers-Danlos Support Group (UK). *Am J Med Genet,* 77, 31-7.

BENHAMOU, S., LEE, W. J., ALEXANDRIE, A. K., BOFFETTA, P., BOUCHARDY, C., BUTKIEWICZ, D., BROCKMOLLER, J., CLAPPER, M. L., DALY, A., DOLZAN, V., FORD, J., GASPARI, L., HAUGEN, A., HIRVONEN, A., HUSGAFVEL-PURSIAINEN, K., INGELMAN-SUNDBERG, M., KALINA, I., KIHARA, M., KREMERS, P., LE MARCHAND, L., LONDON, S. J., NAZAR-STEWART, V., ONON-KIHARA, M., RANNUG, A., ROMKES, M., RYBERG, D., SEIDEGARD, J., SHIELDS, P., STRANGE, R. C., STUCKER, I., TO-FIGUERAS, J., BRENNAN, P. & TAIOLI, E. (2002) Meta- and pooled analyses of the effects of glutathione S-transferase M1 polymorphisms and smoking on lung cancer risk. *Carcinogenesis,* 23, 1343-50.

BERGSTRÖM, A., PISANI, P., TENET, V., WOLK, A. & ADAMI, H. (2001) Overweight as an avoidable cause of cancer in Europe. *Int J Cancer,* 91, 421-30.

BERWICK, M. & VINEIS, P. (2000) Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review. *J Natl Cancer Inst,* 92, 874-97.

BLANKS, R. G., MOSS, S. M., MCGAHAN, C. E., QUINN, M. J. & BABB, P. J. (2000) Effect of NHS breast screening programme on mortality from breast cancer in England and Wales, 1990-8: comparison of observed with predicted mortality. *BMJ,* 321, 665-9.

BLATT, J. & HUNTLEY, D. (1989) Enhancement of in vitro activity against neuroblastoma by doxorubicin and deferoxamine. *J Natl Cancer Inst,* 81, 866-70.

BROOKES, A. J. (1999) The essence of SNPs. *Gene,* 234, 177-86.

BROWN, I., HEYS, S. D. & SCHOFIELD, A. C. (2003) From peas to "chips" - the new millennium of molecular biology: a primer for the surgeon. *World J Surg Oncol,* 1, 21.

BURN, J., CHAPMAN, P. D., BISHOP, D. T., SMALLEY, S., MICKLEBURGH, I., WEST, S. & MATHERS, J. C. (2001) Susceptibility markers in colorectal cancer. *IARC SCi Publ,* 154, 131-47.

BYERS, T. (2000) Diet, colorectal adenomas, and colorectal cancer. *N Engl J Med,* 342, 1206-7.

BYUN, D. S., CHAE, K. S., RYU, B. K., LEE, M. G. & CHI, S. G. (2002) Expression and mutation analyses of P53R2, a newly identified p53 target for DNA repair in human gastric carcinoma. *Int J Cancer,* 98, 718-23.

CAMPBELL, H. & RUDAN, I. (2002) Interpretation of genetic association studies in complex disease. *The Pharmacogenetics Journal,* 2, 349-360.

COLDITZ, G. A., CANNUSCIO, C. C. & FRAZIER, A. L. (1997) Physical activity and reduced risk for colon cancer: implications for prevention. *Cancer Causes Control,* 8, 649-67.

COLLINS, F. S., BROOKS, L. D. & CHAKRAVARTI, A. (1998a) A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res,* 8, 1229-31.

COLLINS, F. S., GUYER, M. S. & CHARKRAVARTI, A. (1997) Variations on a theme: cataloguing human DNA sequence variation. *Science,* 278, 1580-1.

COLLINS, F. S., PATRINOS, A., JORDAN, E., CHAKRAVARTI, A., GESTELAND, R. & WALTERS, L. (1998b) New goals for the U.S. Human Genome Project: 1998-2003. *Science,* 282, 682-9.

CONNOR, J. M. & FERGUSON-SMITH, M. A. (1993) Chapter 11. *Essential Medical Genetics.* 4th ed.

COOPER, D., NUSSBAUM, R. & KRAWCZAK, M. (2002) Proposed guidelines for papers describing DNA polymorphism-disease associations. *Hum Genet,* 110, 207-208.

CORDELL, H. & CLAYTON, D. (2005) Genetic association studies. *The Lancet,* 366, 1121-31.

CORRAO, G., BAGNARDI, V., ZAMBON, A. & ARICO, S. (1999) Exploring the dose-response relationship between alcohol consumption and the risk of several alcohol related conditions: a meta-analysis. *Addiction,* 94, 1551-73.

COUNCIL, G. M. *Research: The role and responsibilities of doctors.*

CRANDALL, C. J. (1999) Estrogen replacement therapy and colon cancer: a clinical review. *J Women's Health Gen Based Med,* 1155-66.

CUNNINGHAM, C. & DUNLOP, M. G. (1996) Molecular genetic basis of colorectal cancer susceptibility. *British Journal of Surgery,* 83, 321-329.

DAHLMAN, I., EAVES, I. A., KOSOY, R., MORRISON, V. A., HEWARD, J., GOUGH, S. C., ALLAHABADIA, A., FRANKLYN, J. A., TUOMILEHTO, J., TUOMILEHTO-WOLF, E., CUCCA, F., GUJA, C., IONESCU-TIRGOVISTE, C., STEVENS, H., CARR, P., NUTLAND, S., MCKINNEY, P., SHIELD, J. P., WANG, W., CORDELL, H. J., WALKER, N., TODD, J. A. & CONCANNON, P. (2002) Parameters for reliable results in genetic association studies in common disease. *Nat Genet,* 30, 149-50.

DALY, A. (2003) Candidate gene case-control studies. *Pharmacogenetics,* 4, 127-139.

DAY, D. A. & TUITE, M. F. (1998) Post-transcriptional gene regulatory mechanisms in eukaryotes: an overview. *J Endocrinol,* 157, 361-71.

DELVAUX, M. (2003) Diverticular disease of the colon in Europe: epidemiology, impact on citizen health and prevention. *Aliment Pharmacol Ther,* 18 Suppl 3, 71-4.

DENG, Z. L., XIE, D. W., BOSTICK, R. M., MIAO, X. J., GONG, Y. L., ZHANG, J. H. & WARGOVICH, M. J. (2005) Novel genetic variations of the p53R2 gene in patients with colorectal adenoma and controls. *World J Gastroenterol,* 11, 5169-73.

DESHPANDE, A. V., OLIVER, M., YIN, M., GOH, T. H. & HUTSON, J. M. (2005) Severe colonic diverticulitis in an adolescent with Williams syndrome. *J Paediatr Child Health,* 41, 687-8.

DIETEL, M. (1998) Molecular pathology of colorectal cancer: From phenotype to genotype. *Recent results in cancer research,* 146, 3-19.

DOXEY, B. W., KUWADA, S. K. & BURT, R. W. (2005) Inherited Polyposis Syndromes: Molecular Mechanisms, Clinicopathology, and Genetic Testing. *Clinical gastroenterology and hepatology,* 3, 633-641.

DUNLOP, M. G. (2002a) Guidance on gastrointestinal surveillance for hereditary non-polyposis colorectal cancer, familial adenomatous polyposis, and Peutz-Jeghers syndrome. *Gut,* 51, v21-v27.

DUNLOP, M. G. (2002b) Guidance on large bowel surveillance for people with two first degree relatives with colorectal cancer or one first degree relative diagnosed with colorectal cancer under 45 years. *Gut,* 51, v17-v20.

EADEN, J. A. & MAYBERRY, J. F. (2002) Guidelines for screening and surveillance of asymptomatic colorectal cancer in patients with inflammatory bowel disease. *Gut,* 51, v10 - v12.

EASTWOOD, M. (1987) Dietary fibre and the risk of cancer. *Nutr. Review,* 45, 193-198.

ELIASHAR, R., SICHEL, J. Y., BIRON, A. & DANO, I. (1998) Multiple gastrointestinal complications in Marfan syndrome. *Postgrad Med J,* 74, 495-7.

ELLIS, N. A., GRODEN, J., YE, T.-Z., STRAUGHEN, J., LENNON, D. J., CIOCCI, S., PROYTCHEVA, M. & GERMAN, J. (1995) The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell,* 83, 655-666.

FEARON, E. R. & VOLGESTEIN, B. (1990) A genetic model for colorectal tumorigenesis. *Cell,* 61, 759-67.

FRANCESCHI, S. & VECCHIA, C. L. (1998) Colorectal cancer and hormone replacement therapy: an unexpected finding. *Eur J Cancer Prev,* 7, 427-38.

FRIEND, S. H. & STOUGHTON, R. B. (2002) The magic of microarrays. *Sci Am,* 286, 44-9, 53.

FUKUSHIMA, T. & TAKENOSHITA, S. (2001) Colorectal carcinogenesis. *Fukushima Journal of Medical Science,* 47, 1-11.

GALIATSATOS, P. & FOULKES, F. D. (2006) Familial Adenomatous Polyposis. *American Journal of Gastroenterology,* 101, 385-398.

GEAR, J. S., WARE, A., FURSDON, P., MANN, J. I., NOLAN, D. J., BRODRIBB, A. J. & VESSEY, M. P. (1979) Symptomless diverticular disease and intake of dietary fibre. *Lancet,* 1, 511-4.

GIBBS, W. W. (2003) Untangling the roots of cancer. *Sci Am,* 289, 56-65.

GIOVANNUCCI, E. (2001) An updated review of the epidemiological evidence that cigarette smoking increases the risk of colorectal cancer. *Cancer Epedemiol Biomarkers Prev,* 10, 725-31.

GLADE, M. J. (1999) Food, nutrition, and the prevention of cancer: a global perspective. American Institute for Cancer Research/World Cancer Research Fund, American Institute for Cancer Research, 1997. *Nutrition,* 15, 523-6.

GOLDBOHM, R. A., BRANDT, P. A. V. D., VEER, P. V. T., BRANTS, H. A. M., DORANT, E., STURMANS, F. & HERMUS, R. J. J. (1994) A prospective cohort study on the relation between meat consumption and the risk of colon cancer. *Cancer Research,* 54, 718-723.

GOLDSTEIN, F., NEWCOMB, P. A. & STAMPFER, M. J. (1999) Post menopausal hormone therapy and the risk of colorectal cancer: a review and meta-analysis. *Am J Med,* 106, 574-82.

GOODMAN, J. E., MECHANIC, L. E., LUKE, B. T., AMBS, S., CHANOCK, S. & HARRIS, C. C. (2006) Exploring SNP-SNP interactions and colon cancer risk using polymorphism interaction analysis. *Int J Cancer,* 118, 1790-7.

GORDON, D., FINCH, S., NOTHNAGEL, M. & OTT, J. (2002a) Power and Sample Size Calculations for Case-Control Genetic Association Tests when Errors are Present: Applications to Single Nucleotide Polymorphisms. *Hum Hered,* 54, 22-33.

GORDON, D., FINCH, S. J., NOTHNAGEL, M. & OTT, J. (2002b) Power and sample size calculations for case-control genetic association tests when errors are present: application to single nucleotide polymorphisms. *Hum Hered,* 54, 22-33.

GOTO, Y., YUE, L., YOKOI, A., NISHIMURA, R., UEHARA, T., KOIZUMI, S. & SAIKAWA, Y. (2001) A novel single-nucleotide polymorphism in the 3'-untranslated region of the human dihydrofolate reductase gene with enhanced expression. *Clin Cancer Res,* 7, 1952-6.

GRAY, J. R. & DAVIES, S. J. (1996) Marfan syndrome. *J Med Genet,* 33, 403-8.

GROGAN, L., BEHAN, K. A. & JOHNSTON, P. G. (1996) Recent advances in the biology of colorectal cancer. *Histology and histopathology,* 11, 807-820.

GUITTET, O., HAKANSSON, P., VOEVODSKAYA, N., FRIDD, S., GRASLUND, A., ARAKAWA, H., NAKAMURA, Y. & THELANDER, L. (2001) Mammalian p53R2 protein forms an active ribonucleotide reductase in vitro with the R1 protein, which is expressed both in resting cells in response to DNA damage and in proliferating cells. *J Biol Chem,* 276, 40647-51.

HANSON, C. A. & MILLER, J. R. (2005) Non-traditional roles for the Adenomatous Polyposis Coli (APC) tumour suppressor protein. *Gene,* 361, 1-12.

HARKIN, D. P. (2006) Genomics and the impact of new technologies on the management of colorectal cancer. *Oncologist,* 11, 988-91.

HATTERSLEY, A. & MCCARTHY, M. (2005a) What makes a good genetic association study? *The Lancet,* 366, 1315-23.

HATTERSLEY, A. T. & MCCARTHY, M. I. (2005b) What makes a good genetic association study? *Lancet,* 366, 1315-23.

HAYASHI, H., FURIHATA, M., KUWAHARA, M., KAGAWA, S., SHUIN, T. & OHTSUKI, Y. (2004) Infrequent alteration in the p53R2 gene in human transitional cell carcinoma of the urinary tract. *Pathobiology,* 71, 103-6.

HEINEN, C. D., GOSS, K. H., CORNELIUS, J. R., BABCOCK, G. F., KNUDSON, E. S., KOWALIK, T. & GRODEN, J. (2002) The APC Tumor Suppressor Controls Entry Into S-Phase Through Its Ability to Regulate the Cylcin D/RB Pathway. *Gastroenterology,* 123, 751-763.

HENRY, M., AUBERT, H., MORANGE, P. E., NANNI, I., ALESSI, M. C., TIRET, L. & JUHAN-VAGUE, I. (2001) Identification of polymorphisms in the promoter and the 3' region of the TAFI gene: evidence that plasma TAFI antigen levels are strongly genetically controlled. *Blood,* 97, 2053-8.

HEWITSON, P., GLASZIOU, P., IRWIG, L., TOWLER, B. & WATSON, E. (2007) Screening for colorectal cancer using the faecal occult blood test,

Hemoccult. *Cochrane Database of Systematic Reviews*, Art. No.: CD001216. DOI: 10.1002/14651858.CD001216.pub2.

HIGASHI, Y., TANAE, A., INOUE, H., HIROMASA, T. & FUJII-KURIYAMA, Y. (1988) Aberrant splicing and missense mutations cause steroid 21-hydroxylase [P-450(C21)] deficiency in humans: possible gene conversion products. *Proc Natl Acad Sci U S A,* 85, 7486-90.

HOSKING, L., LUMSDEN, S., LEWIS, K., YEO, A., MCCARTHY, L., BANSAL, A., RILEY, J., PURVIS, I. & CHUN-FANG, X. (2004a) Detection of genotyping errors by Hardy-Weinberg equilibrium testing. *European Journal of Human Genetics,* 12, 395-399.

HOSKING, L., LUMSDEN, S., LEWIS, K., YEO, A., MCCARTHY, L., BANSAL, A., RILEY, J., PURVIS, I. & XU, C. F. (2004b) Detection of genotyping errors by Hardy-Weinberg equilibrium testing. *Eur J Hum Genet,* 12, 395-9.

HOULSTON, R. S. & PETO, J. (2004) The search for low-penetrance cancer susceptibility alleles. *Oncogene,* 23, 6471-6.

HOULSTON, R. S. & TOMLINSON, I. P. (2001) Polymorphisms and colorectal tumor risk. *Gastroenterology,* 121, 282-301.

HOWE, J. R. & GUILLEM, J. G. (1997) The Genetics of Colorectal Cancer. *Surgical Clinics of North America,* 77, 175-195.

IMYANITOV, E. N., TOGO, A. V. & HANSON, K. P. (2004) Searching for cancer-associated gene polymorphisms: promises and obstacles. *Cancer Lett,* 204, 3-14.

IOANNIDIS, J., NTZANI, E., TRIKALINOS, T. & CONTOPOULOS-IOANNIDIS, D. (2001) Replication validity of genetic association studies. *Nature Genetics,* 29, 306-309.

IRELAND, T. A. O. C. O. G. B. A. (2001) Guidelines for the management of colorectal cancer.

ISSA, J.-P. J., OTTAVIANO, Y. L., CELANO, P., HAMILTON, S. R., DAVIDSON, N. E. & BAYLIN, S. B. (1994) Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nature Genetics,* 7, 536-540.

JARVINEN, H. J. (2004) Hereditary cancer: guidelines in clinical practice. Colorectal cancer genetics. *Annals of Oncology,* 15, iv127-iv131.

JENSON, O. M., MACLENNAN, R. & WARENDORF, J. (1982) Diet, bowel function, fecal characteristics, and large bowel cancer in Denmark and Finland. *Nutr. Cancer,* 4, 5-19.

JO, W.-S. & CHUNG, D. C. (2005) Genetics of Hereditary Colorectal cancer. *Semin Oncol,* 32, 11-23.

JONG, M. M. D., NOLTE, I. M., TEMEERMAN, G. J., GRAAF, W. T. A. V. D., VRIES, E. G. E. D., SIJMONS, R. H., HOFSTRA, R. M. W. & KLEIBEUKER, J. H. (2002) Low-penetrance Genes and Their Involvement in Colorectal Cancer Susceptibility. *Cancer Epidemiology, Biomarkers & Prevention,* 11, 1332-1352.

JUN, S. & STOLLMAN, N. (2002) Epidemiology of diverticular disease. *Best Pract Res Clin Gastroenterol,* 16, 529-42.

KANEKO, H., FUKAO, T. & KONDO, N. (2004) The function of RecQ helicase gene family (especially BLM) in DNA recombination and joining. *Adv Biophys,* 38, 45-64.

KANG, J. Y., MELVILLE, D. & MAXWELL, J. D. (2004) Epidemiology and management of diverticular disease of the colon. *Drugs Aging,* 21, 211-28.

KARUMAN, P., GOZANI, O., ODZE, R. D., ZHOU, X. C., ZHU, H., SHAW, R., BRIEN, T. P., BOZZUTO, C. D., OOI, D., CANTLEY, L. C. & YUAN, J.

(2001) The Peutz-Jegher (sic) gene product LKB1 is a mediator of p53-dependent cell death. *Molecular Cell,* 7, 1307-1319.

KAUPPI, B., NIELSEN, B. B., RAMASWAMY, S., LARSEN, I. K., THELANDER, M., THELANDER, L. & EKLUND, H. (1996) The three-dimensional structure of mammalian ribonucleotide reductase protein R2 reveals a more-accessible iron-radical site than Escherichia coli R2. *J Mol Biol,* 262, 706-20.

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, R177-85.

KIM, Y. I. (2000) AGA technical review: impact of dietary fiber on colon cancer occurrence. *Gastroenterology,* 118, 1235-57.

KIMURA, T., TAKEDA, S., SAGIYA, Y., GOTOH, M., NAKAMURA, Y. & ARAKAWA, H. (2003) Impaired function of p53R2 in Rrm2b-null mice causes severe renal failure through attenuation of dNTP pools. *Nat Genet,* 34, 440-5.

KIRK, B. W., FEINSOD, M., FAVIS, R., KLIMAN, R. M. & BARANY, F. (2002) Single nucleotide polymorphism seeking long term association with complex disease. *Nucleic Acids Res,* 30, 3295-311.

KNUDSON, A. G. (1971) Mutation and cancer: a statistical study of retinoblastoma. *Proc Natl Acad Sci USA,* 68, 820-3.

KOMAZAWA, N., SUZUKI, A., SANO, S., HORIE, K., MATSUURA, N., MAK, T. W., NAKANO, T., TAKEDA, J. & KONDOH, G. (2004) Tumorigenesis facilitated by Pten deficiency in the skin: evidence of p53-Pten complex formation on the initiation phase. *Cancer Sci,* 95, 639-43.

KONDO, Y. & ISSA, J.-P. J. (2004) Epigenetic changes in colorectal cancer. *Cancer and Metastasis Reviews,* 23, 29-39.

KOUNTOURAS, J., BOURA, P. & LYGIDAKIS, N. J. (2000) New Concepts of Molecular Biology for Colon Carcinogenesis. *Hepato-gastroenterology,* 47, 1291-1297.

KRONES, C. J., KLINGE, U., BUTZ, N., JUNGE, K., STUMPF, M., ROSCH, R., HERMANNS, B., HEUSSEN, N. & SCHUMPELICK, V. (2006) The rare epidemiologic coincidence of diverticular disease and advanced colonic neoplasia. *Int J Colorectal Dis,* 21, 18-24.

KWAK, E. L. & CHUNG, D. C. (2007) Hereditary Colorectal cancer Syndromes: An Overview. *Clinical Colorectal Cancer,* 6, 340-344.

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. & VOLGESTEIN, B. (1997) Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. *Nature Genetics,* 17, 79-83.

LAMPE, J. W., SLAVIN, J. L., MELCHER, E. A. & POTTER, J. D. (1992) Effects of cereal and vegetable fiber feeding on potential risk factors for colon cancer. *Cancer Epidemiology, Biomarkers & Prevention,* 1, 207-211.

LE, N. T. & RICHARDSON, D. R. (2002) The role of iron in cell cycle progression and the proliferation of neoplastic cells. *Biochem Biophys Acta,* 1603, 31-46.

LESLIE, A., CAREY, F. A., PRATT, N. R. & STEELE, R. J. C. (2002) The colorectal adenoma-carcinoma sequence. *British Journal of Surgery,* 89, 845-860.

LEWIS, C. (2002a) Genetic association studies: Design, analysis and interpretation. *Briefings in Bioinformatics,* 3, 146-153.

LEWIS, C. M. (2002b) Genetic association studies: design, analysis and interpretation. *Brief Bioinform,* 3, 146-53.

LIN, Z. P., BELCOURT, M. F., CORY, J. G. & SARTORELLI, A. C. (2004) Stable suppression of the R2 subunit of ribonucleotide reductase by R2-targeted short interference RNA sensitizes p53(-/-) HCT-116 colon cancer cells to DNA-damaging agents and ribonucleotide reductase inhibitors. *J Biol Chem,* 279, 27030-8.

LIPTON, L. & TOMLINSON, I. (2006) The genetics of FAP and FAP-like syndromes. *Familial Cancer,* 5, 221-226.

LITTLE, J. & FAIVRE, J. (1999) Family history, metabolic gene polymorphism, diet and risk of colorectal cancer. *Eur J Cancer Prev,* 8, S61-S72.

LOFFELD, R. J. L. F. & PUTTEN, A. B. M. M. V. D. (2001) Diverticular disease of the colon and concomitant abnormalities in patients undergoing endoscopic evaluation of the large bowel. *Colorectal Disease,* 4, 189-192.

LOWER, G. M., JR., NILSSON, T., NELSON, C. E., WOLF, H., GAMSKY, T. E. & BRYAN, G. T. (1979) N-acetyltransferase phenotype and risk in urinary bladder cancer: approaches in molecular epidemiology. Preliminary results in Sweden and Denmark. *Environ Health Perspect,* 29, 71-9.

MAPLE, J. T. & BOARDMAN, L. A. (2006) Genetics of Colonic Polyposis. *Clinical gastroenterology and hepatology,* 4, 831-835.

MARTIN, S. G. & JOHNSTON, D. S. (2003) A role for Drosophila LKB1 in anterior-posterior axis formation and epithelial polarity. *Nature,* 421, 379-384.

MATSUBARA, D., NIKI, T., ISHIKAWA, S., GOTO, A., OHARA, E., YOKOMIZO, T., HEIZMANN, C. W., ABURATANI, H., MORIYAMA, S., MORIYAMA, H., NISHIMURA, Y., FUNATA, N. & FUKAYAMA, M. (2005) Differential expression of S100A2 and S100A4 in lung adenocarcinomas: clinicopathological significance, relationship to p53 and identification of their target genes. *Cancer Sci,* 96, 844-57.

MCCALLUM, A., EASTWOOD, M. A., SMITH, A. N. & FULTON, P. M. (1987) Colonic Diverticulosis in Patients with Colorectal Cancer and in Controls. *Scand J Gastroenterology,* 23, 284-6.

MCCARTHY, E. M. & PHILLIPS, J. A., 3RD (1998) Characterization of an intron splice enhancer that regulates alternative splicing of human GH pre-mRNA. *Hum Mol Genet,* 7, 1491-6.

MERG, A., LYNCH, H. T., LYNCH, J. F. & HOWE, J. R. (2005a) Hereditary Colon Cancer - Part I. *Current Problems in Surgery,* 42, 185-256.

MERG, A., LYNCH, H. T., LYNCH, J. F. & HOWE, J. R. (2005b) Hereditary Colon Cancer - Part II. *Current Problems in Surgery,* 42, 257-334.

MICHELS, K. B., GIOVANNI, E., JOSHIPURA, K. J., ROSNER, B. A., STAMPFER, M. J., FUCHS, C. S. & AL, E. (2000) Prospective study of fruit and vegetable consumption and the incidence of colon and rectal cancers. *J Natl Cancer Inst,* 92, 421-30.

MIMURA, T., EMANUEL, A. & KAMM, M. A. (2002) Pathophysiology of diverticular disease. *Best Pract Res Clin Gastroenterol,* 16, 563-76.

MITRUNEN, K. & HIRVONEN, A. (2003) Molecular epidemiology of sporadic breast cancer. The role of polymorphic genes involved in oestrogen biosynthesis and metabolism. *Mutat Res,* 544, 9-41.

MORINI, S., ANGELIS, P. D., MANURITA, L. & COLAVOLPE, V. (1988) Association of Colonic Diverticula with Adenomas and Carcinomas: A Colonoscopic Experience. *Dis Colon Rectum,* 31, 793 - 796.

MORINI, S., HASSAN, C., ZULLO, A., FRANCESCO, V. D., FESTA, V., BARBERANI, F., FALEO, D. & STROFFOLINI, T. (2002) Diverticular

disease as a risk factor for sigmoid colon adenomas. *Digest Liver Dis,* 34, 635-9.

MORT, R., MO, L., MCEWAN, C. & MELTON, D. W. (2003) Lack of involvement of nucleotide excision repair gene polymorphisms in colorectal cancer. *British Journal of Cancer,* 89, 333-337.

NAKAJI, S., DANJO, K., MUNAKATA, A., SUGAWARA, K., MACAULEY, D., KERNOHAN, G. & BAXTER, D. (2002) Comparison of etiology of right-sided diverticula in Japan with that of left-sided diverticula in the West. *Int J Colorectal Dis,* 17, 365-73.

NAKAMURA, Y. (2004) Isolation of p53-target genes and their functional analysis. *Cancer Sci,* 95, 7-11.

NAKANO, K., BALINT, E., ASHCROFT, M. & VOUSDEN, K. H. (2000) A ribonucleotide reductase gene is a transcriptional target of p53 and p73. *Oncogene,* 19, 4283-9.

NANDA, K., BASTION, L. A., HASSELBLAD, V. & SIMEL, D. L. (1999) Hormone replacement therapy and the risk of colorectal cancer: a meta-analysis. *Obstet Gynecol,* 93, 880-8.

NAVARATNAM, R. V., CHOWANIEC, J. & WINSLET, M. C. (1999) The molecular biology of colorectal cancer development and the associated genetic events. *Annals of the Royal College of Surgeons of England,* 81, 312-319.

NELSON, K. A. & WITTE, J. S. (2002) Androgen receptor CAG repeats and prostate cancer. *Am J Epidemiol,* 155, 883-90.

NORAT, T., LUKANOVA, A., FERRARI, P. & RIBOLI, E. (2002) Meat consumption and colorectal cancer risk: dose response meta-analysis of epidemiological studies. *Int J Cancer,* 98, 241-56.

O'CONNELL, J. B., MAGGARD, M. A. & KO, C. Y. (2004) Colon cancer survival rates with the new American Joint Committee on Cancer Sixth Edition staging. *J Natl Cancer Inst,* **96, 1420-1425.**

OHASHI, J., YAMAMOTO, S., TSUCHIYA, N., HATTA, Y., KOMATA, T., MATSUSHITA, M. & TOKUNAGA, K. (2001) Comparison of statistical power between 2 * 2 allele frequency and allele positivity tables in case-control studies of complex disease genes. *Ann Hum Genet,* 65, 197-206.

OHKI, R., NEMOTO, J., MURASAWA, H., ODA, E., INAZAWA, J., TANAKA, N. & TANIGUCHI, T. (2000) Reprimo, a new candidate mediator of the p53-mediated cell cycle arrest at the G2 phase. *J Biol Chem,* 275, 22627-30.

OKUMURA, H., NATSUGOE, S., MATSUMOTO, M., MATAKI, Y., TAKATORI, H., ISHIGAMI, S., TAKAO, S. & AIKOU, T. (2005) The predictive value of p53, p53R2, and p21 for the effect of chemoradiation therapy on oesophageal squamous cell carcinoma. *Br J Cancer,* 92, 284-9.

PETRUZZIELLO, L., IACOPINI, F., BULAJIC, M., SHAH, S. & COSTAMAGNA, G. (2006) Review article: uncomplicated diverticular disease of the colon. *Aliment Pharmacol Ther,* 23, 1379-91.

POTTER, J. D. (2003) Epidemiology, cancer genetics and microarrays: making correct inferences, using appropriate designs. *Trends in Genetics,* 19, 690-695.

PRICOLO, V. E., FINKELSTEIN, S. D. & BLAND, K. I. (1997) Topographic genotyping of colorectal carcinoma: From a molecular carcinogenesis model to clinical relevance. *Annals of Surgical Oncology,* 4, 269-278.

REDDY, B. S. & WYNDER, E. L. (1977) Metabolic epidemiology of colon cancer. Fecal bile acids and neutral sterols in colon cancer patients and patients with adenomatous polyps. *Cancer,* 39, 2533-9.

RICHMOND, R. C. & POWELL, J. R. (1970) Evidence of heterosis associated with an enzyme locus in a natural population of Drosophila. *Proc Natl Acad Sci U S A,* 67, 1264-7.

RISCH, N. & MERIKANGAS, K. (1996) The future of genetic studies of complex human diseases. *Science,* 273, 1516-7.

ROSSOUW, J. E., ANDERSON, G. L., PRENTICE, R. L., LACROIX, A. Z., KOOPERBERG, C., STEFANICK, M. L., JACKSON, R. D., BERESFORD, S. A., HOWARD, B. V., JOHNSON, K. C., KOTCHEN, J. M. & OCKENE, J. (2002) Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *Jama,* 288, 321-33.

RUEFF, J., GASPAR, J. & KRANENDONK, M. (2002) DNA polymorphisms as modulators of genotoxicity and cancer. *Biol Chem,* 383, 923-32.

SALANTI, G., AMOUNTZA, G., NTZANI, E. E. & IOANNIDIS, J. P. (2005) Hardy-Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power. *Eur J Hum Genet,* 13, 840-8.

SANDHU, M. S., WHITE, I. R. & MCPHERSON, K. (2001) Systematic review of the prospective cohort studies on meat consumption and colorectal cancer risk: a meta-analytical approach. *Cancer Epedemiol Biomarkers,* 10, 349-46.

SASIENI, P., ADAMS, J. & CUZICK, J. (2003) Benefit of cervical screening at different ages: evidence from the UK audit of screening histories. *British Journal of Cancer,* 89, 88 - 93.

SATO, N., FUKUSHIMA, N., MAITRA, A., MATSUBAYASHI, H., YEO, C. J., CAMERON, J. L., HRUBAN, R. H. & GOGGINS, M. (2003) Discovery of novel targets for aberrant methylation in pancreatic carcinoma using high-throughput microarrays. *Cancer Res,* 63, 3735-42.

SCHOLEFIELD, J. H. & STEELE, R. J. (2002) Guidelines for follow up after resection of colorectal cancer. *Gut,* 51, v3-v5.

SCHORK, N. J., FALLIN, D. & LANCHBURY, J. S. (2000) Single nucleotide polymorphisms and the future of genetic epidemiology. *Clin Genet,* 58, 250-64.

SEIDEGARD, J., PERO, R. W., MILLER, D. G. & BEATTIE, E. J. (1986) A glutathione transferase in human leukocytes as a marker for the susceptibility to lung cancer. *Carcinogenesis,* 7, 751-3.

SHAO, J., ZHOU, B., ZHU, L., QIU, W., YUAN, Y. C., XI, B. & YEN, Y. (2004) In vitro characterization of enzymatic properties and inhibition of the p53R2 subunit of human ribonucleotide reductase. *Cancer Res,* 64, 1-6.

SHEILDS, P. G., SHEILDS, C. L., JR, R. C. E. & SINGH, A. D. (2001) Adenocarcinoma arising from congenital hypertrophy of retinal pigment epithelium. *Archives of Ophthalmology,* 119, 597-602.

SHIELDS, P. G. & HARRIS, C. C. (2000) Cancer risk and low-penetrance susceptibility genes in gene-environment interactions. *J Clin Oncol,* 18, 2309-15.

SILVERMAN, E. K. & PALMER, L. J. (2000) Case-Control association Studies for the Genetics of Complex Respiratory Diseases. *Am. J. Respir. Cell Mol. Biol,* 22, 645-648.

SIMPSON, J., SCHOLEFIELD, J. H. & SPILLER, R. C. (2002) Pathogenesis of colonic diverticula. *Br J Surg,* 89, 546-54.

SIMPSON, J., SCHOLEFIELD, J. H. & SPILLER, R. C. (2003) Origin of symptoms in diverticular disease. *British Journal of Surgery,* 90, 899-908.

SIOMI, H. & DREYFUSS, G. (1997) RNA-binding proteins as regulators of gene expression. *Curr Opin Genet Dev,* 7, 345-53.

SLAGER, S. L. & SCHAID, D. J. (2001a) Case-control studies of genetic markers: power and sample size approximations for Armitage's test for trend. *Hum Hered,* 52, 149-53.

SLAGER, S. L. & SCHAID, D. J. (2001b) Case-Control Studies of Genetic Markers: Power and Sample Size Approximations for Armitage's Test for Trend. *Hum Hered,* 52, 149-153.

SMEDS, J., BERGGREN, P., MA, X., XU, Z., HEMMINKI, K. & KUMAR, R. (2002) Genetic status of cell cycle regulators in squamous cell carcinoma of the oesophagus: the CDKN2A (p16(INK4a) and p14(ARF)) and p53 genes are major targets for inactivation. *Carcinogenesis,* 23, 645-55.

SMEDS, J., NAVA, M., KUMAR, R. & HEMMINKI, K. (2001) A novel polymorphism (-88 C>A) in the 5' UTR of the p53R2 gene. *Hum Mutat,* 17, 82.

STATISTICS, O. F. N. (2007) Cancer incidence and mortality in the United Kingdom.

STEFANSSON, T., EKBOM, A., SPAREN, P. & LPAHLMAN (1993) Increased risk of left sided colon cancer in patients with diverticular disease. *Gut,* 34, 499-502.

STEFANSSON, T., EKBOM, A., SPAREN, P. & LPAHLMAN (1995) Cancers among Patients Diagnosed as Having Diverticular Disease of the Colon. *Eur J Surg,* 161, 755-760.

STEINBERG, M. H. & ADAMS, J. G., 3RD (1982) Thalassemia: recent insights into molecular mechanisms. *Am J Hematol,* 12, 81-92.

STOLLMAN, N. & RASKIN, J. B. (2004) Diverticular disease of the colon. *Lancet,* 363, 631-9.

STREMMEL, C., WEIN, A., HOHENBERGER, W. & REINGRUBER, B. (2002) DNA microarrays: a new diagnostic tool and its implications in colorectal cancer. *Int J Colorectal Dis,* 17, 131-136.

STUMPF, M., CAO, W., KLINGE, U., KLOSTERHALFEN, B., KASPERK, R. & SCHUMPELICK, V. (2001) Increased distribution of collagen type III and reduced expression of matrix metalloproteinase 1 in patients with diverticular disease. *Int J Colorectal Dis,* 16, 271-5.

SUZUKI, M., SHIGEMATSU, H., SHIVAPURKAR, N., REDDY, J., MIYAJIMA, K., TAKAHASHI, T., GAZDAR, A. F. & FRENKEL, E. P. (2006) Methylation of apoptosis related genes in the pathogenesis and prognosis of prostate cancer. *Cancer Lett.*

SUZUKI, M., SHIGEMATSU, H., TAKAHASHI, T., SHIVAPURKAR, N., SATHYANARAYANA, U. G., IIZASA, T., FUJISAWA, T. & GAZDAR, A. F. (2005) Aberrant methylation of Reprimo in lung cancer. *Lung Cancer,* 47, 309-14.

TABOR, H., RISCH, N. & MYERS, R. (2002) Candidate-gene approaches for studying complex traits: practical considerations. *Nature Reviews Genetics,* 3, 391-7.

TAKAHASHI, T., SUZUKI, M., SHIGEMATSU, H., SHIVAPURKAR, N., ECHEBIRI, C., NOMURA, M., STASTNY, V., AUGUSTUS, M., WU, C. W., WISTUBA, II, MELTZER, S. J. & GAZDAR, A. F. (2005) Aberrant methylation of Reprimo in human malignancies. *Int J Cancer,* 115, 503-10.

TAKAYAMA, T., MIYANISHI, K., HAYASHI, T., SATO, Y. & NIITSU, Y. (2006) Colorectal cancer: genetics of development and metastasis. *J Gastroenterol,* 41, 185-192.

TANAKA, H., ARAKAWA, H., YAMAGUCHI, T., SHIRAISHI, K., FUKUDA, S., MATSUI, K., TAKEI, Y. & NAKAMURA, Y. (2000) A ribonucleotide reductase gene involved in a p53-dependent cell-cycle checkpoint for DNA damage. *Nature,* 404, 42-9.

TAYLOR, W. R. & STARK, G. R. (2001) Regulation of the G2/M transition by p53. *Oncogene,* 20, 1803-15.

TERDIMAN, J. P. (2000) Genomic events in the adenoma to carcinoma sequence. *Seminars in Gastrointestinal Disease,* 11, 194-206.

THORTON, J. R. (1981) High colonic pH promotes colorectal cancer. *Lancet,* 1, 1081-1082.

TIANEN, M., VAAHTOMERI, K., YLIKORKALA, Y. & MAKELA, T. P. (2002) Growth arrest by the LKB1 tumour suppressor: induction of p21 (WAF1/CIP1). *Human Molecular Genetics,* 11, 1497-1504.

VASEN, H. F., WATSON, P., MECKLIN, J. P. & LYNCH, H. T. (1999) New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology,* 116, 1453-6.

VINEIS, P., MARINELLI, D., AUTRUP, H., BROCKMOLLER, J., CASCORBI, I., DALY, A. K., GOLKA, K., OKKELS, H., RISCH, A., ROTHMAN, N., SIM, E. & TAIOLI, E. (2001) Current smoking, occupation, N-acetyltransferase-2 and bladder cancer: a pooled analysis of genotype-based studies. *Cancer Epidemiol Biomarkers Prev,* 10, 1249-52.

VINEIS, P. & MCMICHAEL, A. (1998) Bias and confounding in molecular epidemiological studies: special considerations. *Carcinogenesis,* 19, 2063-2067.

VINEIS, P., VEGLIA, F., BENHAMOU, S., BUTKIEWICZ, D., CASCORBI, I., CLAPPER, M. L., DOLZAN, V., HAUGEN, A., HIRVONEN, A., INGELMAN-SUNDBERG, M., KIHARA, M., KIYOHARA, C., KREMERS, P., LE MARCHAND, L., OHSHIMA, S., PASTORELLI, R., RANNUG, A., ROMKES, M., SCHOKET, B., SHIELDS, P., STRANGE, R. C., STUCKER, I., SUGIMURA, H., GARTE, S., GASPARI, L. & TAIOLI, E. (2003) CYP1A1 T3801 C polymorphism and lung cancer: a pooled analysis of 2451 cases and 3358 controls. *Int J Cancer,* 104, 650-7.

VOGELSTEIN, B., FEARON, E. R., HAMILTON, S. R., KERN, S. E., PREISINGER, A. C., LEPPERT, M., NAKAMURA, Y., WHITE, R., SMITS, A. M. & BOS, J. L. (1988) Genetic alterations during colorectal-tumour development. *N Eng J Med,* 319, 525-32.

WANG, D. G., FAN, J. B., SIAO, C. J., BERNO, A., YOUNG, P., SAPOLSKY, R., GHANDOUR, G., PERKINS, N., WINCHESTER, E., SPENCER, J., KRUGLYAK, L., STEIN, L., HSIE, L., TOPALOGLOU, T., HUBBELL, E., ROBINSON, E., MITTMANN, M., MORRIS, M. S., SHEN, N., KILBURN, D., RIOUX, J., NUSBAUM, C., ROZEN, S., HUDSON, T. J., LIPSHUTZ, R., CHEE, M. & LANDER, E. S. (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science,* 280, 1077-82.

WANG, Z. & MOULT, J. (2001) SNPs, protein structure, and disease. *Hum Mutat,* 17, 263-70.

WEBB, E. L., RUDD, M. F., SELLICK, G. S., EL GALTA, R., BETHKE, L., WOOD, W., FLETCHER, O., PENEGAR, S., WITHEY, L., QURESHI, M., JOHNSON, N., TOMLINSON, I., GRAY, R., PETO, J. & HOULSTON, R. S. (2006) Search for low penetrance alleles for colorectal cancer through a scan of 1467 non-synonymous SNPs in 2575 cases and 2707 controls with validation by kin-cohort analysis of 14 704 first-degree relatives. *Hum Mol Genet,* 15, 3263-71.

WESS, L., EASTWOOD, M. A., EDWARDS, C. A., BUSUTTIL, A. & MILLER, A. (1996) Collagen alteration in an animal model of colonic diverticulosis. *Gut,* 38, 701-6.

WESS, L., EASTWOOD, M. A., WESS, T. J., BUSUTTIL, A. & MILLER, A. (1995) Cross linking of collagen is increased in colonic diverticulosis. *Gut,* 37, 91-4.

WHITEWAY, J. & MORSON, B. C. (1985) Elastosis in diverticular disease of the sigmoid colon. *Gut,* 26, 258-66.

WICKENS, M., ANDERSON, P. & JACKSON, R. J. (1997) Life and death in the cytoplasm: messages from the 3' end. *Curr Opin Genet Dev,* 7, 220-32.

WILLETT, W. C., STAMPFER, M. J., COLDITZ, G. A., ROSNER, B. A. & SPEIZER, F. E. (1990) Relation of meat, fat and fibre intake to the risk of colon cancer in a prospective study among women. *N Eng J Med,* 323, 1664-1672.

WILSON, J. M. G. & JUNGER, G. (1968) Principles and practice of screening for disease. *WHO Chronicle,* 22, 473.

WITTKE-THOMPSON, J., PLUZHNIKOV, A. & COX, N. (2005a) Rational Inferences about departure from Hardy-Weinberg Equilibrium. *Am.J. Hum. Genet,* 76, 967-986.

WITTKE-THOMPSON, J. K., PLUZHNIKOV, A. & COX, N. J. (2005b) Rational inferences about departures from Hardy-Weinberg equilibrium. *Am J Hum Genet,* 76, 967-86.

WONG, J. J. L., HAWKINS, N. J. & WARD, R. L. (2007) Colorectal cancer: A model for epigenetic tumorigenesis. *Gut,* 56, 140-148.

WONG, T. S., KWONG, D. L., SHAM, J. S., WEI, W. I. & YUEN, A. P. (2005) Methylation status of Reprimo in head and neck carcinomas. *Int J Cancer,* 117, 697.

XUE, L., ZHOU, B., LIU, X., QIU, W., JIN, Z. & YEN, Y. (2003) Wild-type p53 regulates human ribonucleotide reductase by protein-protein interaction with p53R2 as well as hRRM2 subunits. *Cancer Res,* 63, 980-6.

YAMAGUCHI, T., MATSUDA, K., SAGIYA, Y., IWADATE, M., FUJINO, M. A., NAKAMURA, Y. & ARAKAWA, H. (2001) p53R2-dependent pathway for DNA synthesis in a p53-regulated cell cycle checkpoint. *Cancer Res,* 61, 8256-62.

YANAMOTO, S., IWAMOTO, T., KAWASAKI, G., YOSHITOMI, I., BABA, N. & MIZUNO, A. (2005) Silencing of the p53R2 gene by RNA interference inhibits growth and enhances 5-fluorouracil sensitivity of oral cancer cells. *Cancer Lett,* 223, 67-76.

YANAMOTO, S., KAWASAKI, G., YOSHITOMI, I. & MIZUNO, A. (2003) Expression of p53R2, newly p53 target in oral normal epithelium, epithelial dysplasia and squamous cell carcinoma. *Cancer Lett,* 190, 233-43.

YE, Z. & PARRY, J. M. (2002a) The discovery and confirmation of single nucleotide polymorphisms in the human p53R2 gene by EST database analysis. *Mutagenesis,* 17, 361-4.

YE, Z. & PARRY, J. M. (2002b) Identification of polymorphisms in the human Reprimo gene using public EST data. *Teratog Carcinog Mutagen,* 22, 485-93.

YEOWELL, H. N. & PINNELL, S. R. (1993) The Ehlers-Danlos syndromes. *Semin Dermatol,* 12, 229-40.

YOSHIDA, T., HAGA, S., NUMATA, Y., YAMASHITA, K., MIKAMI, T., OGAWA, T., OHKUSA, T. & OKAYASU, I. (2006) Disruption of the p53-p53r2 DNA repair system in ulcerative colitis contributes to colon tumorigenesis. *Int J Cancer,* 118, 1395-403.

ZHANG, Z., LI, J., LANTRY, L. E., WANG, Y., WISEMAN, R. W., LUBET, R. A. & YOU, M. (2002) p53 transgenic mice are highly susceptible to 1, 2-dimethylhydrazine-induced uterine sarcomas. *Cancer Res,* 62, 3024-9.

ZHOU, B., LIU, X., MO, X., XUE, L., DARWISH, D., QIU, W., SHIH, J., HWU, E. B., LUH, F. & YEN, Y. (2003) The human ribonucleotide reductase subunit hRRM2 complements p53R2 in response to UV-induced DNA repair in cells with mutant p53. *Cancer Res,* 63, 6583-94.

ZHOU, S., BUCKHAULTS, P., ZAWEL, L., BUNZ, F., RIGGINS, G., DAI, J. L., KERN, S. E., KINZLER, K. W. & VOGELSTEIN, B. (1998) Targeted deletion of Smad4 shows it is required for transforming growth factor beta and activin signalling in colorectal cancer cells. *Proc Natl Acad Sci,* 95, 2412-2416.

BIBLIOGRAPHY

BALDING, D. J., BISHOP, M. & CANNINGS, C. (2003) *Handbook of Statistical Genetics*, Wiley.

EVANS, D. & GRUBA, P. (2002) *How to Write a Better Thesis*, Melbourne University Press.

JONES, D. (1998) *ABC of Colorectal Diseases,* London, BMJ Books.

KEIGHLEY, M. R. B. & WILLIAMS, N. S. (1999) *Surgery of the Anus, Rectum & Colon*, Balliere Tindall.

KERR, D. J. (2001) *ABC of colorectal cancer,* London, BMJ Books.

PHILLIPS, R. K. S. (2005) *Colorectal Surgery,* Philadelphia, Elsevier.

STRACHAN, T. & READ, A. P. (1999) *Human Molecular Genetics 2,* Oxford, BIOS Scientific Publishers Ltd.

Web resources:

M. Tevfik Dorak's Website: www.dorakmt

National Center for Biotechnology Information: www.ncbi.nlm.nih.gov

APPENDICES

APPENDIX A: PATIENT INFORMATION SHEETS

Appendix A1

INFORMATION SHEET FOR COLORECTAL CANCER PATIENTS

Contact: Dr William Beasley, Research Registrar, Singleton Hospital.

BOWEL CANCER GENETICS RESEACH

p53R2 and Reprimo gene single nucleotide polymorphism frequencies in random, diverticular disease and colorectal cancer populations.

You are being invited to take part in a research study. Before you decide 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 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, P.O. Box 1365, LONDON N16 0BW.

Thank you for reading this.

What is the purpose of the study?

We know that cancer develops due to damaged DNA within cells and that in some cases the risk of developing bowel cancer is inherited due to defective genes being passed from generation to generation.

The aim of this study is to see whether different forms of two newly discovered genes, p53R2 and Reprimo, influence a persons risk of developing bowel cancer.

The results of the research may be used, together with other information, to screen people for bowel cancer and help to catch the disease at an early, curable stage.

It should take two years to complete the research and we will need to collect DNA samples from around two hundred people.

INFORMATION SHEET FOR COLORECTAL CANCER PATIENTS

Why have I been chosen?

As someone who has had bowel cancer, we are interested to see what forms of the two genes, p53R2 and Reprimo, you have and whether people who develop bowel cancer have different forms of the genes from people who do not develop bowel cancer.

Do I have to take part?

Your taking part in this research is entirely voluntary. If you decide to take part you will be given this information sheet to keep. This will not affect the standard of care you receive.

What will happen to me if I take part?

All we ask of you is to let us take a sample of cells from the inside of your cheek using a special brush. We will do this before you go home after your appointment; it takes no more than a couple of minutes to collect the sample.
We will also collect a few details about you and your diagnosis from your hospital records. These details will be kept confidential.
The sample you give will be analyzed to see what forms of the genes you have.
Once we have collected enough samples from different people who do and do not have bowel cancer we will be able to compare the results to see whether the different forms of the genes affect peoples' risk of developing bowel cancer.

How will the results affect me?

The results of this research will not affect any subsequent treatment you may need. However the results may help us to treat future patients with bowel cancer.

Will my taking part in this study be kept confidential?

All information that is collected about you during the course of the research will be kept strictly confidential. Any information about you that leaves the hospital will have your name and address removed so that you cannot be recognized from it.

INFORMATION SHEET FOR COLORECTAL CANCER PATIENTS

What will happen to the results of the research?

The research is being conducted as part of a Master of Surgery Degree by Dr. William D Beasley and will be used to produce a thesis.
The results may also be published in medical or scientific journals.

Who is organizing and funding the research?

The research is being organized and funded by Dr William D Beasley, with support from the colorectal surgery department, Singleton Hospital, and the Centre for Molecular Genetics and Toxicology, School of Biological Sciences, University of Wales, Swansea.
Additional funding has been applied for in the form of a Surgical Research Fellowship from the Royal College of Surgeons of England.

Who has reviewed the research?

The Iechyd Morgannwg Health Local Research Ethics Committee has reviewed the research.

Thank you for taking the time to read this letter and taking part in this research.

Appendix A2

INFORMATION SHEET FOR DIVERTICULAR DISEASE PATIENTS

Contact: Dr William Beasley, Research Registrar, Singleton Hospital.

BOWEL CANCER GENETICS RESEACH

p53R2 and Reprimo gene single nucleotide polymorphism frequencies in random, diverticular disease and colorectal cancer populations.

NB: WE WOULD LIKE TO MAKE IT CLEAR THAT THERE IS NO SUSPICION THAT YOU HAVE BOWEL CANCER.

You are being invited to take part in a research study. Before you decide 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 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, P.O. Box 1365, LONDON N16 0BW.

Thank you for reading this.

What is the purpose of the study?

We know that cancer develops due to damaged DNA within cells and that in some cases the risk of developing bowel cancer is inherited due to defective genes being passed from generation to generation.

The aim of this study is to see whether different forms of two newly discovered genes, p53R2 and Reprimo, influence a persons risk of developing bowel cancer.

The results of the research may be used, together with other information, to screen people for bowel cancer and help to catch the disease at an early, curable stage.

It should take two years to complete the research and we will need to collect DNA samples from around two hundred people.

INFORMATION SHEET FOR DIVERTICULAR DISEASE PATIENTS

Why have I been chosen?

You have been diagnosed as having diverticular disease. This is thought to develop partly due to lack of fibre in the diet. A lack of fibre is also thought to increase the risk of developing bowel cancer. People also tend to develop diverticular disease at the same age as those who develop bowel cancer, and tend to live in the same areas.

However having diverticular disease does not increase your risk of developing bowel cancer.

We want to see whether this difference is due to people with diverticular disease and people with bowel cancer having different forms of the genes p53R2 and Reprimo.

Do I have to take part?

Your taking part in this research is entirely voluntary. If you decide to take part you will be given this information sheet to keep. This will not affect the standard of care you receive.

What will happen to me if I take part?

All we ask of you is to let us take a sample of cells from the inside of your cheek using a special brush. We will do this before you go home after your appointment; it takes no more than a couple of minutes to collect the sample.

We will also collect a few details about you and your diagnosis from your hospital records. These details will be kept confidential.

The sample you give will be analyzed to see what forms of the genes you have. Once we have collected enough samples from different people who do and do not have bowel cancer we will be able to compare the results to see whether the different forms of the genes affect peoples' risk of developing bowel cancer.

How will the results affect me?

The results of this research will not affect any subsequent treatment you may need. However the results may help us to treat future patients with bowel cancer.

INFORMATION SHEET FOR DIVERTICULAR DISEASE PATIENTS

Will my taking part in this study be kept confidential?

All information that is collected about you during the course of the research will be kept strictly confidential. Any information about you that leaves the hospital will have your name and address removed so that you cannot be recognized from it.

What will happen to the results of the research?

The research is being conducted as part of a Master of Surgery Degree by Dr. William D Beasley and will be used to produce a thesis.
The results may also be published in medical or scientific journals.

Who is organizing and funding the research?

The research is being organized and funded by Dr William D Beasley, with support from the colorectal surgery department, Singleton Hospital, and the Centre for Molecular Genetics and Toxicology, School of Biological Sciences, University of Wales, Swansea.
Additional funding has been applied for in the form of a Surgical Research Fellowship from the Royal College of Surgeons of England.

Who has reviewed the research?

The Iechyd Morgannwg Health Local Research Ethics Committee has reviewed the research.

Thank you for taking the time to read this letter and taking part in this research.

APPENDIX B: CONSENT FORMS

Appendix B1

CONSENT FORM FOR COLORECTAL CANCER POPULATION.

Patient identification number for this trial............
N.B. Three copies should be made for (1) patient, (2) researcher, (3) hospital notes

Title of project :p53R2 and Reprimo single nucleotide polymorphisms in random, diverticular disease and colorectal cancer populations.

Name of researcher :Dr William D Beasley

Contact telephone number:01792 205666

Please initial Box.

1. I confirm that I have read and understood the information sheet for the above study and have had the opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected. ☐

3. I understand that sections of any of my medical notes may be looked at by the research team or from regulatory authorities where it is relevant to my taking part in research. I give permission for these individuals to have access to my records
   . ☐

4. I agree to take part in the above study. ☐

Name of Patient   Date   Signature

Name of Person taking consent   Date   Signature

Researcher   Date   Signature

Appendix B2

CONSENT FORM FOR DIVERTICULAR DISEASE POPULATION.

Patient identification number for this trial..............................

N.B. Three copies should be made for (1) patient, (2) researcher, (3) hospital notes

Title of project :p53R2 and Reprimo single nucleotide polymorphisms in random, diverticular disease and colorectal cancer populations.

Name of researcher : Dr William D Beasley

Contact telephone number:01792 205666

Please
initial
box.

1.   I confirm that I have read and understood the information sheet
   for the above study and have had the opportunity to ask questions.   ☐

2.   I understand that my participation is voluntary and that I am free
   to withdraw at any time, without giving any reason, without my
   medical care or legal rights being affected.   ☐

3.   I understand that sections of any of my medical notes may be looked
   at by the research team or from regulatory authorities where it is
   relevant to my taking part in research. I give permission for these
   individuals to have access to my records
   .   ☐

4.   I agree to take part in the above study.   ☐

Name of Patient   Date   Signature

Name of Person taking consent   Date   Signature

Researcher   Date   Signature

APPENDIX C: PUBLICATIONS ARISING FROM THIS STUDY

Appendix C1

PEER REVIEWED PAPER

Reprimo 824 G>C and p53R2 4696 C>G Single Nucleotide Polymorphisms and Colorectal Cancer: A Case – Control Disease Association Study.

William D Beasley (Corresponding author), Department of Colorectal Surgery, Singleton Hospital, Swansea NHS Trust Sketty, Swansea, SA2 8QA, Wales and Centre for Molecular Genetics and Toxicology, School of Biological Sciences, University of Wales Swansea, Singleton Park, Swansea SA2 8PP, Wales. E-mail: williamdbeasley@hotmail.com

John Beynon, Department of Colorectal Surgery, Singleton Hospital, Swansea NHS Trust Sketty, Swansea, SA2 8QA, Wales. E-mail: johnbeynon56@hotmail.com,

Gareth J S Jenkins, Centre for Molecular Genetics and Toxicology, School of Biological Sciences, University of Wales Swansea, Singleton Park, Swansea SA2 8PP, Wales. E-mail: g.j.jenkins@swansea.ac.uk,

Jim M Parry, Centre for Molecular Genetics and Toxicology, School of Biological Sciences, University of Wales Swansea, Singleton Park, Swansea SA2 8PP, Wales. E-mail: jmp@swan.ac.uk

Abstract

Background

Improved survival from colorectal cancer (CRC) may result from screening for inherited genetic risk factors. Reprimo and p53R2 are p53 inducible genes involved in cell cycle surveillance and DNA repair. Single nucleotide polymorphisms (SNPs) of these genes have been discovered but their effects on the genes' function and association with CRC is not known.

Methods

90 healthy controls, 52 diverticular disease controls and 96 colorectal cancer cases were genotyped. DNA was extracted from buccal brush biopsies. Genotyping was performed by polymerase chain reaction (PCR) or polymerase chain reaction/restriction fragment length polymorphism (PCR/RFLP) methods. Tests for Hardy-Weinberg equilibrium and allelic- and genotype-disease association were performed on line using the Finetti program.

Results

All three populations were in Hardy-Weinberg equilibrium with respect to p53R2 4696C>G SNP and no CRC associations were demonstrated with this SNP. The healthy and CRC populations were in Hardy-Weinberg equilibrium with respect to the Reprimo 824G>C SNP but the diverticular disease population was not (p= 0.03). No CRC were demonstrated with Reprimo 824G>C.

Conclusion

No association between p53R2 4696C>G or Reprimo 824G>C with colorectal cancer was shown by this study. An association between the Reprimo 824G>C heterozygote and diverticular disease may exist on the basis of deviation from Hardy-Weinberg equilibrium.

Background

Colorectal cancer (CRC) is the second most common cause of death from malignant disease accounting for nearly 400,000 deaths per year worldwide [1]. The molecular pathology of colorectal carcinogenesis is well understood along the adenoma-carcinoma sequence [2, 3], resulting in the development of genetic paradigms outlining the fundamental molecular steps along the pathway from normal colonic epithelium to adenoma to carcinoma [4-6].

The majority of CRC cases are sporadic [6] and the genetic susceptibility factors involved are unknown but it is thought that common, low penetrance genetic variants, including SNPs, combined with environmental factors (a western style diet rich in red meat and fat and low in fibre and vegetables, obesity and smoking) may influence CRC risk and modulate the genetic steps in the adenoma-carcinoma sequence [7]. Screening for genetic susceptibility factors could potentially have a profound impact on reducing CRC mortality by allowing clinicians to assess individual's' genetic risk of developing CRC and tailoring prevention strategies, screening and treatment to the individual.

p53-inducible ribonucleotide reductase small subunit 2 homologue (p53R2) is located at chromosome 8q23.1 It is a direct target for the tumour suppressor gene p53, and its induction in response to DNA damage assists in G2 arrest and provides

DNA precursors for DNA repair [8, 9]. Dysfunction of p53R2 could result in a failure of DNA damage repair and so lead to gene mutation or cellular apoptotic activation [10].

Impaired p53R2 induction and RR formation with hRRM1 in response to UV has been demonstrated in p53 mutant cell lines [11] and studies have suggested a possible association of p53R2 with oral carcinogenesis [12].

Three SNPs in the p53R2 gene have been identified with in the 3' untranslated region (UTR) of the gene [13]. The effects of these SNPs on p53R2 function are not known but it is easy to see that if they do interfere with function, DNA repair could be compromised and therefore would have a potential role in carcinogenesis. An 8 base pair polymorphism in the 5' UTR of p53R2 has also been detected but no association was found with a variety of cancers [14].

Reprimo is a p53 inducible gene located at chromosome 2q23 [15]. Reprimo expression leads to arrest of the cell cycle at G2/M by inhibition of CDK1 activity and nuclear translocation of cyclin B1 [16]. Cell cycle arrest at this stage in the cell cycle allows repair of damaged DNA or to permanent arrest of severely damaged cells. Defective Reprimo function could result in failure to repair DNA damage and increased cell proliferation. Lowered Reprimo expression and an increased incidence in uterine sarcomas has been shown in p53$^{val135/wt}$ mice exposed to 1,2 dimethylhydrazine [17]. Aberrant Reprimo Methylation has been shown in pancreatic cancer [18].

Two 3' UTR SNPs of the *Reprimo* gene have been identified [19]; It is not known whether these polymorphisms have any effect on Reprimo expression or function.

To date many studies have looked for and identified associations between polymorphisms, including SNPs, and CRC [7, 20, 21]. On the basis of the understanding of the functions of Reprimo and p53R2 it is biologically reasonable

to select p53R2 and Reprimo polymorphisms as candidates for CRC association studies.

This paper presents the results of a population based CRC – SNP association study of Reprimo 824 G>C and p53R2 4696 C>G using a young healthy unmatched population and an aged matched diverticular disease control population.

MATERIALS AND METHODS.

Subjects.

Three populations were studied:

1. A healthy young adult control population (N = 90; 60 female, 30 male. Average age 19.1 years. Ethnicity: 100% Caucasian).

2. A Diverticular disease, age matched, population. (N = 52; 37 female, 15 male. Average age 67.5 years. Ethnicity: 100% Caucasian). All had either endoscopically or radiographically proven diverticular disease.

3. Colorectal cancer population. (N = 96; 32 female, 64 male. Average age 68.3 years. Ethnicity: 100% Caucasian). All had histologically proven colorectal adenocarcinoma.

All subjects were volunteers and gave fully informed consent to participate in the study. Ethical approval was obtained from the local ethics committee.

DNA collection and extraction.

DNA was extracted by standard methods (Puregene™ DNA Purification System, Gentra Systems) from buccal cell brush biopsies.

Genotyping.

The Reprimo 824G>C SNP was identified by an allele specific polymerase chain reaction (PCR) method. Oligonucleotide primers (Upstream; 5′-AGAGGGCGATTAGGGCGCAG-3′, 5′-AGAGGGCGATTAGGGCGCAC-3′, Downstream; 5′-AGGAGAAGAGTGGGAGCGC-3′) as described by Ye and Parry [19] were used and supplied by MWG. PCR was performed in a PTC-225 Peltier Thermal Cycler (MJ research) under the following conditions; five minutes initial denaturation at 94°C, followed by 32 cycles of 94°C for 45 seconds, 62°C for 30 seconds, 72°C for 45 seconds and a five minute final extension at 72°C.

The p53R2 4696 C>G SNP was identified by a polymerase chain reaction/ restriction fragment length polymorphism method. The SNP creates a digestion site for the *Hind III* restriction enzyme. Primers (5′-TTCTGTACATGCAGGCTTGG-3′, 5′-AGATGGAAGGCTGGGAGAAT-3′) were as described by Ye and Parry [13]. The PCR took place under the following conditions; Five minutes initial denaturation at 94°C followed by 32 cycles of 94°C for 10 seconds, 57°C for 20 seconds and 72°C for 45 seconds and a final extension at 72°C for 5 minutes. PCR products were digested with *Hind III* at 37°C overnight.

The reaction mix used for both genes was; 1.5 mM MgCl$_2$, 10 mM Tris-HCl Buffer (pH 8.8), 100 µM each dNTP, 100 pmols upstream primer, 100 pmols downstream

primer, 2.5 U *Taq* DNA Polymerase (Promega corp.), 2 µl template DNA solution (average concentration 2.5ng/µl) in a total volume of 50µl.

All PCR and PCR/RFLP products were visualised on a 2.5% agarose gel with ethidium bromide staining.

Validating genotyping fidelity.

Samples of PCR products were sent for independent sequencing in order to verify the efficacy of the PCR technique.

Statistical analysis.

The SNPs were tested for Hardy-Weinberg equilibrium and genotype and allelic disease association on-line using the de Finetti programme [22] accessed via the Genestat Statistical Genetics web site[23]. The programme generates expected genotype frequencies, allele frequencies and inbreeding coefficient from observed genotype frequencies. Deviation from Hardy-Weinberg equilibrium is tested by Pearson's goodness-of-fit, Log likelihood ratio and exact chi-square tests (with 1 degree of freedom). Tests for association are performed using chi square and odds ratio for allele-, heterozygote- and homozygote- frequency differences, allele positivity and Armitage's trend test. The tests were carried out twice for each SNP using both control populations. P values of less than 0.05 were significant.

Results

Gel electrophoresis patterns for both SNPs are shown in Figures two and three. Reprimo homozygotes are visualised as a single band and heterozygotes as a double band, each band representing either the G or C allele. P53R2 GG homozygote PCR products do not contain the *Hind III* digestion site and were not cut, showing as a single 228 bp electrophoresis band. P53R2 CC homozygote PCR products contain the *Hind III* digestion site and so show as 78bp and 149 bp electrophoresis bands. The p53R2 C>G 4696 heterozygotes are represented by three bands, at 228, 149 and 78 bp.

Figure 1: Electrophoresis gel visualisation of Reprimo 824G>C genotypes.
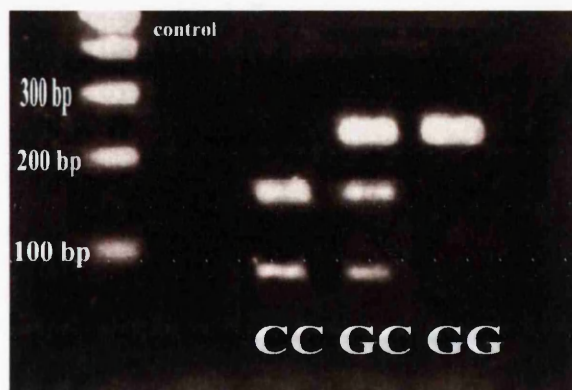
Figure 2: Electrophoresis gel visualisation of p53R2 4696C>G genotypes.

Sequencing of representative PCR product samples from both genes and all

populations confirmed the techniques to identify the SNPs were specific.

Chromograms showing the SNPs are shown in table one.
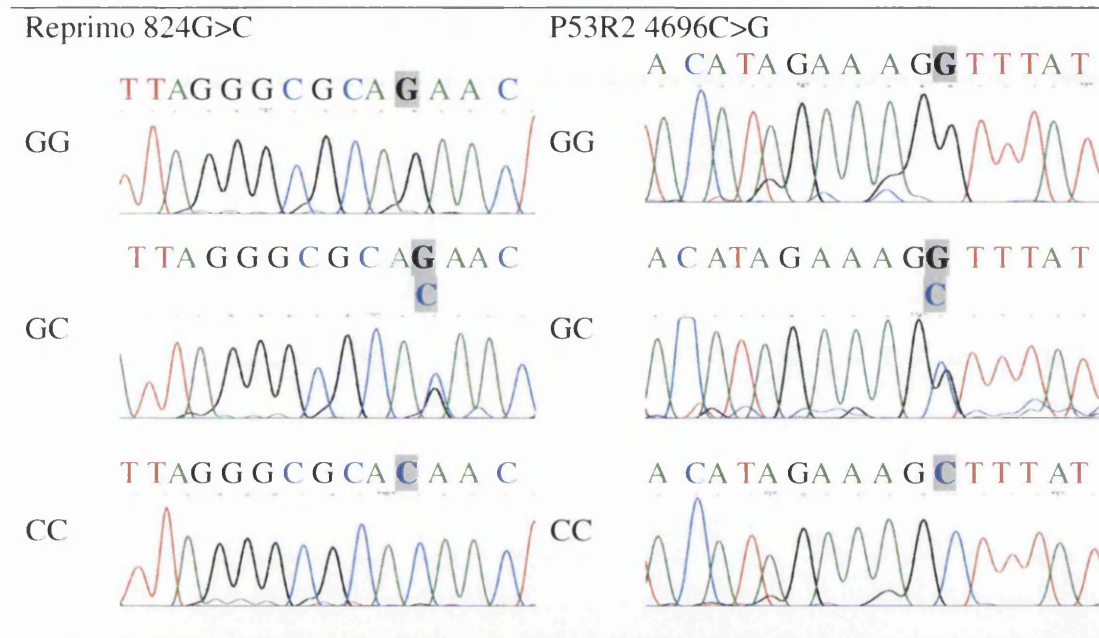
Table 1. Reprimo and p53R2 sequence chromograms.

Tests for Hardy-Weinberg equilibrium (Table 2).

The Hardy-Weinberg equilibrium is the relationship between gene frequencies and

genotype frequencies found within large, stable, randomly mating populations

where $p^2+2pq+q^2=1$ and p and q are the frequencies of the alleles occurring in that

population. Deviation from Hardy-Weinberg equilibrium can indicate disease

association.

	STUDENT POPULATION	DIVERTICULAR DISEASE POPULATION	COLORECTAL CANCER POPULATION
Reprimo 824 G>C	Observed (expected) genotypes		
	GG=24(23) GC=41(43) CC=21(20)	GG=10(14) GC=34(25) CC=7(11)	GG=13(17) GC=55(46) CC=26(30
	Allele frequencies		
	G=0.52 C=0.48	G=0.53 C=0.47	G=0.43 C=0.57
	Exact test (1 degree of freedom)		
	P=0.67	P=0.03	P=0.09
P53R2 4696C>G	Observed (expected) genotypes		
	GG=2(1) GC=11(14) CC=77(76)	GG=1(0) GC=2(4) CC=46(45)	GG=1(1) GC=12(13) CC=82(82)
	Allele frequencies		
	G=0.08 C=0.92	G=0.04 C=0.96	G=0.07 C=0.93
	Exact test (1 degree of freedom)		
	P=0.10	P=0.06	P=0.40

Table 2. Tests for Hardy-Weinberg Equilibrium. Statistically significant results are shown in red.

All three populations were in Hardy-Weinberg equilibrium with respect to the p53R2 C>G 4696 SNP. The student and colorectal cancer populations were in Hardy-Weinberg equilibrium with respect to the Reprimo G>C 824 SNP but the diverticular disease population was in deviation from Hardy-Weinberg equilibrium (p=0.03) with an excess of heterozygotes and lack of both homozygotes.

Tests for disease association (Tables 3 and 4).

No associations between Reprimo 824 G>C or p53R2 C>G SNPs and CRC were

demonstrated in this study.

TESTS FOR ASSOCIATION (95% CONFIDENCE INTERVALS)						
SNP	Populations tested	Allele frequency difference (1 df)	Heterozygous	Homozygous	Allele positivity	Armitage's trend test
Reprimo **824** **G>C**	**CRC v. DD (matched control)**	C ALLELE				
		G v. C	**GG v. GC**	**GG v. CC**	**GG v. GC+CC**	**Common odds ratio**
		O.R.=1.49 C.I.=0.92-2.41 X^2=2.58 P=0.11	O.R.=1.24 C.I.=0.49-3.15 X^2=0.21 P=0.64	O.R.=2.86 C.I.=0.88-9.23 X^2=3.18 P=0.08	O.R.=1.52 C.I.=0.61-3.76 X^2=0.83 P=0.36	O.R.=1.70 X^2=3.37 P=0.07
		G ALLELE				
		C v. G	**CC v. GC**	**CC v. GG**	**GG+GC v. CC**	**Common odds ratio**
		O.R.=0.67 C.I.=0.42-1.09 X^2=2.58 P=0.11	O.R.=0.44 C.I.=0.17-1.11 X^2=3.11 P=0.08	O.R.=0.35 C.I.=0.11-1.13 X^2=3.18 P=0.08	O.R.=0.42 C.I.=0.17-1.10 X^2=3.65 P=0.06	O.R.=0.59 X^2=3.37 P=0.07
	CRC v. young healthy (unmatched control)	C ALLELE				
		G v. C	**GG v. GC**	**GG v. CC**	**GG v. GC+CC**	**Common odds ratio**
		O.R.=1.23 C.I.=0.81-1.87 X^2=0.97 P=0.33	O.R.=2.17 C.I.=0.97-4.83 X^2=3.65 P=0.06	O.R.=1.75 C.I.=0.72-4.25 X^2=1.54 P=0.21	O.R.=2.01 C.I.=0.94-4.33 X^2=3.29 P=0.07	O.R.=1.29 X^2=1.05 P=0.31
		G ALLELE				
		C v. G	**CC v. GC**	**CC v. GG**	**GG+GC v. CC**	**Common odds ratio**
		O.R.=0.81 C.I.=0.54-1.23 X^2=0.97 P=0.33	O.R.=1.24 C.I.=0.62-2.46 X^2=0.37 P=0.54	O.R.=0.57 C.I.=0.24-1.39 X^2=1.54 P=0.21	O.R.=1.01 C.I.=0.53-1.95 X^2=0.00 P=0.97	O.R.=0.78 X^2=1.05 P=0.31

Table 3. Tests for CRC association of Reprimo 824 G>C SNP.

TESTS FOR ASSOCIATION (95% CONFIDENCE INTERVALS)						
SNP	Populations tested	Allele frequency difference (1 df)	Hetero-zygous	Homo-zygous	Allele positivity	Armitage's trend test
p53R2 4696 C>G	CRC v. DD (matched control)	**C ALLELE**				
		G v. C	GG v. GC	GG v. CC	GG v. GC+CC	Common odds ratio
		O.R. = 0.54 C.I.=0.17-1.67 χ^2=1.19 p=0.27	O.R. = 6.00 C.I.=0.26 – 140.05 χ^2 = 1.47 P=0.23	O.R. = 1.78 C.I.=0.11-29.18 χ^2=0.17 p=0.68	O.R. = 1.96 C.I.=0.12-32.00 χ^2=0.23 p=0.63	O.R. = 0.75 χ^2=1.02 p=0.31
		G ALLELE				
		C v. G	CC v. GC	CC v. GG	GG+GC v. CC	Common odds ratio
		O.R. = 1.87 C.I.=0.60-5.84 χ^2=1.19 p=0.28	O.R.=3.37 C.I.=0.72-15.70 χ^2= 2.64 p=0.10	O.R.=0.56 C.I.=0.03-9.18 χ^2=0.17 p=0.68	O.R. =2.43 C.I.=0.66-8.98 χ^2=1.87 p=0.17	O.R.=1.43 χ^2= 1.02 p=0.31
	CRC v. young healthy (un-matched control)	**C ALLELE**				
		G v. C	GG v. GC	GG v. CC	GG v. GC+CC	Common odds ratio
		O.R.=1.14 C.I.=0.54-2.44 χ^2=0.12 p=0.73	O.R.=2.18 C.I.=0.17-27.56 χ^2= 0.38 p=0.54	O.R.=2.13 C.I.=0.19-23.97 χ^2= 0.39 p=0.53	O.R.=2.14 C.I.=0.19-23.98 χ^2= 0.40 p=0.53	O.R.=1.20 χ^2= 0.10 p=0.75
		G ALLELE				
		C v. G	CC v. GC	CC v. GG	GG+GC v. CC	Common odds ratio
		O.R.=0.88 C.I.=0.41-1.87 χ^2=0.12 p=0.73	O.R.=1.02 C.I.=0.43-2.46 χ^2=0.00 p=0.96	O.R.=0.47 C.I.=0.04-5.28 χ^2=0.39 p=0.53	O.R.=0.94 C.I.=0.41-2.15 χ^2=0.02 p=0.88	O.R.=0.86 χ^2= 0.10 p=0.75

Table 4. Tests for CRC association with p53R2 4696 C>G.

Discussion

This study did not find any statistically significant association between the p53R2 4696C>G or the Reprimo 824 G>C SNPs and CRC. The possibility that the SNPs may influence the age of onset of CRC should be considered, however this aspect was not examined in this study.

An unexpected finding, the Diverticular disease control population was shown to be in deviation from Hardy-Weinberg equilibrium with respect to the Reprimo 824 G>C SNP, with an excess of heterozygotes and lack of both G and C homozygotes. The usual causes of deviation from Hardy-Weinberg equilibrium, inbreeding, genetic drift, small population size and selection pressure [24] cannot explain this finding in a large, randomly mating, static and late onset disease population. Deviation from Hardy-Weinberg may also be due to genotyping errors or population stratification [25], however genotyping was shown to be accurate by independent laboratory sequencing of DNA in all sample populations. If none of these factors are causing the deviation from Hardy-Weinberg equilibrium seen in the Diverticular disease population it may be due to genetic association [26]. The mechanism of this association is obscure, however. Heterozygotes have been shown to alter gene function [27] and survival advantage due to increased heterozygote fitness is termed heterosis, however diverticular disease does not affect biological fitness. However in this study it is the combination of the G and C alleles in the heterozygote that is associated with the diverticular disease phenotype.

Diverticular disease is not generally considered a "genetic disease", however association between Far Eastern populations and right-sided diverticula [28] and diverticulosis in young adults with connective tissue disorders such as Marfan's and Ehlers-Danlos syndromes suggest that molecular pathology may play a role [29, 30]. Similarly, extra cellular matrix alterations, colonic wall elastosis and altered collagen structure in the pathogenesis of diverticular disease [31] may be due to altered matrix metalloproteinases activity [32]. Genetic variants in the human metalloproteinases-2 gene have been discovered and SNPs located within the promoter region of the gene display functional variation [33]. A role for Reprimo in diverticular disease remains subject to further research. However it could be speculated that impaired DNA repair due to augmented Reprimo function could

alter the colonic extracellular matrix favouring diverticulosis formation.

Alternatively Reprimo may have a role in symptom development in Diverticular

disease via its effects on DNA repair and cellular proliferation.

Therefore it may be that diverticular disease has complex gene-environment

aetiology and further studies are required to investigate its molecular pathology.

Conclusions

There are no associations between the p53R2 4649 C>G or Reprimo 824 G>C SNP

SNPs and colorectal cancer demonstrated in this study.

There is an association between the Reprimo 824 G>C heterozygote and

diverticular disease on the basis of deviation from Hardy-Weinberg equilibrium.

Competing interests

The authors declare that there are no competing interests.

Authors' contributions

WB, JB, GJ and JP conceived and designed the study and collected tissue samples.

WB performed DNA extraction, genotyping, statistical analysis and drafted the

manuscript. GJ and JP helped draft the manuscript. All authors read and approved

final manuscript.

Acknowledgements

References

1. Boyle, P. and J.S. Langman, *ABC of colorectal cancer: Epidemiology.* BMJ, 2000. **321**(7264): p. 805-8.
2. Fearon, E.R. and B. Vogelstein, *A genetic model for colorectal tumorigenesis.* Cell, 1990. **61**(5): p. 759-67.
3. Leslie, A., et al., *The colorectal adenoma-carcinoma sequence.* Br J Surg, 2002. **89**(7): p. 845-60.
4. Terdiman, J.P., *Genomic events in the adenoma to carcinoma sequence.* Semin Gastrointest Dis, 2000. **11**(4): p. 194-206.
5. Fukushima, T. and S. Takenoshita, *Colorectal carcinogenesis.* Fukushima J Med Sci, 2001. **47**(1): p. 1-11.
6. Yang, V.W., *The molecular genetics of colorectal cancer.* Curr Gastroenterol Rep, 1999. **1**(5): p. 449-54.
7. Houlston, R.S. and I.P. Tomlinson, *Polymorphisms and colorectal tumor risk.* Gastroenterology, 2001. **121**(2): p. 282-301.
8. Tanaka, H., et al., *A ribonucleotide reductase gene involved in a p53-dependent cell-cycle checkpoint for DNA damage.* Nature, 2000. **404**(6773): p. 42-9.
9. Nakano, K., et al., *A ribonucleotide reductase gene is a transcriptional target of p53 and p73.* Oncogene, 2000. **19**(37): p. 4283-9.
10. Yamaguchi, T., et al., *p53R2-dependent pathway for DNA synthesis in a p53-regulated cell cycle checkpoint.* Cancer Res, 2001. **61**(22): p. 8256-62.
11. Zhou, B., et al., *The human ribonucleotide reductase subunit hRRM2 complements p53R2 in response to UV-induced DNA repair in cells with mutant p53.* Cancer Res, 2003. **63**(20): p. 6583-94.
12. Yanamoto, S., et al., *Expression of p53R2, newly p53 target in oral normal epithelium, epithelial dysplasia and squamous cell carcinoma.* Cancer Lett, 2003. **190**(2): p. 233-43.
13. Ye, Z. and J.M. Parry, *The discovery and confirmation of single nucleotide polymorphisms in the human p53R2 gene by EST database analysis.* Mutagenesis, 2002. **17**(5): p. 361-4.
14. Smeds, J., R. Kumar, and K. Hemminki, *Polymorphic insertion of additional repeat within an area of direct 8 bp tandem repeats in the 5'-untranslated region of the p53R2 gene and cancer risk.* Mutagenesis, 2001. **16**(6): p. 547-50.
15. Ohki, R., et al., *Reprimo, a new candidate mediator of the p53-mediated cell cycle arrest at the G2 phase.* J Biol Chem, 2000. **275**(30): p. 22627-30.
16. Taylor, W.R. and G.R. Stark, *Regulation of the G2/M transition by p53.* Oncogene, 2001. **20**(15): p. 1803-15.
17. Zhang, Z., et al., *p53 transgenic mice are highly susceptible to 1, 2-dimethylhydrazine-induced uterine sarcomas.* Cancer Res, 2002. **62**(11): p. 3024-9.
18. Sato, N., et al., *Discovery of novel targets for aberrant methylation in pancreatic carcinoma using high-throughput microarrays.* Cancer Res, 2003. **63**(13): p. 3735-42.
19. Ye, Z. and J.M. Parry, *Identification of polymorphisms in the human Reprimo gene using public EST data.* Teratog Carcinog Mutagen, 2002. **22**(6): p. 485-93.
20. de Jong, M.M., et al., *Low-penetrance genes and their involvement in colorectal cancer susceptibility.* Cancer Epidemiol Biomarkers Prev, 2002. **11**(11): p. 1332-52.

21. Mort, R., et al., *Lack of involvement of nucleotide excision repair gene polymorphisms in colorectal cancer.* Br J Cancer, 2003. **89**(2): p. 333-7.

22. Dietel, M., *Molecular pathology of colorectal cancer: From phenotype to genotype.* Recent results in cancer research, 1998. **146**: p. 3-19.

23. Navaratnam, R.V., J. Chowaniec, and M.C. Winslet, *The molecular biology of colorectal cancer development and the associated genetic events.* Annals of the Royal College of Surgeons of England, 1999. **81**: p. 312-319.

24. Connor, J.M. and M.A. Ferguson-Smith, *Chapter 11,* in *Essential Medical Genetics.* 1993.

25. Silverman, E.K. and L.J. Palmer, *Case-Control association Studies for the Genetics of Complex Respiratory Diseases.* Am. J. Respir. Cell Mol. Biol, 2000. **22**: p. 645-648.

26. Balding, D.J., M. Bishop, and C. Cannings, *Chapter 19,* in *Handbook of Statistical Genetics.* 2001.

27. Richmond, R.C. and J.R. Powell, *Evidence of heterosis associated with an enzyme locus in a natural population of Drosophila.* Proc Natl Acad Sci U S A, 1970. **67**(3): p. 1264-7.

28. Nakaji, S., et al., *Comparison of etiology of right-sided diverticula in Japan with that of left-sided diverticula in the West.* Int J Colorectal Dis, 2002. **17**(6): p. 365-73.

29. Eliashar, R., et al., *Multiple gastrointestinal complications in Marfan syndrome.* Postgrad Med J, 1998. **74**(874): p. 495-7.

30. Beighton, P., et al., *Ehlers-Danlos syndromes: revised nosology, Villefranche, 1997. Ehlers-Danlos National Foundation (USA) and Ehlers-Danlos Support Group (UK).* Am J Med Genet, 1998. **77**(1): p. 31-7.

31. Ludeman, L., B.F. Warren, and N.A. Shepherd, *The pathology of diverticular disease.* Best Pract Res Clin Gastroenterol, 2002. **16**(4): p. 543-62.

32. Mimura, T., A. Emanuel, and M.A. Kamm, *Pathophysiology of diverticular disease.* Best Pract Res Clin Gastroenterol, 2002. **16**(4): p. 563-76.

33. Price, S.J., D.R. Greaves, and H. Watkins, *Identification of novel, functional genetic variants in the human matrix metalloproteinase-2 gene: role of Sp1 in allele-specific transcriptional regulation.* J Biol Chem, 2001. **276**(10): p. 7549-58.