QTL mapping of Apc modifiers in an ApcMin/+ mouse model of spontaneous and irradiation-induced intestinal adenomas.

Elahi, Eiram.

Queen Mary
University of London

QTL mapping of *Apc* modifiers in an $Apc^{Min/+}$ mouse model of spontaneous and irradiation-induced intestinal adenomas

---

Eiram Elahi

Cancer Research UK funded PhD thesis

Supervisor: Prof Andrew Silver, Professor of Colorectal Cancer Genetics Group

Submitted for the degree of PhD

University of London
2013

TABLE OF CONTENTS

STATEMENT OF ORIGINALITY

The work described in this thesis was performed at St Marks Hospital and the Centre for Academic Surgery, Queen Mary School of Medicine and Dentistry under the supervision of Professor Andrew Silver.

This thesis is my own work. Where work has been carried out with the co-operation of others, or use has been made of their observations, due acknowledgement has been given in the Acknowledgements section or in the relevant parts of the text.

I hereby declare that this thesis entitled "***QTL mapping of Apc modifiers in an Apc$^{Min/+}$ mouse model of spontaneous and irradiation induced intestinal adenomas***" has not been submitted for a degree or diploma, or any other qualification, at any other university. I further state that no part of this thesis has already been or is currently being submitted for any such qualification.

Eiram Elahi

Singapore

e.elahi@qmul.ac.uk

ACKNOWLEDGEMENT

I want to acknowledge my supervisor Professor Andrew Silver for his understanding throughout the several stages of my research. I am truly grateful for this. He provided me with a listening ear and support during my personal and professional trials. I am grateful for his guidance and acceptance of the work carried out. I also recognise that this research would not have been possible without the financial support of Cancer Research UK.

I want to express my deep gratitude to Dr Nirosha Suraweera who has been instrumental in the direction and success of this project. I have always said 'there is magic in her methods' and I am thankful for the time she spent teaching me these. I would like to thank Mr Kevin Pack, Mr Emmanouil Volikos and Miss Victoria Johnson; 'the old St Marks crew' before the imminent relocation to Queen Mary University of London. The occasional Friday evening with colleagues at St Marks Social Club (aka 'So-sh') will always remain as fond memories.

I would also like to thank the following members of Colorectal Cancer Research Group: Dr Amy McCart, Dr Nikki Vicryakous, Mrs Guadalupe Echevry-Polanco, Dr Sarah McDonald, Dr Tammie MacFie, Dr James Robinson and Miss Alexandra Parker.

Miss Meriem Galima, my much loved and valued helper in Singapore, without whom I would not have been able to juggle thesis writing and bringing up a young family. Her help and patience have helped me the most. Thank you.

I thank my parents Fatima and Mohd. Sardar for giving me the seeds of experience. I appreciate how tough it has been for them migrating to a new country very young and with very little. They have managed to give me and my 3 brothers everything, most of all their love, concern and support. To my brothers; Imran, Faisal and Omar, I am grateful for the endless brotherly advice, humour and encouragement they have given me throughout my life.

My wonderful husband Dinesh who turned my world around, I thank him for his encouragement, love and providing me with everything required to complete my work. Without him by my side I would have struggled at the start, middle and end. My love and thanks go to him.

Last but not least, my lovely children; Fayha (6 years old) and Hami (10 months) who have kept mummy sane through this process. Fayha who first thought my work was on Minnie mouse has developed a huge interest in this PhD. Her scientific mind is brilliant and has started a PhD of her own titled; 'the good and bad virus which infected a deer'. She has become familiar with many scientific words and her comments never cease to amaze me. I will always remember when she first began to read and looked at my screen and said "mummy, there are so many made up words in your PhD, like this one" and pointed at 'polyps'.

ABSTRACT

BACKGROUND: Radiation exposure to the abdominal region causes intestinal toxicity and is also capable of inducing colorectal cancers (CRC). Genotype-phenotype studies provide some evidence explaining the variation in familial adenomatous polyposis (FAP) patients caused by modifiers of adenomatous polyposis coli (*APC*). This study aims to extend our understanding of irradiation-induced modifiers of $Apc^{Min/+}$ mice and CRC.

METHODS: By using a pre-existing backcross between recombinant inbred line of $Apc^{Min/+}$ mice to the irradiation sensitive inbred BALB/c mouse, we obtained panels of 2Gy-irradiated and sham-irradiated N2 $Apc^{Min/+}$ mice for genotyping with a genome-wide panel of microsatellites markers. Using the number of adenomas in different intestinal segments to represent polyp multiplicity, we carried out a genome wide quantitative trait loci (QTL) scan followed by statistical epistasis modelling and bioinformatics analysis.

RESULTS: We identified five significant QTLs responsible for radiation induced tumour multiplicity in the upper small intestine defined as *Mom* (*Modifier of Min*) radiation-induced polyposis (*Mrip1-5*) on chromosome 2 (LOD 2.8, p = 0.0003), two regions within chromosome 5 (LOD 5.2, p=0.00001, 6.2, p=0.00001) and two regions within chromosome 16 (LOD 4.1, $p=4x10^{-5}$ and 4.8, p=0.00001). Suggestive QTLs were found for sham-irradiated mice on chromosomes 3, 6 and 13 (LOD 1.7, 1.5 and 2.0 respectively; p,0.005). Two significant QTLs were detected in the 2large intestine on chromosome 2 and 7 (LOD 2.7, $p=1.2x10^{-3}$ and 2.2, $p=1.2x10^{-3}$,

respectively). Using statistical epistasis modelling and logical selection of target genes though *in silico* sequence based on BALB/c specific non-synonymous polymorphisms which are predicted deleterious we selected target genes and further eliminated genes by sequencing and mRNA expression.

CONCLUSIONS: Our study locates the QTL regions responsible for increased radiation-induced intestinal tumorigenesis in $Apc^{Min/+}$ mice and identifies candidate genes with predicted functional polymorphisms that are involved in spindle checkpoint and chromosomal stability (*Bub1b, Bub1r,* and *Casc5*), Wnt pathway (*Tiam1, Rac1*), DNA repair (*Recc1* and *Prkdc*) and inflammation (*Duox2, Itgb2l* and *Cxcl5*).

PUBLICATIONS

Published (Chapter 2 and 4)

Elahi E, Suraweera N, Volikos E, Haines J, Brown N, Davidson G, Churchman M, Ilyas M, Tomlinson I, Silver A. **Five quantitative trait loci control radiation-induced adenoma multiplicity in Mom1R Apc Min/+ mice**. PLoS One. 2009;4(2):e4388.

Submitted to BMC Genomics 2013 (Chapter 3 and 4)

Eiram Elahi, Nirosha Suraweera, Ian Tomlinson, Andrew Silver. **Epistasis modifies adenomatous polyposis in an *Apc$^{Min/+}$* mouse model.**

LIST OF ABBREVIATIONS

AC	Adenoma carcinomas
ACF	Aberrant crypt formation
AFAP	Attenuated FAP
ANOVA	Analysis of variance
APC	Adenomatous polyposis coli
ASPs	Affected sib pairs
BALB/c	BALB/cByJ - Albino coat, inbred strain of mouse
B6	C57BL/6 – Black coat, inbred strain of mouse
B6$^{Min/+}$	C57BL/6 (Carrying the Min mutation in *Apc*)
BC	Backcross
BER	Base excision repair
CBC	Crypt base columnar cells
CD	Crohn's disease
CHRPE	Congenital hypertrophy of the retinal pigment epithelium
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
CpG	Cytosine-Phosphate-Guanine
CRC	Colorectal cancer
CSC	Cancer stem cells
CT	Computer tomography
cT	Cycle threshold
Cyclin D1	CCND1
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferases
DSB	Double strand break
DSH	Dishevelled (phosphoprotein)
DSPs	Discordant sib pair
eQTL	Expression QTL
FAP	Familial adenomatous polyposis
FRZ	Frizzled(protein receptor for Wnt)
GEMM	Genetically engineered mouse model
GI	Gastrointestinal
GWAS	Genome wide association study
Gy	Grays
HMM	Hidden Markov Model
HR	Homologous repair
H_2O_2	Hydrogen peroxide
IBD	Identify by descent
IBD	Inflammatory bowel disease
IM	Interval mapping
JPS	Juvenile polyposis syndrome

KRAS	Kirsten rat sarcoma
LI	Large intestine
LOD	Logarithm of Odds
LOH	Loss of heterozygosity
LRS	Likelihood ratio statistic
LS	Lynch Syndrome
LSI	Lower small intestine
MACS	Microsatellite and chromosome stable
MAP	MYH-associated polyposis
MCR	Mutation cluster region
Min	Multiple intestinal neoplasia
MMR	Mismatch repair
Mom	Modifiers of Min
Mom1^{R}	Mom1 resistant
Mom1^{S}	Mom1 susceptible
MQM	Multiple QTL mapping
Mrip	*Mom* radiation-induced susceptibility
mRNA	Messenger RNA
MSI	Microsatellite instability
MT	Microtubules
NADPD	Nicotinamide adenine dinucleotide phosphate
NER	Nucleotide excision repair
NSAIDS	Non-steroidal anti-inflammatory drugs
NHEJ	Non-homologous end joining
NOX	NADPH oxidase
PBS	Phosphate buffer solution
PCR	Polymerase chain reaction
PJS	Peutz-Jeghers syndrome
PSIC	Position specific independent count
PTEN	Phosphatase and tensin homologue
QTL	Quantitative trait loci
qRT-PCR	Quantitatie reverse transcription (PCR)
RC	Rectal cancer
RER^{+}	Replication error positive
RF	Recombination fraction
RFLP	Restriction fragment length polymorphism
RI	Recombinant inbred
RNA	Ribonucleic acid
ROS	Reactive oxygen species
SEM	Standard error of the mean
SNPs	Single nucleotide polymorphisms
SSLP	simple sequence length polymorphism
TDT	Transmission disequilibrium test
TSG	Tumour suppressor gene

UC	Ulcerative colitis
USI	Upper small intestine

LIST OF GENES

ALDH1	Aldehyde dehydrogenase 1 family, member A1
Apc	Adenomatous polyposis coli
ATF2	Activating transcription factor 2
ATM	Ataxia telangiectasia mutated
β-catenin	Beta-catenin
BAX	BCL2-associated X protein
Bmi1	BMI1 polycomb ring finger oncogene
BMP	Bone morphogenic protein
BMPR1A	Bone morphogenic protein receptor, type 1A
BRAF	v-raf murine sarcoma viral oncogene homolog B1
Bub1	Budding uninhibited by benzimidazoles 1 homolog
Bubr1	Budding uninhibited by benzimidazoles 1 homolog, beta
Casc5	Cancer susceptibility candidate 5 (mouse)
CD133	Prominin 1
CD166	ALCAM, activated leukocyte cell adhesion molecule
CD29	ITGB1, integrin, beta 1
CD44	CD44 molecule
Arap2	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 2
c-MYC	Myelocytomatosis viral oncogene homolog
CHK1	Checkpoint homologs 1
CHK1	Checkpoint homologs 2
COX-2	Cyclooxygenase-2
Cxcl5	Chemokine (C-X-C motif) ligand 5 (mouse)
DCC	Deleted in colorectal carcinoma
Dll4	Delta-like 4 (mouse)
Duox2	Dual oxidase 2
DUOXA2	Dual oxidase maturation factor 2
EpCAM	Epithelial cell adhesion molecule
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
Gpx	glutathione peroxidase
Gdac1	glutathione peroxidase deficiency-associated colitis 1
GSK3β	Glycogen synthase kinase 3β
IFNγ	Interferon gamma
Itgb2l	integrin beta 2-like
KRAS	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
LEF1	Lymphoid enhancer factor-1
LGR5	Leucine-rich repeat containing G protein-coupled receptor 5
LKB1	Liver kinase B1
MAPK	Mitogen-activated protein kinase
Mdm2	Murine double minute 2
Mertk	C-mer proto-oncogene tyrosine kinase (mouse)
Med15	Mediator complex subunit 15
MMP7	Metalloproteinase 7
MLH1	MutL homolog 1
MSH2	MutS homolog 2
MSH6	MutS homolog 6

MTHFR	Methylenetetrahydrofolate reductase
MYH	MutY homolog
NFκB	Nuclear factor Kappa- Beta light chain
PGE2	Prostaglandin E2
Pla2g2a	Secretory type II non-pancreatic phospholipase A2
PMS2	Post meiotic segregation increased 2
PPARG	Peroxisome proliferator-activated receptor gamma
PRKDC	Protein kinase, DNA-activated, catalytic polypeptide
Rac1	Ras-related C3 botulinum toxin
RASSF1A	RAS association domain family 1
Rcf11	Replication factor C (activator 1) 1
Rb	Retinoblastoma
Recc1	Replication factor C (activator 1) 1 (mouse)
RUNX3	Runt-related transcription factor 3
SMAD2	SMAD family member 2
SMAD4	SMAD family member 4
STK11	Serine/threonine kinase 11
TGFBR1	Transforming growth factor, beta receptor I
TGFBR2	Transforming growth factor, beta receptor II
TGFβ	Transforming growth factor-beta
Tiam1	T lymphoma invasion and metastasis 1
TLR	Toll-like receptor 5
TCF4	T-cell factor -4
TNF- α	Tumour necrosis factor-α
TP53	Tumour protein 53
VEGF	Vascular endothelial growth factor
VHL	von Hippel-Lindau
WNT	Wingless-type MMTV integration site family

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1 COLORECTAL CANCER

1.1.1 Incidence of colorectal cancer Worldwide, Europe and in the UK

Colorectal cancer (CRC), which affects the areas of the colon and rectum, is the third most commonly diagnosed cancer worldwide, affecting men and women fairly equally with a ratio of 1.2:1 (Parkin *et al*., 2005). The incidence rate is just over 1 million cases per year with a mortality rate of approximately 50% making CRC the second leading cause of cancer mortality in the Western world (Parkin *et al*., 2005; Jemal *et al*., 2009; Herszényi and Tulassay, 2010). In Europe, the incidence of CRC is increasing, particularly in the Southern and Eastern European regions with 334,000 new CRC cases reported in 2008, accounting for 12.9% of all European cancer cases. Cancer Research UK reported just under 40,000 new cases of CRC were diagnosed and approximately 16,000 deaths reported in 2008 (Cancerstats Incidence, 2008).

Parkin *et al* reported at least a 25-fold variation in global CRC incidence rates, with the highest rates observed in North America, Western Europe, Australia, New Zealand, and Japan, intermediate in South America, and low in Africa and South and Central Asia (Parkin *et al.,* 2005). Stirbu *et al* show that immigrants from low risk to high risk regions and their descendents reach the cancer incidence rates of their host country (Stirbu *et al*., 2006) underscoring the importance of environmental and dietary differences.

1.1.2 Cancer survival

While the prevalence of CRC has steadily increased the mortality rates have declined as a result of efficient screening and early detection. About 60% of all patients with cancer will receive radiotherapy during the course of their disease as part of cancer treatment and survival (Halperin *et al*., 2009). The 5-year relative survival rates from the US Surveillance, Epidemiology and End Results (SEER9) database, shows an overall increase of 8.3% from patients diagnosed in 1980 to those diagnosed in 2004 (Howlader *et al*., 2009). Recent estimates from Maddams and colleagues show that there are approximately 2 million cancer survivors (59% women and 41% men) in the UK (recorded to the end of 2008) of which 13% are aged 65 years and above (Maddams *et al*., 2009, Maddams *et al*., 2012). 'Cancer 'survivors' were defined as 'living individuals who were diagnosed with cancer at some point in their life' and distinctions of whether these patients are 'having active disease', 'in remission' or 'cured of their cancer' were not possible. These conservative estimates show that the number of long term cancer survivors in the general population of the UK is substantial and increasing.

1.1.3 Cancer burden in the UK following medical radiation exposure

Radiotherapy targets clonogenic cancer cells however there is an increased risk of damage to healthy, non neoplastic tissue and the therapeutic window between tumour cure and normal tissue damage is thought to be quite narrow resulting in radiotherapy exposed cancer survivors experiencing adverse side effects which include secondary malignancies (Hall and Wu., 2003; Suit *et al*., 2007; Brenner *et al*.,

2003; Hall, 2006). Maddams *et al* provided the first estimate for the burden of secondary cancers caused by radiotherapy. She calculated an overall estimate of 1,346 cases (0.45%) of the 298,000 new cancers registered in the UK in 2007, to be attributable to radiation therapy for a previous cancer. The largest numbers of radiotherapy-induced secondary cancers were observed in the lung (23.7% of the total), oesophageal (13.3%), and then female breast tissue (10.6%). 15% of secondary cancers were associated with radiotherapy among survivors of Hodgkin's disease and oral and pharyngeal tumours (Maddams *et al*., 2011). Further evidence for the risk of radiotherapy comes from some 6,000 UK patients a year who are affected by toxicity and inflammatory responses in the gut, which are well known risk factors for CRC (Andreyev *et al*., 2005; Andreyev 2007).

In addition, a sharp increase in the use of computer tomography (CT) scans is observed since the 1980s, with a total of three million scans performed in 2008 in the UK (Brenner *et al*., 2010). It is believed that 35mSv (equivalent to four abdominal CT scans) is the lowest dose of irradiation absorbed by human tissue to possibly cause cancer (Sountoulides *et al*., 2010). The lifetime risk of cancer incidence as a result of CT colonography scans performed on individuals between 55–70 years every 5 years is likely to affect 2.6 males and 1.9 females in 1000 from a recent press release by the Health Protection Agency (Public Health England, August 2011). Therefore, the collective radiation dose from CT scans and other diagnostic tools probably accounts for a large proportion of the total amount of medical radiation exposure to the adult population of UK which presents a considerable health risk (Frush *et al*., 2008; Pearce *et al*., 2012).

At present, no account is taken of an individual's genetic background and susceptibility to radiation-induced CRC. A major uncertainty is the prevalence and impact of susceptibility loci involved in heritable cancer disorders which may influence or modify radiation exposed damage. Furthermore, nothing is known about IR damage incidence in different ethnic communities or social classes. Identifying these genes would allow pre-screening advantage to distinguish risk prior to medical or occupational irradiation exposure. Thus the study of the genes involved in the molecular mechanisms of CRC not only has the potential of elucidating the pathways and interactions which operate within a CRC framework, but also of identifying candidates for irradiation induced intestinal toxicity.

1.1.5 Intestinal early and late radiation injury

The first case of radiation induced damage to GI was reported by Walsh in 1897 just two years after X-rays were discovered by Wilhelm Roentgen in 1895, when a man developed diarrhoea after several weeks of occupational X-ray irradiation exposure. Walsh concluded that X-rays caused inflammation of the intestinal mucus membrane (Walsh 1897). Radiation effects on normal intestinal tissues can be divided broadly into early or acute, sub-acute and late reactions. Symptoms of early effects and intestinal dysfunction occur between 60–90% patients and can last as long as two years in cancer patients receiving abdominal radiotherapy (Yeoh *et al*., 1995). Furthermore, studies show there is a higher (8.2%) occurrence of late complications developing from patients who initially experience early complications compared to the 3% that develop only late complications (Bourne *et al*., 1983). In general, early damage occurs within 24 hours and is not dose-dependent but

sensitive to the time (dose-rate) of exposure, typically ~0.1Gray (Gy)/hour is sufficient to cause radiation damage.

1.1.6 Radiation induced colitis

Radiation induced colitis and related proctitis involve the inflammation and damage to the lower parts of small intestine, sigmoid colon and rectum. It is a progressive disease and depending on the sensitivity of the gut and the dose of radiation the extent of the injury varies, and is therefore difficult to predict in patients (Cho *et al*., 1995). Complications may develop within a few months with late effects manifesting up to 5 years after radiation. A small number of patients will develop chronic radiation colitis of which 50% may require surgical intervention, which has a high complication rate and death rate of 10.7%, often due to the poor condition of the patient (Kennedy and Heise, 2007). More chronic forms of colitis include the development of fibrosis, edema, perforations, intestinal obstructions, and precancerous lesions which may manifest as deep ulcers or masses and eventually into adenocarcinomas.

It appears that the sequential loss of epithelial cell mitosis following radiation exposure to the intestine ablates the ability to produce enough cells to repopulate the villi, this results in the denudation of the mucosal surface and diminution in villus height followed by blood vessel damage, depletion in nutrient absorption, loss of fluid, electrolytes and oxygen supply and the overall impact is the loss of mucosal protective barrier. The gut is left permeable to xenobiotics, bacteria and other antigens leading to an inflammatory response (Figure 1-1).

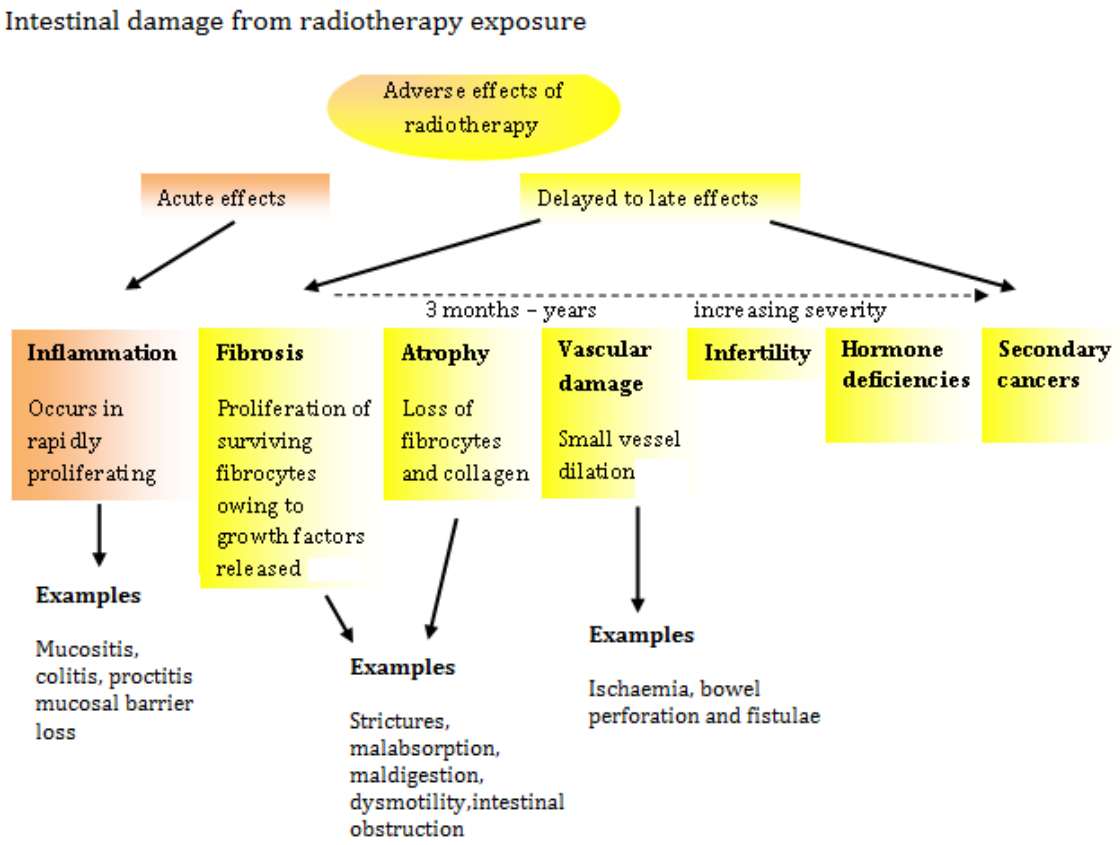

Figure 1-1 The biphasic early and acute late effects of radiation toxicity with examples of intestinal radiation injury. Damage to the stromal tissue affects the vasculature and connective cells and reactions mediated by vascular deficiencies such as the fibrosis, malabsorption, strictures, obstruction, perforation and carcinogenesis.

1.1.7 Radiation-induced secondary CRCs

There is significant evidence for the increase in bladder cancer and to a lesser extent of CRC, following prostatic radiotherapy probably due to the direct radiation and some low scatter emissions (Baxter *et al.*, 2005, Rapiti *et al.*, 2008). Baxter reported that long term prostate survivors who received radiotherapy had an increased risk of rectal cancer (RC) (a hazard ratio of 1.7) similar to having a first degree relative with CRC (Baxter *et al.*, 2005). Injury to the rectum is more common possibly owed

to its fixed position in the pelvis and is rendered more susceptible to radiation injury. A retrospective study by Yang and colleagues (2008) identified 5 patients who had developed CRC following successful treatment for their primary cancers; cervix (n=3), bladder (n=1) and one for rectal with radiotherapy (Yang *et al*., 2008). More, recently, a study looked at the age standardised incidence ratios (SIR) of RC in patients diagnosed with prostate cancer between 1982–2005 comparing groups treated with irradiation or surgery. A significant increased risk was observed in the development of RC with the radiotherapy group (SIR 1.8) compared to those managed by surgery (SIR 1.21) (Margel *et al*., 2011). On the contrary, Kendal and colleagues reported that infrequent incidence of RC was balanced by the incidence of spontaneous RC (Kendal and Nicholas, 2007).

These data suggest early intestinal injury is an almost inevitable consequence following abdominal radiation therapy with an additional and significant risk of patients showing late effects. There appears to be sufficient grounds to assume that cancer predisposed individuals may be at a higher risk for irradiation induced intestinal toxicity and or secondary CRCs. It is estimated that more than 1 million patients in the US have GI related dysfunction related to radiotherapy, similarly, a significant proportion of patients in the UK suffer from some form of intestinal injury following pelvic radiotherapy. The incidence is likely to continue and even increase and so there is a growing need for research behind CRC and irradiation..

1.1.5 Biological effects of irradiation

Tissue damage caused by irradiation involves direct or indirect absorption of high transfer energy within cells. It has been proposed that the radiation-induced effects are caused, in part, by the induction of free radicals and reactive oxygen species (ROS) via chronic oxidative stress (Hernandez-Flores *et al*;, 2005) and inflammation (Matsunami *et al*., 1999). Though the mechanisms have not been fully elucidated, it is generally accepted that an increase in ROS leads to the lipid peroxidation, oxidation of DNA and proteins as well as inactivation of pro-inflammatory factors (Poli *et al*., 2008).

Cells usually show alterations in the genetic processes and favour abnormal cellular growth and proliferation via mutational effects. These processes affect double strand breaks (DSBs), activation of intracellular signalling pathways, DNA repair mechanisms and genomic instability involving genetic and epigenetic processes between the initiated cells and the microenvironment (Rothkamm *et al*., 2001; Vilenchik and Knudson 2003; Harper and Elledge, 2007). DNA repair activities are monitored and processed by five different pathways: homologous repair (HR); non-homologous end joining (NHEJ); nucleotide excision repair (NER); base excision repair (BER); and mismatch repair (MMR) (Hoeijmakers 2001). Defects in DNA repair pathways are associated with higher risk of cancer (Mohrenweiser *et al*., 2003). Repair of DSBs is predominantly carried out via the NHEJ pathway via DNA-dependent protein kinase, catalytic subunit (DNA-PKcs) belonging to the phosphatidylinositol 3-kinase-related kinase (PIKK) protein family encoded by *PRKDC* and function in cell cycle. Mutations in any of its subunits that form a

hetreodimer complex with Ku70/80 result in radiosensitivity as displayed by DNA-PKcs null mice with signs of hypersplasia, dysplasia and aberrant crypt formation (ACF) (Bailey *et al*., 1999).

1.1.8 Radiotherapy risk factors

Factors such as inflammation, poor nutrition, previous operations and presence of infections are factors known to exacerbate radiation induced injury, however, dose, rate, volume and mode of delivery remain the major risk determinants involved in the induction of early to late intestinal toxicity. Intestinal radiation sequelae is more frequent in patients (11.3%) treated with a beam split into two sagittal halves when compared with a three or four field beam (2.3%) (Cerrotta *et al*., 1995) and tissue toxicity increases significantly when the dose exceeds 45 Gys, with the incidence of marked rectal changes rising from 8-51% compared to 33% changes in asymptomatic prostate cancer patients (Sugimura *et al*., 1990). A clear dose-response relationship for radiation associated malignancies is difficult to establish and tends to vary from organ to organ, type of tumour, age and the genetic variability in irradiation response. Therefore, there is no upper or lower threshold that can be applied to radiation induced toxicity and or cancer induction, making it one of the most difficult carcinogens to study.

1.2 COLORECTAL CANCER RISK FACTORS

1.2.1 Genetics and family history

Genetics and family history are the most important risk factors of CRC, and play a key role in predisposition accounting for ~25% of all CRC patients. Despite this relatively large estimate, mutations in single high penetrance genes account for only 5-6% of all CRC (Table 1-1) (Lichtenstein *et al*., 2000; de la Chapelle, 2004).

Table 1-1 Heritable colorectal cancer syndromes and lifetime risk of CRC development. Familial adenomatous polyposis (FAP), Attenuated FAP (AFAP), Lynch Syndrome (LS), Aberrant crypt foci (ACF), Adenomatous polyposis coli (*APC)*, Mismatch repair (MMR), MYH associated polyposis (MAP)

Syndrome	Clinical presentation and notes	High risk gene(s)	Absolute lifetime risk of CRC development
FAP	100-1000 colonic polyps and ACF – polyposis onset age of Age of polyposis onset 16, age of CRC onset 39	*APC*	100% by age 45
Attenuated FAP	<100 polyps proximal colon Difficult to distinguish clinically from LS Age of polyposis onset 36, age of CRC onset, 54	*APC*	69% by age 80 (Neklason *et al*., 2004)
LS	Clinical features encompassed by the Amsterdam Criteria	*MMR genes (MLH1 and MSH2, MSH6, PMS2)*	80% by age 75 (Vasen *et al*., 1999)
MAP	Multiple colonic polyps. Indistinguishable from FAP and AFAP with hyperplastic polyps. Predisposition to duodenal adenomas. Autorecessive disease	*MYH > APC*	80%, CRC often presented same time diagnosis usually by 60 years of age (Farrington *et al*., 2005)
Peutz-Jeghers	Small bowel multiple harmatomatous polyps. Mucocutaneous pigmentation	*STK11*	39% by age 64 (Giardiello *et al*., 2000)

Juvenile polyposis	Self-limiting benign colonic and stomach polyps. Rare and early onset	*SMAD4/PTEN*	17-68% by age 60 (Desai *et al*., 1995)

The remaining genetic burden is therefore attributable to a complex interplay between low penetrance alleles (Webb and Houlston, 2007), environmental variability and genetic modifiers (Figure 1-2) (Houlston and Tomlinson 2000). Modifiers alter disease severity by guiding the genetic and phenotypic consequence through direct or indirect interactions with the primary mutation and are supported by the varying degrees of clinical outcome as observed in human familial adenomatous polyposis (FAP) and Lynch Syndrome (LS) patients (de la Chapelle, 2004).

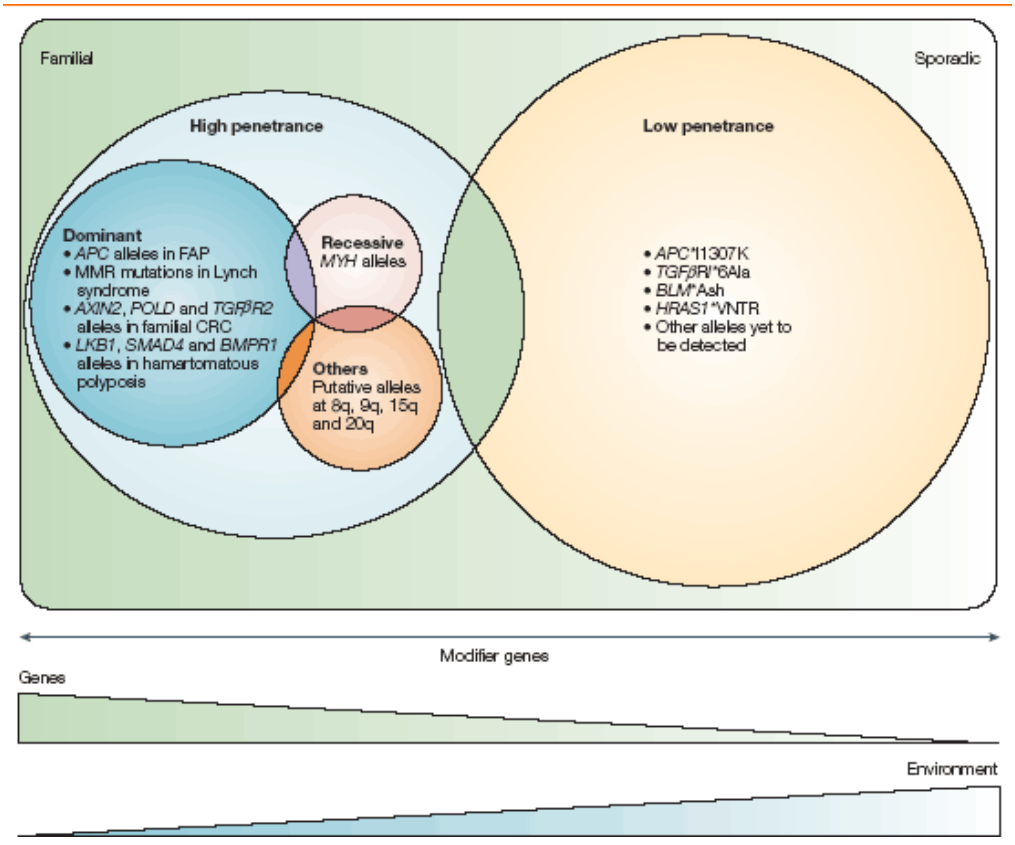

Figure 1-2 The influence of genetics and environmental factors in inherited and sporadic CRCs (de la Chapelle, 2004).

One of the earliest family based studies of CRC and familial inheritance was conducted in Utah families. It was observed that the occurrence of polyps was influenced by genetic factors with a dominant mode of inheritance (Woolf, 1958). It was also reported, that a higher number of deaths from CRC (3.9%) occurred within first degree relatives of CRC patients, compared with sex and age matched controls (1.2%). This difference has since been replicated in numerous studies of various designs, (sample sizes, data validation, analytic methods and countries of origin) and the magnitude of CRC risk has been consistently found to be two to three-fold higher amongst first degree relatives of CRC affected cases. This risk increases with the number of affected family members from 25% to 52% in second and third degree relatives (Slattery and Kerber, 1994) and more so if CRC is diagnosed at a younger age (<50 years) (Fuchs *et al*., 1994). The relative risks from a systematic review and meta-analysis of familial CRC risk was reported by Johns *et al*, in a population study and summarised in Table 1-2 below.

Table 1-2 Estimated relative risk of developing CRC from a positive family history, calculated using weighted average pooled from 27 case control and cohort studies (Johns *et al*., 2001).

Relative risk of CRC and family history	
No family history	1
One first degree relative with CRC	2.3
More than one first degree relative with CRC	4.3
Family history of 1st degree relatives affected < 45 years of age	3.9
Family history of 1st degree relatives affected between 45-59	2.3
Family history of 1st degree relatives affected > 60 years of age	1.8

1.2.2 Inflammatory bowel disease

As mentioned earlier, radiation injury from post pelvic radiotherapy leads to inflammation of the gastrointestinal tract a well known precursor of CRC (Balkwill and Mantovani, 2001; Folkman, 2002). The relationship between inflammation and cancer has been known since 1925, when it was first described by Burrill Crohn (Crohn 1925). Environmental factors (such as geography, cigarette smoking, sanitation and hygiene) are known to influence the ecosystem of the intestinal mucosal microflora and increase the occurrence of intestinal inflammation (Guarner 1998).

Inflammatory bowel disease (IBD) which includes Crohn's disease (CD) and ulcerative colitis (UC) are idiopathic conditions of prolonged inflammation of the colon. Patients with IBD have an inordinately high risk of severe intestinal toxicity following abdominal irradiation (Willet *et al*., 2000) with shared pathological features, such as the activation of mucosal cytokines, crypt distortion, and inflammation of crypts and cells of the lamina propria. CRC accounts for 15% deaths in patients with IBD (Munkholm 2003; Vagefi and Longo, 2005) that increases by 0.5% to 2% annually, 8 to 10 years after IBD is diagnosed (Herszényi *et al*., 2007, Ahmadi *et al*., 2009). IBD CRCs arise from flat dysplastic tissue or dysplasia-associated lesions or masses (DALMs) leading to intestinal lesions and the risk is higher for patients with childhood onset of IBD than those who develop the disease during adulthood (Itzkowitz and Harpaz, 2004). Furthermore, genetic susceptibility to IBD was demonstrated in three European twin studies which showed an increased concordance rate of 20 to 50% in monozygotic twin compared to 10%

concordance rates in dizygotic twins (Tysk *et al*., 1988, Orholm *et al*., 2000, Halfvarson *et al*., 2003).

1.2.3 Ulcerative colitis

UC was rediscovered by Sir Samuel Wilks and Moxon in 1875, when it became clear that UC was different from other known large bowel (ulcerating) diseases which tended to occur from infectious agents such as dysentery and tuberculosis. UC is restricted to the large bowel and characterised by severe ulcerations, predominantly located in the rectum (90% of the patients) and recto-sigmoid (40%), but sometimes affecting the entire colon (20%). The peak incidence of UC is detected between 15-30 years of age, affecting females more than males (Hanauer, 1996). A 20-fold higher risk is associated with CRC in patients with UC which increases 1-2% per year after 10 years affected. The length of time with UC, patient's age and extent of disease at diagnosis are strong independent risk factors for CRC in UC (Lashner 1994).

In the acute phase of inflammation, the mucosa typically becomes hyperaemic and friable and in most cases small mucosal haemorrhages and ulcers coalesce leading to irregular large ulcers. In the more serious cases parts of the mucosa may disappear and be replaced by granulation tissue which, in a later stage, may be covered with new epithelium, that may protrude into the lumen giving a “cobblestone” appearance (pseudopolyps). The mucosa bleeds easily giving rise to the main characteristic of UC; of blood with or without faeces. Particularly important in active colitis are the epithelial changes that appear in the margins of ulcers leading to the

depletion of the mucus, inflammatory metaplasia and atypical dysplasia which is classified as low to high grade progression towards carcinoma. Colitis-associated cancer has been investigated in mouse models which have highlighted the role of toll-like receptors (*Tlr*) and tumour necrosis factor-α (*TNF-α*) in the activation of nuclear factor κB (*NFκB*), which induces transcription of genes involved in tumourigenesis, including cyclooxygenase-2 (*Cox2*) (Westbrook *et al*., 2010) and defective tumour protein 53 (*Tp53*) signalling may be an early event in the progression of colitis induced dysplasia to cancer (Dirisina *et al*., 2011).

1.2.4 Crohn's disease

In 1932 Crohn's and associates defined a pathological condition 'regional ileitis' which today is more commonly known as CD and may involve any portion of the GI tract from mouth to anus, but most often the small intestine and colon are affected (Triantafillidis *et al*., 2009). Though CD can occur at any stage of life, it is most commonly seen in patients between 15 – 30 years of age. A higher occurrence is detected in smokers regardless of sex and in people of Ashkenazi heritage (Seksik *et al*., 2009, Podolsky, 2002). As in the UC, a similar 20-fold higher risk of CRC are observed in CD patients compared to the general public and the risk in CD is also related to the amount of colon involved (Gillen *et al*., 1994). There is an increased prevalence of IBD among first degree relatives of patients suffering from UC or CD (Orholm *et al*., 1991)

In contrast to UC, in CD the inflammation extends through the entire wall of the intestine into the mesentery. The affected colon shows mucosal inflammation, with

cell population of neutrophils, along with mononuclear cells, that may infiltrate the crypts, leading to inflammation or abscess. Non-caseating granulomas, which are aggregates of macrophage derivatives, are found in 50% to 70% of cases and are specific for CD, indicating that they are not infectious. The mucosa may give rise to numerous linear ulcers creating deep fissures which are separated by nodules, creating a "cobblestone" appearance. More chronic mucosal damage is caused by atypical branching of the crypts, blunting of the intestinal villi and a change in the tissue type.

1.2.5 Diet and smoking

As already mentioned, dietary nutrients and carcinogens appear to modulate the biology of the intestine via the colonic mucus barrier within the colonic lumen. Physiological factors which determine the transition rate and the continuous exposure of dietary carcinogens to the lumen gives rise to epithelial cell abnormalities and apoptosis. There is some evidence that DNA methylation can be influenced by the availability of methyl donors, such as folate and methionine from a fibre rich diet and therefore reduce the risk of CRC (Kim and Mason, 1996). Folate and methionine levels are controlled by methylenetetrahydrofolate reductase (*MTHFR*) and methionine synthetase enzymes. The most common polymorphism in the *MTHFR* gene, is Ala677Val (MTHFR C677T) which codes for reduced enzymatic activity. Homozygotes for this variant have about 30% of normal enzyme activity and are associated with a higher risk of CRC (Hubner and Houlston 2007; Zhong *et al*., 2012). Among micronutrients, red meat fat and processed meats have been implicated in the causation of CRC. Recent meta-analysis have shown relative risks

below 1.50 and concluded that red meat and CRC are not statistically significant (Alexander *et al*., 2011). However, the different study designs have led to different outcomes and it appears that the relationship between nutrient intake and human cancer risk is difficult to quantify, with aspects between diet and other risk factors being difficult to discern and isolate as independent factors.

A causal relationship between smoking and CRC has been investigated and though not clearly demonstrated, a significant association between smoking and CRC in a Newfoundland and Labrador population was reported. Both former and current smokers had an increased risk of CRC (odds ratio of 1.36 and 1.9, respectively) compared to non-smokers (Zhao *et al*., 2010). However, the association with CRC varies with the number of years smoked, age of initiation and tumour site. Other studies have shown that smokers have twice the risk of developing colon tumours. About a fifth of tumours (21%) showed microsatellite instability (MSI), which reflects an impairment of the cells ability to repair DNA damage (Slattery *et al*., 2000). In the case of CRC more research is needed to determine whether these associations are causal and if by avoiding certain risk behaviours or increasing protective behaviours would actually reduce cancer incidence.

1.2.6 Drugs, vitamins and minerals

Non-steroidal anti-inflammatory drugs (NSAIDS) and aspirin are thought to have chemopreventative benefits against the development of most cancers and provide relief from the symptoms of inflammation and pain. Aspirin taken daily for 4 or more years is associated with an 18% reduction in overall cancer deaths, and if

taken for several years at 75 mg can be effective in reducing long-term incidence and mortality to CRC, with benefit being greatest for cancers of the proximal colon (Rothwell *et al*., 2013). The mechanism is thought to involve interventions by drugs such as sulindac, celecoxib which inhibit adenoma growth by the suppression of *COX-2* (Chell *et al*., 2005; Markowitz, 2007; Rothwell *et al*., 2013). *COX-2* is involved in the synthesis of proinflammatory prostaglandins with an extensive role in inflammation; prostaglandin E_2 (PGE_2), the product of COX2, is a target for treatment to prevent inflammation-associated cancers. Evidence for the role of *Cox-2* in CRC came from polyp regression observed mouse models of FAP with NSAIDs for anti-inflammatory therapy (Beazer-Barclay *et al*., 1996). In addition, clinical studies and mouse models show that probiotics are associated with the restoration of gut bacteria by producing conjugated linoleic acid to activate peroxisome proliferator-activated receptor gamma (*PPARG*) which in turn inhibits *COX-2* synthesis (Bassaganya-Riera *et al*., 2012).

Hormone replacement therapy and the oral contraceptive pill have been shown to reduce the risk of CRC development, and this protective effect is thought to arise because of changes in bile synthesis and secretion, but remains controversial (Hoffmeister *et al*., 2007). Vitamins, minerals and dietary supplements are advocated as preventative tools in cancer on the premise that antioxidants prevent oxidative DNA damage and are a part of the body's natural defense against free radicals and reactive oxygen. It has been long presumed that those with low levels of antioxidants are at higher risk of developing certain cancers including CRC.

1.3 MOLECULAR PATHOLOGY OF COLORECTAL CANCER

1.3.1 Clonal expansion and cancer stem cells

CRC is a disease characterised by the clonal expansion of cells that eventually become metastatic and immortal. Adenomas and carcinomas are presumed to arise from single epithelial cells (or small subset of cells), which initiates tumour formation (Ponder and Wilkinson, 1986). For a cell to gain a growth advantage over neighbouring cells and become neoplastic six fundamental properties must be acquired: self-sufficiency, proliferation, insensitivity to anti-proliferative signals, evasion of apoptosis, unlimited replicative potential, the maintenance of vascularisation and tissue invasion and metastasis for malignancy (Hanahan and Weinberg, 2000).

The initial steps in the process of clonal expansion are currently a source of controversy and how early lesions develop from normal mucosa and then progress to cancer are debatable. There are two main models of clonal expansion; the first, referred to as the 'bottom up approach', where mutant cells exist at the crypt base or stem cell compartment and migrate to the crypt apex to expand (Preston *et al*., 2003). The second model was proposed following a study on early non familial FAP adenomas, where dysplasia of cells were seen at the openings and luminal surfaces of colonic crypts (Shih *et al*., 2001). The pathogenesis of this model suggests that mutant cells appear in the intracryptal zone between crypt cavities and, as the clone expands, the cells migrate laterally and downwards to displace the normal epithelium of adjacent crypts (Shih *et al*., 2001) in a 'top-down' fashion. There are

studies which support both mechanisms but these remain highly debatable (Preston *et al*., 2003). Despite the clonal origin of most tumours, their tremendous heterogeneity suggests that cancer progression springs from the combined forces of both genetic and epigenetic events. These are triggered within a single cell to produce a diploid tumour cell, followed by clonal expansion. The acquisition of irreversible genetic alterations creates a tumour formed of genetically heterogeneous cells with a growth survival advantage over normal cells (Kinzler and Vogelstein, 1996, Kleinsmith, 2006, Gerlinger *et* al., 2012, Martinez, *et al*., 2013).

Cancer stem cells (CSC) are a proposed small subset of cells as clusters of tumour cells in a heterogeneous cancer cell population. These are ultimately responsible for self-renewal, tumour initiation, metastasis and also confer resistance to therapy (Dean *et al*., 2005). The therapeutic implications of CSCs are numerous; because these cells have the ability to repopulate the tumour either through enhanced repair mechanisms, hypoxic conditions (a state of low oxygen in tissues) or production of anti-apoptotic proteins. Specific therapeutic drug targets have been partially successful and current research is being invested in this emerging field (Fisher, *et al*., 2013)

1.3.2 Intestinal stem cells

The intestinal epithelial barrier is shaped into distinct compartments dedicated to the bidirectional cell lineage and differentiation of stem cells. Stem cells are self renewing cells with a high proliferative capacity that produce cells that develop into all the differentiated cells of the tissue. These processes depend on actively dividing

intestinal stem cells responsible for the maintenance of tissue homeostasis and any direct damage to the stem cells result in the loss of their dependent cell lineages and subsequent collapse of tissue.

The identity of intestinal active stem cells has been proposed by the discovery of markers. The, most notable is the Leucine-rich repeat-containing G-protein coupled receptor 5 (*LGR5*), a marker discovered from studying the mouse intestine (Barker *et al*., 2007). Other markers including glycoproteins, CD133, CD44, CD166, alcohol dehydrogenase a1 (*ALDHa1*), polycomb ring finger oncogene (*Bmi1*) and epithelial cell adhesion molecule (*EpCAM*) have also been described in the intestinal stem cells (Lugli *et al*., 2010; Sanders *et al*., 2011). *Lgr5* positive crypt base columnar cells (CBC) have broad basal surface and are located interspersed among terminally differentiated Paneth cells (Barker *et al*., 2007, Barker and Clevers, 2012). Loss of *Lgr5* positive stem cells in murine studies of CRC with activated Wnt signalling increased tumourigenesis and produced an invasive phenotype of colorectal carcinomas suggesting an important role for *Lrg5* in restricting stem cell expansion (Walker *et al*., 2012).

Hua and colleagues reported that stem cells from bone marrow and hair follicle cells show radioresistance to p53 mediated cell damage with an increase in NHEJ mediated repair of radiation induced DNA DBSs (Hua *et al*., 2012). Presumably this allows the stem cells to protect the bowel from radiation sensitivity and from accumulating potentially damaging mutations. Therefore, it is of importance to study the mechanisms underlying Wnt signalling and NHEJ mediated DNA repair and the

underlying genes which may confer resistance or sensitivity to intestinal cells during radiotherapy.

Genetic instability has long been postulated to be an essential process of cancer development especially as most human malignancies are recognised cytogenetically by marked aneuploidy and chromosomal rearrangements which include translocations, DNA fragmentation and chromosomal fusion (Hackett *et al*., 2001; Feldser *et al*., 2003). There are three major molecular subtypes of CRC instability pathways, the chromosomal instability (CIN) model, MSI pathway and CIMP pathway, defined as; (cytosine phosphate guanine) CpG Island Methylator Phenotype. A fourth category of microsatellite and chromosome stable (MACS) cancers defines cancers which are neither CIN or MSI.

1.3.3 Chromosomal instability

The most common CRC pathway is driven by CIN and occurs in 80%–85% of CRCs (Grady and Carethers 2008). Tumours that exhibit CIN phenotype are most often located at the distal side of the colon and are associated with better prognosis for the patient (Grady and Carethers, 2008). They are often linked to genomic instability as a consequence of APC driven tumourigenesis where tumours arise through a stepwise combination of mutational activations of oncogenes and inactivation of tumour suppressor genes (TSGs) that drive the adenoma to carcinoma sequence (Shih *et al*., 2001, Alberici *et al*. 2007; Fearon and Vogelstein, 1990). Alterations in oncogenes are instrumental in promoting cancer while TSGs are involved in processes to prevent dysregulated growth. Alterations in these

genes lead normal colonic epithelium through hyperproliferation to dysplasia, adenoma and eventually to carcinoma (Figure 1-3) (Fearon and Vogelstein 1990; Su *et al*., 1993; Lengauer *et al*., 1998).

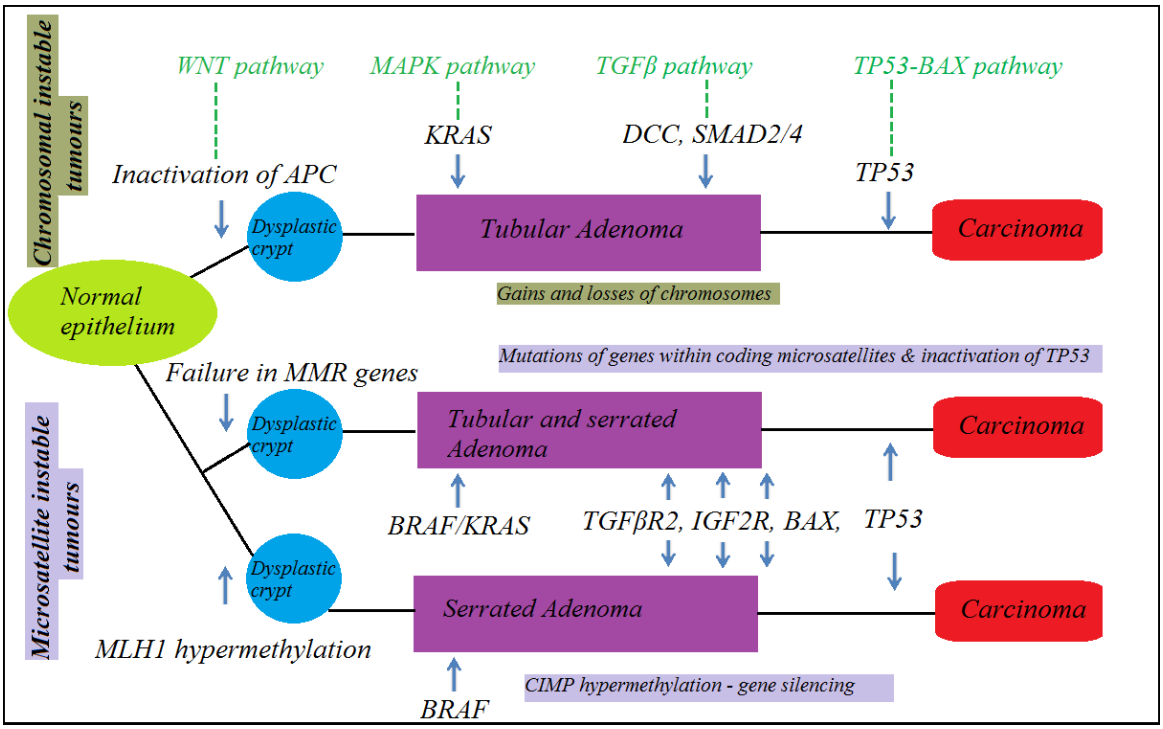

Figure 1-3 A modified 'Vogelgram' showing the three main molecular changes involved in the sequential adenoma to carcinoma pattern in CRC: CIN, MSI and CIMP tumours

Evidence for the adenoma-carcinoma sequence is substantiated by a number of observations: firstly, the incidence of adenomatous polyps precedes that of adenoma carcinomas (AC) secondly, small carcinomatous foci are found within advanced adenomatous polyps and finally cancer risk is directly related to the number of adenomas, as shown by the very high incidence of carcinoma in FAP patients with >100 polyps (Bujanda *et al*., 2010). Furthermore, in a landmark study by Vogelstein of CRC specimens with a range of neoplastic stages, allelic deletions were assayed for chromosomes 5, 17, 18 and for mutations of CRC susceptibility genes. Table 1-3

shows a summary model of the genes known to be implicated in CRC tumour progression and are discussed below in milieu of the adenoma-carcinoma sequence and clinical outcome (Vogelstein *et al*., 1988).

Table 1-3 A summary of tumour suppressor genes implicated in CRC progression. (PM=point mutation, FS=frameshift, MS=missense, del=deletion, ins=insertion).

TSG	**Function**	**Germline mutations**	**Somatic mutation**	**Clinical Condition**
APC *Gatekeeper*	cell adhesion, anterior-posterior pattern formation, axis specification, cell cycle, cell migration, apoptosis, chromosome segregation and spindle	80-85% are PM resulting in stop codon, small del, LOH, 1-2 bp ins result in truncated protein	50% sporadic CRC FS, MS – resulting in premature stop codons mutations clustered in Exon 15 20% mutations observed with TP53 mutation	FAP, AFAP and related conditions
TP53 *Gatekeeper*	Cell cycle, G1 DNA damage checkpoint Regulated transcriptional genes involved in cell growth, proliferation and apoptosis	Arg>Pro at codon 72 -No clear conclusive evidence this mutation is linked with cancer susceptibility	MS, del, ins and splicing variants in exons 4-6 (hotspots at codons 175, 245, 248, 273 and 282) with 80% mutations from methylation of CpG region at 175, 248 and 273 (in bold) 40-50% observed in CR carcinoma not adenoma	Li-Fraumeni syndrome and greatly increase in susceptibility of all cancers
MLH1 *Caretaker*	DNA Mismatch repair gene	Del of codons 578 and 632 Mutations give rise to microsatellite instability	Methylation of promoter	Mainly responsible for LS
MSH6 *PMS2* *Caretakers*	Both function in mismatch repair binding	-CT del (exon 4) -in frame del, PM		LS via microsatellite instability
AXIN2 *Gatekeeper*	protein has role in regulation of β-catenin	Finnish Family study	Is mutated in MMR deficient CRC cells,	Hypodontia and CRC

	pathway via GSK-3β	(1966 C > T mutation)	possibly associated with epigenetic silencing Mutations result in stabilisation of β-catenin and T-cell factor signalling	
STK11 Gatekeeper	TSG of serine-threonine kinase sunction in cell cycle arrest	FS, MS	Give rise to polps in GI tract	Peutz-Jeghers syndrome
PTEN Landscaper	negatively regulates AKT/PKB signalling known as mutated in multiple advanced cancers	Results in cancer prone syndrome - Cowden Syndrome	Results of MLH1-MSH2 deficiency in LS patients FS in MMR deficient cells	Juvenile polyposis, Cowdens disease
BMPR1A Gatekeeper	Transmembrane serine-threonine kinase, type1 receptor	FS, MS and SS mutations	Rare substitution mutations observed in the protein kinase domain of large intestine tumours.	Juvenile polyposis, Cowdens disease
SMAD4 Gatekeeper	TGF β signalling, mediating signals from cell surface to nucleus	Results in juvenile polyps	Minimal LOH on chromosome 18q-21, though 4 bp del of exon 9 in CRC resulting in stop codon. Del of SMAD4 in epithelial cancers	FAP, Juvenile polyposis, Cowdens disease
MYH Caretaker	DNA glycosylase. Functions in oxidative DNA damage repair (BER)	Tyr 82-Cys and Gly253-Asp affect glycosylase activity resulting in APC mutations in somatic cells	MS and truncating protein mutations cause 3-fold increase in CRC risk	MAP
DCC Conditional Gatekeeper	receptor for neptin-1 (axonal chemoattractant). In absence of ligand, it induces apoptosis but when neptin-1 is bound, it prevents apoptosis	Rare germline mutations - 74% loss in CRC also observed in hepatocellular metastatic cancers	Rare homogygous del 120 to 300 bp expansion in exon 7, present in 10-15 % of tumours that display genomic microsatellite instability	Mainly neurological role Advanced CRC

1.3.4 Tumour suppressor genes

TSGs were first recognised and described in epidemiological studies on familial and non-familial cases of human retinoblastoma by Knudson and colleagues (Knudson, 1971). He proposed the 'two hit' theory that both copies of the retinoblastoma gene (*Rb*) are inactivated for familial cancer cases to develop and the inheritance of one allele predisposes the individual to retinoblastoma when the second allele undergoes mutations. This was later supported when *Rb* was discovered with mutations in both alleles of *Rb* tumours (Friend *et al*., 1986; Dryja *et al*., 1986). Bert Vogelstein grouped TSGs into two classes of cancer susceptibility genes; gatekeepers and caretakers (Kinzler and Vogelstein 1997). In 2004 a third classification of landscapers was proposed later (Michor 2004).

APC is a classic example of a TSG. FAP patients inherit one mutated *APC* allele and inactivation of the remaining normal copy of the *APC* gene, by deletion or mutation, completely removes the 'gatekeeper' tumor suppressive function of *APC,* thus initiating the growth of adenomatous polyps. The loss of functional mutations of *APC* occurs in over 80% of CRCs, which abrogates the normal function of the APC protein in controlling of β-catenin activity (Miyoshi *et al*., 1992; Kinzler and Vogelstein, 1996; Pietenpol and Volgelstein, 1993). A*PC* second mutational hits are not randomly selected, but interdependent and distributed according to the resulting levels of residual β-catenin involved in Wnt activation favourable for tumourigenesis, a process referred to as the 'just right' theory (Lamlum *et al*., 1999; Albuquerque *et al*., 2002). According to the position of the germline mutation and the number of 20 amino acid repeats (20AARs) (β-catenin down regulations sites)

retained, three possible 'second hits' that can be expected (see 1-4). Analysis of somatic or "second hit" events has led to the discovery of a new theory in which *APC* requires a 'third hit'. The most frequent third hit was copy number gain or deletions (Segditsas *et al.*, 2009). These hits are associated with the genotype-phenotype differences observed in FAP patients that affect the severity of colonic polyps and will be explained later in section 1.4.4 of this chapter.

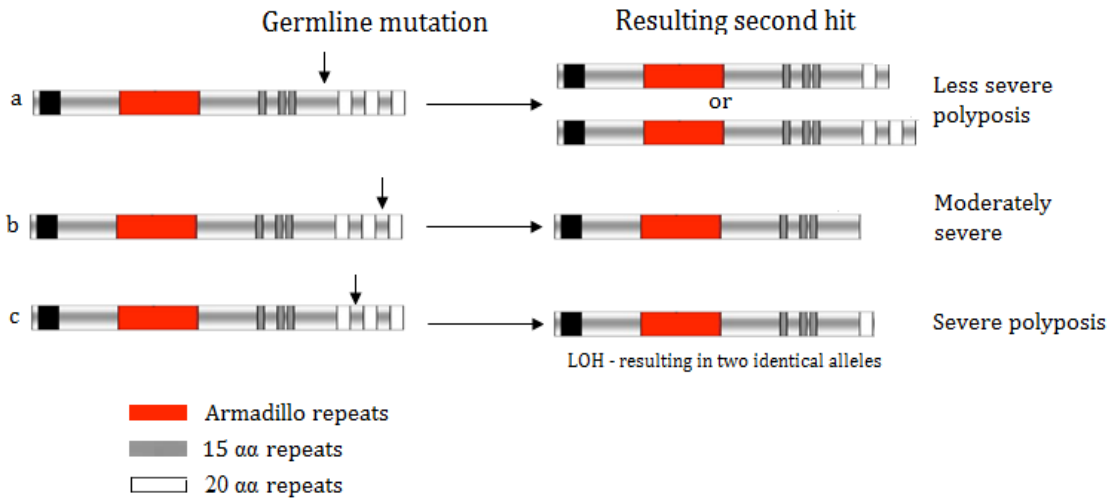

Figure 1-4 The interdependence of germline and somatic APC mutations associated with FAP, where the position and type of the second hit in FAP depends on the location of the germline mutation and the number of 20 amino acid repeats still present in the truncated protein. Germline mutations leading to a truncated protein lacking all the 20AARs (a) acquire somatic second hits, resulting in truncated proteins with one or two repeats. Somatic mutations and truncated APC protein with a loss of all 20AARs is observed in the second hit (b) and germline mutation with one 20AARs, the second hit will result mainly in allelic loss and produce a second homologous allele (c) (Albuquerque *et al.*, 2002).

1.3.5 Oncogene

The first oncogenes to be discovered were *Ras* by Jennifer Harvey and Werner Kirsten in the 60s and 70s (Harvery 1964, Werner 1972) and *Src* by Noble Prize laureates Michael Bishop and Harold Varmus in 1976 (Stehelin *et al*., 1977). These were observed in the retroviral DNA of chicken and rats respectively. By 1976 it was understood that oncogenes are in fact malfunctioning proto-oncogenes and found in many organisms with the ability to induce malignant transformation of human cells *in vivo*. The same genes were later identified in humans with the same cell regulation, cell signalling, growth and differentiation functions (Parada *et al*., 1982, Knudson, 1985).

Ras proteins; HRAS, NRAS and KRAS, are all localised to the inner leaflet of the plasma membrane which is important for their biological function as monomeric GTPases. They operate in cellular signal transduction pathways to regulate cell growth, differentiation and apoptosis. Almost a third of human cancers carry a *RAS* point mutation at codons 12, 13 or 61 rendering the protein constitutively active and oncogenic. HRAS and NRAS traffic to the plasma membrane via a series of post-translational modifications, while the KRAS trafficking pathway is currently poorly characterised. Mutations in *KRAS* are the most frequently observed of cancers: occurring in the pancreas (90%); the colon (50%); and the lung (30%); thyroid tumours (50%); and in myeloid leukaemia (30%), (Salhab *et al*., 1989; Roberts and Der, 2007).

Assays for *KRAS* mutations in CRC specimens indicate that the frequency of *KRAS* mutation increases from 9% in adenomas just under 1 cm in size to 58% in those over 1 cm, providing evidence that KRAS is involved in tumour growth promoting aspects and not CRC initiation (Vogelstein *et al.*, 1988; Leslie *et al.*, 2003). The chronological order of mutations is important in CRC development, with the loss of *APC* representing the earliest and rate-limiting genetic event in *CRC* tumour initiation (Powell *et al.*, 1992) leading to β-catenin nuclear localization. *KRAS* mutation is then a following step to promote adenoma progression to carcinomas (Vogelstein and Kinzler, 2004; Phelps *et al.*, 2009).

1.3.6 Deleted in colorectal cancer and SMAD tumour suppressor genes

Further malignant progression along the adenoma to carcinoma sequence involves the allelic losses of chromosomes 18q and then 17p, which have been found to occur in a large proportion of both colorectal adenomas (CRA) and carcinomas (Fearon *et al.*, 1987; Fearon and Vogelstein, 1990; Powell *et al.*, 1992).

Chromosome 18q is lost in 10 to 30% of early adenomas, 60% of late adenomas and 70% of carcinomas (Leslie *et al.*, 2003). The first candidate TSG in this chromosomal interval to be described was the "deleted in colorectal cancer" (*DCC*) gene which encodes 1447 amino-acid transmembrane protein with four immunoglobulin-like and six fibronectin III-like domains (Hedrick *et al.*, 1994). *DCC* acts as a transmembrane receptor for netrins, which are key components involved in the lateralisation of the human nervous system via axon growth in neurons (Srour *et al.*, 2010). The expression is detected in many tissues including the lung, colon, and

testis with the highest expression in normal brain tissue. In addition to CRC, *DCC* is considered a TSG target gene for pancreatic (Murty *et al*., 1994) and testicular cancers (Strohmeyer *et al*., 1997). In the absence of netrin or at low levels *DCC* induces apoptosis and may be involved in activating downstream signalling pathway that involves MAPK activation. Subsequent activation of the transcription factor Elk-1 and SRE regulated gene expression are pivotal to transducing extracellular mitogenic signals into a nuclear response and contribute to the pathogenesis of CRC

Other TSGs found in this interval are SMAD family members (SMADs), which are signal transducers that mediate intracellular transforming growth factor-beta (TGFβ) signalling pathways. SMAD proteins bind to TGFβ receptor 2 (TGFBR2) receptor to promote the formation of heterodimer with the TGFβ receptor 1 (TGFBR1) and binds to its DNA translocation factors where the complex activates the transcription of genes responsible for a broad range of cellular functions including cellular growth inhibition, apoptosis, differentiation, and matrix production (Heldin *et al*., 1997; Duff and Clarke, 1998). Mutations in *SMAD2* or *SMAD4* are mutated, TGFβ signal is not transduced into the nucleus of the cell.

In addition, mutations in TGFβR1 have been identified in CRC cell lines (Ku *et al*., 2007) and the most common variant TGFβR1*6A, (*6A) is located in exon 1, has a deletion of 3 alanines within a stretch of 9 alanines and has been suggested to be implicated in increased risk for several cancers, for example breast and ovarian cancer (38% and 41% respectively) (Kaklamani *et al*., 2003), and CRC (Pasche *et al* 2004). A significantly higher *6A frequency was observed amongst cancer patients than healthy controls (Pasche *et al*., 1998) and meta-analysis of 12 case control

studies in 7850 Caucasians, showed that 14% were carriers of the *6Ala genotype and are at a 20% higher risk of CRC and based on this the population attributable risk is predicted to be 3.2% (Pasche *et al*., 2004). These findings have been confirmed by Zang *et al* showing 44% higher frequency of *6A in CRC cases than in controls (Zhang *et al*., 2005) which highlights that *6A susceptibility allele is common in the general population and the contribution to the total CRC burden is relatively high in spite of its modest relative risk (1.20). TGFBR2 mutations are also frequently found to affect TGF-β signalling in CRC, mainly among MSI tumours but also in approximately 55% of microsatellite stable (MMS) tumours (Grady *et al*., 1999). Overall, there is convincing mutational evidence for the major role of TGF-β pathway inactivation in the adenoma-carcinoma transition and for SMAD2/4 as primary TSGs of CRC (Roberts and Derynck *et al*., 2001).

Germline mutations in the *SMAD4* and bone morphogenetic protein receptor, type IA (*BMPR1A*) have been shown to be associated with Juvenile polyposis syndrome (JPS) (He *et al*., 2004; Howe *et al*., 1998). In addition, phosphatase and tensin homologue (*PTEN*) mutations are also implicated in JPS predisposing to benign and self-limiting hamartomatous polyps in the upper gastrointestinal tract, small bowel and colon (Chen and Fang, 2009). The lifetime risk of developing GI cancers ranges from 9% to 50% (Howe *et al*., 2004; Schwenter *et al*., 2011). *PTEN is* a landscape TSG which acts indirectly through intracellular signalling and control of the microenvironment in which cells grow and contribute to the neoplastic growth of cells by creating an environment suitable for unregulated cell proliferation (Michor, 2004; Lengauer *et al*., 1997).

1.3.7 TP53 tumour suppressor gene

Most malignant CRC are characterised by loss of the short arm of chromosome 17 (17p) which harbours the *TP53* TSG (17p13.1) in humans. The gene spans 20 kb, with coding sequence in exons 2, 5, 6, 7 and 8, which show a high degree of conserved regions between vertebrates (May and May 1999). Germline mutations of the *TP53* TSG are responsible for the autosomal dominant condition Li-Fraumeni syndrome which is known to greatly increase susceptibility to cancer (Varley 2003). The associated risk of developing cancer is 90% by the age of 70, including sarcomas, leukaemia and malignancies of the breast, brain, and adrenal cortex (Petitjean *et al*., 2007).

It is mutated in over 50% of all CRC (Hollstein *et al*., 1991, Soussi *et al*., 2007) and is very rarely found in benign lesions suggesting that functional inactivation of the *TP53* gene is a late genetic event associated with the transition from adenoma to carcinoma (Vogelstein *et al*., 1988; Rodrigues *et al*., 1990). *TP53* is referred to as the "master watchman", due to ability to block cell proliferation via transcriptional activation of cyclin inhibitors, like p21 in the presence of DNA damage (Waldman *et al*., 1995). In unstressed cells, *TP53* is usually bound to murine double minute (*Mdm2)* oncogene, which promotes continuous degradation by acting as ubiquitin ligase and thus maintaining low levels in cells. *TP53* is activated by the phosphorylation of the N-terminal domain, involving 2 types of protein kinases: MAPK family which responds to several types of stress such as membrane damage and oxidative stress; or the second group; ATM, DNA-PKs, checkpoint homologs 1 and 2 (CHK1, CHK2) family which responds to DNA damage initiated by genotoxic

stress and DNA double strand damage caused by UV, IR or carcinogens via DNA integrity checkpoints. The major events that occur upon *TP53* activation lead to an increase in its half-life and then a conformational change as a transcription factor to regulate stressed cells.

1.3.8 Serine/threonine kinase 11 tumour suppressor gene

Peutz-Jeghers syndrome (PJS), is caused by an autosomal dominant condition, with 90-100% of all cases with a specific genetic mutation in Serine/threonine kinase 11 (*STK11*) TSG also known as liver kinase B1 (*LKB1*) (Jenne *et al* 1998). The gene encodes a member of the serine/threonine kinase and is involved in regulating cell polarity and functions also as a gatekeeper (Hemminki *et al*., 1998). PJS Individuals carry a 20% lifetime risk of developing CRC (Lai *et al*., 2011). The condition is associated with multiple harmatomatous polyps in the digestive tract and hyperpigmentation on the face, around the lips and hands which is often the first sign for diagnosis. An increasing number of sporadic mutations in *LKB1* are being reported in diverse cancers (Sanchez-Cespedes 2007); 30% of lung cancers show mutations in the LKB1 gene (Sanchez-Cespedes *et al*., 2002).

1.3.9 Caretakers and microsatellite instability

Mutations of 'caretaker' genes affect the maintenance of genomic integrity which occurs via cell cycle checkpoints and DNA repair and maintenance pathways, such as NER, BER, MMR and NHEJ (Levitt and Hickson, 2002; Michor *et al*., 2004). The MSI pathway is associated with loss of DNA MMR proteins, which leads to the

accumulation of mutations. MSI or replication error positive (RER^+) has been identified amongst microsatellites, which are repetitive genetic loci prone to slippage during DNA replication resulting in small alterations at microsatellite loci. Inefficient MMR genes cannot repair these allelic changes and this allows for the formation of different sized alleles at the next round of replication (Aaltonen *et al.*, 1993). Genetic and epigenetic changes will in turn cause an increased genomic mutation rate and may facilitate malignant transformation and so MSI has also been referred to as the mutator phenotype. MSI caused by deficient MMR functions was found in about 15% of all sporadic tumours (Fridrichova, 2006) and 90% in LS tumours (Aaltonen *et al.*, 1993, Wang *et al.*, 1999). However, not all LS patients have MSI and about 15% of non LS CRCs show MSI (Julié *at al.*, 2008).

1.3.10 Lynch syndrome

The predisposition of some families to develop cancer at a young age was first recognised by Alfred Warthin in 1895, when his seamstress, predicted her own death from cancer from her family history. By studying her family (Warthin's family G) he found that gynaecological, colonic and stomach cancer occurred frequently. It was not until the 1980s that Lynch developed the idea of cancer family syndromes, which became a fully accepted principle and the term LS was used for this condition (Lynch and Krush, 1971). Initially, it was divided into type I and type II depending on whether individuals developed colonic cancers or not. Later, it was discovered that both of these syndromes were due to inherited defects in the genes that regulate the excision of errors occurring during DNA replication of the mismatch genes. The germline mutation was therefore classed as an autosomal dominant gene

with a high risk for familial predisposition to CRC and extracolonic cancers of the GI, urological, stomach and female reproductive tracts (Lynch and de la Chapelle, 2003).

LS is the most common form of inherited CRC, accounting for 3-5% of all CRCs. These cancers do not display aneuploidy or the characteristic mutations nominated by Fearon and Vogelstein (Fearon and Vogelstein 1990). It is usually diagnosed based on the Amsterdam II and Bethesda criteria (Rodriguez-Bizas *et al.,* 1997; Vasen *et al*., 1999). That is, familial aggregation of three or more relatives with histological verified CRCs, one of which is the first degree relative of the other two. CRC involving at least two successive generations, at least one relative diagnosed with CRC under the age of 50, exclusion of FAP, inclusion of endometrium, ureter, renal pelvis and small bowel cancers and to test the patient's tumours for MSI (Peltomäki 2001, Scott *et al*., 2001). LS patients have an 80% occurrence risk of developing CRC at around 45 years of age compared to 60 for other familial CR cancers (Rustgi *et al*., 2007).

LS is caused by the inheritance of one defective *MMR* gene; MutL homolog 1, (*hMLH1*), MutS homolog 2 (*hMSH2*), MutS homolog 6 (*hMSH6*), and to a lesser extent postmeiotic segregation (*PMS2*), with more than 400 different predisposing *MMR* gene mutations reported, approximately 50% affect *hMLH1*, 40% *hMSH2*, 10% *hMSH6* and less than 5% affecting *PMS2* (Peltomaki 2001). MMR proteins are required to recognise mismatches and are named according to their homology to the bacterial proteins, known as MutS, MutL and MutH and PMS2 (Leach *et al*., 1993). These proteins form complexes and have specific roles: hMSH2 and HMSH6 form MutSα complex to identify single base mismatches, where hMSH2 and hMSH3

together form the MutSβ complex that identifies short insertions or deletion errors. hMLH1 and hPMS2 combine in a heterodimer that interacts with the MutS complexes to excise and repair DNA (Plotz *et al*., 2006).

While some studies report that CIN and MSI are mutually exclusive pathways of genetic instability in CRC (Lengauer *et al*., 1998), more recent studies indicate shared pathways, for example mutations in exon 3 of *β-catenin* are associated with malignant CRCs in LS tumours, and appear not to occur as a result of MMR deficiencies, indicating that genetic pathways maybe linked with an interplay between the Wnt pathway occurring at later stages of tumour development (Johnson *et al*., 2005). The extent of overlap and shared features are unclear, and are likely to involve very small numbers of genes (Muleris *et al*., 2008, Pino and Chung 2010).

1.3.11 Epigenetic events – CIMP and DNA methylation

The third pathway CIMP was suggested by Toyota *et al* in 1999 and cancer development occurs via hypermethylation at several CpG loci converting cytosine to 5-methylcytosine (Toyota *et al*., 1999; Shiraishi *et al*., 2002; Wheeler 2005; Jass, 2007). Approximately 1% of human DNA bases undergo DNA methylation, which involves the addition of a methyl group to DNA (Slagboom and Meulenbelt, 2002) and in adult somatic tissues. DNA methylation typically occurs at between 60-90% of all CpGs (Jones and Baylin, 2007; Calvanese *et al*., 2009).

In cancer, CpG islands, are present in the promoter regions of TSGs, where DNA methyltransferases (DNMT1, DNMT3a, DNMT3b) physically block the binding of transcriptional proteins causing inactivation or silencing of the gene (Vallbohmer *et al*., 2006; Butcher and Rodenhiser, 2007). Findings from several studies have shown that tumours can stably maintain mutations in one allele of a TSG while the other allele is hypermethylated, leading to a functional inactivation of the gene (Myohanen *et al*., 1998; Grady *et al*., 2000). This phenomenon has been described as the 'third pathway' in Knudson's model of TSG inactivation in cancer (Kondo and Issa, 2004). Toyota and co-workers showed that CIMP positive tumours commonly occur in 35-40% sporadic CRCs with MSI related to *MLH1* hypermethylation affecting the mismatch repair process (Baylin *et al*., 1998, Toyota *et al*., 1999, Herman and Baylin, 2003, Kondo and Issa, 2004). Furthermore, CIMP has been showed to be strongly associated with *BRAF* mutation (V600E) (Weisenberger *et al*., 2006), where hierarchal clustering has identified three groups with distinct genetic and epigenetic profiles; CIMP negative tumours display rare methylation and *TP53* mutation. CIMP1 tumours are methylated at multiple loci and display MSI and *BRAF* mutations. In contrast, CIMP2 is methylated at a limited number of age-related genes and displays mutations in *KRAS* (Weisenberger *et al*., 2006, Shen *et al*., 2007).

1.3.12 Haploinsufficiency

Another interesting phenomenon involves the loss of only one functional allele, which may also be sufficient to produce a tumour phenotype in a diploid cell. This is termed haploinsufficiency and at a TSG locus may overcome the need for the somatic loss or mutation of its wild type allele (Payne and Kemp 2005). The onset of a

tumour phenotype occurs when the remaining copy does not provide sufficient messenger RNA (mRNA) to encode for a sufficient amount of functional protein (Payne and Kemp 2005; Fodde and Smiths, 2002). To date, experimental evidence for haploinsufficiency in cancer predisposition comes from the analysis of tumours obtained from mouse models or from hereditary cancer patients carrying heterozygous *null* mutations at known TSGs. The absence of the second hit in a subset of these tumours has been attributed to inactivation of the remaining allele by alternative mechanisms such as epigenetic silencing or mutations in non-coding sequences. For example, mice heterozygous for the TSG gene *Smad4,* develop tumours as a consequence of halpoinsufficiency and initial retention of the wild type allele. Although loss of a single gene copy is sufficient for tumour formation, LOH of the remaining *Smad4* allele does occur later on in tumourigenesis (Alberici *et al*., 2006; Xu *et al*., 2000). *Bona fide* haploinsufficiency has been demonstrated for a subset of tumour suppressor loci including *ATM* (Lu *et al*., 2006), *PTEN* (Kwabi-Addo *et al*., 2001), *LKB1* (Nakau *et al*., 2002) and*TP53* (French *et al*., 2001).

1.4 FAMILIAL ADENOMATOUS POLYPOSIS

FAP was first documented by Menzel in 1872 and later Virchow described a 15 year old boy with multiple polyps as "polyploidy colitis" in 1863. Cockayne reported on the autosomal dominant mode of transmission in 1927 which was later confirmed by Dukes and Lockhart-Mummery who devised a clinical registry of cases and reported on the familial association and confirmed the disease process as Mendelian dominant (Lockhart-Mummery, 1934). By 1955, FAP was well described and defined as multiple polyposis, a condition where hundreds and thousands of adenomatous polyps are found in the large intestine and rectum (Woolf 1958). Lipton and Tomlinson further reported the development of thousands of ACF within the intestinal tract of FAP patients (Lipton and Tomlinson, 2006). FAP patients are at high risk of developing adenomatous carcinomas that tend to occur between the ages of 20 to 45 leading to virtually 100% occurrence of cancer if left untreated (Rustgi *et al*., 2007). It is also generally accepted that the larger polyps transform to adenomatous carcinomas (Debinski *et al*., 1996).

1.4.1 The structure and function of Adenomatous Polyposis Coli gene

The genetic basis for FAP lies in germline mutations of the *APC*. The discovery and physical mapping of the *APC* gene came from a patient with colorectal polyposis and mental retardation who had a deletion in the 5q sub-chromosomal region (Herrera *et al*., 1986, Bodmer *et al*. 1987). Further segregation and linkage analysis in FAP families led to the mapping of the *APC* gene to 5q21 (Bodmer *et al*., 1987) which was cloned and characterised in 1991 to consist of 8535 base pairs that encode a 2843

amino acid multi-domain protein (Figure 1-5) (Groden *et al*., 1991; Nishisho *et al*., 1991; Kinzler *et al*., 1991). The main role of APC is to regulate β-catenin concentration in the cell cytoplasm via Wnt pathway but is also involved in cell-cell adhesion, microtubule stability, cell cycle regulation and apoptosis (Behrens, 2005). It is present in a variety of epithelial tissues, often diffusely distributed in the cytoplasm and can sometimes be found in the apical or lateral regions of epithelial cells (Midgley *et al*., 1997) suggesting *APC* mediates apoptotic control of cellular migration within colonic mucosa regulating cell formation and cell death.

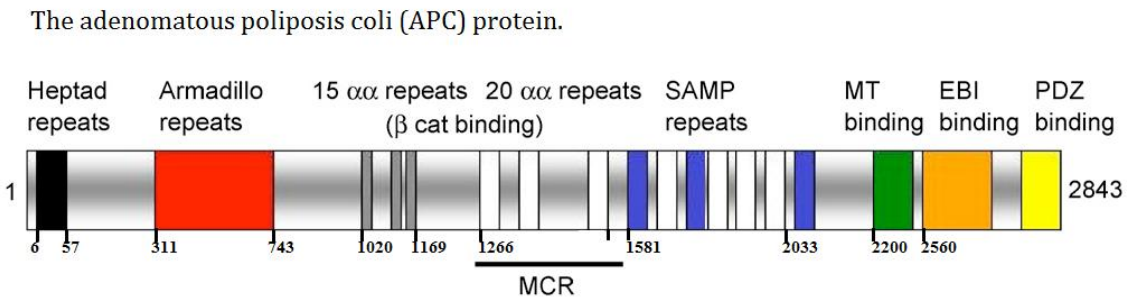

Figure 1-5 The 2843 amino acid (αα) adenomatous polyposis (APC) protein. Conserved domains within APC such as armadillo repeat and regions interacting with other proteins are shown. N-terminus binds to the B56 regulatory subunit of protein phosphatase 2A. β-catenin binding occurs at three 15 αα repeats (between 1020 and 1169 αα) (in blue) and a second binding region with seven 20 αα repeats (in white, between 1342 and 2075 αα), which are interspersed by axin binding SAMP motifs. The mutation cluster region (MCR) (approx 1250 to 1550 αα) encompasses β-catenin binding sites and mutations in this region produce the most severe phenotype in humans. The C terminal portion of the APC protein is involved in binding to cytoskeletal microtubules (MT) and tubulin-binding protein (EB1) (McCart *et al*. 2008).

1.4.2. The Canonical Wingless-Int pathway

The canonical Wnt pathway is an evolutionary conserved signalling pathway that plays numerous developmental and cellular roles in animals (Logan, 2004). The deregulation of this pathway can lead to various types of cancers in particular CRC.

The sequence of events that occur in the CIN pathway to cause cancer is fairly well documented. One of the first events is the activation of the Wnt signal transduction pathway via the loss of *APC* function (Figure 1-6). In the intestine, differentiation of cells as they move up the crypt is under the control of Wnt and bone morphogenic protein (BMP) signaling: as Wnt decreases and BMP increases, cells differentiate and move towards the apex of the villi. When Wnt is high in the bottom and the middle of the crypt, cells tend towards proliferation, but when Wnt is low and BMP is high, cells differentiate and eventually slough off from the top of the villus (Kosinski *et al*., 2007). This suggests that BMP signaling suppresses Wnt signaling, which is central to ensure a balanced control of stem cell self-renewal.

In a normal cell, the absence of the Wnt signal, the wild type APC protein mediates its tumour suppressor function with the downstream binding of APC to β-catenin, glycogen synthese kinase-3β (GSK3β) and axin (Figure 1-6 a). This complex phosphorylates β-catenin and leads to it degradation, regulating its concentration in the cytoplasm and promoting its breakdown. Ultimately the degradation causes the suppression of the Wnt signal and prevents the transcriptional activation of β-catenin target genes (Su *et al*., 1993; Sparks *et al*., 1998; Bienz and Clevers, 2000; van Es *et al*., 2001; Giles *et al*., 2003; Logan and Nusse 2004;).

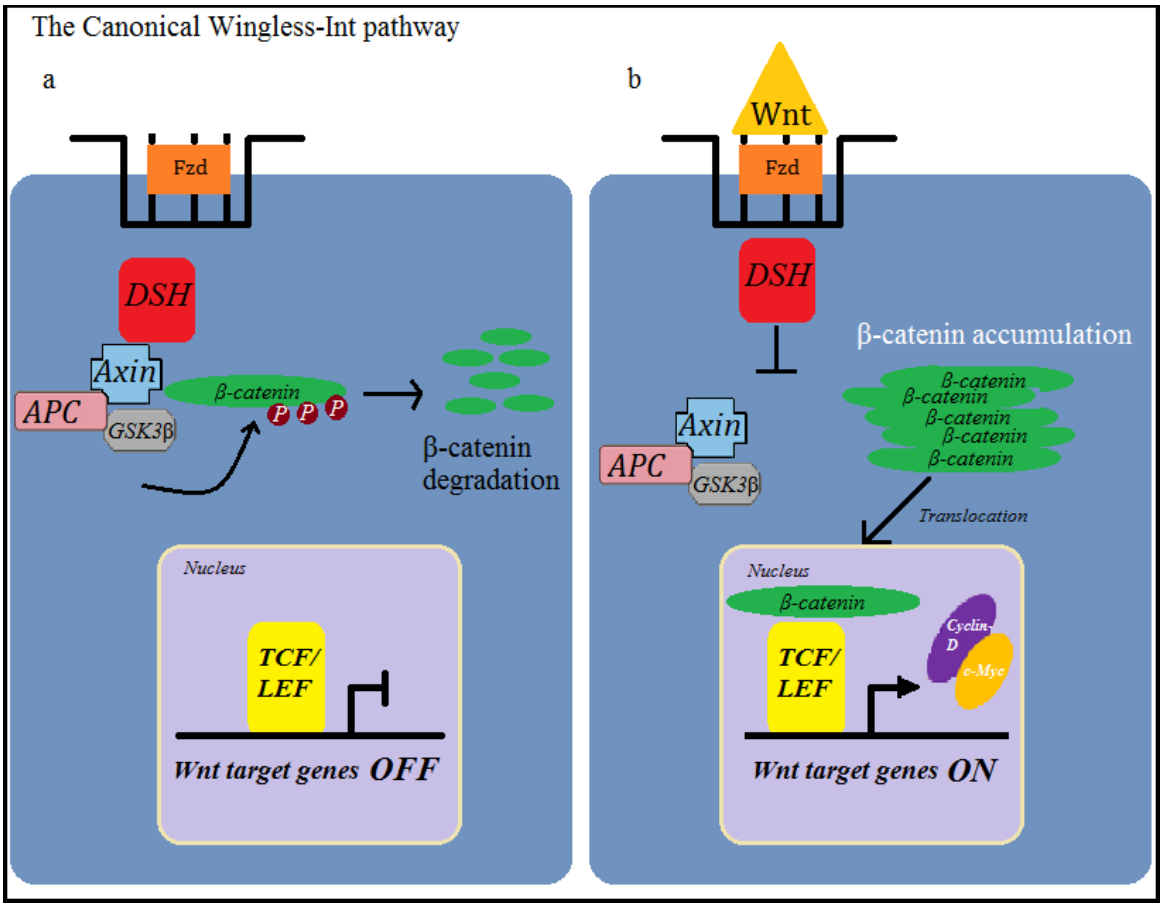

Figure 1-6 The Wnt pathway: off (a) and on (b). Cytoplasmic β-catenin is phosphorylated and degraded by a complex of *APC*, axin and GSK-3β in the absence of Wnt. In the presence of Wnt signal ligand bound to its receptor Frizzled (FRZ), Dishevelled (DSH) is recruited and inactivates the APC-axin-GSK-3β complex. β-catenin may then escape degradation and enter the nucleus where, in conjunction with TCF/LEF, it up-regulates POL ribonucleic acid polymerase for the transcription of multiple target genes.

In the presence of the Wnt signal, or with dysfunctional APC, Wnt binds to the cell surface receptors of the Frizzled family (FRZ), which recruit activated Dishevelled (DSH) family of proteins (Figure 1-6 b). DSH is a key component of the Wnt receptor complex and inhibits the APC-β-catenin-GSK3β-axin complex from forming and initiating the 'normal' proteolytic degradation of the β-catenin (Bienz and Clevers, 2000; Grohmann *et al.*, 2007). The pool of cytoplasmic β-catenin increases and

stabilises, leading the β-catenin to translocate to the nucleus where it heterodimerises with T-cell factor/lymphoid enhancer factor (TCF4/LEF1) family transcription factors to promote activity of a variety of genes involved in proliferation and malignant transformation: the proto-oncogene *c-MYC*, the cell cycle activator CCND1 (*cyclin D1*), gastrin, fibronectin, matrix metalloproteinase 7 (*MMP7*), and vascular endothelial growth factor (*VEGF*) amongst others (Mann *et al*., 1999; Tetsu and McCormick, 1999, Behrens, 2005, Senda *et al*., 2007). This is a key step in tumourigenesis and accumulation of nuclear β-catenin is associated with disease progression (Hugh *et al*., 1999; Bienz and Clevers, 2000; Clevers, 2006).

1.4.3 Adenomatous Polyposis Coli mutations

Mutations in the *APC* gene are very heterogenous, dispersed along 15 exons and more than 800 different germline APC mutations have been identified in FAP tumours and appear to be confined to the 5′ half of the *APC* gene (Miyoshi *et al*., 1992; De Filippo *et al*., 2002; Nagase and Nakamura 1993; Fearon and Vogelstein, 1990). Approximately, 90% of all germline *APC* mutations result in the alteration of the reading frame and introduce premature stop codons leading to truncated proteins as a consequence (de la Chapelle 2004). Béroud and Souss pointed out that 30% inactivation of the APC protein was due nonsense point mutations (usually C>T changes) and 68% by frameshift mutations (Béroud and Souss, 1996). Deletions involving exons or the entire *APC* are seen in 10% of FAP and 20% arise *de novo*. There are two frameshift mutational hotspots of 5 bp deletions at codons 1061 and 1309; these are prone to spontaneous slippage and comprise of 12% and 18% of all

germline APC mutations, respectively. Other frameshift mutations include 2 bps and 1 bps deletions at codons 565 and 1436, respectively.

Mutational analyses of sporadic CRC tumours, like germline *APC* mutations, consist mainly of frameshift and nonsense but are clustered in the center of *APC*; known as the mutation cluster region (MCR) and dominate the sporadic mutational spectrum. This region spans codons 1286–1513 of exon 15 (Nagase and Nakamura 1993; Polakis 1997; SU *et al*., 2002) and accounts for 68% to 77% of all somatic mutations (Miyoshi *et al*., 1992).

1.4.4 Familial Adenomatous Polyposis genotype-phenotype variations

Genotype-phenotype correlation studies for the *APC* gene have shown that the location of a germline mutation provides some information about the polyp burden and the health problems associated with FAP. Several reports have shown a correlation between disease severity and the site of *APC* mutation (Nagase and Nakamura 1993; Nugent *et al*., 1994; Groves *et al*., 2002). In addition, a proportion of FAP patients are affected by extracolonic conditions such as osteomas, epidermoid cyst, congenital hypertrophy of the retinal pigment epithelium (CHRPE), upper gastrointestinal (UGI) polyps or adenomas, desmoid tumours, brain, thyroid, and biliary tract malignancies all which depends on the position of the inherited *APC* mutation.

Classical polyposis leads to thousands or more adenomas, a polyp burden that is usually associated with mutations in codons 169 and 1600. Patients with mutations

in codons 1250 and 1464 are associated with a severe polyposis phenotype (>1000 polyps) which tend to present earlier, are more often symptomatic, and may be more likely to develop CRC (Debinski et a., 1996) Figure 1-7. Similarly, the profuse phenotype, which is characterised by numerous intestinal tumours, also results from mutations between codon 1250 and 1464 (Nagase *et al.*, 1992). Truncations between codons 1-436 and in the extreme 5' end of APC produce AFAP. This is associated with a smaller number of adenomas, which appear later in life (Lynch *et al.*, 1995). Recent genotype-phenotype correlations showed that the age at polyposis onset and years of survival differed significantly by genotype. Patients with *APC* mutation between codons 0-178 or 312-412 developed 'late' polyposis and survived longer, whereas patients with mutations in *APC* 1249-1549 developed polyposis earlier and did not survive as long (Newton *et al.*, 2012).

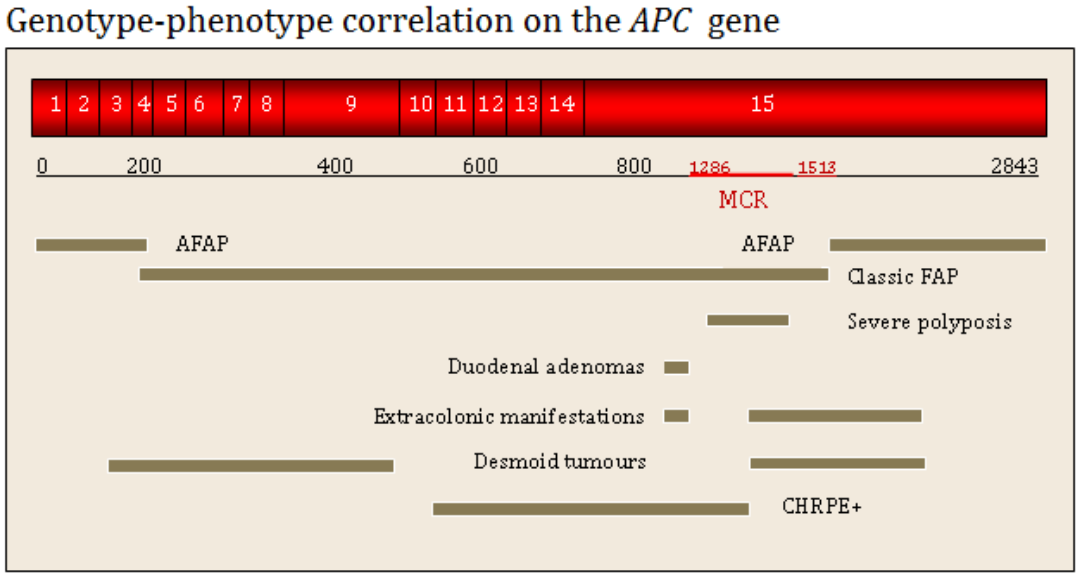

Figure 1-7 The location of mutant codons within the *APC* gene associated with the variation in polyp burden, development of extracolonic tumours in FAP patients. 1-15 represents the exons of *APC* gene and the mutation cluster region (MCR).

CHRPE have been seen in some cases of FAP and tend to occur in patients with mutations in codons 543 and 1309. Duodenal adenoma risk and extracolonic manifestations are highest between codons 976 and 1067. The second highest frequency of extracolonic developments occurs between codons 1310 and 2011 (Bertario *et al*., 2003). The higher risk for desmoid disease amongst FAP patients arises due to biallelic *APC* mutations, with one change occurring distal to codon 1399 of the *APC.* Codons between 1310 and 2011 are also associated with a six-fold increase risk in desmoid tumours and considered high-risk region compared to the low risk region 159 and 495. Desmoids tend to cluster in susceptible individuals, more in females than males and in patients with a positive family history of desmoids (Sturt *et al*., 2004, Bhandari *et al*., 2011). Furthermore, aggressive fibromatosis is marked by the presence of desmoid tumours, a rare condition in the general population. These tumours are benign, well differentiated, and overgrowths of fibrous tissue. Though these are slow growing tumours without any metastatic potential, they are locally invasive leading to organ damage and dysfunction.

There are a few germline mutations after codon 1700 which produce little or no phenotype (Muraoka *et al*., 1994). The clinical variant Gardner's syndrome is typically caused by mutations in the small region between codon 1403 and 1578 (Fearnhead *et al*., 2001) and is also associated with osteomas (tumours of the bone); jaw, extra teeth, and soft tissue tumours including lipomas (fatty tissue) and fibromas.

1.4.5 APC mutations in MYH-associated polyposis

Individuals with an autosomal recessive form of FAP were known as MYH-associated polyposis (MAP) syndrome are linked to biallelic mutations in the MutY homolog (*MYH*) gene (Al Tassan *et al*., 2002). Presently, there is much clinical variability in features of *MYH* mutation carriers, however MAP patients are characterised by multiple colonic polyps, have a milder disease than FAP, but more severe than AFAP and carry a high risk of CRC progression. *MYH* mutations are thought to cause cancer due to the accumulation of G:C to T:A transversions in *APC* suggestive of a defect in DNA repair system (Al Tassan *et al*., 2002; Seiber *et al*., 2003; Lipton and Tomlinson, 2004). Moreover, a few individuals have been diagnosed with CRC at 50 years or younger with few or no polyps (Wang and Cooper *et al*., 2004). Aretz and colleagues also noted between 4% and 25% of individuals with biallelic *MYH* mutations also show duodenal polyposis (Aretz *et al*., 2006).

MYH encodes DNA glycosylase involved in DNA damage repair via BER process. Oxidative damage can cause the mutant base 8-oxoguanine to be incorrectly included into DNA in place of guanine. *MYH* removes adenine residues that are incorporated opposite 8-oxoguanine which have a tendency of mismatch at these bases. Following replication in *MYH* deficient cells a thymidine can be incorporated opposite the adenine and consequentially lead to deficient glycosylase activity and therefore is a caretaker TSG which follows the 'just right' model.

1.4.6 The Adenomatous Polyposis Coli-I1307K gene variant

Other variants of the *APC* gene (first described by Laken *et al* in 1997) such as I1307K are thought to act as a low penetrance allele associated with an increased CRC risk in Ashkenazi Jews and to cause AFAP like phenotype (Groves *et al*., 2002). The frequency of heterozygosity for this allele is 6.1% and 10.4% in Ashkenazim healthy and CRC affected individuals, respectively. When stratified for family history, approximately 28% of all probands are carriers of the *APC*-I1307K variant (Laken *et al*., 1997). Several studies have reported an increased CRC relative risk ratio of 1.58 (95% CI: 1.21–2.07) for carriers of the *APC*-I1307K variant (Houlston *et* al., 2008). The missense mutation occurs in the MCR region of the *APC* at codon 1307, which is thought to occur due to a hypermutable region prone to slippage of the A_8 tract, responsible for replication-induced errors and gain/loss of an adenine resulting in a frameshift change. However, the A_8 is not observed in all polyps of these carriers and suggests additional genetic components may be involved.

1.4.7 Modifiers of Adenomatous Polyposis Coli

As previously mentioned, FAP patients with the same mutation in the *APC* gene can also show an intra-familial variation in polyp numbers and differences in disease occurrence (Su *et al*., 2002, Giardiello *et al*., 2000; Friedl *et al*., 2001; Houlston and Tomlinson, 2001). Hence, the position of the germline mutation alone cannot explain this variation in disease severity. The difference can be striking with some carriers with only a few polyps, while others may have more than 100 and still others may manifest a variety of neoplasias, including osteomas, sarcomas, and

carcinomas (Leppert *et al*., 1990; Spirio *et al*., 1992). Such diversity is thought to be due to environmental factors, especially diet which are highly likely to influence the onset of CRC. It is also suggested that these variations are probably due to genetic modifier genes, which may have a modulating influence on the outcome of *APC* mutation. This scenario has been repeated in mouse models of *APC*, which under laboratory controlled genetic and environmental conditions yield differences in polyp burden, thus substantiating the evidence for modifiers. In animals, modifier effects are usually attributed to the genetic background (distinct from the disease locus) and are inherited as polygenic traits leading to quantifiable differences in disease phenotype with quantitative trait loci (QTL) studies providing examples of genetic modifiers of *APC*.

1.5 THE USE OF MOUSE MODELS

1.5.1 Historical introduction

The concept of inbred mice developed from the popular hobby or 'fancying' of collecting and breeding of mice with different coloured coats, which started in Japan in the 1700s. The popularity grew into Europe by 1800s. William Castle, along with his student Clarence Little began research in 1902 and produced a series of seminal papers on coat colour genetics by buying mice from a local mouse fancier. This led to the development of 'lab mice' or the inbred strains that are used today. Little's first inbred mouse, dilute brown non-agouti (DBA) was developed in 1909 and the more famous strain C57BL/6 was developed in 1921. By 1929 Little had set up the Jackson Laboratory in Maine, USA, which is the most renowned centre for mouse genetics and plays a major role in research into mammalian biology.

Mouse models are powerful tools in cancer research as tumours arise under genetically and environmentally controlled conditions, the supply of tumours is plentiful and allows for significant analysis and follow-up studies on the influence of mutations which drive tumourigenesis. Mice are small in size, relatively easy to handle and to take care of. Their short generation time, rapid tumour growth and high tumour incidence make them ideal for cancer research. On the other hand, selective pressures used to perpetuate laboratory mouse strains have resulted in limited genetic diversity as a consequence of ancestry and many regions in the genomes of the laboratory mouse are highly correlated, resulting in spurious false associations and real ones being missed (Su *et al*., 2010). Outbreeding strategies

and designing of collaborative crosses to create genetically more distinct mice do offer increased allelic variation and can remove the possibility of population structure to prevent false positive associations. Such outbred mice also have low levels of linkage disequilibrium (LD) and high minor allele frequency averages (Yalcin and Flint, 2010). Recent years have seen an increase in genetically engineered mouse models (GEMMs) in research and preclinical studies. Their application to human complex genetics is not always straightforward, but with the use of a number of genetic and predictive tools it is possible to interrogate the findings and bind to relevant human genes and or proteins.

1.5.2 Apc$^{min/+}$ (Min Mouse) model of intestinal tumourigenesis

Many *Apc* GEMMs have been generated in attempt to accurately model the mutations common in FAP families and those implicated in sporadic CRCs of which *Apc$^{Min/+}$* is the first model used to investigate multiple intestinal neoplasia (*Min*). The *Apc$^{Min/+}$* mouse was identified following random mutagenesis with ethylnitrosurea (ENU) and described by Moser *et al*, (1990) (Moser *et al*., 1990). These mice were subsequently found to have a single change mutation from leucine to a stop codon at position 850 of *Apc* (Su *et al*., 1992). The *Apc$^{Min/+}$* mouse polyposis differs from the human condition as adenomas (>90%) are generally localised to the small intestine whereas in FAP patients, polyps are much more common in the colon (Caldwell *et al*., 2007; Silverman *et al*., 2002; Halberg *et al*., 2000; Moser at al., 1990). Adenoma formation in *Apc$^{Min/+}$* mouse is thought to occur within the first few weeks of life, with a full complement of adenomas being attained by approximately 9 weeks of age (60–67 days of age). After this time, adenoma

multiplicity is thought not to change, although adenoma size may still increase. *Apc*$^{Min/+}$ mice become anaemic and have a shortened life span of roughly 120 days (Moser AR., 1990; Shoemaker *et al*., 1995; Chiu *et al*., 1997).

The polyp multiplicity in *Apc*$^{Min/+}$ mouse is greatly influenced by genetic background, and was discovered as a result of linkage studies to map the *Apc*$^{Min/+}$ mutant. Mice heterozygous for *Apc* mutation developed approximately 29 small intestinal polyps when on the C57BL/6 background. William Dove and colleagues performed a series of backcrosses and discovered that the *Apc* phenotype is fully penetrant a characteristic shared in human FAP patients. In addition, backcrosses between F1 offspring from C57BL/6 (B6) *Apc*$^{Min/+}$ mice with AKR/J strain (AKR) or castaneus (CAST) mice demonstrated a significant decrease in polyp number as well as life span (Dietrich *et al*., 1993). The difference in polyp number suggests that certain inbred strains carry modifier genes that can reduce or increase polyp multiplicity in *Apc*$^{Min/+}$ mice.

1.5.3 Modifiers of Min (Mom)

Tracing the inheritance of genetic markers in the backcross (BC) by QTL analysis enabled the first of the modifiers (Table 1-4), modifiers of *Min* (*Mom1)*, to be localised to the distal region of mouse chromosome 4. This was further refined to a 4cM interval on distal chromosome 4 using congenic strains (Gould *et al* 1996b). Additional analysis of inbred strains showed that certain inbred mice carried the *Mom1* resistant (*Mom1*^{R}) alleles; AKR, CAST, SWR/J, DBA/2J and BALB/c, while B6, BTBR and 129/SvJ each carry a *Mom1* susceptible (*Mom1*^{S}). *Mom1* is estimated to

account for approximately 50% of polyp number variation and is classed as a semi-dominant modifier of intestinal polyposis, and suggests that other modifier genes with weaker effects exist (Dietrich *et al*., 1993, Gould *et al*., 1996a). Candidate gene analysis identified that secretory type II non-pancreatic phospholipase A2 (*Pla2g2a*) mapped to the same region as *Mom1* and showed 100% concordance between *Pla2g2a* genotype and *Mom1* phenotype in mouse strains (Dietrich *et al*., 1993, MacPhee *et al*., 1995).

Table 1-4 Summary of selected modifiers of *Min* phenotype. TSG=tumour suppressor gene, LOH=loss of heterozygosity, RI=recombinant inbred, B6=C57BL/6J, MA=MA/MyJ, CAST=CAST/EiJ, BTBR=BTBR/Pas and AKR=AKR/J mice

Mom	**Location**	**Phenotype/comments**	**Candidate gene/marker**	**References**
Mom1	Mapped to 5cM interval to distal chromosome 4, syntenic to humans 1p35-36 (high frequency for LOH)	Semi-dominant allele Control 50% genetic variation in 3 different strains (AKR, CAST and MA)	*Pla2g2a* and D4mit64 - involved in phospholipid metabolism and inflammatory responses	MacPhee *et al*., 1995
Mom2	Mapped to distal chromosome 18 which is syntenous to human 18q	Resistant allele. 8-11 fold reduction in polyp multiplicity	*Atp5a1*. 4bp duplication. excluded Smad genes, *Dcc* and *Tcf4*	Silverman *et al*., 2002
Mom3	Mapped to chromosome 18 – linked to Apc	Broad range of polyps (39-494) median of 107. Polyp variation segregated into 6 RI lines, also linked to pregnancy	-	Haines *et al*., 2005; Suraweera *et al*., 2006
Mom7	Mapped to chromosome 18, syntenic to human chromosome 5. 7.4Mbp interval	2 alleles of *Mom7*- enhancer increases intestinal tumours by 3 -fold (in BTBR and AKR) and suppressive effects in B6	-	Kwong *et al*., 2007

High levels of *Pla2g2a* expression were detected in the intestines of mouse strains resistant to polyp formation, such as AKR and CAST, while extremely low levels were detected in the susceptible B6 strains (MacPhee *et al*., 1995). Resistant *Mom1*^{R} allele carried a wild type *Pla2g2a*, while susceptible B6 strain failed to suppress polyp numbers showed that *Pla2g2a* was disrupted by a frameshift mutation in exon 3, resulting in an inactive gene product, lacked in enzymatic activity and carried the null *Pla2g2a* gene (Gould *et al*., 1996a, Gould *et al*., 1996b, Kennedy *et al*., 1995, MacPhee *et al*., 1995). This data strongly implicated that *Pla2g2a* is a dominant inhibitor of polyp multiplicity in *Apc*$^{Min/+}$ mice. Transgenic mice were constructed where the *Mom1*^{R} AKR allele of *Pla2g2a* was introduced into a susceptible B6 background, the *Apc*$^{min/+}$ offspring over expressed the *Pla2g2a* cosmid transgene and showed evidence that the *Pla2g2a* gene reduces the polyp development by two-fold with a significant reduction in polyp size (Cormier 2000). Also, the reduction in polyp number was not restored to wild-type levels, which suggests either a failure to restore transgene expression to wild type, or the *Mom1* phenotype is the result of several closely linked genes. This would indicate, therefore, that wild type copy of the *Pla2g2a* gene could greatly reduce polyp multiplicity in *Apc*$^{Min/+}$ mice, indicating *Pla2g2a* gene encodes the *Mom1* locus and to act as a modifier.

The homologous human *Pla2g2a* locus has been mapped to the human chromosome 1p35-p36 (Praml *et al*., 1995, Riggins *et al*., 1995, Tomlinson *et al*., 1996). This region is frequently found to be deleted with LOH in various human cancers including CRC, as well a large proportion of sporadic colorectal tumours (Praml *et al*.,1995; Dobbie *et al*., 1997, Tomlinson *et al*., 1996). Human functional studies did not confirm *Pla2g2a* as a major modifier of CRC risk in humans because polymorphic

variation did not appear to account for significant variation in susceptibility to CRC (Riggins *et al*., 1995; Tomlinson *et al*., 1996). Nevertheless, the detection of *Pla2g2a* has been informative, as it plays a part in the synthesis of prostaglandin pathway. Asprin and sulindac and other NSAIDs markedly reduce susceptibility to polyps probably by interfering with *Cox-2* activity within this pathway (Torrance *et al*., 2000; Takeda *et al*., 2003). The involvement of *Cox-2* is supported by the observation that $Apc^{Min/+}$ mice deficient for *Cox-2* have reduced polyp numbers (Oshima *et al*., 1996). Although, *Mom1* did not identify a gene that modulates the susceptibility of CRC in humans, it has revealed a pathway that has led to pharmacological advancements in treatment and management of intestinal tumourigenesis,

Likewise, a second locus, *Mom2 (Mom2)*, can suppress ~90% of polyps in $Apc^{Min/+}$ mice and was traced to a B6 'founder male' who had been mated with DBA/2J female (Silverman *et al*., 2002). $Apc^{Min/+}$ males which do not carry the $Mom2^{R}$ mutation exhibit on average 90 polyps and a 95% incidence of colon polyps by 6 months of age. The effects of reducing polyps in the small and large intestine are stronger in *Mom2* locus than the *Mom1* locus. Genetic linkage studies of the *Mom2* locus resides within 10cM interval and is syntenic with human chromosome 18q21 and 18q23 (Silverman *et al*., 2002, Radice 2000). Chromosome 18q21 frequently undergoes LOH in human CRC (Takagi *et al*., 1996) and family of SMAD genes have been identified in this region, two of which have been shown to result in intestinal neoplasia; SMAD2 and SMAD4 (Koyama *et al*., 1999, Miyaki *et al*., 1999). *Atp5a1* gene was also identified following a spontaneous mutation in this gene in mice (Baran *et al*., 2007), but not observed in human CRC cell lines (Seth *et al*., 2009).

The *Mom3* modifier has been mapped to the proximal region (within the first 25cM) of chromosome 18 in an interval closely linked to the $Apc^{Min/+}$. This modifier was detected following selective breeding to establish two recombinant strains that demonstrated a high and low polyp range. However, the high scoring line had an increase in polyps (range 289-494 polyps per mouse) and a shorter life span (Haines *et al*., 2005). The lack of additional polymorphic markers, along with the unknown genetic background prevented further positional refinement. The significant variation was probably caused by strain contamination, with the B6 not on the supposed B6 background. In a separate study *Mom7* locus mapped to a similar location as *Mom3*, located within the first 7.4Mb of chromosome 18 (Kwong *et al*., 2007). There are probably 5 segregating alleles, which amongst the different inbred strains. It has been speculated that both Mom3 and Mom7 may influence LOH of distal elements of chromosome 18, but whether these represent the same underlying modifier requires resolution by complementation testing and a thorough candidate gene search through sequenced regions of the mouse genome. The nature of any candidate genes and or functional polymorphisms remains unclear for both *Mom3* and *Mom7* (Haines *et al*., 2005; McCart *et al*., 2008).

1.5.4 Genetically engineered Apc mice

Following on from the Apc^{Min}/+ mouse, mouse models with various mutations in the Apc gene have been engineered to reflect the mutations found in FAP kindreds and sporadic CRC. A summary of the different Apc mouse models and their polyp burdens are shown in Table 1-5 and Figure 1-8. The quantifiable nature of Apc mutation phenotypes and the use of different inbred background provides a suitable

way to investigate the susceptibility loci of moderate to low penetrance and link these observation to CRC initiation and progression in humans (McPeek 2000, Mott *et al*., 2000; Johnson and Todd 2000).

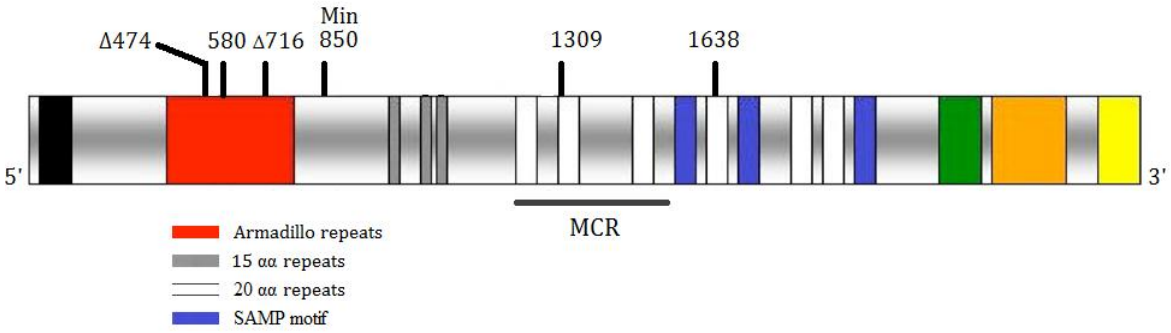

Figure 1-8 Genetically engineered mutations used to create adenomatous polyposis coli mouse models used in the study of intestinal polyposis. The bars represent induced mutations and numbers indicate where the codon affected and Apc protein is truncated.

The histology of intestinal tumours between $Apc^{Min/+}$, $Apc^{\Delta 716/+}$ and $Apc^{1638N/+}$ are similar but considerable differences in tumour numbers show that the genetic mutation affects the tumour burden despite being on the same genetic background. However, localised affects as demonstrated when the targeting cassette is not excised from the $Apc^{\Delta 716/+}$, may influence gene expression up and downstream of Neo (Colnot *et al*., 2004). Furthermore, Oshima *et al* noticed failure in tissue building with the additional layer of villous epithelium indicated that this may be responsible for the increase in adenoma numbers (Oshima *et al*., 1996). These data suggest that other factors that may account for polyp variation may exist, though conclusive evidence is not present.

Table 1-5 Tumour multiplicity models in adenomatous polyposis coli mice: comparison to the phenotype observed in the classical familial adenomatous polyposis model *Apc*$^{Min/+}$. SPF= specific pathogen free, ACF=Aberrant Crypt Foci, AA=amino acid, MCR=mutation cluster region.

GEMM	Mutation	Polyp burden	Histology/pathology of neoplasia	References
Apc$^{Min/+}$ Armadillo repeat	Truncating mutation at codon 850. Chemical mutagenesis; ENU induced	~30 ->100	Polypoid, papillary and sessile polyps. Cystic crypts no ACF 10% incidence of mammary tumours	Moser *et al*., 1990
Apc$^{1638N/+}$ SAMP repeats	*Apc* protein truncated at codon 1638 Neomycin inserted into exon 15	<10	Colonic polypoid hyperplastic lesions. Intestinal tumours with infiltration into mucosa and submucosa. Elongated colonic crypts. Desmoids and spontaneous ACFs	Fodde *et al*., 1994 and Smits *et al*., 1998
Apc$^{\Delta 716/+}$ Armadillo repeat	*Apc* protein truncated at codon 716. Neomycin inserted in antisense into exon 15	~300	Nascent polyps, covered by normal epithelium of a single villus. Dysplastic crypts. No colonic ACF	Oshima *et al*., 1996
Apc^{580S} Conditiona l allele	frameshift mutation at codon 580. Intronic LoxP insertions around exon 14 and exposure to adenovirus encoding Cre recombinase	~5	Adenomas localised near to the anus	Shibata *et al*., 1997
Apc$^{1638T/+}$ SAMP repeats	*Apc* protein truncated at codon 1638. Hygromycin inserted in *Smal* site in sense orientation into exon 15	Tumour/pol yp free	Absence of preputial glands and nipple-associated cutaneous cysts associated with abnormal follicular development	Smits *et al*., 1997
Apc$^{\Delta 474/+}$ Armadillo	Insertional target vector to duplicate	~>100 similar to	Sessile with central depression. 18.5% incidence	Sasai *et al*., 2000

repeat	codons 7-10 (7-8-9-10-7-8-9-10) and produce frameshift at codon 474	*Apc*$^{Min/+}$ and *Apc*$^{1309/+}$ phenotypes	of mammary adenomas; adenoacanthomas (2cm formed between 3-5months. Hyperproliferation of intestinal glands	
Apc$^{1309/+}$ 15AA repeats (MCR)	Frameshift mutation at codon 1309	~35-100	Similar polyps to *Apc*$^{Min/+}$ (throughout the duodenum to colon)	Niho *et al*., 2003
Apc$^{\Delta 14/+}$	Frameshift at codon 580 following deletion of exon 14 using Cre-loxP	~65 More severe than *Apc*$^{Min/+}$ Earlier death and rectal prolapse observed from 4 months	Increase in polyp numbers in animals housed in SPF a condition, which was also associated with, ACF and rectal prolapse. Tubular adenomas and invasive carcinomas observed in >12 month old animals. Shift in distal and colon adenomas	Colnot *et al*., 2004
Apc^{neoR} *Apc*^{neoF} Full-length hypomorphic alleles	Neomycin inserted into an enhancer site of intron 13 in the reverse (R) and forward (F) orientation	1 and 0.3	Dysplastic polyps and reduction of 50 and 19% of tumour incidence within 15 months of *Apc*^{neoR} and *Apc*^{neoF}, respectively	Li *et al*., 2005
Apc^{580} Armadillo repeat	Keratin 4, promoted excision of exon 14, frameshift at codon 580 and truncated protein at 605	~120 Similar phenotype to *Apc*$^{\Delta 14/+}$	Intestinal polyposis with extracolonic manifestations in the skin, thymus and dental	Kuraguchi *et al*., 2006
Apc $^{\Delta 15/+}$	*Apc* flanked by LoxP sites at codon 15 creating a truncated protein at codon 667	~185	Small (2mm) polypoid or flat adenomas, 77% developed in small intestine, 1-2 cutaneous cysts and infrequent desmoids	Robanus-Maandag *et al*., 2010

However, Apc mouse models have provided considerable insight into the role of the Wnt pathway components and thus elucidating the molecular pathways essential in phenotypic diversity. *Cyclin D1*$^{-/-}$ inducible *Apc* deficient models showed a reduction in adenoma burden, indicating *Cyclin D1* is not involved in tumour initiation but tumour progression. Compound Apc mouse models have shown that *Myc* plays an important part in the early *Apc*-driven tumour initiation step; inducible loss of both *Apc* and *Myc* demonstrated how the loss of *Myc* abolishes the crypt development associated with *Apc* deficiency (Sansom *et al*., 2007). Furthermore, *Apc*$^{Min/+}$ mice have been used to show how *Apc* can be inactivated by epigenetic silencing via promoter methylation of RAS association domain family 1 (*RASSF1A*) and its subsequent loss mediates degradation of β-catenin and leads to increased and accelerated intestinal tumourigenesis (Weyden and Adams, 2007).

In summary, the *Apc*$^{Min/+}$ model has identified phenotypic aspects specific to the strain background, FAP related genotype-phenotype and the involvement of the Wnt signalling pathway. It is reasoned that the *Apc*$^{Min/+}$ mouse model, on different genetic backgrounds and using controlled environmental exposures, is a very useful tool to isolate modifiers that modulate the number of adenomas appearing in the mouse intestine. QTL analysis using linkage based on the *Apc*$^{Min/+}$ mouse crossed with a mouse with discordant genetic and traits features is a well-established protocol for quantitative trait genetics leading to the identification of equivalent genetic elements in humans.

1.6 QTL AND LINKAGE ANALYSIS

1.6.1 Historical introduction

The Laws of inheritance (Laws of segregation and independent assortment) were published by Gregor Mendel (1866) and remained largely ignored until 1900 when the mode of inheritance of traits with strong allelic effects was studied by Carl Correns, Hugo de Vries and Eric von Tschermak-Seysenegg. Francis Galton and Karl Pearson developed statistical foundations to quantify heritability including correlation, regression, standard deviation, principle component analysis and partitioning of genetic variance into additive and non-additive components (Galton 1869; Galton 1889; Pearson 1990).

The discovery of gene mapping is credited to Bateson and colleagues for their work on the sweet pea (1905) (Bateson *et al*., 1906), which led Alfred Sturtevant to construct the first genetic map covering six loci in Drosophila (Sturtevant 1913). East in 1910 (East 1910) proposed the concept of quantitative traits and, subsequently, QTL mapping, was envisaged by Sax (Sax 1923) after Sir Ronald Fisher (1918) who described the combinational action of multiple genes. Penrose (1935) described the expected frequency allele sharing of affected sib pairs (ASPs). Haldane and Smith converted the occurrence of recombination events or recombination fraction (rf) into physical distance and by estimating the rf between markers and QTLs which led to the idea behind QTL mapping. Newton Morton tested linkage analysis in humans in 1955 using the logarithm of the odds (LOD) score method (Morton 1955). By 1961 Thoday had conceptualised the idea of

quantitative traits as a complex 'polygenic' model. Following the discovery of molecular markers, the very first genome map was created in maize and tomato using restriction fragment length polymorphism (RFLP) (Helentjaris *et al*., 1986) and the idea has since been used in families to trace quantitative inheritance of complex traits. Risch and Merikangas (1996) demonstrated that complex traits involving several predisposing variants of small effects are difficult to unravel with linkage (Risch and Merikangas, 1996) and so recent research has to turned to investigate the common disease/common variant (CDCV) hypothesis summarised by Peto and Houlston (Peto and Houlston, 2001) and Pharoah *et al* (Pharoah *et al*., 2002). They state that the disease phenotype results from the aggregate effects of modifiers and low penetrance that individually confer a small risk, but in combination act additively or multiplicatively. It is believed that if such alleles occur at high frequency within a population and are involved in multiple biological pathways this may explain some of the 'unknown' genetic risk for heritable cancers.

1.6.2 Linkage and association studies

The study of complex traits can be mapped in healthy and affected human populations, or using animal models with linkage or association studies. While linkage studies seek to identify loci that co-segregate with the trait within families, association studies seek to identify a causal variant which is associated with the phenotype are usually conducted at the population level. The two are not mutually exclusive but complimentary to one another and both have been used in human studies. Work on familial aggregation and patterns of inheritance date back to Mendels first hybridisation of peas experiments and these have dominated linkage

studies. The human genome project and the availability of genetic markers have since enabled widespread mapping and association studies based on populations leading to the discovery of susceptibility variants.

1.6.3 Linkage using human sib pairs

Sib pair linkage studies have been a widely used technique to map modifier genes and qualitative traits in pedigreed populations (Rogus and Krolewski 1996; Fulker *et al*., 1996; Kruglyak and Lander 1995). These studies are based on sharing and the probability of identify by descent (IBD) between siblings that express similar quantitative trait values as a phenotype. The power to detect linkage depends on the deviation of allele penetrance of loci contributing to the genetic trait and the deviation from sharing IBD from its null expectation and also discordant sib pairs (DSPs). The prospects further depend on the number of loci and whether loci act multiplicatively or additively. The selection and size of family sampling is important, larger families present higher costs and higher possibility of siblings contributing different information. This is further implicated by considerations of late disease onset, large number of phenocopies, gene-environmental interactions and locus heterogeneity. Though sibships and smaller families which are easier to identify and recruit are often preferred, it is estimated that 500-1000 families are required to produce adequate power to detect genetic risk factors that are consistent with population and familial prevalence of CRC (Kerber *et al*., 2008). As CRC is associated with environmental factors, higher sample sizes are required which represent a major recruitment challenge and in turn accrue higher costs, taking many years before any conclusions can be drawn.

1.6.4 Transmission disequilibrium test - family based quantitative trait loci mapping

Family based designs to study association are based on trios (two parents and one affected child) and hypothesis testing that genetic association is present between marker and disease (HA) against the null hypothesis (H0) that states no association between marker and disease exist (Rubenstein 1981). Later Ott (1989) added a second alternative hypothesis that both linkage and association must exist, followed by the third null hypothesis by Spielman *et al* when they introduced transmission disequilibrium test (TDT) method (Ott, 1989; Spielman *et al*., 1993). They tested for linkage in a setting of known association and therefore (HO: association but no linkage). There has been much success with the analysis based on TDT and this has been extended in context of quantitative mapping in extended pedigree, nuclear families and mapping studies. This has also formed the basis for haplotype association mapping (Spielman and Ewens, 1998). The test assumptions are based on the parental genotypes being accurate and correct and therefore incorrect genotyping can lead to false inferences (Hirschhorn and Daly, 2005). The other hampering feature includes the lack and reliability of statistical methods to detect and test multiple covariates, genes and interactions under a family-based design.

1.6.5 Genome wide studies based on Association

Modern complex trait mapping in humans utilises LD based on genome wide association (GWA) studies involving the correlation of several hundred thousand markers spaced throughout the genome with trait variation in the population based

samples. The scalability is much greater than with genome wide QTL scans. Risch and Merikangas were the first to note that small genetic effects could be detected with greater power by association analysis and proposed GWAs (Risch and Merikangas, 1996). Since the Human HapMap and the development of high-resolution LD maps and single nucleotide polymorphism (SNP) chips, GWA studies have so far detected 21 CRC risk associated loci (Table 1-6), individually these are estimated to contribute to less than 1% of the CRC but may have up to 6% collective CRC risk (Tomlinson *et al*., 2007; Tomlinson *et al*., 2008; Tenesa *et al*., 2008; Dunlop *et al*., 2012). Most CRC risk variants map to regions of the genome that do not contain known genes or reside in intronic, intergenic regions, however, a small proportion have been mapped at or close to genetic components which are involved in TGFβ signalling pathway and bone morphogentic protein pathway; SMAD7, GREM1, BMP2, RHPN2 and LAMA5. These findings suggest variants could be in strong LD, involved in complex epistasis and or additive effects with neighbouring and distant genes and risk mechanisms. One promising approach is to delineate gene and protein expressions as quantifiable traits and use these in QTL mapping strategies such as expression QTL (eQTL) analysis. Such studies have provided an abundance of genome wide patterns at a transcriptional level and several studies have shown an overlap between trait associated and eQTL variants suggesting possible interactions between near (*cis*) and distant genes (*trans*) via eQTL (Loo *et al*., 2012; Nica *et al*., 2010).

Table 1-6 The colorectal cancer susceptibility risk variants detected by genome wide association studies to date. bp=position, MAF= minor allele frequency extracted from 1000 Genomes and variants mapped to NCBI dbSNP Build 134. Ancestral allele in bold

Locus	SNP / position (bp)	Candidate /closest gene	Major / minor allele	MAF	Reference
1q41	rs6691170 22045446	Intergenic	G/T	0.25 (T)	Houlston *et al.*, 2010
3q26.2	rs10936599 169492101	MYNN	C/T	0.30 (T)	
6p21	rs1321311 36622900	CDKNA	C/A	0.27 (A)	
8q23.3	rs16892766 117699864	EIF3H (intron)	A/C	0.08 (A)	Tomlinson *et al.*, 2008
8q24.21	rs6983267 128413305	Intergenic	G/T	0.43 (T)	Tomlinson *et al.*, 2007
	rs10505477 128407443	intergenic	A/G	0.46 (G)	
	rs7014346 128424792	POU5F1B	G/A	0.32 (A)	
9q24.1	rs719725 6365683	Intergenic	A/C	0.31 (C)	Le Marchand *et al.*, 2007
10p14	rs10795668 8701219	Intergenic	G/A	0.26 (A)	Tomlinson *et al.*, 2008
11q23.1	rs3802842 111171709	ORF 93	C/A	0.31 (C)	Tenesa *et al.*, 2008
11q13.4	rs3824999 74345550	POLD3	T/G	0.38 (G)	Dunlop *et al.*, 2012
12q13.13	rs7136702 50880216	Intergenic	T/C	0.49 (C)	Houlston *et al.*, 2010
	rs11169552 51155663	Intergenic/ATF1	C/T	0.21 (T)	
14q22.2	rs4444235 54410919	BMP4 (intron)	T/C	0.45 (C)	Houlston *et al.*, 2008
15q13.3	rs4779584 32994756	GREM1	C/T	0.48 (T)	Tenesa *et al.*, 2008
16q22.1	rs9929218 68820946	CDH1 (intron)	G/A	0.24 (A)	Houlston *et al.*, 2008
18q21.1	rs4939827 46453463	SMAD7 (intron)	C/T	0.38 (T)	Tenesa *et al.*, 2008
19q13.1	rs10411210 33532300	RHPN2 (intron)	C/T	0.22 (T)	Houlston *et al.*, 2008
20p12.3	rs961253 6404281	Intergenic	C/A	0.29 (A)	Houlston *et al.*, 2008
20q13.33	rs4925386 60921044	LAMA5 (intron)	C/T	0.42 (T)	Houlston *et al.*, 2010
Xp22.2	rs5934683 9751474	SHROOM	T/C	0.37 (C)	Dunlop *et al.*, 2012

More often than not, GWA studies depend on extensive collaborative efforts and require large sample sets to overcome the low statistical power of associations, but the greatest challenge presented by GWA studies is to elucidate the influence these risk variants have on the chances of developing CRC. Other problems include refining the estimates of heritability and the mode of inheritance. Another key question asked is how best to identify networks of genes that act together to influence cancer susceptibility, which may provide more information to understanding the remaining trait variance and in particular the rare small-effect size variants (Dunlop *et al*., 2012; Houlston *et al.,* 2010). The most notable reason for these drawbacks is the inconsistent phenotype penetrance due to varied environmental factors and confounding susceptibility alleles that exist in human populations. In addition, it is not possible to analyse susceptibility alleles as quantitative traits in human studies even if they are of sizeable effects. Experiments in mice have shown the additive, epistatic interactions affect tumour susceptibility (Demant *et al*., 2003) and using such models in a QTL genome wide scan may in part reduce some of the major complexities presented by human GWAS.

1.6.6 Linkage using quantitative trait loci mapping in mice

QTL mapping in experimental crosses involves comparing means of progeny inheriting specific parental alleles and is the main focus of the work carried out here. This is a much simpler and more powerful approach and is supported by many statistical methods (Kruglyak and Lander 1995). The main principle goal of QTL genetics in mice is to perform statistical analysis of the association between the phenotype and genotype to identify genomic regions that may harbour candidate

gene(s) and then using the close homology between mouse and human genomes to subsequently search for human homologs (Bedell *et al*., 1997; Balmain 2002) further mention of comparative analysis between mouse and humans in mentioned under section 1.6.10 of this thesis. QTL mapping using inbred mice for CRC relies on the genetic analysis between a highly susceptible strain, which carries a Mendelian mutation and is crossed to a resistant strain that confers phenotypic differences. These are either spontaneously, chemically or irradiation induced to enhance genetic predisposition and recombination events allow the segregation of susceptible and resistant alleles affecting different phenotypic outcome (Dragani *et al*., 1995, Balmain 2002). The high degree of polymorphism at microsatellite loci among inbred laboratory strains allows for the construction of genetic linkage maps.

Genetic mapping is determined by the recombination events which occur between sister chromatids during meiosis and the linkage is based on the frequency between two markers with which they recombine. The closer they are to each other, the less likelihood of chromosomal crossover will occur between them and genes far apart are unlinked, and assort independently. The rf is the probability for recombination to occur ranging from 0 for completely linked to 0.5 for unlinked loci and is used to calculate the genetic distances. The backcross used in this study expects rf to be <0.5. By using markers that co-segregate with the phenotype can be an indication that the allele responsible for the phenotype is located close to the marker. Mice are phenotyped for the tumour number, location, size, and genotyped at each marker loci and the correlation between the distribution of marker alleles and susceptibility phenotypes allows the detection of chromosomal locations which may confer susceptible loci.

1.6.7 Quantitative trait loci mapping approaches

The basic laws of QTL mapping were formulated by Soller in 1979 (Soller *et al*., 1979) and Lander and Botstein strengthened these in 1989 (Lander and Botstein, 1989) and are referred to as parametric analysis. The main mathematical rule that underlies QTL mapping is that the power to detect a QTL is proportional to the size of its effect, which is the difference between the alleles that are present in the cross and expressed as a portion of the total variance.

Techniques include single marker mapping, interval mapping (IM), composite interval mapping (CIM) and multiple QTL mapping (MQM). Single marker regression is the simplest method for the analysis of QTL mapping data by considering each marker individually, splitting the individuals into groups, according to their genotypes at the marker and comparing the groups' phenotypes averages using analysis of variance (ANOVA) or *t* test. Evidence for a QTL is measured by LOD score: the $\log_{10}$ likelihood ratio comparing the hypothesis that there is a QTL at the marker to the hypothesis that there is no QTL anywhere in the genome; the null hypothesis. This process offers statistical simplicity, but fails to provide separate estimates of the QTL location and QTL effects between markers. The QTL location is confounded and based on the marker which indicates the highest difference between genotype averages and the QTL effect will be less than the true QTL effect. The most important disadvantage of single marker is the limited ability to separate linked QTL and to assess possible interactions among QTL. However, in the absence of missing marker information, single marker regression is robust and highly suitable approach.

1.6.8 Interval, composite and multiple QTL mapping

In IM, the recombination model is based on Hidden Markov model (HMM) and forms the core methodology for probabilistically connecting unobserved genotype to observed genotype. Interval mapping is similar to single marker regression, instead the target locus is replaced by a position (hypothetical QTL Q) in a target interval where the expected QTL genotypes are calculated using the flanking markers. IM fixes (Q) between marker intervals, calculating the rf at incremental steps and thus generating a genetic map (Lander and Botstein, 1989). This method allows for the inference of QTLs between markers and provides better estimates of the QTL effects addressing the issues set out by single QTL mapping. Furthermore, IM accounts for missing data at markers by considering data from the nearest flanking markers for which genotype is available, making it more robust to genotyping errors than simple marker mapping and has become the most commonly used mapping technique. The probability that each individual has the genotype AA or AB at Q is calculated based on four conditional probabilities central to the rf between the left flanking marker (L) and Q and Q and the right flanking marker (R) Table 1-7.

The expression to calculate the genotype probabilities of Q depends on genetic mapping function; Kosambi, which is the same as human inference, and assumes intermediate interference: where the chance of one or more break points in an interval depends on the occurrence of break points in another interval (Kosambi 1944). Carter-Falconer mapping function is the closest to that of a mouse, which shows extreme crossover inference with crossovers seldom being separated by <20cM. Other mapping functions include Haldane (assumes no inference, therefore

allows for multiple crossovers) and Morgan (assumes complete interference and no multiple crossovers.

Table 1-7 Inference of QTL genotype probabilities given the marker genotypes at two flanking markers: left (1) and right (2) and markers separated by recombination fraction (rf) for individuals in a backcross.

Marker genotype		QTL genotype	
Left	Right	AA	AB
AA	AA	(1-r1Q)(1-r2Q)/(1-r12) probability of no recombination in both L-Q and Q-R in a non-recombinant interval	r1Qr2Q/(1-r12) probability of recombination in both L-Q and Q-R in a non-recombinant interval
AA	AB	(1-r1Q)r2Q/r12 probability of recombination between Q-R only in a recombinant interval	r1Q(1-r2Q)/r12 probability of recombination in L-Q only in a recombinant interval
AB	AA	r1Q(1-r2Q)/r12 probability of recombination in L-Q only in a recombinant interval	(1-r1Q)r2Q/r12 probability of recombination between Q-R only in a recombinant interval
AB	AB	r1Qr2Q/(1-r12) probability of recombination in both L-Q and Q-R in a non-recombinant interval	(1-r1Q)(1-r2Q)/(1-r12) probability of no recombination in both L-Q and Q-R in a non-recombinant interval

Once the genotype probabilities are determined, individuals are split into genetically identical genotype groups contributing to the phenotype (two for a backcross and three for intercross). The residual variance (σ^2) for each group will be entirely non genetic and due to measurement error, environmental variation and individual noise. Three simplifying assumptions are made; first, the residual variance is constant (homoscedastic) within each group: second it follows a normal curve distribution with different phenotype averages: and third, QTL act additively and the effect of QTL remains regardless of the genotype at the other loci. Deviation from additivity implies epistasis: in a statistical sense epistasis is measured in individuals

of a backcross, split into four groups according to their joint genotypes at two QTL and when the genotype at one QTL depends on the genotype of another QTL, the QTLs are said to interact.

The phenotype average and the phenotypic variance are estimated by a maximum likelihood function (calculates the observed/expected ratio) using the expectation and maximisation (EM) iterative algorithm. The E-step calculates the conditional probability that the individual is in QTL genotype groups given its marker data, phenotype and current estimates of residual variance and phenotype averages. In the M-step, the estimates are repeated using weights until they converge (and estimates stop changing). Once the maximum likelihood scores are calculated the LOD under the null hypothesis that there is no QTL anywhere in the genome is compared. Though IM is more powerful than single marker approaches, there is one major disadvantage with IM that it still only considers a single QTL interval and the possibility that the LOD profile can indicate multiple QTLs or ghosts QTLs. The one dimensionality of IM does not allow interactions between multiple QTLs to be considered.

Statistical approaches for locating multiple QTLs are presented with the ultimate challenge to search every position in the genome simultaneously for linked and or interacting epistatic QTLs. The statistical implementation of a simultaneous search is quite difficult due the large number of possible QTLs involved that can be involved along with the heavy computational demand. CIM introduced by Zhao-Bang Zeng in (1994) and multiple QTL mapping by Ritsert Jensen (1993) extend the idea of IM to include a subset of marker loci as covariates outside a defined window, with the

purpose to remove any association of linked QTLs and reduce the residual variance effects. The selection of markers to serve as covariates is the key problem and the markers closest to the QTL are the obvious choice. This approach is however restricted to a one dimensional searching and the multiplicity of epistatic QTL effects remains a challenge. Though, CIM procedures generally increase the resolution of IM by accounting for linked QTLs or false QTLs while reducing the residual variation. The ideas underlying CIM have influenced the development for MQM. MQM has three advantages: by controlling for QTLs of large effect, the residual variance may be reduced and so the better to identify QTLs of modest effect; the separation of linked QTL is achieved by comparing the fit of two-QTL model to the best single-QTL model in an exhaustive 2-dimensional search;, and finally epistasis can be considered via models that take multiple QTLs in to account simultaneously. Epistasis is considered by exploring 'models' using both a forward and backward elimination search. T he former starts with no QTLs in the model, then adds each QTLs beginning with a QTL with the highest LOD score. Additional QTLs are then selected on to give the greatest increase in LOD score. The reverse of the forward selection is backward elimination, which starts with all QTLs in the model, and then removes each QTL one at a time so as to give the smallest decrease in the LOD score. These processes involve testing at each step for improvements to an additive only and an additive-interactive QTL model to allow the identification and assessment of a subset of loci. This permits consideration of multiple acting QTLs that may control trait phenotype (Baum *et al*., 1970).

1.6.9 Statistical significance

Non-parametric bootstrapping resampling (Efron and Tibshirani 1993) and permutation resampling methods (Good 2005) are useful means to randomise the trait data for the purpose of evaluating the test statistic under a null hypothesis that tests for a true QTL. False-positives in QTL mapping indicate linkage merely as a result of chance within the genome and are generally high if the size of the study and data are large. Permutation testing is the random assignment of trait values to no more than one individual compared to bootstrapping method which differs by more than one individual might receive the same random trait value, changing the mean variances and trait summary information. Both techniques are computationally demanding and require resampling in excess of 1,000 times to generate precise correction factors (Churchill and Deorge 1994; Sen and Churchill 2001) and set stringent significance thresholds to avoid false positives. Nominal significance criteria of 99.8% or more for any single QTL are usually necessary to assure a genome-wide threshold significance at 5% genome error.

When a QTL is defined, the genomic regions that are identified are usually in range of 10-30 Mbp. Identifying the genes underlying QTLs is the biggest hurdle and a challenging task (Nadeau and Frankel, 2000; Korstanje and Paigen, 2002; Flint *et al.*, 2005). One of the common ways to increase precision toward QTL gene identification is to breed congenic mice. In this approach, breeding transfers the susceptibility allele at a particular locus from one strain on to the genetic background of the other. This process is time taxing, involves extensive breeding that significantly increases cost. Though congenic mice offer a more enhanced

mapping approach it is, however, not as successful as expected due to the reduction in genetic heterogeneity and segregating alleles often missing QTLs.

1.6.10 Comparative genetics.

Another approach to narrow the QTL region is to apply comparative genetics. Since the Mouse Genome Sequencing Consortium first published a high-quality draft sequence of the genome of C57BL/6J mouse in 2001 and researchers of the International Human Genome Project (HGP) subsequently sequenced the entire mouse genome (Waterston *et al.* 2002). Comparison of general features between the human and mouse genome showed that there are striking similarities; the estimated size (bp) of the human genome is calculated at 3 billion bps (over 46 chromosomes) and the mouse at 2.9 billion bps over 40 chromosomes with approximately the same number of genes (~30,000) and this provides an entry point into comparative genetics. Further to this, about 40% of the human and mouse genomes can be directly aligned with each other with 80% of human genes having equivalent orthologues in the mouse genome (Reymond *et al*., 2000; Gitton *et al*., 2002, Lander *et al*., 2001) (Figure 1-9). With this information to hand and the application of the vast comprehensive genomic data, comparisons of discrete segments and aligning homologous DNA from different species to discover biologically relevant genes is made possible.

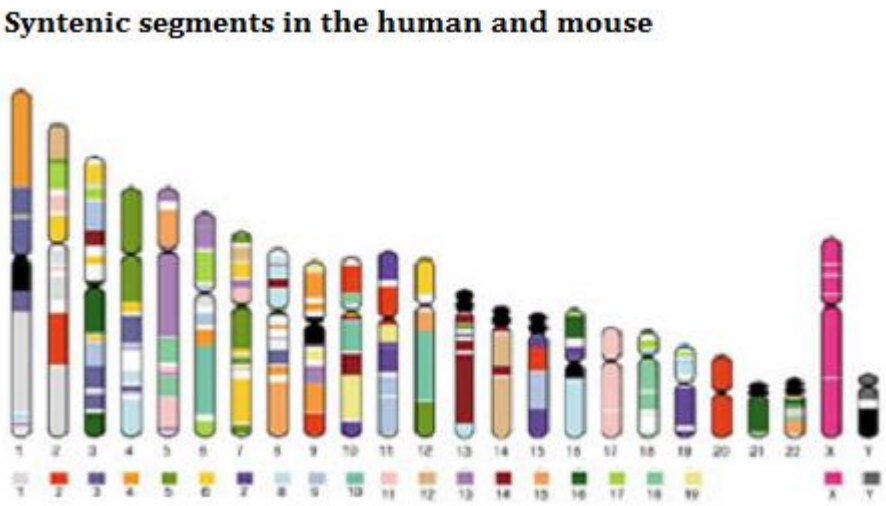

Figure 1-9 Conserved segments between the human and mouse chromosomes. A minimum of two genes whose order is conserved in the mouse genome are shown as block colours for e.g. human chromosome 20 corresponds entirely to a portion of mouse chromosome. (Image from International Human Genome Sequencing Consortium; Lander *et al*. 2001)

The integration of QTL analysis and genomics is crucial in order to identify underlying DNA sequences, which are involved in controlling trait phenotype. A single QTL can consist of more than a hundred genes and so meta-analysis combines data from different sources to assist in QTL candidate gene discovery (Goffinet and Gerber, 2000). QTL mapping studies carry wide applications to construct gene pathways in the elucidation of underlying causal genes; the application of functional genomics and classical bioinformatics will lend a hand in the biological interpretation of the quantitative genetic models in our work.

1.6.11 Tumour quantitative trait loci identified in mice

The use of mouse models and computational genetics in QTL has so far successfully been used to map resistant and susceptibility QTLs for tumours arising in many tissues such as the lung (*Slucs*), colon (*Css*), intestinal (*Mom1, Ssic1*) and skin (*Skts*) (Demant, 2003; Dragani, 2003, Ruivenkamp *et al*., 2002, Fijneman *et al*., 1996). Tumour susceptibility QTLs; *Pctrl* and *Mom1* loci were among the first of tumour susceptibility and resistance phenotypes identified by genome scans using RFLP and simple sequence length polymorphism (SSLP) markers, respectively. *Ptprj* is a candidate gene for CRC *susceptibilty to colon cancer 1* (*Scc1*) in the mouse where 50% of human cancers also show LOH of this gene (Ruivenkamp *et al*., 2002). More recently, Lei Quan and colleagues have shown common linked QTLs which exist in both lung and CRC and suggest that underlying genes could be identical or related (Quan *et al*., 2011). Galvan and colleagues support these findings with the loci *Lsktm1* which is presumed to modulate both lung and skin tumourigenesis. These data suggest that by comparing organ specific susceptibility genes greater knowledge of tumour susceptibly mechanisms can be determined (Galvan *et al*., 2012).

1.6.12 Radiosensitive QTLs identified in mice

Mouse QTL mapping approaches have been applied to dissect the genetic basis of radiosensitivity and tissue injury. Two radiation-induced pulmonary fibrosis (*Radpf1* and *Radpf2*) QTLs have been detected to reveal regions of irradiation induced lung responses: pneumonitis, alveolitis and fibrosis to chromosome 17 and

chromosome 1 respectively (Haston *et al*., 2002). Currently, no other radiation induced QTLs have been mapped in mice, and strain dependent contribution towards radiation induced intestinal damage has been investigated in only a few studies.

So it can be seen that QTL methodology is a powerful tool to identify genomic regions that are associated with traits. While the majority of the QTLs are large genomic regions, these studies continue to develop and enhance the understanding of complex traits, quantitative variation and also the complex interactions. Newer methodologies and software are being written to tackle issues such as QTLs and epistasis. Most importantly, the use of comparative genomics and bioinformatic techniques facilitated by databases has made it possible to narrow the search for candidate genes and QTL candidate genes. The combination of QTLs and bioinformatics should accelerate candidate gene identifications. After which comparisons can be investigated between mouse and human genomes due to the degree of homology, co-localisation and synteny of genes that exist on the same chromosome (Dermitzakis *et al*., 2002; Burgess-Herbert *et al*., 2008; DiPetrillo *et al*., 2005) and tested in human based studies and or functional studies for confirmation (DeBry and Seldin, 1996).

1.7 APPROACH TO THE STUDY

Predisposed genetic susceptibility and exposure to environmental factors contribute significantly to an individual and population's risk of CRC development (Lichtenstein *et al*., 2000). One key example involves the exposure of radiation which affects the cellular processes of DNA repair and chromosomal instability. Both are an intrinsic feature of cancer. Primary radiation responses show variation in humans (BEIR V 1990) and this scenario is repeated in mice populations (Followill *et al*., 1993). As radiation toxicity to the gastrointestinal system is a major side effect, observed in 6,000 individuals undergoing pelvic radiation therapy (Andreyev 2007). This suggests that the increased susceptibility may result from radiation-induced DNA damage and that deficiencies in DNA repairing may be responsible for the pronounced radiosensitivity.

It is well established that the heterogeneity of *APC* mutations lead to varying degrees of adenoma multiplicity and disease severity of CRC. Furthermore, inter-individual variation in polyp numbers is observed in human FAPs demonstrating *APC* modifier genes with different gradients of penetrance and effects thought to be involved in complex genotype-phenotype correlation. As previously mentioned the $Apc^{Min/+}$ mice mimics the human condition FAP and the germline mutation at nucleotide 2549, codon 850 leads to the onset of *Min*, a condition of numerous adenomas throughout the intestines, though mainly in the small intestine.

Our earlier studies, which compared the effects of irradiation on a population derived between $Apc^{Min/+}$ on C57BL/6J ($B6^{Min/+}$) and radiosensitive BALB/c strain,

showed two segments of chromosome 16 (D16Mit189 and *Prkdc*) to modify adenoma multiplicity within the small intestine (Degg *et al.,* 2003). The recombinant inbred line (Line I) on a B6$^{Min/+}$ background was chosen as the resistant line out of 6 with limited intra line variation in adenoma numbers (Haines *et al.,* 2000; Haines *et al.,* 2005). Polyp multiplicity was investigated in N2 *Apc*$^{Min/+}$ mice using a BC strategy involving B6$^{Min/+}$; donor parent of the dominant *Apc* transgene carrying the mutation at codon 850 to be transferred onto the BALB/c background. This mutation will generate the underlying trait of interest and produce variations in polyp multiplicity for the detection of possible modifiers involved in intestinal adenomas.

1.7.1 Breeding program and induction of polyps

Female BALB/c mice were obtained from Harlan United Kingdom Ltd. (Bicester, United Kingdom) and crossed with C57BL/6J (B6$^{Min/+}$) males from a recombinant inbred colony derived from a separate cross (Haines *et al.,* 2005). Briefly, a single B6$^{Min/+}$ male mouse (Imperial Cancer Research Fund, Clare Hall, South Mimms, U.K.) bred with two female C57BL/6J mice (Radiation and Genome Stability Unit, Medical Research Council, Harwell, U.K.). Subsequent intercrossing gave rise to six distinct polyp scoring recombinant lines showing interline variability in polyp numbers. The lowest polyp scoring line (Line I) was selected from the six recombinant *Apc*$^{Min/+}$ lines. N1 males carrying the *Apc*$^{Min/+}$ mutation were then BC to female BALB/c to produce N2 *Apc*$^{Min/+}$ offspring. The recombinant Line I carried the susceptible alleles of *Mom1* (*Mom1*^{S}) . Studies and confers a resistance to polyp multiplicity and was selected as the resistant strain. Studies involving crosses between the *Apc*$^{Min/+}$ mice

and other inbred mice showed fewer numbers of polyps were scored when crossed with BALB/c (Moser *et al.,* 1990) and suggests that BALB/c carried the dominant tumour resistance alleles of *Mom1* (*Mom1^{R})*. To eliminate the modifier effect of *Mom1^{S}* we used the principle of BC to ensure that only *Mom1^{R}* derived from BALB/c in the filial (F1) would be selected. These were crossed with the recurrent BALB/c parent ensuring that each N2 *Apc$^{Min/+}$* was positive for at least one copy of the BALB/c derived modifier of *Min* (*Mom1*) dominant resistance allele (*Mom1^{R}*). Therefore, any effects of genetic modification will be independent of *Mom1^{S}*. BALB/c is also susceptible to IR-induced tumours (Degg *et al.,* 2003) and carries two natural polymorphisms of *Prkdc* involved in DNA double strand break (DSB) repair (Mori *et al*., 2001; Yu *et al.,* 2001). In this setting, our BC scheme favoured the transfer the *Apc$^{Min/+}$* mutation onto the BALB/c background and thus the isolation of chromosomal region(s) susceptible to irradiation and, because of the nature of the cross, the N2 *Apc$^{Min/+}$* will consist of a random set of genes with 75% alleles from BALB/c (maternal parent) and 25% from B6$^{Min/+}$ (parent) allowing the investigation of BALB/c alleles in linkage with the *Apc* gene.

Irradiation was used to detect susceptibility loci, which co-segregate with BALB/c alleles that are possibly involved in DNA damage or repair following irradiation. The BALB/c mouse is model of sensitivity to radiation-induced damage as it is more susceptible to tumour development. The BALB/c carries two natural functional polymorphisms in the DNA repair gene of *Prkdc,* one of which shows a decrease in the DNA-protein kinase activity (Smith *et al*., 1999; Bailey *et al*., 1999; Yu *et al*., 2001; Degg *et al*., 2003) and leads to defective DNA repair of DSBs. The partial defects in *Prkdc* are also considered to be involved in the susceptibility to mammary

tumours and lymphatic neoplasia induced by irradiation in the BALB/c strain (Roderick, 1963). More importantly, there is sufficient evidence provided by Degg and colleagues that two segments of chromosome 16 (D16Mit189 and *Prkdc*) from BALB/c mouse interact with $Apc^{Min/+}$ by specifically enhancing IR-induced adenoma development in the upper part of the small intestine (Degg *et al.*, 2003), This observation forms the basis of our investigations.

143 N2 $Apc^{Min/+}$ mice at 2 days of age were split into two groups: 84 males and 58 females and received 2Gys of irradiation; and 111 N2 $Apc^{Min/+}$ mice (60 males and 51 females) were sham irradiated (0Gy) as the control group (Degg *et al.,* 2003).

1.7.2 Intestinal polyposis

Immediately after sacrifice the small and large intestine was removed and opened longitudinally. The intestinal digesta was washed from the mucosal surface with ice cold phosphate buffered saline (PBS), PH 7.0 and the small intestine tracts of all N2 $Apc^{Min/+}$ mice were separated into the following three sections; small intestine (SI) which was divided into two equal portions with the upper small intestine (USI) segment containing all of the duodenum and jejunum and the lower small intestine (LSI) segment containing the proximal and distal ileum. The third section was the ascending and descending part of the large intestine (LI). Surface intestinal polyps were counted using a light microscope and 0.5% methylene blue in 70% ethanol stain in sections of the irradiated and sham irradiated mice and recorded against their age and sex. Other data recorded included whether the mice suffered from rectal prolapse, mammary tumours and had rectal bleeding. No polyps were

identified in the oesophagus or stomach and are generally a rare occurrence in the *Apc*$^{Min/+}$ mouse.

By using tissue collected from 253 N2 mice and the phenotyped data (number of polyps) collected by Jackie Haines, we aim to further the work carried out by Degg *et al* and carry out a genome wide scan by increasing the number of genotyped data to uncover irradiation-induced QTLs through genetic mapping to identify modifiers involved in spontaneous and irradiation-induced polyp multiplicity. By applying this approach we aim to expand the possibility of identifying other modifiers involved in addition to the ones investigated by Degg *et al* on chromosome 16 and to establish to what extent these genes are involved. Furthermore, by looking at the different parts of the gut, we will investigate for the first time whether radiation sensitivities are the same along the intestine and if they share common loci.

The identification *APC* modifier genes and discovery of their roles in the causation of radiation induced carcinogenesis will ultimately be important for the development of methods for prediction of risk and radiotherapy for human cancers, based on their genotype. This also provides a rationale for the design and development of approaches for the treatment of radiation-induced late normal tissue injury and may reduce uncertainties for radiosensitive exposed groups. Therefore, the identification and characterisation of such modifier genes or susceptibility loci associated with IR-induced DNA damage repair pathways continues to be important.

1.7.3 Hypothesis

The hypothesis is that QTLs exist in the mouse genome in addition to those on chromosome 16 that control radiation induced adenoma multiplicity in the intestine of $Apc^{Min/+}$

To test the hypothesis I aim to carry out a genome wide QTL scan based on the correlation of genotype-phenotype data using the $Apc^{Min/+}$ mouse model to identify modifiers which may interact directly or indirectly with the *Apc* mutation. For this investigation new microsatellite markers were selected and genotyped to create a genome wide panel across the N2 mice. With the use of newly emerging bioinformatics and statistical techniques I plan to reduce the usage of animals in research and provide a methodical approach to identify underlying genetic determinants of spontaneous and radiogenic elements associated with polyp burden and CRC.

1.7.4 Aims

Thus, the key objectives of this thesis are to

Investigate QTL regions which control interactions of the main effects QTLs

- To confirm QTLs detected by Degg *et al* (Chapter 2, section 2.3.4)
- Identify new QTLs responsible for spontaneous and irradiation-induced polyposis (Chapter 2, section 2.3.4 and 2.3.6)
- To construct a gene model of quantitative traits in the different segments of the intestine (Chapter 2, section 2.3.5)
- To construct multiple gene model of quantitative traits involving main-effect QTL (Chapter 2, section 2.3.7)
- Haplotype analysis around detected QTLs (Chapter 2, section 2.3.9)

Investigate epistatic interactions of the main effects QTLs

- 2D, two QTL pair wise scan on all traits (Chapter 3, sections 3.3.1-3.3.5)
- To investigate *cis*-epistasis between QTL pairs (Chapter 3, section 3.3.3)
- To create QTL-objects and investigate the effects of QTL pairs (Chapter 3, sections 3.3.6 – 3.3.8)
- Explore multiple mapping models using forward and backward selection model (Chapter 3, section 3.3.9)

To identify candidate genes that are implicated in CRC mechanisms using genomic data and integrative bioinformatics tools;

- *In silico* sequence modelling based on BALB/c specific SNPs and predictive SNP analysis (Chapter 4, section 4.3.1)
- Assess the profile of all SNPs within genomic regions (Chapter 4, section 4.3.2)

To test and validate underlying genes to further select target genes using

- Sequencing all target genes and nsSNPs (Chapter 5, section 5.3.1)
- mRNA expression of targets in wild type parental mice (Chapter 5, section 5.3.2)
- Literature based evidence summary of candidate genes (Chapter 5, sections 5.4.1 & 5.4.2)
- Discuss possible mechanisms of cancer risks (Chapter 6)

CHAPTER 2 GENETIC MAPPING OF QUANTITATIVE TRAIT LOCI FOR MOM RADIATION-INDUCED POLYP SUSCEPTIBILITY OF THE MOUSE INTESTINE

2.1 INTRODUCTION

This chapter describes the work carried out to identify additional and novel candidate regions in a genome wide approach using tissue from the pre-existing intraspecific BC; [**BALB/cByJ x C57BL/6J *Apc*$^{Min/+}$] x BALB/cByJ**) to extract DNA, perform genotyping and analysis using microsatellite markers to achieve genome wide coverage. In further analysis of the the backcross using interval mapping (IM) and permutation testing I will search for genomic regions that affect polyp multiplicity in irradiated mice to define modifiers as *Mom* (Modifier of *Min*), radiation-induced polyp susceptibility (*Mrip)* loci and also include the dissection of loci based on the different segments of the mouse intestine using a panel of N2 *Apc*$^{Min/+}$ mice. Similarly I will search for spontaneous modifiers of *Apc* using the sham irradiated which will also be used to compare against any *Mrip* loci identified.

The identified QTL regions are highly likely to contain genes of susceptibility to polyp multiplicity in the *Apc*$^{Min/+}$ mouse. Within these regions candidate genes may be selected

2.2 MATERIALS AND METHODS

2.2.1 Genotyping the N2 Apc$^{Min/+}$ population

Microsatellite markers are a subset of simple sequence (di-, tri-, or tetra-nucleotide) repeats found in non-coding regions such as introns and intergenic sequences that tend to exist at high frequency within the genome (Hamada *et al.* 1982). A complete set of microsatellite markers (n=112) (Table 2-1) were selected on the basis of a 12bp predicted allele size difference between the founder strains of mice as indicated in the Whitehead Institute at MIT Center for Genome Research (WICGR), (Genetic and Physical Maps of the Mouse, Database Release 1999); http://www.broad.mit.edu /cgibin/mouse/sts_info). The database was searchable using the genetic distances determined by the MIT F2 intercross by Dietrich (Dietrich *et al.,* 1996) and used with data from the Centre for Inherited Disease Research (CIDR) http://www.cidr.jhmi.edu/mouse and http://www.informatics.jax.org, for the physical (Mbp) and genetic position (cM) of each marker.

Initially 60 markers were selected for genotyping using the Cancer Research UK genotyping facility in Oxford. The genetic data of these markers was analysed using Genescan Software v3.0 (Applied Biosystems) and set up in a relational database using FileMaker Pro version 8.5. A scan was performed The preliminary genome wide scan based on marker regression and IM using Map Manager QTX (Manly *et al* 2001) revealed broad QTLs covering large genomic regions. An additional 52 markers were then selected and genotyped in house and added to the original database of 60 markers to increase marker density between 15-20cM intervals.

Table 2-1 Microsatellite markers used for genotyping N2 *Apc*$^{Min/+}$ BC progeny. Marker positions are based on The Jackson Laboratory's online Mouse Genome Informatics resource (http://www.informatics.jax.org). Location for syntenic markers were taken from CONTIG data* (MGI database, 1999); CONTIG13 (D3Mit51), CONTIG17 (D7Mit362), CONTIG8 (D16Mit5), CONTIG11 (D19Mit6) or published QTL papers¥; D6Mit198 (Griff *et al.*, 1995); D12Mit17 (Pataer *et al.*, 1997); D18Mit208 (Drayton *et al.*, 2006)

Locus	Position (cM)	Locus	Position (cM)	Locus	Position (cM)	Locus	MGI Position (cM)
D1Mit430	4.3	D5Mit168	76.2	D9Mit336	35.4	D14Mit127	18.7
D1Mit169	9.9	D5Mit409	85.2	D9Mit335	32.0	D14Mit174	19.4
D1Mit132	39.5	D6Mit1	6.7	D9Mit347	55.1	D14Mit60	24.6
D1Mit495	55.8	D6Mit83	5.7	D9Mit201	69.4	D14Mit102	34.4
D1Mit17	95.0	D6Mit123	27.8	D9Mit151	72.4	D14Mit107	66.1
D1Mit292	96.3	D6Mit284	41.1	D10Mit189	7.7	D15Mit100	19.3
D2Mit327	40.9	D6Mit287	52.1	D10Mit106	11.7	D15Mit70	38.0
D2Mit395	60.0	D6Mit194	62.9	D10Mit194	24.5	D15Mit159	42.0
D2Mit411	80.0	D6Mit15	77.7	D10Mit31	35.3	D16Mit182	2.6
D2Mit229	89.0	D6Mit198¥	67.0	D10Mit42	39.7	*Prkdc*	10.1
D2Mit148	100.5	D7Mit178	2.0	D10Mit95	47.1	D16Mit4	25.4
D2Mit230	103.0	D7Mit117	17.3	D10Mit233	61.6	D16Mit5*	32.8
D3Mit178	14.3	D7Mit228	25.6	D10Mit103	73.6	D16Mit189	46.8
D3Mit51*	25.5	D7Mit276	34.4	D11Mit216	19.2	D16Mit106	57.7
D3Mit320	66.8	D7Mit323	54.5	D11Mit339	39.5	D17Mit143	4.9
D3Mit352	75.0	D7Mit101	69.0	D11Mit177	40.4	D17Mit51	19.7
D4Mit227	4.4	D7Mit362*	67.8	D11Mit179	54.6	D17Mit180	26.7
D4Mit18	6.0	D8Mit155	2.1	D12Mit182	5.5	D17Mit20	29.7
D4Mit193	14.0	D8Mit124	7.6	D12Mit91	30.1	D17Mit142	49.7
D4Mit348	39.4	D8Mit289	16.5	D12Mit143	36.4	D18Mit222	8.1
D4Mit308	57.7	D8Mit292	21.2	D12Mit17¥	55.0	D18Mit208¥	38.0
D4Mit256	86.2	D8Mit45	42.2	D13Mit115	7.3	D18Mit49	51.3
D5Mit387	14.4	D8Mit211	52.0	D13Mit179	21.7	D18Mit4	57.3
D5Mit352	18.4	D8Mit213	57.5	D13Mit142	32.5	D19Mit28	10.8
D5Mit201	39.6	D8Mit49	72.4	D13Mit107	50.2	D19Mit106	20.2
D5Mit157	48.5	D9Mit250	2.5	D13Mit260	63.7	D19Mit90	36.0
D5Mit188	57.5	D9Mit90	17.8	D13Mit78	67.2	D19Mit103	48.5
D5Mit139	64.2	D9Mit285	21.4	D14Mit126	11.9	D19Mit6*	57.9

2.2.2 DNA preparation

For the additional 52 markers, genomic DNA was extracted from ear snips using a QIAamp® DNA Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer's protocols. Approximately 12mg was cut into small pieces and underwent protein degradation by incubation overnight in Buffer ATL (180 µl) and Proteinase K (20 µl) at 56ºC. This was then spiked with a further 20µl of Proteinase K and incubated overnight at 56ºC.

Cells were lysed in Buffer AL (200 µl) by incubation for 10 minutes at 70ºC. 100% ethanol (200 µl) was then added to each sample and after mixing by pulse vortexing, samples were added to a spin column, with a silica-gel membrane, in a 2 ml micro centrifuge tube. The column, sample and tube were then centrifuged at 8,000 rpm for 1 minute to facilitate the adsorption of DNA onto the membrane. The filtrate was discarded and 500 µl of wash buffer (AW1) was added to the column, prior to centrifugation at 8,000 rpm for 1 minute. Again, the filtrate was discarded, and washing was repeated with 500 µl of wash buffer (AW2) with centrifugation at 13,000 rpm for 3 minutes. The spin column was then placed in a new 1.5 ml eppendorf tubes and 200 µl of elution buffer was added to the column and left to incubate for 10 minutes at room temperature. After centrifugation at 8,000 rpm for 1 minute, the eluate containing purified DNA was collected and stored below -20ºC for future use.

The concentration and purity of the double stranded (ds) DNA samples were performed on a NanoDrop™ ND-1000 spectrophotometer (NanoDrop Technologies,

Wilmington, USA). The optical density (OD) readings at wavelength (λ) 260nm to measure DNA concentration and λ 280nm for protein concentration was used as a means to assess the purity of the sample. A $OD_{\lambda 260}$ / $OD_{\lambda 280}$ ratio of 1.8 to 2.0 was considered adequately pure for subsequent PCR reaction. Qiagen company protocol yielded high quality DNA ranging from 25-200ng/µl.

2.2.3 Oligonucleotides and PCR amplification.

For the additional 52 markers, primers were selected from the database release 1999 (as mentioned above) and purchased from Sigma Genosys UK. The lyophilized primers were diluted to their supplied optical density readings to a stock concentration of 100 µM and stored at 20 ºC. Working dilutions were then made to a concentration of 10 µM. PCR amplification were carried out in a Perkin-Elmer 9600 thermal cycler and was performed with 1 µl genomic DNA in a total volume of 25 µl using the PCR solutions shown in Table 2-2.

Table 2-2 Polymerase chain reaction reagents for preparation of 25µl reaction using 0.1µl Taq DNA Polymerase for genotyping N2 *Apc*$^{Min/+}$ mouse DNA.

Stock Solution	Volume per 25µl reaction (µl)	Final concentration per reaction
10x Buffer Containing [15mM] $MgCl_2$	2.5	1X
Forward Primer [10 µM]	1.25	1 µM
Reverse Primer [10 µM]	1.25	1 µM
dNTP (each dNTP [10mM])	2	200 µM of each dNTP
DNA template (25-200ng/µl)	1	1-20ng
Q solution	5	1X
Sterile PCR H_20	12	-

The melting temperature (T_M) for each primer varies and was calculated at 2ºC below suppliers T_M which considers factors such as total number of nucleotides in the oligo; (GCs and TAs) and the presence of ~50mM monovalent cations (Mg^+) per PCR reaction. These calculations determine the optimal annealing temperature for the PCR procedure and determine the conditions of cycling. We used PCR annealing temperatures in the range of 55-65ºC, but typically the primers annealed at 60 ºC. A PCR program where sample mix is denatured for 15 minutes at 95ºC, followed by 37 cycles (95ºC for 10 seconds, annealing at 55-65ºC for 45 seconds and elongation at 72ºC for 45 seconds) and final elongation at 72ºC for 10 minutes.

2.2.4 Agarose gel electrophoresis

The DNA fragments were resolved by 2% denaturing agarose gel electrophoresis. Agarose (UltraPURE Gibco BRL) was added to 0.5X TBE buffer (44mM Tris, 45mM boric acid, 1mM EDTA) PH8 and melted in a microwave medium power for 16 minutes and ethidium bromide was added to make a final concentration of 0.5µg/ml to allow visualisation under ultraviolet light. DNA molecular weight markers –100 kilobase pair DNA ladder (SigmaAldrich, UK) was loaded in the first row of the gel. DNA samples were loaded with 6X gel loading dye, blue (0.015% bromophenol blue, 11mM EDTA, 2.5% Ficoll 400) obtained from New England BIO labs. Electrophoresis was carried out at 120 volts for 1- 1.5 hours and viewed under UV transilluminator (Bio Rad Gel Doc™ XR+ System) and images were captured using Bio Rad Image Lab™ Software as shown in Figure 2-1.

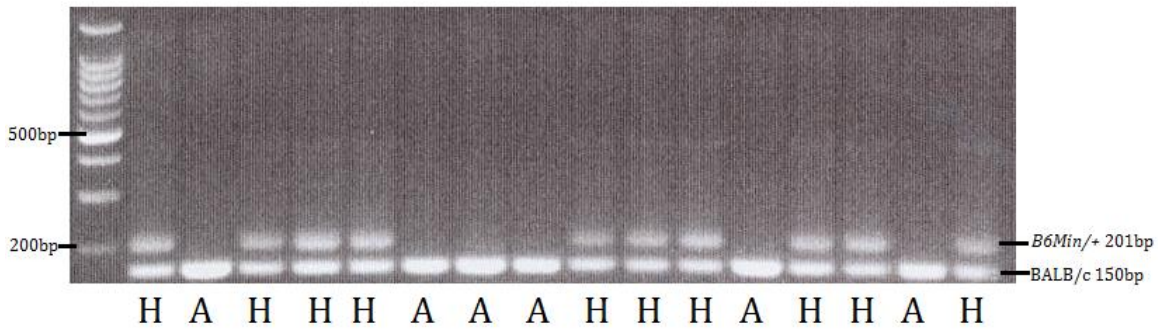

Figure 2-1 Representative gel electrophoresis (2%). image of D7Mit178 microsatellite marker. Lane 1=1-100 bp DNA ladder, Lanes 2-17=N2 (*Apc* $^{Min/+}$) mice genotyping and verification of D7Mit178 polymorphism between B6$^{Min/+}$ (165 bp) and BALB/c (201 bp allele). A=homozygote, H=heterozygote alleles

2.2.5 Linkage analysis

Map Manager QTX analysis (Manly *et al.,* 2001) was used to perform IM. The genotype data were abbreviated to either 'A' or 'H' which references the homozygous and heterozygous genotypes respectively. Missing data was encoded using '0' which Map Manager QTX either excluded from interval mapping and permutation testing unless it can be estimated as an expected genotype effect using the flanking markers. Map Manager QTX places the markers in order, defines the intervals and calculated the segregation pattern of alleles based on the genotype code. Physical and genetic maps were generated and the crossover events were calculated using one of the four conditional probabilities to determine the rf as intended by the Kosambi's mapping function (Kosambi 1944).

Three quantitative traits (adenoma frequency in the USI, LSI and LI) were simulated for each of the 19 autosomes in the mouse genome. The likeliest position of a QTL on a chromosome and the association with a given trait were determined by the

'peak' statistic value described as the likelihood ratio statistic (LRS). LRS is a measure of the ratio between likelihoods under the null and test hypotheses and can be converted to Logarithm of Odds (LOD) by dividing by 4.61 (twice the natural logarithm of 10). The linkage search criterion was set to $p < 0.05$, as a cut off value to assess the initial marker to trait association and a constrained additive-effects model was used to validate the additive effect of alleles at a QTL using one degree of freedom.

For significance the observed LRS was tested against the null hypothesis: that there are no QTLs and the ones observed in this study did not occur by chance in the genome using permutation testing for each interval. Empirical *p*-values for the highest LRS were calculated under the null hypothesis and ranked for significance based on threshold values. Permutation tests were performed to determine three thresholds for QTL detection (Churchill and Doerge, 1994) of 1%, 5%, and 63%. QTLs with LOD scores above the 1% threshold were highly significant, while those above 5% and 63% were significant and suggestive, respectively (Lander and Kruglyak, 1995). We identified QTLs at a 5% genome wide error rate for each individual trait (USI, LSI, and LI). IM analysis was also performed using R-qtl Version 1.25-15 (Broman *et al.*, 2003) for corroboration and further analysis of results.

Linkage maps were generated using Map Manager QTX. Failure of markers during regression can occur if the population shows extreme segregation distortion or non-recombinant markers which are redundant and not separated by recombination events and the genotype data is the same for the marker pair and therefore

uninformative. Redundant markers were observed on chromosome (Chr) 3 between D3Mit320 and D3Mit352 and were hidden or removed from IM analysis and did not appear to have any significant impact on QTL mapping with Map Manager.

The MQM approach was explored initially to investigate linked QTLs and was conducted using R-qtl. This model used the QTL with large effects as either an additive or interactive covariate to enhance the identification of multiple QTLs, to increase precision of QTL positions and their effects, and to identify any interaction patterns between linked QTLs (epistasis). A more in-depth analysis of MQM to construct multiple QTLs and epistasis will be discussed in the next chapter.

2.3 RESULTS

2.3.1 Data visualisation

Positioning of the 112 genome-wide microsatellite markers constructed a genetic map using Map Manager QTX and confirmed using R-qtl. The average interval space between the markers was calculated at 17 cM based on Kosambi's distances for linkage analysis (Figure 2-2).

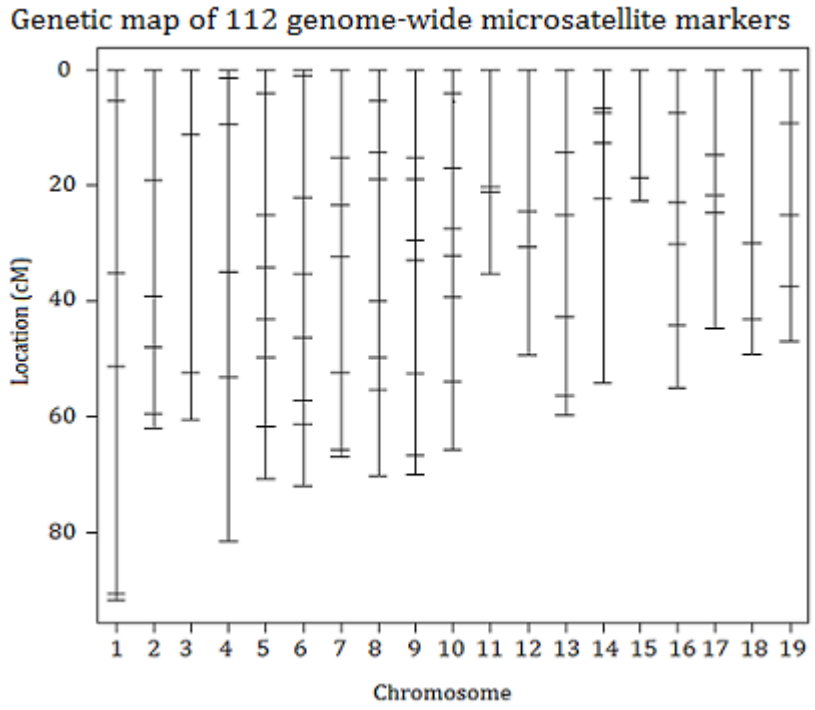

Figure 2-2 Genetic map, constructed using Rqtl makers. This involved the following markers in genome wide mapping; chromosome (Chr) 1, D1Mit430, 169, 132, 495, 17 and 292; Chr2, D2Mit327, 395, 411, 229, 148, and 230; Chr3, D3Mit178, 51, 320 and 352; Chr 4, D4Mit227, 18, 193, 348, 308 and 256; Chr5, D5Mit387, 352, 201, 157, 188, 139, 168 and 409; Chr6, D6Mit1, 83, 123, 284, 287, 194, 15 and 198; Chr7, D7Mit178, 117, 228, 276, 323, 101 and 362; Chr8, D8Mit155, 124, 289, 292, 45, 211, 213 and 49; Chr9, D9Mit250, 90, 285, 336, 355, 347, 201 and 151; Chr10, D10Mit189, 106, 194, 31, 42, 95, 233 and 103; Chr11, D11Mit216, 339, 177 and 179;

Chr 12, D12Mit182, 91, 143 and 17; Chr13, D13Mit115, 179, 142, 107, 260 and 78; Chr13, D13Mit; Chr14, D14Mit126, 127, 174, 60, 102 and 107; Chr15, D15Mit100, 70 and 159; Chr16, D16Mit182, *Prkdc*, 4, 5, 189 and 106; Chr17, D17Mit143, 51, 180, 20 and 142; Chr18, D18Mit222, 208, 49 and 4; Chr19, D19Mit28, 106, 90, 103 and 6.

The percentage of individuals genotyped were 89.1% and 88.1% for 0Gy and 2Gy, respectively and the composition was 50.1% (A) and 49.9% (H) for 0Gy, 49.3% (A) and 50.9% (H) for the 2Gy indicating an overall good coverage of the observed marker data across the N2 $Apc^{Min/+}$ panel (Figure 2-3). The proportion of missing values appeared to be low and was randomly distributed in the population except in the cases of Chr 3, 7, 9, 15 and 18. We inspected the areas of missing genotypes for Chr 3 (10.5%), 7 (21.6%), 9 (14.7%), 15 (10.4%) and 18 (34.8%) of the 2Gy dataset using plot.geno function of R-qtl to indicate any likely errors (data not shown). No apparent genotyping errors were noted from the observed data. It was therefore postulated that the effects of this missing information will not affect QTL analysis. Overall investigations of 0Gy and 2Gy genotyping errors were conducted using the 'errorlod' function of R-qtl as described by Lincoln and Lander (1992).

A LOD score was calculated for each individual mouse at each marker; large error LOD scores indicated higher possibility of genotyping errors with error LOD score < 4, which can be ignored. The cutoff was set to the recommended p= 0.01 and no errors (LOD score set >4) were observed in both datasets. The *Prkdc* status for each N2 $Apc^{Min/+}$ was noted as homozygote (B/B) ($Prkdc^{BALB/c/BALB/c}$) or heterozygote ($B6^{Min/+}$ /B) ($Prkdc^{B6Min/+/BALB/c}$) and the expected 50:50 ratio was observed in both the 0 and 2Gy groups (Degg *et al*., 2003).

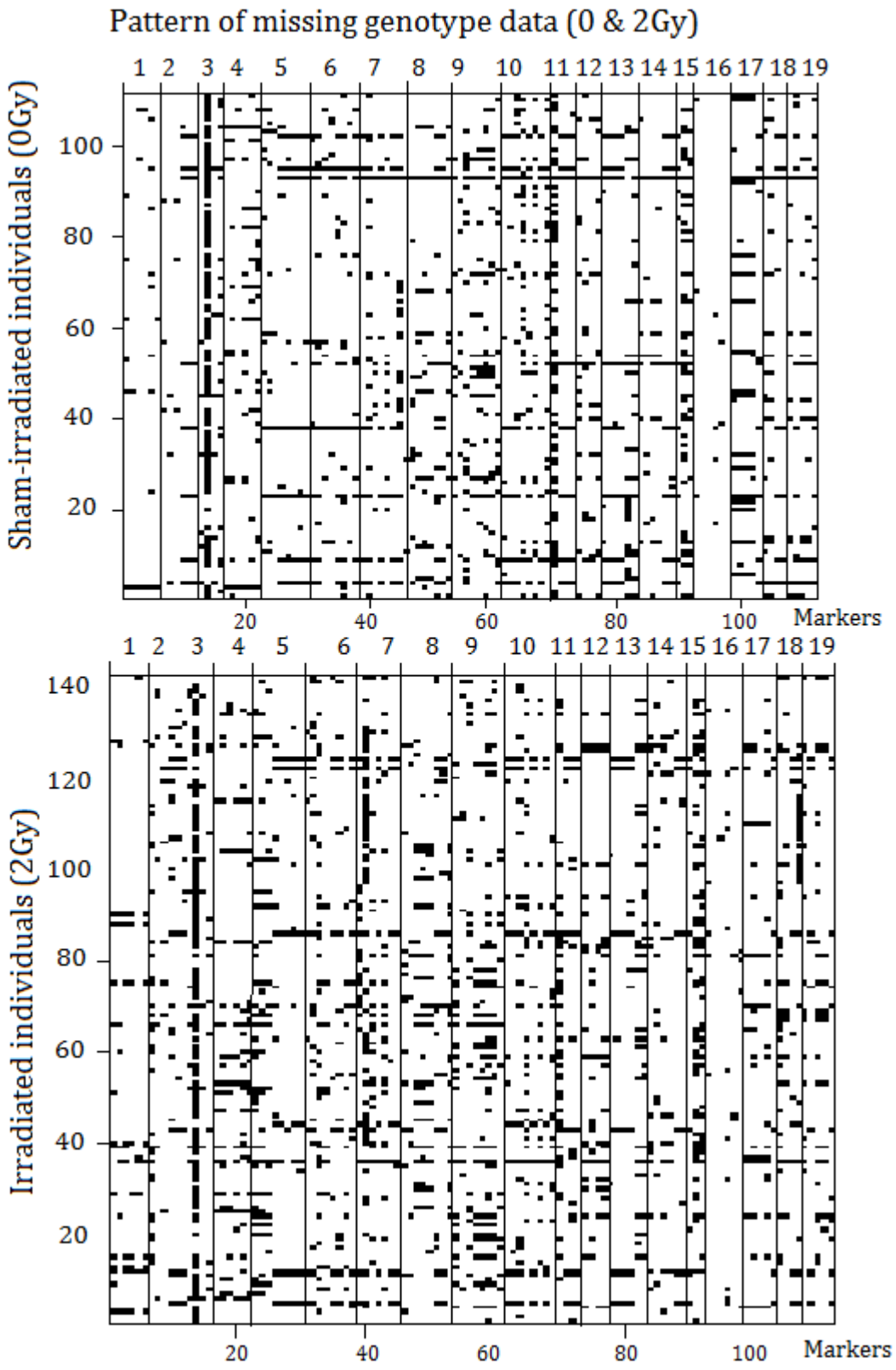

Figure 2-3 Missing genotype for irradiated (0Gy) and sham-irradiated (2Gy) data. Missing values are shown in black

2.3.2 Irradiated N2 Apc$^{Min/+}$ mice have increased intestinal polyp multiplicity.

The sham irradiated control group (0Gy) were used to identify regions linked to spontaneous modifiers and to compare these regions and polyp multiplicity with the radiation (2Gy) group. Animals were assessed for adenoma multiplicity in USI, LSI and LI at time of sacrifice (all mice >80 days; mean: 0Gy, 241 days; 2Gy, 180 days). The 2Gy-irradiated animals had a 1.8 fold higher polyp incidence in the USI compared with the unirradiated mice 0Gy thus illustrating that radiation causes an increase in polyp burden (Degg *et al.,* 2003). Furthermore, the distribution pattern of our phenotype data showed properties that adenoma multiplicity in our dataset is a quantitatively inherited trait and controlled by QTLs (Figure 2-4). In N2 *Apc*$^{Min/+}$ mice, the (USI) (duodenum and jejunum) was more sensitive than the LSI (ileum) to radiation-induced adenoma formation indicating that there may be modifiers that influence tumour multiplicity.

2.3.3 Linkage analysis identified quantitative trait loci

Linkage analysis was conducted by Map Manager QTX to calculate empirical p-values for the highest 'peak' LRS value using permutation scans. The individual trait (USI, LSI, and LI) threshold values were defined using Kosambi's mapping function using informative progeny. Individual permutation tests were performed for each chromosome showing evidence of linkage from marker regression; a 'suggestive' threshold value was defined if the observed LRS value appears ≥20, if the value appears ≥200 then this was considered 'significant and 'highly significant if ≥500.

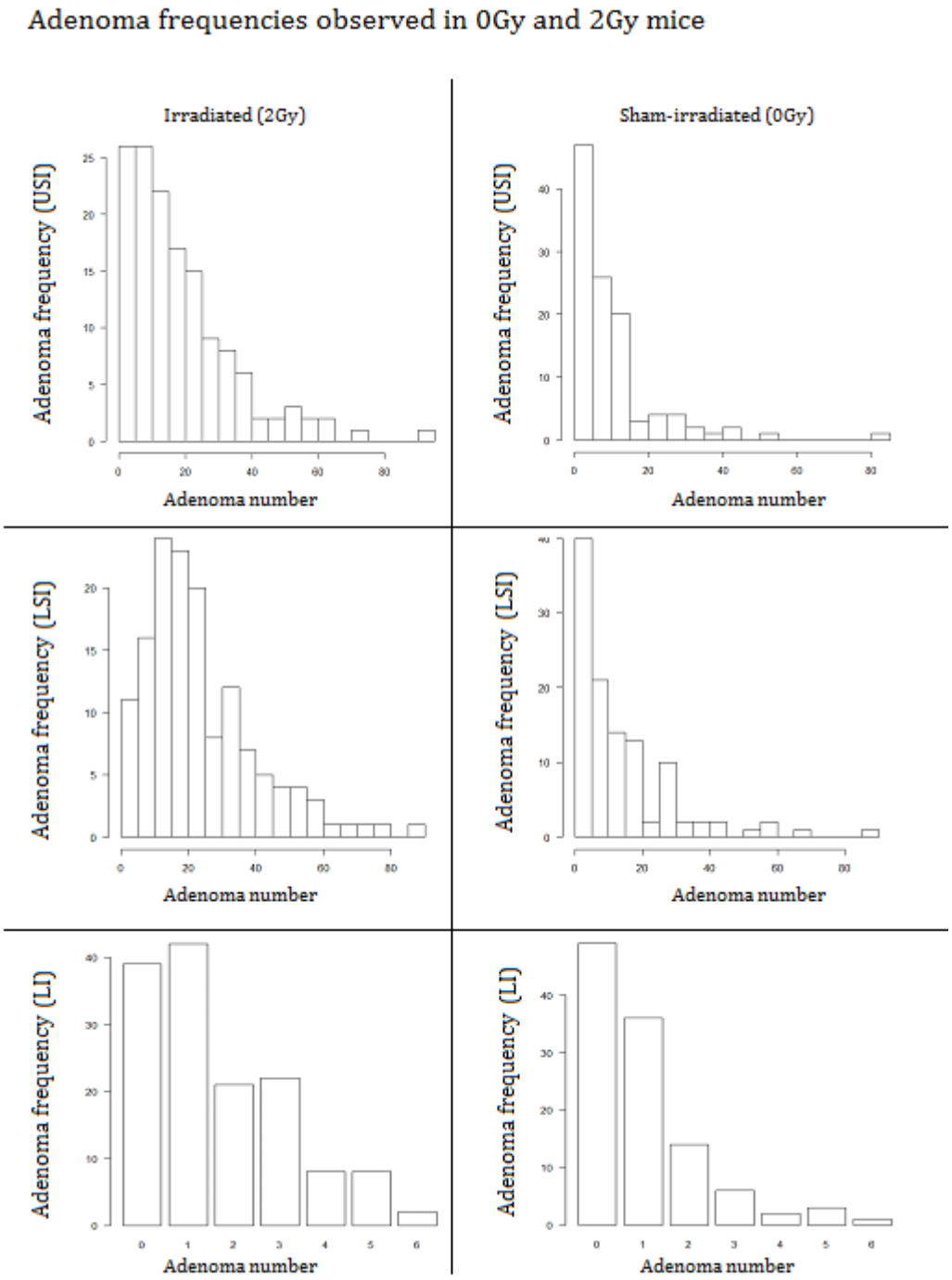

Figure 2-4 Phenotypic frequency of adenomas to demonstrate irradiation influence on polyp multiplicity**.** Adenoma number shown for the duodenum and jejunum; USI=upper small intestine, proximal and distal ileum; LSI=lower small intestine and LI=large intestine for irradiated (2Gy) and sham-irradiated (0Gy).

The overall genome wide threshold value of 2.5 LOD was calculated from over 18 separate permutation tests ($p < 0.05$) and was set to attain significant association of QTLs (shown in Figure 2-5).

We identified several main-effect QTLs, which contributed to polyp multiplicity in the irradiated sample group and found only suggestive QTLs in the sham irradiated group. No apparent over lapping genomic positions were detected between the 0Gy and 2Gy thus, implying at this stage of our analysis that radiation-induced loci are important and specific to the effects of irradiation using this model.

2.3.4 Modifier loci map to chromosome 2, 5 and 16

In the irradiated cohort of mice, markers on chromosome 2, 5 and 16 segregated with high polyp multiplicity and a total of 5 QTLs; one significant and four highly significant QTLs were detected on these chromosomes. These loci were designated *Mrip* (*Mom* radiation-induced susceptibility); *Mrip*1-5 and were distinctly associated with the USI (Table 2-3). Susceptibility loci on chromosome 16, designated *Mrip* 4 and 5 were identified previously and reported by Degg *et al* (Degg *et al.,* 2003), and re-analysed for the completeness of the genome scan undertaken here. Two highly significant *Mrip* loci (*Mrip2, 3*) were detected on chromosome 5, in the intervals between D5Mit201-D5Mit157 at position 22cM of the genetic map constructed by QTX, ($p < 0.00001$, 95% confidence interval (CI) of 13.5cM) and D5Mit188-D5Mit139, ($p < 0.00001$, 95% CI of 10cM) at position 49cM of the genetic map. A significant *Mrip* locus was located on chromosome 2 and designated *Mrip1*, at marker D2mit395 QTX derived genetic map position of 48cM, ($p < 0.0003$, CI

21.5cM) and one suggestive locus on chromosome 17 (p<0.008, CI 39M). Suggestive QTLs were not assigned as *Mrip*, however these were included in further analysis. These QTLS contributed an overall 68% phenotypic variation observed in polyp multiplicity in the 2Gy cohort and confirm the trend towards worsened polyposis in the duodenum probably involving differential expression of genes. Results are summarised in Table 2-3 and Figure 2-5.

Table 2-3 Quantitative trait loci detected by interval mapping for 2Gy irradiated mice. LRS=likelihood ratio statistic, empirical p values calculated for each trait-genotype data, %=percentage variance, CI=confidence intervals (95%) and the association level attained from interval mapping. USI=upper small intestine, LSI=lower small intestine, LI=large intestine, HiSign=highly significant, Sign=significant, Sugg=suggestive based on LRS thresholds assigned using separate permutation tests. cM calculated by QTX genetic map

Trait	Chr	Locus or Interval (cM)	p-value (LRS)	%	95% CI	Sign (threshold)
USI	Chr2 *Mrip1*	D2Mit395 (48)	$3.2x10^{-4}$ (12.9)	9	±21.5	Sign (≥ 12)
	Chr5 *Mrip2*	D5Mit201-157 (22)	<0.00001 (24)	15	±13.5	HiSign (≥ 20.2)
	Chr5 *Mrip3*	D5Mit188-139 (49.0)	<0.00001 (28.6)	19	±10	HiSign (≥ 20.2)
	Chr16 *Mrip4*	*Prkdc* (10)	$4x10^{-5}$ (19)	11	±16.5	Sign (≥ 12.6)
	Chr16 *Mrip5*	D16Mit189 (29)	<0.00001 (22.2)	14	±13.5	Sign (≥ 12.6)
	Chr17	D17Mit180 (18)	$8.2x10^{-3}$ (7)	5	±39	Sug (≥ 6.76)
LSI	Chr1	D1Mit495-17 (85)	$4.4x10^{-3}$ (8.6)	6	±33.5	Sug (≤ 12)
	Chr6	D6Mit194 (56)	$2.6x10^{-3}$ (8.8)	6	±30	Sug (≤ 12.8)
	Chr7	D7Mit178-117 (8)	$7x10^{-3}$ (12.2)	6	±38	Sug (≤ 12.5)

	Chr8	D8Mit211-213 (62)	0.007 (9.5)	7	±38	Sug (≤ 12.7)
LI	Chr1	D1Mit292 (95)	$6x10^{-3}$ (8.4)	6	±32.5	Sug (≤ 13.1)
	Chr2 *Mrip1*	D2Mit395 (48)	$1.2x10^{-3}$ (12.7)	8	±24	Sign (≥ 11.6)
	Chr7	D7Mit228-276 (29)	$1.2 x10^{-3}$ (9.9)	7	±35	Sug (≤ 13.1)
	Chr8	D8Mit124-289 (12)	$2.7x10^{-3}$ (13.6)	6	±30.5	Sign (≥13.2)
	Chr14	D14Mit126 (0)	$3.8x10^{-3}$ (8.4)	6	±33	Sug (≤ 12.6)

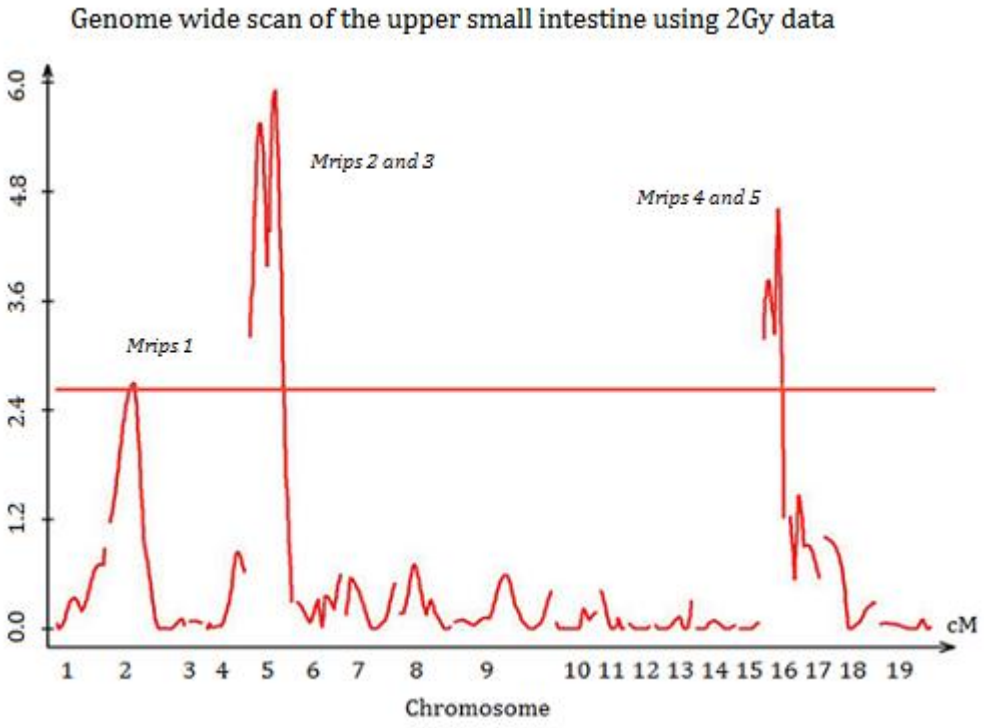

Figure 2-5 Genome scan shows significant quantitative trait loci for adenoma multiplicity identified in the upper small intestine of 2Gy irradiated mice. Strong effect QTLs were found on chromosomes 2, 5 and 16 and determined by MapManager QTX. The horizontal line above the *y* axis represents a 2.5 LOD significant threshold score (p=0.05).

The USI results were investigated for corroboration with R-qtl for chromosomes 1, 2, 3, 5, 16, 17, and 18 (Figure 2-6). No major differences were observed between the two software packages except in chromosome 3: as already mentioned in the case of chromosome 3 (D3Mit320 and D3Mit352) markers failed regression and where removed from analysis. Single marker regression was possible with Rqtl.

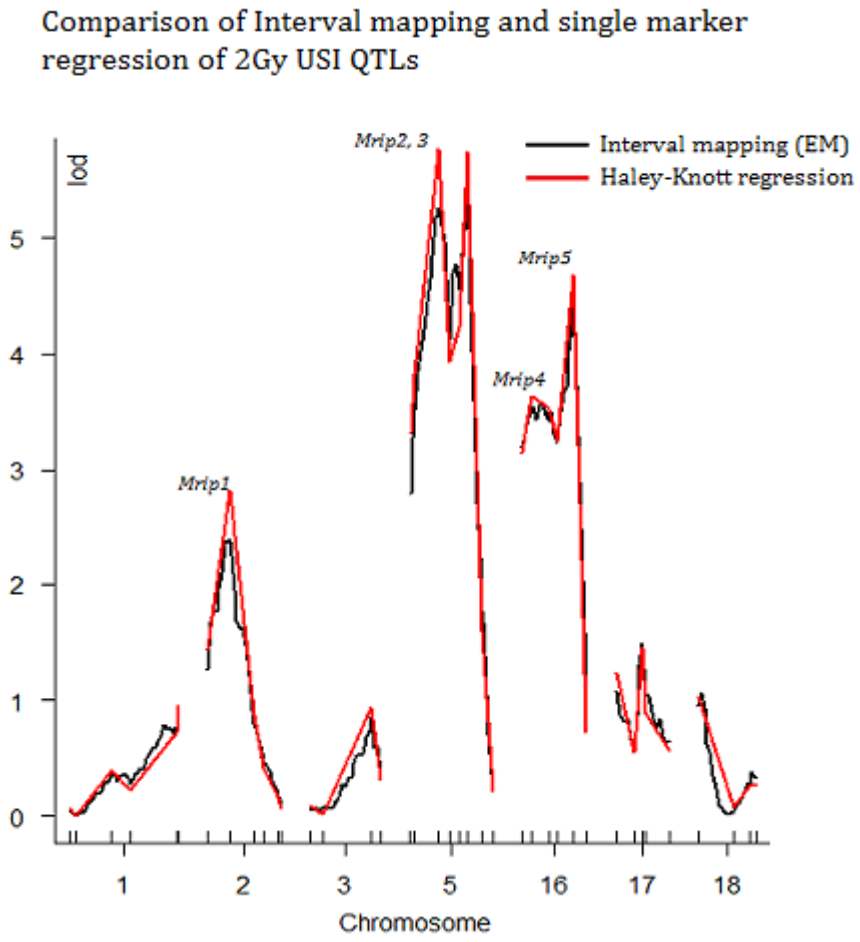

Figure 2-6 Comparison of Interval mapping single marker regression for the QTLs detected for adenoma multiplicity in the upper small intestine of 2Gy irradiated mice. Interval mapping was performed using Expectation-Maximisation (EM) (black) algorithm and single QTL analysis using Haley-Knott (red). Minor differences were noted concerning *Mrip1* and *3* with slightly higher LOD scores with single QTL analysis.

Four suggestive QTLs were detected in the LSI on Chr: 1 at an interval marked by D1Mit495-D1Mit17 (85cM) with peak LRS of 8.6 (p=0.0044; CI, 33.5); Chr 6, at 56cM at D6Mit194 locus with peak LRS of 8.8 (p=0.0026; CI, 30cM); Chr 7 at 8cM within D7Mit178 – D7Mit117 interval with peak LRS 12.2, (p=0.007, CI, 38cM); and Chr 8, D8Mit211-D8Mit213, peak LRS of 9.5 (p=0.007; CI, 38cM) (Table 2-3).

2.3.5 Common modifier loci mapped to chromosome 2 in USI and LI

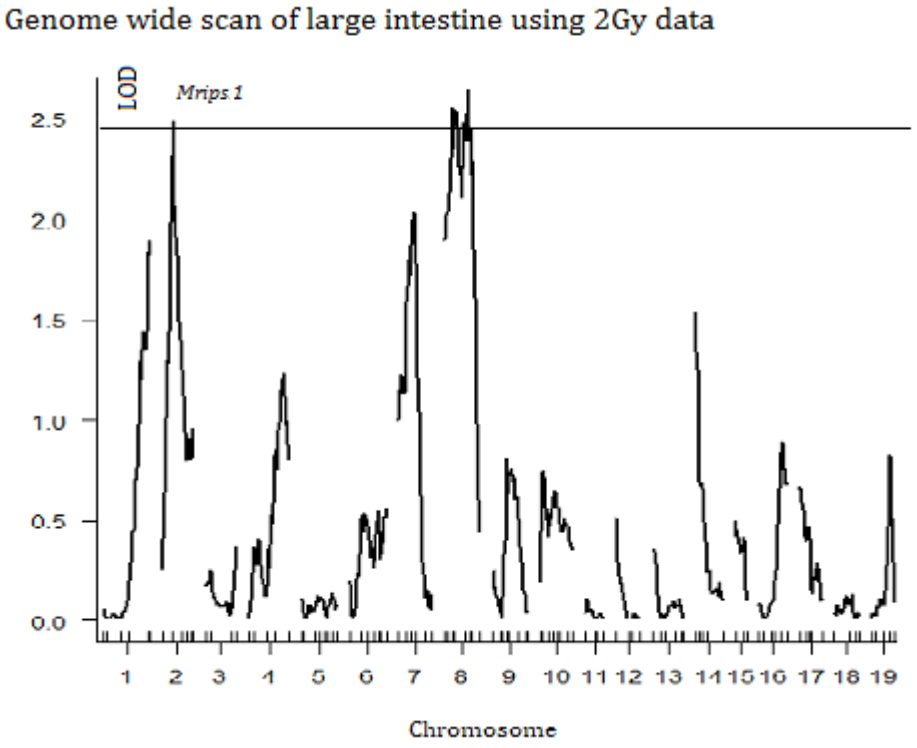

Figure 2-7 Genome scan shows significant quantitative trait loci for adenoma multiplicity identified in the large intestine of 2Gy irradiated mice. Main-effect QTLs detected on chromosome 2 and 8, suggestive QTLs on Chr 1, 7 and 14. The horizontal line above the *y* axis represents a 2.45 LOD significant threshold score (p=0.05).

For the LI trait in this cohort, significant QTLs associated with the LI were detected on chromosome: 2, D2Mit395 (p=0.0012; CI, 24 cM) and 8, D8Mit124 (p=0.0043, CI,

30cM) (Table 2-3 and Figure 2-7). Suggestive QTLs emerged on Chr: 1, D1Mit292 (p=0.006; CI, 32.5cM); 7, D7Mit276 (p=0.0012; CI, 35); and 14, D14Mit126 (p=0.0067; CI, 33cM). *Mrip1* appears to be a common loci involved in the USI and LI as shown in Figure 2-8

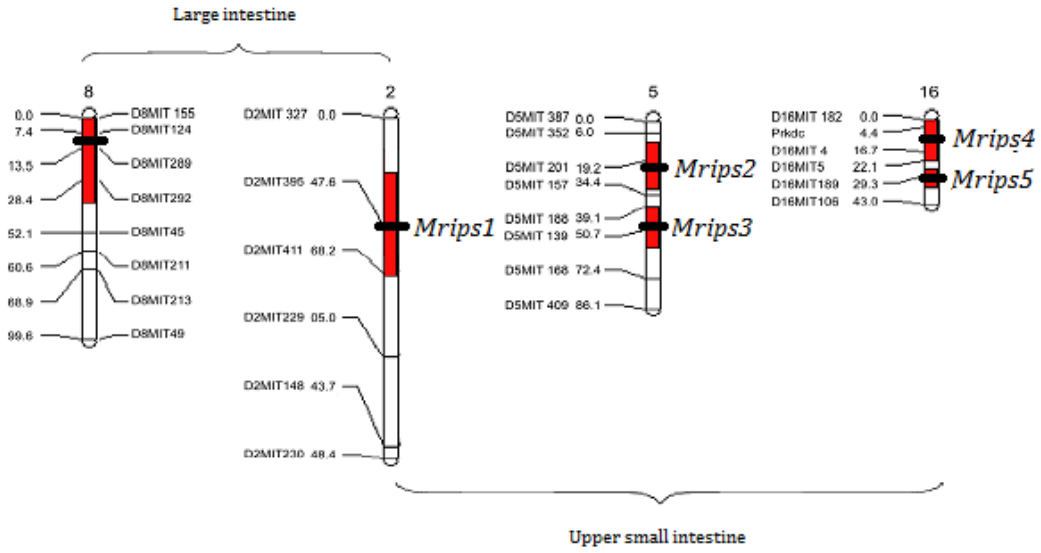

Figure 2-8 Six significant quantitative trait loci detected in upper small intestine and large intestine of 2Gy irradiated mice**.** Strong effect QTLs associated with upper small intestine adenoma multiplicity on chromosomes (Chr) 2, 5 and 16; with Chr 2 common to USI and LI.

IM analysis was repeated for the irradiated cohort of mice using Cartographer QTL (Basten *et al.* 2004; Basten *et al*., 1994) to confirm QTLs from Map Manager QTX. Permutation tests (1,000 tests at 2cM window) with all markers for each of the four traits for 2Gy dataset provided similar LRS threshold values as QTX. The only exception was D2Mit395, which showed borderline significance when analysed using Map Manager, but with Cartographer the marker fell just below the significance threshold (LOD of 2.7; significance ≥ a LOD score of 2.8). The

computations for 2Gy were also analysed using R-qtl to compare Haley–Knott (HK) regression (1992), expectation–maximization (EM) algorithm (Lander and Botstein 1989) and multiple imputation (Sen and Churchill 2001). Similar LOD profiles were obtained as Map Manager QTX, however the large effect QTL*s* *Mrip*2 and *Mrip3* (Chr5) appeared to be sharper peaks with overall higher LOD scores (data not shown). The borderline suggestive QTL detected by QTX on chromosome 7 of the LSI between markers D7Mit178–D7Mit117 was defined as significant (p=0.03) at D7Mit117 (17.3cM) by R-qtl (LOD score of 2.84 and p=0.05 threshold 2.63).

2.3.6 Suggestive loci mapped in the 0Gy mice data

The 0Gy dataset yielded only suggestive QTLs on Chr 3 and 13 for the USI; on chromosome 6 for the LSI; and chromosome 1 for the LI as shown in Table 2-4 and Figure 2-9).

Table 2-4 Quantitative trait loci detected by interval mapping for 0Gy, sham-irradiated mice Data presented shows the trait and loci marker position. LRS=likelihood ratio statistic, empirical p values were calculated for each trait-genotype data, %=percentage variance, CI=confidence intervals (95%) and the association level attained from interval mapping.

Trait	Chr	Locus(cM)	P value (LRS)	%	95% CI	Sign (threshold)
USI	Chr 3	D3Mit352 (100)	$3.5x10^{-3}$ (8)	8	±32.5	Sug (12.2)
	Chr 13	D13Mit179-142 (25)	$3.7x10^{-3}$ (9.2)	8	±33	Sug (≤12.2)
LSI	Chr 6	D6Mit83 (6)	$7.8x10^{-3}$ (7.1)	7	±39	Sug (≤12.3)
LI	Chr 1	D1Mit495-117 (72)	0.04 (7.5)	4	±56.5	Sug (12.1)

For the USI, no definite overlap of main-effect QTLs were detected (Figure 2-9). Chromosome one revealed a suggestive QTL at 72cM associated with LI of the 0Gy. Similarly, 2Gy mice also showed suggestive association at 95cM (D1Mit292) for LI and at 85cM between D1Mit395-D1Mit17 interval for LSI and could suggest possible overlapping loci are involved.

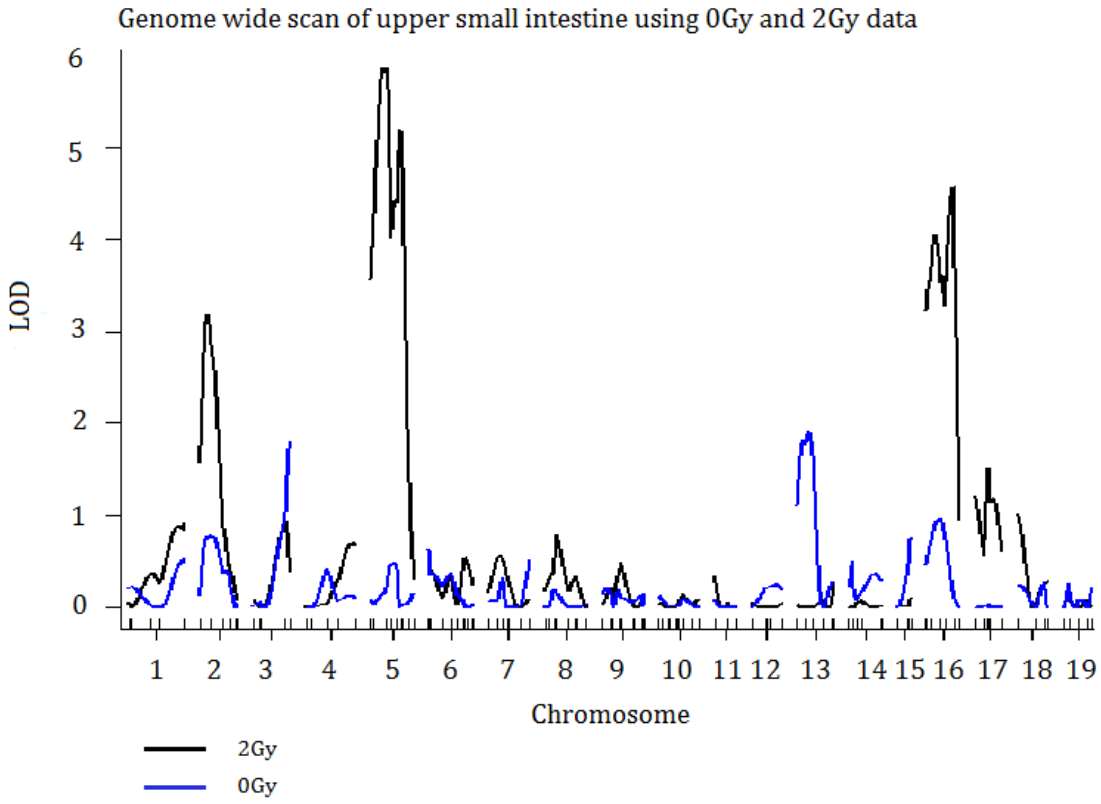

Figure 2-9 Genome wide scan comparison of 0Gy and 2Gy for the upper small intestine trait. Suggestive novel QTLs were detected on chromosomes 3 and 13 in the 0Gy mice (shown in blue) with no overlapping QTLs in the 2Gy mice data.

2.3.7 Genome scan while controlling for the effects of Mrip3

The presence of a large-effect QTL (*Mrip3*) was seen on chr5 (D5Mit139) of the USI, the genome scan was repeated while controlling for this locus to see if loci with low-moderate effects can be made more apparent. The results are plotted together in

Figure 2-10 against the original genome scan. By controlling for *Mrip3* a decrease in the LOD values was detected for Chr4, *Mrip2* and *3* (Chr5), *Mrip4* and *5* (Chr16) and increase was observed in chromosome 1 and 7 which were observed in the LI and substantive suggesting additional common loci between USI and LI.

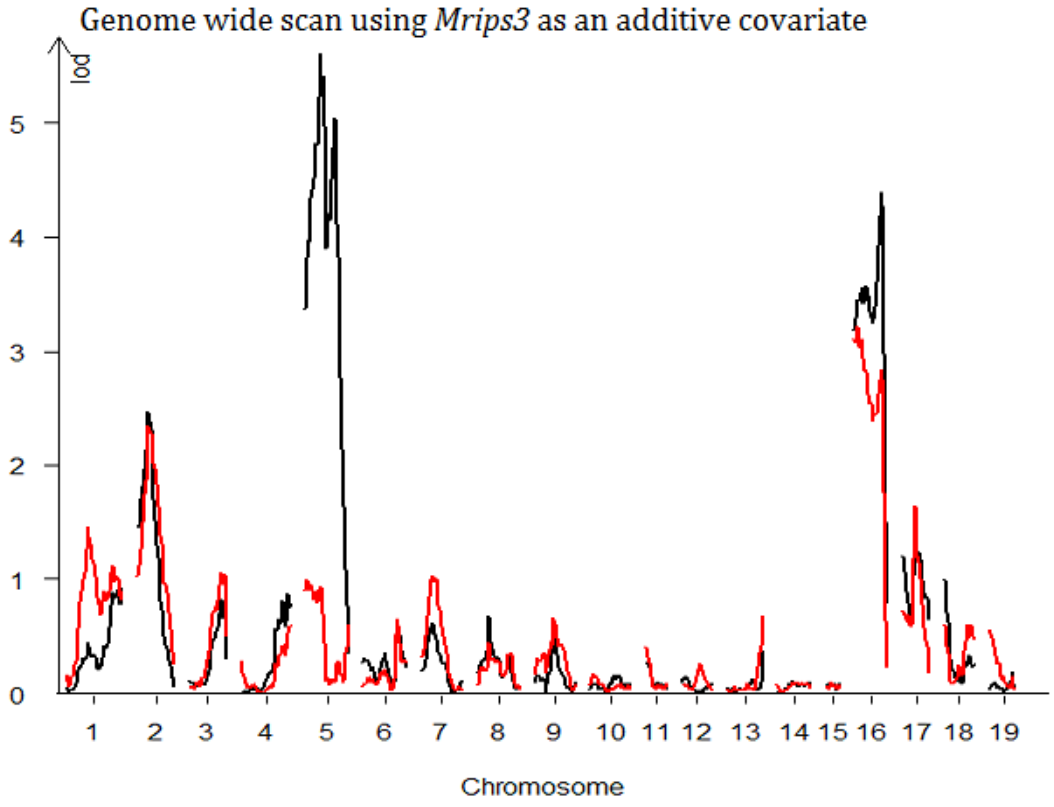

Figure 2-10 LOD curve changes when *Mrip3* was used as an additive covariate (red) from the standard genome scan based on multiple imputation method (black). The majority of the chromosomes and *Mrip1* were unchanged

2.3.8 Calculation of physical and genetic maps

Physical and genetic maps are determined as the number of base pairs (bp) which separate two adjacent loci and the amount of recombination occurring between them (cM), respectively. The reliability of genetic maps depends on the number of

crossovers observed in meiosis, transmission of alleles, the physical distance between loci and the statistical methodology applied. The use of genetic maps has increased advances in the development of identifying susceptibility genes which influence a trait. It is therefore important that the differences between the physical and genetic maps are reconciled which can otherwise present challenges in accuracy and relating the physical and genetic distances when searching for underlying genes. As the exchange of genetic information during meiosis to form recombinant chromosomes is the principle component, therefore the rf is then in part a function of the physical distance between two loci. We calculated the recombination ratio (Mb/cM) of the physical and genetic distances for each *Mrip1-5* interval to compare with the average recombination rate for each chromosome.

Mrip1 candidate interval flanked by D2Mit327 – D2Mit395 with 50Mbp interval and 47.6cM (D3Mit327 starts at 0cM) from our genetic mapping and N2 $Apc^{Min/+}$ cross data has an average rate of 1.05Mb. Compared to the rate of chromosome 2, which at 118Mb and 103.9cM (Mouse Genome Database) has an average of 1.14Mb/cM, with similar averages it is likely that this interval represents a candidate region representative of chromosome 2.

Using the same principle, the average chromosome 5 rate is calculated at 1.6Mbp/cM. The recombination rate of *Mrip2* and 3 intervals are calculated as 3.0Mpb/cM and 0.65Mbp/cM, respectively and their average is 1.85Mbp/cM.

The average chromosome rate is calculated at 1.69Mpb/cM for chromosome 16 and the recombination rates for *Mrip4* compares at 1.67Mbp/cM and *Mrip5* at 2.19Mbp/cM, like *Mrip1* these are likely to represent candidate regions.

2.3.9 Haplotype analysis at Mrip2 and Mrip3

Haplotype analysis at *Mrip2* and *Mrip3* inherited in the N2 $Apc^{Min/+}$ irradiated mice was investigated and shown by the haplotype diagram in Figure 2-11. There remain 17 and 14 BC mice exhibiting recombination between *Mrip2* and *Mrip3*, respectively. This indicates that the BC still holds the potential to narrow these candidate regions further by mapping more polymorphic markers located within the flanking markers.

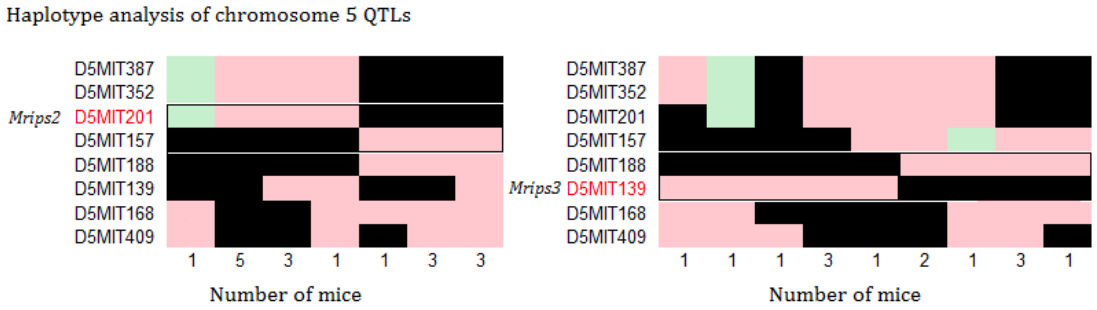

Figure 2-11 Haplotype analysis of recombinant mice at *Mrip2* and *Mrip3* intervals of chromosome 5 using the 8 markers listed. Numbers indicate the number of individuals with that particular genotype (heterozygote in black, homozygote in pink and missing data in green).

At the time we screened potential markers using WICGR database which would lie between marker intervals for *Mrip2* and *Mrip3*. Three out of four markers located within the critical regions were not selected because the polymorphism between B6 and BALB/c alleles did not meet our criteria of exceeding 12bp allelic difference between our strains. D5Mit139 was selected and genotyped in 242 mice, interval

mapping of the initial candidate region was reduced from 33.3cM to 21.6cM, the flanking markers positioned according to MGI at 118.3Mbp and 137.5Mbp saw a reduction in the interval from 19.2Mbp to 7.6Mbp. The paucity of markers in this region and the potential remaining in the BC led this study to explore bioinformatics tools and analysis on existing available data to serve as alternative methods for candidate gene identification.

2.4 DISCUSSION

We were able to identify five independent *Mrip* loci ($p < 0.05$) (*Mrip*1-5) that influenced adenoma multiplicity in the USI irradiated $Apc^{Min/+}$ recombinant mice (Table 2-3). *Mrip1* revealed a marginally significant QTL with co-dominant allele effect. $Apc^{Min/+}$ mice, which are both homozygote and heterozygote for the *Mrip1* locus are involved in polyp multiplicity. *Mrip* 2-5 were detected as highly significant QTLs with recessive allele effects, where the homozygote $Apc^{Min/+}$ mice showed a stronger influence on polyp numbers. These loci account for approximately 68% of the observed phenotypic variance indicating that they play a major role in irradiation induced susceptibility to intestinal adenoma development in $Apc^{Min/+}$ mice. A further two significant QTLs were detected in LI trait of the 2Gy which are considered to be of importance as FAP is characterised by polyps in the large intestine.

Interestingly, no QTLs in the 0Gy sham irradiated mice were detected which overlapped with any of the *Mrip* regions from the 2Gy following a QTL scan, presumably due to the contribution of irradiation induced affects on $Mom1^{R}$ in $B6^{Min/+}$ and BALB/c.

Additionally, it is unclear why the interval recombination rate average is so varied for *Mrip2* and *Mrip3* compared to the average recombination rate for chromosome 5. One possible explanation is that this could be due to segregation distortion (SD) of certain alleles. It appears that the chromosomes bearing the underlying genes of *Mrip2* and *Mrip3* could be over-represented in offspring or involved in meiotic drive,

in favour of one parental allele over the other, which is supported by the fact that *Mrip*2 and 3 favour the homozygote ApcMin/+ mice allele (BALB/c). These markers were initially tested as segregation distortion loci (SDL) by removing them from QTL scan and assessing the recombination fractions and observing the QTL detection. No major differences were noted other than a reduction in the power to detect *Mrip*2 and *Mrip*3 which rules these markers out as SDL. It should be borne in mind that the theory of QTL mapping with SD and how these affect QTL detection especially where linked loci are involved is unclear with a statistical model yet to be designed. In summary, these results indicate that there are strong genetic component of *Mrip*2 which are inherited by the N2 $Apc^{Min/+}$ offspring from the BALB/c parent.

2.4.1 Previous Moms

Some of the previously reported *Mom* loci (*Mom*1, 2, 3 and 7) on Chr 18 and 4 were detected and confirmed in our study in both the 0Gy and 2Gy samples. However these were detected with very weak effects; low LOD scores ranging from 1.14, (p=0.165) to 0.33 (p=0.83) and therefore not attaining significance levels. Our data would suggest that these loci have no, or a very weak influence, on adenoma multiplicity that would require substantially more mice to investigate. *Mom6* was the exception and not detected in either dataset. The genetic effects are unclear and this finding may not be surprising given the genetic makeup of our BC: N2 $Apc^{Min/+}$ offspring carry at least one copy of the BALB/c-derived modifier of $Apc^{min/+}$ (*Mom1*) dominant resistant allele. *Mom2* on chromosome 18 is a spontaneous mutation that is not carried by the mice used here. *Mom3*, the recessive allele conferring susceptibility to severe spontaneous disease is not carried by the $Apc^{Min/+}$ mice used

here. *Mom6* appears to be carried by AKR and is difficult to distinguish from *Mom1*; there are a number of alleles of *Mom7* segregating among mouse strains and it is unclear how the BALB/c allele would influence adenoma number in our crosses.

The ability to detect weak QTLs is a function of sample size and the detection of very weak influences would require substantially more mice to investigate, the fact that other than the three suggestive spontaneous QTLs none map to chromosome 18 suggest that *Mom*2, 3 and 7 are not involved in determining adenoma multiplicity in this genetic background. This highlights the fact that substantially larger numbers of mice would be needed to distinguish modifiers with very small effects in addition to those we have identified. The number of animals required precludes us undertaking such an investigation and increasing markers would not, in the context of our experiment, increase the power of detecting spontaneous QTLs. It however, remains to be seen whether the effects of the Mom2, 3 and 7 on chromosome 18 are made more distinguished or visible using interactive and additive covariates with MQM mapping using R-qtl.

2.4.2 Multiple QTLs

Several studies have reported two or more QTLs on the same chromosome (Spelman and Bovenhuis, 1998; Heyen *et al.,* 1999, Velmala *et al.* 1999) using a two QTL model designed as an extension of marker regression analysis. The selection of appropriate markers is critical for MQM models in order to analyse two or more QTLs on the same chromosome. In addition the use of genetic markers near a large effect QTL as a covariate helps to reduce residual variance and clarify QTLs.

Establishing further evidence for linked QTLs or interacting QTLs will be investigated in the next chapter within the context of multiple QTL models where we plan to use an automated scan procedure to control for large effects QTLs and select suitable markers as covariates.

The present study has investigated susceptibility of BALB/c mice carrying the *Apc*$^{min/+}$ (*Mom1*) dominant resistant allele to irradiation induced adenoma multiplicity by examining the incidence of polyps in the progeny of mice. We have identified numerous QTLs with differing degrees of QTL effects. These QTLs begin to provide insight in whether there is a link in reduced *Prkdc* activity, polyp multiplicity and potentially CRC susceptibility. Thus overall, it appears from our data that certain loci are important and associated with radiation induced tumourigenesis pathway(s) which may be suggestive of radiosensitive genes involved in secondary CRCs caused by radiotherapy.

As complex traits are under the effect of multiple genes with weak individual effects and possibly strong epistatic interactions, we opted out of fine mapping and and extensive breeding programs to produce congenic mice which is no longer necessary due to the information content available in databases. Instead we decided to improve the definition of our QTLs by modelling statistic epistasis to pave the inference of genetic pathways and identify causal variants within our regions. Therefore, my next chapter investigates the use of advanced mapping techniques to refine QTL regions, search for interactions and begin the identification of genes involved in polyposis.

CHAPTER 3 THE GENETIC EPISTATIC BASIS FOR IRRADIATION INDUCED POLYPS IN *APC*$^{MIN/+}$ MICE: INTERACTION AND MULTIPLE QTL MODELLING OF THE MAIN EFFECT-QTLS.

3.1 INTRODUCTION

Genetic dissection of complex traits attempts to link single variants to a specific phenotype, however the extent to which a phenotype is likely to arise is composed of multiple genetic elements that interact in complex ways (Flint and Mackay 2009). Similarly, no one-to-one causal link exists with CRC genes, instead genes tend to aggregate and cluster in families and the variation in phenotype are likely to be influenced by multiple modifier genes. The interaction between genetic variants that results in a phenotypic effect is conditional on the combined presence of two or more variants known as epistasis (Storey *et al*., 2005; Nagel, 2005; Nagel, 2001; Nagel and Steinberg, 2001). It becomes apparent that the products of genetic epistasis and action of modifier genes are possibly 'one and the same' and therefore it is necessary to model the effect of these genes or modifiers in interacting groups (Clark *et al*., 1997; Cordell 2002; Storey *et al*., 2005).

Since most quantitative traits are due to the affect of multiple genes, MQM allows for more information to be extracted from the data to make correct inferences about the genetic makeup and interactions of trait complexes (Korol *et al*. 1998; Jiang and Zeng 1995; Knott and Haley, 2000; Marlow *et al*., 2003). Traditionally, most statistical models used in QTL mapping do not model polygenic variation despite

considerable evidence that interaction between loci affect quantitative traits (Barton and Keightley 2002). This difficulty arises primarily because it is unclear which variant combinations should be tested and under which model of epistasis. Furthermore, multiple confounding factors such as computational burden, correction for multiple testing, false positives, sample size and defining an interaction model have led to the dissemination of numerous quantitative genetic models and much confusion over definitions and interpretation of epistasis. There is no shortage of epistatic QTL mapping software tools available which have been proposed to model simultaneously the effects of multiple QTLs and their interactions (Kao *et al*., 1999; Jannink and Jansen, 2001; Sen and Churchill, 2001; Carlborg and Andersson, 2002; Boer *et al*., 2002; Kao and Zeng, 2002; Yi and Xu, 2002; Ljungberg *et al*., 2004).

To date, MQM studies have been most feasible in model organisms involving genes that have been well characterised; one of the best examples of epistasis is coat colour in mice, where an adaptive transition from dark to light coat colour involves structural changes to the *agouti* and *Mc1r* loci as mice of dark coloured fur move from the forest to the beach (Steiner *et al*., 2007). Obesity in mice has been discovered to involve a network of epistatic interactions on Chr 4, 17 and 19 controlling regulation of fat pad deposits and body weight (Stylianou *et al*., 2006). Gerlai showed evidence of epistasis by demonstrating that null-mutant mouse phenotype was dependent on modifier genes at many loci (Gerlai 1996) which was also seen in fruit fly (de Belle and Heisenberg 1996). Fijneman and colleagues also showed that the quantitative variation in mice susceptible to lung cancer was significantly influenced by interactions among QTLs (Fijneman *et al*., 1996). The

majority of the published examples involve some knowledge about existing genes, which can be sculpted to display a particular phenotype. In contrast, it is far more challenging to identify and test interactions where biological clues of gene function are unknown (Cordell, 2009). Cordell and colleagues evaluate the degree to which statistical tests of epistasis in experiments can actually elucidate underlying biological interactions and pathways and propose this to be more limited than originally estimated (Cordell *et al*., 2002). In addition the extent of epistasis and identification of loci has highlighted the importance of examining phenotype variance using a more global approach where tests in different genetic backgrounds could possibly uncover new variants. It is also unclear whether epistasis observed at the population level can be useful to infer epistasis at an individual level.

Despite these challenges the prevalence and biological significance of epistasis is of great interest, especially in the new arena of GWAs and eQTL mapping. The rapid and modern development of molecular technologies has seen a renewed interest in mapping QTLs at multiple levels in genome-wide case-control studies (Storey *et al*., 2005; Boone *et al*., 2007). To search for n QTLs in n-dimensional scans impacts on the computational processes significantly, especially when using permutation tests to obtain empirical significance thresholds. Therefore, the framework of most epistatic QTL tools based on this condition can be broadly classified into two groups: exhaustive that explicitly test every possible interaction in a enumerate manner (Marchini *et al*., 2005) and non-exhaustive which uses a genetic algorithm such as DIRECT to scan the genomic space for interactions and is non-iterative. DIRECT was originally presented and adapted by Jones and colleagues to divide the search space systematically into smaller boxes, calculating the LOD scores for the center of each

box and using a genetic optimization algorithm to optimize output of each box (Jones, 1993; Carlborg 2000). This process continues until convergence of a defined LOD value until it locates a region of global optimum. DIRECT is an efficient mapping method for investigating epistatic pairs while reducing the search space size by omitting the calculation of QTL pairs (Carlborg 2000). Programs such as GeneNetwork and Pseudomarker, implement rapid DIRECT algorithms to efficiently search epistatic interactions (Ljungberg *et al*., 2004). These programs use the 'pair-scan' function to evaluate non-parametric significance of joint LRS values between QTL pairs; it has a quicker run time but is sensitive to outliers and sample size. As this approach is based on the assumption that low LOD values are 'not of interest', detection of biologically relevant interactions may be omitted.

Traditional exhaustive search engines calculate all LOD scores, as evidence of suggestive and or marginal QTLs (Dupuis *et al*., 1995). These can be shown to be involved in interactions between epistatic QTL pairs with much larger effects than their individual effects (Gurganus *et al*., 1999; Carlborg *et al*., 2000) and it is feasible to apply this approach without the risk of spurious detection, possibly missing interactions, computer and memory failures. Karl Broman's R-qtl software is free and open source multi-platform software. R-qtl is implemented as an add-on package for the statistical language/software R (R Development Core Team, 2010) and is made available from http://www.rqtl.org/ and implements an exhaustive search using the "scantwo" function. It is designed to cope with large datasets to handle larger calculations with greater provisions for controlling spurious associations which reflective of high through put data generation from eQTLs experiments.

As mentioned in chapter one, the identification of the specific genes underlying QTLs is a major challenge and while critical to localising influential regions this tends to be achieved with relatively low resolution. Intervals are typically localised to 10-40 centimorgans (cM) which can harbour hundreds to thousands of genes, any number of which could play a plausible role in, for example, the irradiation induced polyposis model investigated here. Clearly, additional methods aimed at narrowing the QTL interval are vital if the specific genetic variations are to be realistically determined. Though this resolution can be improved marginally by using more markers, the most effective way to improve resolution is to increase pedigree size and the number of recombination events and use advanced animal crosses including congenics. However, this approach has significant cost implications because of the animal husbandry requirements.

The well documented difference in polyp severity between BALC/c and B6 mice offers the opportunity to study how genetic differences between the two strains influence polyp multiplicity via their additive effects and to identify modifiers that may also influence the trait via gene-gene interactions. We have detected QTLs with significant and suggestive levels of significance within the different segments of the intestine (Table 2-3 and 2-4 for 2Gy and 0Gy, respectively).

Next, we will create sets of two-locus models to look at the possible interactions, encompassing all marker pairs to identify any interaction and their statistical importance in a 'scantwo' 2-dimensional exhaustive scan. I now aim to use our QTL model to study epistasis at a genome wide level. The possibility that main effect QTLs may be involved in an epistatic role can be investigated using this approach

and in a broad statistical sense; we aim to create multiple dimensions between markers with two-locus genotype combinations and measure effects by taking the additive element away to account for a model of interaction. We will consider epistasis as the effects of one or more genes involved in polyp multiplicity and are modified by the presence of other gene(s), therefore to be the deviation from the additivity of the genotype values as outlined by Cordell (Cordell 2002). Our method takes into account the additive effects of two loci that confer risk independently and epistatically. Therefore, this chapter considers the interaction modelling of our main-effects QTL modelling to determine epistatic basis for polyp multiplicity in *Apc*$^{min/+}$ mice.

3.2 MATERIALS AND METHODS

3.2.1 A two-dimensional, two locus genome scan

The possibility of epistatic effects was investigated using R-qtl program written for the R statistical language programming environment as described by Broman and colleagues (2003). The program uses Hidden Markov Model (HMM) (Baum *et al*., 1970) algorithms to estimate the most likely genotype and location of pseudo markers at fixed intervals which are then matched to representative markers or main-effect QTLs identified by IM. The R programming scripts written by Karl Broman were edited and adapted as required to perform analysis according to our data and QTLs detected from chapter one (Broman *et al*., 2003).

Starting with a null model we performed a single-QTL genome scan, using 'scanone' at 256 imputations of unknown genotypes based on observed marker data were made at 2.5cM intervals and considered sufficient to generate a precise map. The position (and representative marker) with the highest LOD score was selected and missing genotype data were filled at this marker using the 'geno.fill' action. A two-dimensional QTL scan using 'scantwo' function was implemented using multiple imputations method to reveal and further evaluate pairs of intervals (loci) across the genome to determine how much of the variability in the trait can be explained jointly by two putative QTLs and may epistatic interactions. This function evaluates epistasis using four linear models: the 'Full' (additive plus interaction between 2 QTLs); 'Additive' (additive of 2 independent QTLs); 'One' (single QTL exists on any one chromosome); and each of these models were tested against the fourth 'Null'

model of no QTL. We considered five maximum LOD scores ($\log_{10}$ likelihood for QTL) over a pair of chromosomes; M_f, M_a, M_i, M_{fv1}, M_{av1} and M_1 that are defined using the models mentioned above. M_f considered the fit of the full two-locus model as $\log_{10}$ likelihood ratio comparing the full model with QTLs on two chromosomes to the null model. M_a was equivalent for a two-locus additive model and indicated the presence of at least one QTL, assuming no interaction. M_i was derived, therefore, to examine interaction by comparing the full model with the additive model ($M_i = M_f - M_a$). M_{fv1} and M_{av1} provided evidence for a second QTL, where M_{fv1} allows for epistasis by compared the full model for a QTL pair to the single model (M_1) that one single QTL exists to the null model ($M_{fv1} = M_f - M_1$) and M_{av1} compared the equivalent additive LOD to a single QTL model indicating a second QTL assuming no epistasis ($M_{av1} = M_a - M_1$). M_{fv1} allows for epistasis between the QTLs in a two-locus model and, in the absence of epistasis M_{av1}, would give greater power to detect a second QTL with additive effects. Hence, the two-dimensional genome scan and 'scantwo' method of pairs of markers across the genome will permit the identification of marginal QTLs for LOD scores that are most likely to be epistatic QTLs using the multiple imputation method. Significance thresholds (T_f, T_a, T_i, T_{fv1}, T_{av1} and T_1 were set at 5% significance level for false positives as suggested at 6.50, 4.34, 4.52, 5.43, 2.93 and 2.51 for M_f, M_a, M_i, M_{fv1}, M_{av1} and M_1 respectively. These are the estimated 95th percentiles of the null distributions of the corresponding LOD scores, obtained by 1,000 permutations based on the USI dataset. In order to identify loci of interest from the 2D two-QTL genome scan, the distribution of five LOD scores under the null hypothesis was set. Using the following criteria a pair of QTL was declared of some interest if either of the following held true: 1. $M_f \geq T_f$ (LOD 6.5) or [$M_{fv1} \geq T_{fv1}$ (LOD 5.43), or $M_i \geq T_i$, (LOD 4.52)]; 2. $M_a \geq T_a$ (LOD 4.34) and $M_{av1}, \geq T_{av1}$, (LOD 2.93).

3.2.2 Statistical modeling using ANOVA and effect plots

In the second part of our strategy following a 2D scan of the main-effect QTLs the interaction of epistatic pairs we used the 'makeqtl' function (Broman *et al*., 2003) to create a model, hereafter referred to as a QTL-object, which consisted of multiple QTLs specified according to QTL pairs. The 'fitqtl' function then fitted a specific QTL model in a two-locus hypothesis testing framework, a full model of interaction and principal additive effects was compared with a reduced model in which QTLs are removed sequentially and evaluated by the LOD score to test for the significance of each QTL and associated interactions within the QTL-object. Up to seven pairs of QTLs were matched and investigated under a multi-component ANOVA model. In order to perform more precise mapping the 'refineqtl' function was used to refine the estimates of QTL positions in the current model where the maximised positions for full and additive models maybe different from the pair positions. Pairs of QTL loci, which surpassed the LOD full threshold and showed statistical evidence for interaction using ANOVA (p value >0.05), were analysed using effect plots (Fox *et al*., 2003) to illustrate the interaction effects between markers. The interaction is measured as the change in the phenotypic mean over the mean of another locus as a function of QTL genotypes and by considering two QTLs at a time. The possibility of an interaction existed when the differences at one genotype depend on the genotype at the second loci visualized on an effect plot when the lines describing the main-effects do not run in parallel to each other.

Pairs of QTL loci which surpassed the LOD full threshold and showed statistical evidence for interaction using ANOVA were analysed using effect plots to illustrate

the effects between markers which are not independent but linked for interesting interactions based on QTL pairs following a 2D genome scan. The interaction is measured as the change in the phenotypic mean over the mean of another locus as a function of QTL genotypes, considering two QTLs at a time. The possibility of an interaction exists if the lines describing the main effects were not parallel.

The R-qtl 'fitqtl' function provided estimates of the QTL effects which are derived as a matter of recoding the homozygous and heterozygous markers to -0.5 and +0.5 respectively and reporting the differences between the averages. While it is useful to consider the QTL effects involved in epsitasis, there is generally considerable bias which was likely to vary from the true QTL effect. QTLS that are above a certain size and of statistical significance are likely to be reported and therefore overestimated; the detection of weak effects QTLs was likely to be missed to due to the low power in detection. In addition, modelling genes in humans using any organism model the QTL effects and the actual alleles might not be involved in the same way functionally and so the relevance in measured QTL effects to human diseases will remain open to question and for these reasons not much emphasis was placed on measuring the epistatic QTL effects.

3.2.3 Exploration of models of multiple interacting QTLS

Multiple interacting QTLs were investigated using a forward and reverse stepwise search and penalised LOD (pLOD) criterion (Manichaikul *et al*., 2009). The pLODs are weighted according to each main-effect QTL and each pairwise interaction to control the rate of false interacting QTLs calculated from 2D scantwo of 10,000

permutation tests simulated on salt-induced hypertension in mice data from Sugiyama and colleagues, which consisted of 250 male backcross individuals and markers spaced at 10cM (Sugiyama *et al*., 2000) and set at 2.69, 2.62, 1.19 LOD for main-effects (T_m), heavy (T_h) and light interactions (T_l), respectively. The pLOD score took into account the initial LOD of the model and calculated the penalty based on the number of QTL in the model that are main-effects (P_m), the number of pairwise interactions that will be given a heavy interaction penalty (P_h) and the number of pairwise interactions that will be given a light interaction penalty (P_l). Differentiation between heavy and light interactions was made based on whether the interactive QTL of the model was disconnected or connected to other QTLs, respectively. A model was chosen that had the largest model comparison criteria using the pLOD score among all the models visited.

The forward search starts with no QTLs in the model, then adds each QTLs beginning with a QTL with the highest LOD score. Additional QTLs are then selected on to give the greatest increase in LOD score. The reverse of the forward selection is backward elimination, which starts with all QTLs in the model, and then removes each QTL one at a time so as to give the smallest decrease in the LOD score. These processes involve testing at each step for improvements to the model and the process is completed when there is no further improvement in the model. Here, two models are considered: an additive only and an additive-interactive QTLs. These models allow the identification of a subset of loci and permits consideration of multiple acting QTLs that may control trait phenotype (Sillanpä *et al*., 1998).

3.3 RESULTS

3.3.1 2D two-QTL scan performed on all trait data of 2Gy

A 2D two QTL scan for epistatic QTL pairs was modelled using the 2Gy data on the USI, LSI and LI. The reported LOD scores for the full model which accounted for the combined main additive effects of the two loci plus interaction between the marker pairs was calculated as the additive affect deducted from the null model. We focused our analysis on the 2Gy to model the main effect QTLs as the 0Gy yielded suggestive QTLs and so the power to detect interaction effects from these marginal QTLs was further limited and subject to caution.

Following a 2D two-QTL scan for epistatic effects 27 QTL pairs were detected for the 2Gy dataset where the LOD_f (full model) value equalled or exceeded the LOD threshold of 6.0 and 1 QTL pair which fell just below this threshold was included due to the inferred location of the second QTL in close proximity to *Mom7* on chr18. Interactions between QTL pairs are shown in Table 3-1 where LOD scores for full models > 6.5 threshold (set at 5% genome wide error rate). All other QTL pairs which did not reach above the 6.0 LOD_f threshold were not investigated.

Table 3-1 28 Quantitative trait loci pairs detected from 2D two QTL genome scan involving main-effect loci. USI=upper small intestine, LSI=lower small intestine, LI=large intestine, cM=R-qtl generated positions, Add=additive model, Int=Interactive model.

	Chromosome (C)	QTL 1 (cM)	QTL 2 (cM)	LOD Full (M_f)	LOD Add (M_a)	LOD Int (M_i)	Modelling
USI	C2:C5	*D2MIT395 Mrip1* (60.0)	*D5MIT139 Mrip3* (64.2)	9.79	8.37	1.41	QTL pairs involving two main-effect QTLs
	C2:C16		*D16MIT189 Mrip5* (46.8)	7.29	6.76	0.53	
	C5:C16	*D5MIT201 Mrip2* (39.6)	*D16MIT189 Mrip5* (46.8)	10.4	9.63	0.77	
	C5:C17	*D5MIT139 Mrip3* (64.2)	*D17MIT180* (26.7)	8.51	7.62	0.89	
	C16:C17	*Prkdc Mrip4* (25.4)	*D17MIT180* (26.7)	6.8	6.03	0.77	
	C5:C9	*D5MIT201 Mrip2*	D9MIT90 (17.8)	8.03	6.83	1.19	QTL pairs central to *Mrip2* (39.6cM)
	C5:C10		D10MIT106 (11.7)	7.1	6.04	1.06	
	C5:C14		D14MIT102 (34.4)	6.9	6.01	0.89	
	C5:C1	*D5MIT139 Mrip3*	D1MIT132 (39.5)	8.64	7.17	1.47	QTL pairs central to *Mrip3* (64.2cM)
	C5:C4		D4MIT308 (57.7)	6.88	6.42	0.46	
	C5:C12		D12MIT91 (30.1)	6.75	6.17	0.58	
	C5:C13		D13MIT78 (67.2)	7.69	6.48	1.21	
	C5:C15		D15MIT100 (19.3)	7.01	5.94	1.06	
	C5:C18		D18MIT222 (8.1)	7.71	6.62	1.1	
	C5:C3	*D5MIT139 Mrip3*	D3MIT320 (66.8)	7.01	7.01	0	No QTL object model due to low LOD Int (M_i)
	C5:C6		D6MIT194 (62.9)	6.93	6.63	0.3	
	C5:C7		D7MIT228 (25.6)	7.25	6.94	0.31	
	C5:C8		D8MIT292 (21.2)	6.53	6.33	0.19	
	C5:C11		D11MIT216 (19.2)	7.01	6.66	0.35	

	C5:C19		D19MIT106 (10.8)	6.59	6.54	0.05	
	C6:C6	D6MIT83 (5.65)	D6MIT1 (6.65)	6.54	5.58	1.96	
	C5:C5	*D5MIT201 Mrip2* (39.6)	*D5MIT139 Mrip3* (64.2)	6.93	6.9	0.03	*cis*-epistasis
	C16:C16	*D16MIT189 Mrip5* (46.8)	D16MIT106 (57.7)	6.95	5.73	1.22	
LSI	C6:C3	*D6MIT194 (63.2)*	D3MIT320 (51.8)	6.95	3.36	3.59	
	C6:C18		D18MIT222 (23.1)	5.33*	2.31	3.02	
	C3:C3	D3Mit51 (34.3)	D3MIT320 (51.8)	6.56	3.2	3.27	
LI	C2:C8	*D2MIT395 Mrip1 (63.4)*	D8MIT292 (27.1)	6.64	5.75	0.89	
	C8:C9	D8MIT211 (47.1)	D9MIT335 (27.5)	6.01	4.01	2	

3.3.2 Most detected QTLS pairs of the upper small intestine involve interactions.

From the 23 detected QTL pairs of the USI involved interaction, 6 putative interactions involved main-effect QTLs on both chromosomes and were investigated as QTL-object 1 a total of 22 QTL pairs involved at least one main effect QTL and second QTL inferred on a non main-effect QTL. 18 QTL pairs were central to Chr5; *Mrip2* (D5Mit201) and *Mrip3* (D5Mit139) and investigated as QTL object models. Three QTL-pairs were found to exist in the same chromosome and investigated for cis-epistasis as outlined in Table 3-1.

2D-2QTL scan of upper small intestine (2Gy)

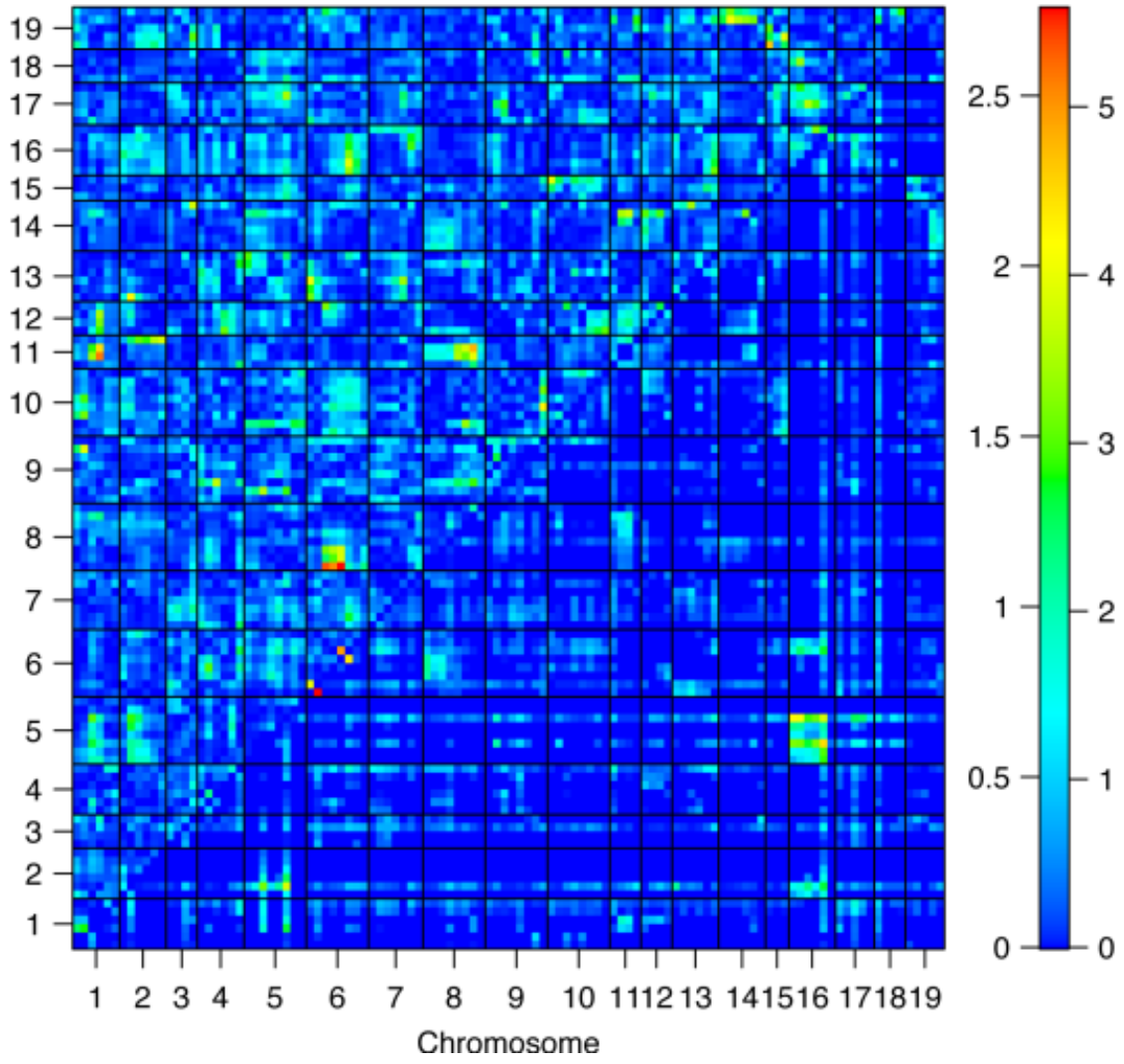

Figure 3-1 LOD scores from a 2D two QTL genome scan of the upper small intestine trait on all chromosomes of 2Gy irradiated mice data. LOD interaction (LOD_i) is displayed in the upper left triangle; LOD full (LOD_{fv1}) is displayed in the lower right triangle which allows for epistasis as evidence for a second QTL by a measure of the LOD_f – M_1 model (a single QTL). The colour scale on the right and numbers to the left and right correspond to the LOD $_i$ and LOD $_f$, respectively.

The lower triangle (LOD_{fv1}) of the 2D scan (Figure 3-1) shows evidence for QTLs on Chr 2, 5 and 16 and other multiple marginal affects QTLs as detected in our previous chapter. Evidence for putative interaction between QTL pairs is shown in the upper

left involving significant or suggestive main-effect QTLs between Chr 2:5, Chr 2:16, Chr 5:16 ,Chr 5:17, Chr 16:17 (QTL object 1).

Interactions central to *Mrip2* or *Mrip3* on Chr 5 were investigated in the following sequence: *Mrip2* with Chr 9, 10, or 14 (QTL object 2) and *Mrip3* with Chr 1, 4, 12, 13, 15 or 18 (QTL object 3). These models were tested using ANOVA analysis and effect plots to confirm interaction in subsequent sections of this chapter. Additional putative QTL interactions between: Chr 6:8, Chr 1:11, Chr 11:14, Chr 14:15, Chr 14:19 which indicated strong interactions were excluded from further analysis due to low LOD_f (ranging between 2.0- 3.0) and similarly, low LOD_{fv1} and LOD_{av1} values which did not sufficiently indicate the presence of second interactive and additive QTLs.

3.3.3 Quantitative trait loci pairs assessed for cis-epistasis

Linked QTLs by repulsion or in coupling which may suggest cis-epistasis were analysed for Chr 5, 16 and 6 of the USI and no evidence for cis-interaction was detected on either of these chromosomes. Chr5 double peaks confirmed separate unlinked QTLs analysis using R-qtl. ANOVA and interaction plots confirmed that these loci are unlinked and act additively (Figure 3-2 A) and the effect of Chr 5 *Mrip2* locus is the same for each of the two genotypes at Chr 5 *Mrip3* and *vice versa* and therefore loci have the same sign with individuals (AA) having higher polyp numbers.

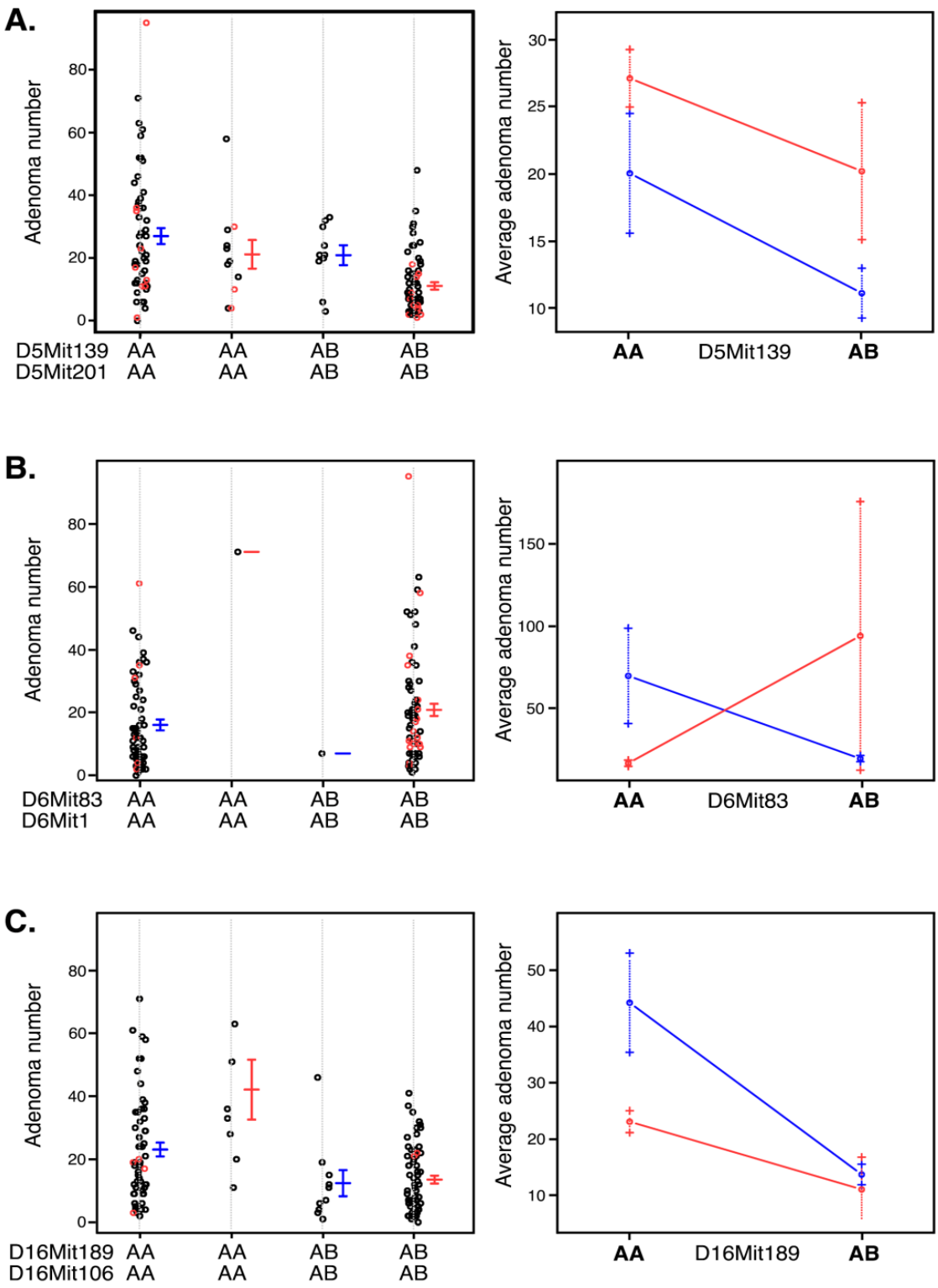

Figure 3-2 Investigation of cis-interactions involving tightly linked QTLs on Chr 5 (A), Chr6 (B) and Chr16 (C) of the USI trait. Dot plots to show the polyp numbers of individuals against genotype clearly showing that very low number of individuals exist between recombinant classes (AA/AB). Effect plots to show phenotype

averages against two-locus genotypes. A=BALB/c allele (homozygous) and B=ApcMin/+ allele (heterozygous). Black circles =observed genotypes, red=imputed for missing data.

The effect plots for Chr 16 (Figure 3-2 C) show that the markers have effects of the opposite sign; individuals AA for D16Mit189 marker and AB for D16Mit106 have the greatest number of polyps, however the inference of the second QTL; D16Mit106 is based on a few individuals. *Cis*-epistasis on Chr6 was investigated between two markers which were typed at a 1cM distance from each other and appeared to be linked in repulsion with the opposite effect; that is whether an individual is AA at one marker and AB at the second or *vice versa* the outcome of high polyp numbers is observed. In both cases, Chr16 and 6 cis-interactions were considered artefacts on the basis of only a few individuals shared the genotype of the second inferred and not substantive and considered an artefact (Figure 3-2 B).

3.3.4 Quantitative trait loci pairs detected for the lower small intestine trait.

A total of nine QTL pairs were detected for the LSI trait; eight were selected with LOD_f >6.0 and 1 QTL where LOD_f = 5.33 (Figure 3-3). One of the two QTL pairs involving a main-effect QTL on Chr 6 (D6Mit194) confirmed *trans*-epistasis (p = 0.0401) using ANOVA with Chr 3 QTL, but no interaction was noted for the QTL pair on Chr18 (region close to the *Mom7*). Seven QTL pairs involved the same chromosome (on Chr 1, 3, 4, 8, 11, 12, 18) and were suggestive of tightly linked QTLs. Except for Chr 3 the QTL pairs were either 5cM or 10cM apart and not typed by a separate marker therefore declared as artefacts and Chr 3 QTL pair did not

confirm *cis*-epistasis despite being typed by a different marker and a distance of 17.5cM the rest (data not shown).

2D-2QTL scan of lower small intestine (2Gy)

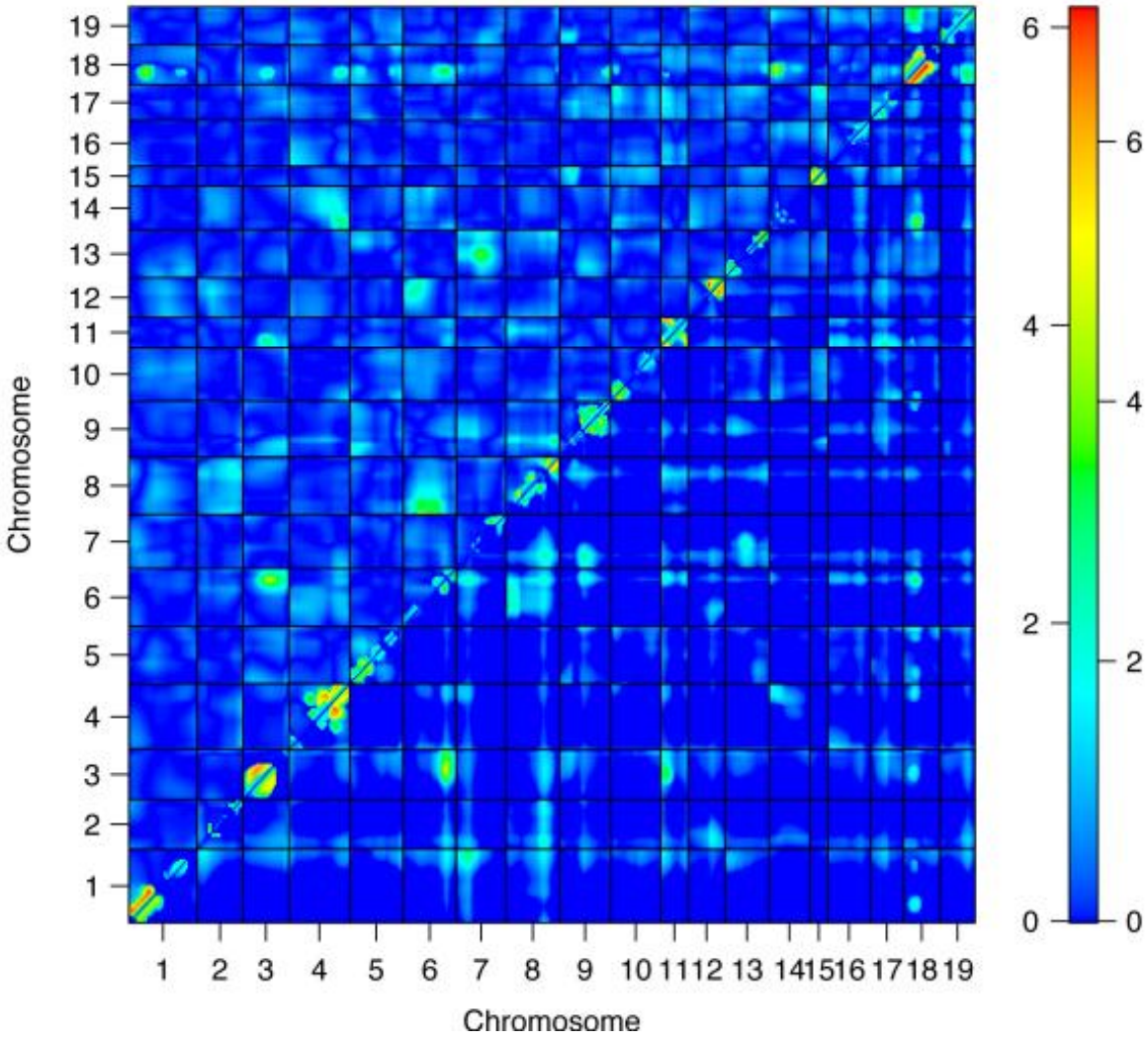

Figure 3-3 LOD scores for all chromosomes, from a 2D, two QTL genome scan of the lower small intestine trait on all chromosomes of 2Gy irradiated mice data. LOD interaction (LOD_i) is displayed in the upper left triangle; LOD full (LOD_{fv1}). The colour scale on the right and numbers to the left and right correspond to the LOD_i and LOD_f, respectively.

3.3.5 Quantitative trait loci pairs detected for the large intestine trait.

Two QTL pairs were detected at LOD_f for the LI trait; one pair involved a significant mains effect QTL on Chr 2 D2Mit295 with Chr 8 (D8Mit292) and the second QTL pair involved novel and previously undetected QTLs, D8Mit211 and D9Mit335 (Figure 3-4). No significant interactions were noted for either of these QTL pairs.

2D-2QTL scan of large intestine (2Gy)

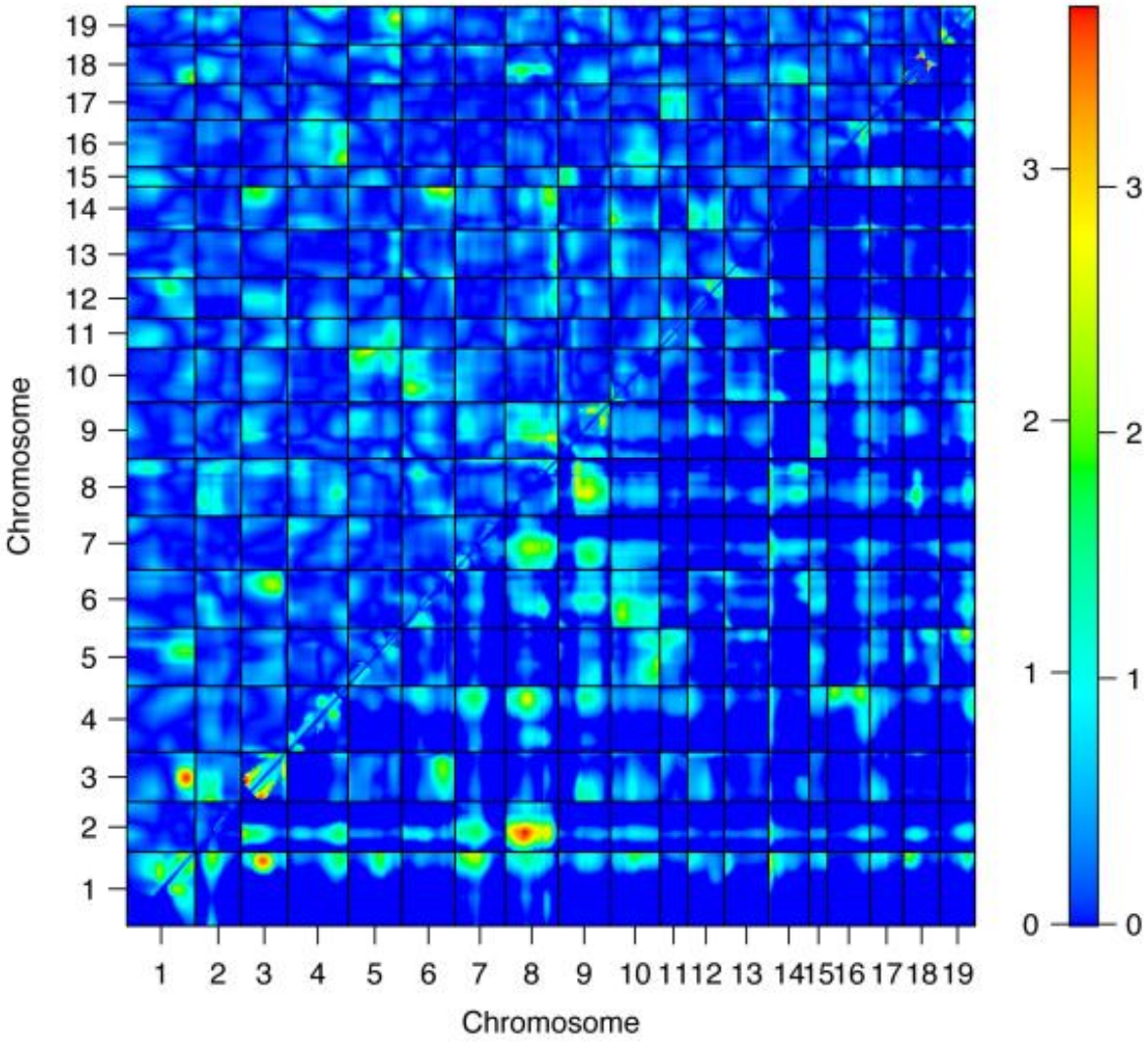

Figure 3-4 LOD scores for all chromosomes, from a 2D, two QTL genome scan of the large intestine trait on all chromosomes of 2Gy irradiated mice data. LOD

interaction (LOD_i) is displayed in the upper left triangle; LOD full (LOD_{fv1}). The colour scale on the right and numbers to the left and right correspond to the LOD $_i$ and LOD $_f$, respectively.

3.3.6 Interaction of main-effect QTL pairs in the USI

The results of scanone and scantwo indicated main effect QTLs on Chr 2, 5, 16 17 with the possibility of second QTLs on Chr 5 (*Mrip3*), 16 (*Mrip4*) and six possible interactions as indicated in Table 3-1. We investigated the possibility of epistasis by fitting a six-QTL model using the 'fitqtl' function taking the positions from scantwo to create QTL object 1 as outlined in Table 3-2.

Table 3-2 ANOVA table for modelling epistasis of main effects QTLs in QTL-object 1. Type III SS= sum of squares, %var=estimated proportion of the phenotype variance explained by the terms in the model. F statistic=the ratio of the mean squares (the sum of squares divided by the df to the error mean square, p values are based on the F statistic.

QTL object 1	df	Type III SS	LOD	%var	F statistic	p values (F) significance	
Chr 2 *Mrip1* (60.0cM)	3	1846.01	2.69	4.98	3.92	0.0104	*
Chr 5 *Mrip2* (39.5cM)	3	1527.61	2.24	4.12	3.24	0.0240	*
Chr 5 *Mrip3* (64.22cM)	4	1964.69	2.85	5.30	3.13	0.0169	*
Chr 16 *Mrip4* (25.4cM)	2	1784.50	2.60	4.81	5.69	0.0042	**
Chr 16 *Mrip5* (46.8cM)	3	1594.85	2.34	4.30	3.39	0.0200	*
Chr 17 D17Mit180 (26.7cM)	3	3108.04	4.40	8.39	6.61	0.0003	***
Chr 2:5 (*Mrip1:3*)	1	998.84	1.48	2.69	6.37	0.0127	*
Chr 2:16 (*Mrip1:5*)	1	6.66	0.01	0.01	0.04	0.8369	
Chr 5:5 (*Mrip2:3*)	1	164.87	0.25	0.44	1.05	0.3069	
Chr 5:16 (*Mrip2:5*)	1	618.49	0.92	1.66	3.94	0.0490	*
Chr 5:17 (*Mrip3*)	1	120.54	0.18	0.32	0.76	0.3820	
Chr 16:17 (*Mrip4*)	1	1732.36	2.53	4.67	11.0	0.0011	**
Error	129	20215.21					

Total	141	37035.75				

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05

The overall fit of the full model (six-QTL) is relative to the null and includes all the terms of the model with 12 degrees of freedom (df) and a LOD score of 18.68. This model explains a substantial proportion of the phenotypic variance (45.4%). Reducing the model by dropping one of the specified QTLs at a time and comparing the LOD score relative to this 'reduced' model (and not the null model) using ANOVA indicated strong evidence for all main effects loci included in this model: Chr 2 (*Mrip1*), Chr 5 (*Mrip2* and *Mrip3*) and Chr 16 (*Mrip5*) ($p \leq 0.05$), Chr 16 (*Mrip4)* ($p \leq 0.01$) and Chr 17 (D17Mit280) ($p \leq 0.001$), significant interactions between; *Mrip1:3* (Chr 2:5), *Mrip2:5* (Chr 5:16) ($p \leq 0.05$) and 6 *Mrip4*:D17Mit280 (Chr 16:17) ($p \leq 0.001$), in total accounted for ~9% of the phenotypic variance. No evidence was detected for the other interactions within this model. The refined position and LOD values of the QTL model showed a 0.07 LOD increase after 2 increments with very minor positional changes of each loci.

QTL pairs, which were shown to have significant interaction determined by ANOVA, were investigated for interactions using effect plots (Figure 3-5). Additively acting QTL pairs are not shown. According to our BC, (AxB) xA, heterozygous AB from the $Apc^{min/+}$ mouse carrying the $Mom1^{R}$, (resistant strain) are expected to have lower polyp numbers than homozygous AA alleles and this was observed by effect plots. The QTL pairs of Object 1 show that Chr 5 (D5Mit139/*Mrip3*) AA allele has an effect in the presence of heterozygous alleles (AB) at Chr 2 D2Mit395/*Mrip1*). Similarly, Chr 2 has an effect only in the presence of the (AA) at Chr 5. Therefore, only

individuals that are AA at Chr 5 and AB at Chr 2 show high polyps multiplicity in the USI. The other 3 genotypes have similar effects on polyp numbers.

Similar findings were observed for QTL pair between *Prkdc* (AA) and heterozygote (AB) at Chr 17 (D17Mit280). Finally, individuals that are AA at Chr 5 (D5Mit201/*Mrip2*) and AA at Chr 16 (D16Mit189/*Mrip5*) have higher polyp numbers and so the QTL favours the homozygosity at both loci.

Interaction plots concerning main-effect QTL pairs

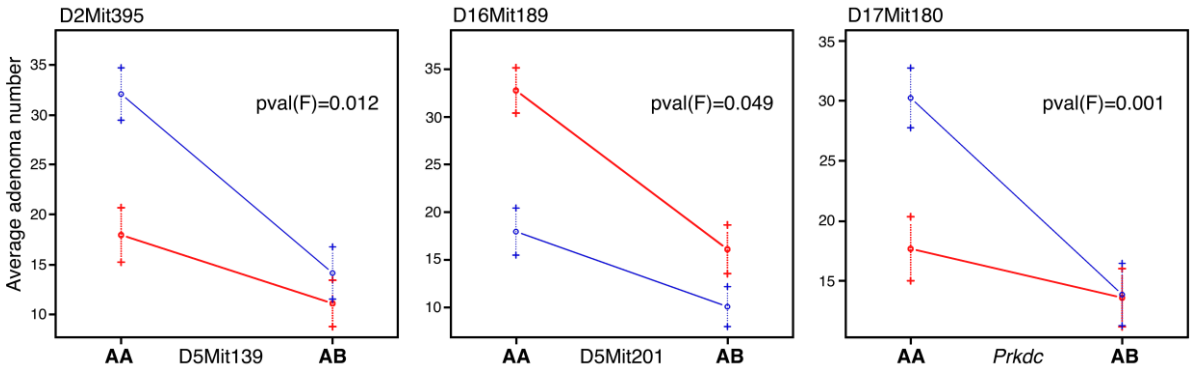

Figure 3-5 Phenotype averages against two-locus genotypes as effect plots for QTL-object 1 of the upper small intestine trait, p value (p val) statistical significance based on F statistic for the six-QTL model as analysed using ANOVA (Table 3-2). A=BALB/c allele (homozygous) and B=Apc$^{Min/+}$ allele (heterozygous).

3.3.7 QTL object 2 pairs central to Mrip2

We next investigated a four-QTL model involving *Mrip2* on Chr 5. The initial full model with LOD score of 9.10 (7df) accounts for 25.5% estimated proportion of the phenotypic variance explained by all terms. The drop one analysis (Table 3-3)

indicated evidence for only one of the four loci, the *Mrip2* locus, but found significant interactions with the marginal effect QTLs on Chr 9, 10 and 14.

Table 3-3 ANOVA for multiple QTL pair modelling and testing significance of epistasis involving *Mrip2*. Type III SS= sum of squares, %var=estimated proportion of the phenotype variance explained by the terms in the model. F statistic=the ratio of the mean squares (the sum of squares divided by the df to the error mean square, p values are based on the F statistic.

QTL object 2	df	Type III SS	LOD	%var	F statistic	P value(F)	significance
Chr 5 *Mrip2* (39.5cM)	4	9014.3	8.72	24.33	10.95	1.03e-07	***
Chr 9–D9Mit90 (17.8cM)	2	1171.2	1.28	3.16	2.84	0.0616	
Chr 10–D10Mit103 (11.7cM)	2	949.3	1.04	2.56	2.30	0.1035	
Chr 14 – D14Mit102 (34.4cM)	2	1175.3	1.28	3.17	2.85	0.0610	
Chr 5: 9	1	833.2	0.91	2.25	4.05	0.0462	*
Chr 5: 10	1	887.5	0.97	2.396	4.31	0.0397	*
Chr 5: 14	1	1088.2	1.19	2.93	5.28	0.0230	*
Error	134	27569.53					
Total	141	37035.74					

Significance codes: '***' 0.001 '*' 0.05

QTL pairs were investigated for interaction using effect plots to confirm interactions involving *Mrip2* (D5Mit201) between markers on Chr 9, 10 and 14 (Figure 3-6). It was confirmed that the heterozygote (AB) at *Mrip2* locus has lowest effect on polyp numbers and homozygous (AA) for *Mrip2* in each QTL pair results in higher polyps: with homozygote individuals at D9Mit90; either AA or AB genotype at D10Mit106 and heterozygote (AB) at D14Mit102.

Interaction plots of QTL pairs central to D5Mit201 (*Mrip2*)

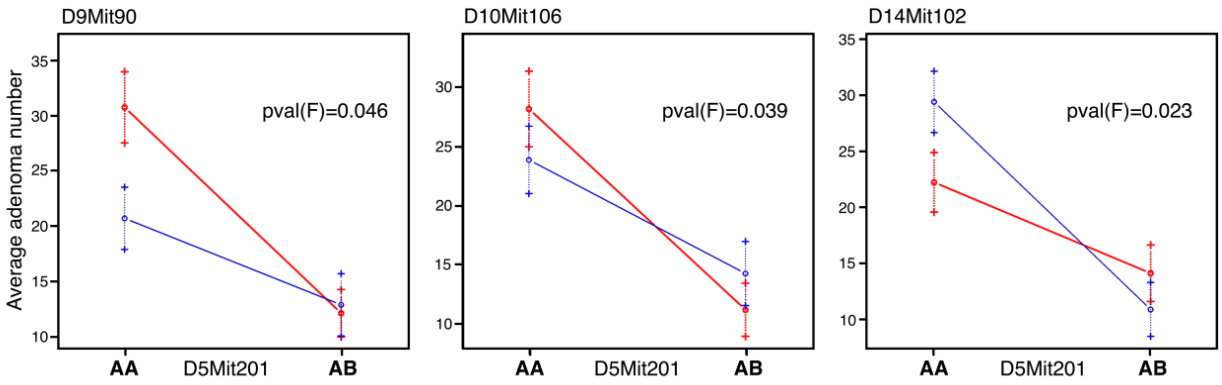

Figure 3-6 Phenotype averages against two-locus genotypes are shown as effect plots for QTL object 2 of the upper small intestine trait. A=BALB/c alleles and B=Apc$^{Min/+}$ allele. P value (p val) statistical significance based on F statistic for the six-QTL model as analysed using ANOVA (Table 3-3).

3.3.8 QTL object 3 pairs central to Mrip3

A seven-QTL model involving Chr 5 (D5Mit139/*Mrip3*) of QTL object 3 was investigated. The initial full model with LOD score of 17.02 (13df) accounts for 42.3% estimated proportion of the phenotypic variance explained by all terms; all loci were significant except Chr 12 and four out of six interactions tested significant under ANOVA (Table 3-4) and shown in effect plots (Figure 3-7).

Table 3-4 ANOVA table for multiple QTL modelling and testing significance of epistasis involving *Mrip3*. Type III sum of squares (SS), estimated proportion of the phenotype variance explained by the terms in the model (%var) F statistic and p values are based on the F statistic, model (M) and interaction (int) as denoted to the QTL pairs.

QTL objects 3	df	Type III SS	LOD	%var	F statistic	P value(F) significance	
Chr 5 *Mrip3* (39.5cM)	7	13147.2	14.8	35.4	11.26	4.56E-11	***
Chr 1 D1Mit132 (39.5cM)	2	2981.1	4.03	8.04	8.94	0.00023	***
Chr 4 D4Mit308 (57.7cM)	2	1078.8	1.52	2.91	3.23	0.04254	*
Chr 12 D12Mit91 (30.1cM)	2	635.2	0.90	1.71	1.90	0.15292	
Chr 13 D13Mit78 (67.2cM)	2	1358.6	1.90	3.66	4.07	0.01923	*
Chr 15 D15Mit100 (19.3cM)	2	1419.6	1.98	3.83	4.25	0.01619	*
Chr 18 D18Mit222 (8.1cM)	2	2411.3	3.30	6.51	7.23	0.00105	**
Chr 5:1	1	2226.9	3.06	6.01	13.3	0.00037	***
Chr5:4	1	622.5	0.88	1.68	3.73	0.05549	.
Chr 5:12	1	540.7	0.77	1.46	3.24	0.07403	.
Chr 5:13	1	985.9	1.39	2.66	5.91	0.01639	*
Chr 5:15	1	1403.6	1.96	3.79	8.42	0.00436	**
Chr 5:18	1	2011.1	2.77	5.43	12.0	0.00070	***
Error	128	21334.40					
Total	141	37035.8					

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

QTL pairs were investigated for interactions using effect plots for evidence of interactions between these loci (Figure 3-7). Additively acting QTL pairs are not shown. The heterozygote (AB) alleles at *Mrip2* locus confirmed that the B allele from B6$^{Min/+}$ yields lower polyps than the A allele from the BALB/c strain. Evidence of epistasis causing higher polyp numbers is observed in homozygote individuals (AA) at Chr 5 (D5Mit139/*Mrip3*) and heterozygotes (AB) at Chr 1 (D1Mit132) and Chr 13 (D13Mit178). In the case of epistasis between Chr 15 (D15Mit100); the

genotype information AA or AB have similar effects on polyp numbers for individuals which are AA for Chr 5 (D5Mit139/*Mrip3*).

Interaction plots of QTL pairs central to D5Mit139 (*Mrip3*)

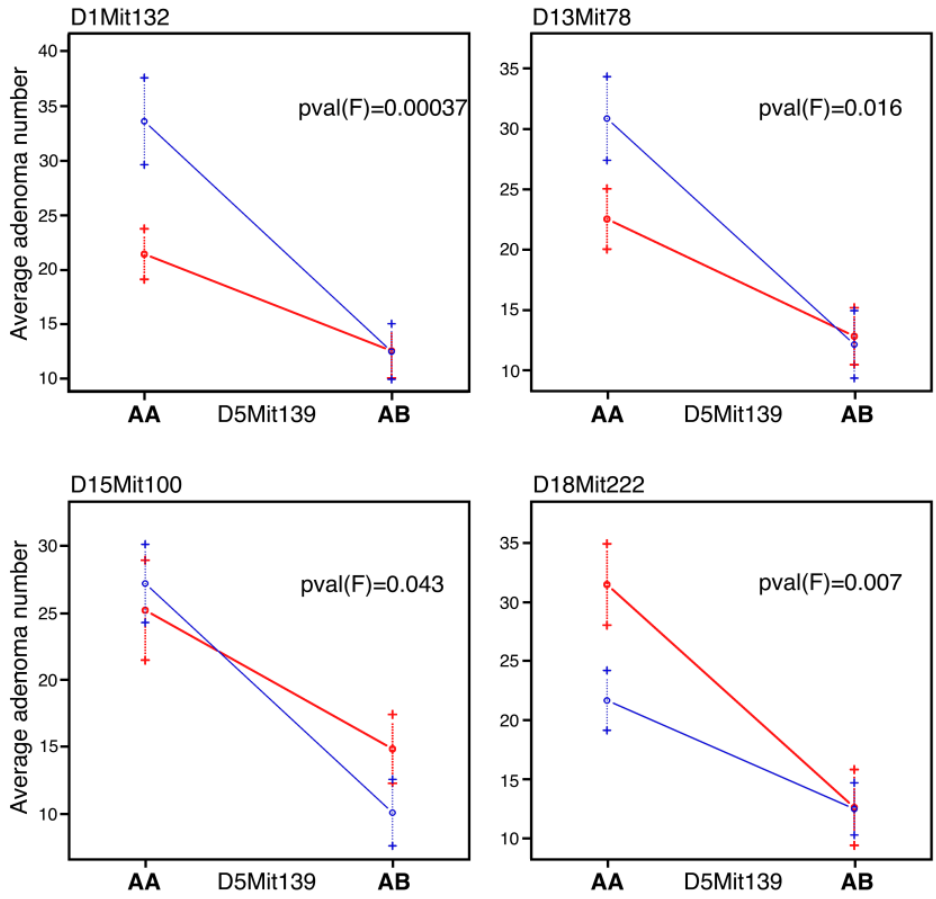

Figure 3-7 Phenotype averages against two-locus genotypes are shown as effect plots for QTL object 3 of the USI trait. A=BALB/c alleles and B=Apc$^{Min/+}$ allele. P value (p val) statistical significance based on F statistic for the six-QTL model as analysed using ANOVA.

Finally, the inference of a QTL interaction between *Mrip3* and Chr 18 (at D18Mit222, 8.08cM) is located close to the modifier of *Min (Mom7)* (3cM), an existing novel modifier of the *Apc*$^{min/+}$ mouse model. Heterozygous individuals carrying the B allele

for D18Mit222 were expected to have lower polyp numbers due to the *Mom7* modifier effect to suppress polyp multiplicity in $B6^{Min/+}$ and this is confirmed in our data. A dot plot (Figure 3-8) of this interaction further shows that the 2 recombinant classes differed in the number of polyps between homozygote (AA) or heterozygote (AB) at D5Mit139 (*Mrip3*). That is homozygote AA at *Mrip3* and D18Mit222 have higher numbers polyp. The inference of the second QTL, D18Mit222, depends approximately on equal numbers of individuals genotyped at AA and AB and shows further reliability of results. However, there is considerably more missing data in the AB genotyped group.

The effect of D5Mit139 and D18Mit222 linked QTLs

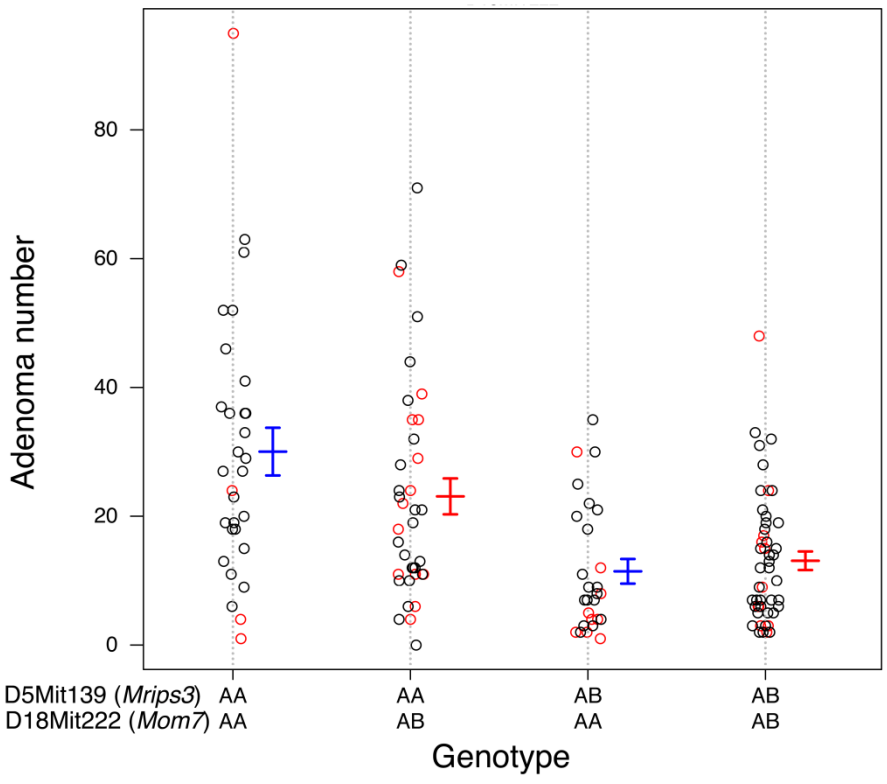

Figure 3-8 The effect of two-linked QTL on Chr5:18 against the upper small intestine polyp phenotype assessed using the 2Gy data. The dot plot of the phenotype as a function of marker genotypes confirms the difference between two

recombinant classes of *Mrip3* AA over AB and D18Mit222 AA over AB for increased polyp severity. Black circles correspond to observed genotypes and red for imputed data where genotype information is missing.

3.3.9 Exploration of additive and additive-interactive models of multiple interacting QTLs

The automated model selection process was applied to identify a subset of loci that influence polyp burden and how these may contribute together to produce the phenotype were evaluated using MQM and a penalized LOD score with penalties fixed for main effects and interactions (Figures 3-9 and 3-10). Starting with a single QTL scan, *Mrip4* (Chr 5) was selected followed by scan for additional additive QTL and any additional interacting QTLs. At each step the location of the QTLs in the current model are refined and repeated until a six-loci model was achieved.

Mrip1, 2, 5 were identified within the first 3 steps of the forward selection process supported by genotype-phenotype data in the USI trait. Additional additive QTLs; Q4 on Chr 6 at 62.9cM (D6Mit194), Q5 on Chr 6 at 53.1cM (D6Mit287) and Q6 on Chr 16 at 55cM (D16Mit106) were selected under the additive model (Figure 3-9) but as these were previously discounted during earlier *cis*-epistasis analysis as linked QTLs not much weight was given to them. Stepwise search for interactive QTLs to add to the model first searched for an additive QTL and then tested for interaction assigning pLODs at each process according to the mains effect or interaction QTL detected.

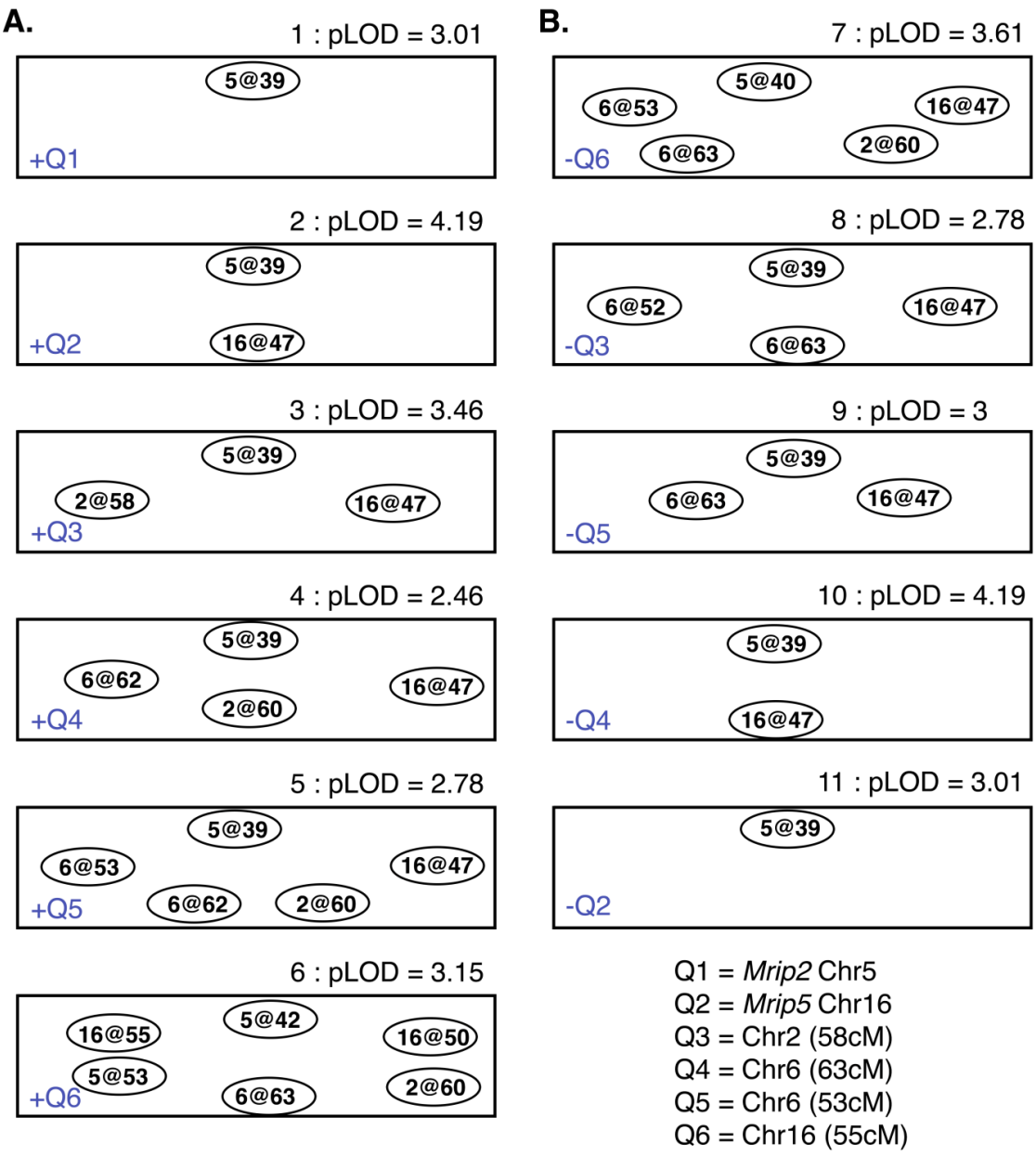

Figure 3-9 The sequence of model selection visited by forward (A) and backward (B) search for additional quantitative trait loci of additive and interaction effects in the upper small intestine using the 2Gy data. The penalised LOD (pLOD) value for adding a QTL (Q) the model as shown in the formula (y) indicating the structure of the additive and/or interactive (connected) QTLs during model selection. The forward selection process is followed by backward elimination set to a maximum of 6 QTLs

A second QTL model was developed while searching for interacting QTLs and involved *Mrip2* and *5* and the suggestive QTL on Chr 17 (Figure 3-10). To this, additional QTLs were added involving Chr 16 (which was previously discounted as linked QTLs) and Chr 3. Chr 3 QTL is a possible overlapping loci between the USI of the 0Gy and 2Gy data, however, caution was applied to the interpretation of these results due to the missing genotype data and redundancy of D3Mit320 and D3Mit352 markers which was detected from Map Manager QTX. R-qtl handled the missing data differently and did not hide these markers from analysis which probably accounts for the difference in analysis. This also indicates that slightly different and dynamic patterns emerge with the use of different software programs all which greatly depend on the statistical inference of missing genotype data.

The model showed an interaction between Chr 5:16 and Chr 17:16 (D16Mit106). The latter interaction differs from the output of the 2D scan and ANOVA analysis which showed a Chr 17:16 interaction to involve *Prkdc* (at 20cM) and not D16Mit106 at 52.6cM. However, in the context of MQM it appears to favour analysis involving peaks distal to the second QTL peak on Chr 16 (D16Mit189/*Mrip5*) based on LOD score evaluation and would seem that the difference in 2D scan and MQM may concern the treatment of missing genotypes. No great difference was noted between the forward or backward elimination approaches.

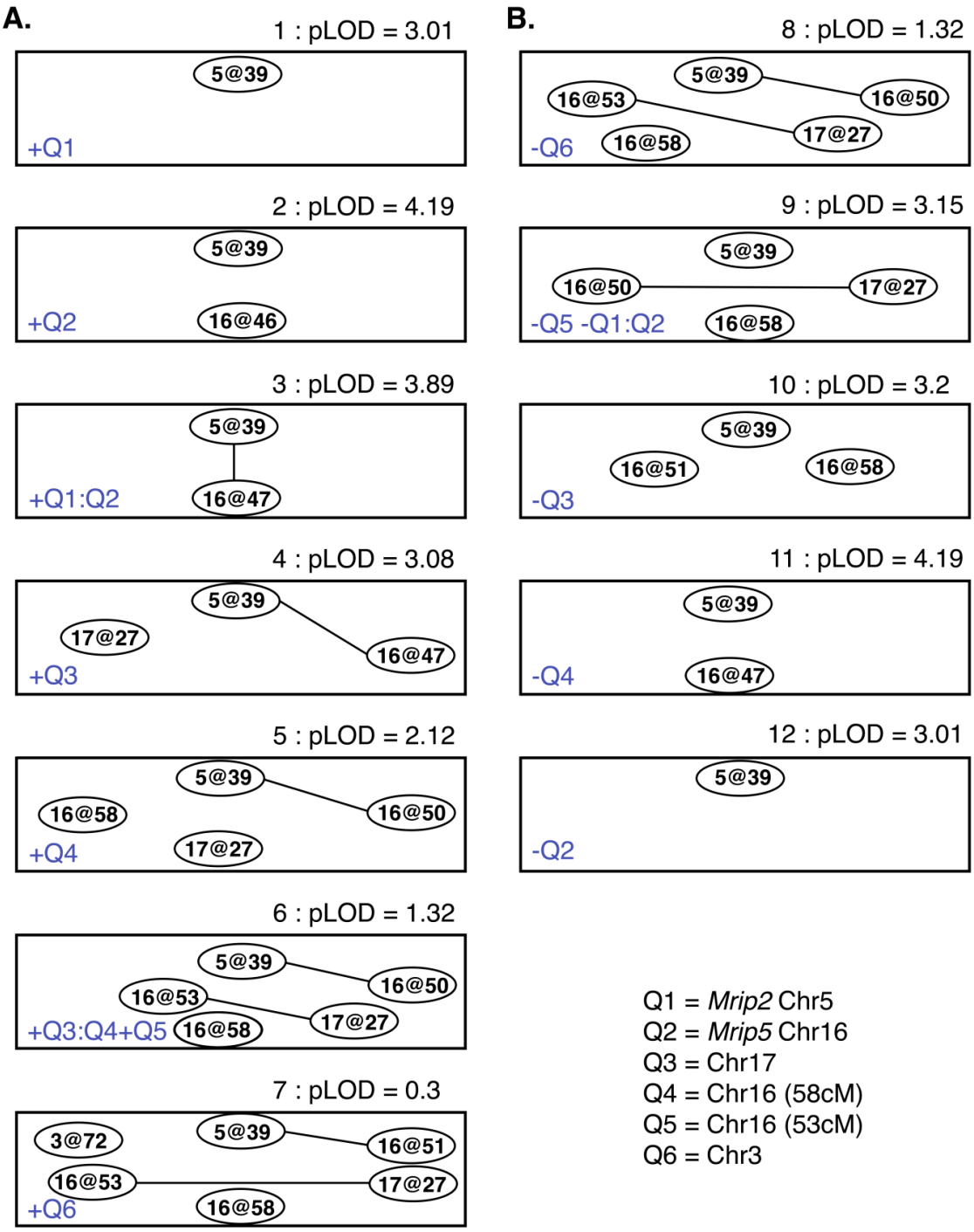

Figure 3-10 The sequence of model selection visited by forward (A) and backward (B) search for additional quantitative trait loci of interaction effects in the upper small intestine using the 2Gy data. The penalised LOD (pLOD) value for adding a QTL (Q) to the model as shown in the formula (y) indicating the structure of the additive and/or interactive (connected) QTLs during model selection. The forward selection process is followed by backward elimination set to a maximum of 6 QTLs.

3.4 DISCUSSION

The findings presented here indicate that statistical epistasis as a property of specific genotypes at two loci whose interaction has an impact on irradiation-induced adenoma multiplicity does exist. The aim of this analysis was to test for the genome-wide effects of interactions based on main-effect QTLs within a statistical framework. The permutation testing which performs 1000 genome scans on permuted dataset to obtain a stable distribution of the model under the null hypothesis of no QTLs existing in the data gave a threshold from 2.53 -2.60 LOD and we were confident that this would not lead to potentially greater numbers of type I errors. We calculated a range of ~8-18% total phenotypic variance overall accounted for by the three QTL-object models of epistasis in that determine adenoma multiplicity. The phenotypic variance of 18% is central to *Mrip2*, 8% central to *Mrip3* and 9% involving two main-effect QTLs.

We used R-qtl to perform an exhaustive search to work through all possible combinations of epistatic QTL pairs using a two-QTL hypothesis testing approach and interactions to test the interdependence between two locus models using ANOVA. The drop one analysis was particularly valuable for studying the support for the individual terms in each model. For the USI trait 10 significant interactions were detected each involved at least one previously detected main-effect QTL. Seven additional markers on Chr9, 10, 14 and Chr1, 13, 15 and 18 were detected by virtue of their interaction with either Mrip2 or Mrip3, respectively and were previously undetected as single locus QTLS. As expected, homozygosity of the BALB/c allele at Mrip2 and Mrip3 increased the average polyp numbers for the USI

with respect to the corresponding heterozygous state and was detected in the background of either homozygous or heterozygous states of the second inferred QTL markers. In addition, Chr6 (D6Mit194) which was previously detected in the LSI trait and not in the USI, showed evidence as a second QTL (LOD_f of 7.23, LOD_a of 6.01 and LOD_i of 1.32 following a 2D scan) and was also selected as an additive QTL from the automatic model selection approach (Figure 3-9) of the USI trait. This would suggest epistatic pleiotropy where the co-variation observed between the USI and LSI may be caused by this single QTL on Chr 6 (D6Mit194). D6Mit194 was therefore involved in significant interactions between *Mrip5* (Chr 16) of the USI (p =0.032) and Chr 3 (D3Mit320) of the LSI (p = 0.011). No other significant interactions were observed in LSI or LI our data.

In summary, we applied exploratory functions within MQM mapping to select subset of loci models that are best supported by a 2-dimensional scan of our 2Gy data. The selection of the appropriate model criterion is important for analysing MQM. Furthermore, the outcome of MQM varies depending on statistical approaches and key difficulties in establishing evidence for linked QTLs or interacting QTLs in the context of multiple QTL models is controlling for large effect QTLs and therefore the selection of suitable markers as covariates is a crucial process. We explored MQM while controlling for major QTL loci to identify QTL pairs that are not artefacts resulting from strong associations at QTLs with strong LOD scores and thus creating an interaction effect between other loci. Higher order epistasis modelling of up to a three-way interaction was not tested as population size requirements of >500 are needed to reliably estimate higher order interactions. A further consideration was

that permutation testing of at least 1,000 replicates would involve multidimensional analyses and suffers from computational burden.

Next, we plan to map these interactions to genomic regions and to evaluate genes involved in the control of irradiation-induced intestinal polyposis. We hope to show that by using *in silico* methodology combined with literature search to begin the process of refining QTLs and the identification of candidate genes of interest.

CHAPTER 4 IN SILICO SEQUENCE MODELLING OF FUNCTIONAL SNPS AS A RATIONAL TO SELECT CANDIDATE GENES WITHIN *MRIP 1-5*.

4.1 INTRODUCTION

4.1.1 In silico modelling

Using unannotated sequences from databases and computational biology tools it is possible to gain further insights into the functions of proteins based on their sequence, structure, evolutionary history and the interactions between proteins. I aim to provide a framework to screen the *Mrip* regions for functional coding region SNP changes and predict using bioinformatics tools the deleterious effects. I also aim to design an initial draft of target genes which will be refined to detect more specific pathways. This will begin the exploratory search of candidate genes as opposed to cherry-picking ones that I consider are important within the identified QTL regions and that may play a role in irradiation induced intestinal polyposis. This will contribute towards our understanding of CRC and irradiation biology, but more specifically to understanding polyposis in the mouse model than can then be extrapolated to humans through comparative mapping.

The aim here, is, to investigate the biological context behind *Mrip* and QTL pairs identified from epistasis using the application of an '*in silico*' sequence search (Grupe *et al*., 2001, Darvasi *et al*., 2001) and literature search based on the epistatic QTL pairs detected in chapter 3. The target genes then serve as the basis to further

explore susceptibility loci and their involvement towards irradiation and CRC progression based on literature mining.

Sequence based selection of target genes is based on polymorphisms and the use of algorithms to predict deleterious effects on protein function. In this section the aim is to look more specifically at underlying SNPs which are likely to represent modifiers of inherited susceptibility. SNPs are partly responsible for individual differences observed in populations affecting disease severity, effectiveness, side-effects and tolerability to drugs and treatments and have become an important process in medical research (Stoneking *et al*., 2001; Chakravarti 2001; Stenson *et al*., 2009). The availability of SNP information from NCBI dbSNP (http://www.ncbi.nlm.nih/gov/SNP/) (Sherry *et al*., 2001) can be used to predict the detrimental effects of amino acid changes in peptide sequence (Chasman and Adams, 2001; Sunyaev *et al*., 2000; Sunyaev *et al*., 2001).

90% of all human genetic variation are made up of SNPs which are randomly distributed every 100-300 bases along human genome and there are ~11 million in the human genome, yet only a small proportion of these SNPs is known to alter the gene function to influence measurable differences in the phenotype (Sachidanandam *et al*., 2001). SNPs within coding regions, non-synonymous SNPs (nsSNPs) (Ramensky *et al*., 2002) and regulatory SNPs (rSNPs) of genes, are most likely to cause an amino acid change in the peptide sequence or influence the overall activity of the protein or the gene expression. Approximately 10–15% of nsSNPs are projected by predictive software to be damaging and, hence, likely to influence phenotype (Ng and Henikoff, 2002). Such SNPS are believed to predispose

individuals to diseases and impact an individual's response to environmental factors such as bacteria, toxins, chemicals and drug and other therapies including, radiotherapy and chemotherapy (Sunyaev *et al*., 2001). SNPs that do not result in a protein change are called synonymous SNPs (sSNPs), the majority of which are intronic, at 3'-UTRs or between genes. Most sSNPs are deemed neutral on the basis of their location and are unlikely to modify proteins and, as a consequence, do not contribute to a phenotype (Zhao *et al*., 2003).

By using the mouse strain data to select nsSNPs that are polymorphic between parental BALB/c and B6 strains within *Mrip,* the functional effect changes of these SNPs using prediction algorithms will be investigated using an *in silico* approach. The high density of single genetic variants provides a convincing rationale to predict the deleterious effects in genes in this manner, furthermore, the large volume of SNP data is a demonstrable platform for comparing human and mouse genomes and is valuable for identifying conserved genes and regulatory elements between the two species.

Several methods use sequence conservation of particular amino acids within a family of sequences or look for particular features of a protein structure to predict whether a substitution affects protein function. Prediction effects of nsSNPs are based on the biochemistry involving the amino acid substitution and its context within the protein sequence, taking into account factors such as hydrophobicity, helical-breaking residue, 3D structure and general rules of protein chemistry (Grantham, 1974) which can complement *in silico* modelling.

I am interested in making use of these features and using empirical prediction rules based on the protein sequence, structural information and homologies in amino acid sequences between species using Polymorphism Phenotyping v1 (PolyPhen) (Ramensky *et al*., 2002) and Sorts Intolerant From Tolerant (SIFT) prediction tools (SIFT) (Kumar *et al*., 2009). Xi *et al*. found that that PolyPhen and SIFT predicted correctly the effect of 96% of amino acid substitutions on protein activity in *APEX1* (Xi *et al*., 2004). In validation studies, PolyPhen and SIFT have been shown to predict successfully the effect of over 80% of amino acid substitutions, benchmarking studies based on the analysis of "known" deleterious substitutions (Ng and Henikoff, 2002; Sunyaev *et al*., 2000). Therefore, *in silico* screening, prediction algorithms and the use of comparative mouse-human genomics forms the basis for further refining the selection of disease loci involved in intestinal polyp multiplicity.

4.2 MATERIAL AND METHODS

4.2.1 In silico

A search of the Mouse Phenome Database MPD for loci that are potentially regulatory such as nsSNPs and splice site ssSNPS was conducted within the candidate regions. The MPD holds ~1400 phenotypic 'phenomics' measurements for 600 strains of mice covering approximately 740 million SNP alleles and is a useful repository for complex trait analysis (Grubb *et al*., 2009). Polymorphic nsSNPs were selected within the same genomic window as used previously for PA and based predominantly on QTL mapping (±15cM) and epistasis analysis using http://phenome.jax.org/db/q?rtn=strains%2 Fcompare2_menu&strains=50,7 using Sangar1 dataset/ The SNPs and the related protein sequence for the selected genes were obtained from dbSNP (NCBI; http://www.ncbi.nlm.nih.gov/SNP/).

Next, the fasta format and the amino acid position and SNP-associated change were submitted to computer algorithms to predict the functional effect of amino acid changes by considering conservation over evolution, the physiochemical differences, and the proximity of the substitution to predicted functional domains (Figure 4-1). The protein sequence together with sequence position and the two amino acid variants is submitted to Polyphen server (which was available at http://coot.embl.de/PolyPhen/ and currently replaced by PolyPhen version 2.2.2 of the software). The analysis was initially set to the original PolyPhen criteria and then assessed using SIFT at http://sift.bii.a-star.edu.sg/ which uses a modified version of PSI-BLAST and Dirichlet mixture regularisation to construct multiple

sequence alignment of proteins that can be globally aligned to a query sequence which belongs to the same clade (Sjolander *et al*., 1996; Altschul *et al*., 1997). This procedure was adopted because of the 90% coverage and 9% rate of false positives rate using PolyPhen compared to 60% coverage and 20% false positives using SIFT (Ng *et al*., 2006). In addition, SIFT predicts the functional importance of amino acid substitutions based the phylogenetic information with the alignment of orthologous and or paralogous protein sequences. This distinguishes residues that are conserved, and thus intolerant to most sequence changes, from those residues that are poorly conserved and tolerant to sequence changes. Sequences within the database, which show between 90-99% sequence similarities to the query are removed from the alignment process in order to exclude proteins containing the substitution of interest and to avoid matching to pseudogenes.

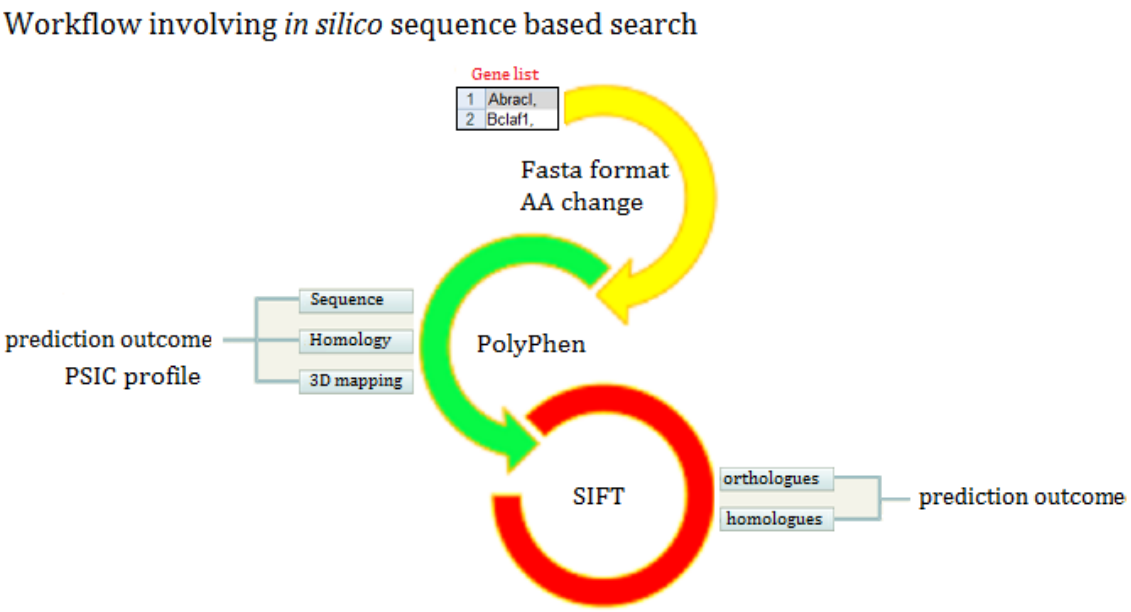

Figure 4-1 A simplified diagram of the *in silico* workflow process. Starting with a gene list derived from MPD of nsSNPs polymorphic between the strains, the fasta format of the genes are submitted to PolyPhen. This combines information on sequence features, multiple alignments of 45 vertebrate genomes using NCBI human

genome, homologous proteins sequences from UniProtKB/UniRef100 and structural parameters based on 3D mapping from PDB/DSSP Snapshot to make a prediction of the amino acid (AA) variant effect on protein function. SIFT predictions are based on orthologues and homologous sequences from UniProt and Swiss-Prot and TrEMBL

Sequence-based characterisation of the substitution site, profile analysis of homologous sequences, and mapping of substitution site to a known protein 3D structure are the parameters taken into account by PolyPhen to calculate a PSIC score for each of the two variants and then computes the PSIC score difference between them. The higher the PSIC score difference the higher the functional impact a particular amino acid substitution is likely to have. PolyPhen scores were designated as probably damaging with a z-score of 2.00, possibly damaging (z = 1.50-1.9), potentially damaging (z = 1.25- 1.49), borderline (z = 1.00-1.24), or benign (z = 0.00-0.99; Xi *et al*., 2004). The SNPs predicted as 'probably or possibly damaging' were selected for SIFT analysis. Similarly, protein sequences were obtained, data was entered and predictions were based on the comparisons of sequences of related proteins which are sourced from UniProt (Universal Protein Resource) consisting of 2,392,904 proteins with 91,269 human proteins from Swiss-Prot and TrEMBL compared to POLYPHEN which uses 70,627 human proteins from Uniprot July 7, 2005 Release 5.0. SIFT scores were classified as intolerant (0.00-0.05), potentially intolerant (0.051-0.10), borderline (0.101-0.20), or tolerant (0.201-1.00) according to the classification proposed by Ng *et al* (2003) and Xi *et al* (2004).

The criteria for selecting target genes was based on the cM position within the QTL interval (±15cM). Target genes must have contained nsSNPs in the regulatory and or functional domains that are predicted to be 'damaging' and be confirmed as variants in our strains of mice and/or linked by citation to other genes for possible genetic epistasis. The suitability of target genes will be based on literature evidence and data mining and the findings presented here.

4.3 RESULTS

4.3.1 In silico results

This *in silico* results section details the large-scale homology and sequence based predictions of 419 fully annotated genes that were selected within the 5 *Mrips.* Also presented are the subsequent analyses of 1,248 nsSNPs to ascertain functional candidate genes involved in irradiation induced intestinal polyposis (detailed below and in Figure 4-2). The goal of the analysis is to predict mutations (or a subset of mutations) which are likely to affect the structure or function of the gene, and thus to identify which of these mutations may have a role in the progression or development of irradiation-induced CRC and related disease. A total of 652 genes (1,870 nsSNPs) (Table 4-1) were highlighted initially on the basis of polymorphism between the two strains within the 5 *Mrip* regions from MPD. The first filter applied to these genes excluded any poorly annotated and hypothetical genes, such as locations (LOCs), variants (GM) or Riken (Rik) (the three major types) and various transcripts or predicted proteins (BOC, ENS, Vnmr). This resulted in omission of 80 genes (12.2% of the total) accounting for 267 nsSNPs (Figure 4-2 and Table 4-1). Similarly, 134 (293 nsSNPs) and 19 genes (63 nsSNPs) were selected within the *Mrip1* (Chr 2) and *Mrip5* (Chr 16) respectively which belonged to the olfactory receptor (*Olfr*) and accounted for 20.5% of the total genes. For Chr 2, these genes are closely located between regions 85.4 – 112.6cM (genes *Olfr996-Olfr1314)* with 280 nsSNPs found clustered within 85.4Mbp and 89.4Mbps of these regions. Chr 16 *Olfr172 – 209* spanned from 58.7Mbps – 59.3 Mbp region of *Mrip5* (Figure 4-2). These are the largest olfactory receptor gene cluster, with over 400 genes expressed

in BALB/c and B6 genome alone. Olfactory genes were not considered to have any important function in adenoma multiplicity and tended to inflate the actual number of genes. As a result, these genes were discarded from predicting functional polymorphisms. In addition, pseudogenes that have lost their protein coding function and are regarded as defective were excluded from further analysis.

Next, 419 genes (fully annotated genes) carrying 1248 nsSNPs were subjected to a second filtering process: a sequence-based prediction method using physical and comparative analysis tools involving PolyPhen v1.12 (based on March 2006 releases of UniProt, NCBI, PDB) and SIFT programs. These softwares were used to understand which sequence changes are likely to either disrupt the structure or function of the protein. This resulted in 156 genes (37%) predicted deleterious and or damaging nsSNPs and 263 genes (63%) predicted to have benign effects and were excluded (Figure 4-2 and Table 4-1). The damaging and or deleterious genes were subjected to a third filter; exclusions were based on genes not involved in tumourigenesis from literature analysis and therefore regarded as not likely candidates. This process eliminated a further 129 genes. The resulting 27 likely candidates were subjected to numerous searches including literature reviews based on genes involved in CRC, IR and polyps, followed by a search for functional domains, conservations of the nsSNPs across species and pathway assocations were all taken into account before final gene targets were selected. This resulted in a final reduction to 12 QTL targets and 21 subsequent nsSNPs as outlined in Table 4-2 as potentially deleterious by PolyPhen and SIFT, and the functional domains associated with some of these polymorphisms were identified. From these 12 genes a subset of 7 genes were identified within *Mrip1*, 2 and 1 genes within *Mrip2* and *3*

and 1 and 1 for *Mrip4* and *5* respectively (including *Prkdc*), were selected for further investigation), details of in silico predictions are found in Table 4-2 and literature supporting target gene selection in Table 4-3

Table 4-1 Selection of genes based on BALB/c specific nsSNPs (in parenthesis) obtained from the MPD. The data presented for each *Mrip* region as defined. A total of 419 genes and 1248 nsSNPS (column 4) were filtered from MPD (excluding hypothetical genes (column 3). Resulting nsSNPs were investigated for predicted affects by PolyPhen and SIFT and deleterious and or damaging nsSNPs (column 5) were subjected to subsequent analysis outlined in column 6. The final number of genes and corresponding nsSNPs were selected as potential candidates relevant to trait of interest.

Mrip Chr Region	**MPD genes selected : polymorphic nsSNPs**	**Poorly annotated genes** (excluded (Filter 1)	**Fully annotated genes** (included for prediction analysis)	**Predicted damaging and or deleterious;** (Filter 2)	Filters 3 & 4	**Final target genes selected**
Mrip1- Chr2 (69.2Mbp - 131.6Mbp)	330 (1003)	159 (386)	171 (617)	52	**Functional domains, (literature based) involvement in tumourigenesis, CRC, IR and polyps -- Structure and function of SNPs -- Amino acid conservation (EMBL-EBI)**	7 (14)
Mrip2- Chr5 (21.3Mbp - 85.1Mbp)	69 (186)	16 (61)	53 (125)	20		2 (2)
Mrip3- Chr5 (90.4Mbp - 138.3Mbp)	167 (435)	28 (98)	139 (338)	60		1 (1)
Mrip4- Chr16 (3.8Mbp - 37.3Mbp)	29 (108)	2 (5)	27 (103)	9		1 (2)
Mrip5- Chr16 (37.3Mbp - 84.7Mbp)	57 (138)	28 (73)	29 (65)	15		1 (2)
Total	**652**	**233**	**419**	**156**		**12**

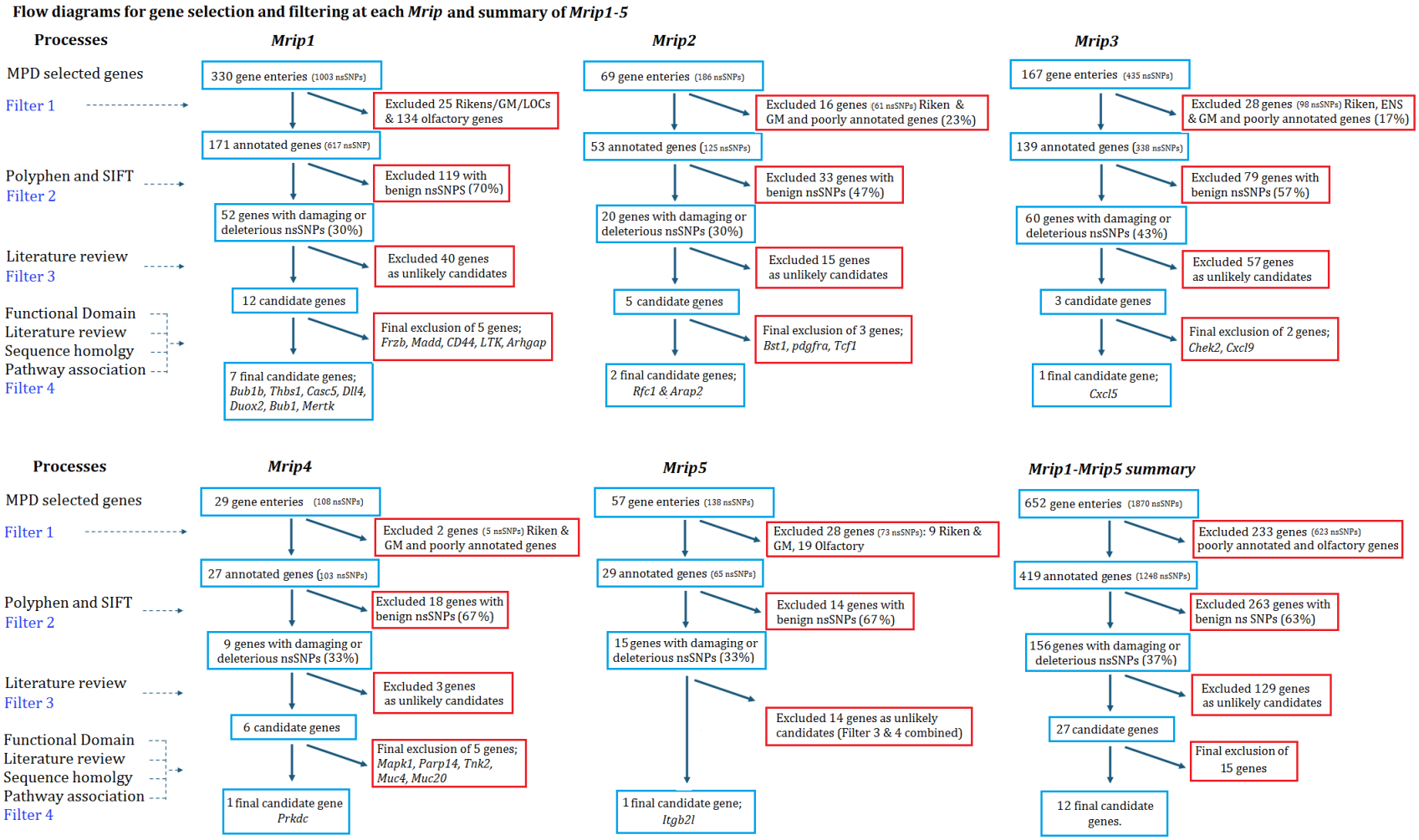

Figure 4-2 Flow diagrams of the processes involved in the selection of target genes for *Mrip1-Mrip5.* Eliminated genes are shown in red boxes and candidate genes in blue.

Table 4-2 Predictive assessment of *BALB* specific amino acid polymorphisms for potentially damaging alterations**.** Polymorphisms within known functional domains were identified. Polyphen and SIFT analysis was carried out to identify potentially damaging alterations. Polyphen predictions by position specific independent count (PSIC). PSIC ≥1.5 is potentially damaging and are in bold (except in the case of * where PolyPhen v2,2 was applied which uses a scale of 1-1.0) SIFT predicts the change to be tolerated (T) or not tolerated (N), if intolerant then the substitution effect is stated. The conservation of sequence alignment at and around the amino acid substitution is compared between different species using Polyphen. The polymorphisms were located within some functional domains and determined using Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) (Jensen *et al.*, 2009) available at; http://string-db.org/

	Gene	Variation polymorphism	Functional domain	PolyPhen (PSIC)	SIFT	Conservation (%)
Mrip1	*Duox2*	H[His] ⇒R[Arg]627 rs27453362	Animal heam peroxidise (aa 60-562)	**1.524**	T	36% (9/25)
		G[Gly] ⇒S[Ser]553 rs27453360		0.114	T	44% (11/25)
		S[Ser] ⇒N[Asn] 378 rs27453345		0.299	T	36% (9/25)
	Bub1b	S[Ser] ⇒G[Gly] 240 rs3144766	CDC20 binding to BUBR1 (Mad2 dependent) Casc5 by sequence similarity	1.32	T	75% (6/8)
		P[Pro] ⇒T[Thr] 278 rs3149979		0.549	T	44% (4/9)
		V[Val] ⇒A[Ala]312 rs3149980		0.072	T	63% (5/8)
		D[Asp] ⇒E[Glu]664 rs27423851		1.241	**N (Acidic to acidic)**	67% (5/8
		N[Asn] ⇒D[Asp]705 rs27423850		0.345	**N- Acidic to uncharged polar)**	67% (5/8)
	Dll4	Q[Gln] ⇒L[Leu]419 rs27423750	EGF like 6	0.431	T	25% (15/59)
		G[Gly] ⇒S[Ser]627 rs27423747		**1.516**	**N- Uncharged polar to non-polar**	38% (6/16)
	Casc5	I[Ile] ⇒T[Thr]1286		**1.643**	T	60% (3/5)

		rs27471474				
	Bub1	P[Pro] ⇒S[Ser]153 rs28269039		**1.569**	T	57% (4/7)
	Mertk	A[Ala] ⇒T[Thr]441 rs27446346	Extracellular domain	0.833*	**N**	75% (6/8)
	Thbs1	Q[Gln] ⇒R[Arg]960 rs27502423	TSP C-terminal	0.99*	T	74% (8/11)
Mrip2	*Arap2*	P[Pro] ⇒R[Arg]278 rs31438432		**2.271**	**N (Basic to non-polar)**	100% (7/7)
	Rcf1	Y[Tyr] ⇒N[Asn]66 rs31457275		**2.440**	T	79% (8/11)
Mrip3	*Cxcl5*	R[Arg] ⇒H[His]10 rs31723268		**1.582**	N/A	100% (2/2)
Mrip4	*Prkdc*	R[Arg] ⇒C[Cys]2140 rs4164952	Pfam domain Nuc194 entry	**2.495**	**N (basic to uncharged polar)**	100% (9/9)
		M[Met] ⇒V[Val]3844 rs48580935	Pi3K/Pi4K domain	0.208	T	50% (4/8)
Mrip5	*Itgb2l*	N[Asn] ⇒D[Asp]509 rs316615		**2.193**	T	100% (45/45)
		G[Gly] ⇒S[Ser]539 rs3164214		**1.662**	T	56% (25/45)

Table 4-3 The function of candidate genes and their role in mediated radiation response and tumourigenesis selected using in silico.

Symbol	Gene	Functional/expression
Duox2	dual oxidase 2	Generate ROS species; Innate immune response; expressed in mucosal surfaces including gut (Bedard and Krause, 2007); hypermethylated in lung cancer (Luxen *et al*., 2008).
Bub1b	budding uninhibited by benzimidazoles 1 homolog, beta	Mitotic checkpoint gene involved in chromosomal instability· Upregulation in CRC (Abal *et al*., 2007; Hanks *et al*., 2004)).
DLL4	delta-like 1	Notch signalling, regulates tumor angiogenesis and tumour growth, expressed in the gut (Katoh and Katoh,

		2006).
Casc5	cancer susceptibility candidate 5, AF15q14	A member of the Spc105/Spc7/KNL-1 family, directly links spindle checkpoint proteins BubR1 and *Bub1* to kinetochores and is required for chromosome alignment (Kiyomitsu *et al*., 2007)
Bub1	budding uninhibited by benzimidazoles 1 homolog	Regulates exit from cell cycle checkpoint, genetic and epigenetic alterations in CRC (Grady, 2004)
Arap2	ArfGAP with rhoGAP domain ankyrin repeat and PH domain	Human homologue involved in focal adhesion and mediates effects of RhoA (Yoon *et al*., 2006)
Rcf11	replication factor C (activator 1) 1	A multi-subunit protein complex needed for proliferating cell nuclear antigen (PCNA)-dependent DNA replication and repair synthesis (Hashiguchi *et al*., 2007).
Cxcl5	chemokine (C-X-C motif) ligand 5	Chemokines as regulators of angiogenesis (Baier *et al*., 2005). Expressed in the colon and variants associated with CRC (Dimberg *et al*., 2007).
Prkdc	protein kinase, DNA activated, catalytic polpeptide	DNA double strand break repair mechanism (Wetering and Chen, 2007).
Itgb2l	integrin beta 2-like	During an inflammatory response, Itg2l help retain CXCL13-expressing cells (Hojgaard *et al*., 2006)

4.3.2 Profile of all SNPs within Mrip regions

At the time of writing the profile of all SNPs within the *Mrip* was extracted from MPD using Sanger dataset (ftp://ftpmouse.sanger.ac.uk/ current snps) as of November 2011 and the consequence was according to NCBI dbSNP mouse build 132 (ftp://ftp.ncbi.nih.gov/snp/ organisms/mouse 10090 as of December 2010). 44,417 SNPs were located within 1,560 genes (including predicted genes and locations) of the *Mrip* regions and briefly summarised in Table 4-4.

Table 4-4 SNP variant profile in 1560 genes within the *Mrip* regions, extracted from MGI build 37. ns= non synonymous, s= synonymous, i=intronic, l=location of unknown consequence, stopL=stop codon loss,U3 and U5=3 and 5 prime untranslated region, ess=essential splice site, ss=splice site, %=percentage of SNPs expressed as a total of 44,417 SNPS (-=percentage value is low)

SNP	Coding genes	Olfactory genes	Hypothetical genes	Total genes	%
i	33926	77	6,854	40857	1.1
U3	818	7	78	903	2.1
s	**732**	144	68	944	92
ns	**331**	101	67	499	0.7
U5	102	14	25	141	-
l	55	37	240	332	2
ss	3		1	4	0.3
stopL	1	-	-	1	-
ess	1		2	3	-
multiple				733	1.6

It was discovered during data cleansing of this list that 16.6% of all SNPs comprised of hypothetical genes, with locations (LOCs), variants (GM) or Riken (Rik) genes as the major three types. Others included expressed sequences (AI*), cDNA (BC*) sequences, DNA segments (D*Ertd*), and various transcripts or predicted proteins (EG, ENS, OTT, RP-23) and this inflated the number of polymorphisms. Rikens, GM and LOC constituted the major proportion of predicted protein genes and or locations, of which 6,854 were intronic and a more equal ratio of sSNPs and nsSNPs was observed (68 and 67 respectively). A total of 147 olfactory genes, with 380 SNPs; 101 non-synonymous (nsSNPs), 144 synonymous SNPs (sSNPs) and 77 intronic SNPs (iSNPs) were observed. Data validation showed that 733 SNPs were assigned more than one consequence which is most likely due to the multiple transcripts used which determined the different consequence.

The total occurrence of mouse SNPs was the highest within introns (40,857 SNPs 91.9%), then 2.1% were sSNPs (944), 1.1% nsSNPs, (499) and 0.7% were found within a locus with an unidentified consequence (332). Some 2% and 0.3% SNPs were detected in the 3' prime UTR (903) and 5' prime UTR (141), respectively. UTR-SNPs, albeit a small fraction, may have functional consequences on mRNA stability and or expression. However, it is less straightforward to predict the functional importance of SNPs within regulatory regions. As expected low numbers of rarer and infrequent SNPs such as essSNPS (3) and ssSNPs (1) were detected mainly in Riken genes. In addition, 1 stop codon gain (olfactory gene) and 1 stop codon loss in leucine carboxyl methyltransferase 2 (*Lcmt2*) were also observed.

The higher rate of iSNPs may be explained on the assumption that these SNPs are not under selective pressure and or due to the length of intronic regions that are in the 100s of kilobases and are generally poorly conserved. Furthermore, the high numbers of polymorphisms suggests that these could be in LD with other variants and leaves the open question of whether a proportion of these are actually gene-associated SNPs. Increasing evidence shows that introns exert a control of genes and are involved in high-level gene expression (Choi *et al*., 1991; Palmiter *et al*., 1991; Millar *et al*., 2010) and so are no longer regarded as 'junk' DNA and, instead, are involved in complex processes during mRNA transcription. In comparing actual genes (excluding all ambiguously annotated or hypothetical genes), a ratio of 2.2 sSNP (732) to every nsSNP (331), was noted and, overall, SNPs were the lowest in exons and highest in introns (33,926 iSNPs) (Table 4-4).

4.3.3 Target genes selected through literature search

Literature analysis was used to investigate the epistatic potential of both *Mrip2* and *Mrip5,* which have the strongest association between adenoma numbers in the USI. Additional gene targets: Ras-related C3 botulinum toxin (*Rac1*) and T lymphoma invasion and metastasis 1 (*Tiam1*) were selected based on the Chr5:Chr16 interaction between *Mrip2:Mrip5. Tiam1* is a metastasis-related gene of T lymphoma that is also involved in the metastasis of CRC. In the presence of the Wnt3a stimulation *Rac1* is proposed to interact with *Tiam1*, which is known to be involved in metastatic potential in CRC (Liu *et al*., 2005). The model suggested that *Tiam1* activates the GTP-bound state of *Rac1* via a GDP-GTP exchange, which mediates Wnt-induced transcription via β-catenin/TCF4 complexes (Esufali *et al*., 2004; Buongiorno *et al*., 2008). *Tiam1* is also thought to activate selectively the *Rac1* GTPase gene, which is expressed in the base of intestinal crypts and involved in the Wnt pathway. *Tiam1* deficiency is also reported to significantly reduce the formation and growth of polyps in *Min* mice (Carter *et al*., 2004, Malliri *et al*., 2006). Further to this, action of *Asef* and *Asef2,* which are guanine-nucleotide exchange factors specific to *Rac1,* are important for adenoma formation in $Apc^{Min/+}$ mice (Kawasaki *et al*., 2009).

4.4 DISCUSSION

The potential biological relevance of the QTL results was investigated using the joint effects of two prediction tools, PolyPhen and SIFT. The *in silico* approach provided an overview of the extent of functional coding sequence variation between mouse inbred strains and has helped to speed the identification of candidate genetic variations that underlie between the two strains and within irradiation induced CRC QTLs. The selection of target genes was narrowed based on a literature review and an known associations with irradiation, tumourigenesis and epistasis from the previous chapter (Chapter 3).

The model built in the work presented here are appropriate for candidate gene selection based on the examination of nsSNPs, which could potentially influence proteins. The SNP search was constrained to coding regions within *Mrips* and only polymorphisms, which were strain-specific were selected. The relevant information was taken from the different integrative biomedical databases and literature to improve data quality and extraction to define subsets of genes that are physiologically important. *In silico* tools further demonstrates a useful approach in providing information about the nature of mutations and functions as a first pass filter to determine the deleterious amino acid substitutions worth pursuing for further validation between the gene variation and trait association. Therefore, the study presented by this chapter demonstrated that by utilising the huge wealth of data already available it is possible to cut the use of animal breeding, selecting targets according to literature based evidence. In the next chapter, the aim is to

confirm and extend these results using sequencing analysis and mRNA gene expression of the selected targets.

CHAPTER 5 SINGLE NUCLETIDE POLYMORPHISMS MUTATION ANALYSIS AND MRNA EXPRESSION OF TARGET GENES.

5.1 INTRODUCTION

DNA sequence polymorphisms affecting either expression and or gene function form the underlying molecular basis for QTLs. Therefore, identifying the sequence variation between strains is an important process to detect the causal gene (Abiola, 2003). Here, QTLs have been associated with adenoma multiplicity using traditional IM analysis and putative target genes within the QTLs have been identified using an *in silico* based sequenced based scan.

To summarise, target gene selection was based on 2 criteria: first, SNPs were selected from sequenced data which may confer partial functional activity (*in silico*) and this was followed with computer-based algorithms to predict functional effects and distinguish disease-associated missense mutations from common variants; second, a selection genes was prioritised on literature search to minimise bias in selection. The outcome revealed 14 target genes (including *Rac1* and *Tiam1*), which may confer potentially susceptible alleles involved in adenoma multiplicity. This chapter aims to characterise the target genes by verifying polymorphisms using sequencing and to distinguish mRNA expression differences in the USI and LSI between BALB/c and B6$^{Min/+}$ in the intestinal mucosal expression of genes that may influence susceptibility to radiation-induced polyposis in BALB/c and B6$^{Min/+}$ mice.

5.2. MATERIALS AND METHODS

5.2.1 Inter-strain sequencing

Genomic DNA was prepared from healthy intestines of BALB/c, B6$^{Min/+}$ and N2 *Apc*$^{Min/+}$ mice using snap-frozen tissue (~25mg) as mentioned in section 2.2.2. PCR conditions were 95ºC for 15 minutes, followed by 24 cycles (95ºC for 1 minute, 55ºC for 1 minute 72ºC for 1 minute) and elongation at 72ºC for 10 minutes.

Primer sets were designed using Primer3 v. 0.4.0 (available from http://frodo.wi.mit.edu/). The PCR reaction products were cleaned to remove residual contaminants such as *Taq* enzyme, nucleotides (dNTPs) and primers using ExoSAP-IT® (USB Corporation, Cleveland, Ohio). ExoSAP-IT treats PCR products ranging in size from less than 100 bp to over 20 kb while conserving the full recovery of the PCR product. A 5 µl sample of the PCR reaction product was mixed with 2 µl of ExoSAP-IT® for a combined 7 µl reaction volume. The sample was incubated at 37ºC for 15 minutes to degrade remaining primers and nucleotides and then incubated at 80ºC for 15 minutes to inactivate ExoSAP-IT®, resulting in a purified PCR product.

The purified product was prepared for the sequence reaction before running on the ABI 3100 genetic sequencer (Applied Biosystems, Warrington, UK). A stock solution of Big Dye Terminator v3.1 (containing buffer, polymerase, and labeled dNTPs) containing 800 µl was diluted with 800 µl 5x Big Dye Terminator v3.1 sequencing buffer (ABI) and 800 µl PCR water. A typical sequencing reaction included 8 µl

diluted Big Dye Terminator v3.1, 1 µl of a 20 µM working solution primer (either the forward or the reverse primer), 1 µl PCR water and 10 µl purified PCR product, giving a total volume of 20 µl. PCRs were performed on an MJ Research Tetrad Thermal Cycler. PCR cycling conditions were 5 minute denaturation at 95°C, followed with denaturation at 96°C for 30 seconds, annealing at 50°C for 45 seconds and extension at 60°C for 4 minutes repeated 34 times, after which samples are kept at 4°C. The annealing temperature varied according to the primer data in the range of 61-64°C, but typically the primers annealed at 60°C.

Unincorporated dye terminators are removed from sequencing reactions before loading onto the ABI 3100 genetic analyzer using a DyeEx 96 well plate kit (QIAGEN). The plate was placed on top of the sample collection plate provided and centrifuged for 3 minutes at 2,000 rpm. After discarding the flow-through, the plate was placed on a 96 well elution plate with a suitable adaptor. 10 µl of the purified sequencing reaction was carefully applied to the gel bed of each well. This was centrifuged for 3 minutes at 2,000 rpm, the dye bound dNTPs bind to the column which consisted of pre-hydrated gel-filtration resin and the eluent contained the purified sequencing reactions.

The spatial calibration for the ABI-Prism 3100 Genetic Analyser was performed using Hi-Di Formamide. Samples run through the machine must be suspended in Hi-Di Formamide in order to produce an accurate sequence reading. Hence, following DyeEx clean up, 1 µl of sequencing product was added to 9 µl of Hi-Di Formamide and then denatured by heating at 95°C for 5 minutes in a MJ Research Tetrad Thermal Cycler. This was immediately placed on wet ice for 5 minutes before

loading into the ABI-Prism 3100 Genetic Analyser. Data analysis was performed using Sequencer version 4.7 for Windows (Gene Codes Corporation, Michigan, USA).

Any identified polymorphisms were checked against the SNP database, (NCBI, Build 36.1) to distinguish between known and novel SNPs. As noted above, SNPs are categorised as synonymous and non-synonymous based on alterations at the amino acid level and nsSNPs investigated for potential effects on protein function.

5.2.2 Reverse transcription

Intestinal musosal scrapings from the USI and LSI were taken from wild type (WT) 8-9 weeks old BALB/c and B6$^{Min/+}$ mice (n=3), and each sample was processed separately and in duplicate. Total RNA was extracted using micro RNAeasy kits (Sigma) and contaminating DNA was removed using RNAse-free DNAse1. RNA samples were checked on 1% agarose gels and concentration determined using NanoDrop®, ND-1000 UV-Vis Spectrophotometer. Total RNA was used to synthesise cDNA using the First-Strand cDNA Synthesis Kit, purchased from Amersham Biosciences. To 5 µg of RNA sample was added 1 µl 50mM Oligo(dT)$_{18}$ primers and denatured in 70°C for 5 min and cooled on ice rapidly. The RNA/primer reaction volume was made up to 20 µl with RNase-free water and a cDNA synthesis mix consisting of 4 µl of RT reaction Buffer (5X), 1 µl [10mM each] dNTPs, 0.25 µl Ribonuclease inhibitor [40U/µl] and 2 µl DTT, 0.5 µl M-MLV reverse Transcriptase [200U/µl]. The solution is gently mixed and incubated for 1hr at 42°C before the reaction was terminated at 70°C for 15 minutes and cooled on ice. cDNA was

cleaned up with the QIAquick PCR purification kit and the purity of cDNA was checked on 1% agarose gel electrophoresis and quantified using nanodrop.

Quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis was performed using the ABI 7900 Sequence Detection System instrument (PE Applied Biosystems). TaqMan polymerase chain reaction primers were purchased from the Applied Biosystems to amplify 13 candidate genes selected from our QTL regions. qRT-PCR was performed and the difference between the cycle thresholds (Ct) of these genes from the reference house keeping gene, glyceraldehydes triphosphate dehydrogenase (*Gapdh*) was calculated for each sample. The results were analysed using the Ct (cycle threshold) value given the cycle number (out of 40) at which the level of fluorescence crossed the threshold value. The threshold value was 'set to above the baseline and sufficiently low to be within the exponential growth region of the amplification curve.

To calculate the relative fold change value, the Ct values were normalised to the endogenous control *Gapdh* using the following: ΔCt = Ct target gene – Ct *Gapdh* and squared to give $2^{-\Delta\Delta Ct}$. Standard error value of the means (SEM) were calculated using the following equation SE = SD/√n where SD is the standard deviation and n the sample size (Livak and Schmittgen, 2001). The $2^{-\Delta\Delta CT}$ was used to statistically measure the expression profiles of the target genes and the fold difference in the intestinal sections between BALB/c and $B6^{Min/+}$.

5.3 RESULTS

5.3.1 Sequencing of BALB/c, B6 and N2 offspring identifies single nucleotide polymorphism

No mutations were detected from the complete gene sequencing done on: *Rac1* and *Tiam1* (selected on the basis of Chr5:Chr16 epistasis), including the coding regions and splice sites, upstream and downstream of the 5' and 3' UTRs for polymorphisms in BALB/c, B6$^{Min/+}$ and N2 *Apc*$^{Min/+}$ offspring. The targets selected from *in silico* showed a more constructive approach towards detecting polymorphisms and facilitating the possibility to filter out SNPs and omitting the need to sequence entire genes. This conserved analysis has shown, that from 44 nsSNPs 23 (52.3%) polymorphisms were confirmed in our mouse model as presented in Table 5-1 and representative sequencing chromatograms for *Dll4* and *Duox2* are shown in Figure 5-1 and 5-2)

Table 5-1 Sequencing of selected nsSNPs identified 23 polymorphisms present between BALB/c and B6$^{Min/+}$within *in silico* targets. Data extracted from MPD or NCBI Build 37.1(*).

Gene ID	Chr:bps	refSNP	Residue change and position	BALB/c	B6$^{Min/+}$
Bub1b	2:118438465	rs3144766	G[Gly] ⇒S[Ser] 240	A	G
	2:118440768	rs3149979	P[Pro] ⇒T[Thr] 278	C	A
	2:118440871	rs3149980	A[Ala] ⇒V[Val] 312	T	C
	2:118456696	rs27423851	D[Asp] ⇒E[Glu] 664	C	G
	2:118457497	rs27423850	D[Asp] ⇒N[Asn] 705	A	G
	* 2:118639687	rs13474781	L[Leu] ⇒F[Phe] 959	A	T
Thbs1	2:117948718	rs27502453	R[Arg] ⇒Q[Gln] 960	G	A

	*	2:118117486	rs8251599	^S[Ser] ⇒A[Ala] 376 or /- ⇒R[Arg]		
	*	2:118113119	rs27502496	D[Asp] ⇒N[Asn] 73	G	A
	*	2:118113233	rs27502494	D[Asp] ⇒Y[Tyr] 111		
	*	2:118117487	rs8251600	S[Ser] ⇒I[Ile] 376 or /- ⇒T[Thr]		
	*	2:118117488	rs8251601	S[Ser] ⇒R[Arg] 376		
	*	2:118117592	rs8251605	N[Asn] ⇒T[Thr] 411 or /- ⇒I[Ile]		
	*	2:118117612	rs8251606	S[Ser] ⇒P[Pro] 418		
Casc5		2:118894273	rs27471494	G[Gly] ⇒S[Ser] 240		
		2:118894395	rs50321001	K[Lys] ⇒N[Asn] 280		
		2:118894607	rs27471489	F[Phe] ⇒Y[Tyr] 351		
		2:118894963	rs27471486	F[Phe] ⇒I[Ile] 470		
		2:118895680	rs3684717	I[Ile] ⇒V[Val] 709		
		2:118897412	rs27471474	I[Ile] ⇒T[Thr] 1286	C	T
		2:118897436	rs29812960	A[Ala] ⇒V[Val] 1294		
		2:118897454	rs29669827	E[Glu] ⇒G[Gly] 1300		
		2:118897832	rs27471471	Y[Tyr] ⇒C[Cys] 1426		
		2:118907258	rs52269307	T[Thr] ⇒P[Pro] 1707		
Dll4		2:119157891	rs27423750	Q[Gln] ⇒L[Leu] 419	T	A
		2:119158514	rs27423747	G[Gly] ⇒S[Ser] 627	A	G
Duox2		2:122117242	rs27453362	H[His] ⇒R[Arg] 627	C	T
		2:122118031	rs27453360	S[Ser] ⇒G[Gly] 553	C	T
		2:122119828	rs27453345	S[Ser] ⇒N[Asn] 378	C	T
		2:122122077	rs48311667	Q[Gln] ⇒R[Arg] 231		
		2:122122455	rs27453333	H[His] ⇒R[Arg] 159	C	T
	*	2:122292682	rs32841329	V[Val] ⇒D[Asp] 495		
Bub1		2:127649702	rs28269040	K[Lys] ⇒R[Arg] 187		
	*	2:127823966	rs28269039	S[Ser] ⇒P[Pro] 153	T	C
Mertk		2:128562357	rs27446565	N[Asn] ⇒D[Asp] 176	G	A
		2:128597122	rs27446346	T[Thr] ⇒A[Ala] 441	A	G
		2:128627269	rs27431508	I[Ile] ⇒V[Val] 951	A	G
Arap2		5:63021396	rs48756961	I[Ile] ⇒L[Leu] 1419 or /- ⇒F [Phe]		
	*	5:62748842	rs31438432	R [Arg] ⇒ P [Pro] 278	G	C
Rfc1		5:65704206	rs31457275	N [Asn] ⇒ Y [Tyr] 66	A	T
Cxcl5		5:91188497	rs31723268	R [Arg] ⇒ H [His] 10	A	G
		5:91188515	rs31723269	I [Ile] ⇒ S [Ser] 16		
Itgb2l		16:96647834	rs3164214	G [Gly] ⇒ S [Ser] 539		
		16:96648384	rs3166158	N [Asn] ⇒ D [Asp] 509	G	A

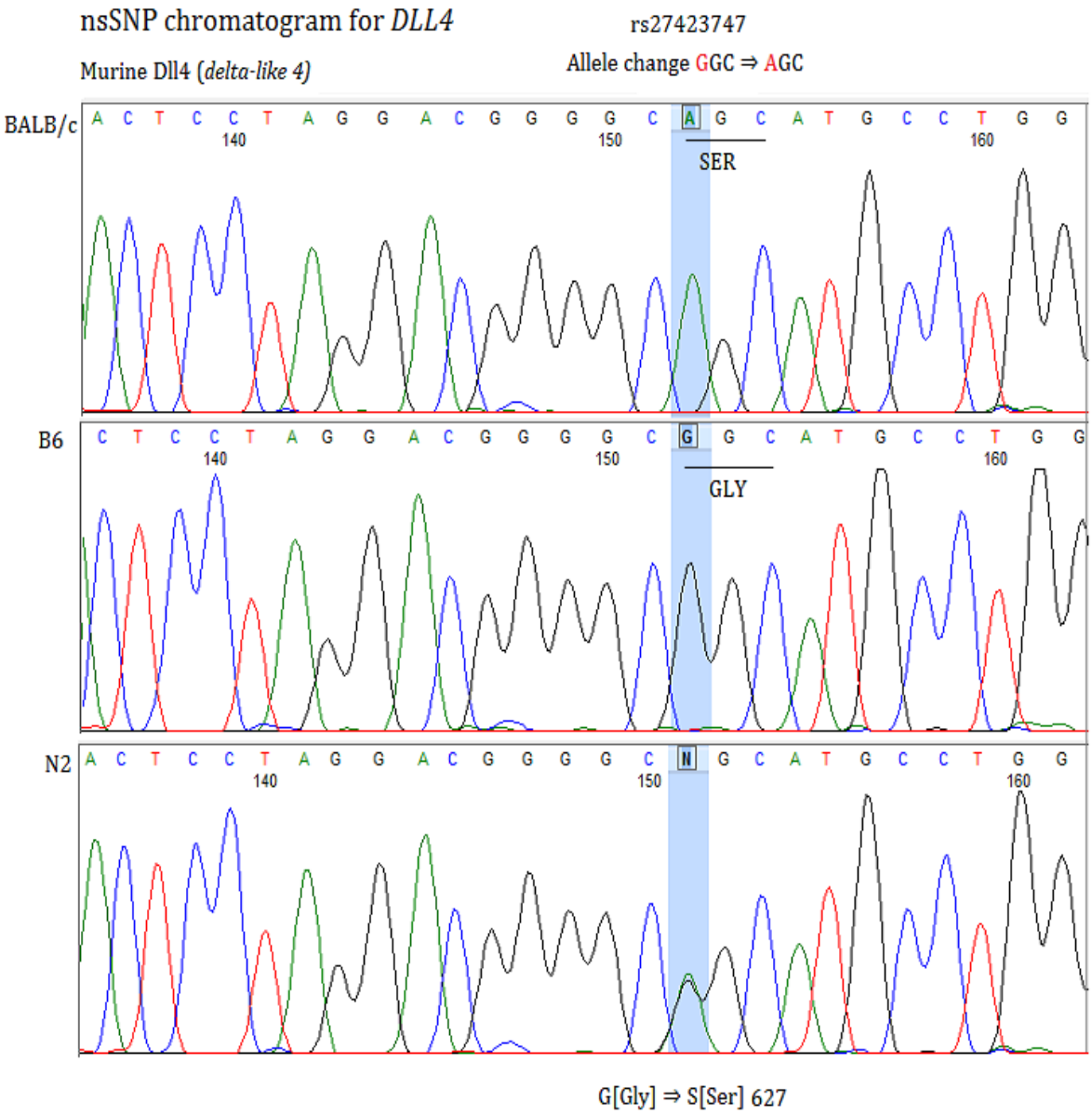

Figure 5-1 Representative DNA sequencing chromatograph for *DLL4* mutation determined by fluorescent DNA sequence analysis. PCR products of genomic DNA were obtained with forward and reverse primers and sequenced to illustrate the G to A transversion resulting in the protein coding sequence change of glycine to serine. Blue highlight indicates the base call marking the nsSNP against the known NCBI dbSNP sequence (Build 36)

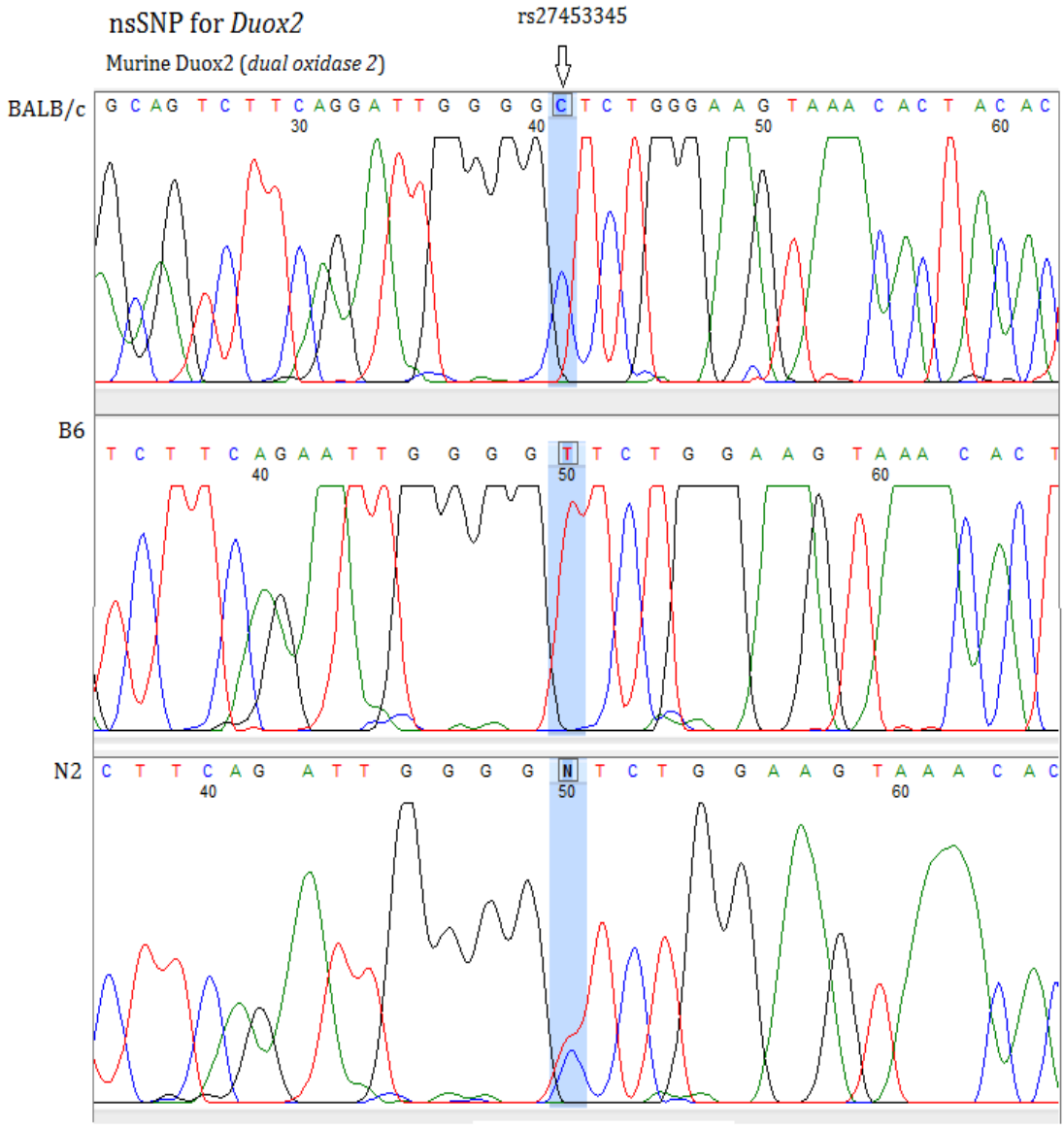

Figure 5-2 Representative DNA sequencing chromatograph for *Duox2* mutation determined by fluorescent DNA sequence analysis. PCR products of genomic DNA were obtained with forward and reverse primers and sequenced to illustrate the C to T transversion resulting in the protein coding sequence change of serine to asparganine. Blue highlight indicates the base call marking the nsSNP compared with known NCBI dbSNP sequence (Build 36)

5.3.2 Analysis of expression by quantitative reverse transcriptase PCR reveals differential gene expression

Based on the study outcome so far; *Duox2* (*Mrip1* on Chr 2) *Rac1* Chr5 (*Mrip2*) and *Tiam1* from Chr16 (*Mrip5*) were selected to test the expression in the USI fragments between BALB/c and B6$^{Min/+}$ (Table 5-2). The expression of these target genes was tested in parental unirradiated strains (due to radiation based experimental constraints at this time animal irradiation facilities were not available) to investigate the difference between the 2 strains in their corresponding segments of the intestine. Our initial target gene expression profiling provides, in part, some understanding of any apparent differences between our strains and segments information on expression levels in normal un-irradiated intestine of BALB/c and B6$^{Min/+}$, allowing comparison between the two strains to assist in linking target genes to risk processes and possible pathways as modifiers.

Total RNA extracted from BALB/c (n=3) and B6$^{Min/+}$ (n=3) mice showed no evidence of degradation on visualisation by agarose gel electrophoresis. The efficiency of amplification was validated using the house keeping probe, *Gapdh,* which gave consistent measurements with little variability. The mRNA expression levels of target genes were determined and the expression in the same intestinal fragment compared (either USI or LSI) between BALB/c and B6$^{Min/+}$ (Table 5-2). All results were validated with repeats and each assay (n=3) was carried out in triplicates (n=9) in separate 96 well plates.

Table 5-2 Summary statistics of *Duox2, Rac1* and *Tiam1* expression in the upper small intestine (USI) and lower small intestine (LSI) of BALB/c (B) or B6$^{Min/+}$ (M). The data include the average gene expression signal (Mean), standard error of the mean (SEM), fold change in expression compared to the highest gene expression value and p values as analysed by ANOVA,

		USI		**LSI**	
		B6$^{Min/+}$	**BALB/c**	**B6$^{Min/+}$**	**BALB/c**
Duox2	Mean	60.54	11308.89	7.12	44.69
	SEM	26.49	4672.02	1.42	9.22
	p value	**B vs M p=<0.0001**		**B vs M p=<0.0001**	
	Fold	186.8		6.3	
Rac1	Mean	11.45	8.89	6.49	3.26
	SEM	2.16	1.55	1.8	0.74
	p value	N/A		**M vs B p=0.001**	
	Fold	1.3		2.0	
Tiam1	Mean	21097.05	6440.46	2709.09	1459.69
	SEM	10920.74	2300.48	1116.7	292.12
	p value	**M vs B p=1.17 E-03**		N/A	
	Fold	3.3		1.9	

Duox2 is expressed preferentially with higher expression in the USI of BALB/c over the LSI from the same mouse and almost 187 times higher expression in the USI of BALB/c compared to the USI of B6$^{Min/+}$ (Table 5-2 and Figure 5-3). These data show highly specific *Duox2* expression in the USI of BALB/c strain.

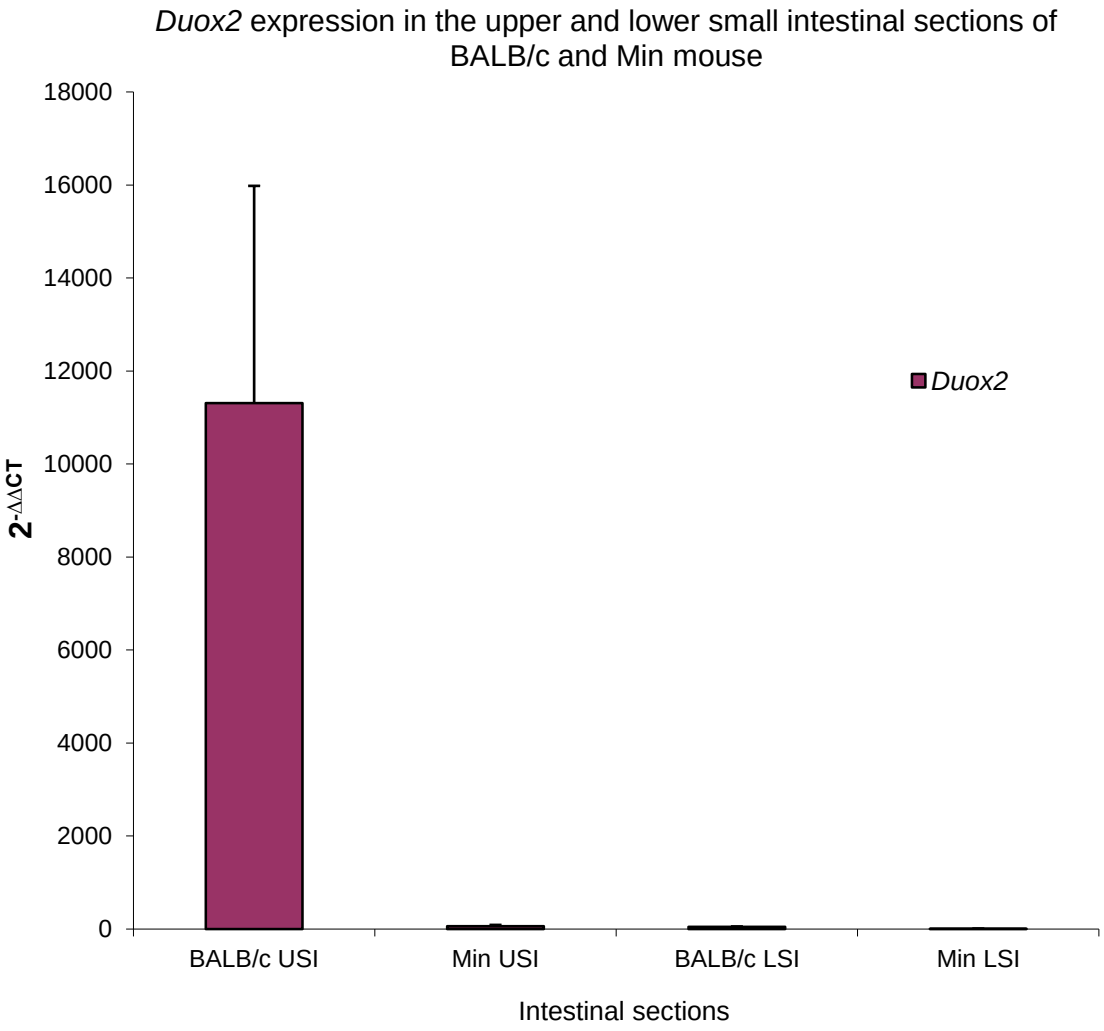

Figure 5-3 *Duox2* shows a significant increase in expression between BALB/c compared to corresponding segments of the B6$^{Min/+}$ A 187-fold increase of *Duox2* was observed. Standard error of the mean (SEM) indicated by error bars.

5.4 DISCUSSION

Following a genome wide scan and faced with a critical challenge of identifying target genes from over 1,000 within our *Mrip* regions made thorough candidate gene selection a near impossible task. We ameliorated this process by combining multiple techniques, namely *in silico* and literature search to present a final analysis that incorporated aspects to provide a summary outcome of high priority target genes.

5.4.1 Chromosome 5 targets

Two main-effect QTLs were identified using IM on chromosome 5. Both showed highly significant associations between adenoma numbers in the USI associated with marker intervals at D5Mit201 and D5Mit139. *Arap2, Rcf1, Cxcl5* were selected from *in silico.* The involvement of *Rac1* (Chr5) and *Tiam1* (Chr16) in radiation-induced multiplicity would be supported by the significant chr5:16 QTL pair interaction (Table 3-2) and the synergy in risk mechanism may be important for the candidate assessment of these genes. In review of our study outcome, no mutations were found in either *Rac1* or *Tiam1* in the N2 $Apc^{Min/+}$ or parental strains. However, *Rac1* (Chr5) which was expressed in the both BALB/c and $B6^{Min/+}$ mice plays a role in the Wnt pathway. Higher *Tiam1* expression was observed in the USI of $B6^{Min/+}$ (p = 1.1 E-03) when compared with BALB/c (Table 5-2) with a 3 fold increase in mRNA compared to USI BALB/c and a 7.8 fold increase when compared to the LSI of the same strain ($B6^{Min/+}$). No significant expression difference was noted for the LSI segment between BALB/c and $B6^{Min/+}$. These data would suggest there is sufficient

evidence to test the hypothesis that under irradiation *Tiam1* expression is up regulated in B6$^{Min/+}$ (of the USI) which possibly correlates with the levels of *Rac1, to* act as a modifier of IR induced polyp formation. Further support for the involvement of *Rac1* also comes from a study that shows an inhibitory effect on *Nox1*/*Nox3* both of which are abundantly expressed in the colon, (Krause *et al*., 2007). These enzymes share the capacity to transport electrons across the plasma membrane and to generate superoxide and other ROS in response to multiple processes including, hypoxia, wound healing, microbial entry, inflammation, and the various forms of abiotic stress similar to *DUOX2*.

Sequence analysis of *Arap2* confirmed 1 nsSNP change with predicted 'possible damaging' effects to the protein function and was selected as a high priority candidate gene. A recent microarray analysis showed *ARAP2* to be strongly up-regulated under conditions of hydrogen peroxide (H_2O_2) production in *SOD1* mutant cultured cells (Boutahar *et al*., 2011). This suggests that *Arap2* could be involved in IR-induced tumourigenesis in our model. H_2O_2 formation depends greatly on the cellular PH level and ARAP2 contains five PH-domains which suggests a possible physiochemical link between *Duox2* and *Arap2* based on the epistasis detected between main effect QTL pairs; Chr2:5 (Table 3-2). ARAP2 is involved in regulating cell shape and adhesion and in the presence of H_2O_2 and SOD1 mutation it is up-regulated which leads to the increased microbial uptake (Gavicherla *et al*., 2010). DUOX2 plays a role in gut innate immunity by generating an oxidation-mediated antimicrobial response in the epithelial mucosal cell which may be impaired due to the mutated *Arap2* and possible interactions with *Duox2* leading to anti inflammatory response.

5.4.2 Chromosome 2 targets

A single main-effects QTL was detected on chromosome 2. QTL mapping identified a significant association with adenoma numbers in the USI and the D2Mit395. The target genes investigated within these regions are shown in Figure 5-4.

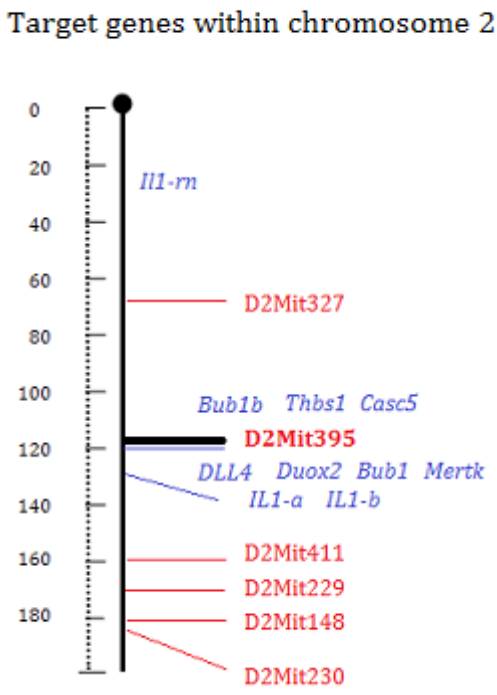

Figure 5-4 Target genes located within *Mrip1* of chromosome 2. The black horizontal bars represent the genetic markers with the highest LRS values against adenoma numbers (QTL). Markers and target genes are displayed in red and blue, respectively. Positions extracted from MPD.

Thbs1 is reported to have some connectivity in the inflammatory response and is thought to be involved in the regulation of adenoma formation in the *Min* mouse model (Gutierrez *et al.*, 2003). Our sequencing data reveals two nsSNPS changes between BALB/c and B6$^{Min/+}$ which are predicted to influence protein function

(Table 5-2). *Thbs1* is strongly expressed in normal colonic epithelial cells and involved in the activation of the Wnt signalling pathway in intestinal epithelial cells (Jo *et al*., 2005). Recently, *Thbs1* was shown to play a role as a negative regulator in IBD associated angiogenesis, (Danese, 2008) and is already established as a low-penetrance susceptibility allele in CRC (Webb *et al*., 2006).

BUB1B plays a critical role in spindle checkpoint activation and deficiencies lead to a compromised response to DNA damage via its interaction with *PARP1*, mediating DNA damage responses via *TP53* activity (Wang *et al*., 2004). *Bulb1b* also demonstrates genomic instability in *Min* mice, and, therefore, plays a role in the development of adenomatous tumours (Rao *et al*., 2005). A total of 6 nsSNPs were confirmed in our mouse model suggesting possible markers for irradiation-induced damage via cell cycle arrest (Table 5-2).

Dll4 is involved in the Notch signalling pathway. In this study 2 nsSNP changes were confirmed, one of which was predicted to have damaging effects on the protein. A recent study revealed that Dll4 is up regulated in conditions of hypoxia and intestinal injury and was expressed in the endothelium of colon cancers (Jubb *et al*., 2009; Fox *et al*., 2009).

Duox2 is a high priority candidate gene and will be discussed in more detail in the next chapter. Here, we discuss some general features and 'gene to promoter' analysis in order to understand how it may be regulated. We used the Frameworker utility from Genomatix; (http://www.genomatix.de/products/index.html) Genomatix Software GmbH, Munich, Germany) (Quandt *et al*., 1995) to apply comparative

analysis to identify gene-gene interactions and investigate alternative transcription elements and promoters linked to the *Duox2*. Based on DNA similarities Frameworker analysis revealed 5 transcription factor binding sites (TFBS) elements to be in common by being on the same strand.

FrameWorker was set to analyse oligo similarities based on different quorum constraint parameters and 66% of the input sequences had to contain the framework. This showed a two promoter element model V$HEAT and V$ETSF where 66.5% of vertebrates have V$HEAT and 99% have V$ETSF. ModelInspector used this 2 element model to detect the number of genes and or loci that share this complex framework and showed 88 genes/promoters from a 51,000 promoter sequence library scanned for DNA matches. These genes were added to the bibliosphere analysis (literature datawarehouse) and the analysis was repeated showing only one gene Mediator complex subunit 15 (*Med15*) with high specificity to share a co-regulatory sequence in the mouse genome with *Duox2*.

Med15 is located on Chr16, 9.9cM. This is located 0.7cM distally from chromosome 16 QTL at *Prkdc* (9.2cM). The function of this gene is not so well defined in the intestine, but appears to play a crucial role in the regulation of transcription. It is also involved in gene 'silencing' by promoter hypermethylation and is associated with nasopharyngeal carcinoma (Kwong *et al*., 2005). *Duox2* is also silenced by promoter methylated in lung cancer (Luxen *et al*., 2008). Mouse Phenome Database (MPD) showed 5 nsSNPs in *Med15* but neither were polymorphic between the BALB/c and B6 strains, which could suggests epigenetic mechanisms affecting the dysregulation of cell expression in *Med15*.

As briefly mentioned previously, *Duox2* is involved in the oxidase system and plays a role in generation of ROS. The DUOX2 protein is located at the apical membrane of the enterocytes in the brush border villa and acts as an immune barrier against pathogenic microbes and is believed to play a potential host defence role in IBD and mucosal inflammation (Rokutan *et al*., 2006). It is a transmembrane glycoprotein with a unique N-terminal heme peroxidase-like domain which protrudes inside the cell as well as a membrane-spanning NADPH oxidase domain with an additional cytosolic segment containing two calcium-binding domains. *DUOX2* is a member of the NADPH oxidase (NOX) family of proteins; NOX1, 2, 3, 4 and NOX5, along with DUOX1 and are membrane bound proteins that generate ROS (Fischer *et al*., 2009). *DUOX2* is expressed in the human colon, small intestine and duodenum, (El Hassani *et al*., 2005) and we also report significant difference in expression in BALB/c over B6 (Figure 5-3). The expression of dual oxidase maturation factor 2 (*DuoxA2*) is mandatory for the translocation of *Duox2* to the plasma membrane to initiate functional enzymatic activity (Grasberger and Refetoff, 2006) with more recent observations of *DuoxA2* mutations affecting the type and amount of ROS response (Hoste *et al*., 2012). *Duox2* as a modifier of CRC will be discussed in the next section.

CHAPTER 6 SUMMARY AND FURTHER WORK

In this study genotype-phenotype associations were investigated to allow detection of QTLs, which may harbour modifiers of IR induced polyposis in the *Min* mouse. This involved assessing the polyp multiplicity in 143 mice in response to 2Gys of irradiation and using numerous bioinformatics tools to identify candidate genes.

My studies found that the majority of genes involved were linked to several missense polymorphisms shown to be associated with susceptibility via their influence on the inflammatory processes. *Bub1b* and *Casc5*, which are involved in known IR response pathways such as cell cycle arrest, proliferation DNA repair pathways should also be followed up. In addition, I propose developing an understanding of the genes involved in transcriptional changes of genes involved in cellular shape and adhesion (*Arap2*), Wnt signalling targets (*Rac1* and *Tiam1*), and an inflammatory associated gene (*Duox2*). Expression analysis of *Duox2* in BALB/c and $B6^{Min/+}$ alongside the detection of nsSNPs by sequence analysis, identified *Duox2* as by far the 'most likely' candidate gene within *Mrip1.* This finding will be discussed further here.

I hypothesise that *Duox2* is a modifier of polyp formation in the *Min* mouse following irradiation induced oxidative stress, which results in cytokine and ROS release. As mentioned previously, *DUOX2* is expressed throughout the colon and recently observed to be overexpressed in human pancreatic cell lines (Wu *et al*., 2011). In the respiratory tract, it has been demonstrated that DUOX2-generated H_2O_2 kills bacteria in conjunction with lactoperoxidase and thiocyanate (synthesised to

hypothiocyanite, a bactericide). In this process, DUOX2 is induced by interferon gamma (IFNγ) and the killing effect is important for host defence. More recently, a linkage study conducted by Esworthy and colleagues defined a region glutathione peroxidase (*Gpx*)-deficiency-associated colitis 1 (*Gdac1*) (Esworthy *et al*., 2012), which overlaps with *Mrip1* detected by us in the USI and LI (Figure 2-8) and the QTL *Cdcs3* detected in inbred mice treated to develop IBD (Farmer *et al*., 2001). This places significant weight on my findings indicating a link between colitis and irradiation induced polyposis. Esworthy and colleagues used a colitis mouse model involving resistant B6 and susceptible 129S1/SvlmJ (129) strain with double knock out technology to produce mice deficient in *Gpx1* and *Gpx2*. Using similar *in silico* techniques for gene selection and elimination they studied the gut microflora in ileocolitis and reported *Duox2* as the most worthy of the 10 colitis associated candidate genes involved in the promotion of inflammation.

DUOX2 produces H_2O_2 as part of the innate immune system and H_2O_2 is a well-known mutagen. It appears to be associated with ROS species in the promotion of inflammation, which is a well-known consequence of gastrointestinal inflammation in IBD patients, leading to dysplasia and CRC and a common adverse reaction to IR exposure. The implications of my findings and those of others suggest a connection between *DUOX2* and inflammation, and GI toxicity caused by IR and subsequent UC-associated dysplasia and adenoma formation. This study of *Duox2* makes it especially relevant for cancers induced by chronic inflammation and also IR induced GI damage. Targeting *DUOX2* and or *DUOXA2* expression and their downstream targets in humans could improve the treatment phase of IBD lessening inflammatory symptoms, while reducing the risk of progression to cancer. This study urges a

better understanding of the role of *DUOX2* and the need to investigate the possibility of missense mutations in human populations at risk of IBD. I propose that *DUOX2* might also be involved in the regulation of Wnt signaling in intestinal cells involving the B6$^{Min/+}$ BC, possibly via the NOD pathway which forms part of innate immunity to sense microbial components derived from bacterial peptidoglycan. *NOD1* has also been shown to participate in host defence and mutations occur in patients of Crohn's disease. Irradiation of mouse intestinal cells carrying IR susceptibility genes possibly from BALB/c has allowed me to link the generation of ROS through *Duox2*, and speculate on possible cross talk between ROS and Wnt β-catenin signaling pathway, which is known to be responsive for polyp formation. This will require experimental verification.

I also propose that oxidative DNA damage has, in part, led to the cellular transformation and development of adenomas in our mouse model due to the defective DNA repair mechanism, introduced by *Prkdc* of BALB/c strain. It can be suggested therefore that ROS and oxidative damage induced by IR may possibly lead to tumour progression via enhanced *Duox2* expression in BALB/c, which is confirmed by the high mRNA expression observed in wild type BALB/c.

Since the completion of the present study's experiments in 2009, advancements in next generation sequencing (NGS) technologies have propelled genomic medicine and research into a new era. Targeted sequencing of protein coding regions (the functional exome) combined with the predictive power of bioinformatics tools now provide high resolution analysis of nsSNPs in cancer genomes. Timmerman and colleagues provided the first whole exome NGS of primary colon cancers and

identified over 50,000 small nucleotide variations (Timmermann *et al*., 2010). Therefore, exome sequencing of specific DNA tracts will play a bigger part in candidate gene identification in the future (Biesecker *et al*., 2010). With proof of concept to localise causative genes behind rare and complex disorders it is believed that NSG techniques will help improve the success rate of QTL to gene identification and the discovery of the underlying pathogenic mutation and with this information novel treatments can then be developed.

6.1 Concluding remarks

The data presented in this thesis supports the hypothesis that additional *Mrip* detected on chromosome 2 and 5 for the USI and 2 and 8 for the LI influence increased polyp multiplicity following IR. The search for candidate genes has provided insights into the importance of oxidative stress caused by *Duox2* as a modifier of $Apc^{min/+}$ mice and the importance of the Wnt pathway in the pathogenesis of IR induced adenoma progression. Additionally, further interpretation of the model used here present mechanistic links between IR-induced GI inflammation and IBD associated CRC. Whilst *Duox2* has emerged as the most prominent candidate gene, other genes are worthy of further investigation. In addition, it is also likely that genes underlying the *Mrip* QTLs are representative of novel loci which when mutated may cause disruption to normal development and or the symbiosis of the intestinal system, and may lead to tumour formation. My study presents opportunities for new discoveries with the application of NGS techniques targeted to the regions identified by QTL mapping and epistasis analysis followed by appropriate functional studies.

REFERENCE LIST

Aaltonen, L. A., Peltomäki, P., Leach, F. S., Sistonen, P., Pylkkänen, L., Mecklin, J. P, Järvinen, H., Powell, S. M., Jen, J., and Hamilton, S. R. (1993). Clues to the pathogenesis of familial colorectal cancer. Science 260, 812-6.

Abal, M., Obrador-Hevia, A., Janssen, K. P., Casadome, L., Menendez, M., Carpentier, S., Barillot, E., Wagner, M., Ansorge, W., Moeslein, G., Fsihi, H., Bezrookove, V., Reventos, J., Louvard, D., Capella, G., Robine, S. (2007) APC inactivation associates with abnormal mitosiscompletion and concomitant BUB1B/MAD2L1 up-regulation. Gastroenterology. 132, 2448–58.

Abiola, O. (2003). The nature and identification of quantitative trait loci: a community's view. Nat. Rev. Genet, 4, 911-916.

Ahmadi, A., Polyak, S. and Draganov, P.V. (2009) Colorectal cancer surveillance in inflammatory bowel disease: the search continues. World J Gastroenterol. 15, 61-6.

Alberici, P., de Pater, E., Cardoso, J., Bevelander, M., Molenaar, L., Jonkers, J. and Fodde, R. (2007). Aneuploidy arises at early stages of Apc-driven intestinal tumorigenesis and pinpoints conserved chromosomal loci of allelic imbalance between mouse and human. Am J Pathol. 170, 377-87.

Albuquerque, C., Breukel, C., van der Luijt, R., Fidalgo, P., Lage, P., Slors, F.J., Leitão, C.N., Fodde, R. and Smits, R. (2002). The 'just-right' signaling model: APC somatic mutations are selected based on a specific level of activation of the beta-catenin signaling cascade. Hum Mol Genet. 11, 1549-60.

Alexander, D. D., Weed, D. L., Cushing, C. A. and Lowe K. A. (2011). Meta-analysis of prospective studies of red meat consumption and colorectal cancer. Eur J Cancer Prev. 4, 293-307

Al-Tassan N., Chmiel N. H., Maynard J., Fleming N., Livingston A. L., Williams G. T., Hodges A. K., Davies D. R., David S. S., Sampson J. R. and Cheadle J. P. (2002).

Inherited variants of MYH associated with somatic G:C>T:A mutations in colorectal tumors. Nat. Genet. 30, 227–232.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-402.

Andreyev, H. J., Vlavianos, P., Blake, P., Dearnaley, D., Norman, A. R., Tait, D. (2005). Gastrointestinal symptoms after pelvic radiotherapy: role for the gastroenterologist? Int J Radiat Oncol Biol Phys. 62, 1464-71.

Andreyev, H. J. (2007). Gastrointestinal problems after pelvic radiotherapy: the past, the present and the future. Clin Oncol. 19, 790-9.

Aretz, S., Uhlhaas, S., Goergens, H., Siberg, K., Vogel, M., Pagenstecher, C., Mangold, E., Caspari, R., Propping, P., Friedl, W. (2006). MUTYH-associated polyposis: 70 of 71 patients with biallelic mutations present with an attenuated or atypical phenotype. Int J Cancer. 119, 807-14.

Bailey, S. M., Meyne, J., Chen, D. J., Kurimasa, A., Li, G. C., Lehnert, B. E., Goodwin, E. H. (1999). DNA double-strand break repair proteins are required to cap the ends of mammalian chromosomes. Proc Natl Acad Sci U S A. 21, 14899-904.

Balkwill, F. and Mantovani, A. (2001). Inflammation and cancer: back to Virchow? Lancet. 17, 539-45.

Balmain, A. (2002). Cancer as a complex genetic trait: tumor susceptibility in humans and mouse models. Cell. 108, 145–152.

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

Barker, N., van Es J, Kuipers, J., Kujalam P, van den Born M., Cozijnsen, M, et al. (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature. 449, 1003-07.

Barker, N., Clevers, H. (2012). Leucine-Rich Repeat-Containing G-Protein-Coupled Receptors as Markers of Adult Stem Cells. Gastroenterology. 138, 1681–96.

Barton, N. H. and Keightley, P. D. (2002). Understanding quantitative genetic variation. Nat Rev Genet. 1, 11-21.

Bassaganya-Riera, J., Viladomiu, M., Pedragosa, M., De Simone, C. and Hontecillas, R. (2012). Immunoregulatory mechanisms underlying prevention of colitis-associated colorectal cancer by probiotic bacteria. PLoS One.7(4).

Basten ,C. J., Weir, B. S., Zeng, Z. B., (1994). Zmap–a QTL cartographer. In Proceedings of the 5th World Congress on Genetics Applied to Livestock Production: Computing Strategies and Software, edited by C. Smith, J. S. Gavora, B. Benkel, J. Chesnais, W. Fairfull, J. P. Gibson, B. W. Kennedy and E. B. Burnside. 22, 65-66. Published by the Organizing Committee, 5th World Congress on Genetics Applied to Livestock Production, Guelph, Ontario, Canada.

Basten, C. J., Weir, B. S. and Zeng, Z. B. (2004). QTL Cartographer, Version 1.17. Departmentof Statistics, North Carolina State University, Raleigh, NC.

Bateson, W., Saunders, E. R. and Punnett, R. C. (1906). Experimental studies in the physiology of heredity. Reports to the Evolution Committee, Royal Society of London. 3.

Baum, L. E., Petrie, T., Soules, G. and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Ann. Math. Statist.. 41, 164–171.

Baxter, N. N., Tepper, J. E., Durham, S. B., Rothenberger, D. A., Virnig, B. A. (2005). Increased risk of rectal cancer after prostate radiation: a population-based study. Gastroenterology. 128, 819-24.

Baylin, S. B., Herman, J. G., Graff, J. R., Vertino, P. M., Issa, J. P. (1998). Alterations in DNA methylation: a fundamental aspect of neoplasia. Adv Cancer Res. 72, 141–196.

Beazer-Barclay, Y., Levy, D. B., Moser, A. R., Dove, W. F., Hamilton, S. R., Vogelstein, B., Kinzler, K. W. (1996). Sulindac suppresses tumorigenesis in the Min mouse. Carcinogenesis. 17, 1757-60.

Bedard, K., and Krause, K. H. (2007) The NOX family of ROS-generating NADPH oxidases: physiology and pathophysiology. Physiol Rev 87: 245–313.

Bedell, M. A., Jenkins, N. A. and Copeland, N. G. (1997). Mouse models of human disease. Part I: techniques and resources for genetic analysis in mice. Genes Dev. 11, 1-10.

Behrens, J. (2005). The role of the Wnt signalling pathway in colorectal tumorigenesis. Biochem. Soc. Trans. 33, 672-675.

BEIR V, Biological Effects of Ionising Radiation. (1990). Health Effects of Exposure to low levels if ionizing radiation. National Academy Press, ISBN-13: 978-0-309-07358-5

Béroud, C. and Soussi, T. (1996). APC gene: database of germline and somatic mutations in human tumors and cell lines. Nucleic Acids Res. 24, 121-4.

Bertario, L., Russo, A., Sala, P., Varesco, L., Giarola, M., Mondini, P., Pierotti, M., Spinelli, P. and Radice, P. (2003). Multiple approach to the exploration of genotype-phenotype correlations in familial adenomatous polyposis. J Clin Oncol. 21, 1698–707.

Bhandari, S., Sinha, A. and Clark, S. A. (2011). Evaluation of management of desmoids tumours associated with familial adenomatous polyposis in Dutch patients. Br J Cancer. 104, 1236.

Bienz, M. and Clevers, H. (2000). Linking colorectal cancer to Wnt signaling. Cell 103, 311-320.

Biesecker, L. G. (2010). Exome sequencing makes medical genomics a reality. Nat Genet. 42, 13–14.

Bodmer, W. F., Bailey, C. J., Bodmer, J., Bussey, H. J., Ellis, A., Gorman, P., Lucibello, F. C., Murday, V. A., Rider, S. H. and Scambler, P. (1987). Localization of the gene for familial adenomatous polyposis on chromosome 5. Nature. 328, 614-616.

Boer, M. P., Ter Braak, C. J., Jansen, R. C. (2002). A penalized likelihood method for mapping epistatic quantitative trait loci with one-dimensional genome searches. Genetics. 162, 951-60.

Boone, C., Bussey, H. and Andrews, B. J. (2007). Exploring genetic interactions and networks with yeast. Nat Rev Genet. 8, 437-49.

Bourne, R. G., Kearsley, J. H., Grove, W. D., Roberts, S. J. (1983). The relationship between early and late gastrointestinal complications of radiation therapy for carcinoma of the cervix. International Journal of Radiation Oncology * Biology * Physics. 10, 1445-50

Boutahar, N., Wierinckx, A., Camdessanche, J. P., Antoine, J. C., Reynaud, E., Lassabliere, F., Lachuer, J. and Borg, J. (2011). Differential effect of oxidative or excitotoxic stress on the transcriptional profile of amyotrophic lateral sclerosis-linked mutant SOD1 cultured neurons. J Neurosci Res. 89, 1439-50.

Brenner , D. J. (2010) Should we be concerned about the rapid increase in CT usage? Rev Environ Health. 25, 63-8.

Brenner, D. J., Doll, R., Goodhead, D. T., Hall, E. J., Land, C. E., Little, J. B., Lubin, J. H., Preston, D. L., Preston, R. J., Puskin, J. S., Ron, E., Sachs, R. K., Samet, J. M., Setlow, R. B. and Zaider, M. (2003). Cancer risks attributable to low doses of ionizing radiation: assessing what we really know. Proc Natl Acad Sci U S A. 100, 13761-6.

Broman, K. W., Wu, H., Sen, Ś. and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. Bioinformatics. 19, 889-890.

Bujanda, L., Cosme, A., Gil, I. and Arenas-Mirave, J. I. (2010). Malignant colorectal polyps. World J. Gastroenterol. 16, 3103-3111.

Buongiorno, P., Pethe, V. V., Charames, G. S., Esufali, S. and Bapat, B. (2008). Rac1 GTPase and the Rac1 exchange factor Tiam1 associate with Wnt-responsive

promoters to enhance beta-catenin/TCF-dependent transcription in colorectal cancer cells. Mol Cancer. 7, 73.

Burgess-Herbert, S. L., Cox, A., Tsaih, S. W. and Paigen, B. (2008). Practical applications of the bioinformatics toolbox for narrowing quantitative trait loci. Genetics. 180, 2227-35.

Butcher, D. T. and Rodenhiser, D. I. (2007). Epigenetic inactivation of BRCA1 is associated with aberrant expression of CTCF and DNA methyltransferase (DNMT3B) in some sporadic breast tumours. Eur. J. Cancer 43, 210-219.

Caldwell, C. M., Green, R. A., Kaplan, K. B. (2007). APC mutations lead to cytokinetic failures in vitro and tetraploid genotypes in Min mice. J Cell Biol. 178, 1109-20.

Calvanese, V., Lara, E., Kahn, A., Fraga,M. F. (2009). The role of epigenetics in aging and age-related diseases. Ageing Res. Rev. 8, 268-276.

Cancer Research UK CancerStats Incidence 2008 – UK April (2011).

Carlborg, Ö., Andersson, L. and Kinghorn,B. (2000). The use of a genetic algorithm for simultaneous mapping of multiple interacting quantitative trait loci. Genetics, 155, 2003–2010.

Carter, J. H., Douglass, L. E., Deddens, J. A., Colligan, B. M., Bhatt, T. R., Pemberton, J. O., Konicek, S., Hom, J., Marshall, M. and Graff, J. R. (2004). Pak-1 expression increases with progression of colorectal carcinomas to metastasis. Clin Cancer Res. 10, 3448–3456.

Cerrotta, A., Gardani, G., Lozza, L., Kenda, R., Tana, S., Valvo, F. and Zucali, R. (1995). Ileal obstruction following radiosurgical treatment for rectosigmoid neoplasm. Radiol Med. 89,643-6.

Chakravarti, A. (2001). Single nucleotide polymorphisms: to a future of genetic medicine. Nature 409, 822-823.

Chasman, D. and Adams, R. M. (2001). Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: Structure-based assessment of amino acid variation. J Mol Biol 307, 683–706.

Chell, S., Patsos, H. A., Qualtrough, D., Zadeh, A. M., Hicks, D. J., Kaidi, A., Witherden, I.R., Williams, A. C., Paraskeva, C. (2005). Prospects in NSAID-derived chemoprevention of colorectal cancer. Biochem. Soc. Trans. 33, 667-671.

Chen, H. M., Fang, J. Y. (2009). Genetics of the hamartomatous polyposis syndromes: a molecular review. Int J Colorectal Dis. 8,865-74.

Chiu, C. H, McEntee, M. F, Whelan, J. (1997). Sulindac causes rapid regression of preexisting tumors in Min/+ mice independent of prostaglandin biosynthesis. Cancer Res. 57, 4267–4273.

Cho, K. H., Lee, C. K., Levitt, S. H. (1995). Proctitis after conventional external radiation therapy for prostate cancer: importance of minimizing posterior rectal dose. Radiology. 3, 699-703.

Churchill, G. A., and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. Genetics 138, 963-971.

Clark, A., and Wang, L. (1997). Epistasis in measured genotypes: Drosophila P-element insertions. Genetics, 147, 157–63.

Clevers, H. (2006). Wnt/beta-catenin signaling in development and disease. Cell 127, 469-480.

Colnot, S., Niwa-Kawakita, M., Hamard, G., Godard, C., Le Plenier, S., Houbron, C., Romagnolo, B., Berrebi, D., Giovannini, M., Perret, C. (2004). Colorectal cancers in a new mouse model of familial adenomatous polyposis: influence of genetic and environmental modifiers, Lab. Invest. 84, 1619–1630.

Cordell, H. J. (2002). Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum Mol Genet. 11, 2463-8.

Cordell, H. J. (2009). Detecting gene-gene interactions that underlie human diseases. Nat Rev Genet. 6, 392-404.

Cormier, T., Bilger, A., Lillich, A., Halberg, R., Hong, K., Gould, K., Borenstein, N., Lander, E., Dove, W. (2000). The Mom1AKR intestinal tumor resistance region consists of Pla2g2a and a locus distal to D4Mit64. Oncogene 19, 3182-3192.

Danese, S. (2008). Negative regulators of angiogenesis in inflammatory bowel disease: thrombospondin in the spotlight. Pathobiology. 75, 22-4.

Darvasi, A. (2001). In Silico Mapping of Mouse Quantitative Trait Loci, Science, . 294,.2424– 2426.

Dean, M,. Fojo, T., Batesman S. (2005). Tumour stem cells and drug resistance. Nat Rev Cancer. 5, 275-84.

De Belle, J. S. and Heisenberg, M. (1996). Expression of Drosophila mushroom body mutations in alternative genetic backgrounds: a case study of the mushroom body miniature gene (mbm). Proc. Natl. Acad. Sci. U S A. 93, 9875–9880.

de la Chapelle, A. (2004). Genetic predisposition to colorectal cancer. Nat Rev Cancer. 4; 769-80

De Filippo, C., Luceri, C., Caderni, G., Pacini, M., Messerini, L., Biggeri, A., Mini, E., Tonelli, F., Cianchi, F. and Dolara, P. (2002). Mutations of the APC gene in human sporadic colorectal cancers. Scand. J. Gastroenterol. 37, 1048–1053.

Debinski, H. S., Love, S., Spigelman, A. D. and Phillips, R. K. (1996). Colorectal polyp counts and cancer risk in familial adenomatous polyposis. Gastroenterology 110, 1028-30.

DeBry, R. W. and Seldin, M. F. (1996). Human/mouse homology relationships. Genomics. 33, 337-51.

Degg, N. L., Weil, M. M., Edwards, A., Haines, J., Coster, M., Moody, J., Ellender, M., Cox, R. and Silver, A. (2003). Adenoma multiplicity in irradiated ApcMin mice is modified by chromosome 16 segments from Balb/c. Cancer Res. 63, 2361-2363.

Demant, P. (2003). Cancer susceptibility in the mouse: genetics biology and implications for human cancer. Nat Rev Genet. 4, 721-734.

Dermitzakis, E. T., Reymond, A., Lyle, R., Scamuffa. N., Ucla, C., Deutsch, S., Stevenson, B. J., Flegel, V., Bucher, P., Jongeneel, C. V. and Antonarakis, S. E. (2002). Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature. 420, 578–582.

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

Dietrich, W. F., Miller, J., Steen, R., Merchant, M. A., Damron-Boles, D., Husain, Z., Dredge ,R., Daly, M. J., Ingalls, K. A and O'Connor, T. J. (1996). A comprehensive genetic map of the mouse genome. Nature (Lond) 380, 149–152.

Dimberg, J., Dienus, O., Lofgren, S., Hugander, A., Wagsater, D. (2007). Expression and gene polymorphisms of the chemokine CXCL5 in colorectal cancer patients. Int J Oncol. 31, 97–102.

DiPetrillo, K., Wang, X., Stylianou, I. M. and Paigen, B. (2005). Bioinformatics toolbox for narrowing rodent quantitative trait loci. Trends Genet. 12, 683-92.

Dirisina, R., Katzman, R. B., Goretsky, T., Managlia, E., Mittal, N., Williams, D. B., Qiu, W., Yu, J., Chandel, N. S., Zhang, L. and Barrett, T. A. (2011). p53 and PUMA independently regulate apoptosis of intestinal epithelial cells in patients and mice with colitis. Gastroenterology. 141, 1036-45.

Dobbie, Z., Heinimann, K., Bishop, D. T., Muller, H. and Scott, R. J. (1997). Identification of a modifier gene locus on chromosome 1p35-36 in familial adenomatous polyposis. Hum. Genet. 99, 653-657.

Dragani, T. A. (2003). 10 years of mouse cancer modifier loci: human relevance. Cancer Res. 63, 3011–3018.

Drayton, M. and Noben-Trauth, K. (2006). Mapping quantitative trait loci for hearing loss in Black Swiss mice. Hear Res. 212, 128-39.

Dryja, T. P., Friend, S. and Weinberg, R. A. (1986). Genetic sequences that predispose to retinoblastoma and osteosarcoma. Symp Fundam Cancer Res. 39, 115-9.

Duff, E. K. and Clarke, A. R. (1998). Smad4 (DPC4)-a potent tumour suppressor? Br J Cancer. 78, 1615-9.

Dunlop, M. G., Dobbins, S. E., Farrington, S. M., Jones, A. M., Palles, C., Whiffin, N., Tenesa, A., Spain, S., Broderick, P., Ooi, L. Y., Domingo, E., Smillie, C., Henrion, M., Frampton, M., Martin, L., Grimes, G., Gorman, M., Semple, C., Ma, Y. P., Barclay, E., Prendergast, J., Cazier, J. B., Olver, B., Penegar, S., Lubbe, S., Chander, I., Carvajal-Carmona, L. G., Ballereau, S., Lloyd, A., Vijayakrishnan, J., Zgaga, L., Rudan, I., Theodoratou, E.; Colorectal Tumour Gene Identification (CORGI) Consortium, Starr, J. M., Deary, I., Kirac, I., Kovacević, D., Aaltonen, L. A., Renkonen-Sinisalo, L., Mecklin, J. P., Matsuda, K., Nakamura, Y., Okada, Y., Gallinger, S., Duggan, D. J., Conti, D., Newcomb, P., Hopper, J., Jenkins, M. A., Schumacher, F., Casey, G., Easton, D., Shah, M., Pharoah, P., Lindblom, A., Liu, T.; Swedish Low-Risk Colorectal Cancer Study Group, Smith, C. G., West, H., Cheadle, J. P.; COIN Collaborative Group, Midgley, R., Kerr, D. J., Campbell, H., Tomlinson, I. P., Houlston, R. S. (2012). Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. Nat Genet. 44, 770-6.

Dupuis, J., Brown, P. O. and Siegmund, D. (1995). Statistical methods for linkage analysis of complex traits from high-resolution maps of identity by descent. Genetics 140, 843-56.

East, E. M. (1910). A Mendelian interpretation of variation that is apparently continuous. Am. Nat. 44, 65–82.

Efron, B. and Tibshirani, R. J. (1993). An introduction to the bootstrap. Monographs on Statistics and Applied Probability, 57. Chapman and Hall, New York. ISBN: 0-412-04231-2.

El Hassani, R. A., Benfares, N., Caillou, B., Talbot, M., Sabourin, J. C., Belotte, V., Morand, S., Gnidehou, S., Agnandji, D., Ohayon, R., Kaniewski, J., Noël-Hudson, M. S., Bidart, J. M., Schlumberger, M., Virion, A. and Dupuy, C. (2005). Dual oxidase2 is expressed all along the digestive tract. Am J Physiol Gastrointest Liver Physiol. 288, 933-42.

Esufali, S. and Bapat, B. (2004). Cross-talk between Rac1 GTPase and dysregulated Wnt signaling pathway leads to cellular redistribution of beta-catenin and TCF/LEF-mediated transcriptional activation. Oncogene. 23, 8260-71.

Esworthy, R. S., Kim, B. W., Rivas, G. E., Leto, T. L., Doroshow, J. H., Chu, F. F. (2012). Analysis of candidate colitis genes in the Gdac1 locus of mice deficient in glutathione peroxidase-1 and -2. PLoS One. 7, e44262.

Farmer, M. A., Sundberg, J. P., Bristol, I. J., Churchill, G. A., Li, R., Elson, C. O. and Leiter, E. H. (2001). A major quantitative trait locus on chromosome 3 controls colitis severity in IL-10- deficient mice. Proc Natl Acad Sci U S A. 98, 13820–13825.

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

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

Fearon, E. R. and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. Cell. 61, 759-767.

Fearon, E. R., Hamilton, S. R. and Vogelstein, B. (1987). Clonal analysis of human colorectal tumors. Science. 238, 193-197.

Feldser, D. M., Hackett, J. A., Greider, C. W. (2003). Telomere dysfunction and the initiation of genome instability. Nat. Rev. Cancer. 3, 623-627.

Fijneman, R.J. de Vries. S. S., Jansen, R.C. and Demant P. (1996). Complex interactions of new quantitative trait loci, Sluc1, Sluc2, Sluc3, and Sluc4, that influence the susceptibility to lung cancer in the mouse. Nat. Genet. 14, 465–467.

Fischer, H. (2009). Mechanisms and Functions of Duox in Epithelia of the lung. Antioxid. Redox Signal. 11, 2453–2465.

Fisher, R., Pusztai, L., Swanton, C. (2013). Cancer heterogeneity: implications for targeted therapeutics. Br J Cancer. 108, 479-85.

Flint, J., Valdar, W., Shifman, S. and Mott, R. (2005). Strategies for mapping and cloning quantitative trait genes in rodents. Nature Reviews Genetics. 6, 271-286.

Flint, J.and Mackay, T. F. (2009). Genetic architecture of quantitative traits in mice, flies, and humans. Genome Res. 19, 723-33.

Fodde, R., Edelmann, W., Yang, K., van Leeuwen, C., Carlson, C., Renault, B., Breukel, C., Alt, E., Lipkin, M. and Khan,P. M. (1994). A targeted chain termination mutation in the mouse Apc gene results in multiple intestinal tumors, Proc. Natl. Acad. Sci. U S A 91, 8969–8973.

Fodde, R.and Smits, R. (2002). Cancer biology: A matter of dosage. Science 298, 761–763.

Folkman, J. (2002). Role of angiogenesis in tumor growth and metastasis. Semin Oncol. 29, 15–18.

Followill, D. S., Kester, D. and Travis, E. L. (1993). Histological changes in mouse colon after single- and split-dose irradiation. Radiat Res. 136, 280-8.

Fox, S. B.and Harris, A. L. (2009) Expression of delta-like ligand 4 (Dll4) and markers of hypoxia in colon cancer. Br J Cancer. 101, 1749-57.

French, J. E., Lacks, G. D., Trempus, C., Dunnick, J. K., Foley, J., Mahler, J., Tice, R. R., Tennant, R. W. (2001). Loss of heterozygosity frequency at the Trp53 locus in p53-deficient (+/−) mouse tumors is carcinogen-and tissue-dependent. Carcinogenesis. 22, 99–106

Fridrichova, I. (2006). New aspects in molecular diagnosis of Lynch syndrome (HNPCC). Cancer Biomark. 2, 37-49.

Friedl, W., Caspari, R., Sengteller, M., Uhlhaas, S. L. C., Jungck, M., Kadmon, M., Wolf, M., Fahnenstich, J., Gebert, J., Möslein, G., Mangold, E. and Propping, P. (2001). Can APC mutation analysis contribute to therapeutic decisions in familial adenomatous polyposis? Experience from 680 FAP families. Gut. 48, 515-21.

Friend, S. H., Bernards, R., Rogelj, S., Weinberg,.R. A., Rapaport, J. M., Albert, D. M., Dryja, T. P. (1986). A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. Nature. 323, 643-646.

Frush, D. P., Goske, M. J. and Hernanz-Schulman, M. (2008). Computed tomography and radiation exposure. N Engl J Med. 358,851.

Fuchs, C. S., Giovannucci, E. L., Colditz, G. A., Hunter, D. J., Speizer, F. E. and Willett, W. C. (1994). A prospective study of family history and the risk of colorectal cancer. N Engl J Med. 331, 1669-74.

Fulker, D. W., Cherny, S. S., Cardon, L. R. (1995). Multipoint interval mapping of quantitative trait loci, using sib pairs. Am J Hum Genet. 5, 1224–1233.

Galton, F. (1869). Hereditary Genius. London. Macmillan.

Galton, F. (1889). Natural Inheritance London. Macmillan & Co

Galvan, A., Colombo, F., Noci, S., Pazzaglia, S., Mancuso, M., Manenti, G., Broman, K. W., Saran, A., Dragani, T. A. (2012). The Lsktm1 locus modulates lung and skin tumorigenesis in the mouse. G3 (Bethesda). 2, 1041-6.

Gavicherla, B., Ritchey, L., Gianfelice, A., Kolokoltsov, A. A., Davey, R. A. and Ireton, K. (2010). Critical role for the host GTPase-activating protein ARAP2 in InlB-mediated entry of Listeria monocytogenes.Infect Immun. 78, 4532-41.

Gerlai, R. (1996). Gene-targeting studies of mammalian behaviour: is it the mutation or the background genotype? Trends Neurosci. 19, 177–181.

Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, Varela I, Phillimore B, Begum S, McDonald NQ, Butler A, Jones D, Raine K, Latimer C, Santos CR, Nohadani M, Eklund AC, Spencer-

Dene B, Clark G, Pickering L, Stamp G, Gore M, Szallasi Z, Downward J, Futreal PA, Swanton C. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing.Engl J Med. 366, 883-92.

Giardiello, F. M., Brensinger, J. D., Tersmette, A. C., Goodman, S. N., Petersen, G. M., Booker, S. V., Cruz-Correa, M. and Offerhaus, J. A. (2000). Very high risk of cancer in familial Peutz-Jeghers syndrome. Gastroenterology. 119, 1447-53.

Giles, R. H., Van Es, J. H. and Clevers, H. (2003). Caught up in a Wnt storm: Wnt signaling in cancer. Biochim. Biophys. Acta. 1653, 1-24.

Gillen, C. D., Walmsley, R. S., Prior, P., Andrews, H. A. and Allan, R. N. (1994). Ulcerative colitis and Crohn's disease: a comparison of the colorectal cancer risk in extensive colitis. Gut. 35, 1590-2.

Gitton, Y., Dahmane, N., Baik, S., Ruiz, I., Altaba, A., Neidhardt, L., Scholze, M., Herrmann, B. G., Kahlem, P., Benkahla, A., Schrinner, S., Yildirimman, R., Herwig, R., Lehrach, H., Yaspo, M. L. (2002). HSA21 expression map initiative. A gene expression map of human chromosome 21 orthologues in the mouse. Nature. 420, 586-90.

Goffinet, B. and Gerber, S. (2000). Quantitative trait loci: a meta-analysis. Genetics. 155, 463-73.

Good, P. (2005). Permutation, Parametric, and Bootstrap Tests of Hypotheses (Springer Series in Statistics). Springer ISBN 0-387-20279-X

Gould, K. A., Dietrich, W. F., Borenstein, N., Lander, E. S. and Dove, W. F. (1996a). Mom1 is a semi-dominant modifier of intestinal adenoma size and multiplicity in Min/+ mice. Genetics. 144, 1769-76.

Gould, K. A., Luongo, C., Moser, A. R., McNeley, M. K., Borenstein, N., Shedlovsky, A., Dove, W. F., Hong, K., Dietrich, W. F. and Lander, E. S. (1996b). Genetic evaluation of candidate genes for the Mom1 modifier of intestinal neoplasia in mice. Genetics. 144, 1777-85.

Grady, W. M. and Carethers, J. M. (2008). Genomic and epigenetic instability in colorectal cancer pathogenesis. Gastroenterology. 135, 1079-99.

Grady, W. M. (2004). Genomic instability and colon cancer. Cancer Metastasis Rev. 23, 11-27.

Grady, W. M., Rajput, A., Myeroff, L., Liu, D. F., Kwon, K., Willis, J. and Markowitz, S. (1998). Cancer Res. 58, 3101-3104.

Grady, W. M., Willis, J., Guilford, P. J., Dunbier, A. K., Toro, T. T., Lynch, H., Wiesner, G., Ferguson, K., Eng, C., Park, J. G., Kim, S. J., Markowitz, S. (2000). Methylation of the CDH1 promoter as the second genetic hit in hereditary diffuse gastric cancer. Nat. Genet. 26, 16-17.

Grantham, R. (1974). Amino acid difference formula to help explain protein evolution. Science. 185, 862–864.

Grasberger, H. and Refetoff, S. (2006). Identification of a maturation factor for dual oxidase. J. Biol. Chem. 281, 18269–18272.

Groden, J., Thliveris, A., Samowitz, W., Carlson, M., Gelbert, L., Albertsen, H., Joslyn, G., Stevens, J., Spirio, L., Robertson, M., Sargeant, L., Krapcho, K., Wolff, E., Burt, R., Hughes, J.P., Warrington, J., McPherson, J., Wasmuth, J., Le Paslier, D., Abderrahim, H., Cohen, D., Leppert, M. and White R. (1991). Identification and characterization of the familial adenomatous polyposis coli gene. Cell. 66, 589-600.

Grohmann, A., Tanneberger, K., Alzner, A., Schneikert, J. and Behrens, J. (2007). AMER1 regulates the distribution of the tumor suppressor APC between microtubules and the plasma membrane. J. Cell Sci. 120, 3738-3747.

Groves, C., Lamlum, H., Crabtree, M., Williamson, J., Taylor, C., Bass, S., Cuthbert-Heavens, D., Hodgson, S., Phillips, R. and Tomlinson, I. (2002). Mutation cluster region, association between germline and somatic mutations and genotype-phenotype correlation in upper gastrointestinal familial adenomatous polyposis. Am J Pathol. 160, 2055-61.

Grubb, S. C., Maddatu, T. P., Bult, C. J. and Bogue, M. A. (2009). Mouse phenome database. Nucleic Acids Res. 37(Database). D720-30.

Grupe, A., Germer, S., Usuka, J., Aud, D., Belknap, J. K., Klein, R. F., Ahluwalia, M. K., Higuchi, R. and Peltz, G. (2001). In silico mapping of complex disease-related traits in mice. Science. 292, 1915-8.

Guarner, F. (1998). Prebiotics in inflammatory bowel diseases. Br J Nutr. 98, 85-9.

Gurganus, M.C., Nuzhdin, S.V., Leips, J.W. and Mackay, T.F. (1999). High-resolution mapping of quantitative trait loci for sternopleural bristle number in Drosophila melanogaster. Genetics. 152, 1585–1604

Gutierrez, L. S., Suckow, M., Lawler, J., Ploplis, V. A. and Castellino, F, J. (2003). Thrombospondin 1-A regulator of adenoma growth and carcinoma progression in the APC(Min/+) mouse model. Carcinogenesis. 24, 199-207

Hackett, J.A., Feldser,D.M., and Greider,C.W. (2001). Telomere dysfunction increases mutation rate and genomic instability. Cell. 106, 275-286.

Haines, J., Dunford, R., Moody, J., Ellender, M., Cox, R. and Silver, A. (2000). Loss of Heterozygosity in Spontaneous and X-Ray–Induced Intestinal Tumors Arising in F1 Hybrid Min Mice: Evidence for Sequential Loss of Apc1 and Dpc4 in Tumor Development. Genes Chromosomes Cancer. 28, 387–394

Haines, J., Johnson, V., Pack, K., Suraweera, N., Slijepcevic, P., Cabuy, E., Coster M., Ilyas, M., Wilding, J., Sieber, O., Bodmer, W., Tomlinson, I., and Silver A. (2005). Genetic basis of variation in adenoma multiplicity in ApcMin/+ Mom1s mice. PNAS, 102: 2868-2873,

Halberg, R. B., Katzung, D. S., Hoff, P. D., Moser, A. R., Cole, C. E., Lubet, R. A., Donehower, L. A., Jacoby, R. F. and Dove, W. F. (2000). Tumorigenesis in the multiple intestinal neoplasia mouse: redundancy of negative regulators and specificity of modifiers. Proc Natl Acad Sci U S A. 97, 3461-6.

Halfvarson, J., Bodin, L., Tysk, C., Lindberg, E. and Järnerot G. (2003). Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. Gastroenterology. 124, 1767-73.

Hall, E.J. and Wu, C.S. (2003). Radiation-induced second cancers: the impact of 3D-CRT and IMRT. Int J Radiat Oncol Biol Phys. 56,83-8.

Hall, E.J. (2006). Intensity-modulated radiation therapy, protons and the risk of second cancers. Int. J. Radiation Oncology Biol. Phys. 65, 1–7.

Halperin, E.C. (2009). Randomized prospective trials of innovative radiotherapy technology are necessary. J Am Coll Radiol. 6, 33-7.

Hamada, H., Petrino, M. G., Kakunaga, T., Seidman, M. and Stollar, B. D. (1984). Characterization of genomic poly(dT-dG) l poly(dC-dA) sequences: 'structure, organization and conformation. Mol. Cell. Biol. 4, 2610-21.

Hanahan, D. and Weinberg, R. A. (2000). The hallmarks of cancer. Cell 100, 57-70.

Hanauer, S. B. (1996). Inflammatory bowel disease. N Engl J Med. 334, 841-8.

Hanks, S., Coleman, K., Reid, S., Plaja, A., Firth, H., Fitzpatrick, D., Kidd, A., Méhes, K., Nash, R., Robin, N., Shannon, N., Tolmie, J., Swansbury, J., Irrthum, A., Douglas, J. and Rahman, N. (2004). Constitutional aneuploidy and cancer predisposition caused by biallelic mutations in BUB1B. Nat Genet. 36, 1159-61.

Harper, J. W. and Elledge, S. J. (2007) The DNA damage response: ten years after. Mol Cell. 28, 739-45.

Harvey, J. J. (1964). An unidentified virus which causes the rapid production of tumours in mice. Nature. 204, 1104–5.

Hashiguchi, K., Matsumoto, Y., Yasui, A. (2007) Recruitment of DNA repair synthesis machinery to sites of DNA damage/repair in living human cells. Nucleic Acids Res. 35, 2913–23.

Haston, C.K., Zhou, X., Gumbiner-Russo, L., Irani, R., Dejournett, R., Gu, X., Weil, M., Amos, C.I. and Travis, E.L. (2002). Universal and Radiation-specific Loci Influence Murine Susceptibility to Radiation-induced Pulmonary Fibrosis Cancer Res. 62:3782-3788.

He, X. C., Zhang, J., Tong, W. G., Tawfik, O., Ross, J., Scoville, D. H., Tian, Q., Zeng, X., He, X., Wiedemann, L. M., Mishina Y. and Li, L. (2004). BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt-beta-catenin signaling. Nat Genet. 36, 1117-21.

Hedrick, L., Cho, K.R., Fearon, E.R., Wu,T.C., Kinzler, K.W. and Vogelstein, B. (1994). The DCC gene product in cellular differentiation and colorectal tumorigenesis. Genes Dev. 8, 1174-83.

Heldin, C.H., Miyazono, K. and Dijke, P. (1997). TGF-beta signalling from cell membrane to nucleus through SMAD proteins. Nature. 390, 465-71.

Helentjaris, T., Weber, D. F., Wright, S. (1986). Use of monosomics to map cloned DNA fragments in maize. Proc Natl Acad Sci U S A. 83, 6035–6039.

Hemminki, A., Markie, D., Tomlinson, I., Avizienyte, E., Roth, S., Loukola, A., Bignell, G., Warren, W., Aminoff, M., Höglund P., Järvinen, H., Kristo, P., Pelin, K., Ridanpää, M., Salovaara, R., Toro, T., Bodmer, W., Olschwang, S., Olsen, A.S., Stratton, M.R., de la Chapelle, A. and Aaltonen, L.A. (1998). A serine/threonine kinase gene defective in Peutz-Jeghers syndrome.Nature. 391, 184-7.

Herman, J. G. and Baylin, S. B. (2003). Gene silencing in cancer in association with promoter hypermethylation. N. Engl. J. Med. 349, 2042-2054.

Hernandez-Flores, G., Gomez-Contreras, P. C., Dominguez-Rodriguez, J.R., Lerma-Diaz, J.M., Ortiz-Lazareno P.C., Cervantes-Munguia, R., Sahagun-Flores, J.E., Orbach-Arbouys, S., Scott-Algara, D. and Bravo-Cuellar, A. (2005). Gamma-irradiation induced apoptosis in peritoneal macrophages by oxidative stress. Implications of antioxidants in caspase mitochondrial pathway. Anticancer Res. 25, 4091–4100.

Herrera, L., Kakati, S., Gibas, L., Pietrzak, E. and Sandberg, A.A. (1986). Gardner syndrome in a man with an interstitial deletion of 5q. Am J Med Genet. 25, 473–476

Herszényi L and Tulassay Z. (2010). Epidemiology of gastrointestinal and liver tumors. Eur Rev Med Pharmacol Sci. 14, 249-58.

Herszényi, L., Miheller, P. and Tulassay, Z. (2007). Carcinogenesis in inflammatory bowel disease. Dig Dis. 25, 267-9.

Heyen, D. W., Weller, J. I., Ron, M., Band, M., Beever, J. E., Feldmesser, E., Da, Y., Wiggans, G. R., VanRaden, P. M., and Lewin, H, A. (1999). A genome scan for QTL influencing milk production and health traits in dairy cattle. Physiol Genomics 11, 165-75.

Hirschhorn, J. N. and Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. 6, 95–108.

Hoeijmakers, J.H. (2001). Genome maintenance mechanisms for preventing cancer. Nature. 411, 366-74.

Hoffmeister, M., Raum, E., Winter, J., Chang-Claude, J. and Brenner, H. (2007). Hormone replacement therapy, body mass, and the risk of colorectal cancer among postmenopausal women from Germany. Br J Cancer. 97, 1486–1492.

Hojgaard. A., et al. (2006). Altered localization of CXCL13 expressing cells in mice deficient in Pactolus following an inflammatory stimulus. Immunology. 119, 212–223.

Hollstein, M., Sidransky, D., Vogelstein, B. and Harris, C. C. (1991). p53 mutations in human cancers. Science 253, 49-53.

Hoste, C., Dumont, J. E., Miot, F. and De Deken, X. (2012). The type of DUOX-dependent ROS production is dictated by defined sequences in DUOXA Exp Cell Res. 318, 2353-64.

Houlston, R. S, Cheadle. J, Dobbins. S. E., Tenesa, A., Jones, A. M., Howarth, K., Spain, S. L., Broderick, P., Domingo, E., Farrington,S., Prendergast, J. G, Pittman, A. M., Theodoratou, E., Smith, C. G., Olver, B., Walther, A., Barnetson, R. A., Churchman, M, Jaeger, E. E., Penegar, S., Barclay, E., Martin, L., Gorman, M., Mager, R., Johnstone, E., Midgley, R., Niittymäki. I., Tuupanen, S., Colley, J., Idziaszczyk, S., COGENT Consortium, Thomas, H. J., Lucassen, A. M., Evans, D. G., Maher, E. R., CORGI Consortium, COIN Collaborative Group, Maughan. T., Dimas. A., Dermitzakis. E.,

Cazier, J. B., Aaltonen, L. A., Pharoah, P., Kerr, D. J, Carvajal-Carmona, L. G., Campbell, H., Dunlop, M. G., Tomlinson, I. P. (2010). Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. Nat Genet.42, 973-7.

Houlston, R. S., Webb, E., Broderick, P., Pittman, A. M., Di Bernardo, M. C., Lubbe, S., Chandler, I., Vijayakrishnan, J., Sullivan, K., Penegar, S.; Colorectal Cancer Association Study Consortium, Carvajal-Carmona L., Howarth, K., Jaeger, E., Spain, S. L., Walther, A., Barclay, E., Martin, L., Gorman, M., Domingo, E., Teixeira, A. S.; CoRGI Consortium, Kerr, D., Cazier, J. B., Niittymäki, I., Tuupanen, S., Karhu, A., Aaltonen, L. A., Tomlinson, I. P., Farrington, S. M., Tenesa, A., Prendergast, J. G., Barnetson, R. A., Cetnarskyj, R., Porteous, M. E., Pharoah, P. D., Koessler, T., Hampe, J., Buch, S., Schafmayer, C., Tepel, J., Schreiber, S., Völzke, H., Chang-Claude, J., Hoffmeister, M., Brenner, H., Zanke, B. W., Montpetit, A., Hudson, T. J., Gallinger, S.; International Colorectal Cancer Genetic Association Consortium, Campbell, H., Dunlop, M. G., (2008). Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet. 40, 1426-35.

Houlston, R. S. and Tomlinson, I. P. (2001). Polymorphisms and colorectal tumor risk. Gastroenterology. 121, 282-301.

Howe, J. R., Roth, S., Ringold, J. C., Summers, R. W., Järvinen, H. J., Sistonen, P., Tomlinson, I. P., Houlston, R. S., Bevan, S., Mitros, F. A., Stone, E.M. and Aaltonen, L. A. (1998). Mutations in the SMAD4/DPC4 gene in juvenile polyposis. Science. 280, 1086-8.

Howe, J. R., Sayed, M. G., Ahmed, A. F., Ringold, J., Larsen-Haidle, J., Merg, A., Mitros, F. A., Vaccaro, C.A., Petersen, G. M., Giardiello, F. M., Tinley, S. T., Aaltonen, L. A. and Lynch, H. T. (2004). The prevalence of MADH4 and BMPR1A mutations in juvenile polyposis and absence of BMPR2, BMPR1B, and ACVR1 mutations. J Med Genet. 41, 484-91.

Howlader, N., Ries, L. A., Stinchcomb, D. G. and Edwards, B. K. (2009). The impact of underreported Veterans Affairs data on national cancer statistics: analysis using population-based SEER registries. J Natl Cancer Inst. 101, 533-6.

Hua, G., Thin, T. H., Feldman, R., Haimovitz-Friedman, A., Clevers, H., Fuks, Z.and Kolesnick, R. (2012). Crypt base columnar stem cells in small intestines of mice are radioresistant. Gastroenterology 143, 1266-76

Hubner, R. A. and Houlston, R. S. (2007). MTHFR C677T and colorectal cancer risk: A meta-analysis of 25 populations. Int J Cancer. 120, 1027-35.

Hugh, T. J., Dillon, S. A., O'Dowd, G., Getty, B., Pignatelli, M., Poston, G. J. and Kinsella, A. R. (1999). beta-catenin expression in primary and metastatic colorectal carcinoma. Int. J. Cancer 82, 504-511.

Itzkowitz, S. H. and Harpaz N. (2004) Diagnosis and management of dysplasia in patients with inflammatory bowel diseases. Gastroenterology. 126, 1634-48.

Jannink, J. L., and Jansen, R. (2001). Mapping epistatic quantitative trait loci with one-dimensional genome searches. Genetics. 2001 157, 445-54.

Jass, J. R. (2007). Hereditary and DNA methylation in colorectal cancer. Gut. 56, 154-155.

Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J.and Thun, M. J. (2009). Cancer statistics. CA Cancer J Clin. 59,225-49.

Jenne, D.E., Reimann, H., Nezu, J., Friedel, W., Loff, S., Jeschke, R., Müller, O., Back, W. and Zimmer, M. (1998). Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. Nat Genet. 18, 38-43.

Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P., von Mering, C. (2009). STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 37 (Database issue), D412-6. .

Jiang, C. and Zeng, Z. B. (1995). Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics. 140, 1111-27.

Jo, W. S., Mizukami, Y., Duerr, E. M., Zukerberg, L. R., Chung, D. C. (2005). Wnt signaling can repress thrombospondin-1 expression in colonic tumorigenesis. Cancer Biol Ther. 4, 1361-6.

Johns, L. E. and Houlston, R. S. (2001) A systematic review and meta-analysis of familial colorectal cancer risk. Am J Gastroenterol 96, 2992-3003,

Johnson, G. C., and J. A. Todd. (2000). Strategies in complex disease mapping. Curr Opin Genet Dev 10, 330.

Johnson, V., Volikos, E., Halford, S. E., Eftekhar Sadat, E. T., Popat, S., Talbot, I., Truninger, K., Martin, J., Jass, J., Houlston, R., Atkin, W., Tomlinson, I. P. and Silver, A.R. (2005). Exon 3 beta-catenin mutations are specifically associated with colorectal carcinomas in hereditary non-polyposis colorectal cancer syndrome. Gut. 54, 264–267.

Jones, D., Perttunen, C. and Stuckman, B. (1993) Lipschitzian optimization without the Lipschitz constant. J. Optim. Theory App., 79, 157–181.

Jones, P. A. and Baylin, S. B. (2007). The Epigenomics of Cancer. Cell, 128, 683-692.

Jubb, A. M., Turley, H., Moeller, H. C., Steers, G., Han, C., Li, J. L., Leek, R., Tan, E.Y., Singh, B., Mortensen, N. J., Noguera-Troise, I., Pezzella, F., Gatter, K. C., Thurston, G., Fox, J., Harris, A. L. (2009). Expression of delta-like ligand 4 (Dll4) and markers of hypoxia in colon cancer. Br J Cancer. 101, 1749–57.

Julié, C., Trésallet, C., Brouquet, A., Vallot, C., Zimmermann, U., Mitry, E., Radvanyi, F., Rouleau, E., Lidereau, R., Coulet, F., Olschwang, S., Frébourg, T., Rougier, P., Nordlinger, B., Laurent-Puig, P., Penna, C., Boileau, C., Franc, B., Muti, C. and Hofmann-Radvanyi, H. (2008). Identification in daily practice of patients with Lynch syndrome (hereditary nonpolyposis colorectal cancer): revised Bethesda guidelines-based approach versus molecular screening. Am J Gastroenterol. 103, 2825-35.

Kaklamani, V.G., Hou, N., Bian, Y., Reich, J., Offit, K., Michel, L. S., Rubinstein, W. S., Rademaker, A., Pasche, B. (2003). TGFBR1*6A and cancer risk: a meta-analysis of seven case-control studies. J Clin Oncol. 21, 3236-43.

Kao, C. H., Zeng, Z. B. and Teasdale, R. D. (1999). Multiple interval mapping for quantitative trait loci. Genetics, 152, 1203–1216.

Kao, C. H., Zeng, Z. B. (2002). Modelling epistasis of quantitative trait loci using Cockerham's model. Genetics. 160, 1243-61.

Katoh, M. and Katoh, M., (2006) Notch ligand, JAG1, is evolutionarily conserved target of canonical WNT signaling pathway in progenitor cells.Int J Mol Med. 17, 681-5.

Kawasaki, Y., Tsuji, S., Muroya, K., Furukawa, S., Shibata, Y., Okuno, M., Ohwada, S. and Akiyama, T. (2009). The adenomatous polyposis coli-associated exchange factors Asef and Asef2 are required for adenoma formation in Apc(Min/+)mice. EMBO Rep. 10, 1355-62.

Kendal, W. S. and Nicholas, G. A, (2007). Population-based analysis of second primary cancers after irradiation for rectal cancer. Am J Clin Oncol. 30, 333-9.

Kennedy, B. P., Payette, P., Mudgett, J., Vadas, P., Pruzanski, W., Kwan, M., Tang, C., Rancourt, D. E.and Cromlish W. A. (1995). A natural disruption of the secretory group II phospholipase A2 gene in inbred mouse strains. J Biol Chem. 270, 22378-85.

Kennedy, D., and Heise, CP. (2007). Radiation Colitis and Proctitis Clin Colon Rectal Surg. 20, 64–72.

Kerber, R. A., Amos, C. I., Yeap, B. W., Finkelstein, D. F., Thomas, D. C. (2008). Design considerations in a sib-pair study of linkage for susceptibility loci in cancer. BMC Med Genet, 9, 64

Kim, Y. I. and Mason, J. B. (1996). Nutrition chemoprevention of gastrointestinal cancers: a critical review.Nutr Rev. 54, 259-79.

Kinzler, K. and Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. Cell 87, 159-170.

Kinzler, K. W. and Vogelstein, B. (1997) Cancer-susceptibility genes. Gatekeepers and caretakers. Nature. 24, 761, 763.

Kinzler, K. W., Nilbert, M. C., Su, L. K., Vogelstein, B., Bryan, T. M., Levy, D. B., Smith, K. J., Preisinger, A. C., Hedge, P., and McKechnie, D. (1991). Identification of FAP locus genes from chromosome 5q21. Science. 9, 661-5.

Kirsten, W. H., Schauf, V. and McCoy, J. (1970). Properties of a murine sarcoma virus. Bibl Haematol. 36, 246–9.

Kleinsmith, L. (2006). Principles of Cancer Biology. Pearson Benjamin Cummings.

Knott, S. A. and Haley, C.S. (2000) Multitrait least squares for quantitative trait loci detection. Genetics. 156, 899-911.

Knudson, A. G., Jr. (1971). Mutation and cancer, statistical study of retinoblastoma. Proc. Natl. Acad. Sci. U S A. 68, 820-823.

Knudson, A. G., Jr. (1985). Hereditary cancer, oncogenes, and antioncogenes. Cancer Res. 45, 1437-43

Kondo, Y. and Issa, J., P. (2004). Epigenetic changes in colorectal cancer. Cancer Metastasis Rev. 23, 29-39.

Korol, A. B., Ronin, Y. I., Nevo, E. and Hayes, P. (1998). Multi-interval mapping of correlated trait complexes, simulation analysis and evidence from barley. Heredity 80, 273–284.

Korstanje, R. and Paigen, B. (2002). From QTL to gene, The harvest begins. Nature Genet. 31, 235–236.

Kosambi, D.D. (1944). The estimation of map distance from recombination values, Annals of Eugenics, 12, 172–175

Kosinski C., Li V. S. W., Chan, A. S. Y., Zhang, J., Ho C, Tsui, W. Y., (2007). Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors. Proceedings of the National Academy of Sciences. 104, 15418-23.

Koyama, M., Ito, M., Nagai, H., Emi, M., Moriyama, Y. (1999). Inactivation of both alleles of the DPC4/SMAD4 gene in advanced colorectal cancers, Identification of

seven novel somatic mutations in tumors from Japanese patients. Mutat Res. 406, 71–77.

Kruglyak, L. and Lander, E. S. (1995). Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am J Hum Genet. 57, 439–454.

Ku, J.L., Park, S. H., Yoon, K. A., Shin, Y. K., Kim, K. H., Choi, J. S., Kang, H. C., Kim, I. J., Han, I.O.and Park, J., G. (2007). Genetic alterations of the TGFbeta signaling pathway in colorectal cancer cell lines, a novel mutation in Smad3 associated with the inactivation of TGFbeta- induced transcriptional activation. Cancer Lett. 247, 283-292

Kumar, P., Henikoff, S.and Ng, P. C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 4, 1073-81

Kuraguchi, ,M., Wang, X. P., Bronson, R. T., Rothenberg, R., Ohene-Baah, N. Y., Lund, J. J., Kucherlapati, M., Maas, R. L., Kucherlapati, R. (2006). Adenomatous polyposis coli (APC) is required for normal development of skin and thymus, PLoS Genet. 2

Kwabi-Addo, B., Giri, D., Schmidt, K., Podsypanina, K., Parsons, R., Greenberg, N., Ittmann, M. (2001). Haploinsufficiency of the Pten tumor suppressor gene promotes prostate cancer progression. Proc Natl Acad Sci U S A. 98, 11563-8.

Kwong, L. N., Shedlovsky, A., Biehl, B. S., Clipson, L., Pasch, C. A. and Dove, W. F. (2007). Identification of Mom7, a novel modifier of Apc(Min/ +) on mouse chromosome 18. Genetics. 176, 1237–1244.

Kwong, J., Lo, K. W., Chow, L. S., Chan, F. L., To, K. F. and Huang, D. P. (2005). Silencing of the retinoid response gene TIG1 by promoter hypermethylation in nasopharyngeal carcinoma. Int J Cancer. 113, 386-92.

Lai, C., Robinson, J., Clark, S., Stamp, G., Poulsom, R., and Silver, A. (2011). Elevation of WNT5A expression in polyp formation in Lkb1+/- mice and Peutz-Jeghers syndrome. J. Pathol. 223, 584-592.

Laken, S. J., Petersen, G. M., Gruber, S. B., Oddoux, C., Ostrer, H., Giardiello, F. M., Hamilton, S. R., Hampel, H., Markowitz, A., Klimstra, D., Jhanwar, S., Winawer, S., Offit, K., Luce, M.C., Kinzler, K. W., Vogelstein, B. (1997). Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. Nat Genet. 17, 79-83.

Lamlum, H., Ilyas, M., Rowan, A., Clark, S., Johnson, V., Bell, J., Frayling, I., Efstathiou, J., Pack, K., Payne, S., Roylance, R., Gorman, P., Sheer, D., Neale, K., Phillips, R., Talbot, I., Bodmer, W., and Tomlinson, I. (1999). The type of somatic mutation at APC in familial adenomatous polyposis is determined by the site of the germline mutation, a new facet to Knudson's 'two-hit' hypothesis. Nat Med. 5, 1071-5.

Lander, E. S, and Botstein, D. (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 121, 185-199

Lander, E. S., Linton, L. M. et al (254 co-authors). Initial sequencing and analysis of the human genome. Nature 409, 860-921 (2001).

Lander, E., Kruglyak, L. (1995). Genetic dissection of complex traits, Guidelines for interpreting and reporting linkage results. Nat Genet. 11, 241–247

Lashner, B. A. (1994). Genetic heterogeneity in ulcerative colitis. Am J Gastroenterol. 89, 1927-8.

Leach, F.S., Nicolaides, N.C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., Peltomäki, P., Sistonen, P., Aaltonen, L. A., Nyström-Lahti, M. (1993). Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. Cell. 75, 1215-25.

Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1997). Genetic instability in colorectal cancers. Nature 386, 623–627.

Lengauer, C., Kinzler, K. W., and Vogelstein, B. (1998). Genetic instabilities in human cancers. Nature 396, 643-649.

Leppert, M., Burt, R., Hughes, J. P., Samowitz, W., Nakamura, Y., Woodward, S., Gardner, E., Lalouel, J. M., White, R. (1990). Genetic analysis of an inherited predisposition to colon cancer in a family with a variable number of adenomatous polyps. N Engl J Med. 29, 904-8.

Leslie, A., Pratt, N. R., Gillespie, K., Sales, M., Kernohan, N. M., Smith, G., Wolf, C. R., Carey, F. A., Steele, R. J. (2003). Mutations of APC, K-ras, and p53 are associated with specific chromosomal aberrations in colorectal adenocarcinomas. Cancer Res. 63, 4656–61

Levitt, N.C., Hickson, I.D. (2002). Caretaker tumour suppressor genes that defend genome integrity. Trends Mol Med. 8,179-86.

Li, Q., Ishikawa, T. O., Oshima, M., Taketo, M. M. (2005). The threshold level of adenomatous polyposis coli protein for mouse intestinal tumorigenesis, Cancer Res. 65, 8622–8627.

Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A., Hemminki, K. (2000). Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med, 343, 78-85

Lipton L., Tomlinson I. (2004). The multiple colorectal adenoma phenotype and MYH, a base excision repair gene. Clin. Gastroenterol. Hepatol. 2, 633–638

Lipton, L., and Tomlinson, I. (2006). The genetics of FAP and FAP-like syndromes. Fam Cancer 5, 221-6.

Liu, L., Wu, D. H., Ding, Y. Q. (2005). Tiam1 gene expression and its significance in colorectal carcinoma.. World J Gastroenterol. 7, 705-7.

Livak, K., and Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods. 25, 402-08.

Ljungberg, K., Holmgren, S., Carlborg, O. (2004). Simultaneous search for multiple QTL using the global optimization algorithm DIRECT. Bioinformatics, 20, 1887–1895.

Lockhart-Mummery, J. P. (1934). The causation and treatment of multiple adenomatosis of the colon. Ann. Surg. 99, 178-184.

Logan, C. Y., Nusse, R. (2004). The Wnt signaling pathway in development and disease. Annu. Rev. Cell Dev. Biol. 20, 781–810.

Logan, C., Y., and Nusse, R. (2004). The Wnt signalling pathway in development and disease. Annu Rev Cell Dev Biol 20, 7812-810.

Loo, L. W., Cheng, I., Tiirikainen, M., Lum-Jonesm A., Seifried, A., Dunklee, L. M., Church, J. M., Gryfe, R., Weisenberger, D. J., Haile, R. W., Gallinger, S., Duggan, D. J., Thibodeau, S. N., Casey, G., Le Marchand, L. (2012). cis-Expression QTL analysis of established colorectal cancer risk variants in colon tumors and adjacent normal tissue. PLoS One. 7(2), e30477.

Lu, S., Shen, K., Wang, Y., Santner, S. J., Chen, J., Brooks, S. C., Wang, Y. A. (2006). Atm-haploinsufficiency enhances susceptibility to carcinogen-induced mammary tumors. Carcinogenesis. 2006 27(4),848-55.

Lugli, A., Iezzi, G., Hostettler, I., Muraro, M., G., Mele, V., Tornillo, L., (2010). Prognostic impact of the expression of putative cancer stem cell markers CD133, CD166, CD44s, EpCAM, and ALDH1 in colorectal cancer. British Journal of Cancer. 103, 382-90.

Luongo, C., Moser, A. R., Gledhill, S., Dove, W. F., (1994). Loss of Apc in intestinal adenomas from Min mice. Cancer Res 54, 5947–5952.

Luxen, S., Belinsky, S., A, Knaus, U.,G. (2008). Silencing of DUOX NADPH Oxidases by Promoter Hypermethylation in Lung Cancer. Cancer Research. 68, 1037-45

Lynch, H. T., de la Chapelle. (2003). A. Hereditary colorectal cancer. N Engl J Med. 6, 919-32.

Lynch, H. T., Krush, A., J., (1971). Cancer family "G" revisited, 1895-1970. Cancer. 27, 1505-11.

Lynch, H. T., Smyrk, T., McGinn, T., Lanspa, S., Cavalieri, J., Lynch, J., Slominski-Castor, S., Cayouette, M. C., Priluck, I., Luce, M. C. (1995). Attenuated familial adenomatous polyposis (AFAP). A phenotypically and genotypically distinctive variant of FAP. Cancer. 15, 2427-33.

MacPhee, M., Chepenik, K. P., Liddell, R. A., Nelson, K. K., Siracusa, L. D., Buchberg, A. M. (1995). The secretory phospholipase A2 gene is a candidate for the Mom1 locus a major modifier of ApcMin-induced intestinal neoplasia, Cell 81, 957–966.

Maddams, J., Brewster, D., Gavin, A., Steward, J., Elliott, J., Utley, M., Møller, H. (2009). Cancer prevalence in the UK, estimates for 2008. Br J Cancer. 101,541-7

Maddams, J., Parkin, D., Darby SC. (2011). The cancer burden in the United Kingdom in 2007 due to radiotherapy. 129, 2885–2893.

Maddams, J., Utley, M., Møller, H., (2012). Projections of cancer prevalence in the United Kingdom, 2010-2040. Br J Cancer. 25, 1195-202

Malliri, A., Rygiel, T. P., van der Kammen, R. A., Song, J. Y., Engers, R., Hurlstone, A. F., Clevers, H., and Collard, J. G. (2006). The rac activator Tiam1 is a Wnt-responsive gene that modifies intestinal tumor development. J Biol Chem. 281, 543-8

Manichaikul, A., Moon, J. Y., Sen, S., Yandell, B. S., Broman, K. W. (2009). A model selection approach for the identification of quantitative trait loci in experimental crosses, allowing epistasis. Genetics. 181, 1077-86.

Manly, K. F., Cudmore, Jr., R. H., and Meer, J. M. (2001). MapManager QTX, cross-platform software for genetic mapping. Mamm. Genome, 12, 930-932.

Mann, B., Gelos, M., Siedow, A., Hanski, M. L., Gratchev, A., Ilyas, M., Bodmer, W. F., Moyer, M.P., Riecken, E. O., Buhr, H. J.and Hanski,C. (1999). Target genes of beta-catenin-T cell-factor/lymphoid-enhancer-factor signaling in human colorectal carcinomas. Proc. Natl. Acad. Sci. U S A. 96, 1603-1608.

Marchini, J., Donnelly, P., Cardon, L. R. (2005). Genome-wide strategies for detecting multiple loci that influence complex diseases. Nat Genet. 37, 413–417.

Margel, D., Baniel, J., Wasserberg, N., Bar-Chana, M., Yossepowitch, O. (2011). Radiation therapy for prostate cancer increases the risk of subsequent rectal cancer. Ann Surg. 254, 947-50

Markowitz, S. D. (2007). Aspirin and colon cancer--targeting prevention? N. Engl. J. Med. 356, 2195-2198.

Marlow, A. J., Fisher, S. E., Francks, C., MacPhie, I. L., Cherny, S., S, Richardson, A. J., Talcott, J. B., Stein, J. F., Monaco. A. P., Cardon, L. R. (2003). Use of multivariate linkage analysis for dissection of a complex cognitive trait. Am J Hum Genet. 72, 561-70

Martinez, P., Birkbak, N.J., Gerlinger, M., McGranahan, N., Burrell, R.A., Rowan, A.J., Joshi, T., Fisher, R., Larkin, J., Szallasi, Z., Swanton, C. N. (2013). Parallel evolution of tumour subclones mimics diversity between tumours. J Pathol. 230, 356-64.

Matsunami, T., Sato, Y., Hasegawa, Y., Ariga, S., Kashimura, H., Sato, T., Yukawa, M. (2011). Enhancement of reactive oxygen species and induction of apoptosis in streptozotocin-induced diabetic rats under hyperbaric oxygen exposure.Int J Clin Exp Pathol. 4, 255-66.

May, P., May, E. (1999). Twenty years of p53 research, structural and functional aspects of the p53 protein. Oncogene. 18, 7621-36.

McCart, A. E., Vickaryous, N. K., Silver, A. (2008). Apc mice, models, modifiers and mutants. Pathol Res Pract. 204, 479-90.

McPeek, M. S. (2000). From mouse to human, Fine mapping of quantitative trait loci in a model organism Proc Natl Acad Sci U S A. 97, 12389–12390.

Michor, F. (2004). Dynamics of Cancer progression. Nature Reviews Cancer. 4, 197-205.

Midgley, C. A., White, S., Howitt, R., Save, V., Dunlop, M. G., Hall, P. A., Lane, D. P, Wyllie, A. H., Bubb, V. J. (1997). APC expression in normal human tissues. J Pathol. 181,426-33.

Miyaki, M., Iijima,T., Konishi, M., Sakai, K., Ishii, A., Yasuno, M., Hishima, T., Koike, M., Shitara, N., Iwama, T., Utsunomiya, J., Kuroki, T., Mori, T.(1999). Higher frequency of Smad4 gene mutation in human colorectal cancer with distant metastasis. Oncogene. 18, 3098–3103.

Miyaki, M., Konishi, M., Kikuchi-Yanoshita, R., Enomoto, M., Igari, T., Tanaka, K., Muraoka, M., Takahashi, H., Amada, Y., Fukayama, M. (1994). Characteristics of somatic mutation of the adenomatous polyposis coli gene in colorectal tumors. Cancer Res. 54, 3011–3020

Miyoshi, Y., Ando, H., Nagase, H., (1992). Germ-line mutations of the APC gene in 53 familial adenomatous polyposis patients. Proc Natl Acad Sci U S A. 89, 4452-6.

Mohrenweiser, H. W., Wilson, D. M., Jones, I. M., (2003). Challenges and complexities in estimating both the functional impact and the disease risk associated with the extensive genetic variation in human DNA repair genes. Mutat Res. 15, 93-125.

Mori, N., Matsumoto, Y., Okumoto, M., Suzuki, N., Yamate, J. (2001). Variations in Prkdc encoding the catalytic subunit of DNA-dependent protein kinase (DNAPKcs) and susceptibility to radiation-induced apoptosis and lymphomagenesis. Oncogene 20, 3609–3619.

Morton, N. M. (1955). Sequential tests for the detection of linkage Am J Hum Genet. 7, 277–318.

Moser, A.R., Pitot, H. C. and Dove, W. F. (1990). A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. Science. 247, 322–324

Mott, R., Talbot, C. J., Turri, M. G., Collins A. C. and Flint J. (2000). A new method for fine-mapping quantitative trait loci in outbred animal stocks Proc Natl Acad Sci U S A, 97, 12649-12654

Muleris, M., Chalastanis, A., Meyer, N., Lae, M., Dutrillaux, B., Sastre-Garau, X., Hamelin, R., Flejou, J. F. and Duval, A. (2008). Chromosomal instability in near-diploid colorectal cancer, a link between numbers and structure. PLoS. ONE. 3, e1632.

Munkholm, P. (2003). Review article, the incidence and prevalence of colorectal cancer in inflammatory bowel disease. Aliment Pharmacol Ther. 18, 1-5.

Murty,V. V., Li, R. G., Houldsworth, J., Bronson, D. L., Reuter, V. E., Bosl, G. J., Chaganti, R. S. (1994). Frequent allelic deletions and loss of expression characterize the DCC gene in male germ cell tumors. Oncogene. 9, 3227-31.

Myohanen, S. K., Baylin, S. B. and Herman, J. G. (1998). Hypermethylation can selectively silence individual p16ink4A alleles in neoplasia. Cancer Res. 58, 591-593.

Nadeau, J. H. and Frankel, W.N. (2000). The roads from phenotypic variation to gene discovery, Mutagenesis versus QTLs. Nat. Genet. 25, 381-384

Nagase H, Miyoshi Y, Horii A, Aoki T, Ogawa M, Utsunomiya J, Baba S, Sasazuki T, Nakamura Y. (1992). Correlation between the location of germ-line mutations in the APC gene and the number of colorectal polyps in familial adenomatous polyposis patients. Cancer Res. 52,4055–7.

Nagase, H. and Nakamura, Y. (1993). Mutations of the APC (adenomatous polyposis coli) gene. Hum Mutat. 2, 425-34.

Nagel, R. L. (2001). Pleiotropic and epistatic effects in sickle cell anemia. Curr Opin Hematol. 8, 105-10.

Nagel, R. L. (2005). Epistasis and the genetics of human diseases. C R Biol. 328, 606-15.

Nagel, R. L., Steinberg, M. H. (2001). Role of epistatic (modifier) genes in the modulation of the phenotypic diversity of sickle cell anemia. Pediatr Pathol Mol Med. 20, 123-36.

Nakau, M., Miyoshi, H., Seldin, M. F., Imamura, M., Oshima, M., Taketo, M. M. (2002). Hepatocellular carcinoma caused by loss of heterozygosity in Lkb1 gene knockout mice. Cancer Res. 15, 4549-53.

Neklason, D. W., Solomon, C. H., Dalton, A. L., Kuwada, S. K., Burt, R. W. (2004). Intron 4 mutation in APC gene results in splice defect and attenuated FAP phenotype. Fam Cancer. 3, 35-40

Newton, K. F., Mallinson, E. K., Bowen, J., Lalloo, F., Clancy, T., Hill, J., Evans, D. G. (2012). Genotype-phenotype correlation in colorectal polyposis. Clin Genet. 81, 521-31.

Ng, P. C. and Henikoff, S. (2006). Predicting the Effects of Amino Acid Substitutions on Protein Function Annu Rev Genomics Hum Genet. 7, 61-80.

Ng, P. C., 2, S. (2003). SIFT, predicting amino acid changes that affect protein function. Nucleic Acids Res.31, 3812-4

Ng, P. C. and Henikoff, S. (2002). Accounting for human polymorphisms predicted to affect protein function. Genome Res. 12, 436–46

Nica, A. C., Montgomery, S. B., Dimas, A. S., Stranger, B. E., Beazley, C., Barroso, I., Dermitzakis, E. T. (2010). Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. PLoS Genet. 2010 Apr 1,6(4), e1000895.

Niho, N., Takahashi, M., Kitamura, T., Shoji, Y., Itoh, M., Noda, T., Sugimura, T., Wakabayashi, K. (2003). Concomitant suppression of hyperlipidemia and intestinal polyp formation in Apc-deficient mice by peroxisome proliferator-activated receptor ligands. Cancer Res. 63, 6090-6095

Nishisho, I., Nakamura, Y., Miyoshi,Y., Miki, Y., Ando, H., Horii, A., Koyama, K., Utsunomiya, J., Baba, S., (1991). Hedge P. Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. Science. 253, 665-9.

Nugent, K. P., Phillips, R. K., Hodgson, S. V., Cottrell, S., Smith-Ravin, J., Pack, K. and Bodmer, W., F. (1994). Phenotypic expression in familial adenomatous polyposis, partial prediction by mutation analysis. Gut 35, 1622-3.

Orholm, M., Binder, V., Sørensen, T. I., Rasmussen, L. P., Kyvik, K. O. (2000). Concordance of inflammatory bowel disease among Danish twins. Results of a nationwide study. Scand J Gastroenterol. 35, 1075-81.

Orholm, M., Munkholm, P., Langholz, E., Nielsen, O. H., Sørensen, T. I., Binder, V. (1991). Familial occurrence of inflammatory bowel disease. N Engl J Med. 324, 84-8.

Oshima, M., Dinchuk, E. J., Kargman, S. L., Oshima, H., Hancock, B., Kwong, E., Trzaskos, J., Evans, J., Taketo, M. (1996). Suppression of intestinal polyposis in Apc (Δ716) knockout mice by inhibition of cyclooxygenase 2 (COX-2) Cell. 87, 803-809.

Ott, J. (1989). Statistical properties of the haplotype relative risk. Genet. Epidemiol. 6, 127–130.

Parada, L. F., Tabin, C. J., Shih, C. and Weinberg, R.A. (1982). Human EJ bladder carcinoma oncogene is homologue of Harvey sarcoma virus ras gene. Nature. 10, 474-8.

Parkin, D. M., Bray, F., Ferlay, J., Pisani., P (2005). Global Cancer Statistics., CA Cancer J Clin. 55, 74-108

Pasche, B., Kaklamani, V., Hou, N., Young, T., Rademaker, A., Peterlongo, P., Ellis, N., Offit, K., Caldes, T., Reiss, M., Zheng, T. (2004). TGFBR1*6A and cancer, a meta-analysis of 12 case-control studies. J Clin Oncol. 15, 756-8.

Pasche, B., Luo, Y., Rao, P. H,, Nimer, S. D., Dmitrovsky, E., Caron, P., Luzzatto, L., Offit, K., Cordon-Cardo, C., Renault, B., Satagopan, J. M, Murty, V. V., Massagué, J. (1998). Type I transforming growth factor beta receptor maps to 9q22 and exhibits a polymorphism and a rare variant within a polyalanine tract. Cancer Res. 58, 2727-32.

Pataer, A., Nishimura, M., Kamoto, T., Ichioka, K., Sato, M., Hiai H., (1997). Genetic resistance to urethan-induced pulmonary adenomas in SMXA recombinant inbred mouse strains. Cancer Research; 57, 2904-8

Payne, S. R. and Kemp, C. J. (2005). Tumor suppressor genetics. Carcinogenesis 26, 2031-2045.

Pearce, M. S., Salotti, J. A., Howe, N. L., McHugh, K., Kim, K. P., Lee, C., Craft, A. W., Berrington de Gonzaléz, A., Parker, L. (2012). CT Scans in Young People in Great Britain, Temporal and Descriptive Patterns, 1993-2002.

Pearson, K. (1990).The Grammar of science. Cousin London. Adam and Charles Black

Peltomäki, P. (2001). DNA mismatch repair and cancer. Mutat Res. 488, 77-85.

Petitjean, A., Achatz, M. I., Borresen-Dale, A. L., Hainaut, P. and Olivier, M. (2007). TP53 mutations in human cancers, functional selection and impact on cancer prognosis and outcomes. Oncogene 26, 2157-2165.

Peto, J., Houlston, R. S. (2001). Genetics and the common cancers. Eur J Cancer. 37, 88-96.

Pharoah, P. D., Antoniou, A., Bobrow, M., Zimmern, R. L., Easton, D. F., Ponder, B. A. (2002). Polygenic susceptibility to breast cancer and implications for prevention. Nat Genet. 31, 33-6.

Phelps, R. A., Chidester, S., Dehghanizadeh, S., Phelps, J., Sandoval, I. T., Rai, K., Broadbent, T., Sarkar, S., Burt, R. W., and Jones, D. A. (2009). A two-step model for colon adenoma initiation and progression caused by APC loss. Cell 137, 623-634.

Pietenpol, J. A. and Vogelstein, B. (1993). Tumour suppressor genes. No room at the p53 inn. Nature 365, 17-18.

Pino, M. S. and Chung, D. C. (2010). Application of molecular diagnostics for the detection of Lynch syndrome. Expert Rev Mol Diagn. 10, 651-65.

Plotz, G., Zeuzem, S., and Raedle, J. (2006). DNA mismatch repair and Lynch syndrome. J. Mol. Histol. 37, 271-283.

Podolsky, D. K. (2002). Inflammatory Bowel Disease. New England Journal of Medicine. 347, 417–29

Polakis, P. (1997). The adenomatous polyposis coli (APC) tumor suppressor. Biochim Biophys Acta. 3, 127-47.

Poli, G., Schaur, R. J., Siems, W. G., Leonarduzzi, G. (2008). 4-Hydroxynonenal: a membrane lipid oxidation product of medicinal interest. Medicinal Research Reviews. 28, 569–631.

Ponder, B. A. and Wilkinson, M. M. (1986). Direct examination of the clonality of carcinogen-induced colonic epithelial dysplasia in chimeric mice. J. Natl. Cancer Inst. 77, 967-976.

Powell, S. M., Zilz, N., Beazer-Barclay, Y., Bryan, T. M., Hamilton, S. R., Thibodeau, S. N., Vogelstein, B., Kinzler, K. W. (1992). APC mutations occur early during colorectal tumorigenesis. Nature. 359, 235-7.

Praml, C., Savelyeva, L., Le Paslier, D., Siracusa, L. D., Buchberg, A. M., Schwab, M., Amler, L. C., (1995). Human homologue of a candidate for the Mom1 locus, the secretory type II phospholipase A2 (PLA2S-II), maps to 1p35-36.1/D1S199. Cancer Res. 55; 5504-6.

Preston, S. L., Wong, W. M., Chan, A. O., Poulsom, R., Jeffery, R., Goodlad, R. A., Mandir, N., Elia, G., Novelli, M., Bodmer, W. F., Tomlinson, I. P., and Wright, N. A. (2003). Bottom-up histogenesis of colorectal adenomas, origin in the monocryptal adenoma and initial expansion by crypt fission. Cancer Res. 63, 3819-3825.

Quan, L., Stassen, A. P., Ruivenkamp, C. A., van Wezel, T., Fijneman, R. J., Hutson, A., Kakarlapudi, N., Hart, A, A, Demant, P. (2011). Most lung and colon cancer susceptibility genes are pair-wise linked in mice, humans and rats. PLoS One. 2011 Feb 24, 6(2), e14727.

Quandt, K., Frech, K., Karas, H., Wingender, E. and Werner, T. (1995). MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res. 23, 4878-84.

Radice, G, L. (2000). Committee Reports, Mouse Genome Database (MGD). Bar Harbor, Maine, Mouse Genome Informatics, The Jackson Laboratory. Chromosome 18.

Ramensky, V., Bork, P., Sunyaev, S. (2002) Human non-synonymous SNPs, server and survey. Nucl Acids Res, 30, 3894.

Rao, C. V., Yang, Y. M., Swamy, M. V., Liu, T., Fang, Y., Mahmood, R., Jhanwar-Uniyal, M., and Dai W. (2005). Colonic tumorigenesis in BubR1+/-ApcMin/+ compound

mutant mice is linked to premature separation of sister chromatids and enhanced genomic instability. Proc Natl Acad Sci U S A. 102, 4365-70.

Rapiti, E., Fioretta, G., Verkooijen, H. M., Zanetti, R., Schmidlin, F., Shubert, H., Merglen, A., Miralbell, R., Bouchardy, C. (2008). Increased risk of colon cancer after external radiation therapy for prostate cancer. Int J Cancer. 123, 1141-5.

Reymond, A., Marigo, V., Murat, B., Yaylaoglu, A., Leoni, C., Scamuffa, N., Caccioppoli, C Emmanouil T. Dermitzakis, Robert Lyle, Sandro Banfi, Gregor Eichele, Stylianos E. Antonarakis & Andrea Ballabio. (2000). Human chromosome 21 gene expression atlas in the mouse Nature. 420, 582-586.

Riggins, G. 0. J., Markowitz, S., Wilson, J. K., Vogelstein, B. and Kinzler, K. W. (1995). Absence of secretory phospholipase a(2) gene alterations in human colorectal cancer. Cancer Res. 55, 5184-5186.

Risch, N. and Merikangas, K. (1996). The future of genetic studies of complex human diseases. Science 273, 1516–1517.

Ritsert, J. C. (1993). Interval mapping of multiple quantitative trait loci. Genetics. 135, 205-211.

Robanus-Maandag, E. C., Koelink, P. J., Breukel, C., Salvatori, D. C., Jagmohan-Changur, S. C., Bosch, C. A., Verspaget, H. W., Devilee, P., Fodde, R., Smits, R. (2010). A new conditional Apc-mutant mouse model for colorectal cancer. Carcinogenesis. 31, 946-52.

Roberts, A. B., Derynck, R. (2001). Meeting report, signaling schemes for TGF-beta. Sci STKE. 18, 113

Roberts, P.J. and Der, C.J. (2007) Targeting the Raf–MEK–ERK mitogen-activated protein kinase cascade for the treatment of cancer. Oncogene 26, 3291–3310

Roderick, T. H. (1963) The response of twenty seven inbred strains of mice to daily doses of whole body x-irradiation. Radiat. Res., 20,614-631

Rodrigues, N. R., Rowan, A., Smith, M. E., Kerr, I. B., Bodmer, W. F., Gannon, J. V., and Lane, D. P. (1990). p53 mutations in colorectal cancer. Proc. Natl. Acad. Sci. U. S. A. 87, 7555-7559.

Rodriguez-Bigas, M. A. (1997). Genetic testing is important in families with a history suggestive of hereditary non-polyposis colorectal cancer even if the Amsterdam criteria are not fulfilled. Br J Surg. 84, 1027-8

Rogus, J. J., Krolewski, A. S., (1996). Using discordant sib pairs to map loci for qualitative traits with high sibling recurrence risk. Am J Hum Genet. 59, 1579–81.

Rokutan, K., Kawahara, T., Kuwano, Y., Tominaga, K., Sekiyama, A., Teshima-Kondo, S. (2006). NADPH oxidases in the gastrointestinal tract, a potential role of Nox1 in innate immune response and carcinogenesis. Antioxid Redox Signal. 8, 1573-82

Rothkamm K, Kühne M, Jeggo PA, Löbrich M. (2001). Radiation-induced genomic rearrangements formed by nonhomologous end-joining of DNA double-strand breaks. Cancer Res. 15, 3886-93.

Rothwell, P. M. (2013). Aspirin in prevention of sporadic colorectal cancer, current clinical evidence and overall balance of risks and benefits. Recent Results Cancer Res. 191,121-42

Rubenstein, P., Walker, M., Carpenter, C., Carrier, C., Krassner, J., Falk, C. and Ginsberg, F. (1981). Genetics of HLA disease associations, The use of the haplotype relative risk (HRR) and the “haplo-delta” (Dh) estimates in juvenile diabetes from three racial groups. Hum. Immunol. 3, 384.

Ruivenkamp, C. A., van Wezel, T., Zanon C. (2002). Ptprj is a candidate for the mouse colon-cancer susceptibility locus Scc1 and is frequently deleted in human cancers. Nat. Genet. 31, 295–300

Rustgi, A. K. (2007). The genetics of hereditary colon cancer. Genes Dev. 21, 2525-38.

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., Mullikin, J. C., Mortimore, B. J., Willey, D. L., Hunt, S. E., Cole, C. G., Coggill, P. C., Rice, C. M., Ning, Z., Rogers, J., Bentley, D. R., Kwok, P. Y., Mardis, E. R., Yeh, R. T.,

Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier. L., Waterston, R. H., McPherson, J. D., Gilman. B., Schaffner. S., Van Etten, W. J., Reich, D., Higgins, J., Daly, M. J., Blumenstiel. B., Baldwin, J., Stange-Thomann, N., Zody, M. C., Linton, L., Lander, E. S., Altshuler, D. (2001). International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature. 409, 928-33.

Salhab, N., Jones, D. J., Bos, J. L, Kinsella, A., Schofield, P. F. (1989). Detection of ras gene alterations and ras proteins in colorectal cancer. Dis Colon Rectum. 32, 659-64.

Sanchez-Cespedes, M. (2007). A role for LKB1 gene in human cancer beyond the Peutz-Jeghers syndrome. Oncogene. 26, 7825-32.

Sanchez-Cespedes, M., Parrella, P., Esteller, M., Nomoto, S., Trink, B., Engles, J. M., Westra, W. H., Herman, J. G., Sidransky, D. (2002). Inactivation of LKB1/STK11 is a common event in adenocarcinomas of the lung. Cancer Res. 62, 3659-62.

Sanders, M., A., Majumdar A., P. (2011). Colon cancer stem cells: implications in carcinogenesis. Front Biosci. 1, 1651-62

Sansom, O. J., Meniel, V. S., Muncan, V., Phesse, T. J., Wilkins, J. A., Reed, K, R., Vass, J. K., Athineos, D., Clevers, J., Clarke, A. R. (2007). Myc deletion rescues Apc deficiency in the small intestine, Nature. 446, 676–679.

Sasai, H., Masaki, M. and Wakitani, K., (2000) Suppression of polypogenesis in a new mouse strain with a truncated Apc (Delta474) by a novel COX-2 inhibitor, JTE-522, Carcinogenesis. 21, 953–958.

Sax, K., (1923) The association of size differences with seed coat pattern and pigmentation in Phaseolus vulgaris. Genetics. 8, 552–560.

Schwenter,F., Ratjen,F., Berk,T., Gallinger,S., Gryfe,R., Gradinger,A.B., Md,M.E., and Durno,C.A. (2011). Juvenile polyposis syndrome, SMAD4 mutations, and hereditary hemorrhagic telangiectasia. J. Pediatr. Gastroenterol. Nutr. 54, 120-2

Scott, R. J., McPhillips, M., Meldrum, C. J. (January 2001). Hereditary nonpolyposis colorectal cancer in 95 families, differences and similarities between mutation-positive and mutation-negative kindreds. Am. J. Hum. Genet. 68, 118–127

Segditsas, S., Rowan, A. J., Howarth, K., Jones, A., Leedham, S., Wright, N. A., Gorman, P., Chambers, W., Domingo, E., Roylance, R. R, Sawyer, E. J., Sieber, O. M., Tomlinson, I. P. (2008). APC and the three-hit hypothesis. Oncogene. 8,146-55.

Seksik, P., Nion-Larmurier, I., Sokol, H, Beaugerie, L. (2009). Cosnes. Effects of light smoking consumption on the clinical course of Crohn's disease. J. Inflamm Bowel Dis. 15, 734-41

Sen. S., and Churchill, G. A. (2001). A statistical framework for quantitative trait mapping. Genetics 159, 371-87.

Senda,T., Iizuka-Kogo, A., Onouchi, T., and Shimomura, A. (2007). Adenomatous polyposis coli (APC) plays multiple roles in the intestinal and colorectal epithelia. Med. Mol. Morphol. 40, 68-81.

Seth, R., Keeley, J., Abu,-Ali G., Crook, S., Jackson, D., Ilyas, M. (2009). The putative tumour modifier gene ATP5A1 is not mutated in human colorectal cancer cell lines but expression levels correlate with TP53 mutations and chromosomal instability. J Clin Pathol. 62, 598-603.

Shen, L., Toyota, M., Kondo, Y., Lin, E., Zhang, L., Guo, Y., Hernandez, N. S., Chen, X., Ahmed, S., Konishi, K., Hamilton, S. R., Issa, J. P. (2007). Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer.Proc Natl Acad Sci U S A. 104, 18654-9.

Sherry, S. T., Ward, M. H., Kholodov, M. M., Baker, J., Phan, L., Smigielski, E. M., Sirotkin, K. (2001) dbSNP, the NCBI database of genetic variation. Nucleic Acids Research 29, 308-11

Shibata, H., Toyama, K., Shioya, H., Ito, M., Hirota, M., Hasegawa, S., Matsumoto, H., Takano, H., Akiyama, T., Toyoshima, K., Kanamaru, R., Kanegae, Y., Saito, I.,

Nakamura, Y., Shiba, K. and Noda T. (1997). Rapid colorectal adenoma formation initiated by conditional targeting of the Apc gene, Science. 278, 120–123.

Shih, I. M., Wang, T. L., Traverso, G., Romans, K., Hamilton, S. R., Ben-Sasson, S., Kinzler, K. W., Vogelstein, B. (2001). Top-down morphogenesis of colorectal tumors. Proc Natl Acad Sci U S A. 98, 2640-5.

Shiraishi, M., Oates, A. J., and Sekiya, T. (2002). An overview of the analysis of DNA methylation in mammalian genomes. Biol. Chem. 383, 893-906.

Shoemaker, A. R., Moser, A. R. and Dove, W. F. (1995). N-ethyl-N-nitrosourea treatment of multiple intestinal neoplasia (Min) mice, age-related effects on the formation of intestinal adenomas, cystic crypts, and epidermoid cysts. Cancer Res. 55, 4479–4485

Sieber, O. M., Lipton, L., Crabtree, M., Heinimann, K., Fidalgo, P., Phillips, R. K., Bisgaard, M. L., Orntoft, T. F., Aaltonen, L. A., Hodgson, S. V., Thomas, H. J., and Tomlinson,I. P. (2003). Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH. N. Engl. J. Med. 348, 791-799.

Sillanpä M., J, Arjas, E. (1998). Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. Genetics. 148, 1373–88

Silverman, K., Korarkor, R., Siracusa, L. D. and Buchberg, A. M. (2002). Identification of the modifier of Min 2 (Mom2) locus, a new mutation that influences Apc-induced intestinal neoplasia.Genome Res. 12, 88-97.

Sjolander, K., Karplus, K., Brown, M., Hughey, R., Krogh, A., Mian, I. S., Haussler, D. (1996) Dirichlet mixtures, a method for improved detection of weak but significant protein sequence homology. Comput Appl Biosci. 12, 327

Slagboom,P. E. and Meulenbelt, I. (2002). Organisation of the human genome and our tools for identifying disease genes. Biol. Psychol. 61, 11-31.

Slattery, M. L., Curtin, K., Anderson, K., Ma, K. N., Ballard, L., Edwards, S., Schaffer, D., Potter, J., Leppert, M., Samowitz, W. S. (2000). Associations between cigarette

smoking, lifestyle factors, and microsatellite instability in colon tumors. J Natl Cancer Inst. 92, 1831-6

Slattery, M. L., Kerber, R. A., (1994). Family history of cancer and colon cancer risk, the Utah Population Database. J Natl Cancer Inst. 86, 1618-26.

Smith, G. C. and Jackson, S. P. (1999). The DNA-dependent protein kinase. Genes & Dev. 13, 916-934

Smits, R., van der Houven van Oordt, W., Luz, A., Zurcher, C., Jagmohan-Changur, S., Breukel, C., Khan, P. M., Fodde, R., (1998). Apc1638N, a mouse model for familial adenomatous polyposis-associated desmoid tumors and cutaneous cysts, Gastroenterology. 114, 275–283.

Smits, R., Kartheuser, A., Jagmohan-Changur, S, Leblanc, V., Breukel, C., de Vries, A., van Kranen, H., van Krieken, J. H., Williamson, S., Edelmann, W., Kucherlapati, R., Khan, P. M., Fodde, R. (1997) Loss of Apc and the entire chromosome 18 but absence of mutations at the Ras and Tp53 genes in intestinal tumors from Apc1638N, a mouse model for Apc-driven carcinogenesis. Carcinogenesis. 18, 321-7

Soller, M. (1976) On the power of experimental design for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor. Appl. Genet. 47, 35–39

Sountoulides, P., Koletsas, N., Kikidakis, D., Paschalidis, K., and Sofikitis, N. (2010). Secondary malignancies following radiotherapy for prostate cancer. Ther Adv Urol. 2, 119-125

Soussi, T. and Wiman, K. G. (2007). Shaping genetic alterations in human cancer, the p53 mutation paradigm. Cancer Cell 12, 303-12.

Sparks, A. B., Morin, P. J., Vogelstein, B., and Kinzler, K. W. (1998). Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer. Cancer Res. 58, 1130-1134.

Spelman, R., Bovenhuis, H. (1998). Genetic response from marker assisted selection in an outbred population for differing marker bracket sizes and with two identified quantitative trait loci. Genetics. 148, 1389–1396.

Spielman, R. S, and Ewens, W. J. (1998). A sibship test for linkage in the presence of association, The sib transmission/disequilibrium test. Am. J. Hum. Genet. 62, 450–458.

Spielman, R. S., McGinnis, R. E., and Ewens, W. J. (1993). Transmission test for linkage disequilibrium, The insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am. J. Hum. Genet. 52, 506–516.

Spirio, L., Otterud, B., Stauffer, D., Lynch, H., Lynch, P., Watson, P., Lanspa, S., Smyrk, T., Cavalieri, J., Howard, L (1992). Linkage of a variant or attenuated form of adenomatous polyposis coli to the adenomatous polyposis coli (APC) locus. Am J Hum Genet. 51, 92-100.

Srour, M., Rivière, J. B., Pham, J. M., Dubé, M. P., Girard, S., Morin, S., Dion, P. A., Asselin, G., Rochefort, D., Hince, P., Diab, S., Sharafaddinzadeh, N., Chouinard, S., Théoret, H., Charron, F., Rouleau, G. A. (2010). Mutations in DCC cause congenital mirror movements. Science. 328, 592.

Stehelin, D., Fujita, D. J., Padgett, T., Varmus, H. E., Bishop, J. M. (1977). Detection and enumeration of transformation-defective strains of avian sarcoma virus with molecular hybridization. Virology. 76, 675-84.

Steiner, C. C., Weber, J. N., Hoekstra, H. E. (2007). Adaptive variation in beach mice produced by two interacting pigmentation genes. Plos Biol. 5, e219.

Stenson, P. D., Mort, M., Ball, E. V., Howells, K., Phillips, A. D., Thomas, N. S., Cooper, D. N. (2009) The human gene mutation database, 2008 update. Genome Medicine

Stirbu, I., Kunst, A. E., Vlems, F. A., Visser, O., Bos, V., Deville, W., Nijhuis, H. G., Coebergh, J. W. (2006). Cancer mortality rates among first and second generation migrants in the Netherlands, Convergence toward the rates of the native Dutch population. Int J Cancer. 119, 2665-72.

Stoneking, M. (2001). Single nucleotide polymorphisms. From the evolutionary past. Nature. 15, 821-2.

Storey, J., Akey, J., and Kruglyak, L., (2005). Multiple locus linkage analysis of genomewide expression in yeast. PLOS Biology, 3, 8

Strohmeyer, D., Langenhof, S., Ackermann, R., Hartmann, M., Strohmeyer, T., Schmidt, B. (1997). Analysis of the DCC tumor suppressor gene in testicular germ cell tumors, mutations and loss of expression. J Urol. 157, 1973-6.

Sturt, N. J. H,, Gallagher, M. C., Bassett, P., Philp, C. R., Neale, K. F., Tomlinson, I. P., Silver A. R., Phillips, R. K. S.(2004). Evidence for genetic predisposition to desmoid tumours in familial adenomatous polyposis independent of the germline APC mutation. 53, 1832–1836

Sturtevant, A. H. (1913). The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. Journal of Experimental Zoology, 14, 43-59.

Stylianou, I. M., Korstanje, R., Li, R., Sheehan, S., Paigen, B., Churchill, G. A. (2006). Quantitative trait loci analysis for obesity reveals multiple networks on interacting loci. Mamm Genome. 17, 22-36.

Su, L. K., Kohlmann, W., Ward, P. A., and Lynch, P. M. (2002). Different familial adenomatous polyposis phenotypes resulting from deletions of the entire APC exon 15. Hum. Genet. 111, 88-95

Su, L. K., Vogelstein, B., and Kinzler, K. W. (1993). Association of the APC tumor suppressor protein with catenins. Science 262, 1734-1737.

Su, W. L., Sieberts, S. K., Kleinhanz, R. R., Lux, K., Millstein, J., Molony. C, Schadt, E. E. (2010). Assessing the prospects of genome-wide association studies performed in inbred mice. Mamm Genome. 21, 143-52.

Sugimura, K., Furukawa, T., Okizuka, H., Yuasa, K., Mihara, Y., Ishida, T. (1990). Post radiation changes in the rectum, assessment by MR imaging. Nihon Gan Chiryo Gakkai Shi. 25, 2693-701

Sugiyama, F,. Churchill, G., A. Higgens, D., C. Johns, C, Makaritsis, K., P. Gavras, H., Paigen, B. (2000). Concordance of murine quantitative trait loci for salt-induced hypertension with rat and human loci. Genomics. 171, 70–77

Suit, H., Goldberg, S., Niemierko, A., Ancukiewicz, M., Hall, E., Goitein, M., Wong, W., Paganetti, H. (2007). Secondary carcinogenesis in patients treated with radiation, a review of data on radiation-induced cancers in human, non-human primate, canine and rodent subjects. Radiat Res. 167, 12-42.

Sunyaev, S., Ramensky, V., Bork, P. (2000) Towards a structural basis of human non-synonymous single nucleotide polymorphisms. Trends in Genetics, 16,198-2000.

Sunyaev, S., Ramensky, V., Koch, I., Lathe, W., Kondrashov, A. and Bork, P. (2001) Prediction of deleterious human alleles. Hum Mol Genet 10, 591–597.

Suraweera, N. ,Haines, J., McCart, A., Rogers, P., Latchford, A., Coster, M., Polanco-Echeverry, G., Guenther, T., Wang, J., Sieber, O., Tomlinson, I. and Silver, A. (2006). Genetic determinants modulate susceptibility to pregnancy-associated tumourigenesis in a recombinant line of Min mice, Hum. Mol. Genet. 15, 3429–3435.

Takagi, Y., Kohmura, H., Futamura, M., Kida, H., Tanemura, H., Shimokawa, K., Saji S. (1996) Somatic alterations of the DPC4 gene in human colorectal cancers in vivo. Gastroenterology. 111, 1369–1372.

Takeda, H., Sonoshita, M., Oshima, H., Sugihara, K., Chulada, P.C., Langenbach, R. Oshima, M. and Taketo M.M. (2003). Cooperation of cyclooxygenase 1 and cyclooxygenase 2 in intestinal polyposis, Cancer Res. 63 4872–4877.

Tenesa, A., Farrington, S. M., Prendergast, J. G., Porteous, M. E., Walker, M., Haq, N., Barnetson, R. A., Theodoratou, E., Cetnarskyj, R., Cartwright, N., Semple, C., Clark, A. J., Reid, F. J., Smith, L. A., Kavoussanakis, K., Koessler, T., Pharoah, P. D., Buch, S., Schafmayer, C., Tepel, J ., Schreiber, S., Volzke, H., Schmidt, C. O., Hampe, J., Chang-Claude, J., Hoffmeister, M., Brenner, H., Wilkening, S., Canzian, F., Capella, G., Moreno,V., Deary, I. J., Starr, J. M., Tomlinson, I. P., Kemp, Z., Howarth, K., Carvajal-Carmona, L., Webb, E., Broderick, P., Vijayakrishnan, J., Houlston, R. S., Rennert, G., Ballinger, D., Rozek, L., Gruber, S. B., Matsuda, K., Kidokoro, T., Nakamura, Y., Zanke,

B. W., Greenwood, C. M., Rangrej, J., Kustra, R., Montpetit, A., Hudson, T. J., Gallinger, S., Campbell, H. and Dunlop,M.G. (2008). Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat. Genet. 40, 631-637.

Tetsu,O. and McCormick, F. (1999). Beta-catenin regulates expression of cyclin D1 in colon carcinoma cells. Nature 398, 422-426.

Thoday, J. M. (1961). Location of polygenes. Nature. 191, 368-370

Timmermann, B., Kerick, M., Roehr, C., Fischer, A., Isau, M. (2010) Somatic Mutation Profiles of MSI and MSS Colorectal Cancer Identified by Whole Exome Next Generation Sequencing and Bioinformatics Analysis. PLoS ONE 5(12)

Tomlinson, I. P., Webb, E., Carvajal-Carmona, L., Broderick, P., Howarth, K., Pittman, A. M., Spain, S., Lubbe, S., Walther, A., Sullivan, K., Jaeger, E., Fielding, S., Rowan, A., Vijayakrishnan, J., Domingo, E., Chandler, I., Kemp, Z., Qureshi, M., Farrington, S. M., Tenesa, A., Prendergast, J. G., Barnetson, R. A., Penegar, S., Barclay, E., Wood, W., Martin, L., Gorman, M., Thomas, H., Peto, J., Bishop, D. T., Gray, R., Maher, E. R., Lucassen, A., Kerr, D., Evans, D. G., Schafmayer, C., Buch, S., Volzke, H., Hampe, J., Schreiber, S., John, U., Koessler, T., Pharoah, P., van, W. T., Morreau, H., Wijnen, J. T., Hopper, J. L., Southey, M. C., Giles, G. G., Severi, G., Castellvi-Bel, S., Ruiz-Ponte, C., Carracedo, A., Castells, A., Forsti, A., Hemminki, K., Vodicka,P., Naccarati,A., Lipton, L., Ho, J. W., Cheng, K. K., Sham, P. C., Luk, J., Agundez, J. A., Ladero, J. M., de la Hoya, M., Caldes, T., Niittymaki, I., Tuupanen, S., Karhu, A., Aaltonen, L., Cazier, J. B., Campbell, H., Dunlop, M. G. and Houlston, R. S. (2008). A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat. Genet. 40, 623-630.

Tomlinson, I., Webb, E., Carvajal-Carmona, L., Broderick, P., Kemp, Z., Spain, S., Penegar,S., Chandler, I., Gorman, M., Wood, W., Barclay, E., Lubbe, S., Martin, L., Sellick, G., Jaeger, E., Hubner, R., Wild, R., Rowan, A., Fielding, S., Howarth, K., Silver, A., Atkin, W., Muir, K., Logan, R., Kerr, D., Johnstone, E., Sieber, O., Gray, R., Thomas, H., Peto, J., Cazier, J. B. and Houlston, R. (2007b). A genome-wide association scan of

tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat. Genet. 39, 984-988.

Tomlinson, I. P., Beck, N. E., Neale, K. and Bodmer, W. F. (1996). Variants at the secretory phospholipase A2 (PLA2G2A) locus, analysis of associations with familial adenomatous polyposis and sporadic colorectal tumours. Ann. Hum. Genet. 5, 369-376.

Torrance, C. J., Jackson, P. E., Montgomery, E., Kinzler, K. W., Vogelstein, B., Wissner, A., Nunes, M., Frost P., Discafani, C. M. (2000). Combinatorial chemoprevention of intestinal neoplasia. Nat. Med. 6, 1024–1028

Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J. G., Baylin, S. B., and Issa, J. P. (1999). CpG island methylator phenotype in colorectal cancer. Proc. Natl. Acad. Sci. U. S. A. 96, 8681-8686.

Triantafillidis, J. K., Nasioulas, G., Kosmidis, P. A. (2009) Colorectal cancer and inflammatory bowel disease, epidemiology, risk factors, mechanisms of carcinogenesis and prevention strategies. Anticancer Res. 29, 2727-37.

Tysk, C., Lindberg, E., Järnerot, G., Flodérus-Myrhed, B. (1988). Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking.Gut. 29, 990-6

Vagefi, P. A., Longo, W. E. (2005). Colorectal cancer in patients with inflammatory bowel disease. Clin Colorectal Cancer. 4, 313-9

Vallbohmer, D., Brabender, J., Yang, D., Schneider, P. M., Metzger, R., Danenberg, K. D., Holscher, A. H.and Danenberg, P. V. (2006). DNA methyltransferases messenger RNA expression and aberrant methylation of CpG islands in non-small-cell lung cancer, association and prognostic value. Clin. Lung Cancer. 8, 39-44.

van Es, J. H., Giles, R. H., Clevers, H. C. (2001). The many faces of the tumor suppressor gene APC. Exp Cell Res. 1, 126-34.

Varley, J. (2003) TP53, hChk2, and the Li-Fraumeni syndrome. Methods Mol Biol. 222, 117-29.

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

Velmala, R. J., Vilkki, H. J., Elo, K. T, de Koning, D. J., Mäki–Tanila, A. V. (1999). Animal Gen. 30, 136-143.

Vilenchik, M. M., Knudson, A. G. (2003). Endogenous DNA double-strand breaks, production, fidelity of repair, and induction of cancer. Proc Natl Acad Sci U S A. 100, 12871-6.

Vogelstein, B. (1988) Genetic alterations during colorectal-tumor development. N. Engl. J. Med. 319, 525–532

Vogelstein, B. and Kinzler, K. W. (2004) Cancer genes and the pathways they control, Nature Medicine 10, 789-799

Vogelstein, B., Fearon, E. R., Hamilton, S. R., Kern, S. E., Preisinger, A. C., Leppert, M., Nakamura, Y., White, R., Smits, A. M., and Bos, J. L. (1988). Genetic alterations during colorectal-tumor development. N. Engl. J. Med. 319, 525-532.

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-tumor development. N Engl J Med. 319, 525-32.

Vogelstein, B., Kinzler, K. W. (1993) The multistep nature of cancer. Trends Genet. 9, 138-41.

Waldman, T., Kinzler, K. W., Vogelstein, B.(1995). p21 is necessary for the p53-mediated G1 arrest in human cancer cells. Cancer Res. 55, 5187-90.

Walker, F., Zhang, H., Odorizzi, A., Burgess, A., W. (2012). LGR5 Is a Negative Regulator of Tumourigenicity, Antagonizes Wnt Signalling and Regulates Cell Adhesion in Colorectal Ca10. Cell Lines. Plos. One 6, 22733

Walsh, D. (1987) Dead tissue traumatism from roentgen ray exposure, BMJ 3,272-273,

Wang, G. S., Cooper, T. A. (2007). Splicing in disease, disruption of the splicing code and the decoding machinery. Nat. Rev. Genet. 8, 749–761

Wang, Q., Lasset, C., Desseigne, F., Saurin, J. C., Maugard, C., Navarro. C., Ruano, E., Descos, L., Trillet-Lenoir, V., Bosset, J.F., Puisieux, A. (1999) Prevalence of germline mutations of hMLH1, hMSH2, hPMS1, hPMS2, and hMSH6 genes in 75 French kindreds with nonpolyposis colorectal cancer. Hum Genet. 105, 79-8

Wang, S., Basten, C. J., and Zeng Z-B. (2007). Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. Avaliable from, (http,//statgen.ncsu.edu/qtlcart/WQTLCart.htm)

Wang, Z., Jin, H., Xu, R., Mei, Q., Fan, D. (2009). Triptolide downregulates Rac1 and the JAK/STAT3 pathway and inhibits colitis-related colon cancer progression. Exp Mol Med. 41, 717-27.

Wang, Q., Liu, T., Fang, Y., Xie, S., Huang, X., Mahmood, R., Ramaswamy, G., Sakamoto, K. M., Darzynkiewicz, Z., Xu, M., Dai, W. (2004). BUBR1 deficiency results in abnormal megakaryopoiesis. Blood. 15, 1278-85.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F. (2002) Initial sequencing and comparative analysis of the mouse genome. Nature. 420, 520–562

Webb, E. L., Houlston, R. S. (2007) Association studies using familial cases, an efficient strategy for identifying low-penetrance disease alleles. Methods Mol Biol.

Weisenberger, D. J., Siegmund, K. D., Campan, M., Young, J., Long, T. I., Faasse, M. A., Kang, G. H., Widschwendter, M., Weener, D., Buchanan, D., Koh, H., Simms, L., Barker, M., Leggett, B., Levine, J., Kim, M., French, A. J., Thibodeau, S. N., Jass, J., Haile, R., and Laird, P. W. (2006). CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nat. Genet. 38, 787-793.

Westbrook, A. M., Szakmary, A., Schiestl, R. H. (2010). Mechanisms of intestinal inflammation and development of associated cancers, lessons learned from mouse models.Mutat. Res. 705, 40-59.

Weterings, E., Chen, D. J. (2007). DNA-dependent protein kinase in nonhomologous end joining: a lock with multiple keys? J Cell Biol. 179, 183–186.

Weyden, L., Adams, D. J. (2007). The Ras-association domain family (RASSF) members and their role in human tumourigenesis. Biochim Biophys Acta. 1776, 58–85.

Wheeler, J. M. (2005). Epigenetics, mismatch repair genes and colorectal cancer. Ann R Coll Surg Engl. 87, 15-20.

Willett, C. G., Ooi, C. J., Zietman, A. L., Menon, V. , Goldberg, S., Sands, B. E., Podolsky, D. K. (2000). Acute and late toxicity of patients with inflammatory bowel disease undergoing irradiation for abdominal and pelvic neoplasms. Int J Radiat Oncol Biol Phys. 46, 995-8.

Woolf, C. M. (1958) A genetic study of carcinoma of the large intestine. Am J Hum Genet 10, 42-7

Wu, Y., Antony, S., Juhasz, A., Lu, J., Ge, Y., Jiang, G., Roy, K., Doroshow, J. H. (2011). Up-regulation and sustained activation of Stat1 are essential for interferon-gamma (IFN-gamma)-induced dual oxidase 2 (Duox2) and dual oxidase A2 (DuoxA2) expression in human pancreatic cancer cell lines. J Biol Chem. 286, 12245-56.

Xi, T., Jones, I. M., and Mohrenweiser, H. W., (2004). Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function. 83, 970-9

Xu, L., Corcoran, R. B., Welsh, J. W., Pennica, D., and Levine, A. J. (2000). WISP-1 is a Wnt-1- and beta-catenin-responsive oncogene. Genes Dev. 14, 585-595.

Yalcin, B. and Flint, F. (2012). Association studies in outbred mice in a new era of full-genome sequencing. Mamm Genome. 23, 719–726.

Yang, C-K., Hsu, T-C., Wang, T-Y. (2008). Radiation-Induced Cancers of the Colon and Rectum-Analysis of Five Cases. J Soc Colon Rectal Surgeon. 19, 41-46.

Yeoh, E., Horowitz, M., Russo, A., Muecke, T., Robb, T., Chatterton, B. (1995). The effects of abdominal irradiation for seminoma of the testis on gastrointestinal function. J Gastroenterol Hepatol. 10, 125-30.

Yi, N., and Xu, S. (2002). Mapping quantitative trait loci with epistatic effects. Genet Res. 79, 185-98.

Yoon, H. Y., et al. (2006) ARAP2 effects on the actin cytoskeleton are dependent on Arf6-specific GTPase-activating-protein activity and binding to RhoA-GTP. J Cell Sci. 119, 4650–66.

Yu, K., Okayasu, R., Weil, M. M., Silver, A., McCarthy, M., Zabriskie, R., Long, S., Cox, R., Ullrich, RL. (2001). Elevated breast cancer in irradiated BALB/c associates with unique functional polymorphism of the Prkdc (DNA-dependent protein kinase catalytic subunit gene). Cancer Res. 61, 1820–1824.

Zhang, H. T., Zhao, J., Zheng, S. Y., Chen, X. F. (2005). Is TGFBR1*6A Really Associated With Increased Risk of Cancer? J Clin Oncol. 23, 7743-7744

Zhao, J., Halfyard, B., Roebothan, B., West, R., Buehler, S., Sun, Z., Squires, .J., McLaughlin, J. R., Parfrey, P. S., Wang, P. P. (2010) Tobacco smoking and colorectal cancer, a population-based case-control study in Newfoundland and Labrador. Can J Public Health. 101, 281-9.

Zhao, Z., Fu, Y.X., Hewett-Emmett, D., Boerwinkle, E. (2003). Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. Gene. 17, 207–213.

Zhao-Bang Zeng. (1994). Precision Mapping of Quantitative Trait Loci. Genetics 136, 1457-68

Zhong, S., Yang, J. H., Liu, K., Jiao, B. H., Chang, Z. J. (2012). Quantitative assessment of the association between MTHFR C677T polymorphism and colorectal cancer risk in East Asians. Tumour Biol. 33, 2041-51