Investigating the Genetics and Pharmacogenetics of Bowel Cancer

Submitted for the degree of Doctor of Philosophy at Cardiff University

Hannah West

2013

DECLARATION

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

Signed - H.West Date - 13th January 2014

STATEMENT 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD

Signed - H.West Date – 13th January 2014

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.

Other sources are acknowledged by explicit references.

Signed - H.West Date – 13th January 2014

STATEMENT 3

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

Signed …… (candidate) Date ……

STATEMENT 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans **after expiry of a bar on access previously approved by the Graduate Development Committee.**

Signed - H.West Date – 13th January 2014

Summary

In this thesis, we aimed to identify genetic factors that influence the risk of colorectal cancer (CRC). We also sought alleles that contribute to the likelihood of extreme adverse reactions to treatment.

We validated five previously identified low penetrance variants using our training phase cohort, consisting of 2,186 advanced CRC (aCRC) from the COIN and COIN-B trials and 2,176 geographically matched controls. Using this cohort we also identified a variant in *RAD1* that was significantly associated with risk (X^2=13.51, P=2x10^{-4}). However, we failed to replicate these findings in an aCRC validation cohort consisting of 1,053 cases and 1,397 geographically matched controls (X^2=2.76, P=0.1), potentially as a result of a lack of power due to insufficient sample numbers.

We identified ten patients from the COIN trial with severe peripheral neuropathy associated with oxaliplatin (PNAO) treatment. Through exome resequencing we identified a novel stop gain variant (Ser613X) in the nucleotide excision repair gene (NER), *ERCC4.* Following analysis of 54 additional patients from the COIN trial with PNAO, we identified three rare nonsynonymous variants (Pro379Ser, Arg576Thr and Glu875Gly) that were predicted to interfere with protein function. Consistent with the rare variant hypothesis of common disease, two of these variants were seen to collectively contribute to the risk of the phenotype (7/63 [11.11%] of patients with PNAO compared to 86/1763 [4.88%] of patients without PNAO; X^2=4.89, P=0.03).

Using the fission yeast, *Schizosaccharomyces pombe,* we sought to elucidate functional effects of these variants in *ERCC4* by creating a model system. Using *cre* recombinase mediated cassette exchange, we introduced the variants of interest into the *ERCC4* homolog, *rad16.* Following treatment with a range of DNA damaging agents, we observed an increased sensitivity following introduction of the novel stop gain, indicating a defect in the NER pathway. Additionally, there was a clear pattern of oxaliplatin-specific sensitivity of strains with the introduced rare nonsynonymous variants, suggesting a defect of XPF in other repair processes associated with interstrand crosslinks.

Acknowledgements

I would like to thank the following;

My supervisors, Prof. Jeremy Cheadle and Prof. Julian Sampson, for their extraordinary supervision, help and encouragement throughout my PhD.

Tenovus and the Kidani memorial trust for funding this project.

Oliver Fleck for his extensive help with the *Schizosaccharomyces pombe* work (and for his patience whilst I got to grips with the genetics of a new organism), as well as his contribution to several aspects of strain construction and phenotype testing. Special thanks to Rebecca Williams for her friendship and help with several parts of the strain construction. Thanks to the entire Hartsuiker group for their kindness and continued help.

Simon Reed and Richard Webster for advice.

Susan Richman, Richard Adams, Tim Maughan, Dave Fisher and other members of the COIN, FOCUS2, FOCUS3 and PICCOLO trials for their help with acquiring samples.

Stephan Buch and Jochen Hampe for their help with the POPGEN collaboration and for hosting me in Kiel, Germany. Thanks to all other members of the lab for making me so welcome.

All the patients whose invaluable donation of DNA made this project possible.

The administration team, Linda, Sherrie, Hannah, Mark and Sathiya, for helping with the technical bits.

Special thanks to Chris Smith and James Colley for their help throughout my PhD, as well as help with the write up process. Thanks to Rebecca Harris for help with lab work.

Shelley, for going out of her way to help with problems, providing a sympathetic ear, kindness and cake when most needed. Special thanks to Laura Thomas for the ‘science chats’, helping with various techniques/protocols and for her friendship.

Mark, Charlie, Lyndsey, Marc, Michelle, Elaine, Kayleigh, Maria and David, and my friends and family ‘back home’ for making my PhD, for the most part, enjoyable.

Simon for, despite everything, putting up with me.

And finally, Mum and Dad. For their love and support.

Abbreviations

A	Adenine
AC	Amsterdam criteria
ACE	Angiotensin converting enzyme
aCRC	Advanced colorectal cancer
ADCC	Antibody-depedant cell-mediated cytotoxicity
ADL	Activities of daily living
AFAP	Attenuated familial adenomatous polyposis
AGT	O^6-alkylguanine DNA alkyltransferase
AGXT	Alanine glycoxylate transferase
Align-GVGD	Align-Grantham Variation/Grantham Deviation
ANOVA	Analysis of variance
ANXA7	Annexin 7
AP	Abasic
APC	Adenomatous polyposis coli
AT	Ataxia telangiectasia
Atl	Alkytransferase like protein
BER	Base excision repair
BMPR1A	Bone morphogenetic protein receptor, type 1A
bps	Base pairs
BRAF	v-raf murine sarcoma viral oncogene homolog B1
BRCA	Breast cancer, early onset
BRIX1	Ribosome genesis protein
BS	Bloom syndrome
C	Cytosine
CCAT2	Colon cancer associated transcript 2
cDNA	Complementary DNA
CI	Confidence intervals
CIMP	CpG island methylation phenotype
CIN	Chromsomal Instability
CMT	Charcot-marie tooth syndrome
COFS	Cerebro-oculo-facio-skeletal syndrome
COIN	Continuous vs intermittent therapy
CPB	Capecitabine
CPD	Cyclobutane pyrimidine dimer
CRA	Colorectal adenomas
CRAC1	Colorectal adenoma and carcinoma
CRC	Colorectal cancer
CS	Cockayne syndrome
CTCAE	Common Terminology Criteria for Adverse Effects
CTS	Contents trade secret
DACH	1,2-diaminocyclohexane group
dH_2O	Distilled water

DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNAJC21	DnaJ homolog, subfamily C, member 21
dNTPs	Deoxyribonucleotides triphosphates
ddNTPs	Dideoxyribonucleotides triphosphates
DPYD	Dihydropyrimidine dehydrogenase
DSB	Double strand break
DSBR	Double strand break repair
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal growth factor receptor
EIF3H	Eukaryotic translation initiation factor 3, H
EMA	European medicines agency
EMM	Edinburgh minimal media
ENG	Endoglin
EPCAM	Epithelial cell adhesion molecule
ERCC(1-6)	Excision repair cross complementation rodent repair deficiency
EXO	Exonuclease I
FA	Fanconi anaemia
FAP	Familial adenomatous polyposis
FDA	US food and drug administration
fdUMP	Flurodeoxyuridine monophosphate
FOLFOX	5-Fluorouracil, leucovorin and oxaliplatin
FOLFORI	5-Fluorouracil, leucovorin and irinotecan
G	Guanine
GATK	Genome analysis toolkit
gDNA	Genomic DNA
GG-NER	Global genomic nucleotide excision repair
GREM1	Gremlin 1
GSTP1	Glutathione-S-transferase-P1
GWAS	Genome wide association study
HMSN	Hereditary motor and sensory neuropathy
HMPS	Hereditary mixed polyposis syndrome
HR	Homologous recombination
HRC	Human randomised control
HNPCC	Hereditary non polyposis colorectal cancer
HU	Hydroxyurea
HWE	Hardy Weinberg equilibrium
ICL	Interstrand crosslink
ICLR	Interstrand crosslink repair
IDL	Insertion/deletion loop
Indel	Insertion or deletion
IPTG	isopropyl-β-D-thio-galactopyranoside
JPS	Juvenille polyposis syndrome
kb	Kilobase

KRAS	Kirsten rat sarcoma viral oncogene homolog
LB	Luria Bertani
LD	Linkage disequilibrium
LiAc	Lithium Acetate
LIG	Ligase
L95	Lower 95% confidence interval
MAF	Minor allele frequency
MAP	MUTYH associated polyposis
MKK3	Mitogen-activated protein kinase kinase 3
MLPA	Multiplex ligation-dependant probe amplification
MMA	Minimal medium agar
MMG	Megamix Gold
MMR	Mismatch repair
MMS	Methyl methanesulfonate
mRNA	Messenger ribonucleic acid
MSI	Microsatellite instability
MT	Mutant
MTHF	5,10-methylenetetrahydrofolate
MTHFR	Methylenetetrahydrofolate reductase
mTOR	Mammalian target of rapamycin
NBS	Nijmegen breakage syndrome
NER	Nucleotide excision repair
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
NO	Nitric oxide
NRP2	Neuropilin 2
OCT1	Organic cation transporter
ODRP	Other DNA repair pathways
OMIM	Online Mendelian inheritance in man
OR	Odd ratio
ORF	Open reading frame
PCA	Prinicipal component analysis
PCIA	Phenol chloroform isoamyl-alcohol
PCR	Polymerase chain reaction
PICCOLO	Panitumumab, Irinotecan & Ciclosporin in COLOrectal cancer therapy
PIP3	Phosphatidylinositol- 3,4,5 – trisphosphate
PI3KCA	Phosphatidylinositol-4,5-bisphosphate 3-kinase
PFS	Progression free survival
PJS	Peutz-Jegher syndrome
PMP22	Peripheral myelin protein 22
PNAO	Peripheral neuropathy associated with oxaliplatin treatment
POL	Polymerase
PolyPhen	Polymorphism Phenotype
PTEN	Phosphatase and tensin homolog

QLQ	Quality of life questionnaire
RMCE	Recombinase mediated cassette echange
RMHNHST	Royal Marsden Hospital NHS Trust
RNase	Ribonuclease
rpm	Revolutions per minute
rs	Reference SNP
RT-PCR	Real time PCR
RTS	Rothmund-Thomson syndrome
SAP	Shrimp alkaline phosphatase
SCN	Sodium channel, voltage gated
SD	Standard deviation
SDM	Site directed mutagenesis
SDS	Sequence detection system
SDSA	Synthesis dependant strand annealing
SGNE1	Secretogranin
SHIP	Study of Health in Pomerania
SIFT	Sorting intolerant from tolerant
SMAD	Mothers against decapentaplegic
SNP	Single nucleotide polymorphism
SSB	Single strand break
STKB11	Serine/threonine kinase 11
STOML3	Stomatin (Epb7.2)-like 3
T	Thymine
TCF4	T cell factor 4
TC-NER	Transcription coupled nucleotide excision repair
thi	Thiamine
TGFβ	Transforming growth factor β
TNM	Tumour, node, metastasis
Tran	Transcript
TS	Thymidate synthetase
TTC23L	Tetratricopeptide repeat protein 23-like
Ub	Ubiquitination
UGT1A1	UDP-glucuronosyltransferase
UKBS	UK blood service
ura4^{+}	Orotidine 5’-phosphate decarboxylase
UV	Ultraviolet
Uve1	UV damaged DNA endonuclease
UVER	UV damaged DNA endonuclease dependent excision repair pathway
U95	Upper 95% confidence interval
VEGF	Vascular epidermal growth factor
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHOPS	World health organisation performance status
WS	Werner syndrome

WT	Wild type
XELIRI	Capecitabine and irinotecan
XELOX	Capecitabine and oxaliplatin
XFE	XPF-ERCC1 progeroid syndrome
X-gal	5-bromo-4-chloro-3-indoyl-D-galactoside
XP	Xeroderma pigmentosum
YEA	Yeast extract agar
YEL	Yeast extract liquid
5-FU	5-Fluorouracil
5-FOA	5-Fluoroorotic Acid
5’UTR	5’ untranslated region
6-4,PP	Pyrimidine (6-4) pyrimidone
8-oxo-G	8-oxo-7,8-dihydro-2’-deoxyguanosine

Contents

List of figures

Chapter One – Introduction

Chapter Two – Materials and method

Chapter Three – Identifying novel low penetrance alleles in DNA repair genes that predispose to CRC

Chapter Four – Identifying genes associated with oxaliplatin-induced peripheral neuropathy in the treatment of aCRC

Chapter Five – Analysis of candidate genes responsible for PNAO

Chapter Six – Construction of a model system to test the functionality of variants identified in *ERCC4*

Chapter Seven – Investigating the functional effects of variants introduced into *rad16*

Chapter Eight – General discussion

List of tables

Chapter One – Introduction

Chapter Three – Identifying novel low penetrance alleles in DNA repair genes that predispose to CRC

Chapter Four – Identifying genes associated with oxaliplatin-induced peripheral neuropathy in the treatment of aCRC

Chapter Five – Analysis of candidate genes responsible for PNAO

Chapter Seven – Investigating the functional effects of variants introduced into *rad16*

Appendices

Codon table

First base in codon	Second base in codon: T			Second base in codon: C			Second base in codon: A			Second base in codon: G			Third base in codon
T	TTT	Phenylalanine (Phe)		TCT	Serine (Ser)		TAT	Tyrosine (Tyr)		TGT	Cysteine (Cys)		T
	TTC			TCC			TAC			TGC			C
	TTA	Leucine (Leu)		TCA			TAA	STOP		TGA	STOP		A
	TTG			TCG			TAG			TGG	Tryptophan (Trp)		G
C	CTT			CCT	Proline (Pro)		CAT	Histidine (His)		CGT	Arginine (Arg)		T
	CTC			CCC			CAC			CGC			C
	CTA			CCA			CAA	Glutamine (Gln)		CGA			A
	CTG			CCG			CAG			CGG			G
A	ATT	Isoleucine (Ile)		ACT	Threonine (Thr)		AAT	Asparagine (Asn)		AGT	Serine (Ser)		T
	ATC			ACC			AAC			AGC			C
	ATA			ACA			AAA	Lysine (Lys)		AGA	Arginine (Arg)		A
	ATG	Methionine (Met)		ACG			AAG			AGG			G
G	GTT	Valine (Val)		GCT	Alanine (Ala)		GAT	Aspartic Acid (Asp)		GGT	Glycine (Gly)		T
	GTC			GCC			GAC			GGC			C
	GTA			GCA			GAA	Glutamic Acid (Glu)		GGA			A
	GTG			GCG			GAG			GGG			G

Chapter One - Introduction

1.1 Colorectal cancer

Colorectal cancer (CRC) is the fourth most common cancer in the UK, with over 40,000 cases diagnosed each year. The overall lifetime risk of developing CRC is around 5%, with 85% of diagnosed cases seen in people over the age of 60 (Ballinger and Anggiansah, 2007). Despite advances in treatment and early screening methods dramatically reducing mortality rates by up to 50% in the last 40 years, approximately 16,000 people still die in the UK each year from the disease (Cancer Research UK, Bowel cancer statistics, 2010).

The rate of development of colorectal adenomas (CRA) and CRC is determined by an individual's exposure to a combination of environmental and genetic factors, although their influence on disease initiation and progression are not exclusive from one another (Kim and Milner, 2007). Current understanding surrounding environmental factors lists a diet high in heterocyclic amines from cooked red and processed meat (Martinez *et al*. 2007; Larsson and Wolk, 2006), obesity (Ning *et al.* 2010), sedentary lifestyle (Wolin *et al*. 2009), smoking (Liang *et al*. 2009) and alcohol intake (Giovannucci, 2004) as some of the risk factors. Many of these are considered to be part of an affluent Westernised lifestyle, the influence of which is mirrored in increased incidences in developing countries adopting said lifestyles (Curado *et al.* 2007). Inflammatory bowel diseases, including ulcerative colitis and Crohn's disease, have been highlighted as risk factors, with a third of deaths related to ulcerative colitis due to the development of CRC (Itzkowitz and Hapraz, 2004).

1.2 Inherited colorectal cancer

The strong heritable component associated with CRC is highlighted by the identification of multiple genetic syndromes. Advances in genetics have led to a better understanding of the underlying molecular dysregulation associated with the phenotypes shown in such conditions, leading to improvements in treatment and increased surveillance for both patients and their family members (Lynch *et al.* 2007).

CRC is typically divided into two sub groups; sporadic and familial (Fig. 1.1). The vast majority of CRC cases are believed to be sporadic, with existing genetic understanding accounting for around 12%. However, the uncharacterised familial risk of CRC is illustrated by twin and sibling studies, which suggest that genetics could account for up to 35% of cases (Lichtenstein *et al*. 2000). The importance of familial contribution to disease burden is further illustrated by the fact that having a first degree relative with the disease increases relative risk two fold. The estimated risk rises further when multiple family members are affected (odds ratio [OR] =4.25) and when an early age of diagnosis is implicated (OR=3.87; Johns and Houlston, 2001).

The so-called 'L-shape' distribution of allelic effects highlights the influence that certain variants have on complex traits such as CRC (Bost *et al*. 2001). The situation arises when a small number of variants with a relatively low minor allele frequency (MAF) have a dramatic effect on risk, whilst, on the contrary, a large number of variants with relatively large MAFs have a modest contribution. Hereditary CRC disorders, such as familial adenomatous polyposis (FAP) and hereditary non-polyposis colorectal cancer (HNPCC), are known to be caused by high penetrance alleles. However, with the emergence of genome wide association studies (GWAS), multiple common low penetrance loci across the genome have been shown to be significantly associated with disease risk, albeit with a small effect size.

1.2.1 High penetrance alleles

High penetrance is assigned to an allele if the presence of at least one of these alleles greatly increases the likelihood of a particular phenotype. These traits tend to be highly heritable and therefore easier to track and determine. In CRC, approximately 6% of all cases are attributable to these types of mutations (Jasperson *et al*. 2010; Patel and Ahnen, 2012; Table 1.1; Fig. 1.1).

A large proportion of the high penetrance, hereditary CRC syndromes are a result of inactivating mutations in tumour suppressor genes. Most loss of function mutations of tumour suppressor genes are recessive, although dominant in nature, and require loss of the second allele in order for a cell to become cancerous. In

	Disease	Contribution to CRC incidence	Gene	Pathway
	Familial adenomatous polyposis (FAP) and attenuated FAP (AFAP)	<1%	*APC*	Wnt signalling
	MUTYH-associated polyposis (MAP)	<1%	*MUTYH*	BER
	Hereditary non-polyposis colorectal cancer (HNPCC)	2-6%	*MSH2; MLH1; MSH6; PMS1; PMS2; MLH3; EPCAM*	MMR
	Polymerase proof-reading associated polyposis	<1%	*POLD1/POLE*	Various DNA repair pathways
Harmartomatous polyposis syndromes	Peutz-Jeghers syndrome (PJS)	<1%	*STKB11*	mTOR
	Juvenilla polyposis syndrome (JPS)		*SMAD4* and *BMPR1A*	TGFβ
	Cowden syndrome		*PTEN*	P13K/Akt/mTOR
	Hereditary mixed polyposis syndrome (HMPS)		*GREM1*	TGFβ

Table 1.1 – High penetrance hereditary CRC syndromes and their associated genes and pathway (BER = Base excision repair; MMR = Mismatch repair; TGF-β = Transforming growth factor-β; mTOR = Mammalian target of rapamycin). Polymerase proof reading associated polyposis is discussed in more detail in section 1.7.2.2.

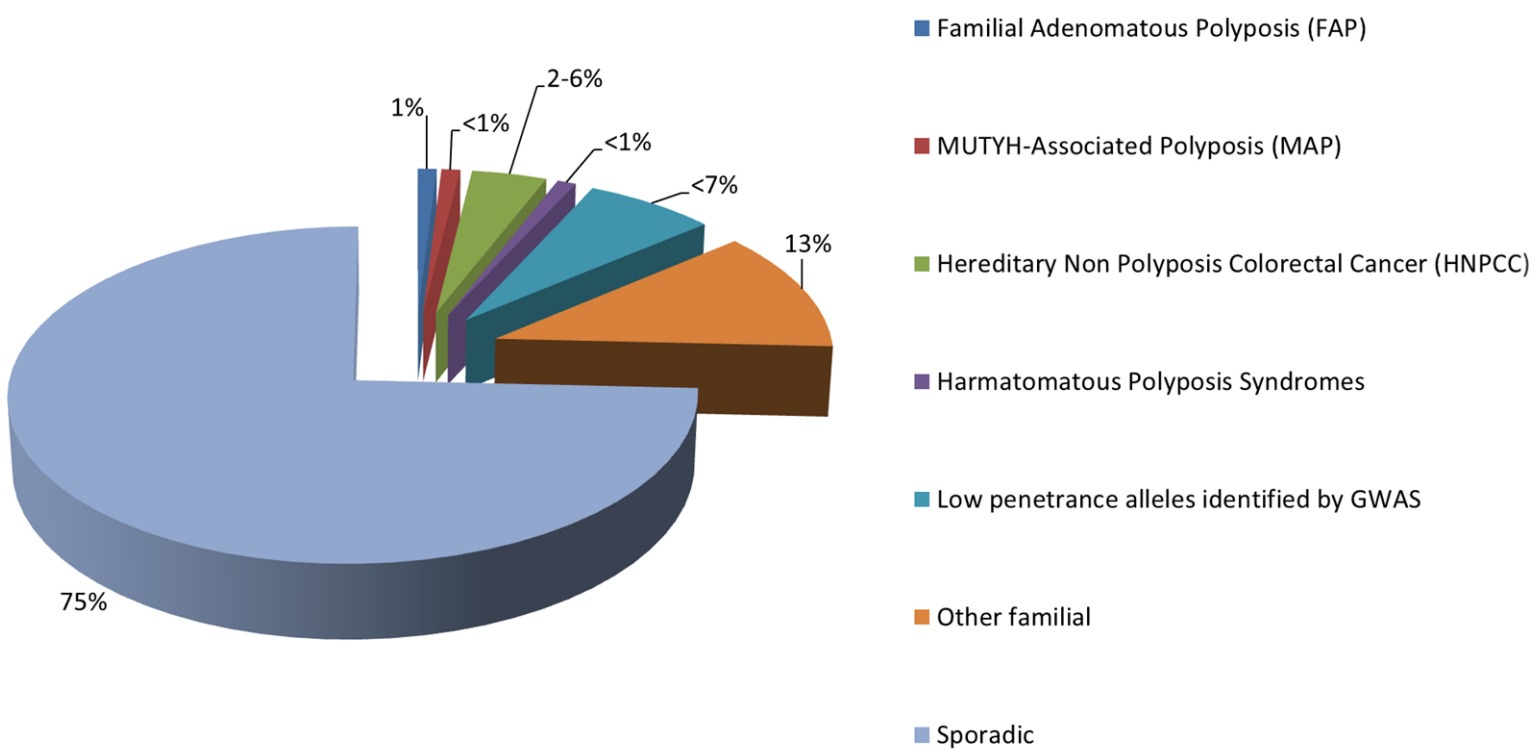


Figure 1.1 - Percentage contribution of known hereditary CRC syndromes to the overall incidences of CRC. A large proportion of cases (~75%) are believed to be sporadic in nature.

accordance with Knudson's two-hit hypothesis of tumour suppressor genes, an initial inherited mutation increases the likelihood of disease as a result of a greater probability of loss of the second allele in somatic cells (Knudson, 1971). In sporadic diseases, a somatic mutation on both alleles must occur (Fig. 1.2).

1.2.1.1 FAP

Accounting for less than 1% of CRC incidence, FAP (OMIM #175100) is an autosomal dominant disease characterised by the formation of hundreds to thousands of variably sized CRAs. It affects 1 in 5,000-10,000 of the population (Nagy *et al.* 2004). Left untreated, it carries an almost 100% risk of CRC usually presenting by the fourth decade of life, with the most common form of treatment being a full colectomy (Thomson, 1990; Galiatsatos and Foulkes, 2006; Half *et al.* 2009). FAP is also associated with allele dependant extra-colonic features, including congenital hypertrophy of the retinal pigment epithelium, dental abnormalities, epidermoid cysts and osteoma's. Additionally, there is an increased risk of thyroid and other endocrine, desmoid, duodeum, brain, liver and pancreatic cancers (Groen *et al*. 2008).

Attenuated familial adenomatous polyposis (AFAP) is a less aggressive form of the disease. It is characterised by the formation of tens to hundreds of CRAs, 69% risk of advancement to CRC, a later age of onset of CRC, and a lower burden of extra-colonic features (van der Luijt *et al.* 1995; Knudsen *et al*. 2003; Burt *et al.*2004).

Both FAP and AFAP are caused by germline mutations in the adenomatous polyposis coli *(APC)* gene. Located on chromosome 5q21-22 (Bodmer *et al*. 1987), it consists of 21 alternatively spliced exons and encodes a 312kDa functional protein (Fearnhead *et al*. 2001). FAP and AFAP can be caused by more than 300 different mutations in *APC*. Although these vary in type, over 90% result in a truncated form of the protein (Miyoshi *et al.* 1992a; Half *et al*. 2009). A large proportion of these mutations are seen in exon 15, the largest exon that contributes over 75% of the coding sequence (Beroud and Soussi, 1996). *APC* mutations of these kinds carry an almost 100% penetrance in carriers. In contrast, a nonsynonymous variant, Ile1307Lys, which is relatively common in the Ashkenazi Jewish population (6%), is thought to carry only a 20% penetrance (Lynch and de la Chappelle, 2003). This

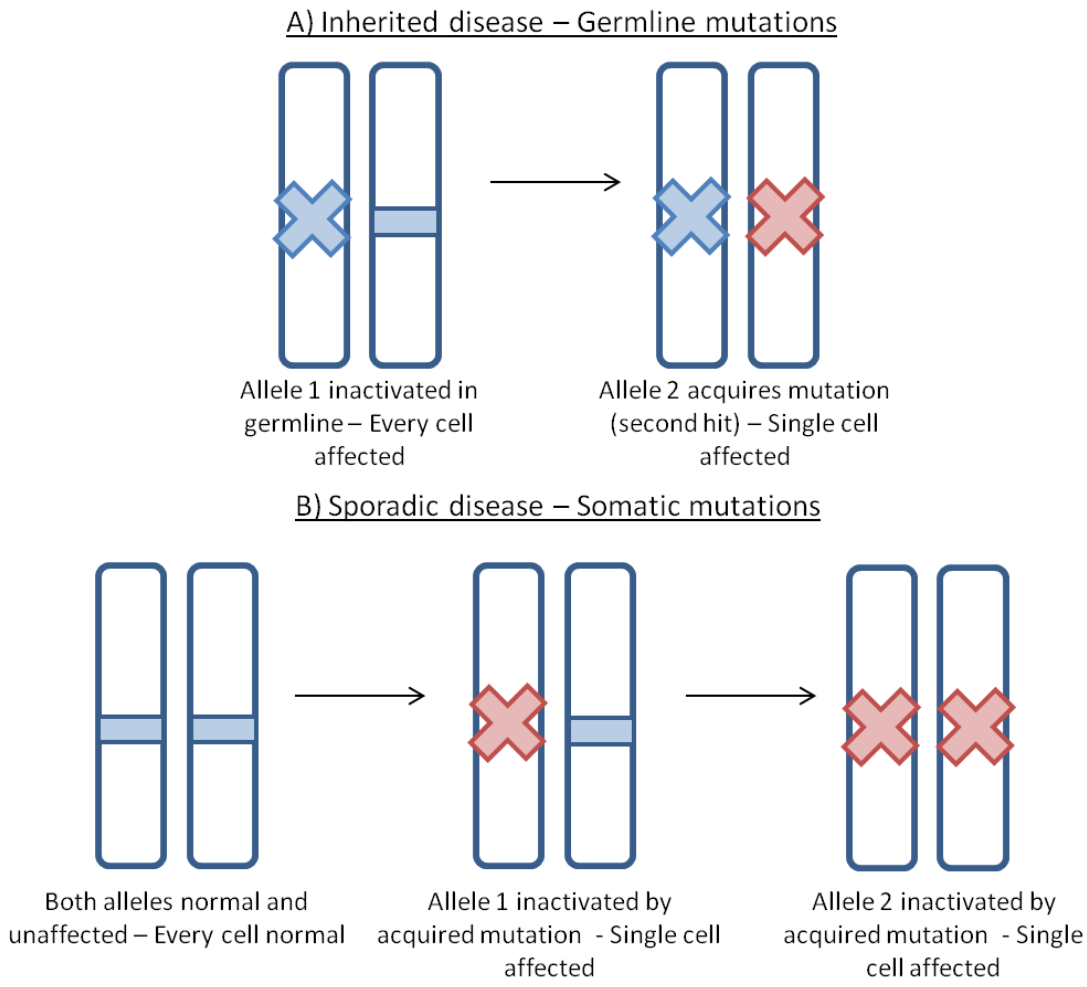


Figure 1.2 – Knudson's two hit hypothesis for loss of tumour suppressor function in tumourigenesis. A) In inherited disease, a mutation of one allele is inherited in every cell, whilst the second allele mutation is acquired in one cell. B) In sporadic disease, two normal alleles are inherited in every cell. One allele is inactivated by sporadic mutation, followed by a second sporadic mutation of the other allele, leading to inactivation of the gene. Loss of tumour suppressor function leads to cellular growth advantage and tumour progression (Knudson, 1985)

variant has been shown to create a hypermutable tract in *APC* that predisposes to somatic mutations (Laken *et al.* 1997). Genetic analysis of families exhibiting AFAP revealed that they had mutations resulting in a truncated form of APC similar to that seen in classical FAP. However, the majority of these mutations were located in the extreme 5' and 3' regions of the gene (before codon158 or after codon 1595) – something which is not common in the classical form of the disease (Spirio *et al.* 1993; Soravia *et al.* 1998).

The exact nature of the somatic 'second hit' has been shown to be highly dependent upon the 'first hit' of germline *APC* allele mutations seen in FAP patients (Lamlum *et al*. 1999), suggesting that there is an 'interdependence' of *APC* mutations that results in a cellular growth advantage (Cheadle *et al*. 2002). Sixty percent of somatic *APC* mutations occur in the 'mutation cluster region' which resides between amino acid 1281 and 1556 of exon 15 (Miyoshi *et al*. 1992b; Cheadle *et al*. 2002).

APC is a critical component of the Wnt signalling pathway, important for the intracellular control of cell growth and survival. Ultimately it is critical in the maintenance of the correct architecture of the colon via its regulation of key target genes (Bienz and Clevers, 2000). Following activation of the frizzled receptor by the Wnt ligand, inhibition of degradation of constitutively active β-catenin occurs as a result of phosphorylation and translocation of proteins key for its normal degradation (Klaus and Birchmeier, 2008). This allows for translocation of β-catenin to the nucleus and transcriptional activity of Wnt target genes to occur. In the absence of ligand, β-catenin degradation is controlled by phosphorylation of the protein by the so-called multi-protein 'destruction complex' of which APC plays a pivotal role (Huelsken and Behrens, 2002; Schneikert and Behrens, 2007). In situations where the destruction complex integrity is compromised, regardless of Wnt ligand binding, β-catenin is not degraded leading to excessive target gene activation (Morin *et al.* 1997; Korinek *et al*. 1997; Mann *et al.* 1999).

1.2.1.2 MUTYH-associated polyposis

MUTYH-associated polyposis (MAP; OMIM #604933) is an autosomal recessive disease, characterised by CRA growth similar to that seen in mild FAP or AFAP (Sieber *et al.* 2003). The development of hundreds of adenomas throughout the colon puts sufferers at a greater risk of carcinoma (Sampson *et al.* 2003). MAP was first identified through somatic analysis of the *APC* gene in patients exhibiting multiple CRA without a germline *APC* mutation. It revealed an excessive proportion of G:C→T:A transversions, resulting in an elevated number of truncation mutations in tumours. Transversion mutations of this kind are commonly due to tautomeric changes that occur as a result of oxidative damage to guanine, which leads to the production of the highly mutagenic 8-oxo-7,8-dihydro-2'-deoxyguanosine (8-oxo-G); a stereometric alteration that readily misbinds with adenine. Guanine is at particular risk of oxidative damage due to its low redox potential (Neeley and Essigmann, 2006). During the repair of oxidative damage by base excision repair (BER; discussed in section 1.3.2), it is typically the role of the DNA glycosylases, *OGG1* and *MUTYH*, to remove 8-oxo-G and the mispaired adenine, respectively. Germline screening of patients revealed biallelic mutations in *MUTYH*, in particular the function impairing nonsynonymous variants Tyr179Cys and Gly396Asp (Al-Tassan *et al.* 2002; Jones *et al*. 2002).

MUTYH is located on chromosome 1p34.1, consists of 16 exons and encodes a protein 535 amino acids in length. Tyr179Cys and Gly396Asp account for approximately 73% of all mutations seen in MAP (Cheadle and Sampson, 2007). Biallelic inactivation of *MUTYH* is the hallmark of MAP, increasing CRC risk by 93-fold, with conflicting reports stating a modest, if any, increased risk observed in monoallelic carriers (Farrington *et al*. 2005; Balaguer *et al.* 2008).

1.2.1.3 HNPCC

HNPCC (OMIM #120435) is an autosomal dominant disease characterised by the formation of many different cancers. One of the most common sites of carcinoma is the colon, with an average age of cancer onset of 44 years (Lynch and de la Chapelle, 1999). Tumours occur more commonly in the proximal colon, can grow synchronously or metachronously and transform up to 7 years more rapidly than sporadic cancers (Jang and Chung, 2010). It carries an 80% risk of CRC in the

patient's lifetime. Regular surveillance after partial colectomy is required since approximately 16% of patients that undergo the procedure will develop a secondary tumour within a ten year period (Nagengast *et al.* 2002). As the most common form of hereditary CRC, it accounts for approximately 2-6% of all CRC cases (Lynch *et al*. 2006).

Diagnosis of HNPCC in families can be subdivided into two catergories using the Amsterdam criteria (AC; Vasen *et al.* 1999); AC1, in which hereditary CRC is predominant and ACII, where multiple other cancer types are observed. These include endometrial, small bowel, renal pelvis and ureter cancers. A clinically distinct form of ACII, Muir Toirre syndrome, is characterised by an elevated risk of sebaceous skin cancers (Hall *et al.*1994). Additionally, Turcot syndrome is a variant associated with an increased risk of brain tumours, notably medullablastomas (Hamilton *et al.* 1995).

Inherited defects in key genes in the mismatch repair (MMR; Section 1.3.1) pathway have been shown to be fundamentally responsible for HNPCC. Up to 90% of patients diagnosed have *MSH2* or *MLH1* inactivating mutations (Peltomäki, 2005). Also implicated in disease etiology are mutations in *MSH6, PMS1, PMS2* and potentially *MLH3*, albeit at a lower frequency (Wu *et al*. 2001; Jasperson *et al.* 2010). Although the protein product is not involved in the MMR process, mutations in epithelial cell adhesion molecule (*EPCAM*) are thought to influence expression of *MSH2* and have been proposed to have a role in less than 1% of cases (Ligtenberg *et al.* 2008; Kovacs *et al*. 2009).

1.2.1.4 Harmartomatous polyposis syndromes

Harmartomatous polyps are benign malformations of the gastrointestinal tract (Calva and Howe, 2008). Although the cellular composition of the polyps is normal, the architecture is disordered and chaotic which results in the presence of a variety of different tissues. Although benign, these abnormalities increase the chance of malignancy in sufferers (Gammon *et al*. 2009).

The harmartomatous polyposes are a heterogeneous group of inherited autosomal dominant conditions that are characterised by an abundance of harmartomatous polyps along the gastrointestinal tract and a marked increased risk

of CRC. Accounting for less than 1% of overall CRC cases, they are collectively rare (Zbuk and Eng, 2007).

1.2.1.4.1 Peutz Jegher syndrome

Peutz Jegher syndrome (PJS; OMIM #175200) is an autosomal dominant disease that predisposes to hamartomatous polyps along the gastrointestinal tract. Approximately 30% of PJS sufferers will develop polyps in the colon, with an estimated relative risk of 84 for progression to carcinoma (Giardiello *et al.* 2000). Genetic studies have implicated the role of serine/threonine kinase 11 (*STK11/LKB1*) at chromosomal location 19q13.3 in the development of the disease. Germline mutations of the gene were identified in approximately 50-90% of patients with PJS. Approximately 70% are truncating or nonsynonymous variants, with the other proportion being attributed to large deletions (Aretz *et al*. 2005).

There is dysregulation of the mammalian target of rapamycin (mTOR) pathway in PJS sufferers. Normally, STK11 phosphorylates AMP activated protein kinase (AMPK) in response to low energy levels. This protein is key in the activation of tuberin, which in turn inhibits mTOR, controlling cellular growth by reducing S6K and 4EBP1 phosphorylation (Corradetti *et al*. 2004). Additionally STK11 may play a role in p53 mediated cell cycle arrest, with low energy levels directly stalling cell cycle progression. The disordered architecture of the harmartomatous polyps seen in PJS supports this; patients with malfunctioning STK11 cannot reduce normal cell growth in reduced energy situations (Shaw, 2006).

1.2.1.4.2 Juvenile polyposis syndrome

Despite appearing outwardly similar to other harmartomatous polyps, juvenile polyposis syndrome (JPS; OMIM #174900) polyps are histologically very different, microscopically appearing as mucous filled glands. Almost all polyps occur in the colon or the rectum, with a 20% likelihood that these will progress to malignancy (Handra-Luca *et al*. 2005).

Two genes from the transforming growth factor (TGFβ) pathway have been implicated in JPS; mothers against decapentaplegic, group 4 (*SMAD4*) at 18q21.1 (Howe *et al*, 1998; Houlston *et al*, 1998) and bone morphogenetic protein receptor, type 1A (*BMPR1A*) at 10q22.3 (Howe *et al*. 2001). Both account for approximately

20% of cases each (Howe *et al*. 2004). The TGFβ pathway is important in the control of the cell cycle. Additionally mutations in endoglin (*ENG*) in the same pathway have been implicated in development of JPS in early childhood (Sweet *et al*. 2005), although its contribution towards disease aetiology is of debate (Howe *et al*. 2007).

1.2.1.4.3 Cowden syndrome

Cowden syndrome (OMIM #158350) is a rare autosomal dominant disease characterised by the formation of multiple harmartomatous polyps along the gastrointestinal tract, with colonic polyps present in up to 90% of cases.

Cowden syndrome is caused by loss of function mutations in the tumour suppressor gene phosphatase and tensin homolog (*PTEN*) in up to 80% of cases. PTEN is a phosphatase protein involved in the regulation of many key signalling pathways through dephosphorylation of phosphatidylinositol-3,4,5-trisphosphate (PIP_3; Blumenthal and Dennis, 2008).

1.2.1.4.4 Hereditary mixed polyposis syndrome

Hereditary mixed polyposis syndrome (HMPS, OMIM #601228) is associated with a predisposition to harmartomatous juvenile polyps. There is an elevated likelihood of CRA and CRC without any extra-colonic symptoms, something which is typically evident in other polyposis syndromes. In the molecular genetics of the disease, identification of a shared haplotype and a possible predisposition locus, so called colorectal adenoma and carcinoma (*CRAC1)* on chromosome 15, followed linkage analysis of two Jewish families (Tomlinson *et al*. 1999; Jaeger *et al*. 2003). Following fine mapping of the region in two families, three genes were identified; gremlin *(GREM1),* secretogranin *(SGNE1)* and formin *(FMN1).* Analysis of the region using a custom array identified a heterozygous, 40kb single copy duplication in all affected individuals. The duplication involves a region spanning the latter part of *SGNE1* to just upstream of *GREM1*. Ectopic overexpression of GREM1 in colorectal crypt cells was observed, with significantly elevated expression of the duplicated allele (Jaeger *et al*. 2012).

1.2.2 Low penetrance allele

Since the majority of inherited CRC cases occur without any known underlying genetic reason, it was proposed that the remaining heritable component could be accounted for by common or rare, low penetrance variants (Fearnhead *et al.* 2005). Low penetrance is attributed to an allele if the effect on phenotype is small, although their contribution to disease burden could be substantial on account of their relative frequency in the general population.

1.2.2.1 'Common disease, common variant' model

The 'common disease, common variant' model is one that helps to explain the variation that arises in many complex diseases such as CRC. In this model, the individual variant risk is relatively small; the OR seen are typically between 1.2 and 1.5 (Bodmer and Bonilla, 2008). However, due to the fact that such alleles are usually relatively common in the population and can more than likely interact with one another in a polygenic manner, they have significant impact on disease likelihood.

1.2.2.1.1 GWAS

The completion of the HapMap project meant that knowledge of linkage disequilibrium (LD) that captures variation across the genome has been made readily available (International HapMap Consortium, 2003). In addition to this, the production of large scale genotyping platforms means that a large number of variants can be screened in thousands of samples at an affordable cost, with quick turnaround. Ultimately, these large scale, hypothesis free, multi stage case control studies have directly identified 12 CRC susceptibility alleles to date.

GWAS are limited in their ability to detect novel variants by several constraints. Of most importance is the difficulty in acquiring the vast amount of samples needed to supply enough power to detect these modest effect sized variants. Pooling of data from multiple cohorts allows for increases in study power and such meta-analysis have identified 8 additional variants, bringing the total number of CRC susceptibility alleles to 20 in cohorts of European ancestry (Table 1.2). Although all have modest contributions to overall risk, with OR ranging from 1.07 to1.28 (Kilpivaara and Aaltonen, 2013), collectively they could account for up to

7% of the familial risk of CRC (Dunlop *et al*. 2012b). Additional problems of GWAS include the need to avoid population stratification by ruling out multiple ethnic groups, as well as the need for validation due to the high rate of false positive associations seen in such studies.

1.2.2.2 'Common disease, rare variant' model

Rare variants have been shown to play roles in the phenotype of complex diseases (Pritchard, 2001). Detection of rare variants is more commonly carried out through candidate searches of genes implicated in disease etiology. Research has indicated a role of rare variants in the Wnt signalling genes *CTNNB1* and *AXIN1,* and the MMR genes *MLH1* and *MSH2* in the collective contribution to a modestly increased risk of CRA development (Fearnhead *et al.* 2004). Additionally, Azzopardi *et al* (2008) have shown that multiple rare but collectively common variants in *APC* contribute towards the development of CRA. Despite previous conflicting reports for the role of the nonsynonymous variant Glu1317Gln in *APC* in the predisposition to CRA, tumourigenesis and CRC (Frayling *et al.* 1998; Lamlum *et al.* 2000; Popat *et al.* 2000; Gismondi *et al.* 2002 Hahnloser *et al.* 2003;) researchers detected a role in CRA predisposition in patients characterised as 'non-FAP, non-MAP'. Additionally they reported that, following exclusion of this variant as well as another low penetrance variant, Ile1307Lys in *APC*, significantly more of these patients carried various other rare nonsynonymous *APC* variants, suggesting a low penetrance effect of these rare alleles (Azzopardi *et al*, 2008).

1.3 DNA repair and cancer

The ability of cells to repair DNA damage is crucial for the integrity and maintenance of genetic material in all organisms, and ultimately in survival. An individual cell is subjected to a plethora of DNA damaging events; up to a million events occur in a single cell each day (Lodish *et al.*2000). DNA damage has the ability to modify the coding sequence of DNA which, if not repaired, can lead to the development of cancer by mutational activation of proto-oncogenes and inactivation of tumour suppressor genes (Hoeijmakers, 2001).

Variant	Loci	Gene	Role/Pathway	Reference
rs6983267	8q24.21	*MYC*	Wnt signalling	(Tomlinson *et al.* 2007; Zanke *et al.* 2007)
rs16969681/ rs11632715†	15q13.3	*GREM1*	TGF-β signalling	(Jaeger *et al.* 2008; Tomlinson *et al.* 2011)
rs4939827	18q21	*SMAD7*	TGF-β signalling	(Broderick *et al*. 2007; Tenesa *et al*. 2008)
rs3802842	11q23			(Tenesa *et al*. 2008)
rs16892766	8q23.3	*EIF3H*		(Tomlinson *et al*. 2008)
rs10795668	10p14			(Tomlinson *et al*. 2008)
rs4444235	14q22.2	*BMP4*	TGF-β signalling	(Houlston *et al.* 2008)
rs9929218	16q22.1	*CDH1*	Cell-cell adhesion	(Houlston *et al.* 2008)
rs10411210	19q13.1	*RHPN2*	TGF-β signalling	(Houlston *et al.* 2008)
rs961253	20p12.3	*BMP2*	TGF-β signalling	(Houlston *et al.* 2008)
rs6691170/ rs6687758†	1q41	*DUSP10*	MAPK signalling	(Houlston *et al.* 2010)
rs10936599	3q26.2	*MYNN*		(Houlston *et al.* 2010)
rs11169552/ rs7136702†	12q13	*LARP4*		(Houlston *et al.* 2010)
rs4925386	20q13.33	*LAMA5*	Cell migration and localisation	(Houlston *et al.* 2010)
rs1321311	6p21	*CDKN1A*	Cell cycle and apotosis	(Dunlop *et al.* 2012a)
rs3824999	11q13.3	*POLD3*	DNA repair	(Dunlop *et al.* 2012a)
rs5934683	Xp22.2	*SHROOM2*		(Dunlop *et al,*.2012a)

Table 1.2 – The 20 variants identified through GWAS at 17 genomic loci associated with CRC risk in Caucasian populations, with the respective genes and pathways (if applicable). †Two variants associated at locus.

There are currently 168 known proteins that are involved in the various pathways of DNA repair (Wood, 2005). Each pathway consists of multiple steps, with specialised proteins involved in the diverse roles required for repair of the different forms of damage that can occur (Lindahl, 1993), although there is considerable overlap between proteins from different pathways. Mutations in DNA repair genes have been shown to have a role in various hereditary cancer predisposing syndromes, with multiple pathways shown to have a role in hereditary forms of CRC (Loeb, 2003; Milanowska *et al.* 2011; Negrini *et al.* 2010; Table 1.3).

1.3.1 MMR pathway

Microsatellites (also known as tandem simple sequence repeats) are short, repeating sequences that are between 1-6 nucleotides in length and found throughout the genome, though are more common in non-coding regions (Beckman and Weber, 1992). They are polymorphic in nature, variable in length between individuals, although are homogenous in all cells in an individual (Boland and Goel, 2010). Microsatellite instability (MSI) is the result of increased variability in the length and frequency of microsatellite repeats throughout the genome commonly due to replication slippage as a result of inefficient binding of DNA polymerases during synthesis (Eisen, 1999; Kunkel, 2004). This results in the formation of insertion-deletion loops (IDL) or base-base mismatches (Schlötterer and Harr, 2001), which broadly deems to have coding effects that can alter expression and/or function of a gene (Nelson and Warren, 1993). It is predominantly the role of the post-replicative MMR pathway to recognise and repair such damage.

Recognition of damage occurs by two heterodimeric complexes consisting of MSH2-MSH6 (hMutSα) or MSH2-MSH3 (hMutSβ), with the former being primarily responsible for recognition of mismatches and single base IDL, and the latter for recognition of all IDLs (Li, 2008). A second heterodimeric complex (hMutL) has the ability to bind to the hMutS complexes in order to recognise damage as well as recruit additional machinery to the area of damage; either MLH1 and PMS2 (hMutLα), MLH1 and PMS1 (hMutLβ) or MLH1 and MLH3 (hMutLγ; Jascur and Boland, 2006). MLH1 carries out incision of the damaged strand. Following recognition of damage, proliferating cell nuclear antigen (PCNA) is loaded onto the DNA by replication factor C (RFC). PCNA is responsible for recruiting exonuclease

Gene	Pathway	Disease (Cancer)
ALKBH3	Direct reversal of damage	Hereditary prostate cancer
ATM	HR	Ataxia-telangiectasia (Lymphomas; leukaemia; breast cancer)
BLM	HR	Bloom syndrome (Multiple; earlier age of onset)
BRCA1, BRCA2, CHEK2 and *RAD51*	HR	Hereditary breast cancer; hereditary ovarian cancer; hereditary prostate cancer
FANCA, -B, -C, -D1 (BRCA2), -D2, -E, -F, -G, -I, -J, -L, -M, -N, -P and *-Q*	ICL repair and HR	Fanconi anaemia (Multiple)
LIG4	NHEJ	LIG4 syndrome (Leukaemia)
MLH1, MLH3, MSH2, MSH3, MSH6, PMS1* and *PMS2	**MMR**	**HNPCC**; Turcot syndrome; Muir Torre syndrome (**CRC**; Endometrial cancer; Small bowel cancer; Renal pelvis cancers; Uterine cancer; Brain cancer; Sebaceous skin cancers)
MUTYH	**BER**	**MAP**
NBN	HR	Nijmegen breakage syndrome (Non-Hodgkin lymphoma; medulloblastoma; glioma; rhabdomyosarcoma)
POLD1/POLE	**DNA repair and synthesis**	**CRA and CRC**
POLH	Translesion synthesis (after UV damage)	Xeroderma pigmentosum, variant (Skin cancer)
RECQL4	HR	Rothmund-Thomson syndrome (Osteosarcoma; skin cancers)
WRN	Telomere maintenaince; HR	Werner syndrome (Various cancers)
XPA, -B, -C, -D, -E, -F, -G	NER	Xeroderma pigmentosum (Skin cancer)

Table 1.3 – DNA repair genes, pathways they are involved in and associated hereditary cancer predisposition syndromes. Pathways associated with CRA and CRC syndromes are given in bold. (BER = Base excision repair; HR = Homologous recombination; ICL = Interstrand crosslinks; NER = Nucleotide excision repair; NHEJ = Non-homologous end joining; UV = Ultraviolet)

(EXO) to degrade the excised strand. Following this, polymerase δ (POLD) and DNA ligase I (LIG1) are recruited to accurately repair the excised strand and repair nicks once DNA synthesis has finished (Vilar and Gruber, 2010; Kunkel and Erie, 2005; Fig. 1.3)

1.3.1.1 MMR gene mutations and cancer

Deficiencies of the MMR system as a result of function impairing mutations predispose patients to various cancers (Lynch and Lynch, 1979; Section 1.2.1.3). HNPCC is a genetically heterogenous disease and multiple genes in the MMR pathway can be mutated. The majority of mutations seen in HNPCC syndromes are observed in *MLH1* and *MSH2,* although many other genes from the MMR pathway have been implicated including *MSH6, PMS1, PMS2* and *MLH3* (Nicolaides *et al.* 1994; Wu *et al.* 2001; Liu *et al.* 2003; Hendriks *et al.* 2004). Before the age of 70, overall risk of cancer in MMR gene mutation carriers is 91% for men and 69% for women. The frequency of different cancers differs between the sexes, with men at a much greater risk of developing CRC (Dunlop *et al.* 1997).

In accordance with Knudson's two-hit hypothesis, somatic inactivation of the second MMR allele in HNPCC patients leads to the formation of a mutator phenotype characterised by an increased rate of MSI; however this does not directly cause tumour growth (Parsons *et al.* 1993). Genes with repeat sequences are commonly affected by MSI in HNPCC patients (Duval and Hamelin, 2002). Such genes include TGFβ receptor 2 (*TGFβR2*)*, AXIN2, β-catenin,* BCL2-associated X (*BAX1*)*, MSH3* and *MSH6* (Wrana *et al*.1994; Lu *et al*. 1995; Malkhoysen *et al*. 1996; Rampino *et al*. 1997; Liu *et al*. 2000; Shitoh *et al*. 2001).

The genetic and allelic heterogenity associated with HNPCC and the wide distribution of mutations throughout genes can make germline screening difficult (Peltomäki and Vasen, 1998). Over 90% of HNPCC patients display MSI, with somatic MSI status at particular repeat sequences commonly used in the diagnosis of germline mutations (Boland *et al.*1998; Lamberti *et al.* 1999). The Bethesde criteria form a comprehensive set of guidelines in the diagnosis of patients with expected HNPCC that should be sent for genetic MSI analysis (Umar *et al.* 2004).

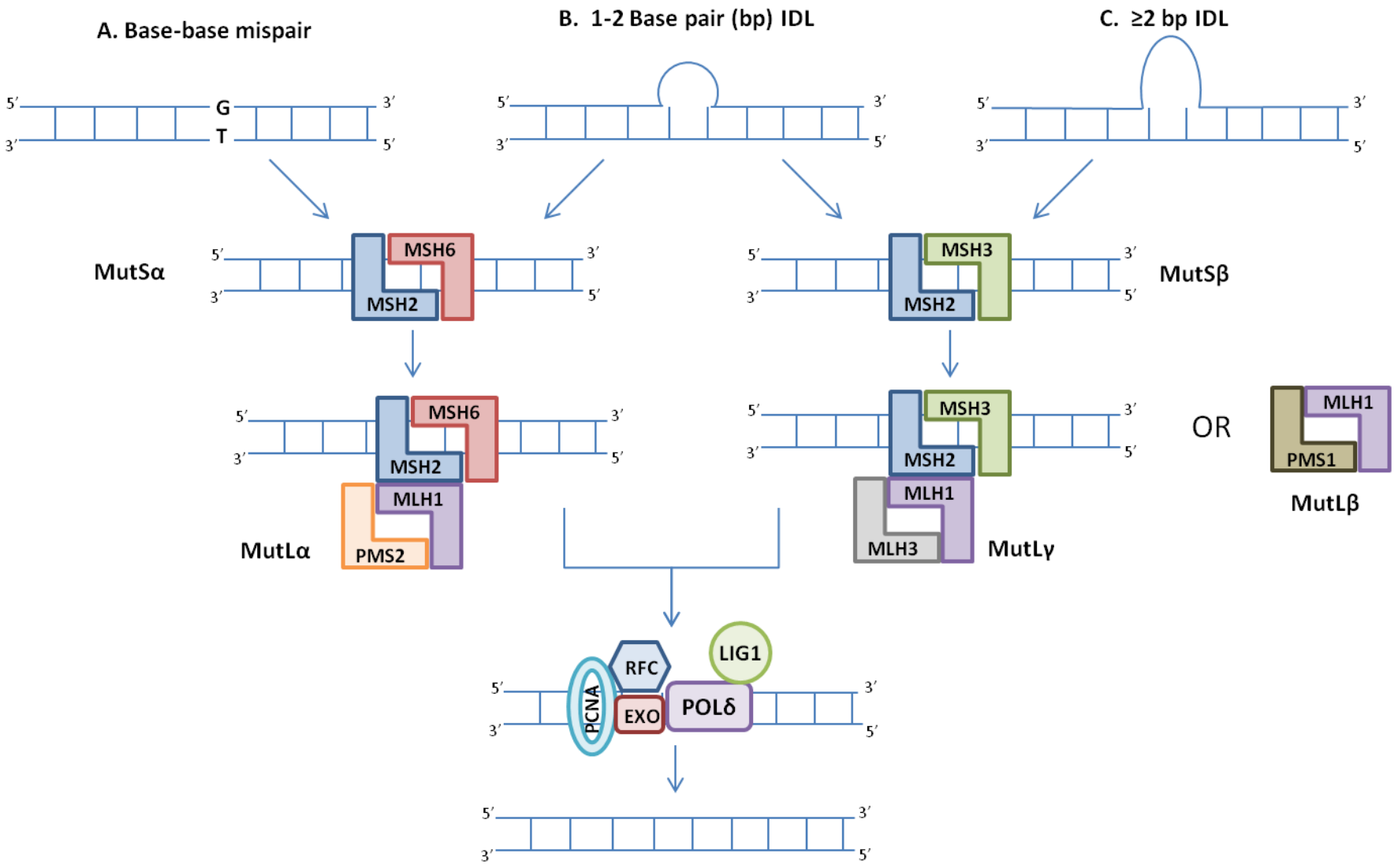


Figure 1.3 – Involvement of the MMR pathway in the repair of various forms of DNA damage.

1.3.2 BER pathway

DNA damage as a result of oxidative stress has been shown to have an important role in the development of degenerative syndromes, such as cancer and aging (Hoeijmakers, 2009). It has been proposed that oxidative stress could be responsible for up to half of all cancers (Beckman and Ames, 1997). The primary source of oxidative damage is from reactive oxygen species, created through both endogenous and exogenous sources (David *et al.* 2007). It is the role of the multi-step BER pathway to remove and repair such damage, as well as repair other forms of damage such as abasic (AP) sites, alkylation and deamination to maintain the integrity of DNA (Lindahl and Wood, 1999).

Following single base DNA damage, DNA glycosylases recognise and initiate repair by excising the damaged base. If monofunctional, the glycosylase removes the base through hydrolysis of the N-glycosidic bond. This results in an AP site, which is incised to form a single strand break (SSB) by apurinic/apyrimidinic endonuclease (APEX1). This leaves a 5'-deoxyribose 5'-phosphate residue (dRP) and a normal 3'hydroxyl (3'OH) group. DNA polymerase β (POLB) is involved in removal of the dRP overhang via the proteins integral lyase activity If bifunctional, the DNA glycosylase first removes the base and then incises the phosphodiester DNA backbone (Fromme *et al.* 2004). Depending on the glysocylase involved and the group left at the 3' end of the break, either APEX1 or polynucleotide kinase 3' phosphatase (PNKP) then processes the strand break (Wallace *et al.* 2012). In short patch repair, POLB repairs the damaged base (Matsumoto and Kim, 1995). X-ray repair cross complementing, 1 (XRCC1) acts as a scaffold protein for DNA ligase III (LIG3; Vidal *et al*. 2001), which then seals the SSB. In long patch repair, either POLB, POLD or polymerase ε (POLE) elongate 2-12 nucleotides from the 3' incision site to create a flap (Dianov *et al.* 2003). Through the action of flap endonuclease (FEN1) this is removed (Liu *et al.* 2005) with the help of PCNA and poly (ADP ribose) polymerase (PARP1) to aid strand displacement. The strand is then ligated by LIG1 (Fig. 1.4; Xu *et al*. 2008).

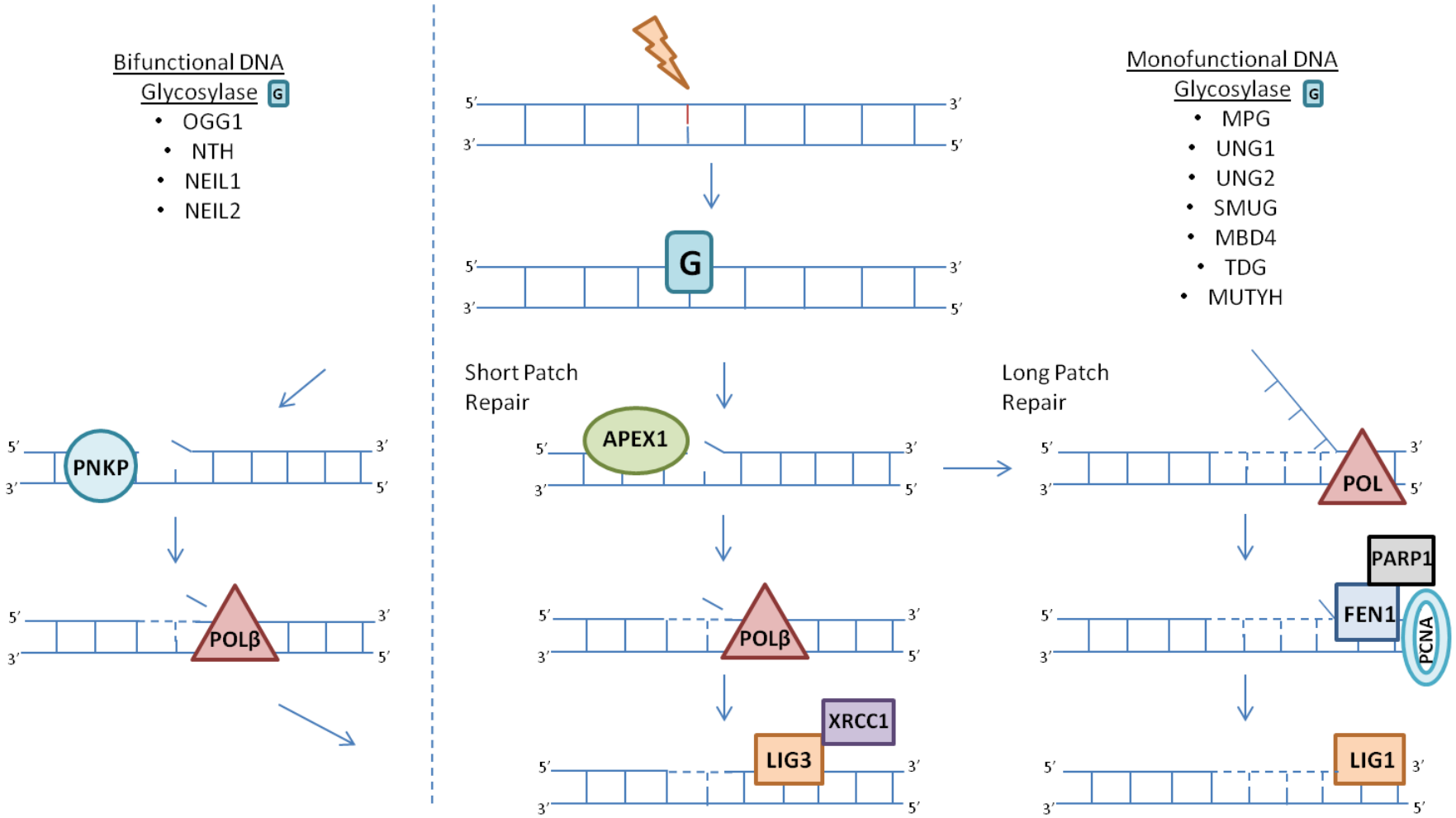


Figure 1.4 – The BER of DNA damage is initiated by the recognition and removal of the damaged base by either a monofunctional or bifunctional glycosylase to create an AP site. This can also occur spontaneously as a result of hydrolysis (Dianov *et al.* 2003). *Either the long or short patch repair pathways will then repair the area, dependant on the type of damage and glycosylase recruited (adapted from* Xu *et al.* 2008).

1.3.2.1 BER gene mutations and cancer

Biallelic mutations in the DNA glycosylase gene, *MUTYH*, have been shown to predispose to the familial CRA condition, MAP (Section 1.2.1.2). Over 30 different mutations of *MUTYH* have been observed in patients with MAP (Wallace *et al.* 2012).

1.3.3 NER pathway

NER is involved in the removal of bulky adducts from DNA that cause distortion of the double helix, hindering replication and transcription. It is typically involved in the repair of ultraviolet B (UV) photoproducts in the forms of cyclobutane pyrimidine dimers (CPD) and 6-4 photoproducts (6-4,PP; Leibeling *et al*. 2006), although it is also involved in the repair of other forms of bulky adducts that occur as a result of exposure to a range of environmental and chemical sources (Gillet and Schärer, 2006).

NER consists of two pathways that differ in the way that the DNA damage is recognised. If the region damaged is protein coding, transcription coupled (TC-NER) repair will occur when RNA polymerase II (RNA pol II) stalls at damaged regions. Cockayne syndrome group A and B (CSA and CSB) are recruited to the stall where they are involved in processing of the damage (Fousteri and Mullenders, 2008). If the damage occurs in a non-coding region of DNA, global genomic repair (GG-NER) is implemented via recognition by the xeroderma pigmentosum, group C (XPC) - HR23B complex and the DNA damage binding proteins, 1 and 2 (DDB1 and DDB2). The proteins involved are dependent on the damage caused and the extent of the distortion of the DNA helix (Hanawalt, 2002; Diderich *et al.* 2011). It is the role of CSB and XPC-HR23B to recruit the ten sub-unit basal transcription factor complex (TFIIH) to the area of damage. This complex includes the helicases XPB and XPD, which facilitates DNA strand unwinding in the presence of ATP (Coin *et al.* 2007). This allows for binding of XPA, replication protein A (RPA) and XPG. XPA ensures the area is damaged, whilst also having roles alongside RPA in protecting the undamaged single stranded DNA and recruiting the XPF and excision repair cross complementing, group 1 (ERCC1) complex. This 5’ endonuclease complex functions, alongside the 3’ incision by XPG, to incise and release the damaged strand (Staresincic *et al.* 2009). This allows for recruitment of repair machinery to

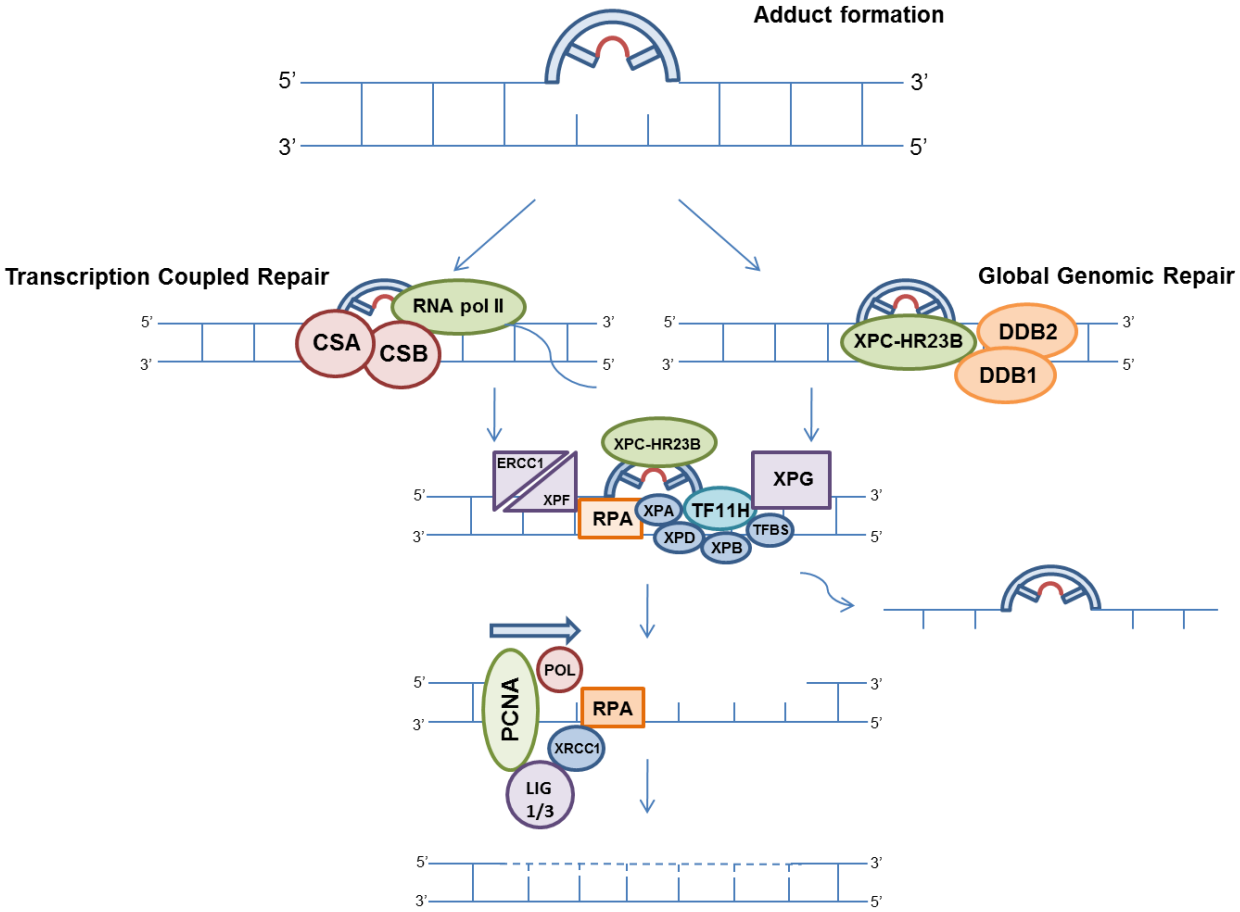


Figure 1.5 – The two branches of the NER pathway. NER functions in the repair of helix distorting lesions. If in a coding region of the genome, TC-NER will occur as a result of RNA pol II stalling. If elsewhere, helix distorting lesions are identified by XPC-HR23B, DDB1 and DDB2, dependent on the damage caused, as part of the GG-NER pathway. Following excision of the damaged strand, repair machinery is recruited. This includes POLD, POLE or polymerase κ (POLK; collectively POL in diagram; Ogi and Lehmann, 2006; Gillet and Scharer, 2006)*, as well as either XRCC1-LIG1 or LIG3 dependant on the stage of the cell cycle a cell is in, respectively* (Moser *et al.* 2007).

the area of excised damage where it carries out repair and ligation, using the undamaged strand as a template (Fig. 1.5).

1.3.3.1 NER gene mutations and cancer

Xeroderma pigmentosum (XP) is an autosomal recessive disorder characterised by extreme sensitivity to UV light and a 1000 fold increased risk of skin cancers due to ineffective repair and accumulation of UV induced DNA damage (Kraemer *et al.* 1987). Skin cancers typically develop 50 years earlier than the general population, with commonly seen lesions including squamous and basal cell carcinomas, and melanomas (Kraemer, 1997). Internal cancer risk is also elevated, albeit to a lesser degree (Kraemer *et al.* 1984). There are currently 8 known complementation groups of XP, all exhibiting similar phenotypes but with varying degrees of sensitivity. Some complementation groups also exhibit signs of neurological degeneration, with 20-30% of all patients' exhibiting symptoms (de Boer and Hoeijmaker, 2000; Anttinen *et al.* 2008). It is more common in Japanese populations, with approximately 1: 40,000 people affected compared to 1:1,000,000 in Western populations (Bhutto and Kirk, 2008).

Seven of the known complementation groups of XP (XPA through to XPG) are as a result of function impairing mutations in NER pathway genes, with the eighth complementation group, XPV, the result of mutations in the replicative bypass polymerase η (*POLH*). In XPV, the NER pathway performs normally; rather it is the inability to carry out DNA replication past regions with UV damage that results in the characteristic XP phenotype (Masutani *et al.* 1999).

1.3.4 Double strand break repair

One of the most detrimental forms of DNA damage are double strand breaks (DSB). DSB are formed following treatment with ionising radiation, X-ray or as a result of chemical damage. They are also formed following replication over a single stranded break, in the repair of interstrand crosslinks (ICL) and following collapse of stalled replication forks. Double strand ends that have become separated are liable to move away from one another. This can make repair difficult and also means there is the opportunity for recombination at other erroneous regions of the genome resulting in chromosome instability (Hoeijmakers, 2001). Chromosomal instability

formed in this way has been shown to be important in the early stages of tumourigenesis (Bartkova *et al.* 2005). There are two pathways involved in the repair of DSB; homologous recombination (HR) and non-homologous end joining (NHEJ).

Which pathway acts to repair DSB is highly dependent on the nature of the break and at which cell cycle stage the affected cell is in. Whilst NHEJ is faster than HR and can occur throughout all stages of the cell cycle, it is a mutagenic process in which split ends are directly ligated (Mao *et al.* 2008; Takata *et al.* 1998). However, HR can only take place in cell cycle phases when the homologous chromosomes are in close proximity. In all other stages HR could lead to dangerous chromosomal translocation as a result of unsuitable selection of homologous regions within similar repetitive sequences in other chromosomes (Lieber *et al.* 2003). There is also a risk of loss of heterozygosity as a result of the HR process conferring mutations from the homologous chromosome (Alexander *et al.* 2001).

1.3.4.1 HR

HR functions in repairing DSB by conferring the correct genetic material from an undamaged strand with which it shares sequence homology, typically the homologous chromosome. It is considered an error free mechanism of DNA repair that occurs in the S and G1 phase of the cell cycle. Various proteins throughout the pathway help to regulate the cell cycle in order to ensure that HR only takes place during these phases to guarantee that repair is carried out safely (Lisby *et al*. 2004). In the main stages of HR, the meiotic recombination 11 (MRE11), nibrin (NBS1) and RAD50 (MRN) nuclease complex, the CTBP interacting protein (CtIP) nuclease, and the bloom syndrome protein (BLM) helicase, are directed to the area of damage (Sartori *et al*. 2007; Mimitou and Symington, 2009; Ouyang KJ *et al.* 2009). Together they process the ends of the DSB to expose the 3’ ends of the strands to create a single strand overlap for efficient recombination via strand cross-over (Wyman and Kanaar, 2006). During this time RAD51, with the aid of RAD52 and BRCA2, displaces RPA that has bound to single stranded DNA. RAD51 is a recombinase that is key in the identification of highly homologous sequences and in guidance of the exposed single strand (Baumann *et al.* 1996). The damaged strand is directed for exchange with a highly homologous sequence, forming a D loop, allowing for a polymerase to use the error free strand to replicate from the area of damage (Scully

*et al.*1997; Fig. 1.6). There are various alternative pathways from this point which include double strand break repair (DSBR) and synthesis-dependant strand annealing (SDSA). In DSBR, the second double strand end forms a Holliday junction. LIG1 then ligates the two ends and the Holliday junctions formed by crossover are cleaved (Holliday, 1964). In SDSA, the newly synthesised strand is displaced from cross over where it is re-ligated. This strand then acts as a template for synthesis and ligation of the other damaged strand (Sung and Klein, 2006). As well as the repair of DSBs, HR is also involved in telomere maintenance, cell cycle control, repair at stalled replication forks and control of meiotic chromosome segregation (Sung and Klein, 2006).

1.3.4.2 NHEJ

NHEJ is another DSB repair pathway in which breaks are simply ligated together; no information from the homologous chromosome is used to repair the break.

The first step in NHEJ is recognition and binding of the Ku heterodimer, consisting of Ku70 and Ku80, onto both strands of DNA either side of the break. Ku helps to maintain the synapsis by keeping the DNA ends in close proximity (Walker *et al*. 2001). The DNA-PKcs-Artemis complex is then recruited to the area of damage where DNA-PKcs phosphorylates Artemis. The complex, via the nuclease activity of Artemis, resects 5' overhangs to produce a blunt end (Ma *et al.* 2002). Recruitment of XLF (Gu *et al.* 2007), and the XRCC4-DNA ligase IV (LIG4) complex occurs (Chen *et al.* 2000), leading to repair and ligation of the separated strands (Fig. 1.6).

1.3.4.3 DSB repair and cancer

1.3.4.3.1 Hereditary breast, ovarian and prostate cancers

Heterozygous *BRCA1* and *BRCA2* mutations are linked to an increased risk of hereditary breast and ovarian cancers (Hall *et al.* 1990; Miki *et al*. 1994; Wooster *et al*. 1995, OMIM #114480). Both act as classical tumour suppressor genes, with a second somatic hit knocking out the genes function in DSB repair (Jasin, 2002; Sung and Klein, 2006). Genetic instability is a hallmark of *BRCA1* and *BRCA2* deficiency and cells display a heightened sensitivity to DNA damaging agents (Gretarsdottir

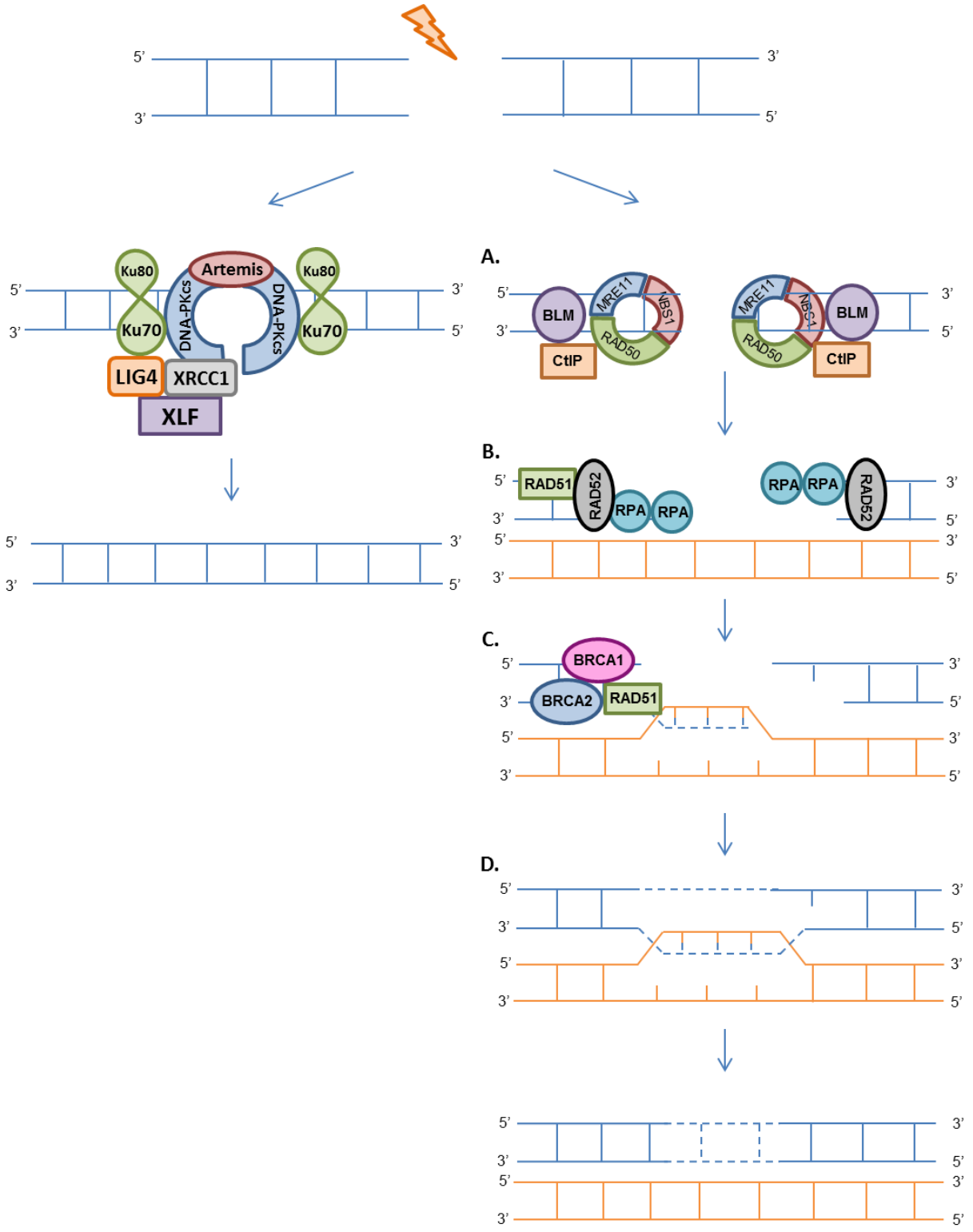


Figure 1.6 – DSB repair pathways. On the left is a schematic of NHEJ which involves simple re-annealing of damaged strands. On the right is HR; the break is processed (A and B), the strand crosses over and the sister chromosome is used as a reference for strand re-synthesis and repair (C and D).

et al. 1998; Moynahan *et al.* 2001). The overall risk of developing ovarian and/or breast cancer by the age of 70 in carriers of mutations in either gene is approximately 27% and 84%, respectively (Ford *et al.* 1998). Mutations of *BRCA1* and *BRCA2* have also been linked to hereditary prostate cancers (Ford *et al.* 1994; Gayther *et al.* 2000; Tischkowitz and Eeles, 2003; Castro *et al*. 2013).

In addition to the *BRCA* genes, other genes involved in HR have been implicated in hereditary breast cancers including checkpoint kinase 2 (*CHEK2;* Bernstein *et al.* 2006). CHEK2 is important in checkpoint signalling and control following DNA damage (Matsuoka *et al.* 1998). Upon activation by phosphorylation by ATM, CHEK2 phosphorylates and activates multiple downstream targets involved in DSB repair and checkpoint signalling, including BRCA1 (Bartek and Lukas, 2003).

1.3.4.3.2 Ataxia telangiectasia

Ataxia telangiectasia (AT, OMIM #208900) is an autosomal recessive disorder that is caused by inactivation of ataxia telangiectasia mutated (ATM). Sufferers of AT are at a 100 fold greater risk of leukaemia and lymphoma than the general population, and *ATM* mutations have also been linked to breast cancer susceptibility in carriers (Athma *et al.* 1999; Gumy-Pause *et al.* 2004). AT cells show an elevated sensitivity to ionising radiation indicative of DSB repair failure (Meyn, 1995). ATM is a protein kinase that is activated following the formation of a DSB. It has many targets in the HR pathway and in checkpoint signalling (Morrison *et al.* 2000).

1.3.4.3.3 Bloom syndrome

Bloom syndrome (BS, OMIM #210900) is an autosomal recessive disease, caused by biallelic loss of function mutations in *RECQL3* (*BLM*). Most of the mutations seen in BLM patients result in the production of a premature stop codon, leading to a truncated protein product (German *et al.* 2007). It predisposes sufferers to many different cancer types, with an early age of onset a hallmark of the disease (German, 1997). BS cells exhibit a high degree of chromosomal rearrangements between sister chromatids (Chaganti *et al.* 1974) which lead to an increased rate of loss of heterozygosity, chromosome rearrangements and deletions (Ouyang *et al.* 2008). BLM is a DNA helicase that is critical in the repair of double strand breaks (Ellis *et al.* 1995).

1.3.4.3.4 Nijmegen breakage syndrome

Nijmegen breakage syndrome (NBS, OMIM #251260) is a rare autosomal recessive disease caused by at least 10 hypomorphic mutations in *NBN* (Weemaes *et al.* 1981; Varon *et al.* 1998; Carney *et al.* 1998). The most common cancer seen in patients is non-Hodgkins lymphoma, although other cancers include medulloblastoma, glioma and rhabdomyosarcoma (van der Burgt *et al.* 2005). NBN is a critical part of the MRN complex involved in the repair of DSBs. It is believed that its role in the complex is in the recruitment of checkpoint proteins and it therefore modulates DNA damage signalling pathways (Kobayashi *et al.* 2004).

1.3.4.3.5 Rothmund-Thomson syndrome

Rothmund-Thomson syndrome (RTS, OMIM #268400) is an autosomal recessive disorder caused by biallelic mutations in *RECQL4* (Taylor, 1957; Kitao *et al.* 1999). At least 39 different mutations have been associated with RTS (Reix *et al.* 2007; Cabral *et al.* 2008; Siitonen *et al.* 2009; Debeljak *et al.* 2009). Sufferers are at a greater risk of osteosarcomas at a much younger age, with 32% of patients displaying symptoms. Additionally, approximately 5% of patients develop skin cancers later in life, with squamous cell carcinoma being the most common lesion seen (Wang *et al.* 2001). RECQL4 is a DNA helicase-like protein that is involved in recruitment of proteins at sites of single strand breaks following MRN processing of DSB (Petkovic *et al.* 2005; Singh *et al.* 2010). Additionally, interactions with proteins from multiple other DNA repair pathways have been reported, implicating a role in the repair of other forms of DNA damage (Woo *et al.* 2006; Fan and Luo, 2008; Schurman *et al.* 2009).

1.3.4.3.6 Werner syndrome

Werner syndrome (WS; OMIM # 277700) is a rare autosomal recessive disorder. It is a result of biallelic loss of function mutations in *RECQL2 (WRN)* (Yu *et al.* 1996). There is an increased incidence of multiple cancers in carriers, with approximately 60% of cancers seen consisting of osteosarcomas, soft tissue sarcomas, thyroid cancers and melanomas (Goto *et al.* 1996). WS cells are prone to large deletions as well as other forms of cytogenetic abnormalities (Fukuchi *et al*. 1989). WRN is a DNA helicase that functions in the ATP dependent unwinding of

DNA (Gray *et al.* 1997). In addition, WRN also possesses a 3'-5' exonuclease domain (Huang *et al.*1998). In HR, WRN has the ability to localise with RPA, recognise branched structures, and dissociate branched recombination structures (Constantinou *et al.* 2000). Additionally, a role of WRN, in complex with BRCA1, has been suggested in the repair of interstrand crosslinks (ICLs; Cheng *et al.* 2006).

1.3.4.3.7 LIG4 syndrome

LIG4 syndrome (OMIM # 606593) is a rare autosomal recessive disorder caused by mutations in *LIG4*. Mutations of these kind decrease the activity of the ligase in NHEJ (Girard *et al.* 2004) and are thought to be hypomorphic, since its knockout in mice is lethal (Frank *et al.* 2000). LIG4 syndrome predispose patients to acute leukaemias (Ben-Omran *et al,* 2005).

1.3.5 ICL repair

ICLs are highly toxic lesions due to the fact that, by binding and effectively joining opposite DNA strands together, they prevent strand separation, critical for replication and transcription (Dronkert and Kanaar, 2001). There are both exogenous and endogenous sources of ICLs but one of the best characterised is by-products of lipid peroxidation, such as malondialdehyde (Niedernhofer *et al.* 2003).

Due to the nature of ICLs, the lesions are only recognised in replicating cells, following stalling of DNA polymerases due to the inseparable DNA strands (Räschle *et al.* 2008). The stalled replication fork is recognised by the FANCM-FAAP24 complex which recruits the Fanconi anaemia (FA) core complex, consisting of seven proteins. The core complex, notably FANCL, ubiquitylates the FANCD2-FANCI complex, leading to retention of the complex. FANCD2-FANCI is responsible for the recruitment of multiple repair enzymes to the area of damage. In addition, FANCM can effectively recruit the Bloom's syndrome complex (BTR) which controls checkpoint activation via RPA and ATR triggered signalling cascades. The presence of RPA also triggers the localisation of HR pathway proteins, notably through BRCA2-FANCN. This allows for HR to control the stalled replication fork via the separation of DNA strands by the helicase activity of FANCJ (Li and Heyer, 2008).

1.3.5.1 ICL repair and cancer

FA (OMIM #227650) is a group of recessive disorders, caused by mutations of one of fourteen different genes involved in the repair of ICLs. The inability of FA cells to repair ICL is highlighted by the severe sensitivity shown to agents that cause ICL (Auerbach, 1988). FA sufferers exhibit a heightened risk of cancer, in particular squamous cell carcinomas, acute myeloid leukaemia, head and neck, oesophageal and gynaecological cancers (Alter, 2003). However, the degree of cancer susceptibility varies between complementation groups (Faivre *et al.* 2000).

All seven of the genes that form the core complex have been implicated in complementation groups of FA (FANCA, FANCB, FANCC, FANCE, FANCF, FANCG and FANCL), with over 90% of reported cases being the complementation groups FANCA, FANCC and FANCG (Deans and West, 2011). Notably, FANCD1 is caused by mutations in *BRCA2*, implicating its importance in multiple DNA repair pathways.

1.4 Treatment of colorectal cancer

The most important prognostic factor in CRC is tumour staging, for which treatment is highly dependent (Table 1.4). Five year survival rates drop to approximately 7% in patients presenting with stage IV CRC, in comparison to 93% in patients presenting with stage I (Cancer Research UK, Bowel cancer survival statistics, 2012). The most common form of curative treatment of stage I-III CRC is through surgery, with approximately 80% of patients undergoing surgical procedures. Adjuvant treatment with radiotherapy or chemotherapy is common. Unfortunately 25% of people present with metastatic CRC and up to 50% of individuals progress to this stage, the treatment for which remains challenging (Van Cutsem and Oliveira, 2009a). Only 20% of patients with hepatic metastasis are applicable for potentially curative surgery (Stangl *et al.* 1994). Chemotherapy therefore remains the mainstay in advanced CRC (aCRC) treatment. There are currently 8 agents that are approved by both the US food and drug administration (FDA) and European medicines agency (EMA) in the treatment of CRC. Additionally, regorafenib has recently received FDA approval, whilst aflibercept has recently received EMA approval following promising results (Table 1.5).

	TNM Staging			
Stage	**Tumour size (T)**	**Lymph nodes (N)**	**Metastasis (M)**	**Description**
0	Tis	N0	M0	(Tis) Cancer *in situ* - confined to mucosa
I	T1	N0	M0	(T1) Tumour invade *submucosa*
	T2	N0	M0	(T2) Tumour invades muscle layer
II	T3	N0	M0	(T3) Tumour invades *subserosa* or beyond
	T4	N0	M0	(T4) Tumour invades adjacent organs
III	T1-2	N1	M0	(N1) Metastasis to 1-3 lymph nodes; (T1-2) either *submucosa* or muscle layer has been invaded
	T3-4	N1	M0	(N1) Metastasis to 1-3 lymph nodes; (T3-4) tumour goes beyond *subserosa* or to nearby organs
	Any	N2	M0	(N2) Metastasis to 4 or more lymph nodes
IV	Any	Any	M1	(M1) Distant metastasis

Table 1.4 – Number stages and corresponding TNM staging of CRC with description of tumour growth given (adapted from Cancer Research UK, http://www.cancerresearchuk.org/cancer-help/type/bowel-cancer/treatment/tnm-and-number-stages-of-bowel-cancer, 2011).

1.4.1 Fluoropyrimidines

Fluoropyrimdines are central in the treatment of aCRC. Fluorouracil (5-FU, Efudex) has been used in the treatment of CRC for over 50 years. It is administered parenterally and, as an analogue of uracil, uses the same cellular transport systems to enter a cell. It can be considered a 'fraudulent' nucleotide; following conversion to flurodeoxyuridine monophosphate (fdUMP) it interacts, alongside reduced folate as a methyl donor (5,10-methylenetetrahydrofolate (MTHF)), and inhibits the action of thymidylate synthetase (TS) in the production of deoxythymine monophosphate, preventing DNA synthesis (Rang *et al*. 2007). It is often administered alongside the folate supplement leucovorin (5'-formyltetrahydrofolate, folonic acid). Leucovorin is anabolised to MTHF and has not only been shown to increase cellular levels of the donor but also to stabilise the TS-FdUMP complex (Radparvar *et al*. 1989). Studies have shown that administration alongside 5-FU results in clinical synergism, with double the response rate in aCRC (Advanced Colorectal Cancer Meta-Analysis Project, 1992).

Capecitabine (CPB, Xeloda) is an oral fluropyrimidine which is readily absorbed through the gut wall and metabolised to 5-FU at a preferential rate in tumour cells (Miwa *et al*. 1998) reducing systemic exposure of 5-FU and thus reducing its associated toxicity (Schüller *et al.* 2000). A three step enzymatic reaction occurs to activate CPB; firstly it is converted by hepatic carboxylesterase to 5'-deoxy-5-fluorocytidine and secondly to 5'deoxy-5-fluorouradine by cytidine deaminase. Finally, it is metabolised to the active metabolite 5-FU by thymidine phosphorylase, of which there is high activity in tumours leading to preferential accumulation (Ishikawa *et al*. 1998). In first line monotherapy treatment, response rates with CPB was significantly superior to those achieved with 5-FU and leucovorin (Van Cutsem *et al*. 2004).

5-FU together with leucovorin, is currently approved for use in the clinic together with oxaliplatin as part of the FOLFOX regimen, whilst CPB is administered alongside oxaliplatin as part of the XELOX regimen. The FOLFOX regimen was shown to double response rates compared to the respective monotherapies, as well as increasing the time of progression free survival (PFS) in the treatment of aCRC

Year	Therapy	Advance
1962	5-Fluorouracil	FDA approve 5-FU in the treatment of aCRC
1990	Adjuvant therapy	Chemotherapy becomes a mainstay as an adjuvant therapy following surgery; shown to improve survival following surgery by 40%
1996-1998	Irinotecan	EMA and FDA approve use of irinotecan together with 5-FU and leucovorin (FOLFIRI) in the first line treatment or as second line monotherapy of aCRC
1996-1999	Oxaliplatin	EMA approval for the use of oxaliplatin together with 5-FU and leucovorin (FOLXFOX) in the second line treatment of aCRC
2001-2004	Capecitabine	EMA and FDA approval for the use of capecitabine, an oral fluoropyrimidine, in the treatment of aCRC together with oxaliplatin and irinotecan as part of the XELOX and XELIRI regimens, respectively.
2002	Oxaliplatin	FDA approve the use of oxaliplatin in the FOLFOX regimen in the second line treatment of aCRC
2004-2005	Bevacizumab	EMA and FDA approval for the use of bevacizumab in the treatment of aCRC together with FOLFIRI and XELIRI
2004	Cetuximab	EMA and FDA approval for the use of cetuximab in the treatment of aCRC alone or in combination therapy with irinotecan
2006-2007	Panitumumab	EMA and FDA approval for the use of panitumumab as a monotherapy, as first line treatment together with FOLFOX and as second line treatment together with FOLFIRI
2008	Cetuximab	Mutations in codon 12 and 13 of the EGFR pathway gene *KRAS* are shown to result in ineffectiveness of treatment (Karapetis *et al.* 2008).
2009-2010	Cetuximab and panitumumab	EMA and FDA revise guidelines for EGFR inhibitors to take into consideration mutations of codon 12 and 13 of *KRAS* known to result in treatment failure
2012	Regorafenib	FDA approval for use of regorafenib in the treatment of aCRC refractory to other approved chemotherapeutics
2013	Aflibercept	EMA approval for the use of aflibercept in the treatment of aCRC that is refractory to oxaliplatin based treatment

Table 1.5 – Main therapeutic advances in the treatment of CRC

(de Gramont *et al*. 2000; Rothenberg *et al*. 2003; Saunders and Iveson, 2006). The FOLFOX and XELOX regimens have both been shown to be effective in the first line treatment of aCRC and as part of adjuvant therapy following surgery (Andre *et al*. 2004; Goldberg *et al.* 2004; Cassidy *et al.* 2004; Twelves *et al*. 2005). Alternatively the two are administered alongside irinotecan as part of the FOLFIRI and XELIRI regimen, again for first and second line treatment of aCRC, although not as adjuvant therapy (Saltz *et al*. 2000; Bajetta *et al.* 2004; Grothey *et al*. 2004). Response rates of XELOX and XELIRI mirrored those of the FOLFOX regimen, verifying that both CPB and 5-FU can be used in various regimens for the effective treatment of aCRC (Grothey *et al*. 2004; Cassidy *et al.* 2004; Cassidy *et al*. 2008; Ducreux *et al.* 2011).

1.4.2 Oxaliplatin

Oxaliplatin (Eloxatin) is a third generation platinum compound that has been used in the treatment of CRC for over 15 years. It consists of a 1,2-diaminocyclohexane (DACH) carrier ligand and a bidentate oxalate ligand (Kidani *et al*. 1978). Non-enzymatic displacement of the oxalate group following absorption allows for the formation of various reactive DACH intermediates that have the ability to react with DNA, notably to guanine and adenine bases. It acts as an alkylating agent of DNA, forming multiple crosslinks (Woynarowski *et al*. 2000). The production of these adducts, as well as secondary lesions that occur as a result of an accumulation of damage, ultimately results in apoptosis (Faivre *et al.* 2003). Approximately 90% of the lesions seen are intrastrand crosslinks, with 60% being between two adjacent guanine residues and the remaining 30% between adjacent guanine and adenine residues (Eastman, 1987). Other lesions observed include interstrand and DNA-protein crosslinks (Zwelling *et al*. 1979; Woynarowski *et al*, 2000). Before the development of oxaliplatin, CRC was considered to have intrinsic resistance to other platinum treatments (Rixe *et al*. 1996).

1.4.3 Irinotecan

Irinotecan (Camptosar) is a plant alkaloid (from the *Camptotheca acuminata* tree) that functions as a topisomerase I inhibitor. Topisomerase I is involved in relaxing super-coiled DNA by creating transient nicks in single stranded DNA during repair and replication (Pommier, 2013). It is readily metabolised by both hepatic and intestinal carboxylesterases to form the active compound SN38 (Adeji, 1999). SN38

functions to stabilise the topisomerase-DNA complex after it has nicked DNA, thus preventing re-annealing. This leads to replication stalling and ultimately apoptosis (Hsaing *et al*. 1985; Kawato *et al.* 1991). As well as in first line combinational treatment regimens, irinotecan is useful as a monotherapy in second line therapy.

1.4.4 Targeted therapies

The rationale behind the stratified treatment of cancer has led to the development of therapies specifically targeted to redundancies or growth advantages displayed by cancer cells. The production of monoclonal antibodies, with epitopes that target cancer cells has increased treatment efficacy and reduced chemotherapy associated side effects. The problem lies with the cost; monoclonal antibodies still remain relatively expensive, meaning that discovering pharmacogenetic reasons for altered response between patients could be critical for adequate use.

1.4.4.1 Cetuximab

Cetuximab (Erbitux) is a chimeric IgG_1 monoclonal antibody first approved in 2004 after successful treatment of aCRC either alone or together with irinotecan (Saltz *et al*. 2004; Cunningham *et al*. 2004; Van Cutsem *et al*. 2009b). However, cetuximab was shown to be ineffective in the first line treatment of aCRC in oxaliplatin based regimens (Maughan *et al*. 2011; Tveit *et al*. 2012), despite some reports suggesting the contrary (Bokemeyer *et al*. 2011).

The epidermal growth factor receptor (EGFR) is involved in regulation of transcription of nuclear targets involved in cell survival and growth through activation of signalling cascades including the Ras/Raf/MEK/MAPK and PI3K-Akt pathways (Krasinskas *et al.* 2011). Cetuximab selectively targets EGFR, competitively blocking ligand binding by EGF and TGFβ, preventing receptor activation (Mendelsohn and Baselga, 2003). Following binding to the extracellular domain of the EGFR receptor, apoptosis occurs as a result of cell cycle stalling in G1 (Huang *et al*. 1999). In addition to blocking ligand binding, as an IgG1 antibody it also has been shown to stimulate antibody-depedent cell-mediated cytotoxicity (ADCC) where the Fc region of the antibody is exposed, recognised as an antigen and the cancer cell targeted by the immune system (Iannello and Ahmad, 2005; Kawaguchi *et al*. 2007). Polymorphisms in receptors on killer cells required for antigen recognition have been

shown to alter the response of patients to cetuximab treatment, suggesting a role for ADCC in successful treatment (Zhang *et al*. 2007).

1.4.4.2 Panitumumab

As well as cetuximab, panitumumab (Vectibix) is also used in the selective targeting of the EGFR. A completely humanised IgG_2 monoclonal antibody, it again targets the extracellular domain of the receptor. Mutational analysis of cetuximab resistant but panitumumab sensitive cell lines suggests that this may be through a slightly different epitope (Montagut *et al.* 2012; Mareike Voigt *et al.* 2012). It is effective as both a monotherapy and in combination with standard chemotherapeutic regimens in the treatment of aCRC (Van Cutsem *et al.* 2007; Hecht *et al.* 2007). It has been shown to be effective at increasing PFS in combination with FOLFOX in the first line treatment of aCRC (Douillard *et al*. 2010) and in combination with FOLFIRI (Berlin *et al.* 2007). In second line treatment with FOLFIRI an increase in response rate of 25% was observed. However, this was dependant entirely on a *KRAS* wild type status (Peeters *et al*. 2010; Section 1.6.4).

1.4.4.3 Bevacizumab

Bevacizumab (Avastin) is a humanised IgG1 monoclonal antibody specifically designed to target the VEGF-A ligand and prevent binding to the VEGF receptor. The VEGF system is chiefly involved in control of endothelial cell proliferation and promotion of angiogenesis, something which tumour cells rely on for sustenance, survival and growth (Kim *et al.* 1993; Lee *et al*. 2000; Ferrara *et al.* 2004). Normalisation of tumour vasculature in bevacizumab treatment is associated with an increase in tumour uptake of irinotecan (Wildiers *et al.* 2003) suggesting a synergistic action in CRC.

Bevacizumab has shown to be effective in increasing overall survival and/or PFS in combination with fluoropyrimidine based treatment regimens (Kabbinavar *et al.* 2005; Hurwitz *et al*. 2005; Giantomio *et al.* 2007; Saltz *et al.* 2008; Van Cutsem *et al.* 2009c; Sobrero *et al.* 2009; Tsutsumi *et al.* 2012; Schmiegel *et al.* 2013; Beretta *et al*. 2013).

1.5 Side effects of CRC treatments (Table 1.6)

1.5.1 Fluoropyrimidines

Infusion of 5-FU is better tolerated than bolus administration since the latter causes no extreme peaks in exposure to chemotherapy (Lokich *et al. 1*989; Hansen *et al.* 1996). Although the degree of toxicity profiles differs between regimens, the main side effects of 5-FU with leucovorin treatment are gastrointestinal epithelial damage resulting in diarrhoea, stomatis, nausea, vomiting and oral mucositis, hand-foot syndrome and neutropaenia (Tsalic *et al.* 2003). In CPB treatment, similar side effects to 5-FU are observed, albeit at a reduced frequency (Cassidy *et al.* 2002; Schmoll *et al.* 2007). However, a hand-foot syndrome is seen at a greater rate. Hand-foot syndrome occurs in 50% of patients undergoing CPB treatment (Van Cutsem *et al.* 2000) and is characterised by erythema, dysthesia and, in extreme cases, swelling, ulceration and blistering of the skin, particularly on the hands and the feet (Barack and Burgdorf, 1991). Although rarely life threatening, it can be interfere with everyday life and compliance of patients undergoing treatment (Cassidy *et al.* 2002). One hypothesis for this increased prevalence is thought to be as a result of raised levels of the CPB metabolising enzyme, thymidine phosphorylase, in skin cells, resulting in an elevation of the metabolite (Asgari *et al.* 1999).

1.5.2 Oxaliplatin

Peripheral neuropathy is the most common dose limiting side effect associated with oxaliplatin treatment. An acute, dose dependant and reversible peripheral neuropathy is reported in 95% of patients undergoing treatment with oxaliplatin. The symptoms consist of parethesia, dysethesia and allodynia in the hands, feet and lips, as well as a laryngospasm or muscle cramps, which are exacerbated by exposure to low temperatures (Extra *et al*. 1998). Fortunately, the acute form appears to be reversible within hours or days (Argyriou *et al.* 2008).

The mechanism of action by which acute neuropathy occurs is not completely understood, however it is thought that it is due to disruption of the voltage gated sodium channels indirectly as an extension of chelation of calcium ions by the oxaliplatin metabolite, oxalate (Grolleau *et al*. 2001). Oxalate is known for causing

Drug	Side effect
Fluoropyrimidines	Gastointestinal epithelial damage; neutropenia; hand foot syndrome (greater incidence with capecitabine)
Oxaliplatin	Acute and chronic peripheral neuropathy
Irinotecan	Hyperstimulation of cholinergic system; neutropenia
EGFR inhibitors (Cetuximab and panitumumab)	Skin rash; trichomegaly; alopecia; hypersensitivity at injection site (with cetuximab)
Bevazicumab	Hypertension

Table 1.6 – Main side effects associated with treatment of CRC

neurotoxic effects in ethylene glycol poisoning, with peripheral neuropathy a symptom (Baldwin and Sran, 2010).

Chronic peripheral neuropathy is reported after several rounds of chemotherapy and has been shown to affect up to 50% of all patients undergoing treatment (Krishnan *et al*. 2006). Symptoms mimic that of cisplatin associated toxicity, consisting of a non-cold associated dysesthesia, paresthesia and sensory ataxia (Grothey, 2003), increasing in intensity following subsequent dosing. Although in 5% of patients symptoms appear to be irreversible following the cessation of treatment, in most cases there is an improvement of symptoms within 2 months (de Gramont *et al.* 2000; Alcindor and Beauger, 2011). It is believed to be due to direct toxicity of nerve cells via the accumulation of platinum adducts in the dorsal root ganglia, affecting DNA transcription and ultimately resulting in enhanced apoptosis in neuronal cells (Ta *et al.* 2006). There are no current treatments to alleviate the symptoms of peripheral neuropathy (Weickhardt *et al*. 2011). Since in most cases neuropathy is reversible, symptoms can be controlled with dose reductions and treatment modifications (de Gramont *et al.* 2000; de Gramont *et al.* 2004; Tournigand *et al.* 2006).

In addition to peripheral neuropathy, an elevated degree of neutropenia, nausea and diarrhoea is associated with the FOLFOX regimen when compared to 5-FU and leucovorin alone (Rothenberg *et al.* 2003).

1.5.3 Irinotecan

Dose limiting side effects of irinotecan consists primarily of a delayed onset of diarrhoea, due to a high concentration of SN38 in the intestine following hepatic elimination (Hecht, 1998). In 40% of patients, the side effect is severe (Pitot *et al.* 2000). Additionally, acute toxicites associated with hyperstimulation of the cholinergic system are commonly observed including emesis, diarrhoea, abdominal cramps, bradycardia and hypotension (Nicum *et al*. 2000; Tobin *et al*. 2004). Experiments in animals have indicated that irinotecan can effectively inhibit acetylcholinesterases, as well as effectively stimulating muscarinic receptors (Kawato *et al.* 1993). Severe neutropenia is also a commonly seen side effect.

The acute cholinergic symtoms respond well to the anti-cholinergic drug, atropine (Pitot *et al.* 2000; Fuchs *et al.* 2003), whilst the delayed onset diarrhoea has shown to be controlled by high dose loperamide (Abigerges *et al.* 1994). However, some patients do not respond and dose modifications or treatment cessation are required (Cunningham *et al.* 1998; Van Cutsem *et al.* 1999; Rothenberg, 2001)

1.5.4 Targeted therapies

1.5.4.1 Cetuximab

One of the most common side effects, seen in 80% of patients treated, is the development of a skin reaction, most notably as an acnieform skin rash. The rash appears to be dose dependant and is seen most commonly on the face, neck, shoulders and chest (Segaert and Van Cutsem,2005). In up to 18% of cases it is severe. In addition, other common dermatological complaints include fissures on the hands and feet, xerosis and changes in hair growth (Agero *et al.* 2006). Other side effects of treatment include trichomegaly, alopecia, diarrhoea, hypomagnesmia and severe hypersensitivity at the site of infusion (Dueland *et al.* 2003; Chung *et al.* 2008).

In most cases treatment of skin rashes is necessary in order to ease discomfort and aid compliance. For acneiform skin rash, topical anti-acne medication or anti-inflammatory medication has been shown to be effective, although the choice of therapy is dependent on the location of rash. If xerosis is also present, a fine therapeutic balance must be struck between acneiform treatment and hydrating lotions, since either treatment can exacerbate the other symptom. In severe cases of acneiform skin rash, high dose oral anti-histamines are effective at reducing the reaction (Segaert and Van Cutsem, 2005).

1.5.4.2 Panitumumab

As an EGFR inhibitor, similar side effects to cetuximab are commonly seen with panitumumab treatment, with dermatological toxicities again being the most common (>90%). Additionally, fatigue, nausea, diarrhoea, hypomagnesmia and neutropeania are all commonly seen (Van Cutsem *et al.* 2007). However, hypersensitivity at the site of injection is rare due to the fact that, unlike cetuximab, panitumumab is a fully humanised antibody (Ranson, 2003).

1.5.4.3 Bevacizumab

The most common side effect of bevacizumab treatment is severe hypertension. Approximately 23% of all patients undergoing treatment will suffer from the side effect, with 8% of these classified as severe (Ranpura *et al.* 2010). It is thought that inhibition of VEGF can lead to a reduced production of vasodilators, such as nitric oxide, lowering normal physiological levels and ultimately resulting in vasoconstriction (Olsson *et al.* 2006; Mourad *et al.* 2008). Additionally a reduced level of nitric oxide also leads to a reduced level of sodium excretion which in turn could contribute to hypertension as a result of water retention in the blood (Granger and Alexander, 2000). Other side effects associated with treatment include an increased risk of arterial and venous embolisms, proteinuria, bleeding, and in rare cases, poor wounding healing and gastrointestinal perforations (Hurwitz *et al.* 2004; Saltz *et al.* 2008).

Hypertension can be treated by the administration of an angiotensin-converting enzyme (ACE) inhibitor or other diuretic, calcium channel blockers, beta blockers or various other anti-hypertensive drugs (Motl, 2005; Pande *et al.* 2006; Saif, 2009). To minimise the chance of bleeding, problems with wound healing and gastrointestinal perforations, it is recommended that bevacizumab treatment as adjuvant to surgery is either discontinued or started at a time point suitable to allow for adequate healing of wounds (Shord *et al*. 2009). In severe cases of all side effects, dose modification and reduction can reduce the severity of the effect seen.

1.6 Pharmacogenetics of CRC treatment

1.6.1 Fluoropyrimidines

Several genetic factors have been attributed to varying response in treatment to the fluoropyrimidine agents in CRC. Polymorphisms in TS have been associated with altered expression of the protein, with increased expression being inversely linked to clinical outcome (Lurje *et al*. 2009). One such polymorphism consists of a 28bp repeat sequence in the 5’ untranslated region (5’UTR) of the gene. Significantly higher expression of TS was associated with 3 such repeats when compared to 2 repeats (Horie *et al.* 1995; Pullarkat *et al*. 2001). Expression was even higher when a G→C polymorphism in the second of the three repeats is present (Mandola *et al*.

2003). Conversely, a 6 base pairs deletion in the 3'UTR significantly decreased mRNA stability, influencing expression of TS (Mandola *et al.* 2004). In terms of side effects to treatment, individuals homozygous for the 2 repeat allele are over ten times more likely to suffer from greater than grade 3 toxicity than individuals homozygous for the 3 repeat allele (Lecomte *et al*. 2004).

Another pharmacogenetic factor in fluoropyrimidine treatment consists of two common polymorphisms in the methylenetetrahydrofolate reductase (MTHFR) gene. MTHFR is important in the production of reduced folate, critical for the action of 5-FU. The polymorphisms Ala222Val and Glu429Ala have been shown to be associated with an increase in response to treatment (Little *et al*. 2003; Etienne-Grimaldi *et al*. 2010).

The main route of 5-FU metabolism is by the enzyme dihydropyrimidine dehydrogenase (DPYD), with up to 80% of the administrated dose degraded by the enzyme (Woodcock *et al*. 1980). Over 15 different polymorphisms correlate with altered DPYD activity, with lowered acivity being associated with a greater degree and a quicker rate of onset of 5-FU associated side effects (van Kuilenburg *et al.* 2000; Collie-Duguid *et al.* 2000; Newton *et al.* 2012). An extreme toxicity phenotype is associated wih a splice site point mutation that results in a 165 base pair deletion consisting of an entire exon of the gene (Wei *et al*. 1996). Although rare in the Caucasian population (MAF <1%), up to 24% of patients with at least one copy of this allele exhibit grade 3 or greater toxicity (Raida *et al.* 2001). Additionally, a rare nonsynonymous variant at position 949, resulting in the subsitution of a valine for an aspartic acid residue has been shown to influence the enzymatic action of DPYD and cause 5-FU toxicity comparable to that seen with the exon skipping mutation (Morel *et al.* 2006).

1.6.2 Oxaliplatin

The efficacy of oxaliplatin in the treatment of aCRC has been shown to be affected by variants in genes involved in its pharmacokinetic and cellular response pathway. For example, a coding variant in glutathione-S-transferase π (*GSTP1*), resulting in an isoleucine to valine substitution at codon 105 of the protein, increases survival in the treatment of aCRC (Stoehlmacher *et al.* 2002) although its reliability as a pharmacogenetic allele is of debate (Fariña Sarasqueta *et al.* 2011). GSTP1 is

involved in the detoxification of reactive intermediates of oxaliplatin by conjugation with glutathione.

Altered expression of *ERCC1,* a gene integral to the NER of platinum adducts, has been shown to affect response to platinum treatment, with increased expression significantly increasing resistance to various treatment regimens in aCRC (Shirota *et al*. 2001; Arnould *et al.* 2003; Seetharam *et al.* 2010; Arora *et al*. 2010; Noda *et al*. 2012; Tentori *et al.* 2013). Concordant with this, increased expression of *ERCC1* is commonly observed following oxaliplatin treatment (Baba *et al*. 2012). Clinical outcome of oxaliplatin treatment has also been associated with a C>T silent polymorphism, encoding Asn118. Homozygosity of the C allele has been shown to be positively correlated with outcome of treatment (Park *et al.* 2003), with presence of the T allele increasing mRNA levels and conferring resistance to treatment (Ruzzo *et al.* 2007).

Another DNA repair gene that has been linked to clinical outcome is the BER gene *XRCC1.* The Arg399Gln polymorphism has been associated with an increased response to treatment (Stoehlmacher *et al.* 2001; Lv *et al.* 2013).

With regards to side effects to treatment, a putative association between chronic peripheral neuropathy and Ile105Val in *GSTP1* has been described (Grothey *et al.* 2005; Ruzzo *et al.* 2007; Peng *et al.* 2013), although the risk allele is of debate (Lecomte *et al.* 2006; Gamelin *et al.* 2007; Inada *et al.* 2010). Particular haplotypes of alanine glycoxylate transferase (*AGXT*), involved in oxalate metabolism, have been shown to predispose towards both acute and chronic forms of peripheral neuropathy (Gamelin *et al.* 2007). Additionally, the silent polymorphism encoding Asn118 in *ERCC1* has been shown to be associated with an elevated rate of onset of peripheral neuropathy in the Japanese population (Inada *et al.* 2010; Oguri *et al.* 2013). Oguri *et al.* also highlighted an association between rs17140129 in phenylalanyl-tRNA synthetase 2 (*FARS2*) and the severity of peripheral neuropathy, and rs10486003 in tachykinin (*TAC1*) and the rate of onset. Both of these variants are in non-coding regions and were originally associated with chronic peripheral neuropathy in a GWAS which also identified 7 other variants as associated with the side effect (Won *et al.* 2012). Also a nonsynonymous variant in sodium channel, voltage gated 10A (*SCN10A*; Leu1092Pro [rs12632942]) and an intronic variant

(rs2302237) in *SCN4A* have been shown, under an overdominant model, to increase the chance of acute peripheral neuropathy, with the latter also influencing the severity of the side effect (Argyriou *et al.* 2013).

1.6.3 Irinotecan

There has been much research into the role of UDP-glucuronosyltransferase (UGT1A1) in response to treatment with irinotecan. UGT1A1 is important in the deactivation of the active metabolite, SN38 (Gupta *et al*. 1997). In patients homozygous for a $[TA]_7$ repeat in the promoter region (referred to as UGT1A1*28) an increased degree of toxicity is observed, particularly diarrhoea and neutropeania (Ando *et al*. 2005; Hoskins *et al*. 2007). Additionally, patients with elevated bilirubin (another substrate of UGT1A1) or with inherited deficiencies in UGT1A1 (Gilberts syndrome; OMIM #143500) have also been shown to be at an elevated risk of irinotecan associated toxicities (Wasserman *et al.* 1997; Lankisch *et al.* 2008).

1.6.4 Cetuximab and panitumumab

Mutations in a downstream effector of the EGFR associated pathway, kirsten rat sarcoma viral oncogene homolog (*KRAS*), are responsible for resistance to EGFR inhibitors. A lack of response in patients with *KRAS* mutations is seen in both monotherapy and combination therapies for both drugs (Lièvre *et al*. 2008; De Roock *et al*. 2008; Freeman *et al*. 2008; Amado *et al*. 2008; Bokemeyer *et al*. 2009; Van Custem *et al*. 2009b). Of note, it was shown that tumours with activating mutations in *KRAS* at codons 12 and 13 had significantly reduced response rates to cetuximab treatment; from 13% to 1.2% (Karapetis *et al*. 2008). Additionally, rarer activating mutations at codon 61 and 146 are associated with a similar lack of clinical response to treatment (Loupakis *et al.* 2009a). As *KRAS* mutations are seen in up to 40% of colorectal tumours, these activating mutations could have major implications in EGFR targeting treatment of CRC.

Following the observation that up to 60% of *KRAS* wild type tumours are unresponsive to EGFR inhibitor treatment, it was proposed that other components of the EGFR pathway could be implicated in lack of response (Linardou *et al.* 2008). In addition to *KRAS* mutations, the presence of the activating v-raf murine sarcoma viral oncogene homolog B1 (*BRAF*) mutation, V600E, was seen to be associated

with a reduction in drug efficacy (Di Nicolantonio *et al.* 2008; Benvenuti *et al*. 2007). *BRAF* mutations are seen in approximately 10% of aCRC (Davies *et al*. 2002; Rajagopalan *et al.* 2002). Similarly, activating mutations in codon 61 of *NRAS,* another isoform of the *Ras* gene, reduces response rates by over 30% in carriers. Both *BRAF* and *NRAS* mutations are considered to be mutually exclusive to any *KRAS* mutation. In addition to this, oncogenic mutations of phosphatidylinositol-4,5-bisphosphate 3-kinase (*PI3KCA*; De Roock *et al*. 2010; Laurent-Puig *et al.* 2009; André *et al*. 2013) and loss of expression of the PI3K pathway inhibitor and tumour suppressor, *PTEN*, are also associated with EGFR inhibitor treatment failure (Frattini *et al.* 2007; Perrone *et al.* 2009; Loupakis *et al*. 2009b; Sood *et al.* 2012). Both *PI3KCA* and *PTEN* mutations can co-occur with other mutations in the EGFR pathway (Sartore-Bianchi *et al*. 2009).

Several studies have also reported a correlation between increased EGFR expression and response to treatment (Moroni *et al*. 2005; Sartore-Bianchi *et al.* 2007; Heinemann *et al*. 2009), although the benefit of testing for overexpression as a biomarker of response is of debate. Recent evidence has emerged suggesting that an acquired mutation, Ser492Arg, found in the extracellular domain of the EGFR receptor could alter binding and therefore lessen effectiveness of treatment of cetuximab but not panitumumab (Montagut *et al.* 2012). Interestingly, the presence of a skin rash as a side effect in either drug treatment is positively correlated with overall response (Saltz *et al.* 2004; Jonker *et al*. 2008; Peeters *et al.* 2009). As EGFR is highly expressed on epidermal surfaces, the characteristic skin rash is thought to be due to direct inhibition of EGFR on the surface of the skin (Giovannini *et al.* 2009)

1.7 Next generation sequencing

Advances in next generation sequencing (NGS) have revolutionised genomics and our understanding of human disease. This ultimately can have implications in the diagnosis and treatment of patients (Gonzalez-Angulo *et al.* 2010). NGS utilises massively parallel sequencing to effectively amplify and sequence the genome reliably and at a low cost; the first genome to be sequenced using a NGS platform was done so at a significantly reduced cost compared to preceding methods (Levy *et al.* 2007; Wheeler *et al*. 2008; Shendure and Ji, 2008). To date, NGS has been used

in the identification of multiple casual alleles in multiple different diseases (Table 1.7).

1.7.1 General workflow

NGS methods consist of three main stages, although the mechanism by which they are carried out can vary greatly depending on the data output desired and platform used (Metzker, 2010). These stages consist of initial sample preparation, massively parallel sequencing and imaging of sequence data, and data analysis. There are multiple NGS platforms currently available. Technologies vary in their amplification method, sequencing method and applications, each with their own advantages and disadvantages (Table 1.8).

Despite 85% of disease causing mutations being located in protein coding regions, only 1% of the entire genome makes up the 'exome' (Ng *et al.* 2009; Choi *et al.* 2009). Considering the cost of whole exome sequencing (WES) is considerably less than whole genome sequencing (WGS) this makes it an appealing alternative when looking for mutations responsible for a given phenotype. In WES, an additional 'target capture' step is carried out during sample preparation in order to select for the protein coding regions of DNA. Following sheering of the DNA, adaptors are ligated to the fragments and hybridisation assays are carried out to isolate the previously defined coding sequences (Pruitt *et al.* 2009). Common techniques include microarray-based (Albert *et al.* 2007; Okou *et al.* 2007) and solution based enrichment assays (Porreca *et al.* 2007).

Following generation of sequencing reads, quality control of reads is carried out to remove errors that can occur during the sequencing process (Pabinger *et al.* 2013). Following this, the reads are aligned with and compared to a reference sequence, ensuring that any differences between the two can be distinguished (Flicek and Birney, 2009). Multiple mapping algorithms for this purpose are available and the choice of tool is dependent on the original platform used and applications required (Bao *et al.* 2011). When analysing samples for variations in relation to the reference genome, multiple tools are available to aid annotation of variants (McKenna *et al.* 2010; Wang *et al.* 2010; Yandell *et al.* 2011). Some alignment tools, such as Mapping and Assembly with Quality (MAQ), have also been developed to also aid in the detection of variants (Li and Durbin, 2009).

Use of NGS technology	Reference
First genome sequenced by WGS	Wheeler *et al.* (2008)
First cancer genome (acute myeloid leukaemia) sequenced by WGS	Ley *et al.* (2008)
First 12 human exomes sequenced using targeted capture technology. Displayed that WES could be used to identify Mendelian disorders by studying four individuals with Freeman-Sheldon syndrome (OMIM #193700)	Ng *et al.* (2009)
First diagnosis of a hereditary disease (congenital chloride losing diarrheao, OMIM #214700) with a previous diagnosis of Bartter syndrome using NGS	Choi *et al.* (2009)
First use of NGS in the discovery of alleles associated with a Mendelian disease trait; WES uncovered *DHODH* mutations in individuals with Miller syndrome by enriching for variants between two siblings and in two unrelated affected individuals (OMIM #263750)	Ng *et al.* (2010)
WES used to uncover the role of *WDR62* mutations in patients with severe brain malformations	Bilgüver *et al.* (2010)
WES was used to uncover autosomal dominant mutations in *SETBP1* in Schinzel-Giedon syndrome, that were shown to be *de novo* following Sanger sequencing of the patients parents	Hoischen *et al.* (2010)
WES used to uncover the role of *MLL2* mutations in patients with Kabuki syndrome (OMIM #147920)	Ng *et al.* (2010)
WES used to identify *de novo* mutations in *POP1* in two siblings with previously unclassified anauxetic dysplasia (OMIM #607095)	Glazov *et al.* (2011)
WES, together with linkage data, used in the discovery of variants in *POLE* and *POLD* (OMIM #615083, #612591, respectively) associated with predisposition to multiple CRA and CRC	Palles *et al.* (2013)
WES used to identify *ERCC4* as a candidate gene for FA in one patient. Its role in an additional patient with previously unclassified FA symptoms was confirmed by Sanger sequencing of the gene (OMIM #615272)	Bogliolo *et al.* (2013)
WES used to uncover a role of *STAMBP* mutations in patients with microcephaly–capillary malformation syndrome (OMIM #614261)	McDonell *et al.* (2013)
NGS technologies used to identify driver mutations and pathways associated with oesphageal adenocarcinoma	Dulak *et al.* (2013)
WES of families with autism uncovers hypomorphic loci in genes implicated in other diseases	Yu *et al.* (2013)

Table 1.7 – A selection of developments and findings from NGS technology

Company	*Instrument (Base error rates)*	*Amplification method*	*Sequencing method*	*Advantages*	*Disadvantages*
Roche	454 FLX Titanium/FLX Titanium +/ GS Jr. Titanium (All 1%)	Emulsion PCR	Pyrosequencing	Long read lengths; fast	Runs are expensive; problems with homopolymer repeats >8bp
Illumina®	GA II/HiSeq ™1000/Hiseq ™2000/ MiSeq /HiScanSQ (All 0.1%)	Solid phase 'bridge' PCR	Sequencing by synthesis	Low running costs; widely used	High start-up costs; difficult to multiplex samples; short read lengths
Life technologies™	SOLiD ™4 (0.06%)/SOLiD ™PI/ SOLiD ™4hq (Both 0.01%)	Emulsion PCR	Sequencing by ligation	Runs are inexpensive; highest accuracy	Slow; short read lengths; high start-up costs
Life technologies™	Ion torrent™ PGM™ 314/316/318 chip (All 1.2%)	Emulsion PCR	H^+ detection synthesis	Fast, platform is inexpensive	Short read lengths; long sample preparation times
Pacific bioscience™	PacBio RS/RS II (13%)	None - sequences single DNA molecules	Real time	Longest read lengths; runs are inexpensive	High start-up costs; high error rates

Table 1.8 – Summary of current available NGS technologies (Glenn, 2011; Henson et al. 2012; Liu et al. 2012)

1.7.2 Gene discovery strategies

NGS has made substantial advances in determining the genetic architecture of many diseases, notably in the discovery of rare variants that previous studies did not have the power to detect (Table 1.7). However, the sheer amount of data produced with NGS can make finding disease-causing variants difficult; between 20,000 and 50,000 variants in a single sample are typically identified through WES (Gilissen *et al.* 2012). This number grows considerably when variation in the whole genome is considered (Pabinger *et al.* 2013). Therefore, techniques to identify potential disease causing alleles are required (Cooper and Shendure, 2011).

The selection of samples can aid greatly in the genetic enrichment process, and help to keep cost down. Two general strategies have been previously outlined; the sequencing of patients exhibiting extreme phenotypes (Li *et al*. 2011) and the sequencing of families. Sequencing of siblings, or other family members, with similar phenotypes can be useful in identifying a common causative allele (Gilissen *et al.* 2012), whilst focusing of family trio's can be helpful when investigating inheritance patterns or in the discovery of *de novo* mutations (Bamshad *et al.* 2011).

Since the vast majority of known Mendelian disease-causing mutations are in protein coding regions, it is rational to consider protein coding variants to be the most deleterious. However, approximately 90% of coding variants identified are known polymorphisms (Robinson *et al.* 2011), and are therefore unlikely to be pathogenic. Filtering for novelty status or by rarity helps to focus the search whilst maintaining power to detect a casual variant. Both can be assessed by using online databases, such as dbSNP (NCBI Resource Coordinators, 2013), the 1000 genome project (1000 Genomes Project Consortium *et al*. 2010) and Ensembl (Flicek *et al.* 2013). Another approach involves assessing whether potential variants are predicted to be deleterious to protein function. For example, truncation, splice site, frameshifting insertions and deletions, and nonsynonymous variants are all likely to have functional implications. Multiple online tools are available to assess how a nonsynonymous variant may affect a proteins function, including SIFT (Ng and Henikoff, 2001), Align-GVGD (Tavtigian *et al.* 2006) and PolyPhen-2 (Adzhubei *et al.* 2010).

Validation of variants identified is important. False positive can often arise as a result of poor mapping of reads or sequencing errors, whilst false negatives can occur as a result of poor coverage, poor calls of variants or poor capture of particular regions, particularly in WES (Majewski *et al.* 2011; Gilissen *et al*. 2012).

1.7.2.1 Complex traits

Complex traits with a known degree of heritability display high degrees of locus heterogeneity, with casual variants present in multiple different genes (Lander and Schork, 1994; Glazier *et al.* 2002). Previously, GWAS have made considerable headway in uncovering common loci that predispose to complex genetic traits (Hindorff *et al.* 2013). However, it has been suggested that additional rare variants could potentially further explain the percentage of heritable cases not explained by current genetic understanding; the so called 'missing heritability' (Manolio *et al*. 2009).

In the 'common-disease-rare variant' hypothesis, rare variants could potentially have a dramatic effect on overall risk (Pritchard, 2001; Bodmer and Bonilla, 2008). Typically, rare variants are not included on the large scale genotyping arrays used in GWAS. Also, due to the low frequency of such potential variants, GWAS are not powerful enough to detect linkage with such variation (McCarthy and Hirschhorn, 2008). Based on the observation that the vast majority of disease causing mutations affects protein coding regions, this suggests that WES could be a useful enrichment tool in rare variant discovery of complex disease (Kiezun *et al.* 2012).

Alternatively, it has been proposed that a low risk, common variant at a given loci uncovered by GWAS could be within haplotypes encompassing rarer variants that individually have a high effect on disease risk and are therefore likely to be the true casual variants (Dickson *et al.* 2010). This indirect association is referred to as synthetic association (Goldstein, 2009) and could be particularly helpful when considering regions of the genome to focus on in NGS. For example, following analysis of the region surrounding a GWAS locus for type 1 diabetes, four rare variants were discovered that were significantly associated with protection against the disease (Nejentsev *et al.* 2009). This highlights the validity of looking at GWA loci as an approach for rare variant discovery using NGS in complex disease.

1.7.2.2 Mendelian disorders

NGS is also important in the diagnosis and discovery of the causes of Mendelian disorders that have previously been missed using traditional approaches (Bamshad *et al.* 2011). The use of NGS as a diagnostic tool was first displayed by Choi *et al.* (2009) who uncovered a homozygous missense variant in solute carrier member 26, member 3 (*SLC26A3*), known to cause congenital chloride diarrhoea, in patients previously diagnosed as having Bartter syndrome.

NGS has also become a powerful tool in the discovery of Mendelian disorders. The first casual variant of a Mendelian disease trait to be uncovered by WES occurred in 2010 by Ng *et al* (2010)*;* the researchers identified the underlying cause of previously undefined Millers syndrome in 6 kindred's. Since then, WES has been used to uncover multiple underlying alleles associated with Mendelian disorders (Table 1.7).

In hereditary CRC, Palles *et al.* (2013) used WGS, together with pre-existing linkage data, to examine 13 families with CRA and CRC without any known hereditary CRC gene mutations. They discovered a nonsynonymous variant, Leu424Val that falls within the catalytic subunit of the POLE complex; important in leading strand DNA synthesis during replication and repair. Additionally, the same research discovered a second predisposition allele in two different families, consisting of the nonsynonymous variant, Ser478Asn, seen in the catalytic subunit of POLD. Again, POLD is involved in DNA synthesis and repair but in the lagging strand.

1.8 Genetic model systems of DNA repair

Adequate DNA repair mechanisms are critical for viable life (Alberts *et al.* 2002). Chemically, the damage that arises in DNA is the same between organisms (Lindahl, 1993). Both prokaryotic and eukaryotic organisms are used as models for various DNA repair pathways and the degree of conservation shown highlights the importance in evolution. The use of genetic modelling systems is invaluable in gaining insight of how genetics influences protein function in a complex system *in vivo* (Table 1.9). The choice of model organism used for genetic manipulation relies

Specie	Advantages	Disadvantages
Escherichia coli (E.coli)	Easy to genetically manipulate; cheap; genome well annotated; well-studied	Not representative of a multicellular organisms; major difference with humans in most DNA repair pathways; prokaryote
Saccharomyces cerevisiae (*S.cerevisiae;* also referred to as 'budding yeast')	Easy to genetically manipulate; cheap; genome well annotated; well-studied; pathways more similar to humans than *E.coli;* as a haploid organism it is useful for studying effects of recessive mutations	Not representative of a multicellular organism; not a mammal; some differences in DNA repair pathways
Schizosccharomyces pombe (*S.pombe*; 'fission yeast')	Easy to genetically manipulate; cheap; genome well annotated; well-studied; pathways more similar to humans than *E.coli;* as a haploid organism it is useful for studying effects of recessive mutations; excises mammalian introns (unlike *S.cerevisiae*)	Not representative of a multicellular organism; not a mammal; some differences in DNA repair pathways; alternative pathway for the repair of UV light
Drosophilla melanogaster (fruit fly)	Representative of a multicelluar organism; genome is well annotated; easy and cheap to use in the laboratory	Differences between DNA repair pathways; not a mammal
Caenorhabditis elegans (round worm)	Representative of a multicellular organism; genome is well annotated; easy and cheap to use; easy to genetically manipulate	Differences between DNA repair pathways; not a mammal
Mus muscularis (mouse)	Mammal; easy to genetically manipulate; genome is well annotated; large proportion of genome (>80%) homologous with humans	Some differences in DNA repair pathways

Table 1.9- Advantages and disadvantage of organisms commonly used as model systems i the study of human DNA repair pathways

heavily on the conservation of proteins and pathways. Also, particular organisms have ‘back up’ DNA repair pathways not seen in other organisms which need to be taken into account when choosing an organism for a genetic study.

1.8.1 MMR pathway

The MMR pathway has been well characterised in *E.coli* (Lahue *et al.* 1989). However, *E.coli* possess only three MMR exclusive proteins (MutS, MutL and MutH) whilst humans and other eukaroytes employ many more (Augusto-Pinto *et al.* 2003) all of which are homologs of MutS or MutL which are essential for MMR in all species (Kolodner, 1996). No homologs of MutH have been identified in humans.

The pathway has also been well studied in *S.cerevisiae*, *S.pombe* and *C.elegans,* and display more similarities to humans. For example, there are multiple homologs of both *MutS* and *MutL* involved in the pathway, though again none have *MutH* homologs (Harfe and Jinks-Robertson, 2000). There are far fewer MMR homologs present in *D.melanogaster,* although at least one of both *MutS* and *MutL* homologs are present (orthologs of *MSH6, MLH1* and *PMS1*; Sekelsky *et al*. 2000).

1.8.2 BER pathway

The BER pathway is well conserved in most organisms, indicating its importance in survival. Studies of *E.coli* have been important in the understanding of mechanisms of repair, with most key proteins conserved from *E.coli* to eukaryotes (Robertson *et al*. 2009). Although most DNA glycosylases are conserved between species, there are some key differences of note. For example, there is no *S.pombe* homolog of OGG1 (Eisen and Hanawalt, 1999; Chang and Lu, 2005), however, it is conserved in *S.cerevisiae*, mice and various other organisms (Arai *et al.* 1997; Radicela *et al*. 1997). Also, the human glycosylase *TDG* is conserved in *E.coli* (*Mug*) but not in *S.cerevisiae* despite being conserved in *S.pombe.* One of the only major differences in *D.melanogaster* is the apparent lack of a *POLB* homolog (Sekelsky *et al*. 2000). The BER pathway is not well conserved in *C.elegans,* with homologs for only a couple of human DNA glycosylases (Eisen and Hanawalt, 1999; Leung *et al.* 2008). All major components of the BER pathway are conserved in mice making it an excellent model organism of the pathway.

1.8.3 NER pathway

There are key differences in the repair of bulky, helix distorting adducts between *E.coli* and eukaryotes .Despite both being able to adequately excise bulky adducts such as those formed following UV treatment, the proteins involved vary greatly. *E.coli* uses a system known as the UvrABCD pathway, which functions in much the same way as the eukaryotic NER system (Hoeijmakers, 1993a; Truglio *et al.* 2006). However, there is little homology with those proteins involved in eukaryotic organisms. Also far fewer proteins are required in *E.coli* excision repair in comparison to eukaryotic repair (Prakash and Prakash, 2000; Cleaver *et al.* 2001).

S.cerevisiae is probably the best studied eukaryotic model organism of NER. There is a very high level of protein homology with humans (Hoeijmakers, 1993b; Prakash *et al.* 1993; Wood, 1997), although there are a few key differences in protein specificity between organisms (Eisen and Hanawalt, 1999). A similar degree of homology is observed in *C.elegans,* however no protein homologous to *DDB2* or *CSA* have been identified (Lans and Vermeulen, 2011).

S.pombe also displays a level of high conservation and homology with human NER proteins (Lehmann, 1996; Egel, 2004). However, *S.pombe* possesses a second UV damage repair pathway which was first recognised in NER knockouts when UV adducts were still repaired at a substantial rate (Birnboim and Nasim, 1975). Additionally *S.pombe* NER knockouts fail to display the same degree of sensitivity as the *S.cerevisiae* counterparts (Lehmann, 1996). The UV damaged DNA endonuclease (Uve1) –dependent excision repair pathway (UVER) has been shown to excise both 6-4,PPs and CPDs much more rapidly than the NER pathway (Yonemasu *et al.* 1997).

D.melangoster appears to lack a TC-NER pathway, since no CSA or CSB homologs have been identified, and relies solely GG-NER (Keightley *et al.* 2009). In rodents, GG-NER of CPDs is significantly less efficient than in humans, due to a lack of p48 which is induced to upregulate NER (Tang *et al.* 2000).

1.8.4 DSB repair pathways

In *E.coli*, the only method for the repair of DSB is through HR. Although there is some degree of homology of the proteins involved, the main steps are carried out by proteins quite different to human HR proteins. One gene that maintains a high level of conservation throughout various species, is *RAD51*; a protein key in the recognition of homology between strands and for strand guidance. Its retention throughout evolution highlights its importance in the repair of DSB (Modesti and Kanaar, 2001). Similarly, there are at least five human homologs of the *E.coli* helicase, *RecQ*, with mutations in these causing WS, BS and RTS (Brosh and Bohr, 2007).

There are many similarities in the DSB repair pathways between *S.cerevisiae, S.pombe, C.elegans* and *D.melanogaster* and these organisms have been invaluable in the study of both HR and NHEJ (Sekelsky *et al*. 2000; Krogh and Symington, 2004; Raji and Hartsuiker, 2006; Lemmens and Tijsterman, 2011). However, both yeast organisms only have one homolog of *RecQ* whilst *D.melanogaster* and *C.elegans* both have four (Sekelsky *et al*. 2000).

The main difference in the repair of DSB in mammalian cells compared to yeast is that the majority of repair in mammalian cells occurs via the NHEJ pathway, whilst in yeast it is through HR (Eisen and Hanawalt, 1999).

1.8.5 ICL repair pathway

The repair of ICL in *E.coli* is predominantly carried out via incision of the damaged strand by the NER protein system UvrABC, as well as by the coordination of HR proteins. Similarly, an orchestration of multiple pathways is known to operate in ICL repair in yeast (McVey, 2010).

The main difference in mammalian cells is the presence of the FA pathway for ICL repair. Of the proteins involved, homologs for four have been identified in *C.elegans* and *D.melanogaster* indicating that the pathway may be important in these organisms (Youds *et al.* 2009; McVey, 2010). A high level of conservation of the pathway is observed in mice, in which they have been extensively studied with regards to the effects of mutations on the development of phenotypes of FA (Bakker *et al.* 2013).

1.9 Aims of this project

1. To identify novel low penetrance alleles in DNA repair pathways that predispose to CRC.

2. To utilise exome resequencing in the identification of alleles associated with severe forms of oxaliplatin induced peripheral neuropathy. To independently validate findings.

3. To further examine identified variants and their associated genes genetically.

4. To create a model system to investigate the functional effects of variants associated with oxaliplatin induced peripheral neuropathy. To further investigate phenotypes associated with the introduced variants.

Chapter Two - Materials and methods

2.1 List of suppliers

Materials and equipment were purchased from the following companies:

ABgene Ltd (Surrey, UK)
Acros Organics (See Thermo Fisher Scientific)
Agilent Technologies (Berkshire, UK)
Anachem Ltd (Bedfordshire, UK)
Applied Biosystems (Chesire, UK)
Becton, Dickinson and Company (Oxford, UK)
Bibby Sterlin (See Thermo Fisher Scientific)
Bioquote (York, UK)
Biorad (Hertfordshire, UK)
Corning Incorporated (Flintshire, UK)
Eurogentec (Hampshire, UK)
Fisher Scientific (Leichestershire, UK)
Formedium (Norfolk, UK)
GE Healthcare (Buckinghamshire, UK)
Illumina (California, USA)
Invitrogen Life Technologies (Strathclyde, UK)
Jencon (West Sussex, UK)
Labtech International (East Sussex, UK)
Melford (Suffolk, UK)
Microzone (Haywards Heath, UK)
Millipore (Hertfordshire, UK)
MJ Research (Massachusetts, USA)
Molecular Dynamics (See GE Healthcare)
MWG Biotech (Buckinghamshire, UK)
New England Biolabs (Hertfordshire, UK)
Pharmacia Biotech (See GE Healthcare)
Qiagen (West Sussex, UK)
R&D Systems (Oxford, UK)
Sigma-Aldrich Ltd (Dorset, UK)

Stratagene (California, USA)
Thermo Fisher Scientific (Massachusetts, USA)
Vector (Peterborough, UK)
VWR International (Leicestershire, UK)

2.2 Materials

2.2.1 Chemicals

Analytical grade chemicals were purchased from either Sigma-Aldrich Ltd or Fisher Scientific unless otherwise stated.

2.2.2 Polymerase chain reaction (PCR)

AmpliTaq Gold DNA polymerase along with appropriate buffer and $MgCl_2$ were purchased from Applied Biosciences. Deoxyribonucleotide triphosphates (dNTPs) were purchased from GE healthcare. All primers (unless otherwise stated) were purchased from Eurogentec. Dimethyl sulfoxide (DMSO) was purchased from Sigma Aldrich. Mega mix gold (MMG) was purchased from Microzone.

2.2.3 PCR purification

Exonuclease I (Exo) was purchased from New England Biolabs. Shrimp alkaline phosphatase (SAP) was purchased from GE healthcare. Millipore Montage SEQ_{96} sequencing reaction clean-up kits were purchased from Millipore.

2.2.4 Electrophoresis

Agarose was purchased from Eurogentec. Ethidium bromide was supplied by Sigma Aldrich. For the purpose of safe disposal of running buffer, ethidium bromide destaining bags from Fisher Scientific were utilised. 100bp DNA ladder was purchased from New England Biolabs and 1kb Plus DNA ladder from Invitrogen Life Sciences.

2.2.5 Sanger sequencing

BigDye Terminator cycle sequencing kit v3.1, POP6 polymer and HiDi formamide were all purchased from Applied Biosystems. Capillary electrophoresis buffers were purchased from Sigma Aldrich.

2.2.6 Sanger sequencing clean up

For the isopropanol method, isopropanol was purchased from Fisher Scientific and HiDi formamide was purchased from Applied Biosystems. For the Montage SEQ_{96} sequencing reaction clean up, kits were purchased from Millipore.

2.2.7 TaqMan single nucleotide polymorphism (SNP) genotyping

All assays and TaqMan universal mastermix were purchased from Applied Biosystems. Predesigned assays were used for RAD1 – rs1805327 (C_25617909_10), POLG – rs3087374 (C_15793548_10), REV1 – rs3087403 (C_15793621_10), BRCA1 – rs799917 (C_2287943_10) and ERCC6 - rs2228527 (C_935106_20).

2.2.8 Gene expression analysis

Expression of target genes was analysed using intron spanning primers. Both colon and kidney first strand cDNA was purchased from Stratagene.

2.2.9 Clinical material

All blood samples from COIN, COIN-B, FOCUS2, FOCUS3 and PICCOLO were obtained with patient consent and with ethical approval for bowel cancer research.

2.2.10 Bacteria culture reagents and solutions

All solutions were made using dH_2O water and autoclaved on a liquid cycle at 15lb/sq.in at 121°C for 20 minutes.

Luria Bertani (LB) Culture Medium

1% w/v tryptone, 0.5% w/v yeast extract (Both Becton, Dickinson) and 1% w/v NaCl in 1L dH_2O

LB agar Medium

1.5% w/v bacterial agar (Becton, Dickinson), 1% w/v tryptone, 0.5% w/v yeast extract and 10% w/vNaCl in 1L dH_2O

Ampicillin Stock Solution

50mg/ml of ampicillin sodium salt (Melford) was dissolved in dH_2O, filter sterilised and stored at -20°C

Glycerol (BDH Laboratories) for long term storage

50% glycerol for long term storage of bacterial cultures was made by diluting 250ml of 100% glycerol with 250ml dH_2O.

SOC Medium (Invitrogen)

2% w/v tryptone, 0.5% w/v yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose

2.2.11 Plasmids

pAW1 was first constructed by Watson *et al* (2008), and was generously provided by Oliver Fleck (Bangor University). pAW8-ccdB was constructed and generously provided by Edgar Hartsuiker (Bangor University). pGEM-T easy vector and system were purchased from Promega.

2.2.12 Chemically competent cells

JM109 chemically competent *E.Coli* cells were obtained from Promega.

2.2.13 Plasmid extraction kit

For small scale plasmid extraction, QIAprep mini-preparation (here after termed miniprep) plasmid kits (Qiagen) were used unless otherwise stated.

2.2.14 *Cre Recombinase*

Cre recombinase enzyme and respective buffer were purchased from New England Biolabs.

2.2.15 Site directed mutagenesis (SDM)

QuikChange Lightning site directed mutagenesis kits were purchased from Agilent Technologies.

2.2.16 Restriction enzymes

All restriction endonucleases were supplied with the appropriate buffer by New England Biolab.

2.2.17 *S.pombe* reagents and solutions

All solutions were made using dH_2O water and autoclaved on the liquid cycle at 15lb/sq.in at 121°C for 20 minutes.

Yeast extract liquid (YEL) and Yeast extract agar (YEA)

For YEL, 0.5% w/v yeast extract and 3% w/v glucose is made up to 1L in dH_2O. This was supplementated with 100mg/L of adenine, histidine, uracil (ura), lyseine and arginine (all Formedium). For YEA, in addition to this 1.6% w/v Bacto-agar was added.

Minimal media agar (MMA)

0.17% w/v yeast nitrogen base, 1.8% w/v Bacto-agar, 0.5% ammonium sulphate, 1% glucose in 1L of dH_2O to pH 6.5. Appropriate supplements to a concentration of 100mg/L were added when required.

Edinburgh minimal media (EMM)

14.7mM potassium hydrogen phthalate, 15.5mM disodium phosphate, 93.5mM ammonium chloride, 2% w/v glucose and 2% w/v Bacto-agar.

Malt extract agar (MEA)

3% w/v Bacto-malt extract and 2% w/v Bacto-agar. Appropriate supplements to a concentration of 100mg/L were added as required.

TE – 0.1M Lithium Acetate (LiAc)

10mM Tris, 1mM EDTA, 0.1M LiAc pH 8.0

40% PEG 4000 with 0.1M LiAc in TE (pH8.0)

40% PEG 4000, TE pH8.0/ 0.1M LiAC pH8.0.

2.2.18 Yeast strains

All strains of *S.pombe* were generously provided by Oliver Fleck. These included EH238 (*smt-0 ura4 D18 leu1-32),* J129 (*h- uve1::LEU2 leu1-32 ura4-D18)* and 503 (*h+ leu1-32 ura4-D18 [ade6-704]*).

2.2.19 Extraction of *S.pombe genomic DNA*

Lyticase, proteinase K and ribonuclease (RNase) were purchased from Sigma-Aldrich. Phenol chloroform isoamyl-alcohol (PCIA) was purchased from Fisher Scientific.

2.2.20 Drugs for *S.pombe* treatments

Oxaliplatin was purchased from R&D systems; methyl methanesulfonate (MMS, 99%) was purchased from Acros Organics; hydroxyurea (HU, 1M) was purchased from Formedium.

2.3 Equipment

2.3.1 Plastics and glassware

Plastic eppendorf tubes (0.65ml, 1.5ml and 2ml) were purchased from Bioquote, whilst 1.5ml tubes were purchased from Sigma. Sterile pipette tips and tips for multi-channel pipettes were purchased from Anachem. Sterile stripettes were purchased from Corning Incorporated. Sterile universals were purchased from Bibby Sterilin. Fisher Scientific supplied 96 well Thermo-Fast PCR reaction plates, whilst 4titude adhesive PCR sealing sheets were obtained from ABgene. 0.2ml plastic strip

tubes were also obtained from ABgene. Glass flasks and beakers were obtained from Jencons or Fisher Scientific.

2.3.2 Thermocycling

Thermocycling was carried out using an MJ Research DNA engine tetrad PTC-225.

2.3.3 Electrophoresis

Electrophoresis was carried out in an AB gene AB0708 100V gel tank using a BioRad 200/2.0 power pack. Visualisation of ethidium bromide stained gels was achieved using a BioRad GelDoc XR transluminator.

2.3.4 Taqman SNP genotyping

Taqman SNP genotyping assays were analysed using either the Applied Biosystems 7900HT Real-Time PCR system (in Germany) or Applied Biosystems 7500 Real-Time PCR system (in Cardiff).

2.3.5 Sanger sequencing

Sanger sequencing was carried out on an ABI 3100 Genetic Analyser (Applied Biosystems). All data was analysed and annotated using Sequencer v4.2. Reference sequences were obtained from online databases, including NCBI and Ensembl.

2.3.6 Quantification of nucleic acids

To measure the concentration of DNA, either an UV spectrophometer (NanoDrop ND-800, Labtech International) or a Qubit® 2.0 Fluorometer (Life technologies) with appropriate buffers was used.

2.3.7 Transfer of *S.pombe*

Transfer of *S.pombe* was carried out using a replicating block and a sterile piece of velvet.

2.3.8 UV treatment

S.pombe cells were treated with UV light using a Stratalinker (Stratagene).

2.4 Bioinformatics and statistical software

Genetic statistical analyses were carried out using the online program PLINK v1.07 (Purcell *et al*, 2007; http://pngu.mgh.harvard.edu/purcell/plink/). All variants run through PLINK were tested for accordance with the Hardy-Weinberg equilibrium (HWE; Hardy, 1908). In addition to PLINKs meta-analysis application, meta-analysis was run using Comprehensive meta-analysis, v2.0 (Biostat; http://www.meta-analysis.com/index.html). Other statistical software utilised included IBM SPSS statistics 20. All primers were designed using Primer 3 v0.4.0 (http://frodo.wi.mit.edu/primer3/) and checked for sequence specificity using the online program Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/).

Illumina's GenomeStudio v2009.1 was used to analyse results of Illumina genotyping and produce data plots. It was also used to create reports with specific variant information for analysis using PLINK.

Species alignment of amino acid and nucleotide sequences was carried out using the online tool Clustal-Omega (Goujon *et al*. 2010; http://www.ebi.ac.uk/Tools/msa/clustalo/), with sequences obtained from NCBI.

Restriction enzymes were chosen based on recognition sites in DNA sequences identified via the New England Biolab Cutter, v2.0 (http://tools.neb.com/NEBcutter2/index.php).

In silico analysis of variants effect on protein function was determined using the online algorithm tools Align-Grantham Variation/Grantham Deviation (Align-GVGD; http://agvgd.iarc.fr/agvgd_input.php), Polymorphism Phenotype v2 (PolyPhen-2; http://genetics.bwh.harvard.edu/pph2/) and 'Sorting intolerant from tolerant' (SIFT; http://sift.jcvi.org). LD data was obtained using Haploview v4.2 (Barrett *et al*. 2005).

In analysis of exome sequencing data, FASTQ files were processed using BWA, calibrated using GATK and annotated using ANNOVAR by Dr James Colley (Cardiff University).

2.5 Methods

2.5.1 General reagents

10 x TAE Buffer (for electrophoresis);

400mM Tris, 200mM Acetic acid, 10mM EDTA to pH 8.0

10 x TBS Buffer

1.5M NaCl, 0.05M Tris, pH 7.6

2.5.2 Quantification of nucleic acids

To measure the concentration of DNA, an UV spectrophometer at wavelengths of 260nm and 280nm was used. An absorbance ratio of 1.8 at these wavelengths was considered an indicator of high sample purity. Alternatively, a Qubit®2.0 fluorometer was used, measuring DNA at a wavelength of 260nm. For samples predicted to have a concentration less than 100ng/µl, high specificity standards and buffers were utilised. For samples predicted to have concentrations up to 1000ng/µl, broad range standards and buffers were utilised.

2.5.3 Primer design

All primers were designed using Primer 3, v0.4.0 (Rozen and Skaletsky, 2000). Wherever possible, primers were designed between 18-25 nucleotides in length, had an annealing temperature within 2^{o}C of the respective partner, and had low predicted dimerisation and secondary structure formation. All primers were checked for locus specificity by using the Primer-Blast software.

2.5.4 PCR

PCR allows for rapid and accurate amplification of a chosen region of DNA *in vitro.* The exponential manner of DNA amplification allows for the production of several thousand copies of the region of interest.

Initial stages involve heating the reaction mixture to a temperature sufficient to disrupt hydrogen bonds between opposite bases, resulting in separation of double stranded DNA. This separation and a cooling in temperature allow the binding of primers designed specifically to the region of interest. Upon the action of a

thermostable DNA polymerase, a new strand is synthesised from the primer by incorporating dNTPs. This results in the production of a complimentary strand of DNA. Repetition of this process, usually between 25-40 times, results in the production of a large amount of specific product (Mullis *et al*, 1986).

Unless otherwise stated, standard PCR reaction mixtures consisted of 0.2mM dNTPs, 10pmols forward and reverse primer, GeneAmp 10x buffer (added to a final concentration of 10mM Tris-HCl, 50mM KCl, 1.5mM $MgCl_2$, 0.01% (w/v) gelatine, pH8.3), 1U AmpliTaq Gold DNA polymerase, 5% DMSO and 40ng of DNA (in a final volume of 25µl). Cycling conditions consisted of an initial denaturation step of 95^{o}C for 2 minutes, followed by 35 cycles of 95^{o}C for 30 seconds, annealing temperature of between 50-60^{o}C for 30 seconds, elongation step of 72^{o}C for 30 seconds, and a final elongation of 72^{o}C for 10 minutes.

For PCR amplification that had previously failed using standard procedures, MMG was utilised. 25ng of template DNA was added to 25pmol of respective forward and reverse primers, with half the reaction mixture consisting of MMG reagent (Contents trade secret; CTS).

2.5.5 Agarose gel electrophoresis

Agarose gel electrophoresis is a method used to separate DNA on the basis of size and shape. Agarose, when solid, forms a matrix with pores running through, the size of which is determined by the concentration of the gel. When an electrical charge is applied, negatively charged DNA and RNA fragments will move towards the positive (anode) electrode, pulling them through the agarose matrix. Shorter molecules and those of a smaller size and shape move faster through the matrix than larger, bulkier molecules, resulting in separation of the product (Sambrook *et al.* 1989). The product can be visualised by the addition of ethidium bromide, an intercalating agents that sits between bases in DNA. The compound forms fluorescent complexes in this setting and these can be viewed under UV light at a wavelength of 300nm.

Agarose gels were made with 1XTAE buffer to a concentration of 0.8-2% (dependant on the size of the fragments to be separated). Conical flasks were heated to allow the agarose to melt, cooled slightly and 0.05µg/ml of ethidium

bromide was added in a fume hood. This was poured into a gel tank and allowed to cool until set. The gel was then completely submerged in 1xTAE buffer in an AB0708 100V gel tank. 2µl of loading dye (15% w/v ficol, 10mM Tris pH 8, 1mM EDTA, 0.2% orange G) was added to 8µl of sample and the entire volume was loaded onto the gel. Gels were run at 100V for around 40 minutes with a 100bp or 1kb DNA ladder. UV visualisation was carried out following separation and photographed using Bio-Rad XR system. Ethidium bromide destaining bags were added to running buffer for a minimum of 24 hours to remove the dye before disposal.

2.5.6 ExoSAP PCR purification

ExoSAP degrades any excess primers, ssDNA and phosphate groups from dNTPs. *Exo* is a 3'-5' exonuclease which degrades excess single stranded oligonucleotides from reactions containing double stranded products. SAP is an alkaline phosphatase that removes 5'-phosphates from the PCR product. 1µl of ExoSAP is added directly to the PCR products. The sample is then incubated at 37°C for 60 minutes, followed by an enzyme deactivating stage of 80°C for 15 minutes.

2.5.7 Sanger sequencing

Sanger sequencing utilises dideoxyribonucleotide triphosphates (ddNTPs), which lack the 3'hydroxyl group of the deoxyribose sugars, being incorporated into an emerging strand by DNA polymerase. As a result of the missing group, there is chain termination (Sanger *et al.* 1977). Each ddNTP is labelled with a different coloured fluorophore and capillary electrophoresis can detect nucleotides up to a sequence length of approximately 500bps.

DNA from a PCR product is denatured and a specific primer bound. The action of DNA polymerase extends the chain from the primer, incorporating dNTPs. However, it is the random insertion of a ddNTP that terminates further ssDNA strand elongation. Capillary electrophoresis separates the ssDNA through the polymer, POP-6, on the basis of size. Smaller products travel fastest, and subsequently pass through the laser beam first. This activates the fluorophore and causes the emission of light at a particular wavelength depending on the incorporated ddNTP.

BigDye Terminator v3.1 Cycle Sequencing kit was used to sequence ExoSAP treated PCR products. A reaction mixture was used based on the manufacturer instructions; 5µl purified PCR product was added to 0.2% BigDye v3.1 (CTS), 10pmol desired primer, 1x BigDye sequencing buffer (CTS) and made up to 10µl with dH_2O.

Cycling conditions consisted of 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and finally 60°C for 3 minutes and 30 seconds

Products of BigDye termination sequencing reactions were subsequently cleaned of all unincorporated nucleotides and dyes by either the isopropanol method or using Montage SEQ Sequencing Reaction Clean-up kits.

2.5.8 Isopropanol clean up method

In the isopropanol clean up method, 40µl of 75% isopropanol was added to 10µl BigDye reaction mixture and left to incubate at room temperature for 30 minutes. Samples were then centrifuged at 4000rpm for 45 minutes and inverted on absorbent paper to remove all isopropanol. Samples were then spun inverted at 500rpm for 30 seconds and air dried in a dark place for 10 minutes to evaporate any residual liquid. The pellet was resuspended in 10µl of HiDi formamide.

2.5.9 Montage SEQ_{96} sequencing clean up

Millipore Montage SEQ_{96} sequencing reaction clean-up kits provide an efficient way to remove salts and dye terminators from Big Dye v3.0 reactions in a similar manner to the isopropanol method. They employ size exclusion technology via a filter at the bottom of each well to retain sequencing products. 20µl of injection solution (CTS) was added to 10µl Big Dye product and the entire volume was transferred to a Millipore clean up plate. Suction was applied to the bottom of the plate for 6 minutes, which was then removed and blotted onto absorbent tissue. 25µl of fresh injection solution was added to the wells and suction was applied for another 6 minutes. Once blotted again, 20µl of fresh injection solution was added to the wells and the plate placed on a microplate shaker for 6 minutes. 10µl was transferred from the clean up plate to a 96 well plate to be sequenced.

Following both clean-up methods, samples were analysed using an ABI 3100 analyser. Chromatograms were visualized and analysed using Sequencher v4.2.

2.5.10 TaqMan SNP genotyping

TaqMan SNP genotyping assays make use of specific primer and probe sets in order to successfully assay for SNPs. During thermal cycling, allele specific probes labelled with different dyes (namely VIC and FAM dyes) are allowed to selectively bind to single stranded DNA. AmpliTaq Gold DNA polymerase extends from the primer, and due to the 5' exonuclease activity of DNA polymerase, breaks down any probe that is bound. This results in the release of the allele specific dye from the immediate proximity of a quencher, leading to a measurable emission.

Reaction mixture was made containing 1 x Taqman Universal Mastermix (CTS), 1 x Taqman assay and a minimum of 10ng of DNA, to a final volume of either 5µl (Applied Biosystems 7900HT Real-Time PCR system) or 25µl (Applied Biosystems 7500 Real-Time PCR system). A pre-read run was performed to determine any baseline fluorescence. PCR was then carried out in the real time machine with thermal cycling conditions consisting of an initial denaturation of 95°C for 10 minutes followed by 40 cycles of 92°C for 15 seconds and 60°C for 1 minute. Following amplification, a post read run using the original pre read document was carried out to subtract the baseline fluorescence. The sequence detection system (SDS) software was used to plot the result of the allelic discrimination run on a scatter plot of allele X versus allele Y.

2.5.11 Gene expression analysis

Tissue specific expression of genes of interest was analysed by amplification with intron spanning primers using first strand colon and kidney cDNA as a template. Two sets of primers for each gene were utilised to gauge expression. Primer for *β-actin* from the supplier was used as a positive control. PCR conditions consisted of an initial denaturation of 95°C for 2 minutes, followed by 40 cycles of 95°C for 1 minute, 55°C for 1 minute and 72°C for 2 minutes, with a final elongation of 72°C for 10 minutes.

2.5.12 Bacterial techniques

2.5.12.1 General growth of bacteria

All glassware, equipment and reagents used were autoclaved before use. All bacterial work was carried out in sterile conditions. Cultures were incubated at either 30°C or 37°C in line with optimum conditions for the plasmid used (pAW8 and p-GEM T easy vector, respectively).

2.5.12.2 Preparation of LB and LB-agar

LB and LB-agar were made up as described in section 2.2.10, adjusted to pH 7 and autoclaved on a liquid cycle. In the case of LB-agar, solution was cooled to 50°C and ampicillin added to the appropriate concentration. Where appropriate, 0.5mM of isopropyl-β-D-thio-galactopyranoside (IPTG) and 80µg/ml of 5-bromo-4-chloro-3-inodyl-D-galactoside (X-gal; both Sigma-Aldrich) were also added. Approximately 20ml was poured into the bottom of an 80mm petri dish and allowed to cool. All plates were stored at 4°C.

2.5.12.3 Set up of starter cultures

For each culture, 5ml of LB was added to a universal along with the appropriate concentration of antibiotic. A sterile pipette tip was used to isolate and remove colonies from LB-agar plates and transferred to a universal. Cultures were left on an orbital shaker for 14-18 hours at either 30°C or 37°C at 200rpm.

2.5.12.4 Long term storage of bacteria

For storage at -80°C, 500µl of 50% glycerol was added to 500µl of starter colonies. The solution was vortexed gently to mix and stored at -80°C.

Recovery of bacteria was carried out by thawing on ice, vortexing to mix and using a sterile loop to streak out the glycerol stock onto LB agar plates with ampicillin. Plates were incubated at either 30°C or 37°C overnight.

2.5.12.5 Ligation reaction

Purified PCR products were ligated into the pGEM-T easy vector system (Promega). The amount of PCR product required was calculated as follows;

$$Insert\ (ng) = \frac{Vector\ (ng) x\ Size\ of\ insert (kb)}{Size\ of\ vector\ (kb)} \times insert{:}\ vector\ molar\ ratio$$

A reaction mixture consisting of 3 units of T4 DNA ligase, 1x T4 DNA ligase rapid ligation buffer, 50ng pGEM-T easy vector and the desired amount of PCR product. Ligation was carried out either at room temperature for 1 hour or at 4°C overnight.

2.5.12.6 Transformation of JM109 competent cells

Transformation of JM109 was carried out via heat shock. This process results in semi-permeabilisation of the cell membrane, allowing for uptake of 'naked' DNA molecules into the cell. JM109 cells were thawed on ice and mixed gently by flicking. Two microliters of each ligation reaction and 50µl of cells were placed into eppendorf tubes on ice for 20 minutes. Samples were heat shocked by placing on a 42°C heat block for 50 seconds, followed by immediately incubating on ice for 2 minutes. Each reaction was added to 950µl of SOC medium and incubated at 37°C with agitation for 1.5 hours, following which 100µl was spread onto LB agar plates containing ampicillin, X-gal and IPTG. Plates were incubated at 37°C for 16-18 hours. The pGEM-T vector carries the *LacZ* gene which encodes β-galactosidase; an enzyme which breaks down X-gal, resulting in the production of blue colonies. If the insert had been correctly taken up into the vector, there is disruption of the *Lac*Z gene, resulting in no X-gal breakdown and colonies appear white in colour. This allows for easy selection of colonies with successful uptake of the vector and insert ligation product.

2.5.12.7 Small scale purification of plasmids

All small scale plasmid purifications were carried out using QIAgen miniprep kit unless otherwise stated and following the manufacturer's protocol. Extraction involves alkaline lysis of cells accompanied by gentle mixing; releasing DNA and denaturing proteins. By the addition of a neutralisation agent and adjustment of salt

levels, binding of DNA to a silica column and precipitation of proteins and other cell debris is facilitated.

5ml starter cultures were harvested by centrifuging for 1 minute at 13,000rpm. Supernatant was discarded and the pellet re-suspended in 250µl of buffer P1, containing RNase (100µg/ml), 50mM Tris/HCl, 10mM EDTA. Following this, 250µl of lysis buffer (buffer P2; 200mM NaOH, 1% SDS) was added and the tube inverted 8-10 times to mix, resulting in lysis of the bacterial cells. After approximately 1 minute (no longer than 5 minutes) 350µl of neutralisation buffer (buffer N3; 3M potassium acetate) was added and inverted 8-10 times to prevent the reaction going any further. The tube was centrifuged for 10 minutes at 13,000rpm. A P1000 pipette was used to transfer supernatant to a spin column. This was centrifuged for 1 minute at 13,000rpm and the flow through discarded. A wash step was carried out by adding 750µl of buffer PE, containing ethanol, and centrifuged for 1 minute at 13,000rpm to remove any salt. The flow was discarded, the tube was twisted slightly and centrifuged for 1 minute at 13,000rpm to ensure that all wash buffer had been removed since ethanol could interfere with some downstream applications. The QIAprep column was placed in a clean 1.5ml microcentrifuge tube and 50µl of dH_2O added to the centre of each membrane, left for 1 minute on the bench and then centrifuged at 13,000rpm for 1 minute.

2.5.12.8 Cre recombinase reaction

Cre recombinase is a topoisomerase enzyme which catalyses both the *in vitro* and *in vivo* homologous recombination of DNA between *lox* sites allowing for site specific recombination. *Lox* sites are 34 base pair sequences consisting of 13 base pair inverted repeat sequences with a central 8 base pair spacer region. The efficiency of recombination can be altered by mutating nucleotides in the spacer region on the *lox* sites. By flanking cassette regions with two varying *lox* sites which display inefficient recombination with one another (in this study *loxP* and *loxM3)*, we can efficiently and precisely carry out a double recombination at a particular locus (Hoess *et al*, 1986; Langer *et al*, 2002).

Molar ratios of plasmid to insert were calculated depending on the size of each to calculate the concentration of insert that was required. A total concentration of 250ng of DNA was required for optimal recombination.

Standard reaction mixture consisted of appropriate volumes of insert and vector, 1x*Cre* recombinase reaction buffer (33mM NaCl, 10mM $MgCl_2$, 50mM Tris-HCl, pH 7.5), 1U *Cre* recombinase and 5% PEG 8000, made up to a final volume of 10µl. Solution was mixed thoroughly and incubated at 37°C for 30 minutes, followed by 70°C for 10 minutes to inactivate the enzyme.

2.5.12.9 SDM

SDM is a technique used to introduce mutations of interest into plasmids. Mutant strand synthesis is carried out using primers designed with the mutation of interest incorporated. Using these primers on a suitable template, thermal cycling is carried out using a high fidelity *PfuI* enzyme. Following this, the paternal strand (which does not contain the mutation) is digested using *DpnI*; an endonuclease that degrades methylated and hemi-methylated DNA (Fig. 2.1). DNA that has been isolated from *E.Coli* is *dam* methylated and susceptible to this degradation. Finally, the mutated plasmid is transferred into competent cells (Kunkel, 1985).

All reactions were carried out using the QuikChange Lightning SDM kit, following manufacturers' protocol. Primers consisting of between 30-37 base pairs with the mutation incorporated into both complementary pairs were used. PCR reaction mixture containing 2x QuikChange lightning buffer (CTS), 1.25pmol of both primers, 10mmol dNTPs, 6% QuikSolution reagent, 2.5U of *PfuUltra* HF DNA polymerase and 10ng of target plasmid was made up to a final volume of 50µl with dH_2O. Thermal cycling conditions for all reactions consisted of an initial denaturation of 95°C for 1 minute followed by 18 cycles of 95°C for 50 seconds, 60°C for 50 seconds and 68°C for 1 minute/kb in size. An elongation step of 68°C for 7 minutes finished the cycle and all reactions were placed on ice for 2 minutes to cool the reaction below 37°C. Degradation of the paternal DNA was carried out by adding 1µl of *DpnI* directly to the reaction mixture and mixing thoroughly by pipetting up and down. Products were incubated in a water bath at 37°C for 1 hour

For the transformation, XL-10 Gold ultracompetent cells were thawed on ice. Once thawed cells were mixed by gently flicking the tube and 45µl was pipetted into pre-chilled microcentrifuge tubes. To each reaction 2µl of XL-10 Gold β-mercaptoethanol mix was added, pipetted up and down to mix and left on ice for 10

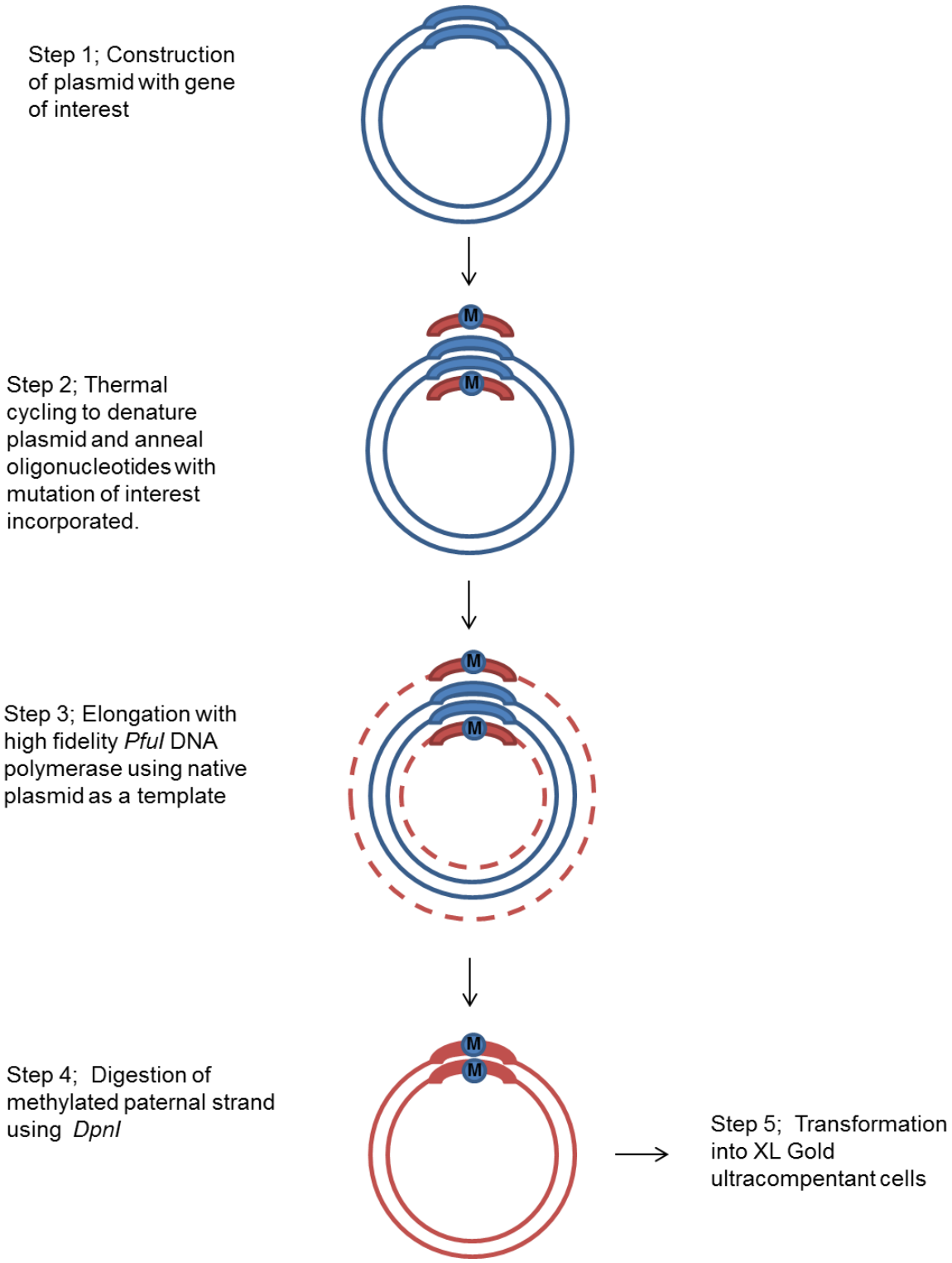


Figure 2.1 – SDM utilising QuikChange lightning SDM kit (M = mutation of interest)

minutes, swirling gently every 2 minutes. 2µl of *DpnI* treated product was added to each aliquot of cells and vortexed slightly to mix, incubating on ice for 30 seconds. To heat pulse, all tubes were placed on a heat block at 42°C for 30 seconds and immediately transferred to ice for 2 minutes. 500µl of preheated (42°C) SOC media was added to each reaction tube and incubated for 1 hour at 37°C on an orbital shaker at 225rpm. After this period 250µl was pipetted onto a LB agar plate with ampicillin incorporated and spread using a sterile spreader. Plates were incubated for 16-18 hours at 30°C (in order to suppress unwanted recombination between the *lox* sites of pAW8-ccdB) and successful colonies harvested.

2.5.12.10 Electroporation

Electroporation is a technique used to electrically induce pores in the cell membrane of bacteria, allowing the passage of solutions which otherwise could not cross the phospholipid bilayer (Neumann *et al.* 1982).

All curvettes and microcentrifgue tubes were placed on ice 5 minute prior to experimentation. DH5α electrocompetant bacterial cells were thawed on ice and mixed by flicking the bottom of the tube. In a cold microcentrifuge tube, 25µl of cells were mixed with 1µl of DNA and stored on ice. Cells were transferred to 2mm curvettes, the outside dried carefully and placed in the micropulser. Following the application of 250 volts, 975µl of SOC medium was added immediately to the cells and transferred to a clean microcentrifuge tube. This was incubated for 2 hours with shaking at 180rpm. 100µl was subsequently pipetted and spread onto pre-warmed plates with selective antibody.

2.5.13 *S.pombe* techniques

2.5.13.1 Growth of S.pombe

All glassware, equipment and reagents used were autoclaved before use. All cultures were incubated at 30°C.

2.5.13.2 Preparation of EMM, MMA, MEA, YEA and YEL

EMM, MMA, MEA, YEA and YEL were made up as described in section 2.2.17 and autoclaved on a liquid cycle. In the case of EMM, MMA, MEA and YEA, the solution was cooled to an appropriate temperature and appropriate supplements

and/or drug treatments being used were added. Approximately 25ml was subsequently poured into a petri dish whilst still in liquid form and allowed to set completely.

When screening selectively for strains without functional orotidine 5'-phosphate decarboxylase *(ura4$^-$),* 0.1% w/v of 5-fluoroorotic acid (5-FOA; Melford) was added to YEA whilst still in liquid form and approximately 25ml poured into a petri dish before setting completely. 5-FOA, otherwise non-toxic to *S.pombe*, is converted to the toxic form 5-fluorouracil by Ura4, resulting in positive selection for *ura4$^-$* strains. With regards to MMA plates, where *in vivo Cre* recombinase was required via controlled expression of *Cre* recombinase from pAW8, thiamine (thi, Acros Organics) was added to a final concentration of 15µM. Approximately 25ml was poured into the bottom of an 80mm petri dish and allowed to cool. All plates were stored at room temperature.

2.5.13.3 Starter cultures

For each culture, 2ml of YEL was added to a sterile glass tube. A sterile loop was used to swab colonies from growing plates and transferred to the glass tube. Cultures were left on an orbital shaker for 18-24 hours at 30°C at 180rpm.

2.5.13.4 Long term storage of S.pombe

Long term storage of *S.pombe* was achieved by mixing 600µl of glycerol with 400µl of overnight culture in YEL and freezing to -80°C. Cultures were restored by freeze thawing on ice, vortexing briefly and streaking onto YEA plates before incubating at 30°C.

2.5.13.5 Colony PCR

A small amount of appropriate colony was taken from plates using a pipette tip and suspended in 25µl of dH_2O. Samples were heated to 100°C for 5 minutes using a PCR machine to break down cell walls and membranes, following which they were centrifuged briefly and placed on ice. Standard Ampli-Taq PCR reaction mixture was added to a final volume of 25µl and placed on the thermal cycler. Thermal cycling conditions consisted of an initial denaturation of 95°C for 30

seconds, followed by 40 cycles of 95°C for 30 seconds, 50°C for 30 seconds and 68°C for 1 minute. A final elongation of 68°C for 5 minutes completed the cycle.

2.5.13.6 Extraction of genomic DNA - PCIA (25:24:1, pH 8)

The extraction of genomic DNA free of excess salts and other impurities was required for downstream applications such as sequencing. A small swab of culture was added to 2ml YEL and left to grow at 30°C with shaking at 180rpm for approximately 16-18 hours until in stationary phase. The culture was centrifuged for 1 minute at 13,000rpm and supernatant disposed of. The pellet was resuspended in 1 ml of solution A (1.2M sorbitol, 40mM EDTA, 20mM citric acid 20mM Na_2HPO_4, adjusted to pH5.6) with 500 units lyticase and incubated at 37°C for 90 minutes. This process acts to breakdown the yeast cell wall and after this time cultures were checked under the microscope to check for disruption.

The reaction was centrifuged for 30 seconds at 13,000rpm and the pellet suspended in 250µl of solution B (50mM Tris-HCl pH 7.5, 50mM EDTA and 1% SDS) and incubated at 65°C for 10 minutes, following which 250µl of solution C (50mM Tris-HCl pH 7.5, 50mM EDTA and 0.2mg proteinase K) was added. The whole reaction mixture was incubated for 90 minutes at 37°C.

1ml of PCIA (Fisher Scientific) was added to each reaction mixture in a fume hood. The reaction was mixed and centrifuged for 10 minutes at 13,000rpm. The aqueous phase was removed with a pipette and added to a clean eppendorf tube.

DNA was the precipitated by adding approximately 45µl of sodium acetate (NaAc) with 0.9ml of absolute ethanol and mixed. The reaction was left at room temperature for 10 minutes, following which DNA was precipitated by centrifuging for 10 minutes at 13,000rpm. The supernatant was removed and left to air dry for 10 minutes at room temperature. The pellet was re-suspended in 200µl of TE (pH 7.5).

Subsequently, 5µl RNase (10mg/ml; Sigma Aldrich) was added and the reaction incubated for 60 minutes at 37°C. 0.4ml of PCIA was added, mixed and centrifuged for 10 minutes at 13,000rpm. The aqueous phase was removed and transferred to a new eppendorf tube. DNA was precipitated again by adding 15µl NaAc alongside 400µl of absolute ethanol and incubating for 10 minutes at room temperature. The reaction mixture was centrifuged for 10 minutes at 13,000rpm and

supernatant discarded. The pellet was washed of excessive salt by addition of 1ml of 70% ethanol and centrifuged for a further 5 minutes at 13,000rpm. The supernatant was removed and any excess ethanol was removed by tapping the inverted eppendorf on absorbent paper and air drying the pellet at room temperature for 10 minutes. The pellet was re-suspended in 50µl of TE (pH7.5).

Extracted genomic DNA was quantified on a Qubit® 2.0 Fluorometer using high specificity buffers and standards.

2.5.13.7 Lithium acetate (LiAc) plasmid transformation

The LiAc method of plasmid transformation allows for adequate permeabilisation of the cell wall of yeast cells through actions of lithium cations to allow for the uptake of plasmid DNA (Ito *et al.* 1983).

Pre-cultures were made by adding a small amount of growing colony from plates to 5ml of YEL and incubating for 16-18 hours at 30°C with shaking at 180rpm. Between 75-150µl of the pre-culture was added to 50ml of YEL and incubated for 16-18 hours at 30°C with shaking to a titre of 1-2 x 10^7 cells/ml.

10ml of culture was added to a clean test tube and centrifuged for 5 minutes at 2,800rpm. The supernatant was removed and the pellet suspended in 20ml dH_2O to wash. A further centrifugation of 5 minutes at 2,800rpm was carried out. The supernatant was removed and the pellet suspended in 5ml TE pH 8.0/0.1M LiAc, and centrifuged for 5 minutes at 2,800rpm. The pellet was resuspended in 100µl TE pH8.0/ 0.1M LiAc, approximately 55µl of DNA sample added and incubated at 30°C for 30 minutes with soft shaking.

Following this, 0.7ml of 40% PEG 4000 and 100µl of TE with 0.1M LiAc (pH8.0) was added and incubated for 60 minutes at 30°C without shaking. 100µl DMSO was added and placed in a water bath at 45°C for 10 minutes to heat shock. This was followed by centrifugation at 2,800rpm for 5 minutes. The supernatant was carefully removed and the pellet was resuspended in 250µl dH_2O. Volumes of 50µl and 200µl were pipetted onto the centre of a selective medium plate and spread sterilely over the plate. Plates were left to air dry in the hood and then placed at 30°C for incubation.

2.5.13.8 Spot test assays – production of plates

Spot test experiments were carried out for MMS, HU and UV treatment. MMS is an alkylating agent that adds methyl groups to nitrogen atoms in purines. HU prevents the production of new nucleotides by inhibiting ribonucleotide reductase. It therefore inhibits DNA synthesis and repair by depleting the dNTP pool.

Approximately 25ml of heated liquid YEA was aliquoted into falcon tubes and allowed to cool to around 50°C. In the case of MMS and HU, appropriate volumes of drug were added to desired concentration and poured into petri dishes. All were allowed to cool and stored at room temperature for 2 days before use.

Chapter Three – Identifying novel low penetrance alleles in DNA repair genes that predispose to CRC

3.1 Introduction

Despite evidence to suggest that up to a third of all CRC cases could be due to underlying genetics, only a proportion are explained by current understanding. Approximately 6% of CRC cases can be explained by rare, high penetrance variants. These include inherited mutations in *APC* (which cause FAP; Fearnhead *et al.* 2001), *MUTYH* (MAP; Al-Tassan *et al.* 2002), *SMAD4/BMP1R1A* (JPS; Howe *et al*. 1998; Howe *et al.* 2002), *STK1/LTB1* (PJS; Aretz *et al*. 2005), *POLE and POLD1* (Palles *et al.* 2013), and various MMR genes (HNPCC; Peltomaki, 2001; as discussed in sections 1.2.1 and 1.7.2.2). It has been proposed that some of the remaining genetic risk could be due to the combined effect of multiple rare, low penetrance alleles; the so-called 'common disease-rare variant' hypothesis (Bodmer and Bonilla, 2008). Previous research has highlighted the role of rare variants in *APC, CTNNB1, AXIN1* from the Wnt signalling pathway and *MSH2* and *MLH1* from the MMR pathway as collectively contributing to an increased risk of CRA (Fearnhead *et al*. 2004; Azzopardi *et al.* 2008; Section 1.2.2.2). In addition to these, GWAS have uncovered common, low penetrance variants that significantly contribute to CRC risk. In total, 20 alleles have been associated with CRC, and despite individual variant risk being relatively low, they are likely to act in concert to significantly alter disease likelihood (Section 1.2.2.1).

Previous research has implicated the importance of DNA damage repair in the development of hereditary cancer syndromes (Section 1.3). With regards to CRC, conditions such as MAP and HNPCC are caused by underlying deficiencies in DNA repair pathways. In addition to the association between hereditary CRC and DNA repair, inactivation of the MMR gene, *MLH1,* has been shown to cause sporadic forms of CRC in up to 12% of cases as a result of the formation of a mutator phenotype (Ionov *et al*. 1993). This is due to epigenetic silencing of *MLH1* via biallelic hypermethylation of CpG islands in the promoter region (Kane *et al*. 1997; Toyota *et al*. 1999).

Using a candidate gene approach to focus on genes in the DNA repair pathways, we sought novel associations between low penetrance variants and CRC risk. To do this, we attempted to genotype every nonsynonymous variant with a MAF ≥4% in DNA repair gene in large case control cohorts.

3.2 Materials and methods

3.2.1 Samples

3.2.1.1 Training phase – aCRC cases and controls

We analysed 2,186 blood DNA samples from unrelated patients with aCRC from COIN (2,073 patients) and COIN-B (113 patients). COIN is a phase III trial, comparing two experimental arms with the control arm of oxaliplatin plus fluoropyrimidine chemotherapy in first line treatment. COIN-B is a phase II trial examining intermittent chemotherapy plus cetuximab. All patients gave fully informed consent for their samples to be used for bowel cancer research. We also analysed 2,176 blood DNA samples from healthy controls from the UK Blood Services collection of Common Controls (UKBS collection; Wellcome Trust Case Control Consortium 2007, Wellcome Trust Case Control Consortium and Australo-Anglo-American Spondylitis Consortium 2007). These samples were selected from a total of 3,092 samples within the UKBS collection that best matched the patients with aCRC in terms of place of residence within the UK (Table 3.1).

3.2.1.2 Validation phase – aCRC cases and controls

We analysed 1,053 blood DNA samples from unrelated patients with aCRC from COIN (10 patients that were not used in the training phase), COIN-B (85 patients that were not used in the training phase), FOCUS2 (361 patients), FOCUS3 (221 patients) and PICCOLO (376 patients that were not recruited into COIN or COIN-B). FOCUS2 is a trial for patients with unpretreated aCRC judged unfit for full-dose combination chemotherapy. FOCUS3 is a trial to determine the feasibility of molecular selection of therapy using *KRAS*, *BRAF* and topoisomerase-1. PICCOLO is a trial for the treatment for fluorouracil-resistant aCRC. We also analysed 1,397 blood DNA samples from unrelated healthy Caucasian controls from the UKBS collection (917 samples that were not used in the training phase) and from the

		Training phase		
		aCRC cases (n=2,186)		Controls (n=2,176)
		COIN (%) n=2,073	COIN-B (%) n=113	UKBS (%) n=2,176
Age at diagnosis (aCRC)/sampling (controls)	Mean	61.5	61.2	43.7
	<20	1 (0.0)	0	64 (2.9)
	20-49	232 (11.2)	13 (11.5)	1,317 (60.2)
	50-59	549 (26.5)	27 (23.9)	602 (27.7)
	60-69	845 (40.8)	49 (43.4)	193 (8.9)
	70-79	435 (21.0)	22 (19.5)	0
	80-89	9 (0.1)	2 (1.8)	0
	Missing	2 (0.1)	0	0
Sex	Female	698 (33.7)	48 (42.5)	1,074 (49.4)
	Male	1,375 (66.3)	65 (57.5)	1,102 (50.6)
WHO-PS	0	969 (46.7)	58 (51.3)	-
	1	951 (45.9)	46 (40.7)	-
	2	153 (7.4)	9 (8.0)	-
Primary Site	Colon	1,119 (54.0)	37 (32.7)	-
	Rectum	653 (31.5)	32 (28.3)	-

Table 3.1 – Clinicopathological data for patient/samples in COIN, COIN-B and the UKBS collection used as part of the training phase cohort

human randomised control (HRC) collection from the Health Protection Agency (480 samples; Table 3.2).

3.2.1.3 Population based analyses

We analysed 2,169 DNA samples from unrelated CRC patients from the POPGEN cohort based in Kiel, Germany. These were in comparison to 2,968 DNA samples from either the POPGEN (n=604) or Study of Health in Pomerania (SHIP; n=2,364) cohorts, based in Kiel or Greifswald, Germany respectively. These samples acted as geographically -matched healthy controls (Table 3.3). Both trials were population-based biobank projects.

In addition, we used publicly available data from another population based cohort consisting of 2,575 CRC cases (1,101 females and 1,474 males; mean age of diagnosis 59 years) recruited through the Institute of Cancer Research/Royal Marsden Hospital NHS Trust (RMHNHST) and 2,707 healthy UK controls (1,871 females and 836 males; mean age at sampling 59 years) recruited as part of the National Cancer Research Network genetic epidemiological studies (n=1,075), the Royal Marsden Hospital Trust/Institute of Cancer Research Family History and DNA Registry (n=1,033) and the UK Study of Breast Cancer Genetics (n= 599; Webb *et al.* 2006).

3.2.2 Genotyping of training phase cohort

Genotyping of the training phase cohort was carried out using Illumina's Fast-Track Genotyping Services (San Diego, CA), using their high throughput BeadArray™ technology on the GoldenGate® platform. Data was analysed and plotted using Illumina GenomeStudio v1.1.

Genes were selected from a comprehensive list of DNA repair genes (http://sciencepark.mdanderson.org/labs/wood/DNA_Repair_Genes.html) and were involved in BER, MMR, NER, HR, NHEJ, ICL repair (ICLR) or other DNA repair pathways (ODRP; Wood *et al*. 2005). Nonsynonymous variants with a MAF ≥4% were chosen through dbSNP (build version 129) or through additional literature reviews. Variant were identified by Christopher Smith and James Colley (Cardiff University; Table 3.4).

		Validation Phase						
		aCRC cases (n= 1,053)					Controls (n= 1,397)	
		COIN (%) n=10	**COIN-B (%) n=85**	**FOCUS2 (%) n=361**	**FOCUS3 (%) n=221**	**PICCOLO (%) n=376**	**UKBS Controls (%) n=917**	**HRC Controls (%) n=480**
Age at diagnosis (aCRC)/ sampling (controls)	Mean	63	62.6	-	-	-	41.3	38.6
	<20	0	0	-	-	-	24 (2.6)	0
	20-49	0	13 (15.3)	-	-	-	567 (61.8)	103 (21.4)
	50-59	2 (20)	16 (25.6)	-	-	-	253 (27.6)	13 (2.7)
	60-69	8 (80)	29 (34.1)	-	-	-	72 (7.9)	1 (0.2)
	70-79	0	24 (28.2)	-	-	-	0	0
	80-89	0	3 (3.5)	-	-	-	0	0
	Missing	0	0	-	-	-	1 (0.1)	358 (74.5)
Sex	Female	5 (50)	38 (44.7)	-	-	-	477 (52)	249 (53.9)
	Male	5 (50)	47 (55.3)	-	-	-	440 (48)	230 (47.9)
	Missing	0	0	-	-	-	0	1 (0.2)
WHO PS	0	3 (30)	38 (44.7)	-	-	-	-	-
	1	7 (70)	41 (48.2)	-	-	-	-	-
	2	0	6 (7.1)	-	-	-	-	-
Primary Site	Colon	3 (30)	55 (64.7)	-	-	-	-	-
	Rectum	7 (70)	30 (35.3)	-	-	-	-	-

Table 3.2 – Clinicopathological data for patient/samples in COIN, COIN-B, UKBS and HRC collections used as part of the validation phase cohort. Clinicopathological data for FOCUS2, FOCUS3 and PICCOLO trials were not available.

		Population based cohort		
		Cases (n=2,169)	Controls (n=2,968)	
		POPGEN (%) n=2,169	SHIP (%) n=2,364	POPGEN (%) n=604
Age at diagnosis (CRC)/ sampling (controls)	Mean	65.5	61.5	63.4
	<20	0	0	0
	20-49	179 (8.3)	302 (12.8)	1 (0.2)
	50-59	469 (21.6)	675 (28.6)	172 (28.5)
	60-69	1,066 (49.1)	761 (32.2)	235 (38.9)
	70-79	182 (8.3)	585 (24.7)	121 (20)
	80-89	46 (2.1)	41 (1.7)	0
	Missing	1 (0.04)	0	75(12.4)
Sex	Female	1,080 (49.8)	1,212 (51.3)	285 (47.2)
	Male	1,089 (50.2)	1,152 (48.7)	319 (52.8)
Primary Site	Colon	1,008 (46.4)	-	
	Rectum	904 (41.6)	-	
	Missing	257 (11.8)	-	

Table 3.3 – Clinicopathological data for the patients/samples used in the POPGEN and SHIP population based collections

Gene	Pathway	Role in pathway	Variants analysed
ATM	ODRP	Essential kinase	rs1800058 [Leu1420Phe]; rs1801516 [Asp1853Asn]; rs35813135 [Thr935Ala]
ATR	ODRP	Essential kinase	rs2227928 [Met211Thr]; rs2229032 [Arg2425Gln]; rs34124242 [Ile1526Val]
BRCA1	HR	Nuclear phosphoprotein	rs16942 [Lys1183Arg]; rs1799950 [Gln356Arg]; rs1799966 [Ser1613Gly]; rs28897674 [Ser153Arg]; rs28897687 [Asn1236Lys]; rs4986850 [Asp693Asn]; rs799917 [Pro871Leu]; rs4986852 [Ser1040Asn]
BRCA2	HR	Involved in RAD51 loading onto DNA	rs144848 [Asn372His]; rs28897708 [Ile505Thr]; rs28897727 [Asp1420Tyr]; rs28897729 [Val1542Met]; rs28897731 [Val1643Ala]; rs28897758 [Leu3101Arg]; rs1046984 [Ser599Phe]; rs28897743 [Arg2336Gln]
BRIP1	HR	Helicase with interactions with BRCA1	rs4986764 [Ser919Pro]
C19orf40	ICLR	Role in the repair of inter-strand cross links	rs2304103 [Ser158Leu]; rs3816032 [Ile192Thr]
CHAF1A	ODRP	Chromatin assembly	rs8100525 [Lys850Arg]; rs9352 [Ala923Val]
CHEK1	ODRP	Effector kinase	rs506504 [Ile471Val]
DCLRE1A	ODRP	DNA crosslink repair	rs3750898 [Asp317His]
DCLRE1B	ODRP	DNA crosslink repair	rs12022378 [His61Tyr]
DCLRE1C	NHEJ	Nuclease	rs12768894 [His243Arg]
EME1	HR	Sub-unit of nuclease	rs12450550 [Ile350Thr]; rs17714854 [Phe63Leu]
ERCC2	NER	5' to 3' DNA helicase	rs13181 [Lys751Gln], rs1799792 [His201Tyr]
ERCC4	NER	5' incision catalytic sub-unit	rs1800067 [Arg415Gln]
ERCC5	NER	3' incision DNA binding sub-unit	rs17655 [Asp1104His]; rs2227869 [Cys529Ser]
ERCC6	NER	Distortion recognition in transcription coupled repair	rs2228527 [Arg1213Gly]; rs2228528 [Gly399Asp]; rs2228529 [Gln1413Arg]; rs2228526 [Met1097Val]
EXO1	ODRP	5' exonuclease	rs12122770 [Ser610Gly]; rs1776148 [Glu670Gly]; rs4149963 [Thr439Met]; rs735943 [His354Arg]; rs9350 [Pro757Leu]
FANCA	ICLR	Part of FA core complex	rs2239359 [Gly501Ser]; rs7190823 [Thr266Ala]; rs1800282 [Val6Asp]; rs11646374 [Ala412Val]; rs7195066 [Gly809Asp]; rs9282681 [Thr1328Ala]
FANCD2	ICLR	Protein recruitment	rs3864017 [Pro714Leu]

Gene	Pathway	Role in pathway	Variants analysed
FANCE	ICLR	Part of FA core complex	rs7761870 [Ser204Leu]; rs9462088 [Ala502Thr]
FANCM	ICLR	Multiple roles in repair of ICL	rs1367580 [Val878Leu]; rs3736772 [Pro1812Ala]
FLJ35220	ODRP	Incision 3' of hypoxanthine and uracil in DNA; inosine in RNA	rs34933300 [Arg112Gln]; rs35549084 [Val29Ile]
HEL308	ODRP	DNA Helicase	rs1494961 [Val306Ile]
LIG1	BER and MMR	DNA ligase – repairs nicks in ssDNA	rs3730947 [Val349Met]
LIG4	NHEJ	DNA ligase – repairs nicks in ssDNA	rs1805388 [Thr9Ile]
MDC1	ODRP	Recruitment of proteins to areas of damage	rs9262152 [Arg268Lys]
MGMT	ODRP	Methyltransferase that directly repairs DNA damage	rs12917 [Leu84Phe]; rs2308321 [Ile143Val]
MLH1	MMR	Part of mismatch and loop recognition heterocomplex MutL	rs1799977 [Ile219Val]
MLH3	MMR	Part of loop recognition heterocomplex MutL	rs175080 [Pro844Leu]; rs28756982 [Val420Ile]; rs17782839 [Ser966Pro]
MMS19	NER	Roles in stabilising and recruiting proteins	rs29001285 [Val197Ile]; rs3740526 [Gly790Asp]
MSH3	MMR	Part of loop recognition heterocomplex MutS	rs184967 [Gln949Arg]; rs26279 [Ala145Thr]; rs1650697 [Ile79Leu]
MSH4	MMR	MutS homolog	rs5745459 [Tyr589Cys]; rs5745549 [Ser914Asn]; rs5745325 [Ala97Thr]
MSH5	MMR	MutS homolog	rs1802127 [Pro786Ser]; rs28381349 [Leu85Phe]
MUS81	HR	Subunit of a structure specific nuclease	rs13817 [Arg37His]; rs545500 [Arg180Pro]

Gene	Pathway	Role in pathway	Variants analysed
MUTYH	BER	DNA glycosylase	rs3219484 [Val22Met]; rs3219489 [Gln335His]
NBN	HR	Acts in complex to repair double strand breaks	rs1805794 [Glu103Gln]
NEIL3	BER	DNA glycosylase	rs13112390 [Gln471His]; rs1876268 [Gly520Arg]; rs34193982 [His286Arg]; rs7689099 [Pro177Arg]
OGG1	BER	DNA glycosylase	rs1052133 [Ser326Cys]; rs17050550 [Ala85Ser]
PARP1	ODRP	Poly-ADP-ribosylation protein	rs1136410 [Val762Ala]
PARP2	ODRP	Poly-ADP-ribosylation protein	rs3093921 [Asp186Gly]; rs3093926 [Arg247Gln]
PMS2	MMR	Part of mismatch recognition heterocomplex MutL	rs2228006 [Lys541Glu]; rs1805321 [Pro470Ser]
POLE	NER, MMR	DNA polymerase	rs5744934 [Asn1396Ser]; rs5744751 [Ala252Val]
POLG	BER	DNA polymerase in mitochondrial DNA	rs3087374 [Gln1236His]
POLI	ODRP	DNA polymerase involved in lesion bypass	rs8305 [Ala706Thr]
POLL	NHEJ	Gap filling DNA polymerase	rs3730463 [Thr221Pro]; rs3730477 [Arg438Trp]
POLM	NHEJ	Gap filling DNA polymerase	rs28382644 [Gly220Ala]
POLN	ODRP	DNA polymerase	rs10011549 [Gly336Ser]; rs11725880 [Pro315Ser]; rs2353552 [Gln121His]; rs9328764 [Arg425Cys]
POLQ	ODRP	DNA polymerase	rs1381057 [Gln2513Arg]; rs3218634 [Leu2538Val]; rs3218649 [Thr982Arg]; rs3218651 [His1201Arg]; rs487848 [Ala581Val]; rs532411 [Ala2304Val]
PRKDC	NHEJ	Catalytic subunit of a DNA kinase	rs8178017 [Met333Ile]
RAD1	BER	Sub-unit of 9-1-1 complex DNA damage sensor	rs1805327 [Glu281Gly]
RAD17	ODRP	DNA damage sensor	rs1045051 [Leu546Arg]
RAD18	ODRP	Ubiquitin ligase	rs373572 [Arg302Gln]

Gene	Pathway	Role in pathway	Variants analysed
RAD23B	NER	Recognise DNA distortion	rs1805329 [Ala249Val]
RAD51L1	ODRP	Involved in recruitment of proteins	rs34594234 [Lys243Arg]
RAD51L3	HR	Role in early stages of DNA strand pairing	rs4796033 [Arg165Gln]
RAD52	HR	Accessory factor in recombination	rs7487683 [Gly180Arg]
RDM1	HR	Repair of double strand breaks	rs2251660 [Cys127Trp]
RECQL5	HR, ODRP	DNA helicase	rs820196 [Asp453Gly]
REV1	ODRP	Scaffold for DNA polymerases	rs3087386 [Phe257Ser]; rs3087399 [Asn373Ser]; rs3087403 [Val138Met]
REV3L	ODRP	Catalytic subunit of POLZ	rs3204953 [Val2986Ile]; rs458017 [Tyr1078Cys]; rs462779 [Thr1146Ile]
RPA1	NER	Pre-incision complex	rs5030755 [Thr351Ala]
TDG	BER	DNA glycosylase	rs2888805 [Val367Leu]
TDP1	ODRP	Repair of DNA topisomerase cross links	rs28365054 [Ala134Thr]
TP53	ODRP	Critical in regulation of cell cycle	rs1042522 [Pro72Arg]
WRN	HR,	Helicase and 3'-exonuclease	rs1346044 [Cys1367Arg]; rs1800391 [Met387Ile]; rs2230009 [Val114Ile]; rs2725362 [Leu1074Phe]
XPC	BER	Recognise DNA distortion	rs2228000 [Arg500Trp]; rs2228001 [Gln940Lys]
XRCC1	BER	Scaffold protein for LIG3	rs1799782 [Arg194Trp]; rs25487 [Gln399Arg]
XRCC2	HR	DNA cross link and break repair	rs3218536 [Arg188His]
XRCC3	HR	DNA cross link and break repair	rs861539 [Thr241Met]
XRCC4	NHEJ	Ligase accessory factor	rs28360135 [Ile134Thr]

Table 3.4 –DNA repair genes with nonsynonymous variants with a MAF ≥4% assayed in the training phase cohort. Variants in each gene shown with rs numbers, followed by the amino acid substitution (in parentheses). Those highlighted in red failed genotyping on the GoldenGate platform

Genotyping of *TTC23L*^{His22Arg} (rs6451173) in the training phase cohort was carried out using KASPar technology by KBioscience (Hoddesdon, Hertfordshire, UK)

3.2.3 Genotyping of validation phase cohort

Genotyping of *RAD1*^{Glu281Gly} (rs1805327), polymerase γ (*POLG)*^{Gln1236His} (rs3087374) and *REV1* ^{Val138Met} (rs3087403) in the validation phase cohort was carried out using KASPar technology.

3.2.4 Genotyping of POPGEN samples

Genotyping of *RAD1*^{Glu281Gly}, *POLG*^{Gln1236His}, *REV1*^{Val138Met}, *BRCA1*^{Leu871Pro} (rs799917) and *ERCC6*^{Arg1213Gly} (rs2228527) in the population cohort was carried out using Taqman genotyping assays. Assays were analysed using either the Applied Biosystems 7900HT Real-Time PCR system (Germany) or Applied Biosystems 7500 Real-Time PCR system (Cardiff) and data was analysed using Applied Biosystems Sequence Detection Software (SDS) Software v2.3.

3.2.5 PCR and Sanger sequencing

The entire open reading frame (ORF), flanking intronic sequences and the 5'UTR of *RAD1,* tetratricopeptide repeat protein 23-like (*TTC23L*)*,* DnaJ homolog, subfamily C, member 21 (*DNAJC21*) and ribosome genesis protein (*BRIX1*) were amplified by PCR. PCR, verification by agarose gel electrophoresis, product purification, Sanger sequencing and sequencing clean up were carried out as described in sections 2.5.4 to 2.5.9. Sequences were analysed using Sequencer v4.6. All primers used are given in Appendices 1-4.

3.2.6 Real time PCR

We carried out real time PCR (RT-PCR) of alanine--glyoxylate aminotransferase 2 (*AGXT2*)*, BRIX1, DNAJC21, RAD1* and *TTC23L* using colon and kidney first strand cDNA. Two set of intron spanning primers for each gene were utilised to gauge expression (Appendix 5; Fig. 3.1). Primers for *β-actin* from Stratagene were used as a positive control. PCR (Section 2.5.4) was carried out with conditions consisting of an initial denaturation of 95°C for 2 minutes, followed by 40

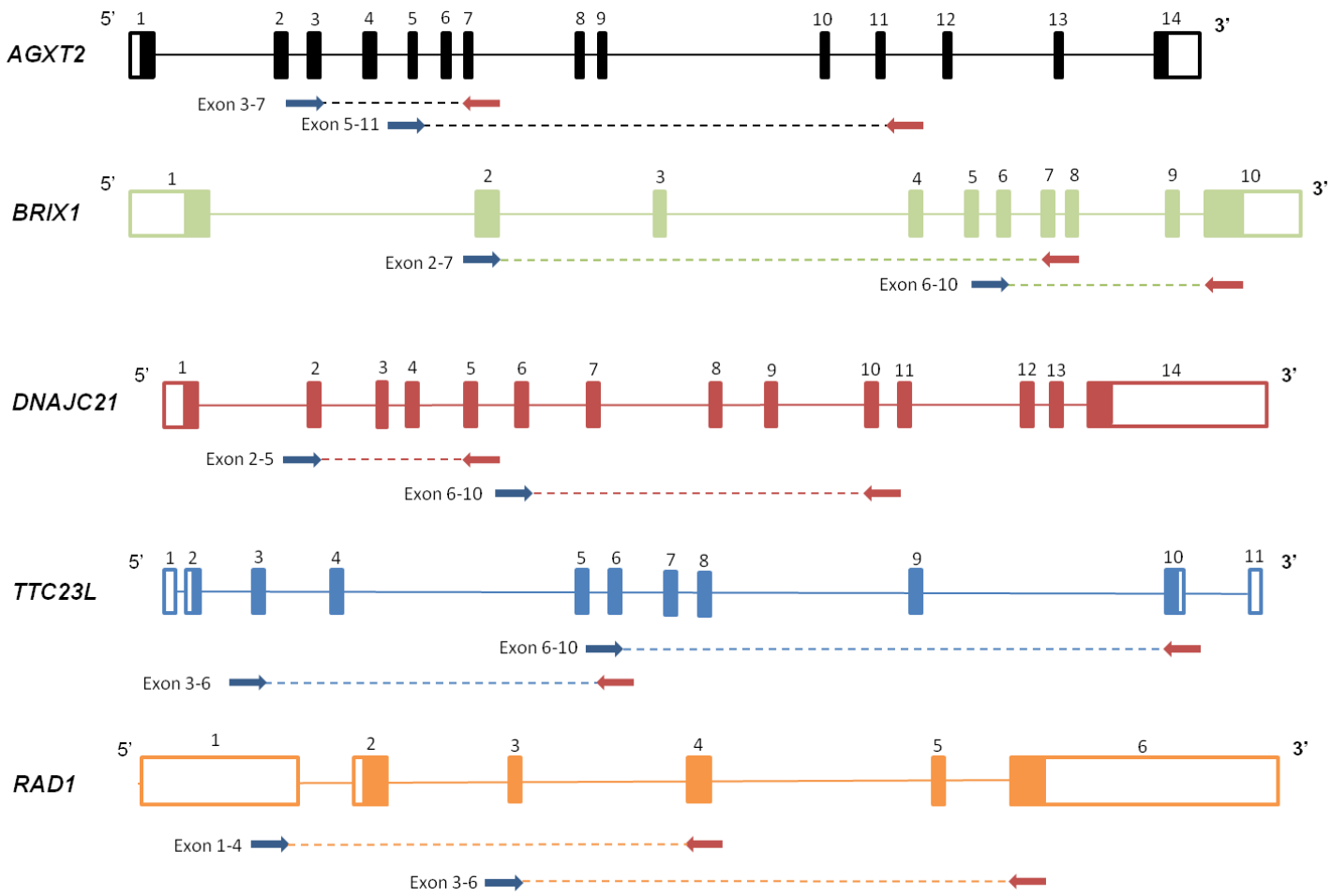


Figure 3.1 – Schematic to demonstrate approximate size and structure of genes analysed with intron spanning primers to assay for gene expression in colonic and renal cDNA. Blue arrow represents forward primers and red arrow represents reverse primers. Closed boxes represent ORF, whilst open boxes represent non-coding exonic regions and horizontal line represents intronic regions

cycles of 95°C for 1 minute, 55°C for 1 minute and 72°C for 2 minutes, with a final elongation of 72°C for 10 minutes. Products were analysed on 1.5% agarose gels (Section 2.5.5).

3.2.7 *In silico* analysis of variants

LD between variants was assessed using Haploview v4.2. Species alignment of all mammals listed on NCBI was carried out using Clustal Omega. A list of common specie names are given in Appendix 20. Prediction of the damaging effects of coding variants on protein function was carried out using SIFT, Polyphen and Align-GVGD.

3.2.8 Statistical analyses

Single marker association analyses and meta-analyses were performed using PLINK v1.07 (Purcell *et al*. 2007). Meta-analysis was also performed using Comprehensive Meta-Analysis program (Biostat). Variants were analysed using Pearsons Chi square (X^2) test for association under an allelic model (1 degree of freedom [d.f]), dominant (1d.f), recessive (1d.f) and genotypic model (2d.f). Violation of the HWE was also assessed. Correction for multiple testing was carried out using the Bonferroni test. Logistic regression was used to analyse data dependant on sex and age for the training phase cohort only since inadequate data was available for validation phase cohort.

3.2.9 Exclusion criteria for samples

Following review of patient notes and medical records, 40 patients of a non-Caucasian background in the training phase cohort were identified and subsequently removed to avoid population stratification.

3.3 Results

3.3.1 Utility of the training phase cohort to identify CRC susceptibility alleles

Our training phase cohort of 2,186 unrelated British patients with aCRC and 2,176 geographically-matched unrelated healthy British Caucasian controls has recently been used to help identify and validate novel CRC-susceptibility loci

(Houlston *et al*. 2010; Dunlop *et al.* 2012a). To further demonstrate the utility of this cohort to identify CRC-susceptibility alleles, we assayed a single genome-wide significant variant from ten of the known loci identified from GWAS of CRC risk alleles. Any cases identified as being of non-Caucasian origin (n=40) were excluded from this analyses. Genotyping concordance rates for duplicate samples (n=55) in the Golden Gate assay were 100% (550/550 genotypes were concordant) and GenTrain scores for the ten variants analysed ranged from 0.68-0.91. The overall genotyping success rate was 98.5% (42,982/43,620 genotypes called successfully; Fig. 3.2).

We independently validated five of these loci using the training phase samples (Table 3.5). The OR observed in this study were all in the same direction to those given in Houlston *et al.* (2008).

3.3.2 Identifying novel variants associated with CRC – Training phase cohort

We attempted to assay every nonsynonymous variant with a MAF ≥4% in the training phase cohort in every DNA repair gene in the human genome (Wood *et al.* 2005). Based on the number of samples in our training phase cohort, we had 72% power to detect a variant with a MAF of at least 4%, with an OR of 1.3 (with 5% significance levels). This effect size was chosen to calculate power because the largest OR seen from current GWAS for CRC was 1.28. We excluded samples known at the time to be of non-Caucasian ethnicity (n=40).

We identified 180 nonsynonymous variants with a MAF ≥4% in DNA repair genes. Of these, 36 failed *in silico* locus conversion. Accordingly, 144 variants were genotyped representing 71 genes, of which 17 failed genotyping, meaning that we successfully genotyped 127 variants representing 68 genes. Genotyping concordance rates for duplicate samples (n=55) was 100% (6,985/6,985 genotypes were concordant), GenTrain scores ranged from 0.47 to 0.97 and the overall genotyping success rate was 99.73% (556,037/557,530 genotypes were called successfully).

Three variants were in violation of HWE (rs175080, rs34193982 and rs34594234 at *P*=0.05). However, when corrected for multiple testing using the

Variant	Chr	Minor allele (A)	Major allele (B)	Cases				Controls				Dom X^2 *P*	Rec X^2 *P*	Geno X^2 *P*	Allelic X^2	*P*	OR	L95	U95
				AA	AB	BB	MAF	AA	AB	BB	MAF								
rs4939827	18q21	C	T	419	1062	661	0.44	504	1113	558	0.49	$1.4x10^{-4}$	$3.8x10^{-3}$	$1.6x10^{-4}$	16.85	$4.04x10^{-5}$	0.84 (0.85)	0.77	0.91
rs16892766	8q23	C	A	13	389	1742	0.10	14	299	1862	0.08	$1.1x10^{-4}$	0.88	$4.1x10^{-4}$	12.8	$3.39x10^{-4}$	1.32 (1.32)	1.13	1.53
rs4779584	15q13	T	C	113	715	1316	0.22	81	677	1413	0.19	$1.5x10^{-2}$	$1.1x10^{-2}$	$8.3x10^{-3}$	9.06	$2.61x10^{-3}$	1.17 (1.19)	1.06	1.3
rs10795668	10p14	A	G	194	904	957	0.31	226	968	889	0.34	$1.1x10^{-2}$	0.13	$3.1x10^{-2}$	6.59	$1.02x10^{-2}$	0.89 (0.89)	0.81	0.97
rs6983267	8q24	T	G	435	1029	675	0.44	483	1067	617	0.47	$2.7x10^{-2}$	0.12	$6.0x10^{-2}$	5.5	$1.9x10^{-2}$	0.90 (0.83)	0.83	0.98
rs961253	20p12	A	C	303	1019	821	0.38	281	1008	886	0.36	0.1	0.24	0.21	3.08	0.08	1.08 (1.13)	0.99	1.18
rs9929218	16q22	A	G	172	865	1105	0.28	181	930	1064	0.30	0.08	0.73	0.21	2.3	0.13	0.93 (0.88)	0.85	1.02
rs4444235	14q22	C	T	507	1044	589	0.48	459	1105	611	0.47	0.68	0.04	0.12	2.16	0.14	1.07 (1.12)	0.98	1.16
rs3802842	11q23	C	A	218	899	1027	0.31	216	874	1085	0.30	0.19	0.8	0.42	1.25	0.26	1.05 (1.21)	0.96	1.15
rs10411210	19q13	T	C	18	362	1764	0.09	23	356	1795	0.09	0.8	0.46	0.7	0.003	0.95	1.00 (0.79)	0.87	1.16

Table 3.5 – Training phase data for variants and chromosomal position (chr) previously identified through GWAS. Variants were analysed using the Chi square test under dominant (dom), recessive (rec), genotypic (geno) and allelic models. Minor allele frequency (MAF), P values (P), odds ratios (OR) and lower (L95) and upper (U95) 95% confidence intervals (CI) were all calculated. Non-Caucasian samples were removed from analysis (n=40). OR from Houlston et al. (2008) shown in parentheses.

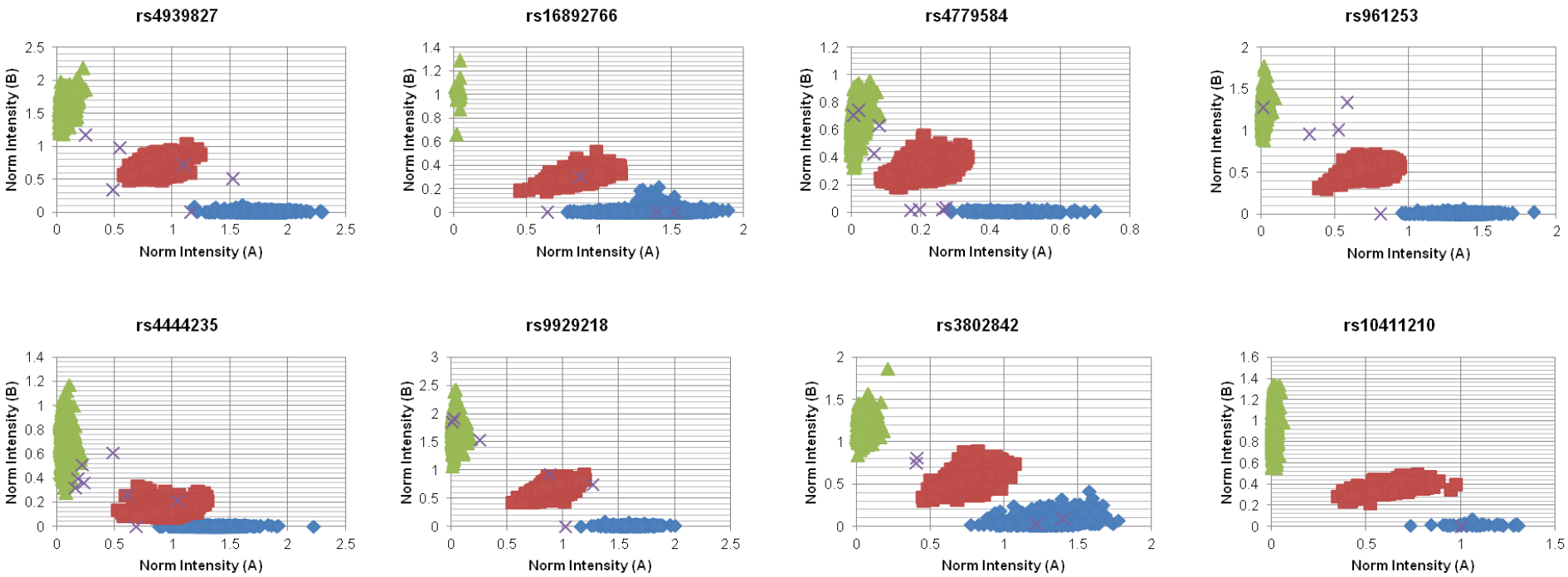


Figure 3.2 – Examples of genotype cluster plots for variants genotyped in the training phase cohort which had previously been identified as alleles associated with CRC risk through GWAS. Plotted using data generated from Illumina GenomeStudio v1.1. Blue are samples identified as AA, red as AB and green as BB.

Bonferroni technique, all variants were within HWE. Under an allelic model, we found that 9 variants, representing 7 genes, were significantly over-represented at the 5% level (genotyping plots given in Fig. 3.3). Only *RAD1*^{Glu281Gly} (rs1805327) remained significant after Bonferroni correction for multiple testing (*P*= 0.03).

LD was assessed between the variants in *ERCC6* (rs2228527 and rs2228529) and *BRCA1* (rs16942, rs799917 and rs1799966) using Haploview. High LD (r^2 and/or D' >0.8) was observed between the two variants in *ERCC6* (r^2=0.99, D'=1.0), as well as between the three variants in *BRCA1* (rs1799966-rs799917, r^2=0.88, D'=0.97; rs1799966-rs16942, r^2=0.99, D' = 0.99; rs16942-rs1799917, r^2=0.9, D'=0.98).

Variants in *RAD1, POLG, REV1* and *FANCA* were all over-represented in controls suggesting a protective effect, whereas variants in *BRCA1* and *ERCC6* were over-represented in cases.

Following adjustment by logistic regression for sex and age, *REV1*^{Val138Met}, *ERCC6*^{Arg1213Gly}, *ERCC6*^{Gln1413Arg}, *BRCA1*^{Pro871Leu}, *BRCA1*^{Lys1183Arg} and *BRCA1*^{Ser1613Gly} remained significant. However, following Bonferroni correction for multiple testing, none remained significant.

3.3.3 Identifying novel variants associated with CRC – Validation phase cohort

We screened our validation phase cohort for the most significant variant identified in the training phase cohort (*RAD1*^{Glu281Gly}). We also genotyped *POLG*^{Gln1236His} and *REV1*^{Val138Met} in 846 aCRC patients from this cohort due to less available PICOLLO samples at the time of genotyping (all controls from the validation cohort were genotyped for these variants). Genotyping was carried out using KBiosciences KASPar technology and data was subsequently analysed using PLINK. Genotyping concordance rates for duplicate samples was 100% (66/66 genotypes were concordant), and overall genotyping success rate was 93.9% (5,756/6,132 genotypes were called successfully). For *RAD1*^{Glu281Gly} (Fig. 3.4), 102 samples failed genotyping using KASPar technology. We therefore determined the genotypes of these samples by directly amplifying the target region by PCR and

Variant	Gene	Amino acid change	Minor allele (A)	Major allele (B)	Cases				Controls				Dom X^2 *P*	Rec X^2 *P*	Geno X^2 *P*	Allelic X^2	*P*	OR	L95	U95
					AA	AB	BB	MAF	AA	AB	BB	MAF								
rs1805327	*RAD1*	Glu281Gly	G	A	10	245	1887	0.06	9	340	1825	0.08	$8.6x10^{-5}$	0.79	$2.9x10^{-4}$	13.51	$2x10^{-4}$	0.73	0.62	0.87
rs3087374	*POLG*	Gln1236His	A	C	13	286	1844	0.07	15	350	1810	0.09	0.001	0.73	0.04	6.21	0.01	0.82	0.70	0.96
rs3087403	*REV1*	Val138Met	A	G	151	842	1149	0.27	172	918	1085	0.29	0.01	0.28	0.04	5.72	0.02	0.89	0.81	0.98
rs2228527	*ERCC6*†	Arg1213Gly	G	A	92	666	1385	0.2	81	624	1470	0.18	0.04	0.34	0.11	4.37	0.04	1.12	1.01	1.25
rs799917	*BRCA1*‡	Pro871Leu	A	G	243	975	925	0.34	237	917	1021	0.32	0.01	0.64	0.04	4.35	0.04	1.10	1.01	1.20
rs16942	*BRCA1*‡	Lys1183Arg	G	A	228	968	945	0.33	222	913	1040	0.31	0.02	0.63	0.05	4.19	0.04	1.10	1	1.20
rs1800282	*FANCA*	Val6Asp	A	T	10	353	1779	0.09	25	384	1766	0.10	0.11	0.01	0.02	4.11	0.04	0.86	0.74	1.00
rs2228529	*ERCC6*†	Gln1413Arg	G	A	92	664	1386	0.20	81	626	1468	0.18	0.05	0.34	0.14	3.97	0.05	1.12	1.00	1.24
rs1799966	*BRCA1*‡	Ser1613Gly	G	A	230	971	942	0.33	225	916	1034	0.31	0.02	0.68	0.06	3.89	0.05	1.10	1.00	1.20

Table 3.6 - Most significant low penetrance DNA repair variants in the training cohort, analysed under dominant, recessive, genotypic and allelic models. A common key is given in table 3.5. † Strong LD was seen between two variants in ERCC6 (r^2=0.99, D'=1.0). ‡ Strong LD was seen between three variants in BRCA1 (rs1799966-rs799917, r^2=0.88, D'=0.97; rs1799966-rs16942. r^2=0.99, D' = 0.99; rs16942-rs1799917, r^2=0.9, D'=0.98)

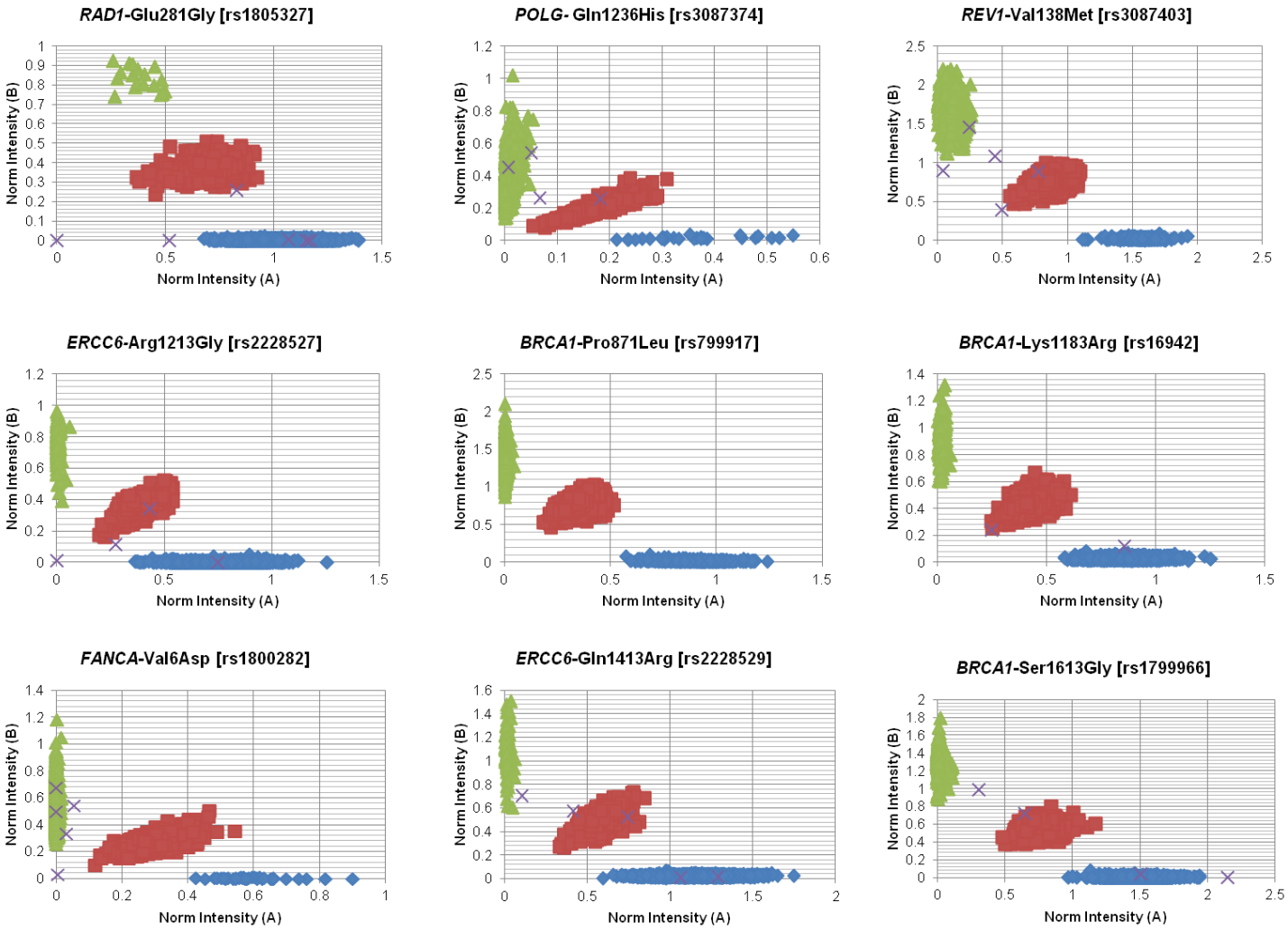


Figure 3.3 – Genotype cluster plots for most significant low penetrance DNA repair variants in the training phase cohort. Plotted using data generated from Illumina GenomeStudios v1.1. A common key is given in figure 3.2.

Sanger sequencing in house. Of the 102 samples analysed, we were able to successfully amplify, sequence and genotype 56.

All variants were within HWE. All were over-represented in controls, concordant with the training phase cohort. However, none reached statistical significance under an allelic model (Table 3.7). We were unable to adjust for sex and age due to insufficient data for cases in this cohort to carry out the analysis.

3.3.4 Population based cohorts – POPGEN and RMHNHST

3.3.4.1 POPGEN

Using TaqMan SNP genotyping assays, we genotyped the POPGEN cohort for the most significant variants from the training phase cohort. These included *RAD1*^{Glu281Gly}, *POLG*^{Gln1236His} and *REV1*^{Val138Met}. In addition, we also genotyped *BRCA1*^{Pro871Leu} and *ERCC6*^{Arg1213Gly}, as these variants were the most significant of the tagging variants in their respective genes.

Taqman assays for *REV1*^{Val138Met} *and POLG*^{Gln1236His} were first set up in Cardiff to gauge the robustness of the technology. A selection of COIN samples from the training phase were chosen based on known genotypes (n=101 for *POLG*; n=105 for *REV1*). All Taqman genotype data was 100% concordant for each variant with the Illumina GoldenGate data.

Genotyping of the POPGEN cohort was carried out in Germany by myself. Overall genotyping success rate was 89.2% (22,923/25,685 genotypes were called correctly). All variants were in accordance with HWE. All variants were more frequent in controls, which was concordant with the training cohort for the variants in *POLG, REV1* and *RAD1* but not for the variants in *ERCC6* and *BRCA1. POLG*^{Gln1236His} was the only variant that was statistically significant under an allelic model (Table 3.8).

3.3.4.2 RMHNHST

Publicly available data for 2,575 CRC cases and 2,707 healthy controls was examined for variants identified as over-represented in the training cohort. In this published study, low penetrance susceptibility alleles were sought by assaying

Variant	Gene	Amino acid change	Minor allele (A)	Major allele (B)	Cases				Controls				Dom X^2 *P*	Rec X^2 *P*	Geno X^2 *P*	Allelic X^2	*P*	OR	L95	U95
					AA	AB	BB	MAF	AA	AB	BB	MAF								
rs1805327	*RAD1*	Glu281Gly	G	A	2	132	900	0.07	13	189	1169	0.08	N/A	N/A	N/A	2.76	0.1	0.83	0.66	1.03
rs3087374	*POLG*	Gln1236His	A	C	1	108	702	0.07	8	198	1071	0.08	N/A	N/A	N/A	3.54	0.06	0.8	0.63	1.01
rs3087403	*REV1*	Val138Met	A	G	55	319	437	0.27	107	562	700	0.28	0.37	0.21	0.4	1.83	0.18	0.91	0.79	1.04

Table 3.7 - Results of genotyping three variants in the validation cohort, analysed under dominant, recessive, genotypic and allelic models, where applicable. A common key is given in table 3.5.

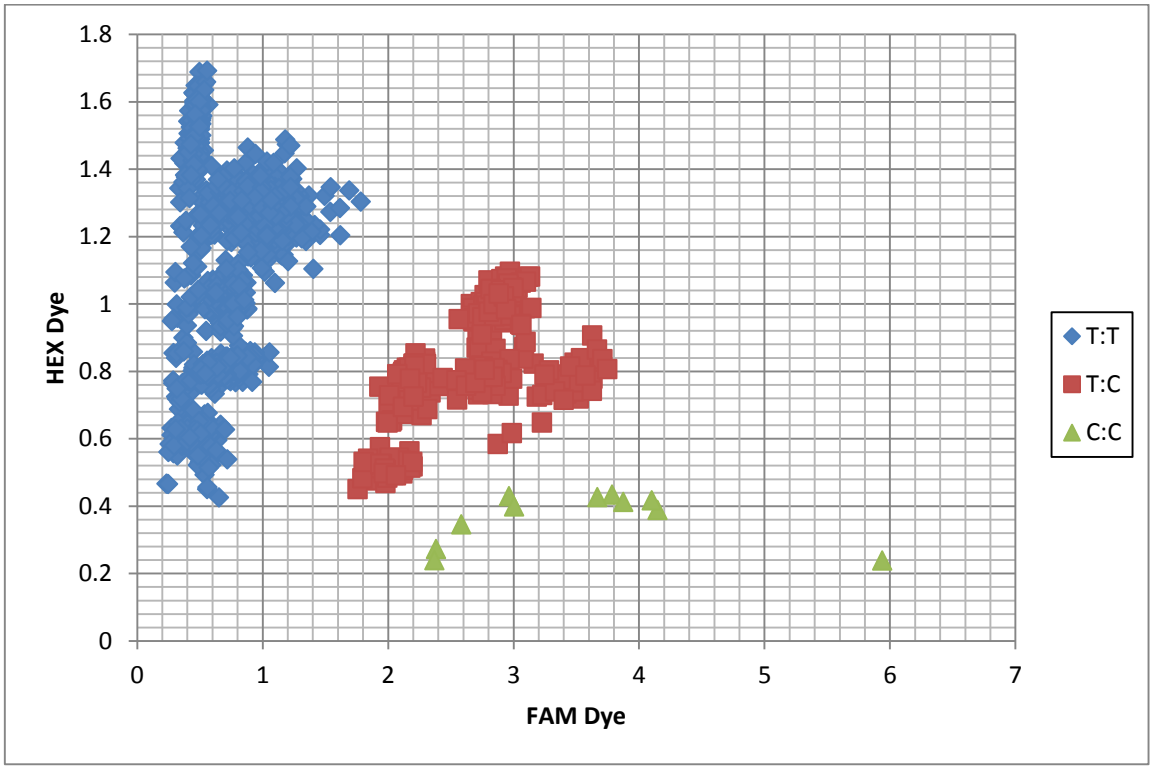


Figure 3.4 – Genotype cluster plot from KASPar genotyping of RAD^{Glu281Gly} (rs1805327)

Variant	Gene	Amino acid change	Minor allele (A)	Major allele (B)	Cases				Controls				Dom X^2 *P*	Rec X^2 *P*	Geno X^2 *P*	Allelic X^2	*P*	OR	L95	U95
					AA	AB	BB	MAF	AA	AB	BB	MAF								
rs1805327	*RAD1*	Glu281Gly	G	A	13	283	1776	0.08	24	361	2313	0.08	0.99	0.31	0.57	0.05	0.82	0.98	0.84	1.15
rs3087374	*POLG*	Gln1236His	A	C	8	291	1813	0.07	22	412	2279	0.08	0.08	0.06	0.06	4.21	0.04	0.85	0.73	0.99
rs3087403	*REV1*	Val138Met	A	G	155	777	1187	0.26	182	1094	1445	0.27	0.04	0.39	0.04	1.61	0.21	0.94	0.86	1.03
rs2228527	*ERCC6*	Arg1213Gly	G	A	106	725	1261	0.22	160	960	1609	0.24	0.36	0.23	0.41	1.49	0.22	0.94	0.86	1.03
rs799917	*BRCA1*	Pro871Leu	A	G	221	906	977	0.32	303	1200	1222	0.33	0.27	0.5	0.51	1.32	0.25	0.95	0.87	1.04

Table 3.8 - Results of genotyping of variants in the POPGEN cohort, analysed under dominant, recessive, genotypic and allelic models. A common key is given in table 3.5.

nonsynonymous variants that were predicted to be deleterious to protein function using the predicted impact of coding variants (PICS) database, PolyPhen and SIFT. Genotyping was carried out by customised Illumina Sentrix bead array assays. In total, 1,467 variants were submitted for genotyping and 1,218 variants were successfully genotyped and analysed. Six variants previously identified in the training phase cohort were analysed (*RAD1*^{Glu281Gly}, *REV1*^{Val138Met}, *BRCA1*^{Lys1183Arg}, *BRCA1*^{Ser1613Gly}, *ERCC6*^{Arg1213Gly} and *ERCC6*^{Gln1413Arg}), which included two variants in both *BRCA1* and *ERCC6* previously shown to be in LD with each other. No variants were over-represented in this cohort (Table 3.9).

3.3.5 Meta-analysis

To enhance the power to detect an association between variants and CRC risk we conducted a meta-analysis of various cohorts for *RAD1*^{Glu281Gly}, *POLG*^{Gln1236His} and *REV1*^{Val138Met}. For each meta-analysis carried out, Cochran's *Q* statistic to test for heterogeneity (*Q)* was used and the I^2 statistic was calculated to determine the proportion of variation due to heterogeneity. A large degree of heterogeneity is typically indicated be an $I^2 \geq 75\%$ and in situations where this arises, a random effects model is typically considered. We observed no significant heterogeneity in any of the meta-analysis carried out and a fixed effects model was used for all.

3.3.5.1 RAD1^{Glu281Gly}

Pooling the data from the four cohorts analysed suggested that *RAD1*^{Glu281Gly} was associated with CRC risk ($P=2x10^{-3}$). A separate analysis of aCRC cohorts only revealed an association between the variant and risk ($P=8.2x10^{-5}$). However, there was no association following the pooling of data from early stage CRC cohorts only ($P=0.43$; Fig. 3.5A).

3.3.5.2 POLG^{Gln1236His}

Pooling the data from the three cohorts analysed suggested that *POLG*^{Gln1236His} was associated with CRC risk ($P=1.9x10^{-3}$). A separate analysis of aCRC cohorts only revealed an association between the variant and risk ($P=2.2x10^{-4}$; Fig. 3.5B).

Variant	Gene	Amino acid change	Minor allele (A)	Major allele (B)	Cases				Controls				Dom X^2 *P*	Rec X^2 *P*	Geno X^2 *P*	Allelic X^2	*P*	OR	L95	U95
					AA	AB	BB	MAF	AA	AB	BB	MAF								
rs1805327	*RAD1*	Glu281Gly	G	A	10	328	2223	0.07	13	363	2314	0.07	0.41	0.61	0.66	0.76	0.38	0.94	0.81	1.09
rs3087403	*REV1*	Val138Met	A	G	189	1008	1364	0.27	222	1085	1385	0.28	0.19	0.24	0.3	2.35	0.13	0.94	0.86	1.02
rs2228527	*ERCC6*	Arg1213Gly	G	A	88	847	1624	0.2	113	862	1718	0.2	0.8	0.15	0.29	0.07	0.79	0.99	0.9	1.08
rs16942	*BRCA1*	Lys1183Arg	A	G	253	1138	1166	0.32	273	1193	1229	0.32	0.98	0.84	0.98	0.02	0.9	0.99	0.92	1.08
rs2228529	*ERCC6*	Gln1413Arg	G	A	87	846	1624	0.2	113	861	1721	0.2	0.79	0.13	0.26	0.08	0.78	0.99	0.9	1.09
rs1799966	*BRCA1*	Ser1613Gly	G	A	256	1141	1160	0.32	277	1196	1222	0.33	0.99	0.81	0.97	0.03	0.87	0.99	0.92	1.08

Table 3.9 – RMHNHST genotyping data available online (ICR - SNPlink database; http://www.icr.ac.uk/research/team_leaders/Houlston_Richard/Houlston_Richard_RES/SNPLINK/index.shtml) from a population based cohort, analysed under dominant, recessive, genotypic and allelic models. A common key is given in table 3.5.

3.3.5.3 REV1^{Val138Met}

Pooling of the data from the four cohorts analysed suggested that *REV1*^{Val138Met} was associated with CRC risk (P=1x10^{-3}). A separate analysis of aCRC cohorts only revealed an association between the variant and risk (P=6.2x10^{-3}). When data from early stage CRC cohorts was pooled, a borderline significant association was observed (P=0.05; Fig. 3.5C).

3.3.6 *In silico* analysis

The glutamic acid at residue position 281 in RAD1 was conserved in multiple species (Appendix 21), although conservation throughout species was not complete. *In silico* analyses suggest that the glycine substitution has an effect on function with a PolyPhen score of 1.586 (possibly damaging), an Align-GVGD score of C65 (GD 97.85) (likely to interfere with function) and a SIFT score of 0.03 (affects protein function).

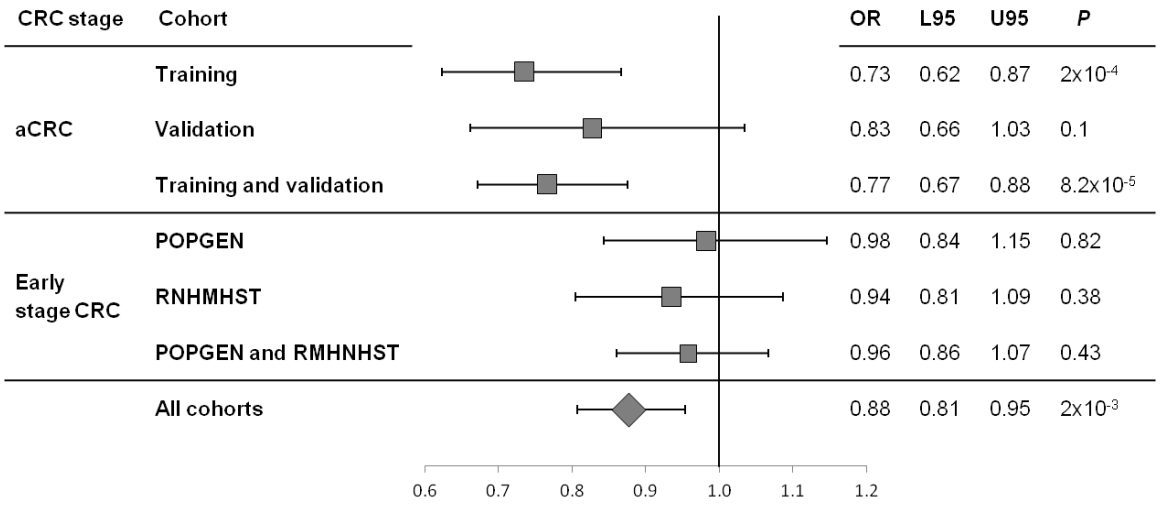
The glutamine at residue position 1236 in POLG was conserved in multiple species (Appendix 22), although conservation throughout species was not complete. However, *in silico* analyses suggest that the histidine substitution is unlikely to affect function, with a PolyPhen score of 0.80 (possibly damaging), an Align-GVGD score of C15 (less likely to interfere with function) and a SIFT score of 0.12 (tolerated).

The valine at residue position 138 in REV1 was conserved in multiple species (Appendix 23) although conservation throughout species was not complete. However, *in silico* analysis suggests that the methionine substitution is unlikely to affect function, with a PolyPhen score of 0.019 (benign), an Align-GVGD score of C15 (less likely to interfere with function) and a SIFT score of 0.11 (tolerated).

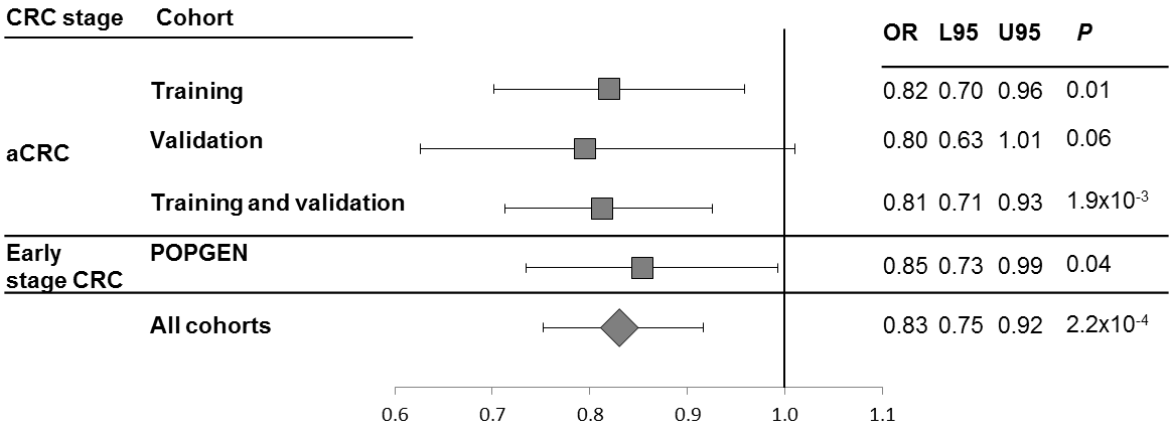
3.3.7 Sequencing of *RAD1*

In order to seek potential casual variants that may be in LD with *RAD1*^{Glu281Gly}, we sequenced the entire ORF, flanking intronic regions and 5'UTR of *RAD1* in twenty five aCRC patients carrying the risk allele. Ten of the patients carried alleles

A.



B.



C.

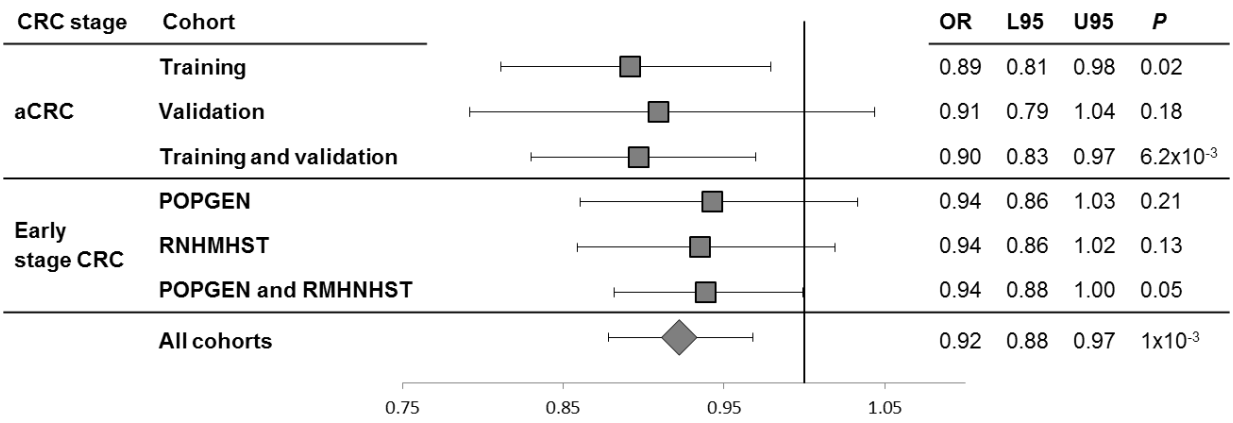


Figure 3.5 – Forest plots of effect size associated with various cohorts and in meta-analyses for A) RAD1^{Glu281Gly} B) POLG^{Gln1236His} C) REV1^{Val138Met}. Closed boxes represent odds ratios (OR) with horizontal lines displaying lower (L95) and upper (U95) confidence intervals (CI) with P values for meta-analysis calculated under a fixed effects model.

encoding Gly/Gly and fifteen carried alleles encoding Glu/Gly. Sample numbers were based on 95% power to detect a variant with a MAF in controls of 8%.

We found two nonsynonymous variants in *RAD1* (Gly114Asp [rs2308957, MAF in dbSNP = 0.5%] and Thr104Ser [rs1805328, MAF in dbSNP = 1.2%]), each in a single sample.

3.3.8 Analyses of genes tagged by *RAD1*^{Glu281Gly}

RAD1^{Glu281Gly} lies in a 62kb LD block that encompasses four other genes (*BRIX1, DNAJC21, TTC23L* and *AGXT2*). We considered whether tagging variants within these genes might be responsible for the association seen for *RAD1*^{Glu281Gly}.

Firstly, we sought expression of these genes within the colon. We observed expression of RAD1, BRIX1, DNAJC21 and TTC23L, but not AGXT2, within the colon. All five were expressed within the kidney.

Secondly, we sought potential causal variants within *BRIX1, DNAJC21* and *TTC23L* that might be in LD with *RAD1*^{Glu281Gly}, by direct sequencing of their entire ORFs, flanking intronic sequences and 5'UTR in twenty five aCRC patients carrying the risk allele.

We found two variants in *BRIX1* that were not likely to affect function (a synonymous variant Thr35 [rs2069465, MAF in dbSNP = 6.6%] and a variant 281bp upstream of exon 1 [rs2069469, MAF in dbSNP = 6.6%]). In *DNAJC21,* we found two synonymous variants that were not likely to affect function (Pro378 [rs17304200, MAF in dbSNP = 7.1%] and Val482 [rs17244979, MAF in dbSNP = 9.3%], and a private nonsynonymous variant Asn561Ser [rs35999194, MAF in dbSNP = 1.4%]). In *TTC23L*, we found four variants (a synonymous variant Thr137 [rs3906383, MAF in dbSNP = 3.4%], one novel variant 452bp upstream of exon 1, one variant 157bp upstream of exon 1 [rs336484, MAF in dbSNP = 41%], and a nonsynonymous variant, His22Arg [rs6451173, MAF in dbSNP = 43%]).

Since there was a high level of LD between *TTC23L*^{Hiss22Arg} and *RAD1*^{Glu281Gly} in the samples analysed (r^2=1.0, D'=1.0), the entire training phase cohort was genotyped for this variant using KASPar genotyping. No association with CRC risk was observed (X^2=0.24, *P*=0.63).

3.4 Discussion

3.4.1 The training phase cohort

Recently, GWAS has uncovered multiple common variants that have a modest contribution to CRC risk. Our training phase cohort has recently been used in the identification and validation of novel CRC susceptibility alleles (Houlston *et al.* 2010; Dunlop *et al.* 2012a). We sought to further demonstrate the ability of the training phase cohort to uncover predisposition alleles by validating alleles previously discovered by GWAS. We successfully validated 5 of these loci. The failure to validate the remaining 5 loci could be due to a lack of power to detect small effect sizes as a result of sample size in the training phase cohort.

3.4.2 Known biological effects of validated variants

A major problem with the interpretation of GWAS results is that the variants discovered rarely appear to be the true casual variants. Steps have been taken to examine GWAS loci in more depth and uncover the underlying biological mechanisms associated with risk loci. Four of the five loci validated by the training phase cohort here have been investigated further, allowing for the biological bases of disease to be alluded to. The variant rs10795668 at the remaining locus, 10p14, appears to be in a region that has no predicted to be protein coding genes.

3.4.2.1 – 18q21 – rs4939827

In the original GWAS carried out by Broderick and colleagues, three variants at the 18q21 locus were identified as being significantly associated with risk of CRC (rs4939827, rs12953717 and rs4464148). The OR for rs4939827 (OR=0.85) given in this study mirrored those seen in the COIN cohort. Association between rs4939827 and CRC risk was replicated in three independent studies (Tenesa *et al.* 2008; Curtin *et al.* 2009; Slattery *et al.* 2010). This variant maps to a distinct LD block in intron 3 of *SMAD7*. Further investigation of this 17kb LD block uncovered a common (MAF=47%) novel variant located in an enhancer element, shown to reduce expression of SMAD7 by 11%, suggesting it was the true contributor to CRC risk (Pittman *et al*. 2009). SMAD7 is a negative regulator of the TGFβ signalling pathway. The TGFβ pathway is involved in the development, prognosis and progression of both hereditary and sporadic forms of CRC, suggesting that the pathway has a key

role in disease aetiology. Upon activation of the TGFβ receptor, SMAD7 binds to the receptor intracellularly and, together with Smurf1, ubiquinates and breakdowns the receptor complex, halting any downstream signalling transduction (Ebisawa *et al.* 2001; Serra, 2002). The TGFβ pathway controls key biological functions that could implicate it in cancer development and progression, such as inflammation, apoptosis, differentiation and cellular adhesion, therefore suggesting that SMAD7 is the most likely candidate gene for CRC at this locus (Shi and Massagué, 2003).

3.4.2.2 – 15q13 – rs4779584

In addition to *SMAD7*, several other components of the TGFβ pathway have been shown to house genetic variants that are significantly associated with CRC risk. This includes the locus validated here, rs4779584, which is seen in the region just upstream of *GREM1.* In the original study, an OR of 1.35 did not meet formal significance after correction for multiple testing. However, following genotyping in three additional cohorts, a meta-analysis revealed that there was an association between the locus and CRC (OR=1.26), similar to the OR reported here (Jaeger *et al.* 2008). A synthetic association of rs4779584 is assumed due to the fact that it tags two functional variants, rs16969681 and rs11632715, which were subsequently shown to also be significantly over-represented in CRC cases (Tomlinson *et al*. 2011). Interestingly, *GREM1* has recently been shown to be associated with the Mendelian colorectal polyposis syndrome HMPS and the two GWAS variants fall within the region duplicated in this condition (Jaeger *et al*. 2012; Section 1.2.1.4.4). GREM1 operates in the bone morphogenetic protein (BMP) pathway. It acts extracellularly on BMP receptors as an antagonist of the signal transduction molecules, BMP2 and BMP4, and therefore reduces signalling.

3.4.2.3 – 8q24 – rs6983267

We also validated the variant rs6983267 at 8q24 originally identified by Tomlinson *et al.* (2007) to be associated with an elevated risk of CRC and CRA (OR=1.21 and 1.22, respectively). An oncogenic mechanism was suggested for the risk allele (G) when it was shown to be amplified in CRC tumours (Tuupanen *et al*. 2008). Following confirmation that the region has a high level of species conservation and contains potential enhancer elements, researchers proposed that the variant could play a role in gene regulation (Yeager *et al*. 2008). Although in a

relative ‘gene desert’, interestingly the nearest coding gene (>300kb away) to the variant is *MYC*; a proto-oncogene key in the Wnt signalling pathway. It was shown that the rs6983267 directly affects the rate of binding of the Wnt related transcription factor, T cell factor 4 (TCF4). In fact, the presence of the causative G allele leads to a 1.5 fold increase in the degree of Wnt signalling response compared to the T allele (Tuupanen *et al*. 2009).

In support of this was the finding that a physical interaction between the risk region and the first half and promoter region of *MYC* occurs in CRC cell lines (Pomerantz *et al*. 2009). The formation with either allele of a chromosomal loop demonstrated that, despite the large genomic distance between the two regions, an interaction is seen. This was supported by Wright and colleagues, who also showed for the first time that the presence of the G allele conferred an increase in *MYC* expression; approximately 2 fold that of the T allele (Wright *et al*. 2012).

Recently, the expression of MYC by TCF4 was shown to be positively regulated by a non-coding RNA transcript, colon cancer associated transcript 2 (CCAT2). *CCAT2* lies in the region of the rs6983267, with the presence of the G allele increasing the transcription rate. It was shown to be over-expressed in CRC, with expression negatively associated with MSI and associated with an increased rate of metastasis (Ling *et al.* 2013).

3.4.2.4 – 8q23.3 – rs16892766

The variant rs16892766 tags a possible causative gene, eukaryotic translation initiation factor 3, H (*EIF3H*). In order to assess the risk associated with rs16892766, a region of LD was investigated further and a tagging variant, rs16888589, was found to be associated with an increase in expression of EIF3H (Pittman *et al.* 2010). EIF3H has previously been shown to increase growth and survival, with over expression linked to other cancer types (Savinainen *et al.* 2006). Additionally, an *in silico* analysis of the region also indicated that rs16888589, in addition to two other variants, were significantly associated with transcript levels of UTP23, suggesting that this was the true target of the functional effect of the locus association (Carvajal-Carmona *et al*. 2011).

3.4.3 DNA repair genes and cancer

Inherited and acquired deficiencies in DNA repair pathway genes have previously been shown as important contributors in the development of multiple cancer types, including CRC (Section 1. 3). We attempted to assay for common (MAF≥4%) nonsynonymous variants in DNA repair genes from multiple pathways in the training phase cohort. We identified one variant, *RAD1*^{Glu281Gly}, which remained statistically significant after correction for multiple testing in the training phase cohort. Despite initial associations in this cohort we failed to replicate findings in an aCRC setting.

3.4.4 Failure to replicate association observed in the training phase

3.4.4.1 The 'winner's curse'

The theory behind the phenomenon of the 'winner's curse' could explain an elevation of the OR seen in our training phase cohort. Winner's curse describes how the effect size of exploratory studies is elevated, conditional on that study being the first to show such an effect (Zöllner and Pritchard, 2007). It is commonly seen in large scale GWAS due to an inability to correct for the large amount of variants tested in a cohort in one go, resulting in a high false positive rate. Similarly, the winner's curse has also been shown to have a role in inconsistencies between candidate gene studies (Ioannidis *et al.* 2001). Consequently, the initial first positive result seen cannot be given as an accurate representation of the true population effect.

With regards to *RAD1*^{Glu281Gly}*,* based on a MAF in controls of 8% at the OR seen in the training phase cohort, we had 66% power to detect the same effect size in our validation phase cohort of 1,053 cases and 1,397 controls. However, in order to compensate for a potential over-estimation in the initial effect size, we would require the validation cohort to be a lot larger. For example, using a more conservative OR of 1.1, we would require over 11,000 cases and controls in order to achieve 80% power at a 5% significance level (Table 3.10).

OR	Power with current validation cohort (cases n=1,053; controls n=1,397)	Number of both cases and controls required for 80% power
1.27	66%	1,697
1.25	59.70%	1,960
1.2	43.10%	2,971
1.15	27.03%	5,237
1.1	14.80%	11,331

Table 3.10 – Sample numbers required in the validation phase cohort to overcome possible elevation in initial effect size of RAD1^{Glu281Gly} due to the 'winner's curse'. Sample numbers for a given odds ratio (OR) were calculated based on a MAF of 8% in controls at a 5% significance level with 80% power

3.4.4.2 Population stratification

Different subpopulations often display different allele frequencies, normally as a result of different ancestral routes. Differences in allele frequencies that occur because of underlying genetic drift are often referred to as population stratification. Population stratification can cause falsely significant results in case control studies of disease when population homogeneity is incorrectly assumed (Freedman *et al.* 2004). As a UK based drug trial, the COIN and COIN-B trials consisted of mostly patients with a known Caucasian background. Despite this we endeavoured to gauge as much information regarding samples as possible before the analysis, using medical records and other notes to rule out any confounding factors of population stratification on the analysis. We identified and removed 40 samples known to be of a non-Caucasian background.

In retrospect of our failure to replicate findings in the validation phase cohort, Fay Hoskins (ICR, London) performed a principal component analysis (PCA) following genotyping of over 280,000 variants on all patients from COIN and COIN-B. It showed that there were 125 COIN or COIN-B patients from a non-Caucasian background. Of these, 37 had previously been identified by us. In total, 128 samples were deemed to be from a non-Caucasian background. Upon removing these from the analysis, we observed very subtle effects on the association in the training phase cohort for *RAD1*^{Glu281Gly} (X^2=13.57, *P*=2x10^{-4}; MAF cases=6%, controls=8%)

3.4.4.3 Linkage disequilibrium

Failure to replicate initial findings could be due to the identified variant being in LD with another true casual variant, meaning that the variant identified is not responsible for the association observed at a locus; there is indirect association (Hirschhorn *et al.* 2002). In order to assess the likelihood that *RAD1*^{Glu281Gly} is in LD with another true casual variant we sequenced the entire ORF, flanking regions and 5'UTR of *RAD1*, as well as three tagging genes within the identified LD block (*BRIX1, DNAJC21 and TTC23L*). We identified one nonsynonymous variant, *TTC23L*^{His22Arg}, which displayed high LD with *RAD1*^{Glu281Gly} in the patients assayed. However, we failed to observe an association in the training phase cohort. Together, these data suggest that *RAD1*^{Glu281Gly} itself is likely to be responsible for the observed association.

3.4.4.4 Meta-analysis

Following rigorous correction for multiple testing, only the Glu281Gly variant in *RAD1* remained significant in our training phase cohort. However, conducting a meta-analysis allowed us to increase the power to assess variants by increasing the sample size. As well as pooling data from all cohorts in meta-analysis for each variant, we also endeavoured to stratify by CRC stage by analysing population based, early stage CRC cohorts (POPGEN and RMHNHST) and aCRC cohorts (training and validation phase cohorts) in separate meta-analysis.

3.4.4.4.1 RAD1^{Glu281Gly}

We observed, when analysed together under meta-analysis, a positive association between *RAD1*^{Glu281Gly} and aCRC. Similarly, meta-analysis of all cohorts revealed a positive association. However, since no association was observed in the other cohorts assessed, it would appear that both meta-analysis results are primarily driven by the original association from the training phase cohort. Since no association was observed in the meta-analysis of early stage CRC cohorts, this suggests that any association may be specific to aCRC. We do not have enough evidence to support a role for *RAD1*^{Glu281Gly} in aCRC predisposition and it warrants further investigation.

3.4.4.4.2 POLG^{Gln1236His}

We observed, when all cohorts were analysed together in meta-analysis, a positive association between *POLG*^{Gln1236His} and CRC. Similarly, an association was observed when data from aCRC cohorts was pooled. Again, since there was little or no association seen in the other cohorts, we feel that the meta-analysis association is again driven by the association from the training phase cohort.

3.4.4.4.3 REV1^{Val138Met}

We observed, when analysed together under meta-analysis, a positive association between *REV1*^{Val138Met} and aCRC. Additionally, an association was observed when all cohorts were analysed together. Again, since there was little or no association seen in the other cohorts, we feel that both meta-analysis associations are driven by the association from the original training phase cohort. No association

was observed in the meta-analysis of early stage CRC, suggesting the association may be specific to aCRC.

Chapter Four – Identifying genes associated with oxaliplatin-induced peripheral neuropathy in the treatment of aCRC

4.1 Introduction

Oxaliplatin (Eloxatin®) is a third generation platinum compound first approved for the treatment of CRC in the EU in 1996. It is commonly used as part of the chemotherapeutic regimens FOLFOX and XELOX (Section 1.4.2). Before the development of oxaliplatin, a proportion of patients with CRC were considered to have an intrinsic resistance to platinum treatments (Kemeny *et al.* 1990; Loehrer *et al.* 1988; Fink *et al.* 1998; Rixe *et al.* 1996). Despite showing different patterns of cancer specific resistance, the platinum drugs are believed to share a common mechanism of action and metabolism. The correct cellular response and pharmacokinetic profile of oxaliplatin is critical for the adequate action of the drug in the treatment of CRC.

4.1.1 Pharmacokinetics of oxaliplatin

Oxaliplatin consists of a central platinum atom, with a DACH carrier ligand and bidentate oxalate ligand (Kidani *et al*. 1978). Oxaliplatin is administered intravenously at a dose of 85mg/m^2 once every two weeks in the first line treatment or 130mg/m^2 once every three weeks in the second line treatment in combination with fluoropyrimidines over the course of 2-6 hours to achieve sufficient plasma C_{max} (Culy *et al.* 2000).

4.1.1.1 Absorption

Upon initial absorption, the oxaliplatin prodrug is non-enzymatically hydrolysed by displacement of the oxalate group by H_2O and chloride ions. This forms the reactive intermediates monochloro-, dichloro- and diaquo-DACH platinum (Desoize and Madoulet, 2002) which bind to amino groups in DNA, RNA and proteins, as well as biotransformation via irreversibly binding to sulphur groups in cysteine, glutathione and methionine (Luo *et al*. 1999).

After a direct 2 hour infusion with oxaliplatin, over 70% of these metabolites will bind irreversibly to plasma proteins, predominantly albumin and erythrocytes, rendering the drug unavailable (Pendyala and Creaven, 1993; Culy *et al*. 2000).

4.1.1.2 Distribution

The DACH compound of oxaliplatin is highly lipophilic and readily distributes from the plasma throughout the body. The high level of distribution is aided by the ability to readily bind to proteins, macromolecules and DNA (Graham *et al*. 2000).

4.1.1.3 Metabolism

Oxalate is produced as a metabolite of oxaliplatin following non-enzymatic displacement by H_2O or chlorine ions. As a chelator of calcium it is thought that oxalate may have an acute role in neuropathy seen in oxaliplatin treatment (Grolleau *et al*. 2001). The metabolism of oxalate is similar to that of glycoxylate, a by-product of amino acid metabolism. Glycoxylate is detoxified and metabolised by AGXT and glyoxylate reductase-hydroxypyruvate reductase (GRHPR) respectively (Holmes and Assimos, 1998).

4.1.1.4 Elimination and excretion

It is believed that renal elimination is the main course of excretion of unbound oxaliplatin, accounting for around 50% of the free concentration. Renal excretion has been shown to occur at a rate of approximately 121ml/min (Kern *et al*. 1999). The proportion of oxaliplatin bound to erythrocytes (approximately 37%) is eliminated from circulation at a rate that is in accordance with the cells half-life (Levi *et al*. 2000).

4.1.2 Cellular processing of platinum agents

4.1.2.1 Cellular influx

The primary mechanism of uptake is passive diffusion; however, several transporter proteins have been implicated in platinum uptake. The copper transporter protein 1 (CTR1; Song *et al.* 2004; Holzer *et al*. 2006) and both organic cation transporters (OCT1 and OCT2) have been shown to increase cellular accumulation

of oxaliplatin (Zhang *et al*. 2006). Knockout of *Oct2* in mice has been linked to an increased rate of oxaliplatin-induced neurotoxicity (Sprowl *et al.* 2013).

4.1.2.2 Trafficking and localisation

Other members of the copper transport system have previously been recognised as having a role in the control of localisation of cellular platinum compounds (Safaei *et al.* 2004). Copper chaperones bind to and distribute platinum drugs throughout the cell. The human antioxidant homologue 1 (HAH1) shuttles platinum compounds to the copper transporting P-type adenosine triphosphatase 7A and 7B (ATP7A and ATP7B) in the Golgi apparatus. Trafficking of both proteins to the plasma membrane is thought to play a role in efflux from the cell (Katano *et al.* 2002). Alternatively, other copper charperones, namely cytochrome C oxidase (COX17) and copper chaperone for superoxide dismutase (CCS), escort platinum compounds to the mitochondria and cytoplasmic superoxide dismutase (SOD), respectively (Plasencia *et al.* 2006).

4.1.2.3 Detoxification

Detoxification of platinum compounds has profound effects on the amount of active drug free to interact with DNA. Direct biotransformation by complex formation with reducing agents rich in thiol groups, such as L-cysteine, L-methionine and glutathione (forming $Pt(DACH)(Cys)_2$, Pt(DACH)Met and $Pt(DACH)(GSH)_2$, respectively) results in unreactive species (Luo *et al.* 1999; Levi *et al.* 2000). Conjugation results in cellular efflux of the platinum, protecting the DNA from damage (Siddick, 2003). Glutathione conjugation is catalysed by GST, a phase II metabolic enzyme. Although many subclasses of GST exist, only a handful have been implicated in platinum detoxification, in particular, GSTP1, GST-τ (GSTT1) and GST-μ (GSTM1) (Stoehlmacher *et al.* 2002; Medeiros *et al.* 2003).

Platinum detoxification is also carried out by metallothioneins (MT); low molecular weight proteins, consisting of mainly cysteine residues. Intrinsically, MT is thought to be important in controlling the exposure of heavy metals, as well as copper. Cancers exhibiting high levels of MT1A and MT2A have been shown to exhibit a reduced response to platinum treatments (Siegsmund *et al.* 1999; Toyoda *et al.* 2000).

4.1.2.4 Efflux

One of the key mechanisms of cellular efflux of platinum compounds is via the ATP7A and ATP7B copper export proteins. Additionally, there are reports of the role of the ATP-binding cassette, subfamily B (ABCB1), ABCG2, ABCC1, ABCC2, ABCC3 and ABCC5 as platinum efflux proteins. Overexpression of several of these has been associated with outcome of platinum treatment (Liedart *et al*. 2003; Oguri *et al.* 2000; Ceckova *et al*. 2008; Theile *et al.* 2009; Pham *et al*. 2012).

4.1.3 Pharmacodynamics of platinum drugs

The anti-neoplastic properties of all of the platinum compounds are based predominantly on their ability to form platinum-DNA adducts in nuclear DNA (Brabec and Kasparkova, 2005). The formation of cross links stalls DNA synthesis (Raymond *et al.* 1998) impairing both replication and transcription and ultimately triggering apoptosis (Faivre *et al.* 2003; Cepeda *et al.* 2007).

Oxaliplatin and cisplatin appear to have similar sequence and regional localisation of DNA damage (Woynarowski *et al*. 1998). Oxaliplatin is believed to form fewer adducts than cisplatin at equimolar concentrations but, in part due to gross modifications of the DNA helix on account of the bulky DACH group, inhibits DNA synthesis at a greater efficiency (Saris *et al*. 1996).

Initially, monoadducts between the platinum adduct and DNA form. However, these adducts are not considered to be integrally damaging (Zwelling *et al*. 1979). It is only following the formation of biadducts that the cytotoxic effects of platinum treatments are evident. It seems that the predominant lesion, constituting about 60% of those seen, consist of intrastrand crosslinks between two guanine residues. Similarly, intrastrand crosslinks between guanine and adenine contribute to around 30% of the lesions (Eastman, 1987; Woynarowski *et al*. 1998). Other DNA adducts include ICLs (Woynarowski *et al*, 2000). Although rare, DNA-protein cross links are also seen (Zwelling *et al*. 1979).

4.1.4 Apoptosis

Following exposure of cells to platinum treatment, cell cycle arrest and intrinsic signalling cascades indicative of that of apoptosis occurs within the first 24 hours of treatment. In response to regulation by p53, a marked increase in BAX leads to the release of cytochrome C from the mitochondria and activation of apoptotic peptidase activating factor 1 (APAF1). This activates the aspartate specific proteases, the caspases. Upstream effector caspase 9 (CASP9) activates CASP3 and CASP7 (Donzelli *et al.* 2004), leading to apoptosis as a result of cleavage of cellular proteins (Arango *et al*. 2004).

4.1.4.1 Cell checkpoints

The process of cell cycle arrest in G2 is critical for the action of platinum drugs by engaging cell death. Cell division cycle 25 (CDC25C) is phosphorylated by the checkpoint kinase proteins, CHEK1 and CHEK2, as part of the DNA damage sensor signalling pathway involving ATM and ataxia telangiectasia and Rad3 related (ATR). Ultimately, the initiator of G2 stalling is in response to an elevation in cell division cycle 2 (CDC2), following translocation from phosphorylated CDC25C (Wang and Lippard, 2005).

4.1.4.2 Damage recognition and cellular transduction

The formation of a shallow but wide structural distortion of the minor groove allows recognition of intrastrand DNA adducts. Initially, as well as other DNA repair proteins, distortion caused by platinum drug treatments is recognised via the binding of high mobility group (HMG) box protein 1 (Wozniak and Blasiak, 2002) and structure specific recognition protein (SSRP1; Yarnell *et al.* 2001).

The role of HMGB proteins is wide in the response to damage, including stimulating site-directed recombination by cleavage of the recombination activating genes 1 and 2 (RAG1/2; van Gent *et al.* 1997), binding and enhancing structural changes of the nucleosome and directly interacting with components of the MMR pathway to stimulate repair (Yuan *et al*. 2004), as well as shielding areas of damage from other repair processes (Huang *et al*. 1994).

HMGB1 directly interacts and localises p53 (Jayaraman *et al.*1998), a crucial component of apotosis and cell cycle arrest triggered by platinum DNA damage. A role for p53 in DNA repair of platinum damage has also been proposed due to interactions with XPC, TFIIH and RPA in the NER process (Dutta *et al.* 1993; Wang *et al.*1995; McKay *et al.* 1999).

Alternatively, SSRP1 binds to suppressor of Ty 16 homolog (SPT16) forming the facilitates chromatin transcription (FACT) complex. The complex recognises 1,2-intrastrand platinum damage and via its HMG domain, recruits the protein to areas of damage (Yarnell *et al*. 2001).

Signal transduction from the nucleus to the cytosol is a key part of the response of a cell to DNA damage in order to control checkpoint progression or trigger apotosis. C-ABL is a nuclear tyrosine kinase that has been shown to be stimulated by platinum drug DNA damage to regulate apoptosis by interactions via a HMG domain. Prevention of this signalling cascade can be controlled by the tumour suppressor retinoblastoma 1 (RB1), which binds to C-ABL and prevents kinase activity following DNA damage signalling. In addition to p53, another pro-apoptotic downstream target of C-ABL, p73, has been shown to be key in response to platinum treatment in MMR-proficient cells only (Shaul, 2000). Since the proficiency of MMR has no effect on oxaliplatin response, this is thought to be specific to cisplatin adducts (Nehmé *et al*. 1999).

C-ABL is also key in activating other protein kinases in response to platinum damage. Firstly, p38-MAPK, important in controlling gene expression and the chromatin environment, had been shown to be activated in platinum treated cells via the mitogen activated protein kinase kinases, MKK3 and MKK6. Downstream target mitogen and stress activated protein kinase 1 (MSK1) phosphoylates histone H3 in response to platinum damage (Wang and Lippard, 2005). Secondly, extracellular signal regulated kinase (ERK) activation following phosphorylation by the mitogen activated protein kinases, MEK1 and MEK2 in response to platinum treatment could have a role in response. Thirdly, a role of the c-Jun N terminal kinase (JNK) signalling cascade (following MKK4/MKK7 mediated phosphorylation) has also been proposed due to observations that activation leads to an increase in cell death following platinum drug treatment (Pandey *et al.* 1996).

However, there are also survival pathways that are key in platinum drug damage. The AKT-pathway is one such example; a part of the PI3K signalling cascade. AKT is activated by the direct binding of PI3K-generated phospholipids and has several anti-apoptotic actions. Firstly, phosphorylation of X-linked inhibitor of apoptosis (XIAP) stabilises the protein and prevents breakdown following platinum drug DNA damage, ultimately resulting in a decrease in activation of apoptotic pathways (Dan *et al.* 2004). Additionally, AKT also prevents apoptosis in response to platinum damage by phosphorylating and increasing activation of nuclear factor *k*B (NF-*k*B), inhibition of which has been shown to increase efficacy of platinum compounds (Mabuchi *et al.* 2004).

Additionally, increase in survival following platinum treatment has been linked to MAP kinase phosphatase (MKP1) which inhibits both JNK and p38MAPK activation (Wang and Lippard, 2005).

4.1.5 DNA repair of platinum induced damage

4.1.5.1 NER pathway

The NER pathway is important in the repair of bulky adducts that alter the helical formation of DNA and cross linking agents, such as those formed in platinum drug treatment (Section 1.3.3).

4.1.5.2 MMR pathway

In platinum treatment, the formation of adducts leads to strand contortion in DNA which the MMR pathway (Section 1.3.1) plays a role in repairing. However, it is the adduct that is recognised by MMR proteins and, as a by-product of this, shielded and protected from other DNA repair processes. Ultimately this results in the removal of the contorted strand and retention of DNA adducts. This process, known as ‘futile cycling’, was first proposed by Goldmacher in 1986 and helps to explain why MMR deficiency increases resistance to platinum treatments (Goldmacher *et al.* 1986).

It is interesting to note that MMR deficiencies confer resistance to cisplatin and carboplatin but not oxaliplatin (Fink *et al.* 1996). This is particularly important in the treatment of CRC, since approximately 15% have MMR deficiencies. The

reasons for the differences between platinum treatments is believed to be as a result of differences in adduct specificity of the MMR pathway (Martin *et al.* 2008).

4.1.5.3 BER pathway

The BER pathway is involved in the removal of non-helix distorting DNA damage (Section 1.3.2). The type of DNA damage caused by platinum drugs means that BER is not thought to be the main mechanism of repair. Despite this, certain BER proteins have been linked to platinum treatment outcome (Stoehlmacher *et al.* 2001; Lv *et al.* 2013).

4.1.5.4 ICL repair

Approximately 5% of the lesions seen in platinum treatment consist of ICL as a result of platinum adducts binding to bases in opposing strands. It is role of the FA pathway to repair these lesions (Section 1.3.5).

4.1.5.5 Replicative bypass

The ability of certain polymerases to skip platinum DNA damage during replication means that there is an opportunity for platinum adducts to accumulate and potential tolerance to develop. Polymerases that have been previously implicated in platinum treatment or could play a role include REV3L, POLB, POLH and POLM (Rabik and Dolan, 2007).

4.1.6 Side effects of oxaliplatin treatment – peripheral neuropathy

As the main dose limiting side effect of oxaliplatin treatment, peripheral neuropathy is a major problem in treatment (O' Dwyer *et al.* 2000). It is more often severe peripheral neuropathy that results in the removal from treatment than disease progression. Additionally, peripheral neuropathy associated with oxaliplatin (PNAO) is not correlated with response to treatment and is therefore considered an avoidable malady (Whinney *et al*. 2009). There are no current treatments to alleviate the symptoms associated with PNAO (Wolf *et al.* 2008). Two clinically distinct forms of neuropathy have been reported and are believed to arise through different pathophysiological mechanisms. An acute form is due to disruption of voltage gated sodium channels indirectly as an extension of chelation of calcium ions by the oxaliplatin metabolite, oxalate (Grolleau *et al*. 2001). The chronic form is due to

direct toxicity of nerve cells via the accumulation of platinum adducts in the dorsal root ganglia (Ta *et al.* 2006).

There is little knowledge surrounding possible risk factors or genetic predisposition to PNAO. Previously, a putative association between chronic PNAO and a coding variant in *GSTP1*, resulting in an isoleucine to valine substitution at codon 105 of the protein, has been described (Grothey *et al.* 2005; Ruzzo *et al.* 2007; Peng *et al.* 2013), although the risk allele is of debate (Lecomte *et al.* 2006; Gamelin *et al.* 2007; Inada *et al.* 2010). Also, particular haplotypes of *AGXT* have been shown to predispose towards both acute and chronic forms of PNAO (Gamelin *et al.* 2007). Additionally, a silent polymorphism which falls within an aspartic acid residue at position 118 of the NER gene *ERCC1* has been associated with an increased rate of onset of chronic PNAO in a Japanese population (Inada *et al.* 2010; Oguri *et al.* 2013). Mutations in genes involved in neuronal function have also been suggested to predispose to PNAO. A nonsynonymous variant in *SCN10A* (Leu1092Pro [rs12632942]) has been shown, under an overdominant model, to increase the chance of acute PNAO (Argyriou *et al.* 2013).

Here, we sought to identify the underlying genetic causes of PNAO in patients exhibiting the most severe phenotypes using exome resequencing. In order to assess the sequencing data, we applied two analysis strategies;

1. Analysis of variants in genes involved in the pharmacokinetics and cellular response to oxaliplatin
2. Analysis of novel variants in genes involved in neuronal function and/or peripheral neuropathy

4.2 Materials and methods

4.2.1 Patient selection

Patients were selected from 2,445 individuals undergoing treatment with 5-fluorouracil or capecitabine, oxaliplatin and potentially cetuximab as part of the COIN trial. PNAO with a grade 3 or greater was observed in 23% of patients with 5-fluorouracil based regimens and 16% of those with capecitabine based regimens over the entire trial period (Maughan *et al*. 2011). Assessment of PNAO was carried out every 6 weeks following the initiation of treatment. The recording of PNAO grade

was carried out by a consultant and clinical nurse using the Common Terminology Criteria for Adverse Effects v3.0 (CTCAE; National Cancer Institute common toxicity criteria for adverse events, Accessed June 19, 2013; Table 4.1). Additionally, patients who reported at least grade 3 neuropathy carried out a Quality of life Questionnaire (QLQ C30) which supported evidence of severe PNAO.

4.2.2 Oxaliplatin administration as part of the COIN trial

With capecitabine, oxaliplatin was given intravenously at 130mg/m^2 over a period of 2 hours at 3 weekly intervals. Capecitabine was given orally twice a day for the three weeks prior to oxaliplatin administration. Initially it was given at 1000mg/m^2 but was reduced to 850mg/m^2 following evidence that there was elevated toxic effects in patients from Arm B of the trial (Adams *et al.* 2009).

With 5-fluorouracil and folinic acid, oxaliplatin was given intravenously at 85mg/m^2 over a period of 2 hours at 2 weekly intervals. This was followed by a bolus injection of 400mg/m^2 of 5-fluorouracil, with a 46 hour infusion of 2400mg/m^2 of the drug. Either 175mg of L-folinic acid or 350mg of D,L-folinic acid was given intravenously over a 2 hour period concurrent to oxaliplatin treatment (Maughan *et al*. 2011).

4.2.3 Exclusion of known neuropathies

Exclusion of known neuropathies in the ten patients sent for exome resequencing was carried out by multiplex ligation-dependant probe amplification (MLPA) at Bristol Genetics laboratory. Samples were analysed with the SALSA® MLPA® kit using the P033-B2 probe mix (Appendix 6; MRC Holland, Amsterdam), following the manufacturer's instructions. Sample were analysed on a Beckman Coulter CEQ 8000 capillary analyser and with the GeneMarker software package. Additionally, exome resequencing data of all genes associated with known neuropathies was examined in all ten patients.

4.2.4 *MUTYH* analysis

Patient 1 was shown previously to carry potentially biallelic mutations in *MUTYH* (Gly396Asp and Arg426Leu). Cloning was carried out by Christopher Smith

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
Peripheral sensory neuropathy	Loss of deep tendon reflexes; Paresthesia that does not affect function	Sensory alteration or parasthesia; interfering with function but not with ADL	Sensory alteration or paresthesia interfering with ADL	Disabling	Death

Table 4.1 – Grading criteria for symptoms of PNAO in accordance with CTCAE v3.0 (ADL – activities of daily living)

to determine if these variants were on the same or opposite strands. Exon 13 was amplified by PCR (Forward primer - 5'-AGGGCAGTGGCATGAGTAAC-3'; Reverse primer – 5'-GGCTATTCCGCTGCTCACTT-3'; Section 2.5.4) followed by verification by agarose gel electrophoresis and PCR purification (Sections 2.5.5 and 2.5.6). Ligation into the pGEM-T easy vector, transformation and plasmid extraction were carried out (Sections 2.5.12.5 - 2.5.12.7). Following PCR and clean up, amplification products were sequenced and cleaned up (Sections 2.5.7 and 2.5.9). Sequences were viewed with Sequencher v4.2.

4.2.5 The platinum pharmacokinetic and cellular response pathway

In order to analyse the exome resequencing data, a pathway approach was initially taken. We concentrated on genes involved in platinum drug pharmacokinetics and cellular response (Sections 4.1.1-4.1.5). Genes were found via literature reviews of platinum pathways and exome resequencing data was filtered accordingly. In total, we identified 104 genes that may play a role including four genes involved in drug influx (*OCT1, OCT2, CTR1* and *hMATE1*), three genes involved in trafficking (*CCS, COX17* and *SOD1*), seven genes involved in detoxification (*MT1A, MT2A, NQO1, GSTT1, GSTP1, GSTM1* and *MPO*), two genes involved in oxalate metabolism (*AGXT* and *GRHPR*), three genes involved in sequestration (*ATP7A, ATP7B* and *HAH1*), thirty two genes involved in DNA damage response and subsequent signalling pathways (*SPT16, SSRP1, HMGB1, RAG1, RAG2, ABL1, RB1, p53, p73, AURKA, CCNG2, p38MAPK, MSK1, MKK3, MKK6, Histone H3, ERK, MEK1, MEK2, JNK, MKK4, MKK7, MPK1, AKT, NF-kB, XIAP, Bax, APAF1, CYC, CASP3,6* and *9*), forty six genes involved in DNA damage repair and the associated response pathways (*POLB, POLH, POLM, REV3L, FANCA, FANCB, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCL, FANCM, FANCN, FAAP100, RM1, FAN1, MLH1, MSH2, MSH6, PMS2, ATM, ATR, CHEK1, CHEK2, BRCA1, BRCA2, GADD45, DDB2, CDC25C, CDC2, CSA, HR23B, RNApolII, RPA1, ERCC1-6, XPA, XRCC1, XRCC3* and *MGMT*), and seven genes involved in drug efflux (*ABCC1-5, ABCB1* and *ABCG2*; Fig.4.1).

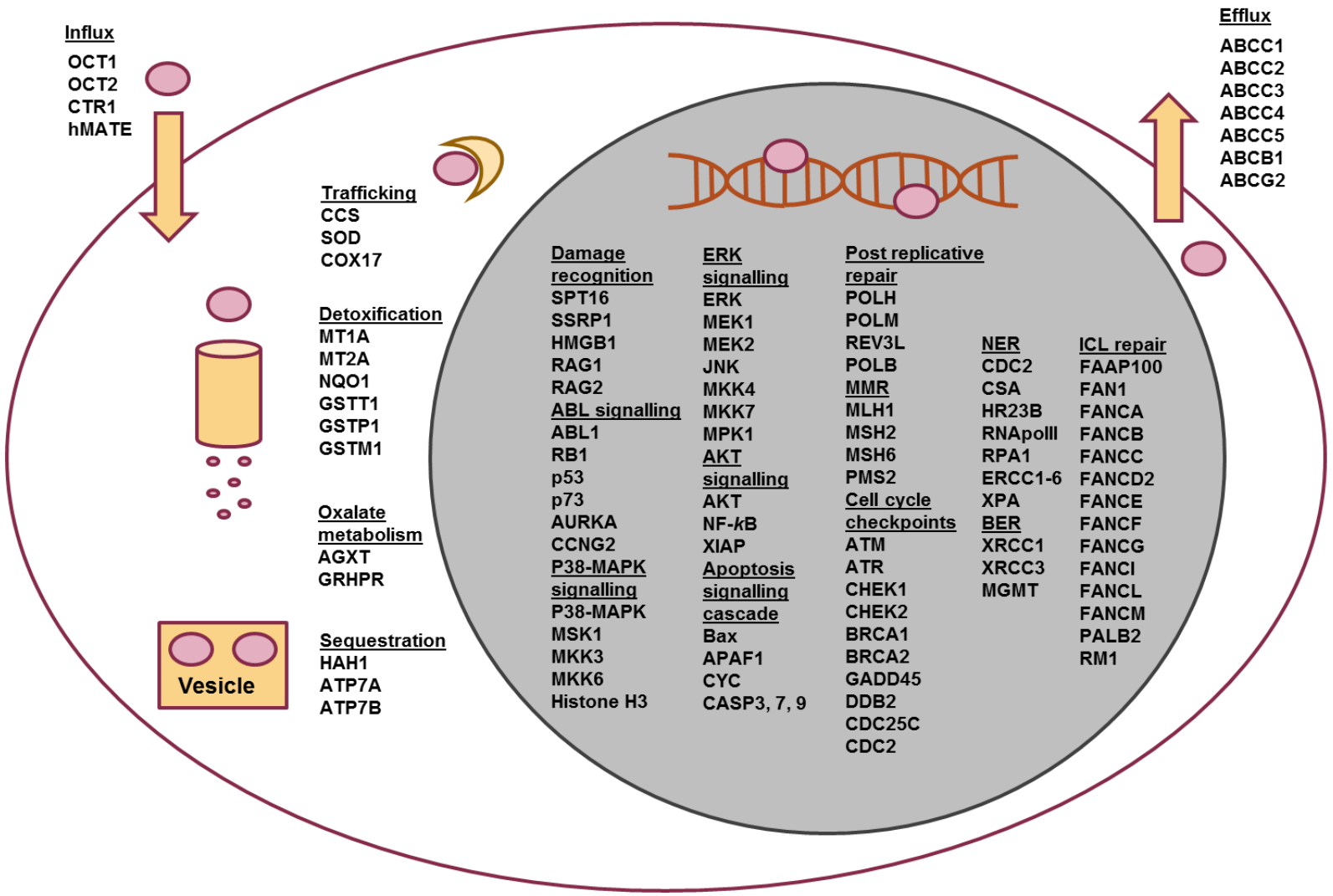


Figure 4.1 – Proteins implicated in the cellular pharmacokinetics and response pathways to platinum drugs.

4.2.6 Exome resequencing

Exome resequencing, read alignment and variant calling was carried out by James Colley. Library fragments containing exomic DNA from our 10 patients with PNAO were collected using the Roche Nimblegen SeqCap EZ Exome Library v2.0 solution-based method. Massively parallel sequencing was performed on the Illumina Genome Analyser at the University of North Carolina. Fastq files were processed through a sequence analysis pipeline using BWA (Li and Durbin, 2009) for sequence alignment and modules from the Broad Institute's Genome analysis Toolkit (GATK) (McKenna *et al.* 2010) to recalibrate quality scores, refine alignments around potential insertion or deletions (indels), eliminate duplicate reads, call indel and SNP genotypes, generate QC metrics, and apply quality filters to the genotype calls. SNP calls were annotated using the Analysis package ANNOVAR (Wang *et al*. 2010).

4.2.7 Analysis of genes involved in neuronal function or peripheral neuropathy

Literature reviews of gene of interest were carried out by searching for a role with 'neurons' or 'peripheral neuropathy' via NCBI and other internet search engines.

4.2.8 PCR and Sanger sequencing

All variants of interest from exome resequencing were validated by Sanger sequencing of an independent PCR product. PCR of specific regions, verification by agarose gel electrophoresis, product purification, Sanger sequencing and sequencing clean up were carried out as described in sections 2.5.4 to 2.5.8. Sequences were analysed using Sequencer v4.6. Primers are given in Appendix 7.

4.3 Results

4.3.1 Patient selection

Nine patients from COIN were identified as having severe PNAO that required removal from treatment within the first 7 weeks. A tenth patient was identified as having PNAO whilst receiving therapy from Professor Timothy Maughan, stopping treatment at the end of the first cycle on account of severe toxicity. This patient was recruited into this COIN Trial Management approved translational project.

4.3.2 *MUTYH* analysis

Cloning and Sanger sequencing of exon 13 of *MUTYH* revealed that Patient 1 was compound heterozygous for the variants Gly396Asp and Arg426Leu. The patient had 'multiple colorectal polyps'.

4.3.3 Exclusion of known hereditary neuropathies

We carried out MLPA of *PMP22* on all ten patients but did not find gene dosage abnormalities. We also examined exome resequencing data in the ten patients with PNAO for mutations in *PMP22* and the other genes associated with rare inherited neuropathies such as *MPZ/PO, SIMPLE/LITAF, EGR2, NEFL, GJB1/CX32, PRPS1, DNM2, YARS, MFN2, RAB7, GARS, HSPB1 (HSP27), HSPB8 (HSP22), GDAP1, LMNA, MED25, MTMR2, SBF2/MTMR13, KIAA1985 (SH3TC2), NDRG1, PRX, FGD4, FIG4, BSCL2, DCTN1, SPTLC1* and *IGHMBP2*. At 20-fold coverage, on average, we covered >50% of the ORF of 85% of these genes. Additionally, 38% of genes had, on average, greater than 90% of the ORF covered. However, 15% of the genes had less than 5% of the ORF covered on average (Table 4.2).

We failed to find any stop-gain mutations or truncating indels in these genes in our 10 patients with PNAO. Although we did find various nonsynonymous variants in *IGHMBP2*, these variants were also found in dbSNP at a similar or greater frequency (Thr879Lys [rs17612126; MAF=15% in patients with PNAO compared to 30% in Caucasians in dbSNP]; Ile275Val [rs10896380; 25% compared to 30%]; Arg694Trp [rs2236654; 25% compared to 30%]; Leu201Ser [rs560096; 15% compared to 11%]; Thr671Ala [rs622082; 40% compared to 30%]) and were therefore considered to be benign polymorphisms. Therefore, we excluded all known genes associated with inherited neuropathies as the likely cause of PNAO.

4.3.4 Exclusion of other known causes of PNAO

4.3.4.1 GSTP1

We examined exome resequencing data for a nonsynonymous variant in *GSTP1* that had previously been associated with PNAO (Grothey *et al.* 2005; Ruzzo *et al.* 2007; Peng *et al.* 2013).The variant consisted of an isoleucine to valine

	Patient ID										
Gene	1	2	3	4	5	6	7	8	9	10	Average
BSCL2	70	79	85	78	74	77	86	74	75	77	77
DCTN1	47	68	77	80	57	69	65	61	72	73	67
DNM2	66	80	81	81	76	81	81	81	83	81	79
EGR2	50	65	82	78	66	71	68	69	71	72	69
FGD4	93	92	96	97	95	98	99	95	97	96	96
FIG4	82	76	96	95	79	85	98	84	85	97	88
GARS	0	0	0	0	0	0	0	0	0	0	0
GDAP1	92	78	100	100	92	98	100	98	100	100	96
GJB1	87	86	88	94	82	91	82	90	97	91	89
HSPB1	0	0	9	0	0	0	4	0	0	3	2
HSPB8	41	18	68	83	68	47	63	54	52	62	55
IGHMBP2	59	82	92	93	71	88	89	77	90	85	83
LITAF	100	96	100	100	100	100	100	100	100	100	100
LMNA	35	73	76	79	56	77	65	58	71	73	66
MED25	27	44	47	50	33	48	52	37	44	44	43
MFN2	91	100	100	100	98	99	100	99	98	100	99
MPZ	69	87	84	77	68	93	83	77	86	76	80
MTMR2	84	82	96	96	88	93	96	96	92	96	92
NDRG1	80	98	97	92	79	97	92	85	96	88	90
NEFL	0	0	0	0	0	0	0	0	0	0	0
PRPS1	92	75	100	100	85	84	91	95	93	100	92
PRX	3	4	5	5	5	5	5	5	5	5	5
SBF2	92	84	98	98	89	97	98	94	89	98	94
SH3TC2	64	82	88	94	66	90	90	69	88	88	82
SPTLC1	96	92	96	96	91	96	96	95	96	96	95
YARS	88	93	100	100	91	100	100	98	99	100	97

Table 4.2 – Percentage of the ORF covered (at 20-fold coverage) of genes previously implicated in hereditary neuropathies. Shades from red through to green represents no to complete coverage, respectively.

substitution at position 105 (rs1695). We found two patients homozygous for the variant (MAF=20% in patients with PNAO compared to 38% in dbSNP).

4.3.4.2 AGXT haplotype

We examined exome resequencing data for a particular haplotype in *AGXT* that consisted of two nonsynonymous variants; Pro11Leu and Ile340Met (rs34116584 and rs4426527, respectively). This particular haplotype, in either the heterozygous or homozygous form, has previously been associated with PNAO (Gamelin *et al.* 2007). We found three patients heterozygous for both variants and another patient heterozygous for just Pro11Leu (Pro11Leu [MAF=20% in both patients with PNAO and dbSNP] and Ile340Met [MAF=15% in both patients with PNAO and dbSNP]).

4.3.4.3 ERCC1

We examined exome resequencing data for a synonymous variant in *ERCC1* that had previously been associated with rate of onset of PNAO (Asp118 [rs11615] Inada *et al.* 2010; Oguri *et al.* 2013). We found five patients heterozygous for the variant (MAF=25% in patients with PNAO compared with 35% in dbSNP).

4.3.4.4 SCN10A

We examined exome resequencing data for a nonsynonymous variant (Leu1092Pro [rs12632942]) that had previously been associated with risk of PNAO (Argyriou *et al.* 2013). We found four patients heterozygous and one patient homozygous for the variant (MAF=30% in patients with PNAO compared with 24% in dbSNP).

4.3.5 Exome resequencing results

On average, across the entire coding genome, we had 54.7% (range 45.7-59.9%) coverage of the ORF at 20-fold coverage. We identified on average 48.9 (range 40-57) stop gains and 87.7 indels predicted to result in frameshift mutations (range 73-111) per patient exome. Variants not present in dbSNP v132 (deemed ‘novel’) were considered to be the most likely to cause PNAO and warranted further investigation. We identified on average 8 (range 2-11) and 28.2 (range 16-57) novel stop gains and frame shifting indels, respectively, per patient (Table 4.3).

	Patient	1	2	3	4	5	6	7	8	9	10
Stop Gains	Total	43	51	46	40	56	52	51	48	45	57
	Novel	2	10	6	7	10	10	11	8	6	10
	Oxaliplatin pathway	1	1	1	2	1	1	1	2	1	2
	Novel and in the oxaliplatin pathway	0	0	0	0	0	0	0	1	0	0
Indels	Total	73	111	80	86	85	99	91	77	93	82
	Novel	16	57	21	20	16	41	28	18	39	26
	Oxaliplatin pathway	2	1	1	1	1	0	3	0	0	0
	Novel and in the oxaliplatin pathway	0	0	0	0	0	0	0	0	0	0

Table 4.3 – Number of stop gain and frameshifting indels identified from exome resequencing in each patient analysed. Variants were filtered based on novelty status, as well as for variants in genes involved in the platinum pathway (Table 4.4).

	1	2	3	4	5	6	7	8	9	10
Stop Gains	*MKK3* - Gly102X (rs55796947)	*MKK3* - Gly102X (rs55796947)	*MKK3* - Gly102X (rs55796947)	*MKK3* - Gly102X (rs55796947); ***BRCA2* - Lys3326X (rs11571833)**	*MKK3* - Gly102X (rs55796947)	*MKK3* - Gly102X (rs55796947)	*MKK3* - Gly102X (rs55796947)	*MKK3* - Gly102X (rs55796947); ***ERCC4* - Ser613X**	*MKK3* - Gly102X (rs55796947)	*MKK3* - Gly102X (rs55796947); ***BRCA2* - Lys3326X (rs11571833)**
Indels	*CASP9* - Val448fs (rs2234723); *OCT1* - Pro425fs (rs113569197)	*OCT1* - Pro425fs (rs113569197)	*OCT1* - Pro425fs (rs113569197)	*CASP9* - Val448fs (rs2234723)	*POLM* - Arg108fs (rs28382645)		*CASP9* - Val448fs (rs2234723); *OCT1* - Pro425fs (rs113569197); *POLM* - Arg108fs (rs28382645)			

Table 4.4 – Stop gain and frameshifting indels in genes in the platinum pathway. Novel variants are highlighted in red. Variants validated by Sanger sequencing of an independent PCR product are in bold font.

4.3.6 Analysis strategy 1 – Analysis of genes in the platinum pathway

We analysed exome resequencing data for the 104 genes identified as important in the platinum pharmacokinetic and cellular response pathways. On average, we covered 74% of the ORF in all of the genes of interest at 20-fold coverage. In addition, over 74% of the genes in the pathway had at least 70% of their ORF covered at this depth, with 32% of genes with at least 90% of the ORF covered. However, 6% of the genes were not covered (Table 4.5).

4.3.6.1 Stop gain mutations

We identified Gly102X in *MKK3* at the same frequency to that reported in dbSNP (rs55796947, MAF=50%) and was therefore considered likely to be a common benign polymorphism.

A stop-gain in *BRCA2* (Lys3326X, rs11571833; MAF in dbSNP=0.1%) was found in two patients and was verified by Sanger sequencing of an independent PCR products.

We identified a single patient (Patient 8) with a novel stop gain Ser613X in exon 9 of *ERCC4* which was verified by Sanger sequencing of an independent PCR product.

4.3.6.2 Frameshifting indels

We identified one frameshifting deletion (Pro425fs; rs113569197) in *OCT1* in four samples. However, the variant was not confirmed in any of the samples upon Sanger sequencing of independent PCR products, suggesting that it was an artefact.

We also discovered Val448fs in *CASP9* in multiple patients with a frequency similar to that reported in dbSNP (rs2234723, MAF in dbSNP=19.60%) and was therefore considered likely to be a common benign polymorphism. Additionally, Arg108fs in *POLM* was observed in two patients and in dbSNP (rs28382645, MAF in dbSNP=2.3%) but was in a transcript associated with nonsense mediated decay so was not considered functional.

		Patient ID										
Pathway	Gene	1	2	3	4	5	6	7	8	9	10	Average
Uptake	*CTR1*	97	59	100	100	75	67	99	98	63	100	86
	SLC22A1	70	77	87	92	74	84	86	80	90	87	83
	SLC22A2	70	67	91	91	77	82	84	79	72	83	80
	MATE1	79	85	92	92	79	89	92	87	90	92	88
Trafficking	*CCS*	58	78	78	78	62	78	86	78	76	77	75
	COX17	44	44	44	44	44	44	44	44	44	44	44
	SOD1	96	81	100	100	99	93	100	100	91	100	96
Efflux	*ABCC1*	0	0	0	0	0	0	0	0	0	0	0
	ABCC2	94	86	100	100	96	95	100	100	93	100	96
	ABCC3	74	82	95	95	78	87	91	83	87	89	86
	ABCC4	96	91	99	100	97	99	100	98	99	99	98
	ABCC5	13	14	14	14	14	14	14	14	14	14	14
	ABCB1	97	87	100	100	97	96	100	97	93	100	97
	ABCG2	95	81	99	99	94	99	100	93	96	99	96
Sequestration	*ATP7A*	97	88	99	100	94	98	99	99	100	100	97
	ATP7B	0	0	0	0	0	0	0	0	0	0	0
	HAH1	0	0	0	0	0	0	0	0	0	0	0
Detoxification	*MT1A*	0	15	73	26	15	15	82	18	15	62	32
	MT2A	0	0	0	0	0	0	0	0	0	0	0
	NQO1	75	52	100	100	64	91	100	71	86	100	84
	GSTT1	0	74	94	83	61	75	0	81	92	100	66
	GSTP1	0	0	0	0	0	0	0	0	0	0	0
	GSTM1	74	0	0	0	58	0	92	28	0	0	25
	MPO	62	81	84	88	72	82	84	69	79	80	78
Oxalate metabolism	*AGXT*	37	76	77	83	63	79	72	65	86	77	71
	GRHPR	83	93	100	99	85	100	97	92	100	100	95
Damage recognition	*SPT16*	85	81	92	91	84	89	96	89	87	93	89
	SSRP1	89	86	99	96	93	91	96	93	91	96	93
	HMGB1	21	4	50	51	25	25	64	25	27	65	36
	RAG1	65	80	94	91	74	92	92	78	88	93	85
	RAG2	84	89	96	96	86	96	97	88	97	99	93
ABL signalling	*ABL1*	42	50	68	67	46	51	64	49	56	61	55
	RB1	63	52	78	81	69	71	89	80	68	93	74
	p53	40	60	85	73	53	75	76	62	63	80	67
	p73	43	87	73	81	59	98	62	63	98	66	73
	AURKA	92	100	98	98	93	100	98	97	99	99	97
	CCNG2	82	83	83	88	83	90	98	86	86	99	88
p38-MAPK signalling	*p38MAPK*	98	82	100	100	99	100	100	99	99	100	98
	MSK1	85	70	96	93	86	89	96	92	85	96	89
	MKK3	63	76	90	95	77	77	85	91	89	85	83
	MKK6	91	84	100	98	83	93	100	95	87	100	93
	Histone H3	31	24	31	31	29	31	31	31	31	31	30
ERK signalling	*ERK*	89	89	89	89	89	89	89	89	89	89	89
	MEK1	88	87	92	94	86	93	92	87	92	92	90
	MEK2	31	60	66	77	40	68	55	48	72	53	57
JNK signalling	*JNK*	91	76	95	94	90	98	100	95	88	99	93
	MKK4	81	65	90	90	82	75	90	82	75	90	82
	MKK7	0	0	0	0	0	0	0	0	0	0	0
	MPK1	43	57	63	66	49	64	54	44	61	52	55
AKT signalling	*AKT*	75	94	99	100	91	98	98	92	99	100	95
	NF-KB	93	79	98	100	93	95	100	96	92	100	95
	XIAP	73	52	86	97	39	62	86	82	89	94	76

		Patient ID										
Pathway	Gene	1	2	3	4	5	6	7	8	9	10	Average
Apoptosis signalling cascade	*Bax*	48	80	95	89	60	82	80	68	77	70	75
	APAF1	89	79	97	99	89	96	100	98	94	100	94
	CYC	0	0	16	0	0	0	54	0	0	27	10
	CASP3	67	94	91	95	69	100	100	91	99	100	91
	CASP7	100	87	100	100	93	100	100	95	99	100	97
	CASP9	68	86	89	89	83	89	89	85	87	89	86
Post replicative repair	*POLH*	55	57	79	80	63	65	81	59	59	79	68
	POLM	32	69	71	73	44	79	56	53	74	63	61
	REV3L	73	80	90	95	87	95	96	90	93	97	90
	POLB	87	91	99	100	93	98	100	96	93	100	96
MMR	*MLH1*	72	81	90	84	73	80	88	75	80	91	82
	MSH2	68	83	94	99	87	93	100	94	92	99	91
	MSH6	67	70	83	87	69	88	91	74	88	88	81
	PMS2	49	50	75	72	50	56	78	61	57	76	62
Cell cycle checkpoints	*ATM*	86	76	96	98	89	95	100	96	88	99	92
	ATR	79	82	91	91	76	93	98	84	89	96	88
	CHEK1	79	69	100	93	80	85	100	82	77	100	86
	CHEK2	51	56	75	71	50	63	83	63	58	73	64
	BRCA1	91	87	98	97	91	96	98	95	94	98	95
	BRCA2	51	44	77	90	64	81	94	76	71	92	74
	GADD45	39	48	52	48	40	66	47	43	62	69	51
	DDB2	70	84	100	99	76	86	96	80	82	99	87
	CDC25C	66	78	86	93	71	84	97	79	79	93	83
	CDC2	71	36	81	81	71	81	81	73	81	81	74
NER	*CSA*	90	90	100	100	85	100	100	99	90	100	95
	HR23B	69	49	87	92	70	68	88	80	62	89	75
	RNApolII	90	86	99	99	82	96	99	90	90	99	93
	RPA	93	93	94	93	92	93	93	93	92	93	93
	ERCC1	76	92	100	100	95	93	94	100	94	100	94
	ERCC2	63	68	77	78	63	70	73	71	75	73	71
	ERCC3	79	88	100	99	90	99	90	91	99	96	93
	ERCC4	55	59	89	87	63	66	93	67	73	96	75
	ERCC5	72	70	95	90	72	81	97	82	75	95	83
	ERCC6	81	77	94	98	86	92	96	89	89	97	90
	XPA	66	66	79	79	67	68	79	78	72	79	73
BER	*XRCC1*	63	90	99	96	78	87	98	85	93	95	89
	XRCC3	39	65	61	53	39	63	52	41	67	46	53
	MGMT	34	76	85	89	72	89	80	76	71	77	75
ICL repair	*FAAP100*	29	49	54	54	34	53	51	43	55	50	47
	FAN1	90	80	100	100	86	90	100	91	84	100	92
	FANCA	76	79	96	92	76	81	90	82	79	90	84
	FANCB	72	19	89	99	53	59	90	92	88	99	76
	FANCC	81	88	98	100	93	99	100	99	92	100	95
	FANCD2	76	76	93	88	76	86	95	82	78	94	84
	FANCE	38	56	69	73	57	66	66	56	63	66	61
	FANCF	15	45	43	40	17	51	38	33	52	48	38
	FANCG	46	66	89	95	50	74	87	53	72	77	71
	FANCI	86	80	92	91	87	87	95	88	84	95	89
	FANCL	81	75	97	99	77	100	100	95	96	99	92
	FANCM	71	57	88	96	78	85	97	87	81	97	84
	PALB2	85	91	94	95	84	96	96	91	93	95	92
	RM1	66	78	96	93	76	96	97	78	93	99	87

Table 4.5 – Percentage of the ORF covered (at 20-fold coverage) of genes implicated in the platinum pharmacokinetic and cellular response pathways. A common key is given in table 4.2.

4.3.7 Analysis strategy 2 – Analysis of genes involved in neuronal function and/or peripheral neuropathy

4.3.7.1 Stop gain mutations

We considered whether stop gains in genes involved in nerve function and/or peripheral neuropathy might also be responsible for PNAO. Therefore, every gene predicted to carry a novel stop gain variant (n=52) from the whole exome analyses was assessed in the literature for a potential role in neuronal function. Literature searches were carried out as described in section 4.2.7.

We identified 2 genes as potentially relevant; stomatin like 3 (*STOML3*) and annexin (*ANXA7*) A stop gain variant in *STOML3* (Arg164X), identified in a single patient and absent from dbSNP, was confirmed in an independent PCR product. However the variant Tyr54X in *ANXA7* was not confirmed upon sequencing of an independent PCR product from the relevant patients genomic DNA and was therefore excluded.

4.3.7.2 Frameshifting indels

We identified 204 novel frameshifting indels from the exome analysis of ten patients with PNAO and every gene was assessed in the literature for a potential role in neuronal function. We identified 5 genes that potentially had a role in peripheral neuropathy; adapter protein containing PH domain (*APPL1* ; Phe472fs), neurofilament, medium (*NEFM*; Tyr63fs), neuropilin 2 (*NRP2*; Ser904fs and Cys907fs), semaphorin-4C (*SEMA4C*; Gly648fs) and protein phosphatase 1 (*PPP1R13L*; Pro562fs). We attempted to validate by carrying out Sanger sequencing of independent PCR products from the relevant patients; only the deletion in *NRP2* was present in two samples. The rest were considered artefacts.

The variant identified in *NRP2* consisted of a CGCA deletion resulting in a frameshift (Ser904fs), as well as an insertion of a single adenine (Cys907fs). One patient was homozygous; another was heterozygous for both variants. Upon sequencing, both were validated in the relevant samples.

4.4 Discussion

4.4.1 Identification of MAP in Patient 1

By cloning and sequencing of exon 13 of *MUTYH* of Patient 1, we identified that the patient was compound heterozygous for the variants Gly396Asp and Arg426Leu. Biallelic mutations of this kind have previously been associated with the inherited CRC condition MAP (Section 1.2.1.2). The patient had 'multiple colorectal polyps', consistent with MAP. There has been no association between peripheral neuropathy and MAP previously made.

4.4.2 Exclusion of hereditary neuropathies

We first attempted to rule out inherited forms of peripheral neuropathy. Charcot-marie tooth syndrome (CMT; also known as hereditary motor and sensory neuropathy [HMSN]) comprises both a clinically and genetically heterogeneous group of disorders. Individual's exhibit distal sensory loss, weakness and wasting of the muscles (Reilly *et al*. 2011). As the most common form of inherited neuropathy, it has an overall population prevalence of 1 in 2,500. Over sixty genes encoding proteins with different cellular functions and localisation have been associated with the disease, accounting for 50% of all cases (Rossor *et al.* 2013). Approximately 75% of patients with CMT1 have a 1.4mb duplication in peripheral myelin protein 22 *(PMP22).* No dosage abnormalities were found following MLPA analysis of *PMP22* in the ten patients with PNAO. Five nonsynonymous variants in *IGHMBP2* (previously associated with hereditary neuropathies), were identified following analysis of exome resequencing data. However, all were seen at similar frequencies to that reported in dbSNP. Therefore, we ruled out all inherited forms neuropathy in our ten patients.

4.4.3 Exclusion of known causes of PNAO

Secondly, we investigated coding variants previously associated with PNAO. These included *GSTP1*^{Ile105Val}*, AGXT*^{Pro11Leu} and *AGXT*^{Ile340Met}*, ERCC1*^{Asp118} and *SCN10A*^{Leu1092Pro}*.* All variants were observed in the ten patients with PNAO at a lesser or similar frequency to that documented in dbSNP suggesting that they did not contribute to PNAO.

4.4.4 Exome resequencing

NGS has allowed researchers to adequately assess large regions of the genome to help identify the underlying causes for disease phenotypes (Section 1.7). Exome resequencing allows researchers to target the protein coding regions of the genome. However, only the regions captured by the exome targeted platform will be sequenced. Here we used the Roche Nimblegen SeqCap EZ Exome Library v2.0 solution-based method for exome enrichment. This capture kit targets 89.8% of the exome annotated in CCDS (Parla *et al.* 2011). We analysed coverage of the ORF of genes involved in hereditary neuropathies and the platinum pathway at 20-fold coverage. We observed that over 90% of the ORFs were covered on average for 38% and 32% of these genes, respectively. Furthermore, two of the genes involved in hereditary neuropathies (*GARS* and *NEFL*) and six of the genes involved in the platinum pathway (*ATP7B, HAH1, MT2A, GSTP1, MKK7* and *ABCC1*) had no coverage in the ten patients. We speculated that this could be as a result of these genomic regions not being well represented by the probes. Analysis of the annotation files of the genomic regions covered by the platform (available online at http://www.nimblegen.com/products/seqcap/ez/v2/index.html#annotation) revealed that all of these genes had probes to cover 100% of their ORFs. We suggest that this could be due to a lack of specificity of probes in some of the earlier capture kits. This could ultimately result in false negative results. Later capture kits have taken steps to overcome this, such as improving probe specificity and increasing probe numbers to cover areas with poor capture.

Here we present the results from the analysis of exome resequencing data of ten patients exhibiting PNAO. We sought to identify variants by taking two analytical approaches; analysis of variants in genes in the platinum pathway and analysis of variants in genes involved in neuronal function or peripheral neuropathy. Variants predicted to be most detrimental to protein function (stop gains and frameshifting indels) were focused on. We identified four genes with a potential role in the development of PNAO (analysed further in Chapter 5).

4.4.4.1 BRCA2

BRCA2 is a tumour suppressor gene which functions to repair DSBs as part of the HR pathway (Roy *et al*. 2011) as well as having roles in the repair of ICLs

(Cipak *et al.* 2006; Section 1.3.4 and 1.3.5). Although DNA repair is critical in the maintenance of neuronal homeostasis (McMurray *et al*. 2005), no previous link of *BRCA2* with peripheral neuropathy has previously been established.

4.4.4.2 ERCC4

ERCC4 encodes the structure specific 5' endonuclease protein XPF which, in complex with ERCC1 (van Vuuren *et al.* 1993; Park *et al.* 1995), plays a role in the NER pathway, the main DNA repair pathway involved in the removal of bulky and DNA distorting adducts (Section 1.3.3), such as those formed by oxaliplatin (Reardon *et al.* 1999). XPF is the catalytic sub-unit of the complex (Enzlin and Schärer, 2002). The complex has also been implicated in ICL repair (Kuraoka *et al.* 2000) and repair of DSBs (Sargent *et al.* 1997; Niedernhofer LJ *et al.* 2004; Ahmad *et al.* 2008; Al-Minawi *et al*. 2009).

4.4.4.3 STOML3

STOML3 encodes a mechanosensory channel, stomatin like protein. It is expressed in the primary sensory neurons in the dorsal root ganglion in mice (Mannsfeldt *et al.* 1999). Deletion of *STOML3* leads to loss of mechanoreceptor function and loss of mechanosensitive currents in isolated neurons from mice (Wetzel *et al.* 2007).

4.4.4.4 NRP2

NRP2 has been shown to have a crucial role in the signalling responsible for peripheral nervous system axonal guidance (Schwarz *et al.* 2009; Roffers-Agarwal and Gammill, 2009). A putative association between variants in *NRP2* has previously been seen with regards to chronic PNAO in GWAS of 96 CRC patients (Lee *et al.* 2010)

Chapter Five – Analysis of candidate genes responsible for PNAO

5.1 Introduction

In Chapter 4, we attempted to uncover a genetic basis for a predisposition to PNAO via exome resequencing of ten patients with extreme forms of the phenotype. By focusing on novel stop gain variants and frameshifting indels involved in the platinum pathway and in neuronal function and/or peripheral neuropathy, we uncovered variants in four candidate genes that potentially had a role in PNAO. Two of those genes are involved in neuronal function. NRP2 has a crucial role in the signalling responsible for peripheral nervous system axonal guidance (Roffers-Agarwal and Gammill, 2009) and *STOML3* encodes a mechanosensory channel (Wetzel *et al.* 2007). The two proteins encoded by genes involved in the platinum pathway are both involved in DNA repair; BRCA2 is involved in the repair of DSBs and ICLs and XPF (encoded by *ERCC4*) is involved in NER, DSB and ICL repair.

Here, we studied these variants and their associated genes to prove a casual role in PNAO, by using a combination of strategies;

1. Analysing control samples.

2. Assaying for additional variants in more patients with PNAO.

3. Analysing functionally related genes.

5.2 Materials and methods

5.2.1 Patient selection

Selection of additional patients with PNAO within the first 12 weeks of treatment was carried out as described in section 4.2.1.

5.2.2 Control samples

We used panels of either 47 or 190 UKBS healthy control subjects to assay for variants in order to assess their frequency in the normal population.

5.2.3 Correlating variants with PNAO

In order to correlate variants with the risk of PNAO, we obtained clinical data regarding the maximum grade of PNAO after 12 weeks of treatment for the entire COIN cohort. We termed 'PNAO' as ≥ grade 3 or removed from treatment within the first 12 weeks, whilst patients graded 0 and 1 were grouped as not suffering from PNAO. Grade 2 patients were not included in any analysis to allow for better discrimination between patients with and without PNAO.

5.2.4 PCR and Sanger sequencing

PCR, verification by agarose gel electrophoresis, product purification, Sanger sequencing and sequencing clean up were carried out as described in sections 2.5.4 to 2.5.8. Sequences were analysed using Sequencer v4.6. All primers used for PCR and Sanger sequencing are given in Appendices 8-11. Primers used for the validation of nonsynonymous variants in NER genes identified by exome resequencing are given in Appendix 7.

5.2.5 Genotyping

Genotyping of Arg415Gln (rs1800067) in *ERCC4,* Asp118 (rs11615) in *ERCC1*, Lys3326X (rs11571833) in *BRCA2* and Gly399Asp (rs2228528), Arg1213Gly (rs2228527) and Gln1413Arg (rs2228529) in *ERCC6* was carried out with Illumina's Fast-Track Genotyping Service using their high throughput GoldenGate technology. Genotyping of Pro379Ser (rs1799802), Arg576Thr (rs1800068), His466Gln (novel), Glu875Gly (rs1800124) and rs1799800 in *ERCC4,* and Asp425Ala (rs4253046), Gly446Asp (rs4253047), Pro694Leu (rs114852424), Ser797Cys (rs146043988), Gly929Arg (novel), Phe1217Cys (rs61760166), Arg1230Pro (rs4253211), Ala1296Thr (rs139509516), Thr1441Ile (rs4253230) and Phe1437Ile (novel) in *ERCC6* was carried out by KBiosciences using their KASPar technology.

5.2.6 *In silico* analysis of variants

LD between variants was assessed using Haploview v4.2. Species alignment of all mammals listed on NCBI was carried out using Clustal Omega. A list of

common species names is given in Appendix 20. The functional consequences of amino acid changes on protein function were determined using Align-GVGD.

5.2.7 Statistical analysis

Differences between patients with and without PNAO and variant status was determined using the Pearsons Chi square test (X^2) or the Fisher exact test if cell counts were <5.

5.3 Results

5.3.1 Patient selection

A second panel of samples, consisting of 54 patients with extreme and dose limiting PNAO after 12 weeks of treatment was selected following review of their toxicity data.

5.3.2 Further analysis of genes implicated in PNAO

5.3.2.1 NRP2 analysis

We screened for the CGCA deletion resulting in a frameshift (Ser904fs), as well as an insertion of a single adenine (Cys907ins) in *NRP2* in a panel of 47 healthy UKBS control subjects. We amplified the region by PCR and carried out Sanger sequencing using primers previously used for validation in Appendix 7. We identified both variants in 3 of the 47 samples; one sample was homozygous and two were heterozygous (2/64 [3.1%] of patients with PNAO compared to 3/47 [4.3%] of healthy controls [*P*=0.65])

5.3.2.2 STOML3 analysis

We carried out PCR and Sanger sequencing to screen the entire ORF, flanking regions and 5'UTR of *STOML3* in 54 additional patients with PNAO. No coding variants were found and intronic variants were observed at frequencies similar to those found in dbSNP (rs9548577 - MAF=0.8% in patients with PNAO compared to 0.8% in dbSNP; rs9574474 - MAF=10.9% compared to 18.1%).

5.3.2.3 BRCA2 analysis

We assayed for Lys3326X (rs11571833) in *BRCA2* in all other available cases from COIN. Genotyping was performed using Illumina's Fast-Track Genotyping Services (San Diego, CA) utilising their high throughput GoldenGate technology. In total, 2,060 samples were genotyped or sequenced successfully. Overall, we found similar proportions of cases with (2/64; 3.1% of patients) and without (36/1,752; 2.1%) PNAO harbouring this variant (X^2=0.35, *P*=0.56).

5.3.2.4 ERCC4 analysis

5.3.2.4.1 Phenotype of patient 8

Through exome resequencing and Sanger sequencing of an independent PCR product, we identified and validated a novel stop gain in *ERCC4* in one patient with PNAO (Patient 8; Chapter 4). We amplified and sequenced the entire ORF and flanking regions of *ERCC4* in this patient and did not find any other coding variants.

The patient was a 79 year old female at the time of undergoing oxaliplatin therapy. She had been diagnosed with metastatic CRC following an ultrasound scan on her liver in March 2006. She had originally presented with right upper quadrant pain and two months of intermittent diarrhoea. Her carcinoembryonic antigen had been 130µg/L following testing, and a computerised tomography scan revealed multiple metastases throughout the liver, as well as a large caecal mass. A biopsy of the liver provided histological diagnosis of adenocarcinoma from a synchronous colonic primary cancer. The patients past medical history included a tubular adenoma which had been excised in 2001, peri-orbital rosacea diagnosed in 2003, excision of a seborrhoeic wart, moderate macular degeneration consistent with her age group and mild osteoarthritis. Allergy skin tests at this time had suggested nickel sensitivity and she was allergic to lidocaine. There was no past medical history of skin cancers, no immunodeficiency disorders or related diseases, no ataxia, memory loss or muscle weakness.

5.3.2.4.2 Analysis of ERCC4 in additional patients with PNAO

We carried out Sanger sequencing of amplified PCR products of the entire ORF, flanking regions and 5'UTR of *ERCC4* in 54 additional patients with PNAO. We identified five nonsynonymous variants: Pro379Ser was found in 3 patients (MAF=4.69%) and was previously documented in dbSNP (rs1799802); Arg415Gln in 9 patients (MAF=14.1%) and in dbSNP (rs1800067); His466Gln in a single patient and not in dbSNP; Arg576Thr in a single patient and in dbSNP (rs1800068) and Glu875Gly in 4 patients (MAF=6.25%) and in dbSNP (rs1800124). Apart from one case that carried both Arg576Thr and Glu875Gly all other cases carried a single *ERCC4* nonsynonymous variant in a heterozygous state.

We also identified 3 synonymous variants (Ala11 [rs3136042], Ser835 [rs1799801] and Thr885 [rs16963255]) and three variants in the 5'UTR (-30T>A [rs1799797], -356C>A [rs6498486] and -69G>C [novel]), all of which were considered unlikely to affect function (Fig. 5.1).

5.3.2.4.3 In silico analysis of nonsynonymous variants in ERCC4

We used Align-GVGD to gauge the potential impact on function of the five nonsynonymous variants identified. Pro379Ser, Arg576Thr and Glu875Gly were all predicted to interfere with function (a score of C65). Arg415Gln was less likely to interfere with function (Class C35) and His466Gln was not predicted to interfere with function (Class C15).

Alignment of all mammalian sequences available on NCBI was carried out using Clustal Omega, revealing that XPF was well conserved across several species. Conservation was seen in all species analysed for Pro379, Arg415, Arg576, Ser613 and Glu875. Although some conservation was observed, His466 was not well conserved (Appendix 24).

We analysed the 5'UTR of *ERCC4* for potential transcription factor binding sites. Although the 5'UTR of *ERCC4* lacks common consensus sequences, there is a TTCGGC(T/C) heptamer repeated ten times within 300bps immediately upstream of the translation start site. This heptamer is moderately conserved between several species potentially validating a regulatory role (Appendix 25). Rs1799797 (MAF=24.4% in patients with PNAO compared to 25% in dbSNP) is seen in the last

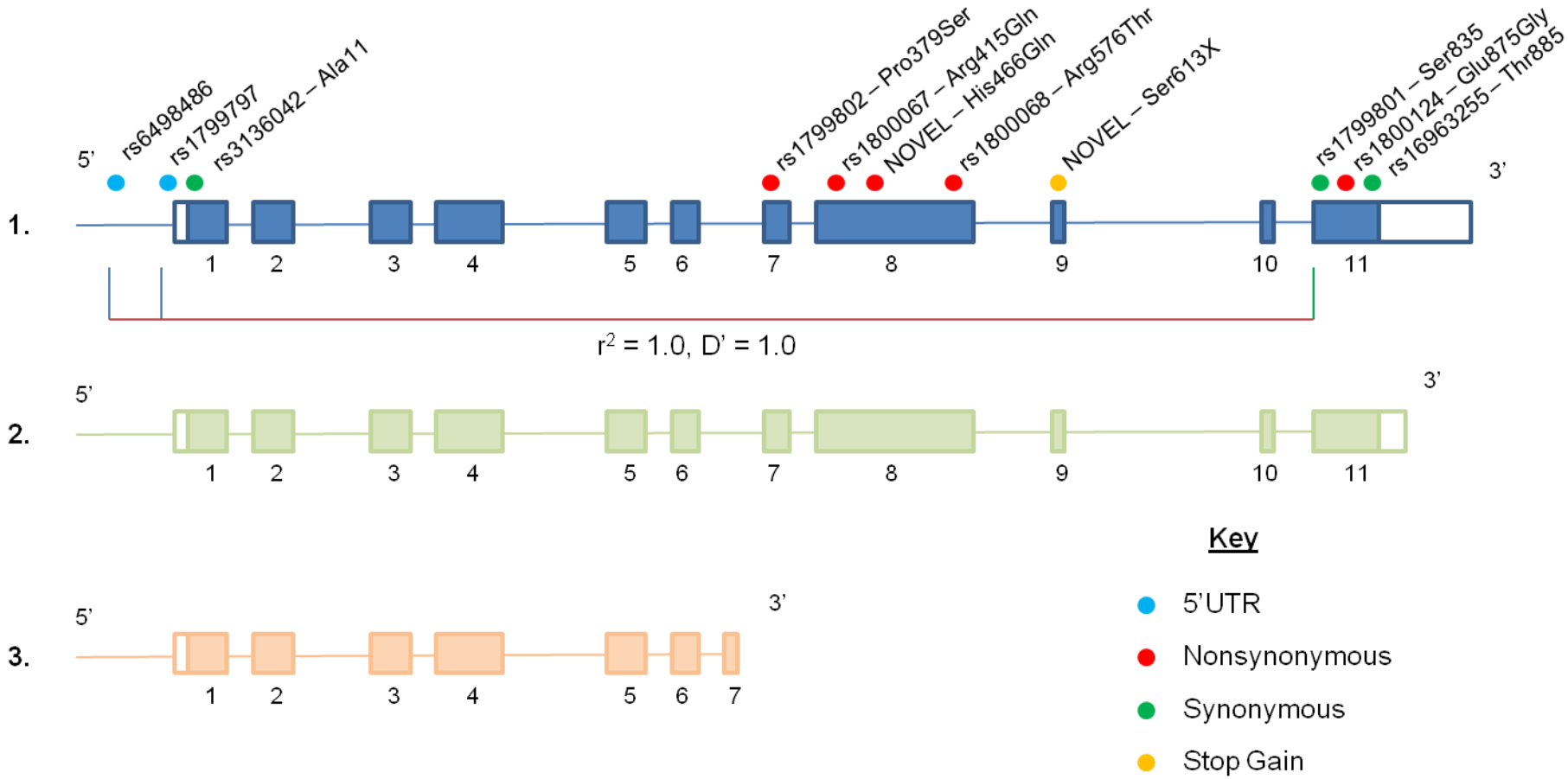


Figure 5.1 – A summary of the transcripts of ERCC4, including exons, ORF (filled in blocks), 5'UTR and variants seen in patients with PNAO

base of the penultimate repeat before the start of exon one and is in high LD with the synonymous variant Ser835 and another variant located in the 5'UTR, rs6498486 (both r^2=1.0, D'=1.0).

5.3.2.4.4 Correlating variants in ERCC4 with PNAO

We genotyped Pro379Ser, His466Gln, Arg576Thr and Glu875Gly in the 2,186 available cases from COIN and COINB using KBiosciences KASPar technology. We also used the same technology to genotype the intronic variant rs1799800 (in strong LD with -356C>A, -30T>A and Ser835; all r^2=1.0, D'=1.0) which has previously been linked with an increased risk of bortezomib induced peripheral neuropathy in the treatment of multiple melanoma (Broyl *et al.* 2010). Arg415Gln was genotyped using Illumina's GoldenGate technology.

For KASPar genotyping of *ERCC4* variants, the overall genotyping success rate was 98.1% (11,075/11,290 genotypes were called successfully) and concordance rate for duplicated samples (n=63) was 100% (315/315 genotypes were concordant). All samples deemed to be heterozygous and homozygous for their respective variant were validated in house via Sanger sequencing (n=33 for rs1799802; n=5 for rs1800068; n=73 for rs1800124). Samples that failed genotyping were Sanger sequenced to determine their genotype (n=26 for rs1799802; n=28 for rs1800068; n=10 for rs1800124). Additionally, following plotting of genotyping data, outliers were identified and sequenced to verify robustness of technology (n=4 for rs1799802; n=1 for rs1800068; n=9 for rs1800124; 100% concordance). In the genotyping of Arg415Gln using Illumina's GoldenGate technology, overall genotyping success rate was 99.95% (2,069/2,070 genotypes were called successfully) and concordance rate for duplicated samples (n=63) was 100% (Fig. 5.2)

We compared the frequencies of individual variants, and variants grouped by their likely effect on function as determined by Align-GVGD, in patients with and without PNAO. Variants predicted to affect protein function included Pro379Ser; Arg576Thr; Glu875Gly. Although each of these rare variants was found more frequently in cases with PNAO as compared to those without (Pro379Ser was in 4.76% cases with PNAO compared with 1.53% of cases without PNAO, Arg576Thr was in 1.59% compared with 0.22%, and Glu875Gly was in 6.35% compared with

3.41%, respectively), none were individually significantly over-represented when analysed using Fishers exact test. However, combined, we found that more patients with PNAO carried a potentially function impairing variant (7/63, 11.11%) as compared to patients without PNAO (90/1,762, 5.11%; X^2=4.23, *P*=0.04). However, there is a potential for novel or private variants in small cohorts to skew the data and therefore we removed Arg576Thr from the analysis. Statistically more patients with PNAO carried Pro379Ser and Glu875Gly than patients without PNAO (7/63, 11.11% compared to 86/1763, 4.88%; X^2=4.89, *P*=0.03).

Arg415Gln, which was predicted by Align-GVGD to be less likely to interfere with function, was found in similar proportions of patients with (9/63, 14.29%) and without (260/1,754, 14.8%) PNAO (*P*=0.91). The novel variant His466Gln was not seen in any further samples and was considered to be private (Table 5.1).

Rs1799800 was not in association with PNAO (24/63, 38.1% of patients with PNAO carried this variant as compared to 834/1,736, 48% without, *P*=0.121).

5.3.3 Analysis of other genes in the NER pathway

5.3.3.1 Analysis of ERCC1

Since XPF and ERCC1 function together to form a 5' incision complex (van Vuuren *et al.* 1993; Park *et al.* 1995), we sought likely causal variants in *ERCC1* via amplification and Sanger sequencing of the ORF, intronic boundaries and 5'UTR in all 64 patients with PNAO. We found three synonymous variants (Thr75 [rs3212947], Asn118 [rs11615], Pro128 [rs139827427]) and five variants in the 5'UTR (-96T>G [rs2298881]; -230C>A [rs41559012]; -303C>T [rs41540513]; -495C>A [rs3212931]; -790T>C [rs3212930]; Fig 5.3).

Rs11615 has previously been associated with response to treatment and, more recently, rate of onset of PNAO in a Japanese population (Ruzzo *et al.* 2007; Inada *et al.* 2010; Oguri *et al* 2013). Therefore, we genotyped the COIN cohort for this variant. Overall, we found similar proportions of cases with (38/64, 59.4% of patients) and without (1,063/1,717, 61.9%) PNAO harbouring this variant, thereby failing to support a casual role (X^2=0.168, *P*=0.682).

	Variant	rs #	Frequency in patients (%)		X^2	*P*	OR (L95-U95)
			+ PNAO	- PNAO			
Predicted to affect function (C65)	**Pro379Ser**	**rs1799802**	**3/63 (4.76%)**	**27/1,763 (1.53%)**	**N/A**	**0.08**	
	Arg576Thr	rs1800068	1/63 (1.59%)	4/1,762 (0.22%)	N/A	0.16	
	Ser613X	Novel	1/63 (1.59%)	-	-		-
	Glu875Gly	**rs1800124**	**4/63 (6.35%)**	**60/1,763 (3.41%)**	**N/A**	**0.28**	
	Total (No private variants)		**7/63 (11.11%)**	**86/1,763 (4.88%)**	**4.89**	**0.03**	**2.44 (1.08-5.51)**
Less likely to affect function (C15-35)	Arg415Gln	rs1800067	9/63 (14.1%)	260/1,754(14.8%)	0.014	0.91	
	His466Gln	Novel	1/63 (1.59%)	0/1,677 (0%)	N/A	0.04	-

Table 5.1 – Nonsynonymous and stop gain variants identified in ERCC4 in patients with and without PNAO analysed with respect to their potential effect on function. Ser613X was not included to determine the total numbers since it was only assayed for in cases with PNAO. Variants seen in more than one PNAO sample (highlighted in bold and shaded) were analysed in a combined analysis (total). We did not include the private variant Arg576Thr in the combined analysis due to the potential to skew the data. Patient 1C (a patient with PNAO) was not included in the analysis since they were not part of the original COIN trial. One patient with PNAO carried both Arg576Thr and Glu875Gly. Another patient without PNAO carried both Pro379Ser and Glu875Gly. Values in the total column reflect the number of patients genotyped. The Chi square (X^2) test was used to test significance or Fishers exact test when values in cells were <5, and respective P value (P) given alongside odd's ratios (OR) with 95% confidence intervals (L95 and U95).

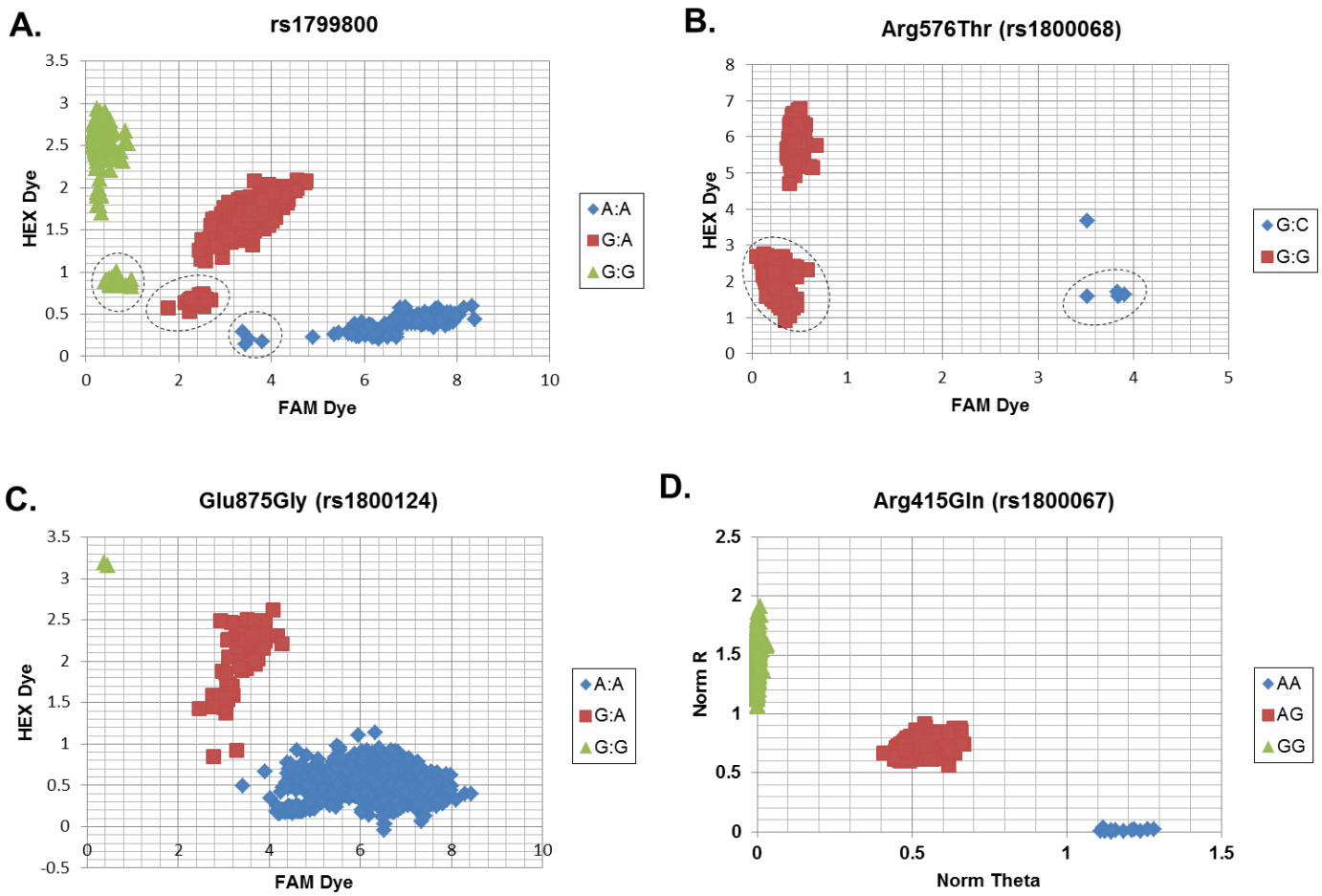


Figure 5.2 – Genotyping cluster plots of A. rs1799800, B. Arg576Thr (rs1800068), C. Glu875Gly (rs1800124) and D. Arg415Gln (rs1800067). Figures A-C were generated through plotting data generated through KASPar technology; Figure D was generated through plotting genotyping data from Illuminas GoldenGate platform. Differential genotype groupings (circled) are due to variance of values as a result of genotyping samples in various batches-this is not outlying data.

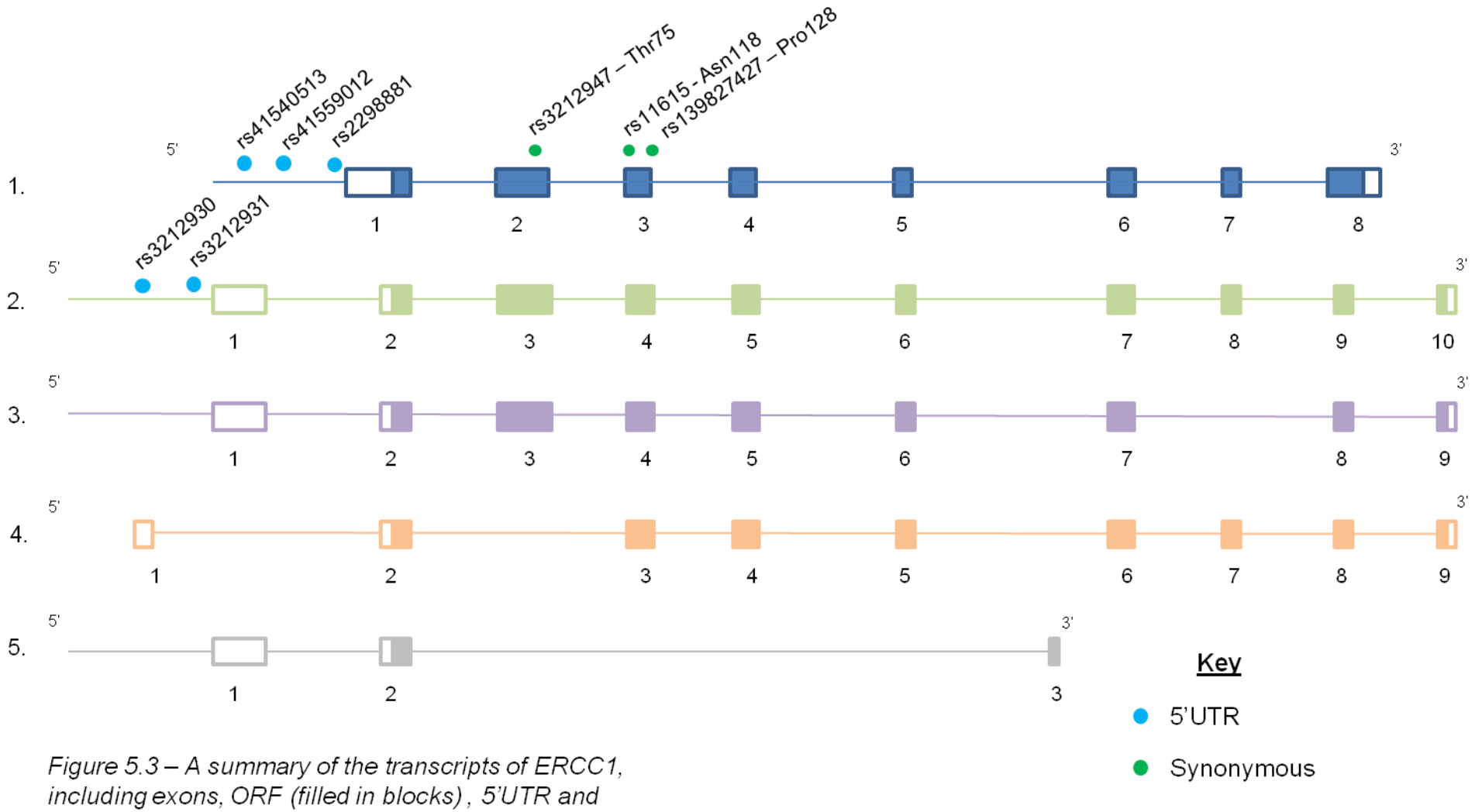


Figure 5.3 – A summary of the transcripts of ERCC1, including exons, ORF (filled in blocks), 5'UTR and variants seen in patients with PNAO.

5.3.3.2 Variants in other ERCC homologs

We examined the exome resequencing data for variants in *ERCC2, ERCC3, ERCC5, ERCC6* and *ERCC8*. We did not find any variants of interest in *ERCC2, ERCC5* or *ERCC8.* However, we identified one novel nonsynonymous variant in *ERCC3* (*XPB* [Arg283Cys]) and three novel nonsynonymous variants in *ERCC6* (*CSB* [Ser797Cys, Gly929Arg and Phe1437Ile]). All were validated by Sanger sequencing of an independent PCR product.

5.3.3.2.1 ERCC3

The region containing Arg283Cys was amplified and Sanger sequenced in 190 UKBS controls. Of the samples successfully sequenced, we discovered that the variant was present in 1 out of 167 control subjects, suggesting that this variant was a low frequency variant.

5.3.3.2.2 ERCC6

The regions containing Ser797Cys, Gly929Arg and Phe1437Ile were amplified and Sanger sequenced in 190 UKBS controls. Of the samples successfully sequenced, we discovered that Ser797Cys was present in 1 out of 155 subjects, suggesting that this variant was a low frequency variant. Neither Gly929Arg nor Phe1437Ile in *ERCC6* were seen in the controls studied. We therefore amplified and sequenced the 5'UTR, ORF and flanking regions of *ERCC6* in the 64 patients with PNAO.

We identified 12 additional nonsynonymous variants; six of these were rare (MAF≤1%) and six were common (MAF >1%). Four of the rare variants identified had a higher MAF in patients with PNAO than that given in dbSNP (Table 5.2). We also identified two synonymous variants (Leu45 [rs2228524] and Gly917 [rs2229760]) and one variant in the 5'UTR (-466G>C), all of which were unlikely to affect function

5.3.3.2.2.1 In silico analysis of nonsynonymous variants

Gly399Asp, Asp425Ala, Gly446Asp, Pro694Leu, Ser797Cys, Arg1215Gly, Phe1217Cys, Arg1230Pro and Thr1441Ile were all predicted to interfere with function (Class C65). Ala1296Thr was predicted as likely to interfere with function

	Align GVGD classification	Exon	Amino acid change [variant ID]	MAF in cases with PNAO	MAF in dbSNP
Rare variants (MAF ≤1%)	N/A	Transcript 2 – Exon 6	Gly929Arg [Novel]	0.80%	Novel
	Class C15	21	Phe1437Ile [Novel]	0.80%	Novel
	Class C55	19	Ala1296Thr [rs139509516]	0.80%	0%
	Class C65	5	Asp425Ala [rs4253046]	2.40%	0.10%
		10	Pro694Leu [rs114852424]	0.80%	0.50%
		13	Ser797Cys [rs146043988]	1.60%	0.10%
		18	Phe1217Cys [rs61760166]	0.80%	0.10%
		21	Thr1441Ile[rs4253230]	1.60%	1.30%
Common variants (MAF >1%)	Class C15	18	Met1097Val [rs2228526] †	20.50%	20%
	Class C35	21	Gln1413Arg [rs2228529] †	21.40%	20%
	Class C65	5	Gly399Asp [rs2228528]	15.10%	16.10%
		5	Gly446Asp [rs4253047]	2.40%	3.10%
		18	Arg1215Gly [rs2228527] †	21.40%	20%
		18	Arg1230Pro [rs4253211]	11.40%	10.80%

Table 5.2 – All nonsynonymous variants identified in ERCC6. Shaded are rare variants that appear to be more common in patients with PNAO compared to the frequency data given in dbSNP. † Variants seen in high LD with each other

(Class C55). Gln1413Arg was less likely to interfere with function (Class C35), and Met1097Val and Phe1437Ile were not predicted to interfere with function (Class C15). We were unable to assess the novel variant Gly929Arg due to it being located in an alternative transcript for which the protein sequence was not available.

5.3.3.2.2.2 Correlating variants in ERCC6 with PNAO

Following the identification of 14 nonsynonymous variants in *ERCC6* we genotyped the 2,186 available cases from COIN and COIN-B. Gly399Asp, Arg1215Gly and Gln1413Arg were genotyped using Illumina's GoldenGate platform; Asp425Ala, Gly446Asp, Pro694Leu, Ser797Cys, Gly929Arg, Phe1217Cys, Arg1230Pro, Ala1296Thr, Thr1441Ile and Phe1437Ile were genotyped using KBioscience KASPar technology. We did not genotype Met1097Val since this variant was in high LD with Gly399Asp and Arg1215Gly ([Met1097Val-Arg1213Gly; r^2 = 0.99, D' = 1.0], [Met1097Val-Gln1413Arg; r^2 = 1.0, D' = 1.0], [Arg1213Gly-Gln1413Arg; r^2 = 0.99, D' = 1.0]).

For KASPar genotyping, the overall genotyping success rate was 97% (22,340/23,020 genotypes were called successfully) and concordance rate for duplicated samples (n=63) was 99.2% (625/630 genotypes were concordant). For Illumina genotyping, the overall genotyping success rate was 99.85% (6548/6558 genotypes were called successfully) and concordance rate for duplicated samples (n=63) was 100% (189/189 genotypes were concordant).

Of the rare variants that were predicted to be damaging, we found that Asp425Ala, Pro694Leu and Ser797Cys were individually statistically overrepresented in patients with PNAO (Asp425Ala, 4.76% in patients with PNAO compared to 0.86% in patients without PNAO, *P*=0.02; Pro694Leu, 1.59% and not present in patients without PNAO, *P*=0.04; Ser797Cys, 3.29% compared to 0.17%, *P*=0.01). One patient with PNAO was heterozygous for both Asp425Ala and Ser797Cys.

Combined we found that these five rare variants were statistically associated with PNAO (11.11% compared to 1.47%, $P=1.7 \times 10^{-8}$). However, there is a potential for novel and private variants in small cohorts to skew associations. Therefore, we conducted the combined analysis without Pro694Leu, Phe1217Cys and Thr1441Ile.

We observed a statistically significant over representation of Asp425Ala and Ser797Cys in patients with PNAO (6.78% compared to 1.04%, P=6x10^{-3}).

The novel nonsynonymous variants from exome resequencing, Glu929Arg and Phe1437Ile, were seen in patients without PNAO. Neither were individually associated with risk of PNAO (Glu929Arg, 1.59% compared to 0.23%, P=0.16; Phe1437Ile, 1.59% compared to 0.06%, P=0.07). Also, the rare variant not predicted to be damaging, Ala1296Thr, was not associated with PNAO risk (1.59% compared to 0.17%; P=0.13; Table 5.3).

None of the common variants predicted to affect function were statistically associated with PNAO (Gly399Asp, 28.57% compared to 30.65%, P=0.73; Gly446Asp, 4.76% compared to 3.1%, P=0.32; Arg1215Gly, 36.51% compared to 34.61%, P=0.76; Arg1230Pro, 20.63% compared to 19.13%, P=0.77). Similarly, the common variant predicted to be less likely to affect function, Gln1413Arg, was not associated with PNAO risk (36.51% compared to 34.46%, P=0.74; Table 5.4)

5.3.3.2.2.3 Combined analysis of ERCC4 and ERCC6 rare variants

We carried out a combined analysis for the two rare variants from *ERCC4* (Pro379Ser and Glu875Gly) and the two rare variants in *ERCC6* (Asp425Ala and Ser797Cys) shown to be associated with PNAO risk. We found that significantly more patients with PNAO had one of these rare variants in *ERCC4* or *ERCC6* in comparison to those without PNAO (11/63 [17.46%] compared to 103/1742 [5.97%]; X^2=13.5, P=2.4x10^{-4}; Table 5.5).

	Variant	rs #	Frequency in patients (%)		*P*	OR (L95-U95)
			+ PNAO	- PNAO		
Predicted to affect function (C65)	**Asp425Ala**	**rs4253046**	**3/63 (4.76%)**	**15/1,756 (0.86%)**	**0.02**	**5.80 (1.64-20.58)**
	Pro694Leu	rs114852424	1/63 (1.59%)	0/1,761	0.04	84.55 (3.41-2096.37)
	Ser797Cys	**rs146043988**	**2/63 (3.29%)**	**3/1,754 (0.17%)**	**0.01**	**19.14 (3.14-116.62)**
	Phe1217Cys	rs61760166	1/63 (1.59%)	3/1,738 (0.17%)	0.13	
	Thr1441Ile	rs4253230	1/63 (1.59%)	4/1,745 (0.22%)	0.16	
	Total (No private variants)		**4/63 (6.78%)**	**18/1,748 (1.04%)**	**6.10E-03**	**39.44 (8.63-180.18)**
Less likely to affect function (C15-C55)	Ala1296Thr	rs139509516	1/63 (1.59%)	3/1,748 (0.17%)	0.13	
	Phe1437Ile	Novel	1/63 (1.59%)	4/1,752 (0.23%)	0.16	
No information	Glu929Arg	Novel	1/63 (1.59%)	1/1,752 (0.06%)	0.07	

Table 5.3 – Rare nonsynonymous variants identified in ERCC6 in patients with and without PNAO analysed with respect to their potential effect on function. Patient 1C (a patient with PNAO) was not included in the analysis since they were not part of the original COIN trial. One patient with PNAO carried two of the predicted to be functional rare nonsynonymous variants (Asp425Ala and Ser797Cys). Variants seen in more than one PNAO sample (highlighted in bold and shaded) were analysed in a combined analysis (total). Fishers exact test was used and respective P value (P) given alongside odd's ratios (OR) with 95% confidence intervals (L95 and U95)

	Variant	rs #	Frequency in patients (%)		X^2	*P*
			+ PNAO	- PNAO		
Predicted to affect function (C65)	Gly399Asp	rs2228528	18/63 (28.57%)	536/1,749 (30.65%)	0.12	0.73
	Gly446Asp	rs4253047	3/63 (4.76%)	54/1,756 (3.1%)	N/A	0.32
	Arg1215Gly	rs2228527	23/63 (36.51%)	607/1,754 (34.61%)	0.1	0.76
	Arg1230Pro	rs4253211	13/63 (20.63%)	335/1,751 (19.13%)	0.09	0.77
Less likely to affect function (C35)	Gln1413Arg	rs2228529	23/63 (36.51%)	604/1,753 (34.46%)	0.11	0.74

Table 5.4 – *Common nonsynonymous variants identified in ERCC6 in patients with and without PNAO analysed with respect to their potential effect on function. Patient 1C (a patient with PNAO) was not included in the analysis since they were not part of the original COIN trial. The Chi square (X^2) test was used to test significance and P values (P) are given.*

	Variants	Frequency in patients (%)		X^2	*P*	OR (L95-U95)
		+ PNAO	- PNAO			
ERCC6	Asp425Ala, Ser797Cys	4/63 (6.78%)	18/1,748 (1.04%)	N/A	$6.1x10^{-3}$	39.44 (8.63-180.18)
ERCC4	Pro379Ser, Glu875Gly	7/63 (11.11%)	86/1,763 (4.88%)	4.89	0.03	2.44 (1.08-5.51)
	Total	**11/63 (17.46%)**	**104/1,743 (5.97%)**	**13.5**	**$2.4x10^{-4}$**	**3.33 (1.69-6.68)**

Table 5.5 – Combined analysis of four rare, predicted to be damaging, nonsynonymous variants in ERCC4 and ERCC6. Novel and private variants (seen in one patient with PNAO) were not included due to the potential of such variants to skew analyses.

5.4 Discussion

5.4.1 Excluding roles of *NRP2, STOML3* and *BRCA2* in PNAO

Two frameshifting indels in *NRP2* originally seen in two patients with PNAO were also identified in 3 UKBS healthy control samples, suggesting these variants were common polymorphisms. We therefore ruled *NRP2* out of future analysis.

Following the identification of a novel stop gain variant in *STOML3* in one patient with PNAO, we attempted to find further rare variants in additional patients with PNAO by sequencing the ORF, flanking region and 5'UTR of *STOML3.* We failed to find any other coding variants to support a role for *STOML3* in PNAO.

We identified a known stop gain variant, Lys3326X, in two patients with PNAO in *BRCA2*. Following genotyping and analysis of this variant in the COIN cohort, we failed to find an association between the variant and PNAO risk. We therefore ruled it out of future analysis.

5.4.2 *ERCC4*

5.4.2.1 Hereditary disease associated with ERCC4

Biallelic mutations in *ERCC4* are known to cause the autosomal recessive UV sensitivity disorder XP, group F (XPF; OMIM #278760) characterised by an elevated risk of cancer, in particular skin and oral cancers (Section 1.3.3.1; Matsumura *et al.* 1998; Lehmann *et al*. 2011). In these patients, expression of XPF is reduced to approximately 5% of that seen in normal cells (Brookman *et al.* 1996). However, in comparison to other XP complementation groups, the XPF phenotype is considered mild, with the majority of cases seen in Japanese patients (Gregg *et al*. 2011). Recently, biallelic mutations in *ERCC4* have also been attributable to the development of FA (OMIM #615272; Bogliolo *et al.* 2013), Cockayne syndrome (CS) and an XP-CS-FA phenotype (Kasiyama *et al.* 2013). FA is characterised by an increased risk of various cancers and bone marrow failure (Section 1.3.5.1). CS sufferers exhibit growth retardation, photosensitivity and impairment of the nervous system. Additionally, homozygosity of the Arg153Pro allele has been shown to cause XPF-ERCC1 (XFE) progeroid syndrome (OMIM #610965), characterised by moderate UV sensitivity and an accelerated rate of aging (Niedernhofer *et al.* 2006).

5.4.2.2 ERCC4 and Patient 8

A stop gain (Ser613X) in *ERCC4* was identified in one patient with PNAO. After reviewing this patient's medical record, we found no indication of XP. We assayed for a mutation in the second *ERCC4* allele in the patient by Sanger sequencing of amplified products of their entire ORF and flanking intronic sequences. We failed to find any other coding variants, indicating that the patient was haploinsufficient for a mutant allele.

5.4.2.3 Variants identified in ERCC4

We identified five nonsynonymous variants in *ERCC4* in patients with PNAO. Two of these variants, which were individually rare and predicted to be damaging (Pro379Ser and Glu875Gly), were shown to collectively contribute to the development of PNAO.

In addition, we also identified two variants in the 5'UTR of *ERCC4.* Although the 5'UTR of *ERCC4* lacks common consensus sequences, there is a heptamer repeat with mild species conservation. Rs1799797 is seen in the last base of the penultimate repeat before the ORF and is in high LD with the synonymous variant Ser835 and another variant located in the 5'UTR, rs6498486. However, since it was seen to be at a similar frequency in patients with PNAO to that reported in dbSNP, it was considered to be a benign polymorphism.

5.4.2.4 ERCC4 in chemotherapy induced peripheral neuropathy

The role of XPF in chemotherapy induced peripheral neuropathy was explored in more detail in the bortezimab treatment of multiple myeloma. Patients carrying the intronic variant rs1799800 and the silent polymorphism Ser835 in *ERCC4*, were at a 2.74 and 2.48-fold greater risk, respectively, of developing late onset peripheral neuropathy after treatment with bortezimab (Broyl *et al.* 2010). These two variants were shown to be in high LD with one another. We genotyped rs1799800 but found no association with PNAO.

5.4.3 Other *ERCC* family members

5.4.3.1 ERCC1

XPF functions in a complex with ERCC1 (van Vuuren *et al.* 1993; Park *et al.* 1995). ERCC1 has roles in binding to single stranded DNA and localising the complex to the area of DNA damage (Tsodikov *et al.* 2005; Tripsianes *et al.* 2005). The complex structure formation is critical since either subunit requires the other for stabilisation, although it is thought this is more crucial for XPF protein stability (de Laat *et al.* 1998; Arora *et al.* 2010). Biallelic mutations in *ERCC1* have been shown to cause XP (Gregg *et al.* 2011), CS (Kasiyama *et al.* 2013) and cerebro-oculo-facio-skeletal syndrome (COFS; OMIM #610758). COFS represents a moderate sensitivity to UV light but with a severe growth failure (Jaspers *et al.* 2007).

Here we investigated a role of *ERCC1* in PNAO by amplifying and Sanger sequencing the 5'UTR, ORF and flanking regions in all patients identified with the phenotype. We identified five variants in the 5'UTR, all of which were unlikely to affect function. We also identified three synonymous variants. Despite these variants being unlikely to affect protein function, researchers have previously found a correlation between the silent polymorphism, *ERCC1*^{Asn118}, and rate of onset of PNAO in a Japanese population (Inada *et al.* 2010; Oguri *et al.* 2013). In chapter 4, we observed the variant at a reduced frequency in the initial ten PNAO patients in comparison to dbSNP. Here we sought and failed to find an association between the variant and PNAO, confirming that it is unlikely to have an effect on Caucasian populations.

5.4.3.2 ERCC6

ERCC6 encodes CSB, a SWI/SNF DNA-dependent related ATPase (Troelstra *et al*. 1992). It is recruited to areas of damage following stalling of RNApolII at DNA lesions and has multiple roles including chromatin remodelling (Citterio *et al.* 2000) and recruitment of other NER proteins (Fousteri *et al.* 2006). Inactivating mutations have been shown to predispose patients to CS, group B (CSB, OMIM #133540; Mallery *et al.* 1998) characterised by physical and mental retardation, premature aging, neurological abnormalities, retinal degeneration, hearing loss and sensitivity to UV light (Nance and Berry, 1992).

We identified fourteen nonsynonymous variants in ERCC6, five of which were rare and predicted to be detrimental to protein function. After exclusion of private variants, we have shown that two of these variants (Asp425Ala and Ser797Cys) collectively contribute to the likelihood of PNAO. These variants have not previously been linked to CSB.

5.4.4 Rare variant hypothesis

When analysed together we observed a collective contribution of two rare variants in *ERCC4* seen in more than one patient (Pro379Ser and Glu875Gly) and two rare variants in *ERCC6* (Asp425Ala and Ser797Cys) in PNAO patients. Although private and novel variants were identified in both genes, we did not include such variants in this analysis as there is a potential for such variants in a small cohort to skew the data and bias the statistics.

The 'rare variant hypothesis' has previously been used to explain how individually rare but collectively common variants could contribute to disease etiology. For example, multiple rare nonsynonymous variants in three genes (*ABCA1, APOA1* and *LCAT*) have been shown to be associated with low levels of HDL-cholesterol (HDL-C); a risk factor of artherosclerosis (Cohen *et al.* 2004). Similarly, rare nonsynonymous mutation in Wnt signalling genes (*APC*, *AXIN* and *CTNNB1*), and mismatch repair genes (*MSH2* and *MLH1*) have been shown to collectively contribute to an increased predisposition risk to CRA (Fearnhead *et al.* 2004*;* Azzopardi *et al*. 2008). The data presented here suggests that variants from multiple genes in the NER pathway could be contributing to the risk of PNAO.

Chapter Six – Construction of a model system to test the functionality of variants identified in *ERCC4*

6.1 Introduction

The ability of oxaliplatin to form DNA crosslinks is critical for the drugs function as a chemotherapeutic in the treatment of CRC (Brabec and Kasparkova, 2005). It is primarily the role of the NER pathway to remove bulky intrastrand adducts, the predominant lesion formed following oxaliplatin treatment (Reardon *et al.* 1999). The XPF (encoded by *ERCC4*) and ERCC1 complex is primarily involved in the 5' incision of DNA during NER of such damage, although the complex also has roles in the repair of ICL and DSBs (Sections 1.3.3 and 4.4.4.2).

In Chapter 4, we identified a novel truncation mutation in *ERCC4* in one patient with PNAO. In Chapter 5, we subsequently characterised two rare nonsynonymous variants in *ERCC4* in seven other patients with PNAO that collectively significantly altered the risk of developing this side effect. Typically, it is desirable to validate biomarker findings in a validation cohort. However, since the variants observed were rare (collective MAF in controls was 2.4%) we would require a very large cohort of patients in order to observe the effect size seen in COIN (OR=2.44). In COIN, we had approximately 60% power at a 5% significance level to detect this OR. In addition to this, the phenonomen of the 'winners curse' could mean that this OR was elevated. Therefore, if we consider a more modest OR of 1.8, with 75% power at a 5% significance level, we would require in excess of 250 samples with PNAO and 7000 samples without PNAO to validate our findings. Since all other trials utilising oxaliplatin as part of the treatment regimen consisted of far fewer patients (n<1000), this was not possible.

Since we cannot easily validate our findings in an independent cohort, we modeled our variants in the organism *S.pombe* in order to gauge their effects on function. *S.pombe* was the organism of choice as it is a simple, eukaryotic model, with a well annotated genome (Section 1.8). In addition to roles in NER, intrastrand crosslink repair, ICL repair and HR, the XPF *S.pombe* homolog, Rad16, has also been shown to play a role in checkpoint signalling and MMR independent from

typical MMR pathway proteins (Carr *et al.* 1994; Fleck *et al.* 1999; Prudden *et al.* 2003; Bøe *et al.* 2012).

In an attempt to develop a model system to test the effect of residue changes on the function of XPF, we genetically manipulated *rad16* using *Cre* recombinase mediated cassette exchange (RMCE). We sought to initially knock out *rad16* in the creation of a *rad16* base strain, followed by restoration of wild type *rad16* with flanking *lox* sites in order to test functionality. Mutations of interest were created by SDM on a constructed vector and RMCE was used to introduce the mutated cassettes into the *rad16* base strain.

6.2 Materials and methods

6.2.1 Construction of the *rad16* deletion base strain (Fig. 6.1)

6.2.1.1 Construction of loxP-ura4⁺-loxM3 PCR product

For the construction of a *rad16* deletion (*rad16Δ*) base strain, we used PCR with primers that incorporated a 100bp region upstream and downstream of the genomic *rad16* locus. The 3' ends of the primers were also designed to incorporate a 20 nucleotide region in pAW1 (Appendix 12; Bähler *et al*.1998), in order for amplification of the *ura4+* gene with flanking lox sites (so called *ura4⁺*F and *ura4⁺*R – Appendix 13; Appendix 14 for *lox* sites), as described by Watson *et al.* (2008). Primers with HPLC purity were purchased from MWG. PCR was carried out on lineralised pAW1 as the target using MMG® (Section 2.5.4). PCR conditions consisted of an initial denaturisation of 95˚C for 2 minutes, followed by 30 cycles of 95˚C for 20 seconds, 55˚C for 30 seconds and 72˚C for 90 seconds. This was followed by a final elongation step of 72˚C for 10 minutes. Reaction mixtures from two identical PCRs were pooled and ethanol precipitation carried out. The pellet was dissolved in dH_2O and an aliquot run on 1.5% agarose gel (Section 2.5.5). The PCR product consisted of *ura4⁺* with flanking lox sites and with a 100bp sequence specific to the *rad16* genomic region at the 5' and 3' ends (hereafter *loxP-ura4⁺-loxM3*).

6.2.1.2 Lineralisation of pAW1

Lineralisation of pAW1 was carried out in order to achieve a more efficient PCR reaction. 40ng of pAW1 was lineralised with 1 unit of *Acc*I (New England Biolabs) in supplied buffer at 37˚C for 2 hours.

6.2.1.3 Transformation of loxP-ura4⁺- loxM3

We transformed the *loxP-ura4*+- *loxM3* PCR product into a wild type strain of *S.pombe* in order to knock out *rad16* and incorporate *lox* sites at the locus to allow for ease of future recombination events. For homologous integration of the *loxP-ura4⁺- loxM3* PCR product at the *rad16* genomic locus in the EH238 wild type strain (*ura4-D18, leu1-32*), the LiAc method of transformation was utilised (Section 2.5.13.7). LiAc reaction product was plated onto MMA supplemented with leucine (MMA +leu) and allowed to grow for 5-7 days. Following this, successfully growing colonies were streaked out onto a MMA+leu master plate and left to grow for 5-7 days.

6.2.1.4 Enrichment by UV sensitivity

Rad16 plays a crucial role in the repair of DNA intrastrand cross links created by UV damage (McCready *et al.* 1993). Enrichment of UV sensitive colonies was an easy way to screen for either knock out or insertion of *rad16*. Oliver Fleck (Bangor University) carried out the enrichment process. Replica plates of the potential transformants were created by transfer using a replicating block covered with sterile velvet onto a YEA plate. A *rad13Δ* (homologous to human *ERCC5/XPG*; Rad13 is another important component in the repair of UV induced intrastrand crosslinks as part of the NER pathway; McCready *et al.* 1993) base strain was streaked onto the plates, as well as the unaltered EH238 wild type strain to act as controls. Transformants and controls were tested for UV sensitivity by treatment with 50-100J/m^2 of UV light using a Stratalinker.

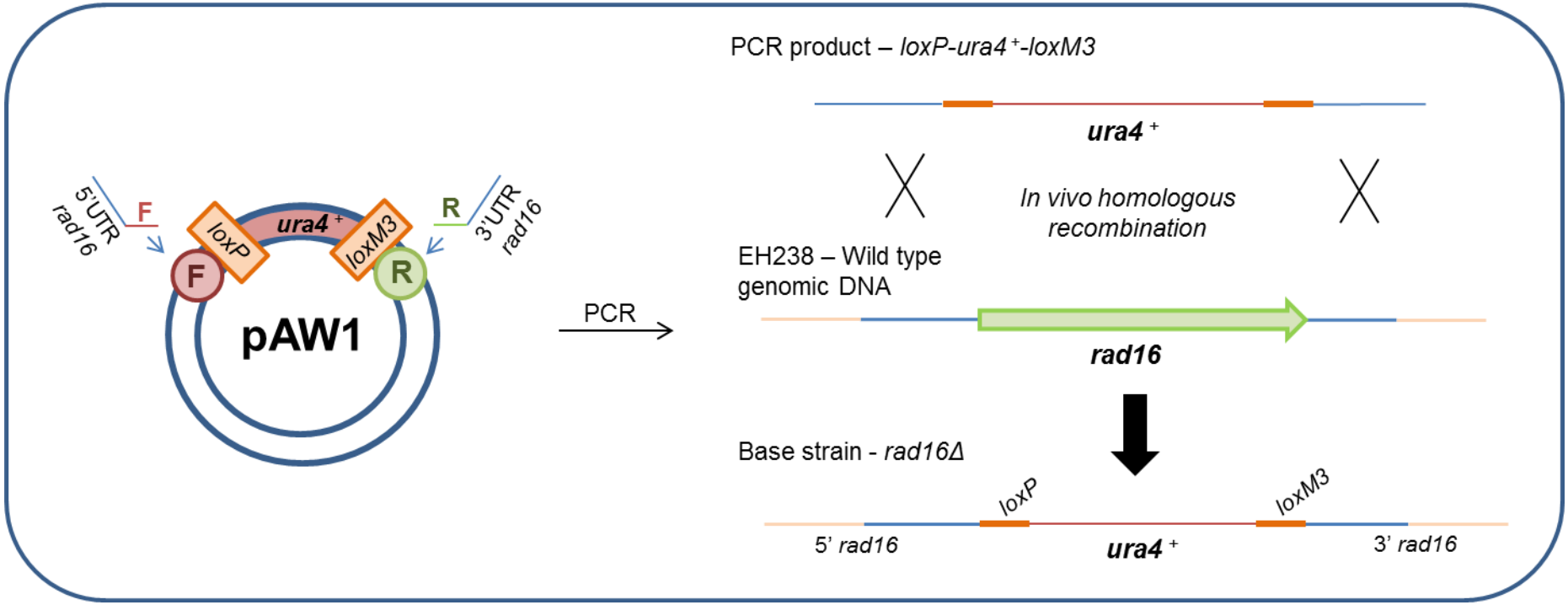


Figure 6.1 – Construction of the rad16Δ base strain. A PCR product was created using primers designed to amplify loxP-ura4⁺-loM3 in pAW1. Primers also incorporated 100bps of sequence upstream and downstream of rad16. Homologous recombination between the PCR product and rad16 in the wild type strain, EH238, allowed for successful knock out of rad16 and incorporation of lox sites to aid future recombination events.

6.2.1.5 Colony PCR of UV sensitive transformants

We carried out colony PCR (Section 2.5.13.5) on transformants that displayed UV sensitivity using primers A, B and C (Appendix 15). Primer A was designed to hybridise to the sequence upstream of the integration site, whereas B and C were designed to be specific to the *rad16* and *ura4*$^{+}$ gene insert respectively. 10pmol of primers A, B and C were all added to the colony PCR reaction mixture and results analysed on 1.5% agarose gels (Section 2.5.5). Successful colonies were transferred to liquid media and frozen at -80˚C.

6.2.1.6 PCR and sequencing of lox sites

PCIA extraction (Section 2.5.13.6) was carried out to isolate genomic DNA to allow for PCR and Sanger sequencing of *lox* sites to ensure integrity. PCR, verification by agarose gel electrophoresis, product purification, Sanger sequencing and sequencing clean up were subsequently carried out (Sections 2.5.4 to 2.5.8). Sequences were analysed using Sequencer v4.6. Primers used are given in Appendix 16.

6.2.2 Cloning of *rad16*$^{+}$ (Fig. 6.2)

6.2.2.1 Construction of the loxP-rad16^{+}*-loxM3 PCR product*

PCR was carried out using MMG (Section 2.5.4). Primers, *rad16*-Forward and *rad16*-Reverse, incorporating the relevant *lox* sites were designed with the 5’ and 3’ ends of the ORF of *rad16* incorporated. Additionally, in *rad16*-Forward, nucleotides encoding an N-terminal histidine tag ($[His]_6$) were integrated (Appendix 17). Primers were from MWG and of HPLC purity. Using previously extracted genomic DNA from *S.pombe,* thermocycling conditions consisted of an initial denaturation of 95˚C for 2 minutes, followed by 30 cycles of 95˚C for 20 seconds, 51˚C for 30 seconds and 72˚C for 3 minutes and 30 seconds. This was completed by a final elongation step of 72˚C for 10 minutes. The reaction was repeated ten times, reaction mixtures pooled and ethanol precipitation carried out. The pellet was resuspended in dH_2O and an aliquot was run on a 1.5% agarose gel.

6.2.2.2 Lineralisation of pAW8-ccdB

Lineralisation of pAW8-*ccdB* was required for a more efficient *in vitro Cre* recombinase reaction by relaxing supercoiled plasmid. 500ng of pAW8-*ccdB* was lineralised with 2 units of *Spe*I (New England Biolabs) in the supplied buffer. The reaction mixture was incubated for 1 hour at 37°C and an aliquot run on an agarose gel alongside an aliquot of undigested plasmid to confirm lineralisation.

6.2.2.3 In vitro Cre recombinase reaction between loxP-rad16⁺-loxM3 and pAW8-ccdB

Recombination of *loxP-rad16⁺-loxM3* was carried out with pAW8-*ccdB*, allowing for switching of *rad16⁺* at the *ccdB* locus via corresponding *lox* sites. Molar ratios of 1:4, plasmid to insert were calculated in order to aid a more efficient *Cre* recombinase reaction. A standard *Cre* reaction was carried out (Section 2.5.12.8) and consisted of 100ng pAW8-*ccdb* and 150ng *loxP-rad16⁺-loxM3* PCR product.

6.2.2.4 Transformation of electrocompetent E.coli cells with Cre recombinase reaction product.

Electroporation was used for transformation of DH5α *E.coli* electrocompetent cells with 1µl of *Cre* recombinase reaction product mixture and 25µl of cells (Section 2.5.12.10). 100µl of transformation reaction was plated out onto LB plates with 100µg/ml amplicillin and incubated at 30˚C for approximately 24 hours. Successful transformants were established in LB with 100µg/ml amplicillin and left to grow at 30˚C overnight with shaking. Following this, plasmid extraction was carried out by Rebecca Williams (PhD student, Fleck group, Bangor) using the Machery-Nagel Nucleospin® plasmid extraction kit.

6.2.2.5 Verification of successful cloning

Verification that *loxP-rad16⁺-loxM3* had been successfully inserted into pAW8 in place of *ccdB* was determined by restriction digest with *Bam*HI by Rebecca Williams. 5µl of extracted plasmid was digested with 2 units of *Bam*HI (New England Biolabs) with the supplied buffer and incubated for 1 hour at 37˚C. Following this, an aliquot of digestion product was run on a 1% agarose gel alongside an aliquot of undigested plasmid (Section 2.5.5).

6.2.3 Construction of *rad16*$^{+}$ strain (Fig.6.2)

6.2.3.1 Transformation of pAW8-rad16^{+} *into rad16Δ base strain*

For homologous integration of the *rad16*$^{+}$ cassette at the genomic locus, pAW8-*rad16*$^{+}$ was transformed into the *rad16Δ* base strain using the LiAc method (Section 2.5.13.7). Reaction products were plated out on EMM supplemented with thiamine (EMM+thi) to induce expression of *Cre* recombinase through activation of the no message in thiamine (*nmt41*) promoter of pAW8. Cells were allowed to grow at 30˚C for approximately 4 days, at which point they were streaked out onto EMM+thi masterplates. The following steps were carried out by Oliver Fleck. Individual colonies were propagated in YEL to allow for removal of the plasmid and subsequently prevent further *Cre* recombinase action. Following 2 days of incubation at 30°C, cells were streaked out on YEA+5-FOA to allow for adequate selection for *ura4*$^{-}$ strains (5-FOA resistant, 5-FOA^{R}; leu-). After incubation for 2 days at 30˚C, YEA masterplates were produced.

6.2.3.2 Enrichment by high dose UV sensitivity

The 5-FOA^{R} transformants were further analysed by assaying for restored DNA damage repair capacity by treating a replica of the masterplate with 200J/m^{2} of UV using a Stratalinker. The wild type strain, EH238, and the *rad16Δ* strain were also treated in the same manner to act as a comparison for repair proficient and deficient strains, respectively.

6.2.3.3 Enrichment by UV and MMS spot test treatment

Transformants showing resistance to high dose UV were further analysed for sensitivity by spot tests treatment of 50-100J/m^{2} of UV and 0.01-0.015% MMS. The wild type strain, EH238, and the *rad16Δ* strain were also treated in the same manner to act as a comparison for repair proficient and deficient strains, respectively.

6.2.3.4 Colony PCR of UV and MMS resistant transformants

Those colonies verified as insensitive to UV and MMS were analysed further by colony PCR (Sections 2.5.13.5 and 6.2.1.5, Appendix 15).

6.2.3.5 PCR and sequencing of the ORF of rad16^{+}

Genomic DNA was extracted using the PCIA method (Section 2.5.13.6). The *rad16*$^{+}$ ORF and *lox* sites were amplified by PCR in order to gauge their integrity by Sanger sequencing (Section 6.2.1.6). Primers used are given in Appendix 18.

6.2.4 SDM of pAW8-*rad16*$^{+}$

6.2.4.1 Mutant plasmid synthesis (rad16^{MT})

Mutant strand synthesis and transformation of electrocompetent cells was carried out using the QuikChange Lightning SDM kit (Section 2.5.12.9), utilising pAW8-*rad16*^{WT} and primers designed to create the variant amino acids. Primers were designed between 30-37 nucleotides in length, <40% GC content and with the mutation of interest in the centre of the primer with at least 10 bases either side to allow for adequate binding to the template. Primers used are given in Appendix 19.

6.2.4.2 Extraction of rad16^{MT} plasmids

Successfully growing colonies were added to liquid LB with 100µg/ml of amplicillin. Following incubation for 16-18 hours at 30°C, plasmid extraction was carried out with Qiagen miniprep plasmid extraction kits following the manufacturers' protocol (Section 2.5.12.7).

6.2.4.3 PCR and Sanger sequencing of the ORF of rad16^{MT}

The ORF and flanking *lox* sites of the extracted plasmids were analysed by Sanger sequencing of an independent PCR product (Section 6.2.1.6, Appendix 18) to verify the integrity of the gene and *lox* sites, as well as ensuring that the appropriate mutation had been introduced.

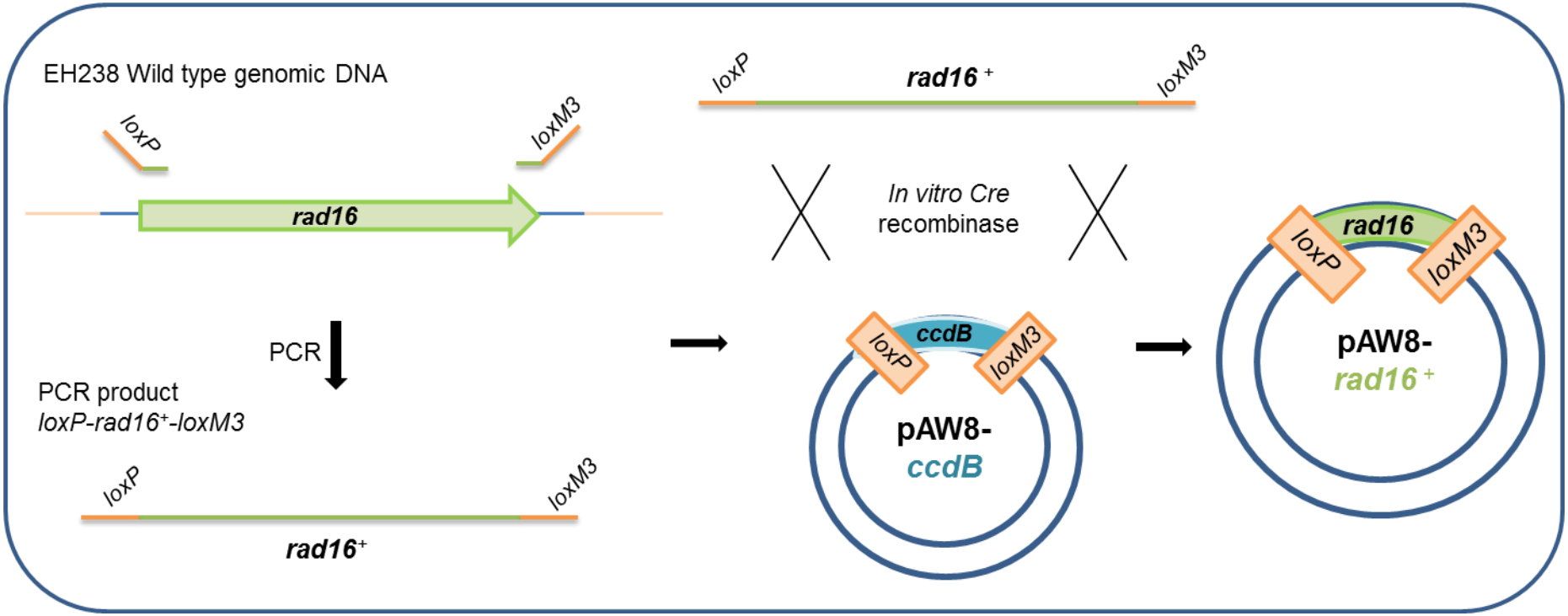


Figure 6.2 – Construction of pAW8-rad16⁺. A PCR product consisting of the entire rad16 gene with flanking lox sites was produced by amplification of EH238 wild type genomic DNA. In vitro RMCE was carried out between pAW8-ccdB and loxP-rad16⁺-loxM3 PCR product, allowing for successful integration of rad16⁺ into the vector.

6.2.5 Construction of *rad16^{MT}* strains (Fig. 6.3)

6.2.5.1 Transformation of pAW8-rad16^{MT} into rad16Δ base strain

For homologous integration of the *rad16^{MT}*cassette, pAW8-*rad16^{MT}* was transformed into the *rad16Δ* base strain using the LiAc method (Section 2.5.13.7). Reaction products were plated out on EMM+thi. Cells were allowed to grow at 30˚C for approximately 4 days, at which point they were streaked out onto EMM+thi masterplates. Individual colonies were propagated in YEL to allow for removal of the plasmid and subsequently prevent further *in vivo Cre* recombinase action. Following growth for 2 days at 30°C, colonies were streaked out on YEA+5-FOA to allow for adequate selection for *ura4^{-}* (5-FOA^{R}; leu-) colonies and allowed to grow for 2 days, followed by production of YEA masterplates.

6.2.5.2 Colony PCR of UV and MMS resistant transformants

The 5-FOA^{R} colonies were further analysed by colony PCR (Sections 2.5.13.5 and 6.2.1.5, Appendix 15).

6.2.5.3 PCR and sequencing of the ORF of rad16^{MT}

The *rad16^{MT}* ORF and *lox* sites were amplified by PCR and sequenced in order to gauge integrity and incorporation of the relevant mutations. PCIA extraction of genomic DNA was carried out (Section 2.5.13.6), followed by PCR and Sanger sequencing (Section 6.2.1.6, Appendix 18).

6.2.6 Construction of *uve1Δ* strains

Strain crossing was carried out by Oliver Fleck. The available *uve1Δ* strain (J129) is a different mating type to the strains used in this study, we first created the correct mating type. To do this, we crossed J129 (*h^{-} uve1::LEU2 leu1-32 ura4-D18)* with 503 *(h^{+}leu1-32 ura4-D18 [ade6-704])*. Strains were mixed on sporation media (MEA) and incubated for two days at 30˚C. After this time, the cells were placed under a microscope to identify asci with 4 spores each; a sign that the crossing of strains had been successful. These were then treated with 30% ethanol to kill the cells; the spores survive. The successful spores were grown on MMA with appropriate

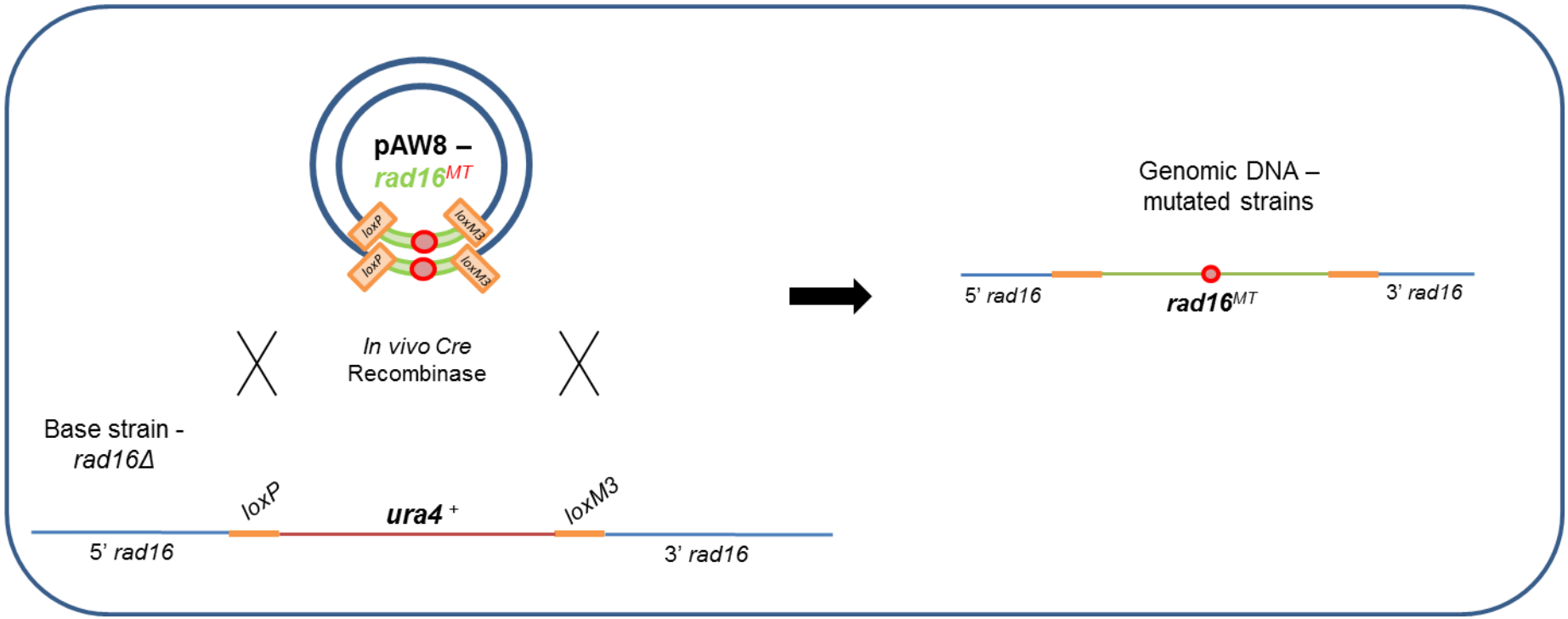


Figure 6.3 – Construction of rad16^{MT} strains. The various mutations of interest were introduced into the rad16Δ base strain by in vivo RMCE

supplements; in this case, uracil. After approximately 4 days growth, individual colonies were streaked onto plates of the same media and left to grow at 30˚C for a further 4 days to create the masterplate. The masterplate was replica plated onto YEA without adenine to cross out the redundant *ade6-704* (a nonsense mutation that results in adenine auxotrophy). Selection of colonies of white colour (strains with *ade6-704* are red in colour) was made. The successful cross was named OL2112 *(h^+uve1::LEU2 leu1-32 ura4-D18)*.

Cross 2 was carried out to combine OL2112 and the *rad16Δ* base strain *(smt-0 rad16::URA4 leu1-32 ura4-D18).* The mating types, h^+ and *smt-0*, will readily cross with one another. Strains were crossed in the same manner as described previously. Strains were grown on MMA. The produced strain was named *uve1Δ-rad16Δ (h^+rad16::URA4 uve1::LEU2 leu1-32 ura4-D18).*

Cross 3 was carried out to cross *uve1Δ* into the *$rad16^+$ and $rad16^{MT}$* strains *(smt-0 $rad16^+$leu1-32 ura4-D18 and smt-0 $rad16^{MT}$leu1-32 ura4-D18)*. Strains were grown on MMA+*ura.* The strains produced were named *uve1Δ-$rad16^+$* and *uve1Δr-$ad16^{MT}$* (Fig. 6.4).

6.2.7 Long term storage of bacterial colonies

Liquid cultures with successfully mutated plasmids were stored in equal volumes of 50% glycerol at -80˚C.

6.2.8 Long term storage of *S.pombe* cultures

Liquid cultures of the *rad16Δ* base strain and successfully mutated strains were stored in 60% glycerol at -80°C.

6.2.9 *In silico* analysis

Alignment of amino acids between species was carried out using Clustal Omega. Restriction enzymes were chosen based on recognition sites within DNA sequence of plasmids using the New England Biolab Cutter, v 2.0.

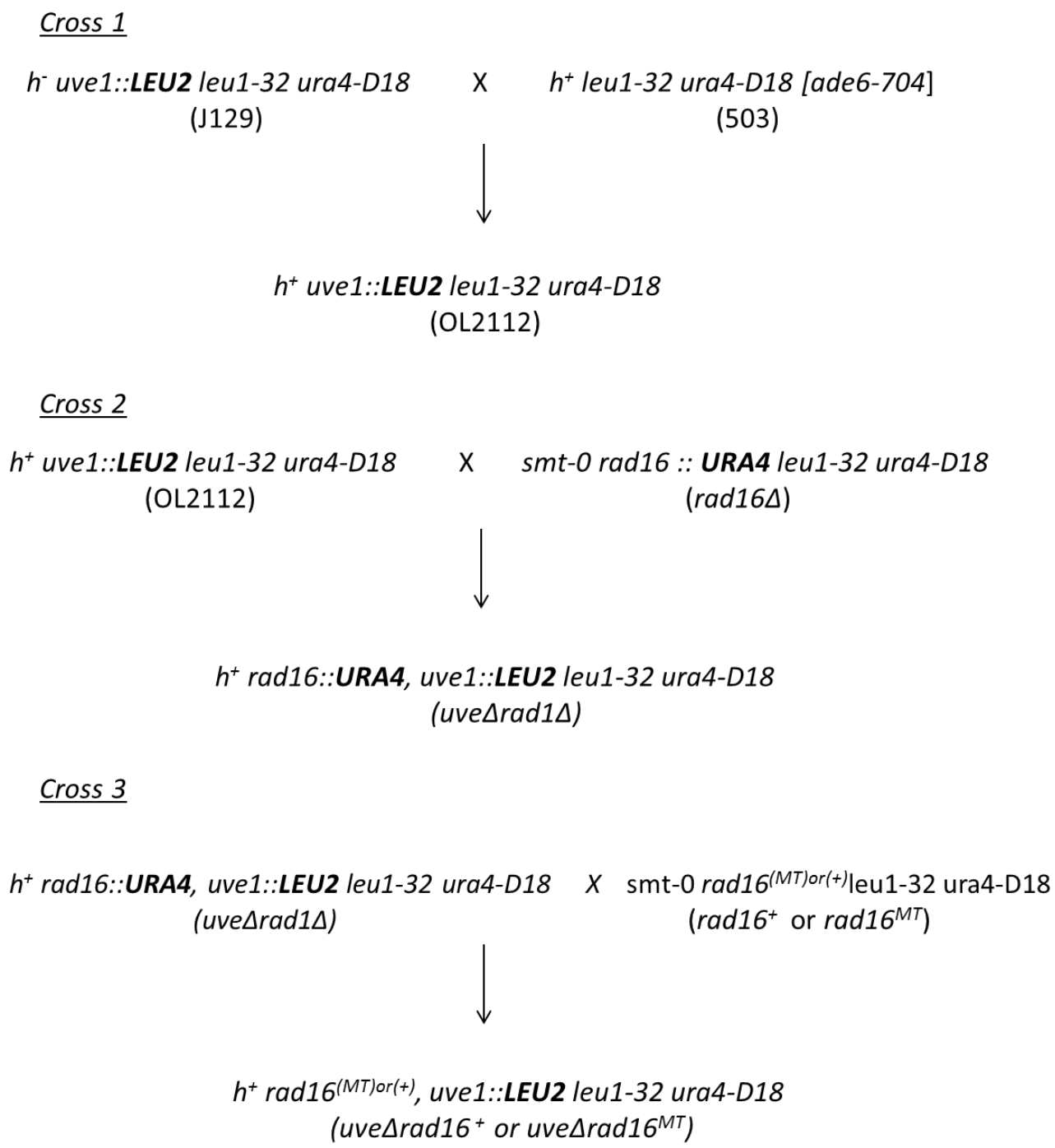


Figure 6.4 – Schematic of strain crosses carried out in order to knock out the uve1 gene in our previously constructed rad16Δ, rad16⁺ and *rad16^MT strains. The final strain is selectable by its ability to grow without leucine but not in the absence of uracil*

6.3 Results

6.3.1 Analysis of conservation between species

Percentage overall amino acid homology between XPF and the yeast homologs, Rad16 (*S.pombe*) and Rad1 (*S.cerevisiae*), was 36% and 31%, respectively. We observed conservation between all residues predicted to functionally affect the protein in XPF (Pro379, Arg576, Ser613 and Glu875) and Rad16 (Pro361, Arg548, Ser585 and Glu844, respectively). Additionally, the residue affected by a variant unlikely to affect function and not associated with PNAO, Arg415 (Rad16 - Arg399) was also conserved.

For XPF and Rad1 we observed conservation between the residues Pro379 and Ser613 only (Rad1 - Pro469 and Ser747, respectively; Fig. 6.5).

6.3.2 Construction of the *rad16Δ* base strain

The *rad16Δ* base strain was successfully constructed by incorporation of the PCR product *loxP-ura4*$^{+}$*-loxM3* at the *rad16* genomic locus (Fig. 6.7A). Recombination was made possible by the integration of 100bp regions at the 5' and 3' end of the PCR product which was homologous to the upstream and downstream regions at the *rad16* genomic locus. UV sensitivity enrichment allowed for identification of colonies likely to have *rad16* replacement by *ura4*$^{+}$ (Fig. 6.6). Following recognition of UV sensitive transformants, verification of replacement of *rad16* with *ura4*$^{+}$ was confirmed by colony PCR; the presence of a 537bp band indicated successful integration (Fig. 6.7B). Sanger sequencing on amplified products from extracted DNA confirmed that both *lox* sites were present and without mutation. However, 75% (3/4 colonies sequenced) contained mutations in the 50 base pairs immediately upstream of the gene.

Figure 6.5 – Alignment of residues implicated in PNAO in XPF (Homo sapiens), Rad16 (S.pombe) and Rad1 (S.cerevisiae). Amino acids highlighted in green are residues of interest.

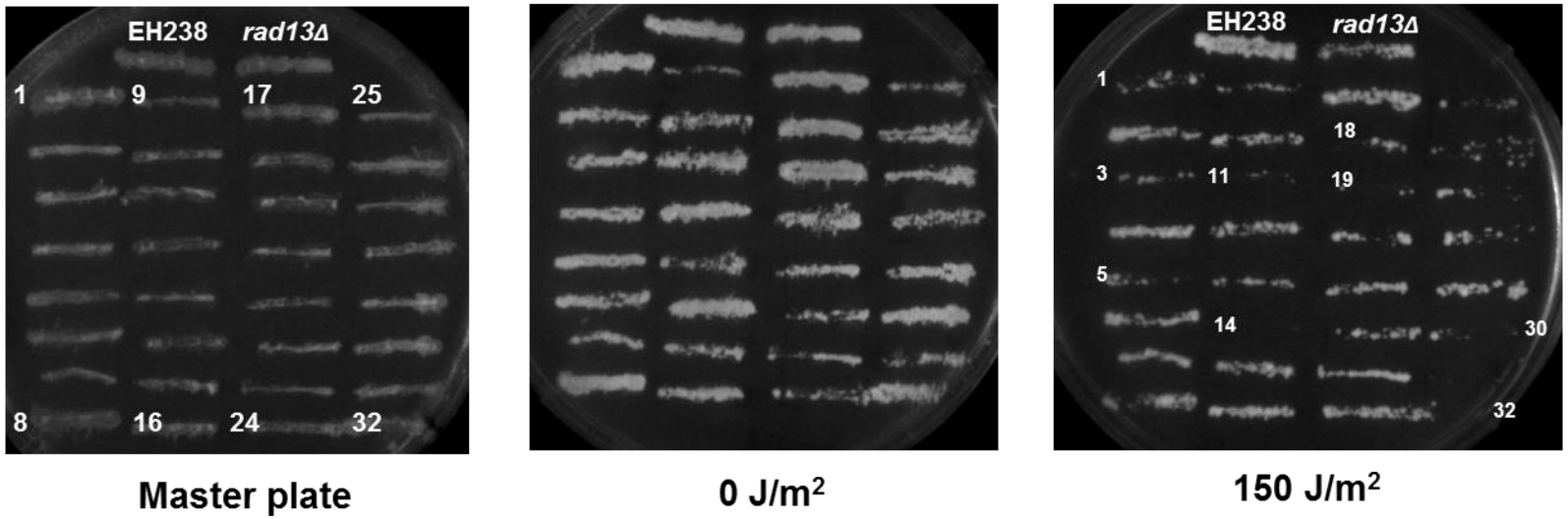


Figure 6.6 – UV enrichment for rad16Δ colonies. Wild type EH238 and a rad13Δ strain were also added to the plate for comparison as an NER proficient and deficient strain, respectively. Colonies identified as sensitive and analysed further are numbered on the plate treated with 150J/m^2.

6.3.3 Construction of *loxP-rad16*$^{+}$*-loxM3* and cloning into pAW8-*ccdB*

The production of the *loxP-rad16*$^{+}$*-loxM3* cassette was achieved by PCR and verified by the presence of a ~3kb band on an agarose gel (Fig. 6.7C). *In vitro Cre* recombinase was carried out with pAW8-*ccdB* and introduced into bacteria cells via electroporation. Subsequently, individual colonies were isolated and plasmid extraction was carried out using a Qiagen miniprep kit (Fig. 6.7D). A restriction digest was carried out using *Bam*HI to verify that the *rad16*$^{+}$ had been successful recombined into the plasmid (Fig. 6.7E). The ORF and *lox* sites were amplified and Sanger sequenced and integrity confirmed in all extracted plasmids.

6.3.4 Transformation of pAW8-*rad16*$^{+}$ into *rad16Δ* base strain, and genetic and phenotype testing

pAW8-*rad16*$^{+}$ was transformed into the *rad16Δ* base strain. Following growth of transformed cultures, enrichment by UV was carried out to confirm successful integration; those with restored *rad16* had restored ability to repair UV damage. We identified four transformants with restored DNA damage repair capacity and further analysed these by spot test treatment with 50-100J/m^2 of UV treatment and MMS treatment (Fig. 6.8). The transformants selected displayed a similar capacity for repair as EH238 wild type strain (Fig. 6.9). Following identification of insensitive colonies, colony PCR was carried out, with a band at 888 bps indicative that there had been successful incorporation at the *rad16* genomic DNA locus of *rad16*$^{+}$ from the plasmid. Additionally, genomic DNA from identified colonies was extracted, and Sanger sequencing of amplified products displayed that the entire ORF was intact in all colonies extracted.

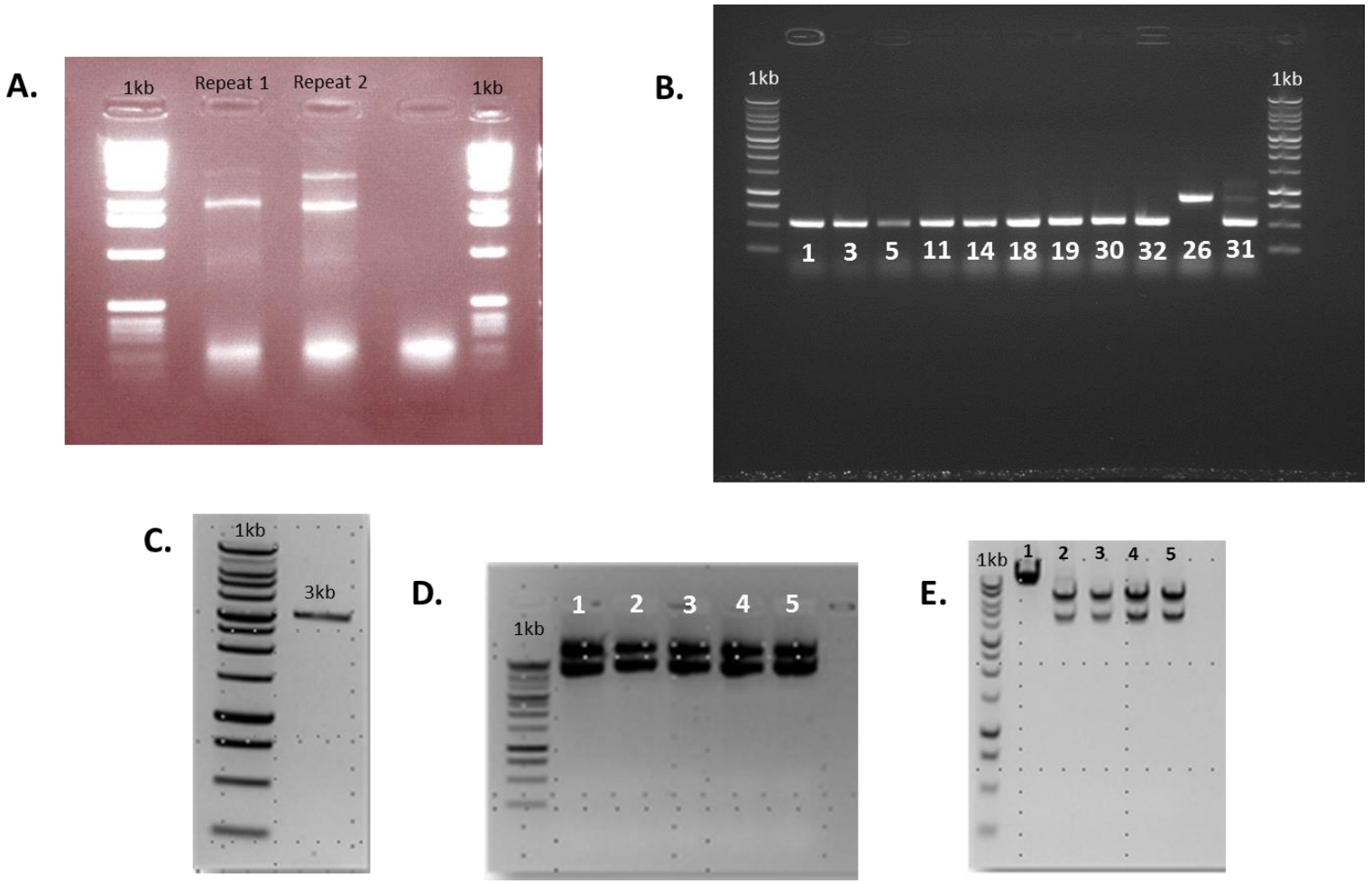


*Figure 6.7 A) Production of a loxP-ura4⁺-loxM3 PCR product from targeted amplification of pAW1 (in duplicate). B) Colony PCR of colonies transformed with loxP-ura4⁺-loxM3 chosen as a result of increased UV sensitivity. Colony 26 acted as a no recombination control (Fig. 6.6). C) Production of loxP-rad16⁺-loxM3 from genomic DNA. D) Extracted pAW8-rad16⁺ from five isolated colonies created by RMCE between pAW8-ccdB and loxP-rad16⁺-loxM3 E) Bam*HI *digestion of extracted pAW8-rad16⁺ from the 5 colonies isolated in D. Figures C-E were produced by Rebecca Williams.*

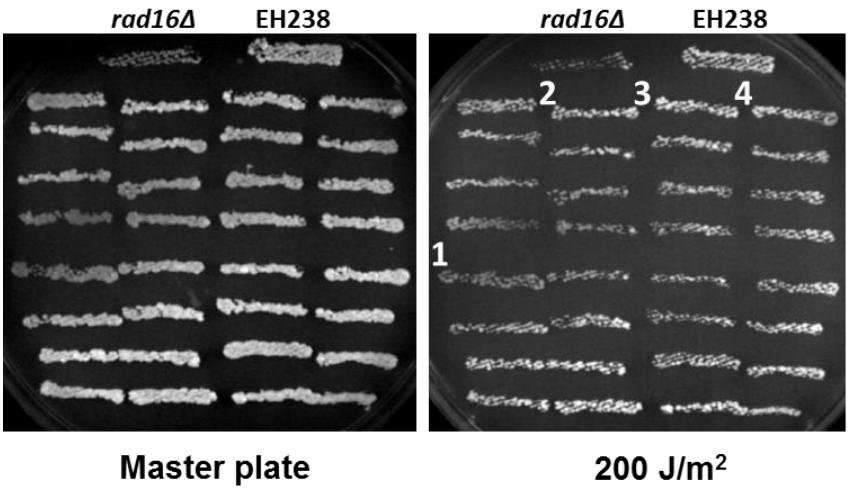


Figure 6.8- UV enrichment for rad16⁺ colonies. The wild type EH238 strain and rad16Δ were also added to the plate for comparison as NER proficient and deficient strains, respectively. Colonies identified as insensitive and analysed further are numbered.

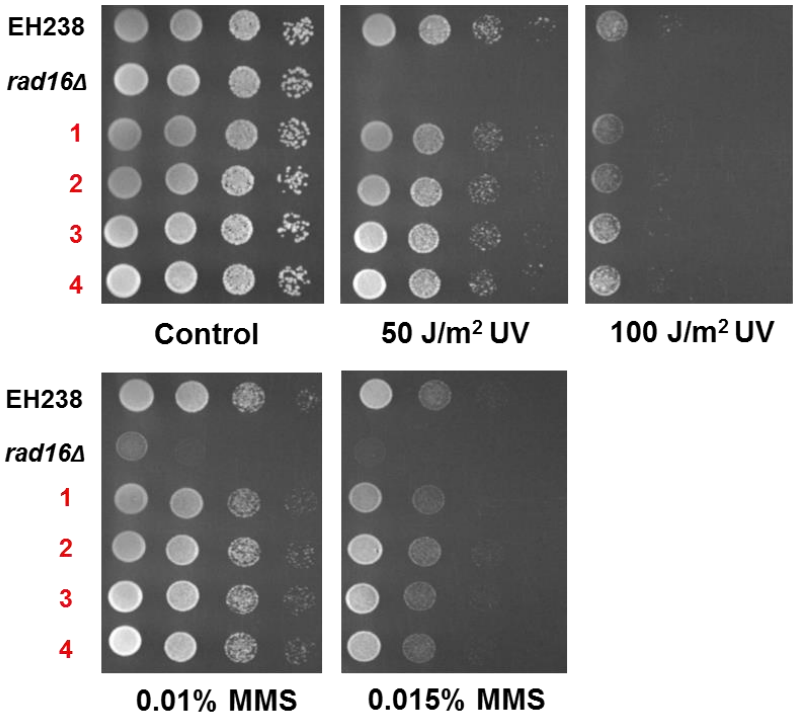


Figure 6.9 – Spot tests on four strains identified from sensitivity analysis to be insensitive to UV treatment and therefore more likely to have successful recombination of rad16⁺. Treatment included various doses of MMS and UV light.

6.3.5 SDM of pAW8-*rad16*$^{+}$

Following SDM on pAW8-*rad16*$^{+}$, plasmids were analysed by Sanger sequencing of amplified PCR products of the entire ORF and flanking *lox* sites to allow for Sanger sequencing, ensuring integrity and successful incorporation of mutations (Fig. 6.10).

In the first instance we had successful integration of Pro361Ser, Arg399Gln, Ser585X and Glu844Gly with no additional mutations in the ORF. However, we failed to integrate Arg548Thr. Following repeat of the SDM process with new primers, we identified colonies with the Arg548Thr with the mutation successfully incorporated, without additional mutations in the ORF.

6.3.6 Transformation of pAW8-*rad16*^{MT}

Mutated plasmids were transformed into *rad16Δ* base strain. Successful colonies were selected and colony PCR was carried out, with a band at 888 bps indicative that there had been successful incorporation at the *rad16* genomic DNA locus of *rad16*^{MT} from the plasmid. Additionally, genomic DNA was extracted, amplified and Sanger sequenced, showing that the entire ORF was intact in all colonies extracted, and all colonies contained their respective introduced mutation.

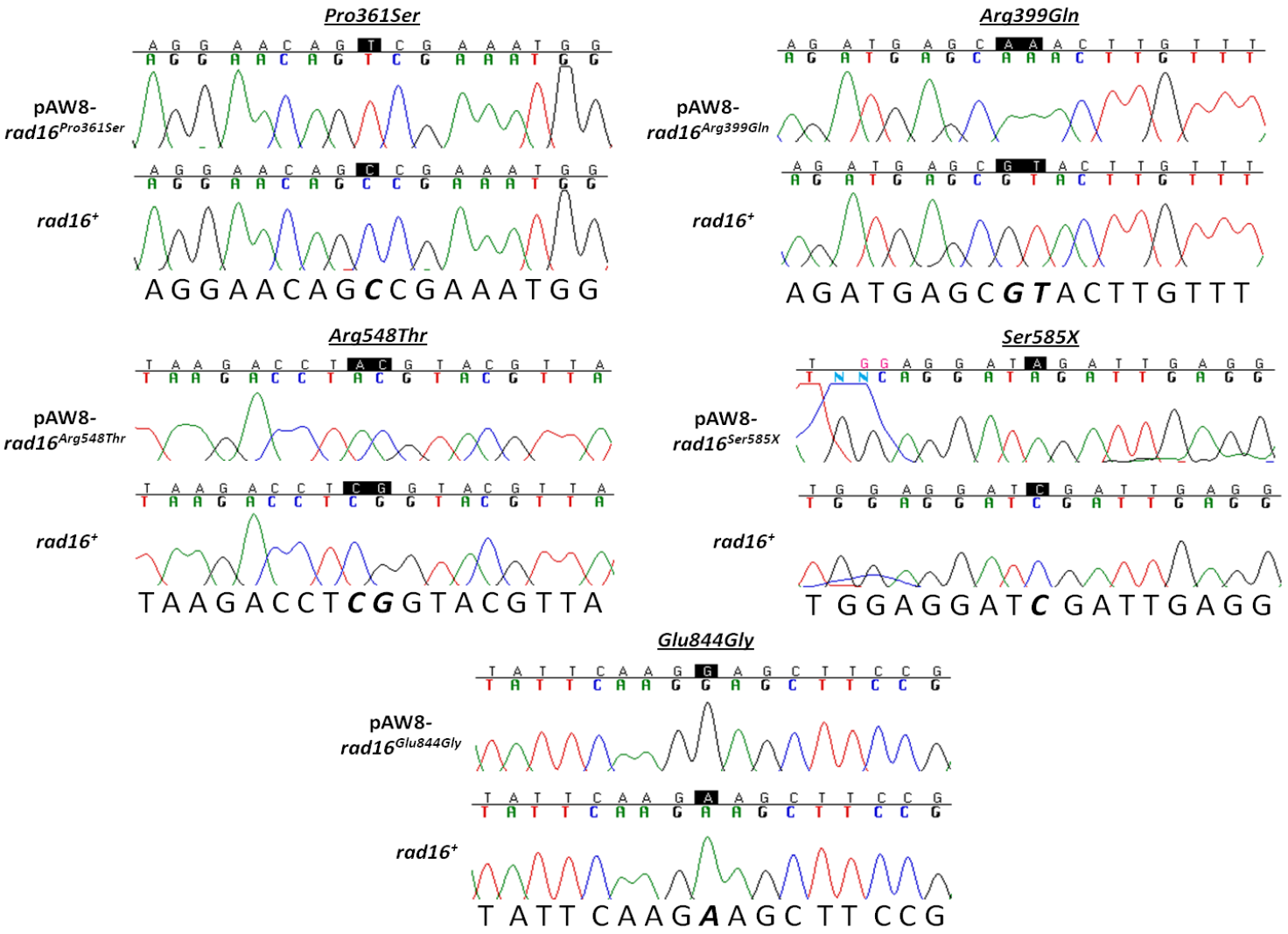


Figure 6.10 – Chromatogram data of successfully introduced mutations in the pAW8-rad16⁺ plasmid using SDM

6.4 Discussion

6.4.1 Species conservation

We attempted to create a model system in order to test effects of various residue changes identified in patients with PNAO in the DNA repair gene *ERCC4*. We observed conservation of all amino acids of interest between XPF and the *S.pombe* homolog, Rad16. Despite *S.cerevisiae* being the better studied model for NER, complete conservation of residues of interest in the XPF homolog, Rad1, was not observed; only two of the five variants were conserved. This meant that *S.pombe* was the ideal candidate for modelling the residue changes of interest.

6.4.2 RMCE

We used PCR based methods to create the *loxP-ura4*$^{+}$*-loxM3* cassette that was used to successfully create the *rad16Δ* strain by recombination into a wild type strain. By replacement of *rad16* with a selectable marker, *ura4*$^{+}$*,* we could easily identify successfully recombined strains.

The *Cre/lox* recombination system is a powerful tool and is used in the genetic manipulation of many organisms. The use of the site specific topoisomerase enzyme, *Cre* recombinase, allows for efficient and accurate cassette exchange between *lox* sites. By introduction of flanking *loxP* and *loxM3* sites at the *rad16* locus, we were provided with a useful tool for site specific and accurate recombination at the *rad16* genomic locus with various constructed cassettes. Differences in the spacer region of these two *lox* sites means that they recombine inefficiently with one another (Langer *et al.* 2002; Watson *et al*. 2008), preventing undesirable recombination events from occuring. Reinstatement of *rad16*$^{+}$ and introduction of mutations was easily achieved by introducing a plasmid with the desired cassette into the *rad16Δ* base strain. This has advantages over PCR based methods of recombination, where homologous integration at a locus of interest can be used for gene deletion or insertion of mutations of interest (Bähler *et al.* 1998). PCR based method can suffer from low recombination efficiency and require the homologous integration process to be repeated if a different gene modification is required (Krawchuk and Wahls, 1999). The incorporation of *lox* sites that recombine with any *lox* flanked cassette efficiently at the locus of interest can overcome this

problem (Watson *et al.* 2008) and requires only one homologous recombination event. However, the insertion of the *lox* sites into genomic DNA could be a disadvantage in itself as they could potentially affect gene expression and/or protein function.

6.4.3 SDM

By carrying out SDM on pAW8-*rad16*$^{+}$, we successfully produced a vector with our mutations of interest incorporated into the *rad16* gene. In the first instance we failed to introduce the mutation that results in Arg548Thr. We theorised this could be due to a high TA content at the 5' end of the original forward primer and therefore replaced this primer (Appendix 19). Following repeat of the SDM process with the new primers, we identified colonies with Arg548Thr successfully incorporated, without additional mutations in the ORF.

6.4.4 Analysis of functionality

In the construction of our model system we tested for functionality of the constructed strains at several points. Firstly, in the construction of the *rad16Δ* base strain, we considered that knockout of an essential NER gene would result in a UV sensitivity phenotype. Therefore we treated transformed cells with UV light to ascertain any heightened sensitivity. Similarly, following reintroduction of *rad16*$^{+}$, we considered that reinstatement of the functional gene should restore a cells ability to repair UV and MMS induced damage. This constituted an efficient screening method when selecting for constructed strains of particular phenotypes.

Secondly, following reintroduction of *rad16*$^{+}$, we attempted to ascertain that the insertion of incorporated *lox* sites and a $[His]_6$ tag did not affect expression of the gene, or function of the protein product. By treating with low dose UV, we confirmed that no effect on *rad16*$^{+}$ role in NER was seen. By comparison to an unaltered strain, we demonstrated that there was no phenotypic effect on the strains ability in the repair of DNA damage.

Additionally an essential gene, *mis18,* lies immediately upstream of *rad16.* Mis18 is involved in the control and regulation of centromeric chromatin and cell division by correct loading of the histone H3 variant, Cnp-1; an essential kinetochore (Hayashi *et al.* 2004; Williams *et al.* 2009). The region that falls between *rad16* and

mis18 is likely to be involved in the transcription of the gene. Parts of this region were involved in the homologous recombination of *loxP-ura4*$^{+}$*-loxM3* during the production of the *rad16* base strain. It is possible that mutations could have been incorporated into this PCR product during amplification, of particular relevance in the 100bps upstream of *rad16* required for homologous integration. Additionally, in the production of the *rad16Δ,* we incorporated *lox* sites at the 5' genomic region of *rad16.* Following recombination of the *loxP-ura4*$^{+}$*-loxM3*, the integrity of the *loxP* site and the region immediately upstream was checked by analysing sequence data. Of the colonies analysed, 75% contained a mutation in this region. We therefore used an error free strain as our base strain.

Additionally, it was important to ascertain that introduction of *lox* sites did not affect the strains function. By observing normal colonies on growing plates and survival of the *rad16Δ* base strains in normal physiological conditions in comparison to the unaltered EH238 strain, we were confident that there was no effect due to the incorporation of *loxP-ura4*$^{+}$*-loxM3* at the *rad16* genomic locus.

We ensured the functionality of the *lox* sites at several stages by confirming their integrity by amplifying and Sanger sequencing extracted DNA and plasmids. We carried this out on genomic DNA following recombination of *loxP-ura4*$^{+}$*-loxM3* and following the production of pAW8-*rad16*$^{+}$ and pAW8-*rad16*^{MT}. By verifying their integrity in this manner, we were satisfied that there would be no downstream problems when carrying out the various RMCE.

6.4.5 Knockout of alternative UV repair pathways

Although it is typically and predominantly the role of the NER pathway in the repair of DNA adducts that occurs as a result to UV light, *S.pombe* possesses a distinct, alternative pathway that has also been shown to participate in this type of DNA repair (McCready *et al.* 1993). The UV damaged DNA endonuclease (Uve1) – dependent excision repair pathway (UVER) has been shown to excise both 6,4PP and CPD's. It has also been shown to excise platinum adducts, although at a reduced efficiency (Avery *et al.* 1999). Uve1 activates the pathway by first nicking the DNA 5' to the adduct, at which point a BER like process will repair the damage much more rapidly than the NER pathway (Yonemasu *et al.* 1997). In order to rule out this

pathway in the repair of DNA damage following UV treatment, we knocked out *uve1* in all strains. This was carried out by crossing of strains with an *uve1Δ* with all *rad16* strains created for this project.

The production of *S.pombe* strains with mutations of interest incorporated into *rad16* provides a useful tool to study the potential of these variants to affect the proteins function. This can allow us to ascertain if the variants, originally identified in *ERCC4,* could affect the repair processes associated with various DNA damaging agents (Chapter 7).

Chapter Seven – Investigating the functional effects of variants introduced into *rad16*

7.1 Introduction

UV light causes direct DNA damage, producing CPD and 6-4,PP lesions. Both result in distortion of DNA, hindering transcription and replication, which can ultimately result in cell cycle arrest and apoptosis (Sinha and Hader, 2002). It is the role of NER to recognise, excise and repair the damaged strand (Section 1.3.3). *S.pombe* has an alternative UV repair system; UVER, which is governed by the endonuclease Uve1 (Section 6.4.5). In chapter 6, we successfully knocked out *uve1* in all constructed *rad16* strains in order to truly assay for the effect of the variants of interest in the repair of UV damage by the NER pathway.

MMS is an alkylating agent that adds methyl groups to nitrogen in purines. Despite the NER pathway being chiefly involved in the repair of bulky DNA adducts and not alkylated bases, mutations in NER genes in *S.pombe* have previously been shown to be sensitive to the actions of MMS (Kanamitsu and Ikeda, 2011). In *S.pombe,* this is believed to be due to the actions of the DNA repair sensor alkytransferase like protein (Atl1), responsible for the repair of alkylation of guanine residues (Pegg, 2000). Atl1 is responsible for recognising alkylation damage of guanine residues and shaping the lesion into what appears to be a bulky adduct, which subsequently recruits NER machinery to the area of damage (Pearson *et al.* 2006; Tubbs *et al*. 2009).

HU inhibits the production of new nucleotides by inhibiting ribonucleotide reductase. It therefore inhibits DNA synthesis and repair by depleting the dNTP pool. This results in replication fork stalling and cell-cycle arrest by inhibiting the build-up of nucleotides that normally occur during S phase (Koç *et al.* 2004; Petermann *et al.* 2010). Previously, NER deficient strains (*rad13Δ*) have been shown to be sensitive to HU (unpublished data; Rolf Kraehenbuehl, Bangor University).

Here, we sought to identify the functional consequences of variants introduced into *rad16* in Chapter 6. By administration of a combination of treatments, we tested

DNA repair pathways that XPF/Rad16 are associated with (Section 6.1). In addition to oxaliplatin, we treated with UV light, MMS and HU.

7.2 Materials and methods

7.2.1 Spot tests

7.2.1.1 Primary cultures

Primary cultures were established by inoculating colonies of strains isolated from a YEA plate in YEL at 30°C with shaking overnight.

7.2.1.2 Cell counts and dilutions

Cell counts of primary cultures were carried out and cells diluted in ten-fold serial dilutions in dH_2O to the appropriate concentrations (ranging from 1×10^4-10^7cells/ml). Either 5µl (*uve1Δ* strain UV spot tests plates) or 7µl (*uve1*$^+$ strain UV, MMS and HU spot test plates) spots of each concentration was pipetted directly onto the plate in ascending order. For all treatments, untreated spot tests with the same cell concentration were used as controls.

7.2.1.3 UV treatment

Once dry, plates were treated with a range of UV doses using a Stratalinker (10, 50 and 100J/M^2 for *uve1*$^+$ strains and 10, 20, 40 and 60J/M^2 for *uve1Δ* strains). Plates were stored at 30°C for four days, at which point photos of cell growth were taken. The experiment was repeated in triplicate for each concentration. In addition to the constructed *uve1Δ* strains, we also tested an *uve1Δ* strain with unaltered *rad16* (J129).

7.2.1.4 MMS and HU treatment

Spot tests with MMS (0.01, 0.015, 0.0175 and 0.02%) and HU (6 and 8mM) were carried out on plates with the desired concentration of reagents incorporated (Section 2.5.13.8). Plates were stored at 30°C for four days at which point photos of cell growth were taken. The experiment was repeated in triplicate for each concentration.

7.2.2 Acute treatments

7.2.2.1 Primary cultures

Primary cultures were established as described in section 7.2.1.1.

7.2.2.2 Oxaliplatin

Cell counts of primary cultures were taken and acute treatment was carried out by incubation of $1x10^7$ cells in YEL with and without 1mM of oxaliplatin for 18 hours at 30°C. In the non-treatment groups, an equivalent volume of DMSO was added in place of oxaliplatin.

Following incubation, cells were counted and ten-fold dilutions were made to a range of appropriate concentrations. Approximately 100µl of the appropriate concentration was plated out in duplicate onto YEA plates, spread sterilely and allowed to dry. For all untreated cells, $1x10^2$ cells were plated out; for treated *rad16Δ* and *rad16*^{Ser585X} 1×10^4 and 1×10^3 cells were plated out and, for all other strains, 1×10^3 and 1×10^2 cells were plated out. Plates were then stored at 30°C for four days, at which point counts of all growing cultures were made and percentage survival determined by comparison with untreated cells.

7.2.2.3 UV treatment of uve1Δ strains

Cell counts of primary cultures were carried out and ten-fold dilution of cells to a range of appropriate concentrations were made.

Approximately 100µl of each concentration was plated out in duplicate onto YEA plates, spread sterilely and allowed to dry. For all untreated cells, $1x10^2$ cells were plated out. Once dry, plates were treated with the appropriate dose of UV using a Stratalinker. A range of UV treatments at various cell concentrations were used dependant on the sensitivity phenotype associated with the strain (Table 7.1). Plates were then stored at 30°C for four days, at which point counts of all growing cultures were made and survival determined by comparison with untreated cells. In addition to all constructed strains, we also plated and treated J129.

7.2.2.4 Statistical analysis

Average survival data for acute exposure experiments were analysed with the ANOVA test using the statistical programme IBM SPSS statistics 20 following transformation using the arcsine function. Correction for multiple testing was carried out using the Bonferroni technique.

A.

UV Dose (J/m^2) / Strain	**Dose 1**	**Dose 2**
	5	**10**
uve1Δ-rad16Δ; *uve1Δ-rad16^{Ser585X}*	1x10^4 and 1x10^5	1x10^6 and 5x10^6

B.

UV Dose (J/m^2) / Strain	**Dose 1**	**Dose 2**
	20	**40**
J129 *uve1Δ-rad16$^+$*; *uve1Δ-rad16^{Pro361Ser};* *uve1Δ-rad16^{Arg399Gln};* *uve1Δ-rad16^{Arg548Thr};* *uve1Δ-rad16^{Glu844Gly}*	1x10^2	1x10^2 (*J129* only) and 1x10^3

Table 7.1 – Amount of cells plated for each strain with dose of UV treatment administered. Amount of cells plated were dependant on sensitivity phenotype previously demonstrated by the strains in UV treatment spot tests

7.3 Results

7.3.1 Spot test

7.3.1.1 UV treatment of UVER proficient strains

We observed at all doses of UV an increase in sensitivity of the *rad16*^{Ser585X} and *rad16Δ* strain. There were no apparent differences in sensitivity between all strains with nonsynonymous variants and *rad16*$^{+}$ (Fig. 7.1A).

7.3.1.2 UV treatment of UVER deficient strains

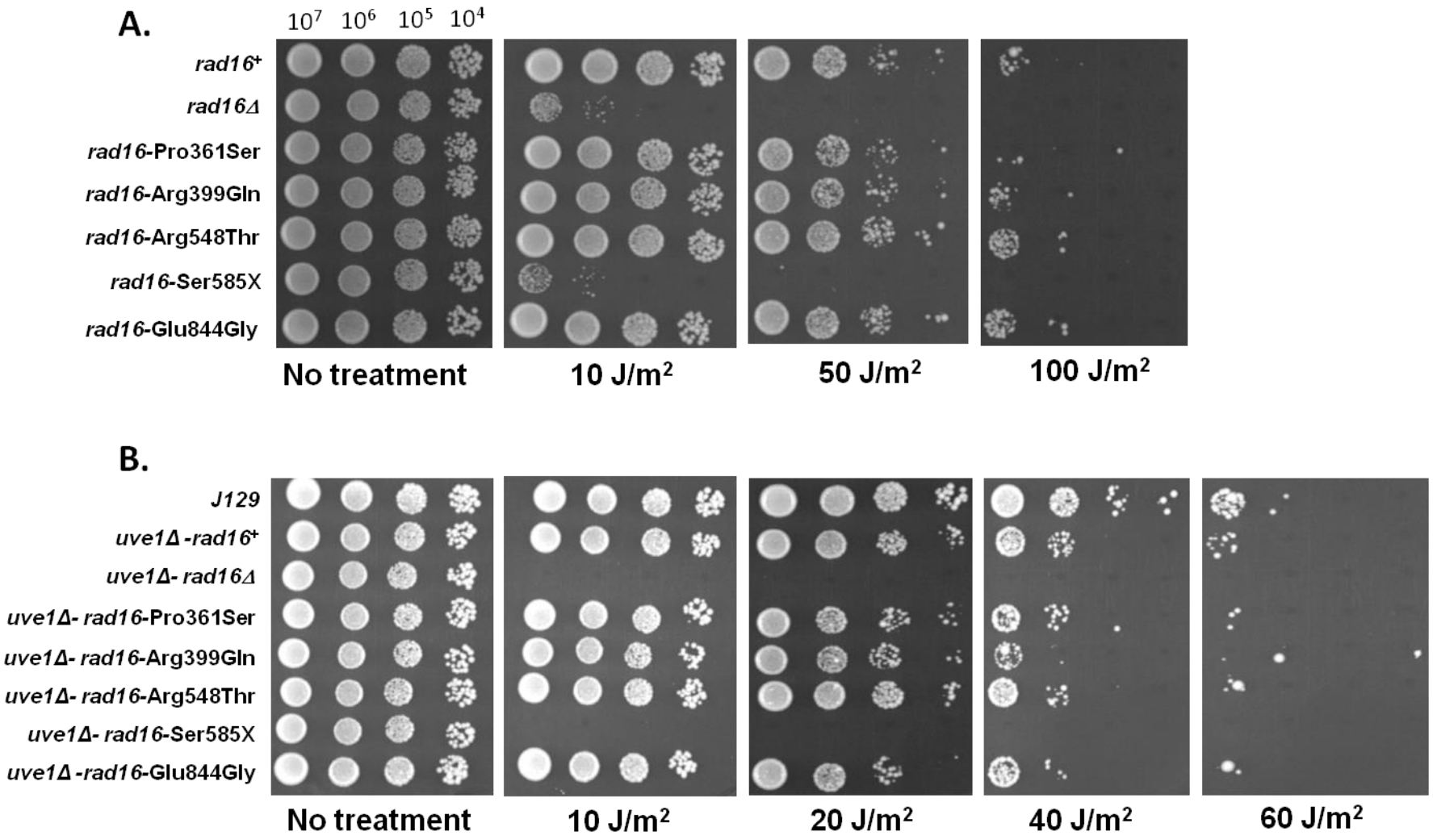
We observed heightened sensitivity of *uve1Δ-rad16*$^{+}$ in comparison to J129. Similar to the results of UV treatment of UVER proficient strains, we observed a heightened sensitivity of the *uve1Δ-rad16*^{Ser585X} and *uve1Δ-rad16Δ* strains at all doses of UV treatment. There were no apparent differences in sensitivity between strains with nonsynonymous variants introduced and *uve1Δ-rad16*$^{+}$ (Fig. 7.1B).

7.3.1.3 MMS treatment

We observed a heightened sensitivity at all concentrations of MMS for the *rad16*^{Ser585X} and *rad16Δ* strain. There were no apparent differences in sensitivity between all strains with nonsynonymous variants and *rad16*$^{+}$ (Fig. 7.1C)

7.3.1.4 HU treatment

We observed a slight sensitivity phenotype of *rad16*^{Ser585X} and *rad16Δ* following HU treatment. There were no apparent differences in sensitivity between all strains with nonsynonymous variants and *rad16*$^{+}$ (Fig. 7.1D).


A.
10^7 10^6 10^5 10^4
rad16+
rad16Δ
rad16-Pro361Ser
rad16-Arg399Gln
rad16-Arg548Thr
rad16-Ser585X
rad16-Glu844Gly
No treatment
10 J/m2
50 J/m2
100 J/m2
B.
J129
uve1Δ-rad16+
uve1Δ- rad16Δ
uve1Δ- rad16-Pro361Ser
uve1Δ- rad16-Arg399Gln
uve1Δ- rad16-Arg548Thr
uve1Δ- rad16-Ser585X
uve1Δ-rad16-Glu844Gly
No treatment
10 J/m2
20 J/m2
40 J/m2
60 J/m2

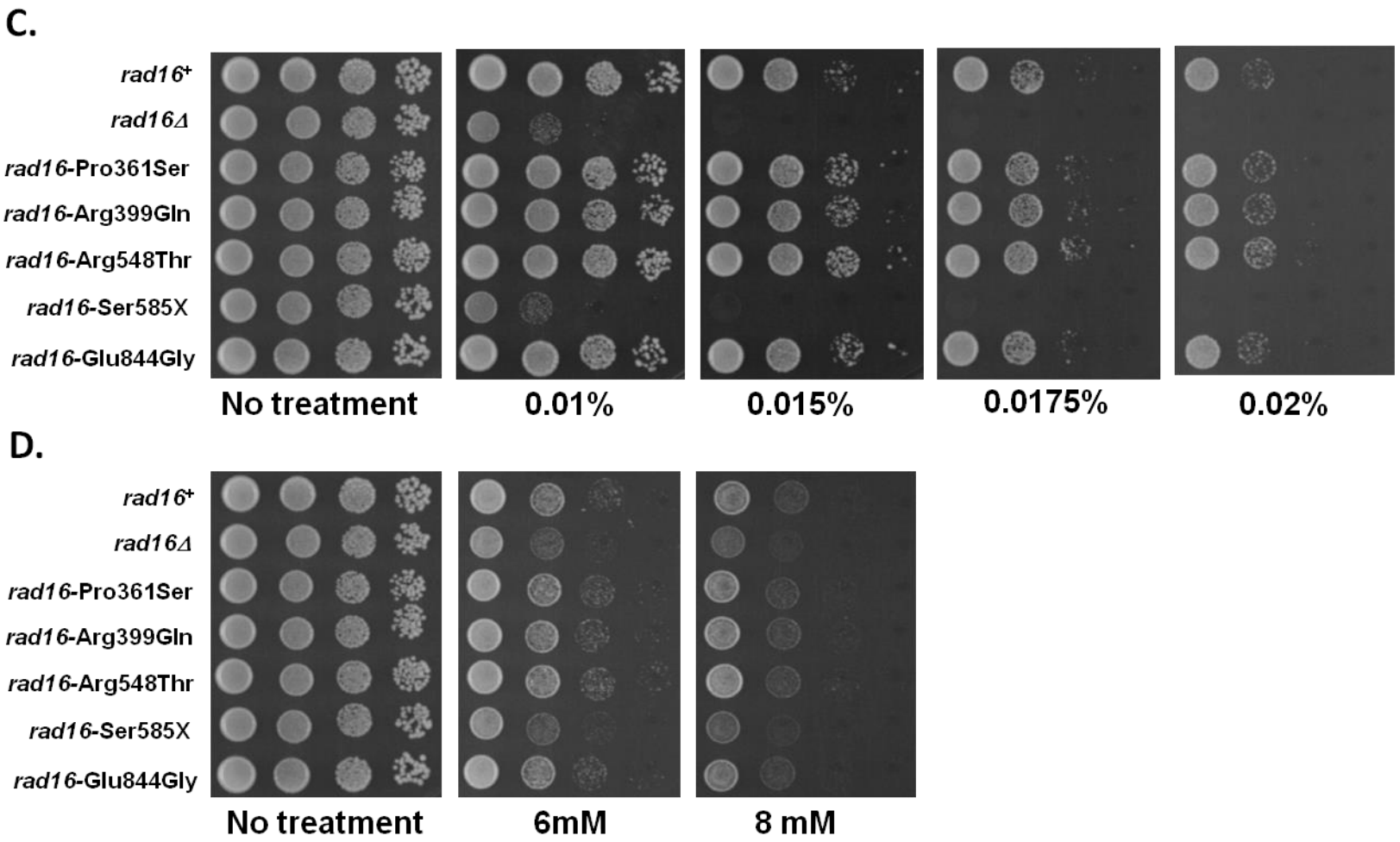


Figure 7.1 – Spot test results for; A. UV treatment of proficient UVER rad16 strains. B. UV treatment of J129 and uve1Δ-rad16 strains. C. MMS treatment of rad16 strains. D. HU treatment of rad16 strains. Concentration of cells plated on every plate is displayed on the 'no treatment' plate in Fig7.1A. No treatment plates in Fig7.1A, C and D are identical. Further repeats of each experiment are given in Appendices 26-29.

7.3.2 Acute treatments

Percentage survival for all strains following treatment was calculated in comparison to untreated plates (for oxaliplatin - Table 7.2; for UV – Table 7.3). Percentages were transformed using the arcsine function and comparisons with *rad16*$^{+}$ (oxaliplatin) or *uve*Δ-*rad16*$^{+}$ (UV) were carried out using ANOVA. Data was corrected for multiple testing using the Bonferroni technique.

7.3.2.1 Oxaliplatin treatments

We observed a statistically significant decrease in survival for the *rad16*^{Ser585X} strain only. However, when plotted separately, a consistent pattern of survival between strains with the introduced nonsynonymous variants was observed for experiments one, three and four (Fig. 7.2A, C and D). Due to variability of values between repeats, data from each experiment was normalised to *rad16*$^{+}$*,* subsequently averaged and plotted (Appendix 30; Fig. 7.2E). We were unable to apply statistics to the normalised data due to no standard deviations for *rad16*$^{+}$ (treated as 100%). In the normalised plot, the *rad16*^{Ser585X} and *rad16Δ* strains had less than 20% of the overall survival displayed by *rad16*$^{+}$*,* whilst all strains with nonsynonymous variants had less than 60% survival compared to *rad16*$^{+}$ (Fig. 7.2E)*.* Experiment two was excluded from the average for the normalised graph due to what appeared to be an outlying data point (*rad16*$^{+}$; Table 7.3; Fig. 7.2B)

7.3.2.2 UV treatments

For the UV treatment, at dose one, we observed a statistically significant decrease in survival for both the *uve1Δ-rad16Δ* and *uve1Δ-rad16*^{Ser585X}. This was not replicated at dose two. Due to variability of values between repeats, data was normalised to *uve1Δ-rad16*$^{+}$, subsequently averaged and plotted at both doses (Appendices 31-32; Fig. 7.3A-B). In both normalised plots, we observed high sensitivity with the *uve1Δ-rad16*^{Ser585X} and *uve1Δ-rad16Δ* strains (less than 1%) in comparison to *uve1Δ-rad16*$^{+}$*.* For the nonsynonymous variants, at dose one the survival was similar to that observed to *rad16*$^{+}$ (between 80-120%; Fig. 7.3A). This was mirrored at dose two for all strains except *uve1Δ-rad16*^{Glu844Gly} (Fig.7.3B). We observed heightened sensitivity of *uve1Δ-rad16*$^{+}$ compared to J129.

	Strains	Experiments				Average	SD	*P*
		1	2	3	4			
Survival of treated strains in comparison to untreated controls (%)	*rad16*$^{+}$	9.22	16.4	55.7	59.4	35.18 (41.44)	26.04 (28)	-
	rad16Δ	0.64	7	16.5	8.1	8.06 (8.41)	6.52 (7.93)	0.142 (1)
	rad16-Pro361Ser	5.09	28.2	28	19.7	20.25 (17.6)	10.85 (11.6)	1 (1)
	rad16-Arg399Gln	3.4	14.3	25.9	23.7	16.83 (17.67)	10.27 (12.4)	1 (1)
	rad16-Arg548Thr	3.63	19.7	16	25.6	16.23 (15.08)	9.29 (11.01)	0.914 (1)
	rad16-Ser585X	0.28	3.3	4.5	5.6	3.42 (3.46)	2.29 (2.80)	0.046 (1)
	rad16-Glu844Gly	6.64	17.1	32.4	28.7	21.21 (22.58)	11.7 (13.93)	1 (1)

Table 7.2 –Percentage survival of cells following treatment with 1mM of oxaliplatin. Averages and standard deviations (SD) were calculated for the four experiments. This was also calculated with experiment two excluded due to an outlying data point (rad16^{+}; in parenthesis). Data was transformed using the arcsine technique and ANOVA was used to assess for differences in survival for each strain in comparison to rad16^{+}. Bonferroni corrected P values (P) are given.

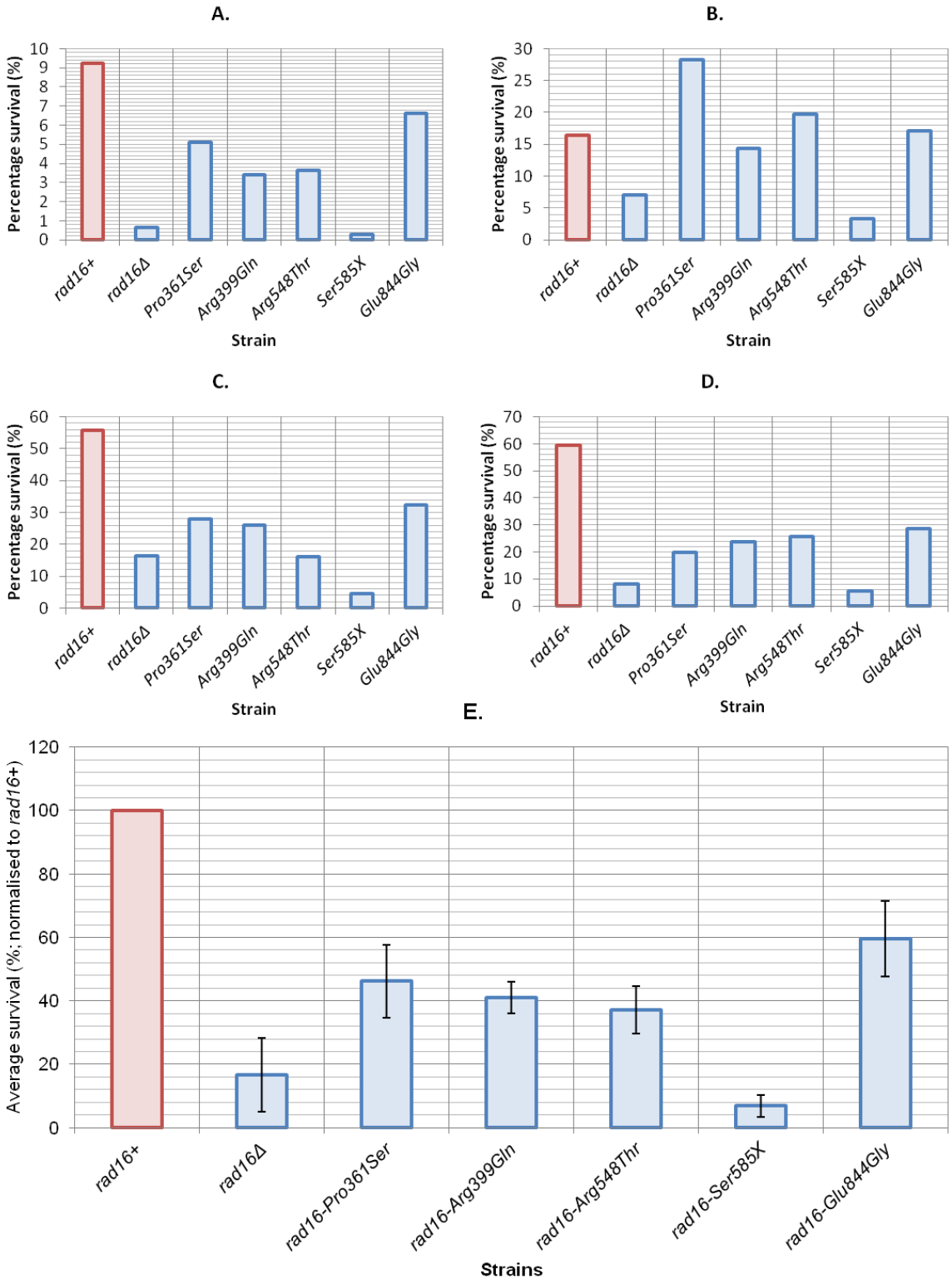


Figure 7.2 – Percentage survival (Table 7.2) for experiments 1 (A), 2 (B), 3 (C) and 4 (D). E. Average normalised percentage survival (for experiments 1, 3 and 4; Appendix 30) of constructed strains in comparison to untreated controls following oxaliplatin treatment. Experiment 2 was not included in the average due to an outlying data point. Standard deviations are displayed as vertical lines.

A.

	Strains	Experiment 1	2	3	Average	SD	*P*
Dose 1	*J129*	116	87.69	62.04	88.58	26.99	-
	uve1Δ-rad16^{+}	58.70	70.39	66.11	65.07	5.92	-
	uve1Δ-rad16Δ	0.0116	0.0003	0.0044	0.0054	0.0057	$4.2x10^{-4}$
	uve1Δ-rad16-Pro361Ser	74.58	55.26	61.08	63.64	9.91	1
	uve1Δ-rad16-Arg399Gln	56.30	86.96	70.48	71.25	15.34	1
	uve1Δ-rad16-Arg548Thr	82.96	82.17	54.35	73.16	16.3	1
	uve1Δ-rad16-Ser585X	0.0096	0.0011	0.0019	0.0042	0.0047	$4.2x10^{-4}$
	uve1Δ-rad16-Glu844Gly	46.59	58.42	58.94	54.65	6.98	1

B.

	Strains	Experiment 1	2	3	Average	SD	*P*
Dose 2	*J129*	60.32	38.96	36.75	45.34	13.02	-
	uve1Δ-rad16^{+}	8.70	4.41	7.78	6.96	2.26	-
	uve1Δ-rad16Δ	0.0114	0	0.0004	0.003933	0.006469	1
	uve1Δ-rad16-Pro361Ser	17.03	1.68	9.22	9.31	7.68	1
	uve1Δ-rad16-Arg399Gln	12.61	2.54	4.76	6.63	5.29	1
	uve1Δ-rad16-Arg548Thr	8.67	4.65	5.54	6.29	2.11	1
	uve1Δ-rad16-Ser585X	0.0025	0.0001	0.0002	0.000933	0.001358	1
	uve1Δ-rad16-Glu844Gly	7.61	1.79	1.69	3.70	3.39	1

Table 7.3 – Percentage survival of cells following treatment with A. Dose 1 B. Dose 2 of UV (Table 7.1). Averages and standard deviations (SD) were calculated for the three repeats. Data was transformed using the arcsine function and ANOVA was used to assess for differences in survival for each strain in comparison to uve1Δ-rad16^{+}*. Bonferroni corrected P values (P) are given.*

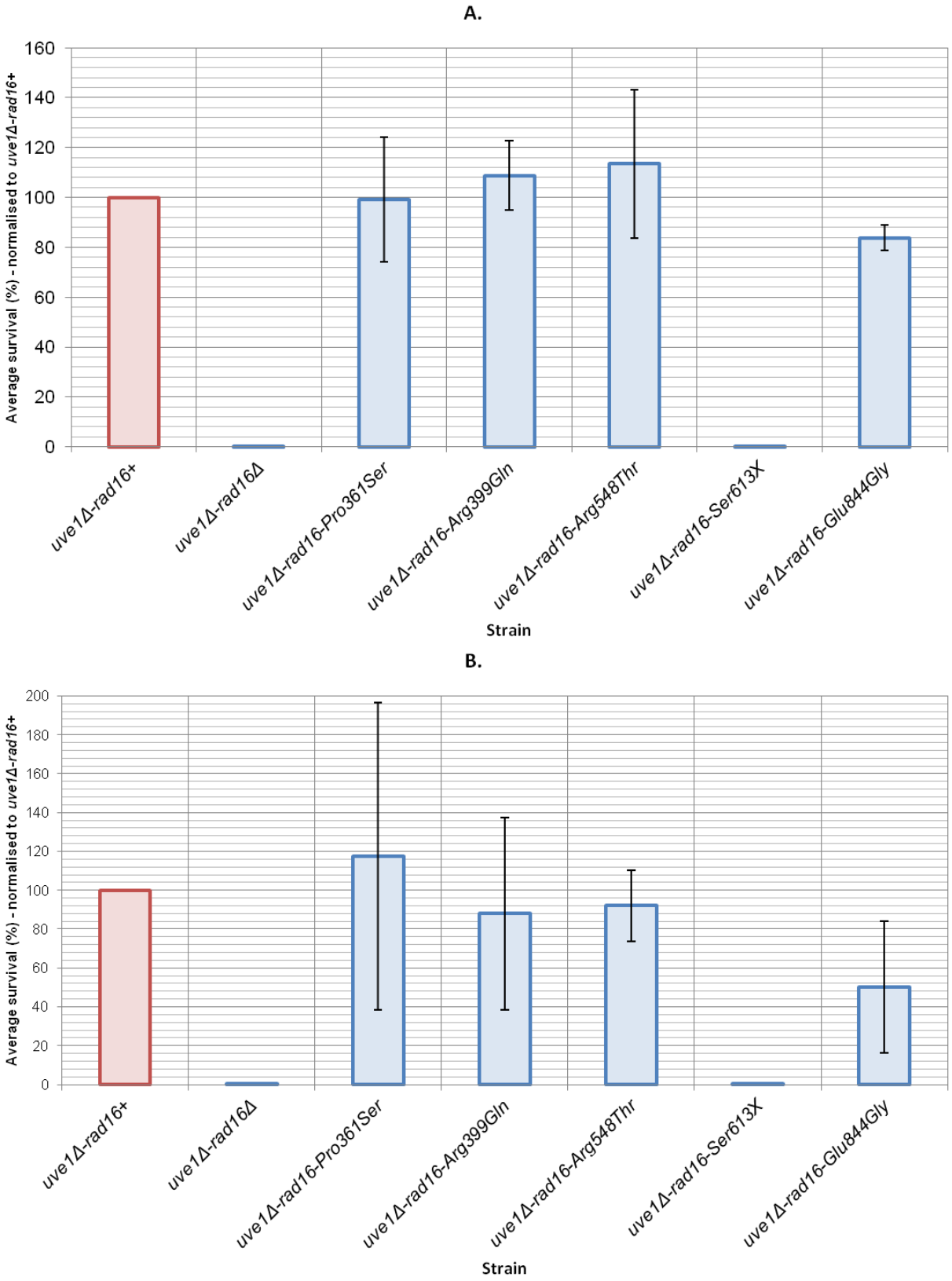


Figure7.3 – A. Average survival normalised to uve1Δ-rad16⁺ of constructed strains in comparison to untreated controls for dose 1 (Appendix 31). B. For dose 2 (Appendix 32). Standard deviations are displayed as vertical lines.

7.4 Discussion

Untreated control plates for all conditions showed that all constructed strains were viable and grew normally (in comparison to *rad16*$^{+}$ or *uve1Δ-rad16*$^{+}$) under normal physiological conditions indicating that there was no functional impact of incorporated *lox* sites and/or $[His]_6$ tag.

7.4.1 UV treatment of *uve1*$^{+}$ strains

We observed, even at low doses of UV (10J/m^2), extreme sensitivity of the *rad16Δ* strain and *rad16*^{Ser585X}. This suggests that the introduction of the truncation mutation, Ser585X, severely impedes the ability of *rad16* to act normally in the repair of UV induced damage, indicative of an NER deficiency.

All nonsynonymous variants introduced displayed a similar UV sensitivity as *rad16*$^{+}$ indicating no significant effect on *rad16* in the repair of UV damage. We saw no difference between the three predicted to be damaging variants associated with PNAO and Arg399Gln, not associated with PNAO.

7.4.2 UV treatment of *uve1Δ* strains

7.4.2.1 Spot tests

To compensate for the increased sensitivity as a result of the loss of the alternative UVER pathway, we treated *uve1Δ* strains with lower doses (10, 20, 40 and 60J/m^2) of UV than administered to the UVER proficient strains.

In addition to the *rad16* constructed strains, we also treated J129. We observed a slightly heightened sensitivity of the *uve1Δ-rad16*$^{+}$ strain in comparison to the J129 strain for all doses of UV, suggesting that there may be an effect of the introduced $[His]_6$ tag and/or *lox* sites. However, since all constructs of *rad16* created here contain the same genetic modification; this deems any associated phenotype comparable.

As displayed by the UVER proficient strains, we observed a heightened sensitivity of both the *uve1Δ-rad16Δ* and *uve1Δ-rad16*^{Ser585X}*.* Again, all strains with nonsynonymous variants displayed sensitivity similar to that observed with *uve1Δ-*

rad16^{+}. We saw no difference between the three predicted to be damaging variants associated with PNAO and Arg399Gln, not associated with PNAO.

7.4.2.2 Acute treatment

We carried out an acute UV treatment on all *uve1Δ* strains. At dose one, as observed with the UV treatment spot tests, a statistically significant increase in sensitivity of the *rad16Δ* and *rad16*^{Ser585X} was observed in comparison to *rad16*$^{+}$. This was despite compensation for increased sensitivity as a result of UVER deficiency by lowering the dose of UV administered. Similarly, there were no significant differences between all nonsynonymous variant in comparison to *rad16*$^{+}$.

We did not see a statistical difference in survival following treatment with dose 2, despite normalisation graphs displaying a clear decrease in survival of both the *rad16*^{Ser585X} and *rad16Δ.* Our inability to prove statistically that there was a difference at this dose could be due to large variability between the repeats.

As with the UV spot tests treatments, we observed a heightened sensitivity of the *uve1Δ-rad16*$^{+}$ strain in comparison to J129 at all doses.

7.4.3 MMS treatment

We observed, at all concentrations of MMS, sensitivity of the *rad16Δ* and *rad16*^{Ser585X} similar to the heightened sensitivity observed with UV treatment. Since the repair of MMS induced alkylation employs NER proteins indirectly through the actions of Atl1, this reinforces the concept of a deficiency of NER proteins in strains with the Ser585X variant.

As observed with the UV treatment, all introduced nonsynonymous variants displayed a similar sensitivity to MMS to that observed for *rad16*$^{+}$, indicating no observable effect of these variants on DNA repair. Similarly, we saw no difference between the three predicted to be damaging variant associated with PNAO and Arg399Gln, not associated with PNAO.

7.4.4 HU treatment

Although some sensitivity of *rad16Δ* and *rad16*^{Ser585X} was observed with HU treatment in comparison to *rad16*$^{+}$*,* the degree of sensitivity observed was not as

severe as that seen in MMS and UV treatments. Since HU depletes dNTPs, it predominantly stalls replication forks which ultimately results in DSB when these forks collapse following prolonged or excessive dosing at the site (Petermann *et al.* 2010). Previous research suggests that Rad16 and the human XPF-ERCC1 complex could have a role in the repair of such DSBs (Sargent *et al.* 2000; Prudden *et al.* 2003; Ahmad *et al*. 2008; Al-Minawi *et al.* 2009; Kikuchi *et al.* 2013). The degree of sensitivity following HU treatment is mirrored between the *rad16Δ* and the *rad16*^{Ser585X} strain, suggesting that *rad16* has some role in the repair of HU specific DNA damage and that the truncation strain is unable to function adequately in the repair of such damage.

7.4.5 Oxaliplatin treatment

We were unable to mimic spot tests treatments for oxaliplatin due to low stock concentration of the drug. Oxaliplatin is only soluble in DMSO at a maximum concentration of 40mM, meaning that to reach concentrations in a 25ml YEL plate required we would have needed to add a high volume, reducing the amount of YEL and possibly affecting the ability of strains to grow normally.

In order to assay for the effects of oxaliplatin, we carried out an acute oxaliplatin treatment. Statistically, we observed a heightened sensitivity of *rad16*^{Ser585X} only. We thought that variability between repeats could be influencing the statistics and our ability to observe statistical significance between the other strains. Therefore, we normalised data to *rad16*$^{+}$ to account for variability between repeats. Since the percentage survival of *rad16*$^{+}$ in experiment two appeared to be an outlying data point, we removed the experiment from the average normalisation. Following normalisation, we observed a clear decrease in survival for *rad16Δ* also. In the normalised plot, unlike the acute UV and various spot test treatments, a heightened degree of sensitivity was also observed with all nonsynonymous variants introduced in comparison to *rad16*$^{+}$.

In summary, we have shown that the introduction of Ser585X (Ser613X) into *rad16* sensitises strains to MMS, HU and UV treatments to the same degree as that seen in the *rad16Δ* strain. Sensitivity was also observed following treatment with

oxaliplatin. This suggests that this truncating mutation is as detrimental to the role of the protein in various DNA repair pathways as deletions of the gene in its entirety. Although not statistically significant, we also observed an oxaliplatin specific decrease in survival for all strains with nonsynonymous variants (discussed further in Chapter 8).

Chapter Eight – General discussion

8.1 CRC predisposition

Genetics has been shown to have an important role in CRC. Highly penetrant mutations have been shown to result in multiple hereditary CRC syndromes, whilst a variety of low penetrance alleles are believed to act in concert with one another to significantly alter an individual's risk. Our training phase cohort has been used in the identification and validation of low and moderate penetrance risk alleles. Although not substantial enough to be presented in this thesis, I helped identify novel low penetrance alleles by GWAS meta-analysis (Appendix 33) and a moderate risk allele in the DNA repair gene *OGG1* (Appendix 34).

Better understanding of the genetics of CRC has led to the realisation that there are multiple proteins in particular pathways that are implicated in CRC risk. For example, multiple high and low penetrance alleles in genes that encode proteins involved in the TGFβ signalling pathway have been shown to be important in inherited forms of the disease. These include high penetrance mutations in *SMAD4* and *BMPR1A,* which are known to predispose to JPS (Section 1.2.1.4.2). Similarly, overexpression of *GREM1* as a result of an upstream 40kb duplication has recently been shown to cause HMPS (Section 1.2.1.4.4). Interestingly, two low penetrance variants (rs16969681 and rs11632715) that fall within this region have been shown, following analysis of a GWAS risk locus, to be associated with disease risk (Section 3.4.2.2). Similarly, a low penetrance risk variant, rs4939827, again identified by GWAS, is associated with over-expression of SMAD7 (Section 3.4.2.1). In addition, GWAS has also uncovered CRC risk loci associated with *RHPN2* (19q13.1, [rs10411210]), *BMP2* (20p12.3, [rs961253]) and *BMP4* (14q22.2, [rs4444235]; Section 1.2.1.2), all part of the TGFβ signalling pathway.

In addition to hereditary syndromes, the TGFβ pathway is important in CRC tumourigenesis. Complete loss of chromosome 18q is seen in approximately 75% of colorectal adenocarcinomas. This region is known to contain both *SMAD2* and *SMAD4* (Mehlen and Fearon, 2004). Additionally, *TGFβR2* contains a microsatellite

repeat that is prone to MSI in MMR deficient cells, such as that seen in HNPCC or in approximately 12% of sporadic cancers (Lu *et al.* 1995; Fig. 8.1).

As well as the TGFβ pathway, various DNA repair pathways are implicated in the genetics of CRC. Recently the identification of high penetrance mutations in *POLE* and *POLD1*, both important in DNA synthesis following excision of damage in multiple DNA repair pathways, were shown to predispose to multiple CRA and CRC (Section 1.7.2.2).

High penetrance mutations in *MUTYH*, involved in the excision of adenine bases erroneously incorporated opposite 8-oxo-G that has formed due to oxidative damage, cause MAP (Section 1.2.1.2). Recently a variant in the oxidative repair protein *OGG1,* which encodes a protein which has roles in the direct repair of 8-oxo-G, has also been shown to act as a low penetrance risk allele for CRC (Smith *et al.* 2013). Other cancer types with an inherited component have been shown to be caused by both low and high penetrance mutations in genes involved in particular DNA repair pathways. For example, hereditary breast cancer is commonly a result of high penetrance mutations in *BRCA1* and *BRCA2* (Section 1.3.4.3.1). However, low penetrance inactivating mutations in *BRIP1,* which encodes a DNA helicase with known interactions with BRCA1 in HR and ICL repair, have also been shown to predispose to the disease (Seal *et al.* 2006).

HNPCC is due to mutations in multiple genes in the MMR pathway (Section 1.2.1.3). The MMR pathway is also important in CRC tumourigenesis. Up to 12% of sporadic tumours exhibit signs of MMR deficiency. This is most commonly as a result of inactivation of the MMR system via silencing of *MLH1* via biallelic hypermethylation of the CpG islands in the promoter region, with tumours showing such methylation patterns being known as having a CpG island methylator phenotype (Kane *et al.* 1997; Toyota M *et al*, 1999). As a functional part of all three hMutL complexes, *MLH1* is critical for functional MMR.

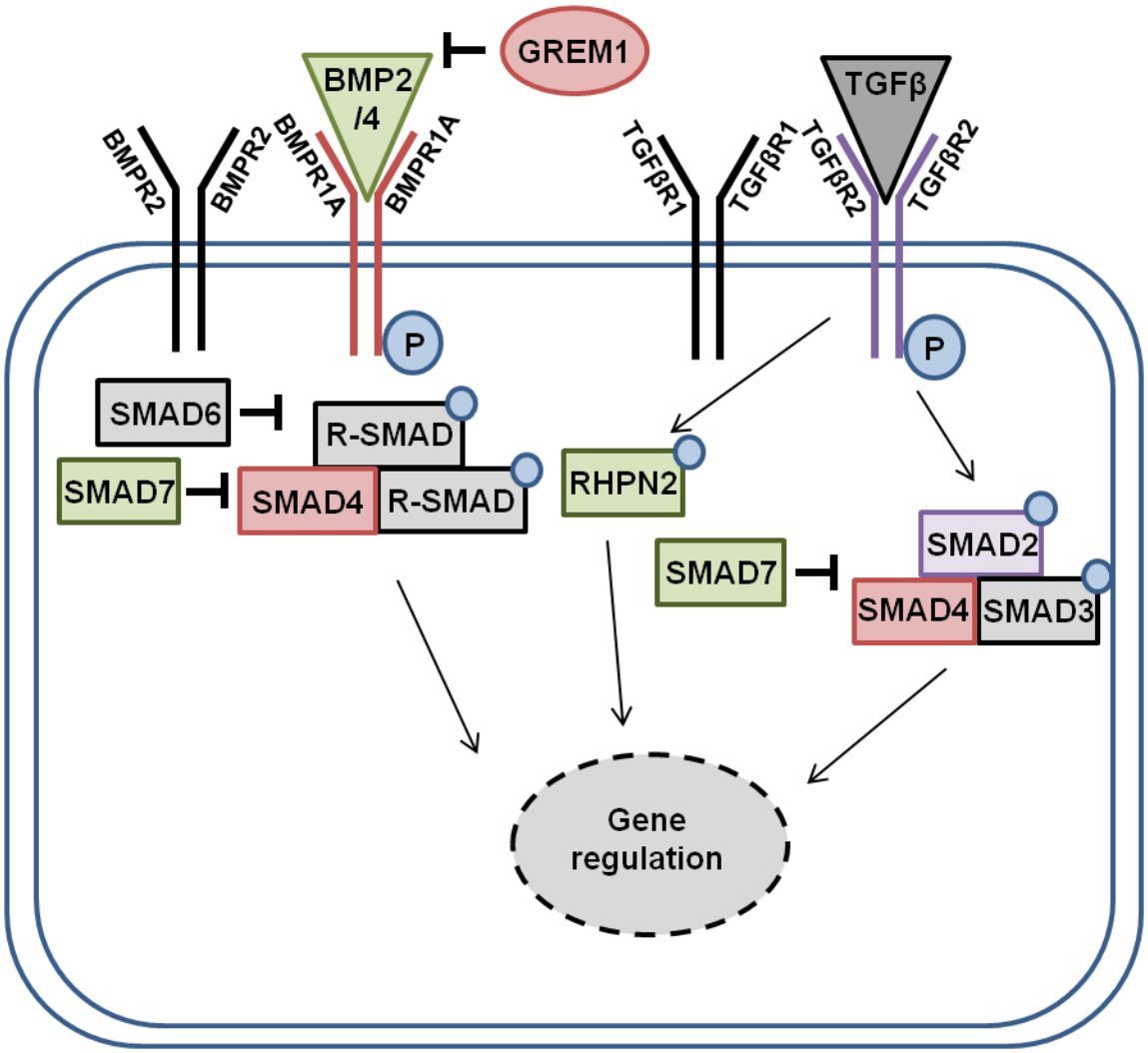


Figure 8.1 – The TGFβ signalling cascade. In TGFβ signalling, following binding of the TGFβ ligand to TGFβR1 and TGFβR2, there is receptor activation by phosphorylation (P). The activated intracellular domain of the proteins phosphorylate SMAD2 and SMAD3 and, following recruitment of SMAD4, relocate to the nucleus in order to regulate gene expression. Also, activated TGFβR activates RHPN2. SMAD7 is involved in the negative regulation of the SMAD2/SMAD3 complex. In the BMP signalling pathway, following binding of either BMP2 or BMP4 to BMPR1A or BMPR2, a dimeric complex involving combinations of the regulatory SMADS (R-SMAD; SMAD1, SMAD5 and SMAD8) is phosphorylated and activated. This complex recruits SMAD4 and relocates to the nucleus to regulate gene expression. This process is negatively regulated by SMAD6 and SMAD7, whilst GREM1 regulates BMP2/4 binding. Shown in green are proteins encoded by genes implicated by GWAS as associated with CRC risk; purple are involved in CRC tumourigenesis; red are involved in inherited forms of CRC. NOTE; GREM1 has been shown to have both high penetrance and low penetrance disease alleles. SMAD4 is associated with a hereditary CRC syndrome and is also involved in tumourigenesis. Adapted from Tenesa and Dunlop, 2009.

Here, we used our training phase cohort to help identify novel disease alleles associated with CRC. Given the importance of mutations in genes from DNA repair pathways in hereditary cancer syndromes, including various hereditary CRC disorders, we took a candidate gene approach focusing on DNA repair pathways. Despite our initial findings in the training phase cohort, we were unable to validate the apparent association between *RAD1*^{Glu281Gly} and aCRC. We suggest that this could be due to the validation phase study being underpowered by the current sample size, and more aCRC and controls in the validation phase could be beneficial in ascertaining the effect of the allele on risk.

RAD1 is a component of the RAD9-HUS1-RAD1 (9-1-1) complex (Burtelow *et al.* 2000), which has roles in translesion synthesis, DSB repair and checkpoint activation in response to DNA damage (Parrilla-Castellar *et al*. 2004; Pichierri *et al.* 2012). Additionally, roles in BER have also been proposed following the observation that the complex interacts with the DNA glycosylases MUTYH, TDG and NEIL1 (Shi *et al.* 2006; Guan *et al.* 2007a; Guan *et al.* 2007b), as well as interacting and regulating FEN1 (Friedrich-Heineken *et al.* 2005) and LIG1 (Song *et al*. 2009). Although there is no previous link to RAD1 in cancer predisposition, knockout in mice leads to an elevated rate on skin cancers (Han *et al.* 2010). With regards to CRC, the interaction with MUTYH is of particular interest, given the links between MUTYH and hereditary CRC. Inefficient binding of MUTYH to the 9-1-1 complex, as a result of nonsynonymous variants in *MUTYH*, has been shown to lead to a repair deficiency phenotype (Turco *et al.* 2013). Although the native *RAD1* allele was mostly conserved throughout evolution and the amino acid change was predicted to be detrimental to protein function, it remains unclear how *RAD1*^{Glu281Gly} could affect protein function. We postulate that potential disruption of protein-protein interactions could be key in the contribution of the *RAD1* variant to the development of CRC. For example, the variant allele could disrupt complex formation by affecting the known C terminal binding of RAD1 to the N terminal of RAD9 (Doré *et al.* 2009). Alternatively, the variant could affect binding and localisation of the various other proteins that the complex has been associated with.

8.2 NGS of patients with adverse drug reactions

Severe chemotherapeutic side effects can lead to a cessation of treatment or dose reductions which could be detrimental in the treatment of cancer. It is well known that there is variability between individuals in the severity of adverse effects experienced with the same treatments, for which genetics has the potential to play a role (Eichler *et al.* 2011). Improved understanding of underlying genetics could further understanding of cellular processes involved in drug reactions. This has the potential to allow clinicians to make better, informed choices when treating patients to improve the chance of treatment success whilst reducing debilitating side effects. For example, in the treatment of CRC, the FDA recommends genotyping for polymorphisms associated with *UGT1A1* before administration of irinotecan. This allows for modifications of the dose administered, reducing the risk of severe diarrhoea and neutropenia in patients with these detrimental polymorphisms. Similarly, genotyping before treatment with the fluoropyrimidines for several variants associated with *DYPD* has been recommended to reduce severe side effects associated with polymorphisms that affect the rate of drug metabolism.

Previously, candidate gene studies and GWAS have proved useful in the study of the pharmacogenetics of adverse events associated with many different drugs. However, candidate gene studies often fail to account for the various different mechanisms that are involved in toxic responses and GWAS of pharmacogenetics struggle to obtain sample sizes sufficient to validate an apparent association due to the often rarity of adverse events. NGS could prove to be useful in the pharmacogenetic study of adverse drug reactions due to its ability to identify rare variants, whilst sufficiently considering large proportions of the genome (Daly, 2010). Several studies have used NGS to study chemotherapy response and resistance. However, there are currently no published NGS studies investigating severe toxic responses

Here, we used exome resequencing to uncover alleles associated with PNAO. One of the main limitations of exome resequencing is that by directly targeting the protein coding region a vast proportion of the genome is not analysed; up to 99% of the human genome is considered ‘non-coding’. A recent GWAS study of chronic PNAO uncovered nine variants in eight genes that appeared to be associated with

risk (Won *et al.* 2012). All of these variants were intronic. Additionally, a variant intronic to *SCN4A* has previously been associated with the severity and rate of onset of chronic PNAO (Argyriou *et al.* 2013). Due to high cost and time constraints, WGS is not as accessible as WES. It is anticipated that as the 'third generation' sequencing technology improves and competition between manufactures increase, there will be reductions in cost and time taken to acquire results. This will allow for WGS to be used more frequently in the discovery of alleles associated with particular phenotypes.

8.3 PNAO

PNAO remains a debilitating side effect in the treatment of aCRC for which, despite improved understanding of the underlying mechanisms of both the acute and chronic forms, there is currently no treatment to alleviate symptoms. In addition to impacting on cancer treatment due to dose modifications, it can affect the overall health and well-being of patients undergoing treatment (Tofthagen *et al.* 2013).

8.3.1 Exome resequencing of patients with PNAO

We initially identified ten patients from the COIN trial and its translational study with extreme and dose limiting PNAO, In order to efficiently assess the exome resequencing data generated, we took two strategies to identify potential casual alleles. Firstly, we considered genes involved in the pharmacokinetics and cellular response to platinum drugs. Previously, variants in *GSTP1, AGXT* and *ERCC1* have been shown to lead to an altered degree of PNAO, suggesting that altered cellular levels and/or effects of oxaliplatin can alter the sensitivity to treatment. We discovered a stop gain in the DNA repair gene, *ERCC4,* involved in the repair of DNA adducts such as that seen in oxaliplatin treatment. A decreased ability to repair DNA adducts could lead to an accumulation of lesions that has the potential to increase the rate of apoptosis.

Secondly, we investigated genes involved in the neuronal function and/or peripheral neuropathy. Although in this thesis we did not find any association

between genes involved in neuronal function and/or neuropathy the recent finding that a nonsynonymous variant associated with the voltage gated sodium channel, *SCN10A*, was associated with an increased incidence of acute PNAO validates the analysis approach (Argyriou *et al.* 2013). Despite being beyond the scope of this project, a complete pathway analysis (much the same as carried out in this project for oxaliplatin) of genes involved in peripheral nerve function and/or neuropathy could be beneficial when considering future studies of PNAO.

An alternative strategy not considered here includes the analysis of the data from the ten patients with PNAO without a prior hypothesis implied. This strategy would avoid any selection bias that comes with focusing on specific pathways such as that which can occur in candidate gene studies. To achieve this, filtering for all novel or low frequency stop gain or frameshifting indels in genes that are seen mutated in two or more of the ten patients could highlight potential genes involved in PNAO pathogenesis.

The avoidance of false positive and false negative results as a result of coverage issues is crucial in the analysis of exome resequencing data. In this study, we failed to validate a proportion of variants that were discovered through exome resequencing. Additionally, we found a small percentage of genes included in various analyses lacked sufficient coverage, potentially resulting in false negative results. An example of false negative results influencing WES studies was shown by Gilssen *et al.* (2012), whom demonstrated a failure to identify the causative gene associated with Kabuki syndrome (*MLL2;* at the time unknown*)* since it was not represented on the enrichment kit used and therefore was not sequenced. This highlights the need for stringent validations and consideration of coverage when analysing data generated.

8.3.2 *ERCC4* and PNAO

In this thesis we report the discovery of the variant Ser613X in *ERCC4* in one patient with PNAO. The patient was heterozygous for the variant. This variant would result in a truncated form of XPF, missing both the nuclease domain and the helix-hairpin-helix (HhH_2) domain, important for binding to ERCC1 (de Laat *et al.* 1998).

Since the DNA binding of the complex and stability of XPF is dependent on *ERCC1* (Tsodikov *et al.* 2005; Tripsianes *et al.* 2005; Arora *et al.* 2010), this would suggest that the haploinsufficiency seen in Patient 8 could be due to a decreased level of active XPF-ERCC1. This could potentially lead to inadequate DNA repair of oxaliplatin induced adducts and an increased rate of apoptosis, characterised by a heightened sensitivity.

We presented the identification of two rare nonsynonymous variants in *ERCC4* which were shown to collectively contribute to PNAO. The rare variant hypothesis of disease aetiology states that individually rare but collectively common variants influence the likelihood of disease. We therefore theorised that these variants could be displaying varying but complementary effects on protein function. Both of these variants fall within proposed functional domains of XPF (McNeil and Melton, 2012). Recently, the Glu875Gly variant has been shown to alter the DNA binding ability of the XPF-ERCC1 complex, despite not effecting the protein-protein interaction between the two (Allione *et al.* 2013). Interestingly, the variant Pro379Ser has previously been identified as a pathogenic mutation in XPF when seen in a compound heterozygous state with other *ERCC4* mutations (Gregg *et al.* 2011). Previously, analysis of XPF and ERCC1 in cells derived from XPE and XPF patients have revealed that there is cytoplasmic mislocalisation of both proteins that potentially contributes to a reduced capacity for DNA repair. In wild type cells, XPF and ERCC1 are never seen solely in the cytoplasm. Two cell lines from patients with one Pro379Ser allele, XP7NE and XP32BR, both displayed cytoplasmic localisation of XPF (Ahmad *et al.* 2010). Previously, cellular mislocalisation of proteins has been linked to other diseases including cystic fibrosis, where mislocalisation of mutant forms of the cystic fibrosis transmembrane conductance regulator (CFTR) chloride channel to the endoplasmic reticulum ultimately results in protein degradation (Welsh and Smith, 1993).

Interestingly, previous research has suggested that altered expression of the sub units of the XPF-ERCC1 complex could affect the response and side effects to platinum treatment, presumably through an altered ability to repair damaged DNA (Section 1.6.2). Typically, increased expression of the subunits is correlated with an increased resistance to treatment, and a worse prognosis. On the contrary, a decreased level of expression leads to a heightened sensitivity to treatment. This

could result in an improved response, as well as potentially elevating the rate of adverse effects due to a reduced capacity for DNA repair, a build up of adducts and an increase in apoptosis.

8.3.3 NER involvement in neuronal function and PNAO

Since neurons are considered terminally differentiated, the need to replicate the genome is obsolete. However, the stability of DNA is paramount for adequate transcription. The various pathways of NER have been shown to be key in the maintenance of neuronal DNA, in particular TC-NER (Jaarsma *et al.* 2011). This is supported by the observation that approximately 20-30% of patients with mutations in several complementation groups of XP (XPA, XPD and XPG) have been reported to exhibit neurological symptoms, including peripheral neuropathy (Thrush *et al.* 1974; Kanda *et al*. 1990; Robbins *et al.* 2002; Anttinen *et al.* 2008). Additionally, reduced capacity of the NER pathway has been linked to amplified adduct levels in the dorsal root ganglion of *Xpa*$^{-/-}$ and *Xpc*$^{-/-}$ mice following cisplatin treatment, implicating a role in the development of peripheral neuropathy (Dzagnidze *et al.* 2007).

With regards to mutations in *ERCC4,* XPF patients have previously exhibited symptoms of a milder neurological condition (Gregg *et al.* 2011), with signs of axonal polyneuropathy reported in one patient (Sijbers *et al.* 1998). This is supported by the findings that reduced expression of the XPF-ERCC1 complex mimicking that seen in XFE syndrome due to functional disruption of *ERCC1* in mice has been shown to cause a distinct phenotype associated with peripheral neuropathy. The phenotype consisted of an accelerated aging related neuronal dysregulation and morphological loss of neurons (Goss *et al.* 2011).

PNAO has previously been shown to be due to direct oxalate toxicity on neuronal cells by altering the action of voltage gated sodium channels (Grolleau *et al.* 2001). However, the genetic findings presented here suggest that DNA repair mechanisms could also contribute to neuropathy in the acute setting. The inability to repair DNA adducts that form following oxaliplatin treatment potentially increases the rate of neuronal apoptosis, synergistically contributing to an elevated rate of PNAO when the direct toxicity of oxalate is considered.

8.4 Assaying the effects of *ERCC4* variants on DNA repair

Using RMCE, we have successfully produced a model system in which to assay the functional effects of variants identified in *ERCC4* in the *S.pombe* homolog, Rad16. We observed a heightened sensitivity with the *rad16*^{Ser585X} strain to all forms of DNA damaging agents tested. The degree of sensitivity was similar to that observed with *rad16Δ*. Rad16 binds to the ERCC1 homolog, Swi10, via the C terminal domain (Carr *et al.* 1994) in much the same manner as XPF binds to ERCC1, and the sensitivity observed here highlights the importance of the C terminal interaction in complex formation for adequate DNA repair.

The primary action of oxaliplatin as a chemotherapeutic is via the formation of inter and intrastrand crosslinks in DNA. The NER pathway is involved in the removal of intrastrand cross links but functions poorly in the removal of ICLs. However, XPF-ERCC1 has NER independent roles in the repair of ICL by unhooking and HR (Niedernhofer *et al.* 2004). Interestingly, we observed between a 40 and 60% decreased survival of all strains constructed with nonsynonymous variants of interest following oxaliplatin treatment in comparison to the *rad16*$^{+}$*.* However, this was not seen in the acute UV treatment, or mirrored in spot test treatments utilising MMS or UV treatment. This suggests that another mechanism distinct of NER was affected. We theorise that a heightened oxaliplatin sensitivity of strains with nonsynonymous variants could be due to a functional effect of the complex that hinders its ability to repair ICL. It is unclear what the exact effect of these variants on ICL repair could be. However, the recent findings that mutations in *ERCC4* in two patients result in a FA phenotype as a result of an inability of the nuclease domain to properly process ICL, despite a relatively intact ability of the NER pathway, indicate that particular mutations could variably affect XPF in different mechanisms of DNA repair (Bogliolo *et al.* 2013).

As mentioned previously, various *ERCC4* mutations have been associated with a cellular mislocalisation of the complex. It could be possible the variants introduced into *rad16* could be affecting localisation of the protein product. During the construction of *rad16*$^{+}$*,* we incorporated a histidine tag at the 5’ genomic region of *rad16*. The histidine tag could therefore prove useful to assay for any mislocalisation effects of the introduced variants in *S.pombe* both before and after treatment with

oxaliplatin. This could be carried out by using antibodies to target the histidine tag in order to compare expression in separated nuclear and cytoplasmic cellular fractions or, alternatively, through immunofluorescence.

8.5 Future directions

8.5.1 Analysis of *ERCC4* variants in human cells

Although *S.pombe* acts as a good model for the variants discovered in *ERCC4*, it is desirable to investigate these variants further in human cells. There are many differences between the two organisms that could mean the effect of the variants observed may not be representative of human cells (Section 1.8). We have genotyped and identified HRC lymphoblastoid cell lines heterozygous for the *ERCC4* variants Pro379Ser, Arg576Thr and Glu875Gly. Work in our lab has begun to investigate the effects of these using various assays. This is being done by analysing the effect of these variants on survival following treatment with oxaliplatin and UV light, localisation by immunofluorescence, as well as assaying for the rate of repair of UV induced adducts.

8.5.2 Functional analysis of *ERCC6*

In chapter five, we identified five predicted to be damaging rare nonsynonymous variants in *ERCC6*. Two of these were seen to collectively contribute to PNAO. This suggests that mutations in other components of the NER pathway could be playing a role in PNAO risk. If we are unable to validate these results in an independent cohort, it would be desirable to introduce these variants into a model organism, such as carried out here for rare variants in *ERCC4.* This could prove useful when assessing for any functional effects of the variants in the repair of oxaliplatin induced DNA damage. However, due to time constraints this is something that is beyond the scope of this thesis.

8.5.3 NGS of patients with other adverse drug reactions

We have shown that NGS could potentially be used to uncover alleles associated with adverse drug reactions in the chemotherapeutic treatment of CRC. Therefore NGS could potentially be used as a tool to find genetic reasons for other adverse events in patients exhibiting severe forms of a given side effects (Summarised in table 1.6)

8.5.4 GWAS of severe adverse events

As well as taking an unbiased approach to disease gene discovery, GWAS encompass a large proportion of variation across the genome by genotyping SNPs in regions of high LD. Previously, GWAS has successfully been used to study severe adverse reactions to cancer treatments. With regards to CRC, a recent GWAS of toxicity associated with 5-FU or FOLFOX treatment uncovered and validated one SNP that was significantly associated with 5-FU associated diarrhoea (Fernandez-Rozadilla *et al.* 2013). However, low statistical power due to problems' reaching sufficient sample size in such studies and the need for adequate replication cohorts has meant that it can prove difficult to ascertain a specific association signal. It has been proposed that samples used in GWAS could be enriched in order to increase the chances of observing a given risk association. This could be done by concentrating on smaller cohorts that display an extreme form of a given phenotype to a given drug. By effectively enriching for a given phenotype we hope to increase the likelihood that a particular signal of high effect will be observed (Gurwitz and McLeod, 2013). Additionally there have been advances in genotyping chip design to cover regions not tagged by the common variants seen on traditional chips (Spencer *et al.* 2009). This means that some rare variants that would have otherwise been missed may now be successfully assayed. Association SNPs identified in GWAS can aid in assaying for the true casual SNP (as long as the association strength between the two is high) by guiding the researcher to the region that should be focused on during sequencing (Freedman *et al.* 2011). Taken together, these advances could prove useful in the discovery of alleles associated with severe adverse events in CRC treatment and broaden pharmacogenetic understanding.

Publications

- Smith CG, **West H**, Harris R, Idziaszczyk S, Maughan TS, Kaplan R, Richman S, Quirke P, Seymour M, Moskvina V, Steinke V, Propping P, Hes FJ, Wijnen J, Cheadle JP (2013). Role of the Oxidative DNA Damage Repair Gene OGG1 in Colorectal Tumorigenesis. *J Natl Cancer Inst*. Jul 12 (Appendix 34)

- Smith CG, Naven M, Harris R, Colley J, **West H**, Li N, Liu Y, Adams R, Maughan TS, Nichols L, Kaplan R, Wagner MJ, McLeod HL, Cheadle JP (2013). Exome Resequencing Identifies Potential Tumor-Suppressor Genes that Predispose to Colorectal Cancer. *Hum Mutat*; 34(7): pp 1026-1034

- Dunlop MG, Dobbins SE, Farrington SM, Jones AM, Palles C, Whiffin N, Tenesa A, Spain S, Broderick P, Ooi LY, Domingo E, Smillie C, Henrion M, Frampton M, Martin L, Grimes G, Gorman M, Semple C, Ma YP, Barclay E, Prendergast J, Cazier JB, Olver B, Penegar S, Lubbe S, Chander I, Carvajal-Carmona LG, Ballereau S, Lloyd A, Vijayakrishnan J, Zgaga L, Rudan I, Theodoratou E; Colorectal Tumour Gene Identification (CORGI) Consortium, Starr JM, Deary I, Kirac I, Kovacević D, Aaltonen LA, Renkonen-Sinisalo L, Mecklin JP, Matsuda K, Nakamura Y, Okada Y, Gallinger S, Duggan DJ, Conti D, Newcomb P, Hopper J, Jenkins MA, Schumacher F, Casey G, Easton D, Shah M, Pharoah P, Lindblom A, Liu T; Swedish Low-Risk Colorectal Cancer Study Group, Smith CG, **West H**, Cheadle JP; COIN Collaborative Group, Midgley R, Kerr DJ, Campbell H, Tomlinson IP, Houlston RS (2012). Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nat Genet*.; 44(7): pp 770-776 (Appendix 33)

Appendix

Appendix 1

Primers used in the amplification and Sanger sequencing of the ORF, flanking regions and 5'UTR of RAD1 (tran – transcript)

Region	Forward primer (5'-3')	Reverse primer (5'-3')	Product size (bps)
Tran 1 – 5'UTR	ATGCAATCCAATCTGGCTCT	AGCGCGGAGTAGGTGATAAG	501
Tran 2 – 5'UTR-A	TGGAAACAATCGCTCAAAAA	TGATTGCGCCACTACATTTC	466
Tran 2 – 5'UTR-B	AGACAGGGTCTTGCTCCTTG	AAGTTGGAGTCAGAGCCTATTTC	389
Tran 3 – 5'UTR-A	GCGAGAAATAACCAAGGAAAA	GCAAGGTAGGAGGGGATGT	347
Tran 3 – 5'UTR-B	TCAAGTAAGTAACCCAAGAAAAGG	GAAGGAGGCGGCACAGAC	361
Exon 2	AGCCCCTTTCCACCTCTC	TTGTCTACTGAAACCTTCCGATT	431
Exon 3	TTCTCATGGGATTAGCACAGTA	TGAACCAATGTTTATGTTCCAA	316
Exon 4	AGGAGAAGCTGAACCCAGAA	TGGGAAGATGGAGTACAGACC	372
Exon 5	TGTGGTTTATTTTTGGATGAATG	ACCTCCTCTTTATCACCAATGA	194
Exon 6	TGGGAGTTCTGAGCAGTGTT	GAAAATCCAATATGAAATGACAAA	339

Appendix 2

Primers used in the amplification and Sanger sequencing of the ORF, flanking regions and 5'UTR of BRIX1.

Region	Forward primer (5'-3')	Reverse primer (5'-3')	Product size (bps)
5'UTR	CTCCTGGGGCCAACAACT	GACCCCACCGCAAAGGTA	508
Exon 1A	TCCAACAAAACAGGCGATG	TCGCTCCTATTTCCGATCTC	417
Exon 1B	GGAGGAAGTGAAGCCAGTCC	TTAACACCCGGGCTACTCTG	513
Exon2	CCTGGGCAACAGTGTGAGA	TGAGAACACTGAACAAATGAAGA	299
Exon3	TCTGGATAGCATATGTGGTTTGA	TGGAAGACCAGTTATTGAAGAGT	273
Exon4	AGCCTGGTTAGGTATTTTTGAGA	TTTGCTTCATTCTCAACCCTTA	292
Exon5	TGGTGAATAGGTGGTAAGCATT	TTTGCACAACTATTTCAAAAGATTA	187
Exon6	AAAAGTAATCTTTTGAAATAGTTGTGC	AAGGTGGGGGTGAAACTAAA	243
Exon7	GCCAAGTTATAGAAACAAATGAGC	GGATGATACCGTGGTGTACTAAA	240
Exon8	GGCTTTTGATGAATTACCACATT	GGGCAACAAGAGCGAGAC	323
Exon9	GGATTATCAATTATTTCAGGCACA	GGCCAAAGGGTTCTGGTA	284
Exon10	CACTGGCTGAAAGGATATATGG	TTTTTCTTTCTTGTAAATGCTACACT	495

Appendix 3

Primers used in the amplification and Sanger sequencing of the ORF, flanking regions and 5'UTR of DNAJC21 (Tran-Transcript).

Region	Forward primer (5'-3')	Reverse primer (5'-3')	Product size (bps)
5'UTR	GCTCCTATCTCCCCCTTCAG	GGCGGTGGTGTAGGTCAGT	463bps
Exon 1	ACAGAGCCCACCCCTAGC	GCCAGGTCCCTCTCTGCT	485bps
Exon2	GTTTGTGGCATTTCTGATGG	CCTATATTGATAACTGCTCCCAAC	242bps
Exon3	TTCAAAAGGAGCAAGAAATCC	GGCAATGCATCTTCAGTTTTC	332bps
Exon4	CCCAAAATTCTTCAACATTAAAA	GCCTGGGTGACAGAGTGAGA	327bps
Exon5	GGTAAAAGATGTTTCGCATCAG	GGCTGATGACTGAACCCAAC	493bps
Exon6	GCTCTCTAGTGGGAATGGATTTT	TTGGGAGATGTCAAATAAGCA	327bps
Exon7	GCATATTTTAGATTTGTGCTCTGA	ACTGTGCCACTGCACTCCT	229bps
Exon8	TTGGTTGCAGTTATCCAGCA	CCTGGGCAACAGAGTGATTC	338bps
Exon9	CCATTGAACTACAGCCTTGTG	CTGAATAAATAAGAGCACTGCAAC	311bps
Exon10 (Tran2-Exon11)	TTGGCAACATAACATAAAAAGC	TGGTCAGTCATGGGAAAGAA	347bps
Exon11 (Tran2-Exon12)	AGAGAGCACTCAAATAATGATGG	GGAATGGCTCACCAAATACA	245bps
Exon12 (Tran2-Exon13)	ACAATTGTTTGATGCTTAATCTTG	TGCAGATCACTGAAATTTTAACTC	353bps
TRAN2-Exon10	TTGAATGTGGGTTGTGTAACAG	CAATGGCAACAACAAACAGG	281bps

Appendix 4

Primers used in the amplification and Sanger sequencing of the ORF, flanking regions and 5'UTR of TTC23L.

Region	Forward primer (5'-3')	Reverse primer (5'-3')	Product size (bps)
5'UTR - A	TGCCAACGAAAGGGTAAGA	AAAACTGGCCTCTGCATATTGT	423
5'UTR - B	CGCTCCCTTCCATCCTTG	ACGTCCCCAGTTACCTTCC	322
Exon 2	CAAACCAGAGGGGGAAAATAG	AAAGCTCCTCACCCAGGTTT	238
Exon 3	TGTCCCTATCCCCTAGTTGG	AATCTGCATCGAAGGCAAAG	369
Exon 4	TTCAGTCCTGTGTTCCAGTGA	GTTGTGTGGGTCAGTTCAGC	273
Exon 5	CCTCCCAGGAATGTTTTTGA	ATCTCCCACCCTCCTGTATG	322
Exon 6	TATCACCTGCTGTCCCTGTG	CATGTAATTCCAAGCCTCATTC	250
Exon 7	CAGTTCCTTTTGTGTCTGCAA	TTCCAGCCCTTGTTCTTCTG	351
Exon 8	GCTGGGACTCAAACTCACCT	CCAATGTGCTTCCCTCATGT	253
Exon 9	TGTCAATTGAGCCAAAGCTG	TGCCTAGTTTTATCTGGGACCT	323
Exon 10	CAGTGCAATGAAAGGAGAGACA	TCCTCCATACACTGCCCTCT	184

Appendix 5

Primers used in the analysis of expression of specified genes in kidney and colon cDNA.

Gene (exons covered)	Forward primer (5'-3')	Reverse primer (5'-3')	Product size (bps)
AGXT2 (Exons 3-7)	CTTGGCTACAACCGTGTCCT	CCACGAAAAACATCTGGACA	521
AGXT2 (Exons 6-11)	GAACTCCCTGGTGGGACA	AGGCATTTCGCCAAAGATT	479
BRIX1 (Exons 2-7)	AACGGATTCTCATCTTTTCTTCC	GCATAATGTGGTAATTCATCAAAA	358
BRIX1 (Exons 6-10)	CCCTCGCTGAACTGAAGATG	CAGTGGGATCATGTGGAAGA	468
DNAJC21 (Exons 2-5)	ATCTGGATAATGCCGCAGAA	TCATCTCTTCGGCTTTCCTC	612
DNAJC21 (Exons 6-10)	TGGTGGAGCAGTACAGAGAACA	ACTCCTTCTCCAGGTCCATTT	492
RAD1 (Exons 1-4)	ACTTCCTCCGCGGTTCCT	AGGCTTGTCAGGAGACATGG	719
RAD1 (Exons 3-6)	TCAGGAGTTTAAAGTTCAGGAAGA	TCAAGACTCAGATTCAGGAACTT	640
TTC23L (Exons 6-10)	TGGCAGAGAAGCCTATTTCAA	TCCTCCATACACTGCCCTCT	644
TTC23L (Exons 3-6)	TCATCCCAAAGAGAAATTAGCC	AGCAAGTGTTAGGTCGTTCTCA	463

Appendix 6

Primers used for MLPA for CMT.

Gene	Forward primer (5' section of probe)	Reverse primer (3' section of probe)	Product size (bps)
IL4	CTACATTGTCACTGCAAATCGACACCTAT	TAATGGGTCTCACCTCCCAACTGCTTCCCCCT	130
KCNJ6	TGACATGCCAAGCTCGAAGCTCCT	ACATCACCAGTGAGATCCTGTGGGGTTACCGG	136
PMP22	CAGTTACAGGGAGCACCACCAGGGAA	CATCTCGGGGAGCCTGGTTGGAAGCTGCAGGCTTAGTCTGT	142
PMP22	GCTACAGTTCTGCCAGAGATCAGTTGC	GTGTCCATTGCCCACGATCCATTGCTAGAGAGAATCAGATA	148
KIF1B	TATGTCGGGAGCCTCAGTGAAGGT	GGCTGTCCGGGTAAGGCCCTTCAATTCTCGAGAGACCAGC	154
PRKCE	CGAGTTCGTCACCGATGTGTG	CAACGGACGCAAGATCGAGCTGGCTGTCTTT	160
PMP22	CCTCTTCCTCAGGAAATGTCCACCACTGTT	TCTCATCATCACCAAACGGTGAGGCTGGTTTTGTGCT	166
PMP22	TGACAGGATCATGGTGGCCTGGA	CAGACTGCAGCCATTCTGGGGGAAAGAGACACTTGGTTAGG	172
CFTR	CTTGTTCCATTCCAGGTGGCTGCTTCT	TTGGTTGTGCTGTGGCTCCTTGGAAAGTGAGTATTCCATG	178
BX089850	CTGCAGTTGGTTGAATCTGAAGAGCCCTT	GGATACGGAAGGCTGACTGTGTATGGCTACTCTGAAGAATG	184
HIPK3	CCTCAAGACCTATGTTACAGCATCCAACT	TATAATATCTCCCATCCCAGTGGCATAGTTCACCAAGTCCC	193
TEKT3	GCCTTGTTAACGAGGTACACGAGGTTG	ACGACACCATCCAGACCCTGCAGCAGCGCCTGAGGGATGC	202
IFNG	TAAGTAGGAACTCATCCAAGTGATGGCTGAACT	GTCGCCAGCAGCTAAAACAGGGAAGCGAAAAAGGTCTAGA	211
KIF1B	GGAGCACAAAGCACCGTGGGGTCCTT	TTGCAGGCCCTCAATGACAAAGACATGAACGACTGGTT	219
PMP22	GGGAGGGTCTTGCCTTAACATCCCTT	GCATTTGGCTGCAAAGAAATCTGCTTGGAAGAAGGGGTTAC	229
PMP22	TCTTCTCAGCGGTGTCATCTATGTGATC	TTGCGGAAACGCGAATGAGGCGCCCAGACGGTCTGTCTGA	238
STCH	CAATGATGTATATGTGGGATATGAAAGCG	TAGAGCTGGCAGATTCAAATCCTCAAAACACAATATAT	247
PMP22	CCAGAATGCTCCTCCTGTTGCTGAGTA	TCATCGTCCTCCACGTCGCGGTGCTGGTGCTGCTGTTCGT	256
SFTPB	GCTCATGCCCCAGTGCAACCAAGT	GCTTGACGACTACTTCCCCCTGGTCATCGACTACTTCCAG	265
LRRC48	CTGAGCTTGTTCAACAACCGGATCTCCAAG	ATCGACTCCCTGGACGCCCTCGTCAAGCTGCAGGTGT	274
DNAH5	GGCTTTCCTGGAGCTACTCAATACATTGAT	AGACGTCACCACGAGGGATCTGAGTTCCACGGAACGA	283
TEKT3	GGGACCGCTTTCCCCACTCCAATT	TGACCCATAGCCTGAGCCTTCCTTGGAGACCCAGCAC	292
NTNG1	CCATGAACATGGCAGTGCTATGACTTTTCT	GACTACTCTTAACCAGTGAGGGCTACCTAGACTCAGGTGC	301
PMP22	CTGTCTCTGTTCCTGTTCTTCTGCCAACT	CTTCACCCTCACCAAGGGGGGCAGGTTTTACATCACTGGAA	310
PTK2	CCAGGTTTCTGGCTACCCTGGTTCACATG	GAATCACAGCCATGGCTGGCAGCATCTATCCAGGTCAGGCA	319
BRCA1	GATGCACAGTTGCTCTGGGAGTCT	TCAGAATAGAAACTACCCATCTCAAGAGGAGCTCATTAAG	328
PMP22	CCGGAGTGGCATCTCAACTCGGAT	TACTCCTACGGTTTCGCCTACATCCTGGCCTGGGTGGCCTTCC	337
ELAC2	CTGACACCCAGCACTTGGTCCTGAAT	GAGAACTGTGCCTCAGTTCACAACCTTCGCAGCCACAAGAT	346
FLJ25830	GTGTAGCAGAACAGCTCAGGTGCTAGAAAT	AGCCAGTCTCATTGACTCAACTGTGTTTCCTCAGAGAATCCCG	355
APC	GCTATGGGAAGTGCTGCAGCTTT	AAGGAATCTCATGGCAAATAGGCCTGCGAAGTACAAGGATGCCA	364
FLJ25830	GACTGTCGTCAGCTCGCCTCCATGGTT	AGAGACTAGAATCGTGGAGCCCAATGTTTCCAACAGTGAG	373
LIMK1	CGTTTCATCTGCCTCACGTGTGGGACCTTT	ATCGGTGACGGGGACACCTACACGCTGGTGGAGCACT	382
COX10	CATGGCCCTTCCCATCAATGCGT	ACATCTCCTACCTCGGCTTCCGCTTCTACGTGGACGCAGAC	391
SECTM1	GGTGGTCACTGCTGTCTTCATCCTCT	TGGTCGCTCTGGTCATGTTCGCCTGGTACAGGTGCCGC	400
ERBB2	TGCACCTTCTACCGCTCACTGCTGGAG	GACGATGACATGGGGGACCTGGTGGATGCTG	409
COX10	GGGAGGAATCCTCTACTCCTGGCAGTT	TCCTCATTTCAACGCCCTGAGCTGGGGCCTCCGTGAAGAC	418
EIF3S3	GCCAGAACATCAAGGAGTTCACTGCCCAA	AACTTAGGCAAGCTCTTCATGGCCCAGGCTCTTCAAGAATA	427
KCNQ1	GCTCTCGGGAATTTGAGGCCTGT	GGCTGCTGTGGACCCTGGGAAAGAGCCTGTGCTTC	436

Appendix 7

Primers used for validation by amplification and Sanger sequencing of various variants identified through exome resequencing

Gene	Variant	Forward primer (5'-3')	Reverse primer (5'-3')	Product size (bps)
ERCC4	*Ser613X*	GCGCTCTAGGTTGCTGATTT	CTTCCTTGCCCTATCCTTCC	287
BRCA2	*Lys3326X*	TGACGAAGAACTTGCATTGA	TTCTTTTCTCATTGTGCAACATA	349
STOML3	*Arg164X*	CTGGGGAGAGGGGTATCAA	GTGTTGGAATTCTCACCGTTT	410
ANXA7	*Tyr54X*	AACAAGCAGGAATGAAGAGGA	TGTTCCTTATTTTTAGATGGGTCA	200
APPL1	*Phe472fs*	TCTGGGATTATGTTTGTACTGAAA	GAAATGCAGACAGGGGATTA	389
NEFM	*Tyr63fs*	TGAGCTACACGTTGGACTCG	ATCTCCGCCTCAATCTCCTT	397
NRP2	*His906fs*	GTGCTGGTGCTGGTCTCC	AACCAAAATGAACCCAAGGA	362
SEMA4C	*Gly648fs*	CGGGTCCTTCCTCTACGA	AGTAGCCTTGGCCCCTTTCT	304
PPP1R13L	*Pro562fs*	CCCCCTACCCACAAGAAAC	GGCTCTTTGCTACAGCTCCT	276
SLC22A1	*Pro425fs*	TTTCCTTTACTCCGCTCTGG	TGATTACAGGCATGAGCCACT	314
ERCC3	*Arg283Cys*	TTTTTGACCATTGGACCTCTT	TTGGCTTTTCAGCAAGGTGT	404
ERCC6	*Ser797Cys*	AAGAAGCTGGTGGAGAACTG	ACTGCCACCTCAGCATCAG	459
ERCC6	*Gly929Arg*	CTCACCCTGTCAACCTCACC	TCATCTCCACCAGAAGGTCA	358
ERCC6	*Phe1437Ile*	CCACTGGAATCAGATGTAGCTTT	TCTTCCTTTTTGGCCAGGTT	465

Appendix 8

Primers used for the amplification and Sanger sequencing of the ORF, flanking region and 5'UTR of ERCC4

Region	Forward primer (5'-3')	Reverse primer (5'-3')	Product size (bps)
5'UTR-A	GGGGATGTGGAAACTCAAAA	TGTTGAGCACCAGCACCA	559
5'UTR-B	AGCCTGGGCAACATATCAAC	AGAGAGCCGAGCCTGAGAA	399
Exon 1A	CTCTCGGACTCGGCTCTCT	GTGCAGCTGGAGAAAGTGG	252
Exon 1B	CCGCTGCTGGAGTACGAG	TGTCATCGCGTAGTGTCAGG	433
Exon 2	TCAGAGAAAGACAGCACATTATTT	TGGAGAAAAATAAAATGGAAATTG	357
Exon 3	CTCTGTTCTGTGCGTGGCTA	CCATCAAATTGCTCTCGACTT	547
Exon 4	TTTGTTGTTTTGCTTTTCGTG	GCTATGTTTTTAAGTGACCTCCA	425
Exon 5	GATACACAGGAAATAATCCTTTTGA	CACACCTGATTCCCCCTAAA	354
Exon 6	CGGTGTGGTTGGTAGGAAGA	TTTCACATGGCCAAAGAAGAC	348
Exon 7	TGATGCTCGTGTTATCTGTTG	AAATAGAGACAGGGTTTCACCA	327
Exon 8A	ATGTCTTCCCTTCGGGTGA	AGCCCGTTCTTTGTTTTGG	314
Exon 8B	GAGCGGAGGCCTTCTTATTG	AGTGAGGGGTTCTTTCAGGA	377
Exon 8C	AAGGAGATGTCGAGGAAGGA	AAGCAGCATCGTAACGGATA	401
Exon 9	GCGCTCTAGGTTGCTGATTT	CTTCCTTGCCCTATCCTTCC	287
Exon 10	TCCTTGTTTTTGTTTTTGTTTTTC	CCAACCCCCATTTTTAAGAG	361
Exon 11A	CCATCCATCAGAGTTAACAACA	CCTCGGGAAGTGAGAGAGAA	403
Exon 11B	TGGAGCGCAAGAGTATCAGT	ATCAAGGAGCGGCAGTTTTT	430
Exon 11C	CTGAAACAAAGCAAGCCACA	TCTGGTCCACCGTACAATCA	442

Appendix 9

Primers used for the amplification and Sanger sequencing of the ORF, flanking regions and 5'UTR of ERCC1

Region	Forward primer (5'-3-)	Reverse primer (5'-3')	Product size (bps)
5'UTR-A	GCTGTCGTTGGTCACTGCT	AGACTGCAGAGGGATCGAG	463
5'UTR-B	CCTGCTCTATGCTCTACTCTCC	AGAGCTCCATAGCGTCAGGT	482
Exon 1	TGCGGGATGAGAACGTAGAC	CCCCATCCTATCCTCTTCGT	237
Exon2	AAAGGGGAGAGGAACTCACA	GGAGAACAAAGTGGCTGGAA	405
Exon3	GTGCAAGAAGAGGTGGAGGA	TCCAGAACACTGGGACATGA	263
Exon4	ATTCCAGTGAGAGGGAAAAGG	CTGCATTTCCTCTTGGAAGG	265
Exon5	CCACCACGCCTGGCTAAT	ACAGGAAGGAGAAGGGAAGG	241
Exon6	GGCAATTCTTATGACTGACCA	TGGAACTGAAGCTCAACCAC	255
Exon7	CAGGCAGTCTGGGGACAC	CAGGGAGATGGAAGGAAATG	260
Exon8	CCCTGGGGAATATCTGAGG	AGGCTGGTCTCCAACTTCTG	351
Exon9	TAAAGAACCAAAACCCCACTC	CAGAATCCCTCCCCAGAGAC	238
TRAN5Ex3	AAGTGATCCTCCTGCCTCAG	CTGGCTACAGGCCAGCTCTT	169

Appendix 10

Primers used for the amplification and Sanger sequencing of the ORF, flanking regions and 5'UTR of STOML3

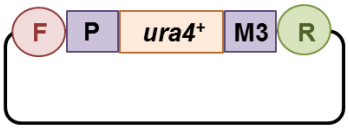
Region	Forward primer (5'-3')	Reverse primer (5'-3')	Product size (bps)
5'UTR-A	TCGAGATTGCAGTGAGCTGT	AAAGACCAAACTCCTAAGTGTCC	209
5'UTR-B	GCTTTGGAGGCACTGATAGG	GCAGTGAGTGAGCTTTTGAAA	397
Exon 1A	TTTCAAAAGCTCACTCACTGC	TGTGAAGAACAGGCAGCAAC	419
Exon 1B	CTTCCCTCACCAGGGTAACT	TGCTACAACTCCTGCTTTGC	293
Exon 2	TTCTATGCAGCCACATCAGG	CCAGACGGAATACAACAGCA	342
Exon 3	CATTACCTTCCCCATCTCCA	AATAGGCACCACCAGGAAAA	281
Exon 4	CATGTATCGCCCCATGTAAA	GCGGGTACTCAGCTCATCTT	302
Exon 5	GCCAGGACAGGTTTTAGGTG	GTGGGGGATGCTTTGAACT	441
Exon 6	TCACTCCAAATGCTGTAAATGC	AACCCCTTTCTCATGCAAAT	381
Exon 7	CTGGGGAGAGGGGTATCAA	GTGTTGGAATTCTCACCGTTT	410

Appendix 11 - Primers used for the amplification and Sanger sequencing of the ORF and flanking regions of ERCC6

Region	Forward primer (5'-3')	Reverse primer (5'-3')	Product size (bps)
5'UTR-A	CTCTCCATCCTTCCCGTGTT	CACAAGGAAACAGAGACGCT	517
5'UTR-B	TTCCTAGTCCCCTTTCCGAC	AACCCTGCAGAGAAAGTCAG	555
Exon 2A	TCTTTTGAGAAATCAGTCTGCTC	CATAGACGTCCACACCCAAA	410
Exon 2B	CCTCTCCTTTCGTTCTGTGG	AAGGGCTGGGACTCAAAGAT	462
Exon 3	GGGACATGTTTTTAGCATTTTTC	CAAAAGAAACCCAAAATTTCTC	350
Exon 4	ACAATAAATCACTTCATTGCTGTT	CTCAAAACCCAGGCAAAGAC	258
Exon 5A	TTCTTTTGTACCCTCGGAAA	TGGGGCTGGAGCTTTTCTA	312
Exon 5B	CCAGATCCCTCAGAAACAGG	CAGAGGCTTCAGCTCATAGTCA	472
Exon 5C	AGGGAGACTCTGAGGGTGAAG	GCTGCAGAAATCCAACCTCT	443
Exon 6	TTTGGGTAATAATGGCAGCA	AGTACCTTCAGGGCCCACAG	370
Exon 7	ATCCCGCATGTTTCTCTGAC	ATTCACAAGACCCTCCTCCA	373
Exon 8	TTGTTTTGGGGGATCTTTTG	CATGGATCAGAGAAAAACCAAA	295
Exon 9	TGTGTCATGATTGAAAGGGTGT	GCAGTTGGCACAAGGAAAG	349
Exon 10	GCAGAGGAGCGTTTTAGGG	CCTGGCCATCTTTCTCACAT	364
Exon 11	AAGGAATGGAAGCAATTGAGAG	AAGAATCCACTGAGGGCAGA	359
Exon 12	ATCGCATGCATTGTGCTCT	ATCTACCATGCGGGACTTCA	343
Exon 13	AAGAAGCTGGTGGAGAACTG	ACTGCCACCTCAGCATCAG	459
Exon 14	CTGTTGGGGAAGGTGTTACC	ATGCTGCTTTGGTGGGTAAG	301
Exon 15	GAGGCAGAGGCTGAGAACTG	TGAAGCCAACCAGAACATCA	284
Exon 16	CTGGCTGTTCATCTTCCTGA	CAAAACAAAAACCAAAGCCTTA	397
Exon 17	TTCCAAGGCTGAAAGTTTGG	CAAAGGCTGGTTGAATCCTT	434
Exon 18A	AAGTGCAATTTTTGCAGGTATTA	TTTGAACATTCCCCATTTCC	424
Exon 18B	AGCAATGATAGGCTTGGAGAA	AGGCTTTTGCTTTGGTCTCA	194
Exon 18C	GCCAGGCTCAAACAGAAGC	ACAGCCCTCTATGCACCATC	362
Exon 19	TTGATTCTCTAGGGTTTTCTAAGTG	CCACACCTGCCCTGATTTTA	415
Exon 20	AGCATGCAGATGGAAATGTT	TGACCATTTTCTTCACATCCA	285
Exon 21A	CTCACCCTGTCAACCTCACC	TCATCTCCACCAGAAGGTCA	358
Exon 21B	TCGCTTCCTCCTCACTCTTG	ATCAAGTGCAGCCAACTTCC	439
Transcript 2			
Exon6A	TGCTATTGTGCGGTAAAAAGA	TTTCCTCCTCCTTGATGGTG	465
Exon6B	TCAAGATGGCTCTGATGCTG	TTCCCAAAACATACGCCTTC	415
Exon6C	CATCTTGGCTTGACTAGCTCTG	CAGCAAATGTAGCCCAGACA	413
Exon6D	TCACGGGTGCAAACAATTTA	ACCAGATGAGGCAACAGTGA	439
Exon6E	CCACTGGAATCAGATGTAGCTTT	TCTTCCTTTTTGGCCAGGTT	465
Exon6F	TGGATTTTCTGGAGTTTCGTC	TGGAACACTAGGCAAAACCA	290

Appendix 12

Schematic of pAW1 with corresponding primer sequences. Primer sequence specific to plasmid are given. These allow for adequate amplification of ura4⁺ with flanking lox sites from pAW1 for the construction of the base strain (adapted from Watson et al. 2008).

Name	Schematic	Primers (5' – 3')	Product Size
pAW1		**F = CGGATCCCCGGGTTAATTAA R = GAATTCGAGCTCGTTTAAAC**	1.9KB

Appendix 13

Primers used in the amplification of loxP-ura4⁺-loxM3 from the pAW1 plasmid. In the forward primer, 100bp region upstream of the rad16 gene, ending immediately 5' to the start codon, was incorporated in the 5' region of the primers. Conversely, in the reverse primers, the 100bp region downstream of the rad16 gene, ending immediately 3' at the stop codon, was incorporated in the 5' region of the primer.

Primer	Sequence (5'-3')
ura4⁺F Primer (100bp upstream of rad16 + F from Appendix 12)	TCCATCCAAATTGGAAAATTTTCGCATCAAAGTATTTAACAGCTTTCAG AAATCAAAATTGCAAATTGGAAAATCTCTACGAATAACACCACCATTAA ATcggatccccgggttaattaa
ura4⁺R Primer (100bp downstream of rad16 + R from Appendix 12)	TTATTAATTAGGTGCGCTTAACATTCTATATATGGTGAACCAATATATAT CAGATGTAGAAGCAAAAATTAAATATATTACAAAATTATAAAAAAATAAA gaattcgagctcgtttaaac

Appendix 14

Sequences of the loxP and loxM3 recombination sites. Cre recombinase catalyses recombination effects at these specific sites. Two 13bp inverted repeats act as the binding sequence for Cre recombinase, whilst the 8bp spacer region helps in the replacement of strands. loxP and loxM3 differ in the spacer region only, allowing for specific directional recombination (Hoess et al. 1986; Langer et al. 2002).

Recombination Site	Sequence (5'-3')
loxP	ATAACTTCGTATA GCATACAT TATACGAAGTTAT TATTGAAGCATAT CGTATGTA ATATGCTTCAATA
loxM3	ATAACTTCGTATA TGGTATTA TATACGAAGTTAT TATTGAAGCATAT ACCATAAT ATATGCTTCAATA

Appendix 15

Primers used for colony PCR

Primer name	**Sequence (5'-3')**	**Product size with A(bps)**
A	TCACGATCCGCCTAAATTCC	-
B (*rad16* specific primer)	TCAAGGGACTCTGCCACAAC	888
C (*ura4*+ specific primer)	TAAAGCAAGGGCATTAAGGC	537

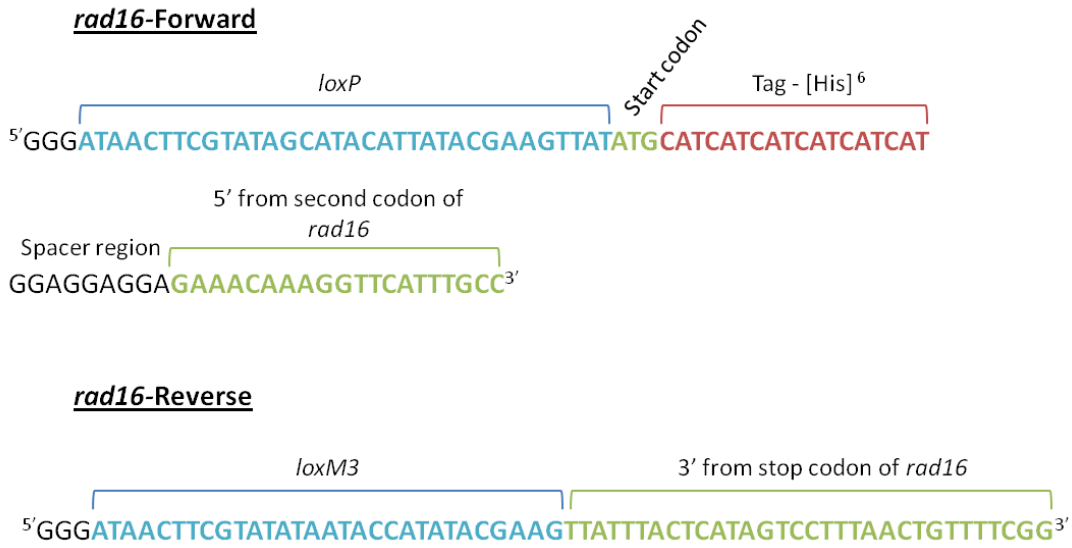
Appendix 16

Primer sequences used in the amplification and Sanger sequencing of lox sites from extracted genomic DNA

Primer set	**A and D (5'-3')**	**C and F (5'-3')**	**Product size (bps)**
A and C (5'-3') - *loxP* site	TCACGATCCGCCTAAATTCC	CTAAAGCAAGGGCATTAAGGC	537
D and F (5'-3') – *loxM3* site	CCCAAGCATTAGAATCACCAA	TGTACAAAGCCAATGAAAGATG	440

Appendix 17

Composition and structure of primers used in the construction of loxP-rad16⁺-loxM3 by PCR



Appendix 18

Primers used in the amplification and Sanger sequencing of the ORF of rad16 from extracted genomic DNA and constructed plasmids

Primer set	Forward primer (5'-3')	Reverse primer (5'-3')	Product size (bps)
1	*For pAW8-rad16 (pAW8-BF);* ATTAGGCACCCCAGGCTTTA; *For genomic DNA;* TCACGATCCGCCTAAATTCC	TGTGTTTTATGGAGGCGTGA	879 537
2	TGCGATTGTATCGTGAAACC	TGCATTAGATGGGTAGCTGGA	568
3	TCAGGTTCGTCTTTTCTTTCC	TGGAAGGCTCTGGTTTCTTG	582
4	CGTGATTATTTGAGTACGGTGA	TTTGCTCCTCAATCGATCCT	592
5	TTCGATTCTGATCCCAATTTT	TGTAAAAGATTGGTGTTGTTCG	642
6	GCCGTCTTTATTCCCAATGT	TCACGCCTCCATAAAACACA	436
7	CTGGGCAAGACTCAACGAA	For pAW8-rad16 (rad16-R); GGGATAACTTCGTATATAATAC CATATACGAAGTTATTTACTCA TAGTCCTTTAACTGTTTTCGG For genomic DNA; CCCAAGCATTAGAATCACCAA	224 538

Appendix 19

Primers used for SDM of pAW8-rad16⁺. Codons of interest are highlighted in grey, whilst nucleotide change/s shown in lower case and bold

Variant in *S.Pombe* Rad16	**Equivalent XPF**	**Native codon in *rad16***	**Forward primer (5'-3')**	**Reverse primer (5'-3')**
Pro361Ser	Pro379Ser (rs1799802)	CCG	GGAGGAACAG**t**CGAAATGG TCTGTCTTGCAAGACG	CGTCTTGCAAGACAGACCATT TCG**a**CTGTTCCTCC
Arg399Gln	Arg415Gln (rs1800067)	CGT	TGTGCGCAGATGAGC**aa**ACT TGTTTACAGTTACGTGA	TCACGTAACTGTAAACAAGT**tt** GCTCATCTGCGCACA
Arg548Thr	Arg576Thr (rs1800068)	CGG	ATTTAAGACCT**ac**GTACGTT ATTATGTTCGAT (Replacement; TTAAATAATTTAAGACCT**ac**G TACGTTATTATGTTCG)	ATCGAACATAATAACGTAC**gt**A GGTCTTAAAT (Replacement; CGAACATAATAACGTAC**gt**AG GTCTTAAATTATTTAA)
Ser585SX	Ser613Stop	TCG	GTACTATGGAGGAT**a**GATTG AGGAGCAAAA	TTTTGCTCCTCAATC**t**ATCCTC CATAGTAC
Glu844Gly	Glu875Gly (rs1800124)	GAA	GAAAGATATTCAAG**g**AGCTT CCGAAACCTCAGA	TCTGAGGTTTCGGAAGCT**c**CT TGAATATCTTTC

Appendix 20 - Scientific and common name of all mammalian species used in sequence alignments

Scientific name	Common name
Ailuropoda melanoleuca	Giant panda
Bos taurus	Cattle
Callithrix jacchus	Common marmoset
Canis lupus familiaris	Domestic dog
Cavia porcellus	Guinea pig
Ceratotherium simum simum	White rhinoceros
Condylura cristata	Star nosed mole
Cricetulus griseus	Chinese hamster
Dasypus novemcinctus	Nine banded armadillo
Echinops telfairi	Lesser hedgehog tenric
Equus caballus	Horse
Felis catus	Domestic cat
Gorilla gorilla gorilla	Western lowland gorilla
Homo sapiens	Human
Jaculus jaculus	Lesser Egyptian jerboa
Loxodonta africana	African elephant
Macaca mulatta	Rhesus monkey
Monodelphis domestica	Opossum
Mus muscularis	Mouse
Nomascus leucogenys	White cheeked gibbon
Ochotona princeps	American pika
Octodon degus	Degu
Odobenus rosmarus divergens	Walrus
Orcinus orca	Killer whale
Ornithorhynchus anatinus	Platypus
Oryctolagus cuniculus	European rabbit
Otolemur garnettii	Greater galago
Ovis aries	Sheep
Pan paniscus	Bonobo
Pan troglodytes	Chimpanzee
Papio anubis	Olive baboon
Pongo abelii	Sumatran orang-utan
Rattus norvegicus	Norwegian rat
Saimiri boliviensis boliviensis	Black capped squirrel monkey
Sarcophilus harrisii	Tasmanian devil
Sorex araneus	Common shrew
Sus scrofa	Wild boar
Trichechus manatus latirostris	West Indian manatee
Tursiops truncatus	Bottle nosed dolphin

Appendix 21- Species conservation of the glutamic acid residue at position 281 in RAD1 (highlighted in grey). Alignment produced with Clustal Omega. A list of common names for species is given in Appendix 20. An asterisk (*) indicates complete conservation of a residue; a colon (:) indicates conservation between residues that are have very similar properties; a period (.) indicates conservation between residues which have weakly similar properties.

```
Homo sapiens - NP_002844.1                                KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVPESES--
Bos taurus - NP_001179419.1                               KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVPDS----
Callithrix jacchus - XP_002745099.1                       KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVPESESSI
Canis lupus familiaris - XP_536505.3                      KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVSESES--
Cavia porcellus - XP_003470270.1                          KVSIRTDNRGFLSLQYMIKNEDGQISFVDYYCCPDEDIPESES--
Ceratotherium simum simum - XP_004422775.1                KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEELPESEP--
Condylura cristata - XP_004678143.1                       KVSIRTDNRGFLSLQYMIRNEDGQVCFVEYYCCPDEEVSESES--
Dasypus novemcinctus - XP_004453398.1                     KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEVVPESES--
Echinops telfairi - XP_004698420.1                        KVSIRTDNRGFLSLQYMIRNEDGQISFVEYYCCPDEEEVSKSEP-
Felis catus - XP_003981515.1                              KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVPESES--
Gorilla gorilla gorilla - XP_004059017.1                  KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVPESES--
Macaca mulatta - NP_001248089.1                           KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVPES----
Mus musculus - NP_035362.2                                KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVPES----
Nomascus leucogenys - XP_004088254.1                      KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVPESES--
Ochotona princeps - XP_004583824.1                        KVSIRTDNRGFLSLQYMIKNEDGQICFVEYYCCPDEEVPESESGV
Odobenus rosmarus divergens - XP_004412959.1              KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVPESES--
Orcinus orca - XP_004266062.1                             KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVPES----
Otolemur garnettii - XP_003792981.1                       KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEVVPESES--
Ovis aries - XP_004017107.1                               KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVPDS----
Pan paniscus - XP_003806716.1                             KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVPESES--
Pan troglodytes - XP_517813.2                             KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVPESES--
Papio anubis - XP_003899614.1                             KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVPES----
Pongo abelii - NP_001126246.1                             KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVPESES--
Rattus norvegicus - NP_001099889.1                        KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVPES----
Saimiri boliviensis boliviensis - XP_003925987.1          KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVPESESRI
Sarcophilus harrisii - XP_003759996.1                     KVSIRTDNRGFLSLQYMIKNEDGQICFVEYYCCPDEDVSDSES--
Sorex araneus - XP_004605611.1                            KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVPESES--
Trichechus manatus latirostris - XP_004385297.1           KVSIRTDNRGFLSLQYMIRNEDGQISFVEYYCCPDEV-PESES--
Tursiops truncatus - XP_004323605.1                       KVSIRTDNRGFLSLQYMIRNEDGQICFVEYYCCPDEEVPES----
                                                          ******************:*****:.**:*******   ..
```

End of protein

Appendix 22- Species conservation of the glutamine residue at position 1236 in POLG (grey). Alignment produced with Clustal Omega. A common key is given in Appendix 21. A list of common name for species is given in Appendix 20.

```
Homo sapiens - NP_001119603.1                    ALDIYQIIELTKGSLEKRSQPGP-----------------------------------
Bos taurus - XP_002696610.1                      ALDIYQIIELTKGSLEKXSQPGL-----------------------------------
Callithrix jacchus - XP_002749184.1              ALDIYQIIELTKGSLEKRSQPGP-----------------------------------
Cavia porcellus - XP_003475296.1                 ALDIYQIIELTKGSLEKXSQPGP-----------------------------------
Ceratotherium simum simum - XP_004431996.1       ALDIYQIIELTKGSLEKRSQPGP-----------------------------------
Condylura cristata - XP_004693042.1              ALDIYQIIELTKGSLERHSQSGL-----------------------------------
Dasypus novemcinctus - XP_004455270.1            ALDIYQIIELTKGSLEKRSQPGP-----------------------------------
Echinops telfairi - XP_004709933.1               ALDIYQIIERTKGSLGR-----------------------------------------
Equus caballus - XP_001503097.1                  ALDIYQIIELTKGSLDKRNQPGP-----------------------------------
Felis catus - XP_003986860.1                     ALDIYQIIELTKGSLEKXSPPGP-----------------------------------
Gorilla gorilla gorilla - XP_004056785.1         ALDIYQIIELTKGSLEKRSQPGP-----------------------------------
Jaculus jaculus - XP_004658240.1                 ALDIYQIIELTKGSLDKRSQPGP-----------------------------------
Loxodonta africana - XP_003413940.1              ALDIYQIIELTKGSLEKRSQPGP-----------------------------------
Macaca mulatta - XP_001092360.1                  ALDIYQIIELTKGSLEKRSQPGP-----------------------------------
Mus musculus - NP_059490.2                       ALDIYQIIELTKGSLEKRSQPGP-----------------------------------
Nomascus leucogenys - XP_003268536.2             ALDIYQIIELTKGSLEKRSQPGP-----------------------------------
Ochotona princeps - XP_004578377.1               ALDIYQIIELTKGSLEKRSQPGP-----------------------------------
Octodon degus - XP_004623348.1                   ALDIYQIIKLTKGSLEKRSQPGP-----------------------------------
Odobenus rosmarus divergens - XP_004398363.1     ALDIYQIIELTKGSLEPXSQPGP-----------------------------------
Orcinus orca - XP_004278297.1                    ALDIYQIIELTKGSLEKXSQPGP-----------------------------------
Otolemur garnettii - XP_003788714.1              ALDIYQIIELTKGSLEKXSQRGP-----------------------------------
Ovis aries - XP_004023438.1                      ALDIYQIIELTKGSLEKXSQPGLYPCLEALYLLPWSFIGAAKAPKTQAFSSAFCKRAGPR
Pan paniscus - XP_003820468.1                    ALDIYQIIELTKGSLEKRSQPGP-----------------------------------
Pan troglodytes - XP_523149.2                    ALDIYQIIELTKGSLEKRSQPGP-----------------------------------
Papio anubis - XP_003901414.1                    ALDIYQIIELTKGSLEKRSQPGP-----------------------------------
Pongo abelii - XP_002825850.2                    ALDIYQIIELTKGSLEKRSQPGP-----------------------------------
Rattus norvegicus - NP_445980.1                  ALDIYQIIELTKGSLEKRSQPGP-----------------------------------
Saimiri boliviensis boliviensis - XP_003921677.1 ALDIYQIIELTKGSLEKRSQPGP-----------------------------------
Sus scrofa - XP_001927099.1                      ALDIYQIIELTKGSLEKRSQPGP-----------------------------------
Trichechus manatus latirostris - XP_004370680.1  ALDIYQIIELTKGSLEKRSQPGP-----------------------------------
                                                 ********: *****
```

Appendix 23 - Species conservation of the valine residue at position 138 in REV1 (grey). Alignment produced with Clustal Omega. A common key is given in Appendix 21. A list of common name for species is given in Appendix 20.

```
Homo sapiens - NP_001032961.1                          PEWIVESIKAGRLLSYIPYQLYTKQSSVQKGLSFNPVCRPEDPLPGPSNIAKQLNNRVNH
Ailuropoda melanoleuca - XP_002920000.1                PEWIVESIKAGRLLSCIPYQLYSKPSNMQKGLNFNPACKAEESVPGPSSVAKQLNNRVNH
Bos taurus - XP_003586659.1                            PEWIVESIKAGRLLSCIPYQLYSKPSSVQKSLNFNPVCKPEDPLPGPSSITKQLNDRVNH
Canis lupus familiaris - XP_538458.3                   PEWIVESIKAGRLLSCIPYQLYSKQSNMQKGLNFNPACKAEEPVPGPSSVAKQLNNRVNH
Cavia porcellus - XP_005004513.1                       PEWIVESIKAGRLLSYIPYQLYTKQSNAQRSLNFNPVCKPEDPVPGPSHVAKQPNNKVNH
Ceratotherium simum simum - XP_004435909.1             PEWIVESIKAGRLLSYIPYQLYSKQSSVQKGLNFNPVCQPEDPVPGPSNIARQLNDRVNH
Condylura cristata - XP_004686221.1                    PEWIVESIKAGRLLSYIPYQLYNKQSSVQKGLNFNPVCKPEDPVPGPSDIAKQLNNRVNH
Dasypus novemcinctus - XP_004458115.1                  PEWIVESIRAGRLLSHVPYQLYTKQPSAQRGLTFHAVCKPEEPAPGPSSVAKPLNNRVNH
Echinops telfairi - XP_004704398.1                     PEWIVESIKAGRLLSCIPYQLYTKQSTVQRALNFNPVCKPEDSMPGPSNISKQLNNRVNH
Equus caballus - XP_001490149.1                        PEWIVESIKAGRLLSYIPYQLYTKQSNVQKGLNFTPVCKPEDPVPGPSNITKQLNNRVNH
Gorilla gorilla gorilla - XP_004031548.1               PEWIVESIKAGRLLSYIPYQLYTKQSSVQKGLSFNPVCRPEDPLPGPSNIAKQLNNRVNH
Macaca mulatta - XP_002799407.1                        PEWIVESIKAGRLLSYIPYQLYTKQSSVQKGLSFNPVCRPEDPVPGPSNIAKQFNNRVNH
Monodelphis domestica - XP_001363717.1                 PEWIVESIKAGRLLSYIPYQLYTKQSSVQKSLNFNPICKPEDSMPGPSNIAIELNNRVNQ
Mus musculus - NP_062516.2                             PEWIVESIKAGRLLSSAPYQLYTKPSAAQKSLNFNPVCKPEDPGPGPSNRAKQLNNRVNH
Nomascus leucogenys - XP_003274893.2                   PEWIVESIKAGRLLSYIPYQLYTKQSSVQKGLSFNPVCRPEDPLPGPSNIAKQLNNRVNH
Ochotona princeps - XP_004594269.1                     PEWIVESIKAGRLLSCLPYQLYSRQSSVQKSLNFNAVCKPEEPLPGPSNTAKQLNNRVNH
Odobenus rosmarus divergens - XP_004400175.1           PEWIVESIKAGRLLSCIPYQLYSKQSNMQKGLNFNPACKAEEPGPGPSSVAKQLNNRVNH
Orcinus orca - XP_004277800.1                          PEWIVESIKAGRLLSYIPYQLYSKQSSTQKGLNFNPVCKPEDPVPGPSSITKQLNDRVNH
Ornithorhynchus anatinus - XP_001507264.2              PEWIVESIKAGRLLPSIPYQLYTKQPGVPKGLNFNTICKPEDPLPGPSNRAKQLNQRVNH
Oryctolagus cuniculus - XP_002709984.1                 PEWIVESIKAGRLLSCIPYQLYNKQSAVQKSLTFSAACKPEDPLPGPSSIAKQFNNRVNH
Otolemur garnettii - XP_003794384.1                    PEWIVESIKAGRLLSYIPYQLYTKQSSVQKGLNFNPVCKPEDPVPGPSNIAKQFNNRVNH
Ovis aries - XP_004006184.1                            PEWIVESIKAGRLLSCIPYQLYSKQSSVQKSLNFNPVCKPEDPLPGPSSITKQLNDRVNH
Pan paniscus - XP_003805448.1                          PEWIVESIKAGRLLSYIPYQLYTKQSSVQKGLSFNPVCRPEDPLPGPSNIAKQLNNRVNH
Pan troglodytes - XP_001160264.1                       PEWIVESIKAGRLLSYIPYQLYTKQSSVQKGLSFNPVCRPEDPLPGPSNIAKQLNNRVNH
Papio anubis - XP_003909074.1                          PEWIVESIKAGRLLSYIPYQLYTKQSSVQKGLSFNPVCRPEDPVPGPSNIAKQFNNRVNH
Pongo abelii - NP_001126930.1                          PEWIVESIKAGRLLSYIPYQLYTKQSSVQKGLSFNPICRPEDPLPGPSNIAKQLNNRVNH
Rattus norvegicus - NP_001101683.1                     PEWIVESIKAGRLLSSVPYQLYTKPSTAQKSLNFTPVCKPEEPVPGPSEIAKQLTNRVNH
Saimiri boliviensis boliviensis - XP_003941371.1       PEWIVESIKAGRLLSYIPYQLYTKQSNVQKGLSFNPVCRPEDPVPGPSNIAKQLNNRVNH
Sarcophilus harrisii - XP_003765700.1                  PEWIVESIKAGRLLSYIPYQLYTKQSSVQKSLNFNPICKPEDPMPGPSNIAIELNNRVNQ
Sorex araneus - XP_004615851.1                         PEWIVESIKAGRLLSCIPYQLYSKQSNAQKGLSFNPVCKPGDPVPGPSNITKQFTNRVNH
Sus scrofa - XP_003481195.1                            PEWIVESIKAGRLLSYIPYQLYSKQSTLQKGLNFNPVCKPEDSLPGPSSIAKQLNNRVNH
Trichechus manatus latirostris - XP_004369371.1        PEWIVESIKAGRLLSYIPYQLYTKQSSVQKGLNFNPVCKPEDPMPGPSNIAKQLNNRVNH
Tursiops truncatus - XP_004323487.1                    PEWIVESIKAGRLLSYIPYQLYSKQSSTQKGLNFNPVCKPEDPVPGPSSITKQLNDRVNH
                                                       ********:*****   ****:.:     :.*.*   *:  :  ****      .::**:
```

Appendix 24 - Species conservation of variants in ERCC4. A. Pro379 (highlighted in grey); B. Arg415 (grey); C. His466 (grey) D. Arg576 (grey) and Ser613 (yellow) E. Glu875. A common key is given in Appendix.21. A list of common name for species is given in Appendix 20.

A.

```
Homo sapiens - NP_005227.1                          RVYHLPDAKMSKKEKISEKMEIKE------GEETKKELVLESNPKWEALTEVLKEIEAEN
Ailuropoda melanoleuca - XP_002918113.1             RVYHLPDAKMNKKGKMSEKMEIKE------EQETKKELVLESNPKWEALTEVLKEIEAEN
Bos taurus - NP_001192289.1                         RVYHVPDAKMSKKSRLPEKLEIKE------EQETKKDLVLESNPKWAALTEVLKEIEAEN
Callithrix jacchus - XP_002755959.2                 RVYHLPDAKISKKGKISEKMEVKE------GQETKKELVLESNPKWEALTEVLKEIEAEN
Canis lupus familiaris - XP_536967.3                RVYHLPDAKMNKKGKMSEKMEIKE------DQEAKKELVLESNPKWEALTEVLKEIEAEN
Cavia porcellus - XP_003477812.1                    RVYHIPDAKMSKKSKMSEKMEIKE------GQETKKELVLESSPKWEALTEVLKEIEAEN
Ceratotherium simum simum - XP_004441913.1          RVYHVPDAKMSKKDKMSEKMEIKE------EQETKKELVLESNPKWEALTEVLKEIEAEN
Condylura cristata - XP_004692174.1                 RVYQIPDAKASKKGKVSEKTEIKE------GQETKRELVLESNPKWEALTEVLKEIEAEN
Cricetulus griseus - NP_001230961.1                 RVYRVPDVKLNKKAKMSES---AE------GQETKKELVLESNPKWEALSEVLKEIEAEN
Dasypus novemcinctus - XP_004466678.1               RVYHVPDAKMSKRSKMSEKMEIEE------GQETQKELVLESNPKWEALSEVLKEIEAEN
Echinops telfairi - XP_004705920.1                  RVYHVPDAKMSKKSKRSEKTEVKE------GQETKKELVLESNPKWEALSEVLKEIEAEN
Equus caballus - XP_001490470.2                     RVYNVPDDKMSKKGKMSEKMEIKE------EQETKKELVLESNPKWEALTEVLKEIEAEN
Felis catus - XP_003998897.1                        RVYHLPDAKMNKKGKVSEKTEIEQ------VQETKKELVLESNPKWEALTEVLKEIEAEN
Gorilla gorilla gorilla - XP_004057285.1            RVYHLPDAKMSKKEKVSEKMEIKE------GQETKKELVLESNPKWEALTEVLKEIEAEN
Jaculus jaculus - XP_004652331.1                    RVYHVPDAKTSKTGKTSEILEIKE------GQETKKELVLENNPKWEALTEVLKEIEAEN
Loxodonta africana - XP_003417763.1                 RVYHVPDAKMSKKSQMSEKMEIKE------GQDTKKELVLESNPKWEALTEVLKEIEAEN
Macaca mulatta - XP_001107209.1                     RVYHLPDAKMSKKGKISEKMEIKE------GQEAKKELVLESNPKWEALTEVLKEIEAEN
Monodelphis domestica - XP_001375532.1              RVYHIADTRMNKKCKMTEKTNIKE------NQEMKRELVLESNPKWEALTEVLKEIETEN
Mus musculus - NP_056584.2                          RVYRVPDVKLNKKAKTSEKTSSPE------VQETKKELVLESNPKWEALTDVLKEIEAEN
Nomascus leucogenys - XP_004091351.1                RVYHLPDAXMSKKEKISEKMEIKERARYLVGLKTKKELVLESNPKWEALTEVLKEIEAEN
Ochotona princeps - XP_004586829.1                  RVYRVPDAKLMKKGKMSENMEIKE------GQETRKELVLECSPKWEALTEVLKEIEAEN
Octodon degus - XP_004644276.1                      RVYHIPDAKMSKKSKIPEKMEIKE------GQEAKKELVLESNPKWEALTEVLKEIEAEN
Odobenus rosmarus divergens - XP_004408974.1        RVYHLPDAKMNKKGQMSEKMEMKE------EQETKKELVLESNPKWEALTEVLKEIEAEN
Orcinus orca - XP_004270213.1                       RVYHVPDAKMSKKGKLPEKLEIKE------EQETRKELVLENNPKWEALSEVLKEIEAEN
Oryctolagus cuniculus - XP_002711804.1              RVYHVPDSKMSKKGKMSEKMEIKE------EQETKKELVLECNPKWEALTEVLKEIEAEN
Otolemur garnettii - XP_003802775.1                 RVYHVPDAKINIKGKISEKMEVKE------GQETKKELVLESNPKWEALTQVLREIEAEN
Ovis aries - XP_004021225.1                         RVYHVPDAKMTRKSKLPEKLEIKE-------QETKKELVLESNPKWAALTEVLKEIEAEN
Pan paniscus - XP_003832737.1                       RVYHLPDAKMSKKEKISEKMEIKE------GQETKKELVLESNPKWEALTEVLKEIEAEN
Pan troglodytes - XP_510831.2                       RVYHLPDAKMSKKEKISEKMEIKE------GQETKKELVLESNPKWEALTEVLKEIEAEN
Papio anubis - XP_003916616.1                       RVYHLPDAKMSKKGKISEKMEIKE------GQEAKKELVLESNPKWEALTEVLKEIEAEN
Pongo abelii - XP_002826187.1                       RVYHLPDAKMSKKEKISEKMEIKE------GQETKKELVLESNPKWEALTEVLKEIEAEN
Saimiri boliviensis boliviensis - XP_003928795.1    RVYHLPDAKISKKGKISEKMEIKE------GQETKKELVLESNPKWEALTEVLKEIEAEN
Sarcophilus harrisii - XP_003761738.1               RVYHIPDIRMNKKCKMTEKTD-QE------GQELKRELVLESNPKWEALTEVLKEIEIEN
Sorex araneus - XP_004604386.1                      RVYHIPDTKLSKKDKIPEKVETKE------KQETKRELVLENSPKWEALTEVLKEIEAEN
Trichechus manatus latirostris - XP_004373324.1     RVYHVPDAKMSTKSKMSEKMEIKE------EQETKKELVLESNPKWEALTEVLKEIEAEN
                                                    ***.: *       :  *     :        . :::**** .*** **::**:*** **
```

B.

```
Homo sapiens - NP_005227.1                          KESEALGGPGQVLICASDDRTCSQLRDYITLGAEAFLLRLYRKTFEK--DSKAEEVWMKF
Ailuropoda melanoleuca - XP_002918113.1             KESEALGGPGQVLICASDDRTCSQLREYITVGAEAFLLRLYRKTFEK--DTKAEEVWMKF
Bos taurus - NP_001192289.1                         KESEVLGGPGQVLICASDDRTCSQLREYLTVGAEVFLLRLYRKTFEK--DSKAEDVWMRL
Callithrix jacchus - XP_002755959.2                 KESEALGGPGQVLICASDDRTCSQLRDYLTLGVEAFLLRLYRKTFEK--DSTAEEVWMKL
Canis lupus familiaris - XP_536967.3                KESEALGGPGQVLICASDDRTCSQLREYITIGAEAFLMRLYRKTFEK--DSKAEEVWMKL
Cavia porcellus - XP_003477812.1                    KESEALGGPGQVLICASDDRTCAQLKDYLTAGAEAFLLQLYRKTFEK--DSKAEEVWMKF
Ceratotherium simum simum - XP_004441913.1          KESEALGGPGQVLICASDDRTCSQLREYIAIGAEAFLLRLYRKTFEK--DSKAEEVWMTF
Condylura cristata - XP_004692174.1                 KESEALGGPGRVLICASDERTCSQLREYLTGGAEAFLLRLYRKAFEK--DGKAQEVWAPP
Cricetulus griseus - NP_001230961.1                 KESEALGGPGQVLICASDDRTCCQLRDYLTAGAEAFLLRLYRKTFEK--DSKAEEVWVNL
Dasypus novemcinctus - XP_004466678.1               KESEALGGPGQVLICASDDRTCSQLREYITIGAEAFLLRLYRKTFEK--DSKAEEVWMKF
Echinops telfairi - XP_004705920.1                  KESEALGGPGQVLICASDDRTCIQLREYITSGAEAFLLALYRKTFEK--DSKAEEVWMKF
Equus caballus - XP_001490470.2                     KESEALGGPGQVLICASDDRTCSQLREYIAVGAEAFLLRLYRKTFEK--DSKAEEVWMKF
Felis catus - XP_003998897.1                        KESEALGGPGQVLICASDERTCSQLREYITIGAEAFLLRLYRKTFER--DSKAEDVWVQF
Gorilla gorilla gorilla - XP_004057285.1            KESEALGGPGQVLICASDDRTCSQLRDYITLGAEAFLLRLYRKTFEK--DSKAEEVWMKF
Jaculus jaculus - XP_004652331.1                    KESEALGGPGQVLICASDDRTCSQLKDYLTAGAEAFLLRLYRKTFEK--DSKAEEVWMQL
Loxodonta africana - XP_003417763.1                 KESEALGGPGQVLICASDDRTCSQLRECITTGAEAFLLRLYRKSFEK--DSKAEEVWMKF
Macaca mulatta - XP_001107209.1                     KESEALGGPGQVLICASDDRTCSQLRDYVTLGAEAFLLRLYRKTFEK--DSKAEEVWMKF
Monodelphis domestica - XP_001375532.1              NNSEALGGPGQVLICASDDRTCAQLREYITIGAKAFLLKIYTKTFGK--DNKAGEVKVKF
Mus musculus - NP_056584.2                          KESEALGGPGRVLICASDDRTCCQLRDYLSAGAETFLLRLYRKTFEK--DGKAEEVWVNV
Nomascus leucogenys - XP_004091351.1                KESEALGGPGQVLICASDDRTCSQLRDYITLGAEAFLLRLYRKTFEK--DSKAEEVWMTF
Ochotona princeps - XP_004586829.1                  KQSEALGGPGQVLVCASDDRTCSQLKDYLTAGPEAFLLRLYRQAFEK--DSLAEEVWVRL
Octodon degus - XP_004644276.1                      KESEVLGGPGQVLICASDDRTCSQLKDYLTTGAEDFLLQLYRKTFEK--DSKAEEVWMKF
Odobenus rosmarus divergens - XP_004408974.1        KESEALGGPGQVLICASDDRTCSQLREYITVGAEAFLLRLYRKTFEK--DSKAEEVWMKF
Orcinus orca - XP_004270213.1                       KESEALGGPGHVLICASDDRTCSQLRECIAVGAEAFLLRLYRKTFEK--DSKAEDVWMRF
Oryctolagus cuniculus - XP_002711804.1              KESEVLGGPGQVLICASDDRTCSQLQGYLTAGAEAFLLGLYRQAFDK--DSRAEEVWARL
Otolemur garnettii - XP_003802775.1                 KESEALGGPGQVLICASDDRTCSQLRDYITLGTEAFLLRLYRKTFEK--DSKAKEVWMKF
Ovis aries - XP_004021225.1                         KESEVLGGPGQVLICASDDRTCSQLREYLTVGAEVFLLRLYRKTFEK--DSKAEDVWMRL
Pan paniscus - XP_003832737.1                       KESEALGGPGQVLICASDDRTCSQLRDYITLGAEAFLLRLYRKTFEK--DSKAEEVWMKF
Pan troglodytes - XP_510831.2                       KESEALGGPGQVLICASDDRTCSQLRDYITLGAEAFLLRLYRKTFEK--DSKAEEVWMKF
Papio anubis - XP_003916616.1                       KESEALGGPGQVLICASDDRTCSQLRDYITLGAEAFLLRLYRKTFEK--DSKAEEVWMKF
Pongo abelii - XP_002826187.1                       KESEALGGPGQVLICASDDRTCSQLRDYITLGAEAFLLRLYRKTYEK--DSKAEEVWMKF
Saimiri boliviensis boliviensis - XP_003928795.1    QESEALGGPGQVLICASDDRTCSQLRDYLTLGVEAFLLRLYRKTFEK--DSKAEEVWMKL
Sarcophilus harrisii - XP_003761738.1               NNSEALGGPGKVLICASDDRTCAQLREYLTIGAKTFLLKLYTKTFGK--D-KAGEVKMKF
Sorex araneus - XP_004604386.1                      QESEALGGPGQVLICASDDRTCTQLREYLSCGAEAFLLRLYRKTFEKEKDGPAEEDWARF
Trichechus manatus latirostris - XP_004373324.1     KESEALGGPGQVLICASDDRTCSQLREYITTGAEAFLLRLYRKTFEK--DSKAEEVWMKF
                                                    ::**.*****:**:****:*** **:  :: * : **: :* ::: :  *  * :
```

C.

```
Homo sapiens - NP_005227.1                          RKEDSSKRIRKSHKRPKDPQNKERASTKE-RTLKKKKRKLTLTQMVGKPE-ELEEEGDVE
Ailuropoda melanoleuca - XP_002918113.1             RHEDGPKRMAKSNKRPKDPQNKQRAATKE-RTLKKKKRRLTLTQMMGKSE-EPEEEGDVK
Bos taurus - NP_001192289.1                         RKEDDSKRIMKSSKRPKDLRDKDRPPAKE-KALRKKKPRLTLTQMMGRTE-ELEGEAGVE
Callithrix jacchus - XP_002755959.2                 RKEDSSNRIMKSHRRPKDPQNKERASTKQ-RTLKKKKRKLTLTQMVGKSE-ELEGEGDVE
Canis lupus familiaris - XP_536967.3                RHEDSPKRMARSNKRPKDPQNKQRAATKE-RTLKKKKRRLTLTQMVGKSE-EPDEEQDVK
Cavia porcellus - XP_003477812.1                    RKEDGSKRTIKSNKRPKDLQHKERGSIKE-RIPKKKKRRLTLTQMKGKSE-EVEEEEDVK
Ceratotherium simum simum - XP_004441913.1          RKEDTSKRRMKSTKRPKDPQNKERTSTKE-KTRKKKKRKLTLTQMMGKSE-ELAEEGDAK
Condylura cristata - XP_004692174.1                 PREDCARRTTKARRRPADRPP------AQGRALKKKPRGLTLTQMVGAAG---PAGGDAP
Cricetulus griseus - NP_001230961.1                 RKGDGPKRTMKSDKRPKDTKNKERASTKKGAPK-RKKRELTLTQVMGTAEEP-PEEGAAE
Dasypus novemcinctus - XP_004466678.1               RKEDNAKRRMKSNKRLKNPQNRERASTKE-RTLKKKRRRLTLTQMIGKSE-ELEAEPGIE
Echinops telfairi - XP_004705920.1                  RKEDRSRRAMRANKRPKGHPSLERASGKG-RPLKRQKQRLTLPQMMGHPE-ELEGQGEAE
Equus caballus - XP_001490470.2                     RQEDSSKRMAKSNKRPKDPHNKEKPSTKE-RTLRKKKRRLTLTQMIGKPE-ELEEEEDVK
Felis catus - XP_003998897.1                        RNEDSAKRMAKSHKRPKDPHSKQRAAPKE-RTLKKKKRRLTLTQMIGKPE-EPEEEGDAK
Gorilla gorilla gorilla - XP_004057285.1            RKEDSSKRIRKSHKRPKDPQNKERASTKE-RTLKKKKRKLTLTQMVGKPE-ELEEEGDVE
Jaculus jaculus - XP_004652331.1                    RKDDSSKRPVKSHKRPKDTHNKDWASNKD-KALKRKRRRLTLTQMMGNSR-ELEGEGDVE
Loxodonta africana - XP_003417763.1                 RKEDSSKRIMKPNKRPKDPQNKERASNKE-RTLKRKKRRLTLTQMMANSE-EHEGEGKVE
Macaca mulatta - XP_001107209.1                     RKEDSSKRIRKSHKRPKDPQNKERASAKE-RTLKKKKRKLTLTQMVGKPE-ELREEGDVE
Monodelphis domestica - XP_001375532.1              KKEDNSKGNQKSSKESKVSKAKVRTSK---R-PPKKKQELTLAQMITKIEDESE------
Mus musculus - NP_056584.2                          RKGDGPKRTTKSDKRPKAAPNKERASAKRGAPLKRKKQELTLTQVLGSAEEP-PEDKALE
Nomascus leucogenys - XP_004091351.1                RKEDSSKSIRKSHKRPKDPQNKEQSSTKE-RTLKKKKRKLTLTQMVGKPE-ELEEEGDVE
Ochotona princeps - XP_004586829.1                  RKEDSSRRTRRSQKRPKDPPVKERPA-KE-RAGRKKKPKLTLTQMMGRPEDELEEEGEAE
Octodon degus - XP_004644276.1                      RKEDGSKRTGKSNKRPKDLQHKEQSSLKE-RIPKKKKRRLTLTQMIGKSE-E---EEEVE
Odobenus rosmarus divergens - XP_004408974.1        RCEDRPKRMAKSNKRPKDPPNKQRAATKE-RTLKKKKRRLTLTQMMRQSE-EPEEAGDVK
Orcinus orca - XP_004270213.1                       RKEDGSKRIMKSNRRPKGLRNKDRASAKE-RAVKKKKPRLTLTQMMEQSE-ELEEEGGVE
Oryctolagus cuniculus - XP_002711804.1              RKADGSSRTRRSDKRPRARA------ARE-RALKKKKPKLTLTQMVGKPEE-PAEEGDVE
Otolemur garnettii - XP_003802775.1                 RKEDSSREIMKSKKRPKNPQNKERSSTKG-RTLKKKKRRLTLTQMIEKSE-ELEEDGDVE
Ovis aries - XP_004021225.1                         RKEDDSKRIMKSNKRPKDLQDKDRPPAKE-KALRKKKPRMTLTQMVGRAE-ALEGEAGAE
Pan paniscus - XP_003832737.1                       RKEDSSKRIRKPHKRPKDPQNKERASTKE-RTLKKKKRKLTLTQMVGKPE-ELEEEGDVE
Pan troglodytes - XP_510831.2                       RKEDSSKRIRKPHKRPKDPQNKERASTKE-RTLKKKKRKLTLTQMVGKPE-ELEEEGDVE
Papio anubis - XP_003916616.1                       RKEDSSKRIRKSHKRPKDPQNKERASAKE-RTLKKKKRKLTLTQMVGKPE-ELREEGDVE
Pongo abelii - XP_002826187.1                       RKEDSSKRIRKSHKRPKDPQNKERASTKE-RTLRKKKRKLTLTQMVGKPE-ELEEEGDVE
Saimiri boliviensis boliviensis - XP_003928795.1    RKEDSSKRIMKSHKRPKDPQNKERASTKE-RTLKKKKRKLTLTQMVGKSE-ELEEEGDVQ
Sarcophilus harrisii - XP_003761738.1               SKEDNSKGNLKSSKESKVSKAKIRTSK---KQRPTKKQELTLTQMVTKIEDEAE------
Sorex araneus - XP_004604386.1                      RKEDCARRGVKSGKGPKGRARAK-------ERTAKKKKKLTLTQMVAKTEEPEGEEGEAP
Trichechus manatus latirostris - XP_004373324.1     RKEDNSKRIMKSNKRPKDPQNKEKASTKE-RTLKRKKRRLTLTQMMGNSE-EHKGEEQVE
                                                       *      :  :                    :   :** *:
```

D.

```
Homo sapiens - NP_005227.1                          EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Ailuropoda melanoleuca - XP_002918113.1             EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Bos taurus - NP_001192289.1                         EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Callithrix jacchus - XP_002755959.2                 EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Canis lupus familiaris - XP_536967.3                EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Cavia porcellus - XP_003477812.1                    EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Ceratotherium simum simum - XP_004441913.1          EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Condylura cristata - XP_004692174.1                 EVEPRYVVLYDAELTFVRQLEIYRASRPGRPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Cricetulus griseus - NP_001230961.1                 EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Dasypus novemcinctus - XP_004466678.1               EMEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Echinops telfairi - XP_004705920.1                  EVEPRYVVLYDADLTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Equus caballus - XP_001490470.2                     EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Felis catus - XP_003998897.1                        EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Gorilla gorilla gorilla - XP_004057285.1            EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Jaculus jaculus - XP_004652331.1                    EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Loxodonta africana - XP_003417763.1                 EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Macaca mulatta - XP_001107209.1                     EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Monodelphis domestica - XP_001375532.1              EVEPRYVVLYDAELTFVRQLEIYRASRPGKPMRVYFLIYGGSTEEQRYLTALRKEKEAFE
Mus musculus - NP_056584.2                          EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Nomascus leucogenys - XP_004091351.1                EVEPRYVVLYDAELTFVRQXEIYRASRPGKPCRVYFLIYGGSTEEQRYLTALRKEKEAFE
Ochotona princeps - XP_004586829.1                  EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Octodon degus - XP_004644276.1                      EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Odobenus rosmarus divergens - XP_004408974.1        EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Orcinus orca - XP_004270213.1                       EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Oryctolagus cuniculus - XP_002711804.1              EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Otolemur garnettii - XP_003802775.1                 EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Ovis aries - XP_004021225.1                         EVEPRYVVLYDAELTFVRQLEIYRATRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Pan paniscus - XP_003832737.1                       EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Pan troglodytes - XP_510831.2                       EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Papio anubis - XP_003916616.1                       EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Pongo abelii - XP_002826187.1                       EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Saimiri boliviensis boliviensis - XP_003928795.1    EVEPRYVVLYDAELTFVRQLEIYRASRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Sarcophilus harrisii - XP_003761738.1               EVEPRYVVLYDAELTFVRQLEIYRASRPGKPMRVYFLIYGGSTEEQRYLTALRKEKEAFE
Sorex araneus - XP_004604386.1                      ELEPRYVVLYDAELTFVRQLEIYRASRPGRPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
Trichechus manatus latirostris - XP_004373324.1     EVEPRYVVLYDAELTFVRQLEIYRAGRPGKPLRVYFLIYGGSTEEQRYLTALRKEKEAFE
                                                    *:**********:****** ***** ***:* ****************************
```

E.

```
Homo sapiens - NP_005227.1                          SETLPESEKYNPGPQDFLLKMPGVNAKNCRSLMHHVKNIAELAALSQDELTSILGNAANA
Ailuropoda melanoleuca - XP_002918113.1             SDTLPESEKYNPGPQDFLLRMPGVNAKNCRTLMHHVKNIAELASLSQDKLAGILGNPGNA
Bos taurus - NP_001192289.1                         SETLPEAEKYNPGPQDFLLKMPGINAKNCRSLMNHVKNIAELASLPFDELASMLGNTASA
Callithrix jacchus - XP_002755959.2                 SETLPESEKYNPGPQDFLLKMPGVNAKNCRSLMHHVKSIAELATLSQEELTSILGNAANA
Canis lupus familiaris - XP_536967.3                SETLPESEKYNPGPQDFLLKMPGVNAKNCRALMHHVKNIAELASLSQDQLVGVLGNAGNA
Cavia porcellus - XP_003477812.1                    SEALPESEKYNPGPQDFLLKMPGVNAKNCHSLMNHVKNIAELATLSQDKLTSILGHAANA
Ceratotherium simum simum - XP_004441913.1          SETLPESEKYNPGPQDFLLKMPGVNAKNCRSLMSRVKNIAELASLSQDELAGILGNAAHA
Condylura cristata - XP_004692174.1                 SDTLPESDRYRPGPQDFLLKMPGVNAKNCHALMNQVGSLAELARLSQAELAAVLGNSANA
Cricetulus griseus - NP_001230961.1                 SETLPESDKYNPGPQDFVLKMPGINAKNCHSLMNHVKNIAELASLSQERLTSILGHAGNA
Dasypus novemcinctus - XP_004466678.1               SETLPEAEKYNPGPQDFLLKMPGVNAKNRRSLMNHVKNIAELATLSQDKLTSILGNAANA
Echinops telfairi - XP_004705920.1                  SATLPESEKYNPGPQDFLLKMPGISAKNCRALMTHVKNMVELATLSQERLTGILGNAANA
Equus caballus - XP_001490470.2                     SETLPESEKYNPGPQDFLLKMPGVNAKNCRSLMNHVKTIAELASLSQDQLAGVLGNAANA
Felis catus - XP_003998897.1                        SETLPESEKYNPGPQDFLLKMPGVNAKNCRALMHHVKNIAELASLSQDKLAGILGNASNA
Gorilla gorilla gorilla - XP_004057285.1            SETLPESEKYNPGPQDFLLKMPGVNAKNCRSLMHHVKNIAELAALSQDELTSILGNAANA
Jaculus jaculus - XP_004652331.1                    SETLPESEKYSPGPQDFLLKMPGVNAKNCRSLMNHVKNIAELATLPQDKLASILGNTGNA
Loxodonta africana - XP_003417763.1                 SETLPESEKYNPGPQDFLLKMPGVNAKNCRSLMNHVKNMAELATLSQDKLTSILGNAANA
Macaca mulatta - XP_001107209.1                     SETLPESEKYNAGPQDFLLKMPGVNAKNCRSLMHHVKNIAELATLSQDKLTSILGNAANA
Monodelphis domestica - XP_001375532.1              AETLPDSEKYNPGPQDFLLKMPGVNAKNCCSLMNHVKSIAELTTLSQDNLSSILGNAANA
Mus musculus - NP_056584.2                          SETLPESDRYNPGPQDFVLKMPGVNAKNCRSLMNQVKNIAELATLSLERLTTILGHSGNA
Nomascus leucogenys - XP_004091351.1                SEALPESEKYNPGPQDFLLKMPGVNAKNCRSLMHHVKNIAELTTLSQDKLASILGNAANA
Ochotona princeps - XP_004586829.1                  SETLPESDKYNPGPQDFLLKMPGVNAKNCRSLMNRVKNIAELASLSQEELTSILGNAANA
Octodon degus - XP_004644276.1                      SEALPESEKYNPGPQDFLLKMPGVNAKNCHSLMNHVKNIAELATLSQDKLTGILGHAANA
Odobenus rosmarus divergens - XP_004408974.1        SETLPESEKYNPGPQDFLLKMPGVNAKNCRPLMQHVQNIAELASLSQDKLAGILGNAGNA
Orcinus orca - XP_004270213.1                       SETLPEAEKYNPGPQDFLLKMPGVNAKNCRSLMTNIKNIAELASLSLDKLVLLLGNAANA
Oryctolagus cuniculus - XP_002711804.1              SESLPESEKYNPGPQDFLLKMPGVNAKNCRSVMNHVKNIAELATLSQDELARILGNAVNA
Otolemur garnettii - XP_003802775.1                 SETLPESDKYNPGPQDFLLKMPGVNAKNCRSLMNHVKNIAELATLSQDKLSSVLGNAANA
Ovis aries - XP_004021225.1                         SETLPEAEKYNPGPQDFLLKMPGINAKNCHSLMNHVKNIAELASLSFDKLASMLGNTASA
Pan paniscus - XP_003832737.1                       SETLPESEKYNPGPQDFLLKMPGVNAKNCRSLMHHVKNIAELAALSQDELTSILGNAANA
Pan troglodytes - XP_510831.2                       SETLPESEKYNPGPQDFLLKMPGVNAKNCRSLMHHVKNIAELAALSQDELTSILGNAANA
Papio anubis - XP_003916616.1                       SETLPESEKYNAGPQDFLLKMPGVNAKNCRSLMHHVKNIAELATLSQDKLTSILGNAANA
Pongo abelii - XP_002826187.1                       SETLPESEKYNPGPQDFLLKMPGVNAKNCRSLMHHVKNIAELATLSQDELTSILGNAANA
Saimiri boliviensis boliviensis - XP_003928795.1    SETLPESEKYNPGPQDFLLKMPGVNAKNCRSLMHHVKSIAELATLSQEELTSILGNAANA
Sarcophilus harrisii - XP_003761738.1               TETLPDSEKYNPGPQDFLLKMPGVNAKNCCSLMNHVKSIAELTTLSQDKLSSILGNAANA
Sorex araneus - XP_004604386.1                      SDTLPESDRYNPGPQDFLLRMPGVSAKNCRALMNRVRSIAELAGLSQAQLADVLGNAASA
Trichechus manatus latirostris - XP_004373324.1     SETLPESEKYNPGPQDFLLKMPGVNAKNCRSLMNHVKNIAELATLAQDKLTSILGNAANA
                                                    : :**:::::*  *****:*:***:.***   :* .: .:.**: *   .*  :**.   *
```

Appendix 25- Alignment of 300 base pairs in the 5'UTR of ERCC4 (5' of exon1 highlighted in grey; start of ORF highlighted in green). A heptamer sequence (TTCGGC(C/T)) is seen repeated ten times in Homo sapiens (yellow), with apparent varying conservation between several species. Rs1799797 was found in the final nucleotide of the penultimate repeat in humans (blue). A common key is given in Appendix 21. A list of common species names is given in Appendix 20.

```
Homo sapiens - NG_011442.1                          GTGACTC-----------CATGAATCTTCGGCTCCACTCGGCTCTTCTTCGGCTGAGTTC
Bos taurus - AC_000182.1                            CGGACTC-----------CACGAATCTTCGATTCTTCTAGGTTCCTCTTCGGCTGCGTTC
Callithrix jacchus - NC_013907.1                    GTGACTC-----------TACGAACCTTCGGCTCCACTCGGCTCCTCTTCGGCTGAGTTC
Dasypus novemcinctus - NW_004482259.1               GAATCTT-----------CGGCCCGCGCGCTTTCGGATTCGGCCTTCTTCGGCTTCCTTC
Equus caballus - NC_0091562                         GTGGCTC-----------TATTAATCTTCGGCTCTGCTCGGCTCCTCTTCGTCTGTGTTC
Gorilla gorilla gorilla - NC_018440.1               GTGACTC-----------TATGAATCTTCGGCTCCACTCGGCTCTTCTTCGGCTGAGTTC
Oryctolagus cuniculus - NC_013674.1                 CTCTTCG-----------GATCATCCTCTTCTCGGCCTCGGCTCCTTTTCGGCTATTTCC
Otolemur garnettii - NW_0038526652                  TTAACTTTTTCCGTTTTTGATGTCCTTCGGGTTACTCCCCTCTCCTTGGTGGATGGTGGC
Pan paniscus - NW_00870645.1                        GTGACTC-----------TATGAATCTTCGGCTCCACTCGGCTCTTCTTCGGCTGAGTTC
Pan troglodytes - NC_006483.3                       GTGACTC-----------TATGAATCTTCGGCTCCACTCGGCTCTTCTTCGGCTGAGTTC
Pongo abelii - NC_012607.1                          GTGACTC-----------TATGAATCTTCGGCTCCACTCGGCTCTTCTTCGGCTGAGTTC
Saimiri boliviensis boliviensis - NW_003943634.1    GTGACTC-----------TACGAACCTTCGGCTCTACTCGTCTTCTCTTCGGCTGAGTTC
Trichechus manatus latirostris - NW_004443960.1     GTGACTT-----------TGCGAATCTTTAGC---TCTCGGCTTCTCTTCGGCTGCGTTC
                                                                    .               .       *    * .*     *

Homo sapiens - NG_011442.1                          GGCCAACGCTTGCCTTCTCAGGCTCGGCTCTCTTCGGCTTTCTGAGAGCGC-GCCTGTCG
Bos taurus - AC_000182.1                            GGCCGGCTCCCGTCCTGTCGGGCTCGGCTCTCTTCGGCTACGTTCGGCGGC-TTTCGTCC
Callithrix jacchus - NC_013907.1                    GGCCGACGCTTGCCTCCTCAAGCTCGGCTCTCTTCGGCTTTCTGAGAGCAC-GCTCGTCC
Dasypus novemcinctus - NW_004482259.1               GGCCAGAGCGAGACTTCTCGGCCTCGGCTCTCTTCGGCCGTCTTTGGCCAGCGCTCGTCC
Equus caballus - NC_0091562                         GGCCAGCGCTCGCCTTCTCGGGCTCGGCTCTCTTCGGCTTCCTTCGGCCAG-CGCGCGCC
Gorilla gorilla gorilla - NC_018440.1               GGCCAACGCTTGCCTTCTCAGGCTCGGCTCTCTTCGGCTTTCTGAGAGCGC-GCCTGTCG
Oryctolagus cuniculus - NC_013674.1                 GGCTATCTCTCGTCTCCTCAGAACCGGCGCTCCTCGGCCGCCTTCGGTCTC--GCTCGTC
Otolemur garnettii - NW_0038526652                  AGGCGGAGGAGGTGTAG--AGCGCCCGAGGACTTGGAAGCACAGAAACAGGGTTCAGGCC
Pan paniscus - NW_00870645.1                        GGCCAACGCTTGCCTTCACAGGCTCGGCTCTCTTCGGCTTTCTGAGAGCGC-GCCTGTCG
Pan troglodytes - NC_006483.3                       GGCCAACGCTTGCCTTCTCAGGCTCGGCTCTCTTCGGCTTTCTGAGAGCGC-GCCTGTCG
Pongo abelii - NC_012607.1                          GGCCAACGCTCCCCTTCTCAGGTTCGGCTCTCTTCGGCTTTCTGAGAGCGC-GCCTGTCC
Saimiri boliviensis boliviensis - NW_003943634.1    GGCCGAAGCTTGCCTTCTCAGGCTCGGCTGTCTTCGGCTTTCTGAGAGCAA-GCTCGTCC
Trichechus manatus latirostris - NW_004443960.1     GGCCCGCGCTCGTCT---CAAACTCGGCTCTTTT--------------------------
                                                    .*    .           ..   * *.  :  *

Homo sapiens - NG_011442.1                          TGGCCTCGGCTCCTCTTCGGCTGCGTTCGGCCCACGAT----------------------
Bos taurus - AC_000182.1                            ------------------------------------------------------------
Callithrix jacchus - NC_013907.1                    TGGTCTCGGCTCCTCTTCGGCTGCGTTCGCCCCACGAT----------------------
Dasypus novemcinctus - NW_004482259.1               CCTGTTCGACTCCTCTTCGGCTTCCTTCGGCCGGCGCT----------------------
Equus caballus - NC_0091562                         TTC-------------------------------------------------------
Gorilla gorilla gorilla - NC_018440.1               TGGCCTCGGCTCCTCTTCGGCTGCGTTCGGCCCACGAT----------------------
Oryctolagus cuniculus - NC_013674.1                 AGGACTCGGCTCCTCTTCGGCTGCGTTCAGCCCGCCCTCTCGGGGTCGGCTTTCTTCGGG
Otolemur garnettii - NW_0038526652                  TGC--AGGAGACAACCGCGGGTCGG-TAGCTCCAGCAT------------------CGCG
Pan paniscus - NW_00870645.1                        TGGCCTCGGCTCCTCTTCGGCTGCGTTCGGCCCACGAT----------------------
Pan troglodytes - NC_006483.3                       TGGCCTCGGCTCCTCTTCGGCTGCGTTCGGCCCACGAT----------------------
Pongo abelii - NC_012607.1                          TGGCCTCGGCTCCTCTTCGGCTGCGTTCGGCCCACGAT----------------------
Saimiri boliviensis boliviensis - NW_003943634.1    TGGCCTCGGCTCCTCTTCGGCTGCATTCGCCCCACGAT----------------------
```

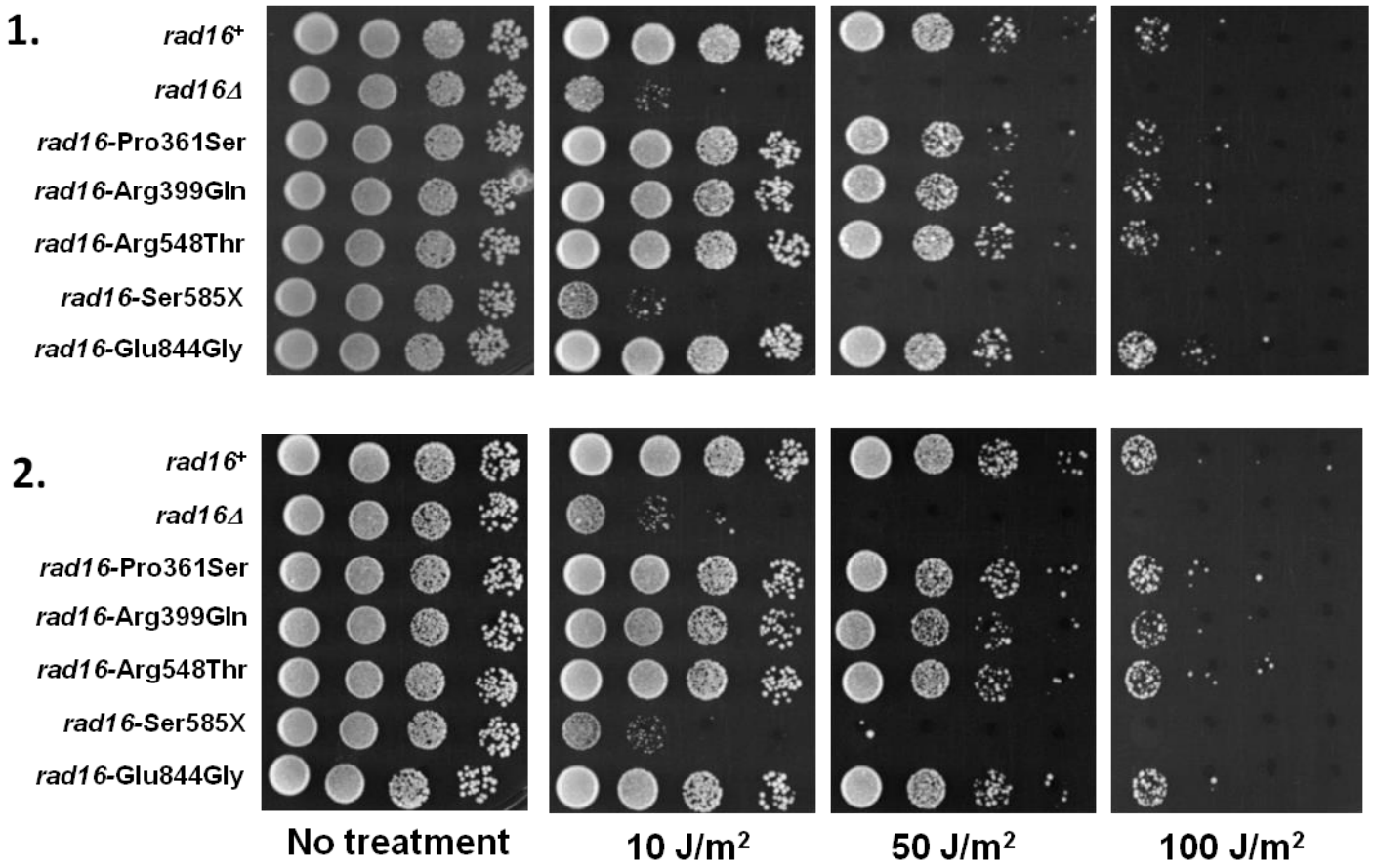
```
Trichechus manatus latirostris - NW_004443960.1     -----------------CGTCTGCGTTCGGCCGGTGCT----------------------
Homo sapiens - NG_011442.1                          ----CA----TCTCAGTCTCAGCTCTCCTCGGCTACGTTCGGCTGGCT-GCCGTCCTCTC
Bos taurus - AC_000182.1                            ---------------GGACTCGGCTCCCTT--------------------------CGGCC
Callithrix jacchus - NC_013907.1                    ----CG----TCTCAGTCTCGGCTCTCCTTGGCTTTTTTCGGCTGGAT-TTCGTCCTTTC
Dasypus novemcinctus - NW_004482259.1               ----CGTTTTTC--GGGCTCGGCTCTCTTCG---GCTATCTTCGGCCA-AAGCTCGTCTG
Equus caballus - NC_0091562                         ------------TCGGGCTCGGCTCTCTTCGGCTCCCTTCG---GCTT-ACGTTCGTCTC
Gorilla gorilla gorilla - NC_018440.1               ----CT----TCTCAGTCTCAGCTCTCCTCGGCTACGTTTGGCTGGCT-GCCGTCCTCTC
Oryctolagus cuniculus - NC_013674.1                 TGATTCTCGTCACGGCCTCGGCTCCTCTTCGGCTGCGGTCGGTCGGCTATCGTCGGATCC
Otolemur garnettii - NW_0038526652                  CGGAGGGCCTATGAATCTTCGGCTCCACTCGGCTCCACTCGGCT-----CCACTCGGCTC
Pan paniscus - NW_00870645.1                        ----CA----TCTCAGTCTCAGCTCTCCTCGGCTACGTTCGGCTGGCT-GCCGTCCTCTC
Pan troglodytes - NC_006483.3                       ----CA----TCTCAGTCTCAGCTCTCCTCGGCTACGTTCGGCTGGCT-GCCGTCCTCTC
Pongo abelii - NC_012607.1                          ----CA----TCTCAGTCTCGGCTCTCCTCGGCTACGTTCGGCTGGCT-GTCGTTCTCTC
Saimiri boliviensis boliviensis - NW_003943634.1    ----CG----TCTCAGTCTCGGCTCTCCACGGCTGCGTTCGACTATAT-TTCGTTGTCTC
Trichechus manatus latirostris - NW_004443960.1     ----CCACTTACCGGACCTGTTTTCTCTTCG---GCTCCGTTCGGCCT-ACGTTAGTCTC
                                                                  .         * . :

Homo sapiens - NG_011442.1                          GGACTCGGCTCTCTTCGGTTGAGTTCGGCCTAC---TCTCCACTAGGAGT----------
Bos taurus - AC_000182.1                            ACACTCGGCTCTCTTCGGCATGGTCCGGCCCAC---TCTCGTCTCAGACT----------
Callithrix jacchus - NC_013907.1                    GGACTCGGCTCTCTTCGGTTAAGTTCGGCCTAC---TCTCGTCTCGGACT----------
Dasypus novemcinctus - NW_004482259.1               GGCTCGGTTCCTCTTCGGCTCCCTTCGGCCGAC---GCTCGCCTCGGACT----------
Equus caballus - NC_0091562                         AGCCTCGGCTCTCTTCGGCTCCCCTCGGCTTAC---GCTCGTCCGGGTTT----------
Gorilla gorilla gorilla - NC_018440.1               GGACTCGGCTCTCTTCGGTTGAGTTCGGCCTAC---TCTGCACTGGGAGT----------
Oryctolagus cuniculus - NC_013674.1                 GACTCGGGTGCTCTTCGGCTGCGTTCGGCCCGT---GCTCGTCTCGGACT----------
Otolemur garnettii - NW_0038526652                  CACTCGGCTCCTCTTCGGCTGAGTTCGGCCGGCCTTCGTCCTCTCCGACTCGGCTTTCCT
Pan paniscus - NW_00870645.1                        GGACTCGGCTCTCTTCGGTTGAGTTCGGCCTAC---TCTCCACTCGGAGT----------
Pan troglodytes - NC_006483.3                       GGACTCGGCTCTCTTCGGTTGAGTTCGGCCTAC---TCTCCACTCGGAGT----------
Pongo abelii - NC_012607.1                          GGACTCGGCTCTCTTCGGTACAGTTCGGCCTAC---TCTCATCTCGGAGT----------
Saimiri boliviensis boliviensis - NW_003943634.1    GGACTCGGCTGTCTTCGGTTAAGTTCGGCCAAC---TCTCGTCTCGGACT----------
Trichechus manatus latirostris - NW_004443960.1     GGACTCGGCTCTCTTCGGTTGCGTTCGGCCCAC---GCTCGTCTCAGACT----------
                                                      .   *    ******* :     ****  .      *   *   *: *

Homo sapiens - NG_011442.1                          CGGCTTCCTTCGGCTGCGTTCGGCTGC-GACCCGGAAGAGCTTCCATG
Bos taurus - AC_000182.1                            CGGCTTCCTTCGGCTACGTTTCGCCGCCGCCCCGGAAGGGCGCCCATG
Callithrix jacchus - NC_013907.1                    CGGCTCCTTTCGGCTGCATTCGGCTGC-GCCCCGGAAGAGCCTCCATG
Dasypus novemcinctus - NW_004482259.1               CGGCTCTCTTCGGCTACGTTCGGCCGCGGCCCCGGAAGGGCGTGCATG
Equus caballus - NC_0091562                         CGGCTCTCTTCGGCTGCGTTCGGCCGCGGCCCCGGAAGGGCGCCCATG
Gorilla gorilla gorilla - NC_018440.1               CGGCTTCCTTCGGCTGCGTTCGGCTGC-GACCCGGAAGAGCTTCCATG
Oryctolagus cuniculus - NC_013674.1                 CG----------GCTCTCTCCGGCTGCGGCCCCGGAAGAGCGCCCATG
Otolemur garnettii - NW_0038526652                  CGGCGACGTTCGGCTGCGTTCG-GCGGCGCCCCGGAAGAGCTTCCATG
Pan paniscus - NW_00870645.1                        CGGCTTCCTTCGGCTGCGTTCGGCTGC-GACCCGGAAGAGCTTCCATG
Pan troglodytes - NC_006483.3                       CGGCTTCCTTCGGCTGCGTTCGGCTGC-GACCCGGAAGAGCTTCCATG
Pongo abelii - NC_012607.1                          CGGCTTCCTTCGGCTGCGTTCGGCCGC-GACCCGGAAGAGCTTCCATG
Saimiri boliviensis boliviensis - NW_003943634.1    CGG---------------------CTAC-GCCCCGGAAGAGCCTCCATG
Trichechus manatus latirostris - NW_004443960.1     CGGCTCTCTTCGGCTACGTTAGGCCGCGGACCCGGAAGAGCATCCATG
                                                    **                       .  *.********.**   ****
```

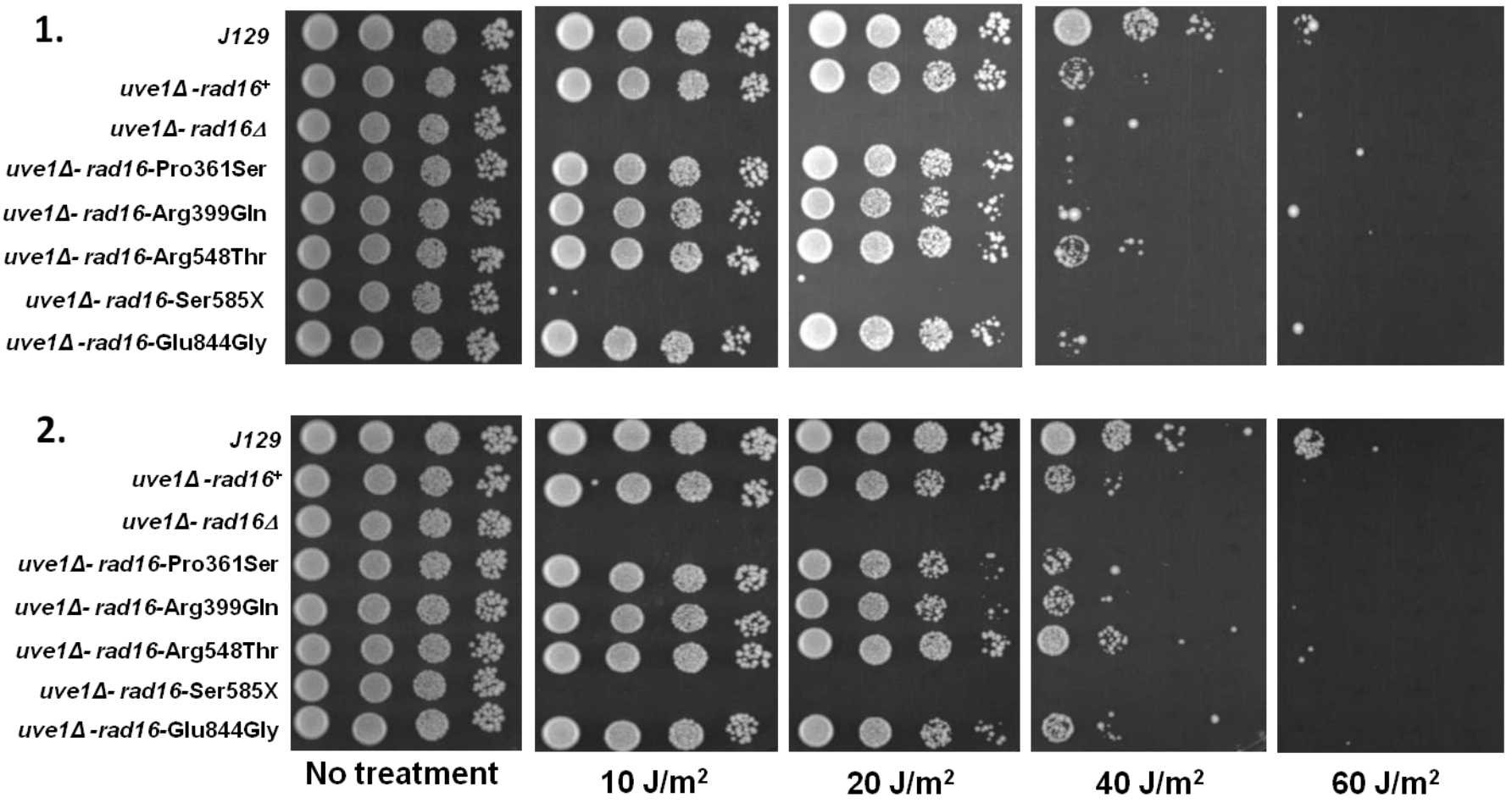
Appendix 26

Two additional repeats of UV spot test treatment of UVER proficient strains.



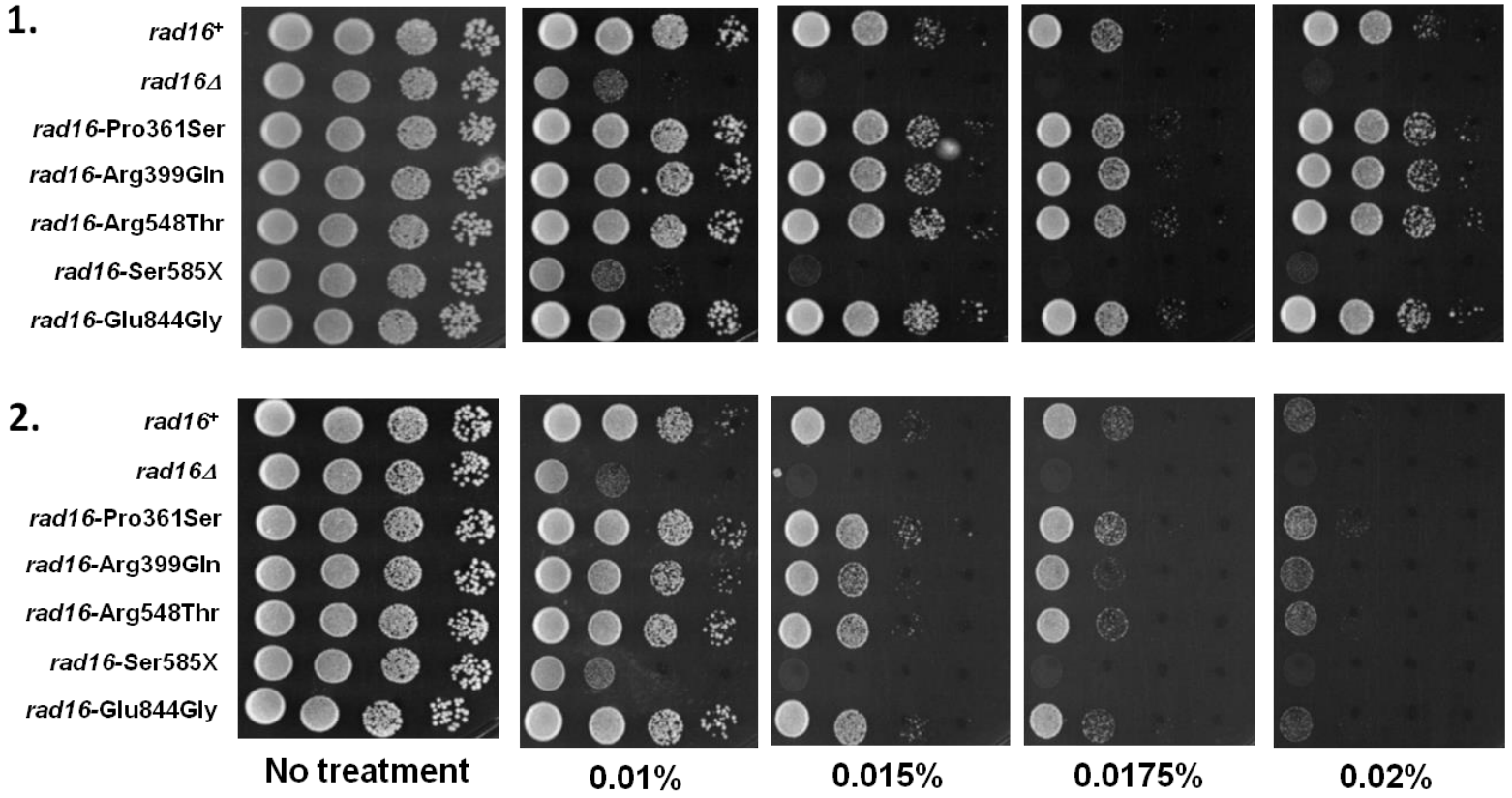
Appendix 27

Two additional repeats of UV spot test treatment of J129 and constructed uve1Δ-rad16 strains



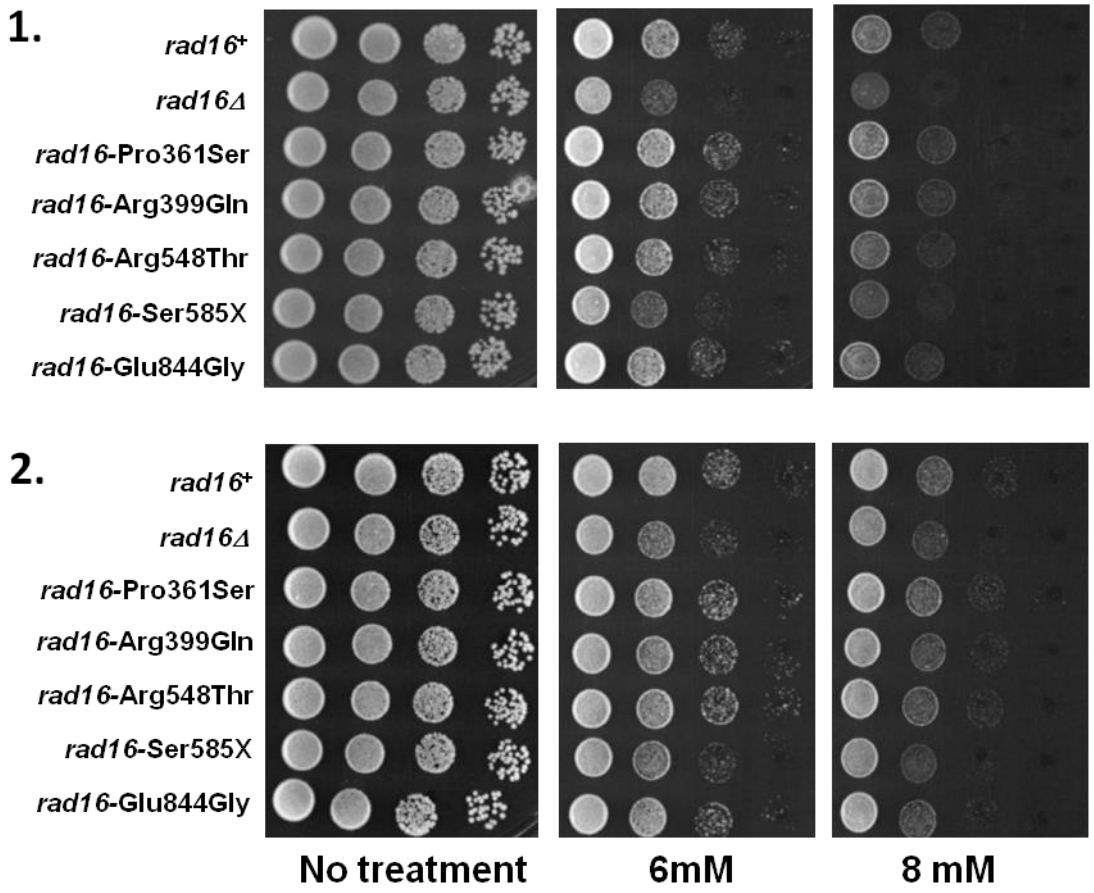
Appendix 28

Two additional repeats of MMS spot test treatment of UVER proficient strains. No treatment plates are the same as given in Appendix 26.



Appendix 29

Two additional repeats of HU spot test treatment of UVER proficient strains. No treatment plates are the same as given in Appendix 26.



Appendix 30

Percentage overall survival, normalised to rad16⁺, of strains treated with 1mM of oxaliplatin in comparison to survival of untreated control. Experiment 2 has been omitted due to outlying data.

		Experiment				
	Strains	**1**	**3**	**4**	**Average**	**SD**
Percentage survival normalised to *rad16⁺* (%)	***rad16⁺***	**100**	**100**	**100**	**100**	**0**
	rad16Δ	6.91	29.62	13.64	16.72	11.66
	rad16-Pro361Ser	55.17	50.27	33.16	46.20	11.56
	rad16-Arg399Gln	36.83	46.5	39.9	41.08	4.94
	rad16-Arg548Thr	39.37	28.73	43.1	37.07	7.46
	rad16-Ser585X	3.08	8.08	9.43	6.86	3.35
	rad16-Glu844Gly	72.03	58.17	48.32	59.51	11.91

Appendix 31

Percentage overall survival, normalised to uve1Δ-rad16⁺, of strains treated with 'dose one' of UV treatment (Table 7.1) in comparison to survival of untreated control plates.

		Experiment				
	Strains	**1**	**2**	**3**	**Average**	**SD**
Dose 1	***uve1Δ-rad16⁺***	**100**	**100**	**100**	**100**	**0**
	*uve1Δ-rad16*Δ	0.020	0.000	0.007	0.009	0.010
	uve1Δ-rad16-Pro361Ser	127.056	78.505	92.387	99.316	25.006
	uve1Δ-rad16-Arg399Gln	95.923	123.540	106.612	108.692	13.926
	uve1Δ-rad16-Arg548Thr	141.344	116.735	82.207	113.429	29.707
	uve1Δ-rad16-Ser585X	0.016	0.002	0.003	0.007	0.008
	uve1Δ-rad16-Glu844Gly	79.377	82.995	89.149	83.840	4.940

Appendix 32

Percentage overall survival, normalised to uve1Δ-rad16⁺, of strains treated with 'dose two' of UV treatment (Table 7.1) in comparison to survival of untreated control plates.

		Experiment				
	Strains	**1**	**2**	**3**	**Average**	**SD**
Dose 2	***uve1Δ-rad16⁺***	**100**	**100**	**100**	**100**	**0**
	*uve1Δ-rad16*Δ	0.131	0	0.005	0.045	0.074
	uve1Δ-rad16-Pro361Ser	195.889	38.113	118.540	117.514	78.893
	uve1Δ-rad16-Arg399Gln	144.957	57.624	61.185	87.922	49.426
	uve1Δ-rad16-Arg548Thr	99.667	105.492	71.272	92.143	18.309
	uve1Δ-rad16-Ser585X	0.029	0.002	0.003	0.011	0.015
	uve1Δ-rad16-Glu844Gly	87.556	40.598	21.738	49.964	33.894

Appendix 33

Dunlop MG, Dobbins SE, Farrington SM, Jones AM, Palles C, Whiffin N, Tenesa A, Spain S, Broderick P, Ooi LY, Domingo E, Smillie C, Henrion M, Frampton M, Martin L, Grimes G, Gorman M, Semple C, Ma YP, Barclay E, Prendergast J, Cazier JB, Olver B, Penegar S, Lubbe S, Chander I, Carvajal-Carmona LG, Ballereau S, Lloyd A, Vijayakrishnan J, Zgaga L, Rudan I, Theodoratou E; Colorectal Tumour Gene Identification (CORGI) Consortium, Starr JM, Deary I, Kirac I, Kovacević D, Aaltonen LA, Renkonen-Sinisalo L, Mecklin JP, Matsuda K, Nakamura Y, Okada Y, Gallinger S, Duggan DJ, Conti D, Newcomb P, Hopper J, Jenkins MA, Schumacher F, Casey G, Easton D, Shah M, Pharoah P, Lindblom A, Liu T; Swedish Low-Risk Colorectal Cancer Study Group, Smith CG, **West H**, Cheadle JP; COIN Collaborative Group, Midgley R, Kerr DJ, Campbell H, Tomlinson IP, Houlston RS (2012). Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nat Genet*.; 44(7): pp 770-776

Appendix 34

Smith CG, **West H**, Harris R, Idziaszczyk S, Maughan TS, Kaplan R, Richman S, Quirke P, Seymour M, Moskvina V, Steinke V, Propping P, Hes FJ, Wijnen J, Cheadle JP (2013). Role of the Oxidative DNA Damage Repair Gene OGG1 in Colorectal Tumorigenesis. *J Natl Cancer Inst ;* 105(16): pp 1249-1253

References

Abigerges D *et al.* (1994). Irinotecan (CPT-11) high-dose escalation using intensive high-dose loperamide to control diarrhoea. *J Natl Cancer Inst;* 86(6): pp 446-449.

Adams RA *et al*. (2009). Toxicity associated with combination oxaliplatin plus fluoropyrimidine with or without cetuximab in the MRC COIN trial experience. *Br J Cancer*; 100(2): pp 251-258.

Adeji AA. (1999). A review of the pharmacology and clinical activity of new chemotherapy agents for the treatment of colorectal cancer. *Br J Clin Pharmacol*; 48(3): pp 265–277.

Advanced Colorectal Cancer Meta-Analysis Project. (1992). Modulation of fluorouracil by leucovorin in patients with advanced colorectal cancer: evidence in terms of response rate. *J Clin Oncol*; 10(6): pp 896-903.

Adzhubei IA *et al*. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, 7(4): pp 248-249.

Agero AL *et al.* (2006). Dermatologic side effects associated with the epidermal growth factor receptor inhibitors. *J Am Acad Dermatol;* 55(4): pp 657-670.

Ahmad A *et al*. (2008). ERCC1-XPF endonuclease facilitates DNA double-strand break repair. *Mol Cell Biol*; 28(16): pp 5082-5092.

Ahmad A *et al*. (2010). Mislocalization of XPF-ERCC1 nuclease contributes to reduced DNA repair in XP-F patients. *PLoS Genet*; 6(3): e1000871.

Alberts B *et al.* (2002). Molecular biology of the cell, 4th Edition. New York: *Garlend Science.*

Albert TJ *et al*. (2007). Direct selection of human genomic loci by microarray hybridization. *Nat Methods*; 4(11): pp 903-905.

Alcindor T and Beauger N. (2011). Oxaliplatin: a review in the era of molecularly targeted therapy. *Curr Oncol*; 18(1): pp 18-25.

Alexander *et al.* (2001). Homologous recombination as a mechanism for carcinogenesis. *Biochimica et Biophysica Acta*; 1471: pp 109-121

Allione A *et al.* (2013). Inter-individual variation in nucleotide excision repair pathway is modulated by non-synonymous polymorphisms in ERCC4 and MBD4 genes. *Mutat Res.*

Al-Minawi AZ *et al*. (2009). The ERCC1/XPF endonuclease is required for completion of homologous recombination at DNA replication forks stalled by inter-strand cross-links. *Nucleic Acids Res*; 37(19): pp 6400-6413.

Al-Tassan N *et al.* (2002). Inherited variants of MYH associated with somatic G:C → T:A mutations in colorectal tumors. *Nat Genet*; 30(2): pp 227-232.

Alter BP. (2003). Cancer in Fanconi anaemia. *Cancer*; 97(2): pp 425-440.

Amado RG *et al.* (2008). Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J Clin Oncol*; 26(10): pp 1626-1634.

Ando M *et al.* (2005). Pharmacogenetics of irinotecan: a promoter polymorphism of UGT1A1 gene and severe adverse reactions to irinotecan. *Invest New Drugs*; 23(6): pp 539-545.

André T *et al.* (2004). Oxaliplatin, fluorouracil and leucovorin as adjuvant treatment for colorectal cancer. *N Engl J Med*; 350(23): pp 2343-2351.

André T *et al.* (2013). Panitumumab combined with irinotecan for patients with KRAS wild-type metastatic colorectal cancer refractory to standard chemotherapy: a GERCOR efficacy, tolerance, and translational molecular study. *Ann Oncol*; 24(2): pp 412-419.

Anttinen A *et al.* (2008). Neurological symptoms and natural course of xeroderma pigmentosum. *Brain*; 131(Pt 8): pp 1979-1989.

Arai K *et al.* (1997). Cloning of a human homolog of the yeast OGG1 gene that is involved in the repair of oxidative DNA damage. *Oncogene*; 14(23): pp 2857-2861.

Arango G *et al.* (2004). Molecular mechanisms of action and prediction of response to oxaliplatin in colorectal cancer cells. *Br J Cancer*; 91(11): pp 1931-1946.

Aretz S *et al*. (2005). High proportion of large genomic STK11 deletions in Peutz-Jeghers syndrome. *Hum Mutat*; 26(6): pp 513-519.

Argyriou AA *et al*. (2008). A review on oxaliplatin-induced peripheral nerve damage. *Cancer Treat Rev*; 34(4): pp 368-377.

Argyriou AA *et al*. (2013). Voltage-gated sodium channel polymorphisms play a pivotal role in the development of oxaliplatin-induced peripheral neurotoxicity: Results from a prospective multicenter study. *Cancer*.

Arnould S *et al*. (2003). Cellular determinants of oxaliplatin sensitivity in colon cancer cell lines. *Eur J Cancer*; 39(1): pp 112-119.

Arora S *et al*. (2010). Downregulation of XPF-ERCC1 enhances cisplatin efficacy in cancer cells. *DNA Repair (Amst)*; 9(7): pp 745-753.

Asgari MM *et al.* (1999). Expression and localization of thymidine phosphorylase/platelet-derived endothelial cell growth factor in skin and cutaneous tumors. *J Cutan Pathol*; 26(6): pp 287-294.

Athma P *et al.* (1999). Molecular genotyping shows that ataxia-telangiectasia heterozygotes are predisposed to breast cancer. *Cancer Genet Cytogenet*; 92(2): pp 130-134.

Auerbach AD. (1988). A test for Fanconi's anaemia. *Blood;* (72): pp 366-367.

Augusto-Pinto L *et al.* (2003). *Escherichia coli* as a model system to study DNA repair genes of eukaryotic organisms. *Genet Mol Res*; 2(1): pp 77-91.

Avery AM *et al.* (1999). Substrate specificity of ultraviolet DNA endonucleace (UVDE/Uve1p) from *Schizosaccharomyces pombe. Nucleic Acids Research;* 27(11): pp 2256-2264

Azzopardi D *et al.* (2008). Multiple rare nonsynonymous variants in the adenomatous polyposis coli gene predispose to colorectal adenomas. *Cancer Res*; 68(2): pp 358-363.

Baba H *et al*. (2012). Up regulation of ERCC1 and DPD expressions after oxaliplatin-based first-line chemotherapy for metastatic colorectal cancer. *Br J Cancer*; 107(12): pp 1950-1955.

Bähler J *et al*. (1998). Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. *Yeast*; 14(10): pp 943-951.

Bajetta E *et al*. (2004). Randomized multicentre Phase II trial of two different schedules of irinotecan combined with capecitabine as first-line treatment in metastatic colorectal carcinoma. *Cancer*; 100(2): pp 279-287.

Bakker ST *et al.* (2013). Learning from a paradox: recent insights into Fanconi anaemia through studying mouse models. *Dis Model Mech*; 6(1): pp 40-47.

Balaguer F *et al.* (2008). Identification of MYH mutation carriers in colorectal cancer: a multi-centre, case-control, population-based study. *Clin Gastroenterol Hepatol*; 5(3): pp 379-387.

Baldwin F and Sran H. (2010). Delayed ethylene glycol poisoning presenting with abdominal pain and multiple cranial and peripheral neuropathies: a case report. *J Med Case Rep*; 4: pp 220.

Ballinger AB and Anggiansah C. (2007). Colorectal cancer. *BMJ*; 335(7622): pp 715-718.

Bamshad MJ *et al.* (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet;* 12(11): pp 745-755.

Bao S *et al.* (2011). Evaluation of next-generation sequencing software in mapping and assembly. *J Hum Genet*; 56(6): pp 406-414.

Barack BR and Burgdorf WHC. (1991). Chemotherapy induced acral erythema. *J Am Acad Dermatol;* 24(3): pp 457-461.

Barrett JC *et al*. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*; 21(2): pp 263-265.

Bartek J and Lukas J. (2003). Chk1 and Chk2 kinases in checkpoint control and cancer. *Cancer Cell*; 3(5): pp 421-429.

Bartkova J *et al.* (2005). DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature*; 434(7035): pp 864-870.

Baumann P *et al.* (1996). Human Rad51 protein promotes ATP-dependent homologous pairing and strand transfer reactions *in vivo*. *Cell*; 87(4): pp 757-766.

Beckman JS and Weber JL. (1992). Survey of human and rat microsatellites. *Genomics*; 12(4): pp 627-631.

Beckman KB and Ames BN. (1997). Oxidative decay of DNA. *J Biol Chem*; 272(32): pp 19633-19636.

Ben-Omran TI *et al.* (2005). A patient with mutations in DNA ligase IV: clinical features and overlap with Nijmegen breakage syndrome. *Am J Med Genet A*; 137A(3): pp 283-287.

Benvenuti S *et al.* (2007). Oncogenic activation of the RAS/RAF signaling pathway impairs the response of metastatic colorectal cancers to anti-epidermal growth factor receptor antibody therapies. *Cancer Res*; 67(6): pp 2643-2648.

Beretta GD *et al.* (2013). FOLFIRI + bevacizumab as second-line therapy for metastatic colorectal cancer pretreated with oxaliplatin: a pooled analysis of published trials. *Med Oncol*; 30(1): pp 486

Berlin J *et al.* (2007). Panitumumab with irinotecan/leucovorin/5-fluorouracil for first-line treatment of metastatic colorectal cancer. *Clin Colorectal Cancer*; 6(6): pp 427-432.

Bernstein JL *et al.* (2006). The CHEK2*1100delC allelic variant and risk of breast cancer: screening results from the Breast Cancer Family Registry. *Cancer Epidemiol Biomarkers Prev*; 15(2): pp 348-352.

Béroud C and Soussi T. (1996). APC gene: database of germline and somatic mutations in human tumors and cell lines. *Nucleic Acids Res*; 24(1): pp 121-124.

Bhutto AM and Kirk SH. (2008). Population distribution of xeroderma pigmentosum. *Adv Exp Med Biol*; 637: pp 138-143.

Bienz M and Clevers H. (2000). Linking colorectal cancer to Wnt signalling. *Cell*; 103(2): pp 311-320.

Bilgüvar K *et al.* (2010). Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature*; 467(7312): pp 207-210.

Birnboim HC and Nasim A. (1975). Excision of pyrimidine dimers by several UV-sensitive mutants of S.pombe. *Mol Gen Genet*; 136(1): pp 1-8.

Blumenthal GM and Dennis PA. (2008). PTEN hamartoma tumor syndromes. *Eur J Hum Genet*; 16(11): pp 1289-1300.

Bodmer WF *et al.* (1987) Localisation of the gene for familial adenomatous polyposis on chromosome 5. *Nature*; 328(6131): pp 614-616.

Bodmer WF and Bonilla C. (2008). Common and rare variants in multifactorial susceptibility to common disease. *Nat Genet*; 40(6): pp 695-701

Bøe CA *et al*. (2012). Induction of a G1-S checkpoint in fission yeast. *Proc Natl Acad Sci USA*; 109(25): pp 9911-9916.

Bogliolo M *et al*. (2013). Mutations in ERCC4, Encoding the DNA-Repair Endonuclease XPF, Cause Fanconi Anemia. *Am J Hum Genet*; 92(5): pp 800-806.

Bokemeyer C *et al.* (2009). Fluorouracil, leucovorin and oxaliplatin with and without cetuximab in the first line treatment of metastatic colorectal cancer. *J Clin Oncol*; 27(5): pp 663-671.

Bokemeyer C *et al.* (2011). Efficacy according to biomarker status of cetuximab plus FOLFOX-4 as first-line treatment for metastatic colorectal cancer: the OPUS study. *Ann Oncol*; 22(7): pp 1535-1546.

Boland CR *et al.* (1998). A national cancer institute workshop on microsatelitte instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatelitte instability in colorectal cancer. *Cancer Res*; 58(22): pp 5248-5257.

Boland CR and Goel A, (2010). Microsatellite instability in colorectal cancer. *Gastroenterology*; 138(6): pp 2073-2087

Bost B *et al.* (2001). Genetic and non-genetic bases for the L-shaped distribution of quantitative trait loci effects. *Genetics*; 157(4): pp 1773-1787.

Brabec V and Kasparkova J. (2005). Modifications of DNA by platinum complexes. Relation to resistance of tumors to platinum antitumor drugs. *Drug Resist Updat*; 8(3): pp 131-146.

Broderick P *et al.* (2007). A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat Genet*; 39(11): pp 1315-1317.

Brookman KW *et al*. (1996). ERCC4 (XPF) encodes a human nucleotide excision repair protein with eukaryotic recombination homologs. *Mol Cell Biol*; 16(11): pp 6553-6562.

Brosh RM and Bohr VA. (2007). Human premature aging, DNA repair and RecQ helicases. *Nucleic Acids Res*; 35(22): pp 7527-7544.

Broyl A *et al*. (2010). Mechanisms of peripheral neuropathy associated with bortezomib and vincristine in patients with newly diagnosed myeloma: a prospective analysis of data from the HOVON-65/GMMG-HD4 trial. *Lancet Oncol*; 11(11): pp 1057-1065.

Burt RW *et al.* (2004). Genetic testing and phenotype in a large kindred with attenuated familial adenomatous polyposis. *Gastroenterology*; 127(2): pp 444-451.

Burtelow MA *et al*. (2001). Reconstitution and molecular analysis of the hRad9-hHus1-hRad1 (9-1-1) DNA damage responsive checkpoint complex. *J Biol Chem*; 276(28): pp 25903-25909.

Cabral RE *et al.* (2008). Identification of new *RECQL4* mutations in Caucasian Rothmund-Thomson patients and analysis of sensitivity to a wide range of genotoxic agents. *Mutat Res*; 643(1-2): pp 41-47.

Calva D and Howe JR, (2008). Hamartomatous polyposis syndromes. *Surg Clin North Am*; 88(4): pp 779–vii.

Cancer Research UK. (2010). Bowel cancer statistics. Retrieved March 14, 2012, from Cancer Research UK: http://www.cancerresearchuk.org/cancer-info/cancerstats/types/bowel/

Cancer Research UK. (2012, March 1). Bowel cancer survival statistics. Retrieved April 16, 2013, from Cancer Research UK: http://www.cancerresearchuk.org/cancer-info/cancerstats/types/bowel/survival/

Cancer Reasearch UK. (2012, March 1). TNM and number stages of bowel cancer. Retrieved July 29, 2013, from Cancer Research UK: http://www.cancerresearchuk.org/cancer-help/type/bowel-cancer/treatment/tnm-and-number-stages-of-bowel-cancer

Carney JP *et al.* (1998). The hMre11/hRad50 protein complex and Nijmegen breakage syndrome: linkage of double-strand break repair to the cellular DNA damage response. *Cell*; 93(3): pp 477-486.

Carr AM *et al*. (1994). The rad16 gene of Schizosaccharomyces pombe: a homolog of the rad1 gene of Saccharomyces Cerevisiae. *Mol Cell Biol*; 14(3): pp 2029-2040.

Carvajal-Carmona LG *et al*. (2011). Fine-mapping of colorectal cancer susceptibility loci at 8q23.3, 16q22.1 and 19q13.11: refinement of association signals and use of in silico analysis to suggest functional variation and unexpected candidate target genes. *Hum Mol Genet*; 20(14): pp 2879-2888.

Cassidy J *et al.* (2002). First line oral capecitabine therapy in metastatic colorectal cancer: a favourable safety profile compared with intravenous 5-fluorouracil/leucovorin. *Ann Oncol*; 13(4): pp 566-575.

Cassidy J *et al*. (2004). XELOX (capecitabine plus oxaliplatin): active first-line therapy for patients with metastatic colorectal cancer. *J Clin Oncol*; 22(11): pp 2084-2091.

Cassidy J *et al.* (2008). Randomised phase III study of capecitabine plus oxaliplatin compared with fluorouracil/folinic acid plus oxaliplatin as first line therapy for metastatic colorectal cancer. *J Clin Oncol*; 26(12): pp 2006-2012.

Castro E *et al.* (2013). Germline *BRCA* mutations are associated with higher risk of nodal involvement, distant metastasis, and poor survival outcomes in prostate cancer. *J Clin Oncol*; 31(14): pp 1748-1757.

Ceckova M *et al.* (2008). Effect of ABCG2 on cytotoxicity of platinum drugs: interference of EGFP. *Toxicol In Vitro*; 22(8): pp 1846-1852.

Cepeda V *et al.* (2007). Biochemical mechanisms of cisplatin cytotoxicity. *Anticancer Agents Med Chem*; 7(1): pp 3-18.

Chaganti RSK *et al.* (1974). A many-fold increase in sister chromatid exchanges in Blooms syndrome. *Proc Natl Acad Sci USA*; 71(11): pp 4508-4512.

Chang DY and Lu AL. (2005). Interaction of checkpoint proteins Hus1/Rad1/Rad9 with DNA base excision repair enzyme MutY homolog in fission yeast, Schizosaccharomyces pombe*. J Biol Chem*; 280(1): pp 408-417.

Cheadle JP *et al.* (2002). Different combinations of biallelic *APC* mutation confer different growth advantages in colorectal tumours. *Cancer Res*; 62(2): pp 363-366.

Cheadle JP and Sampson JR. (2003). Exposing the MYtH about base excision repair and human inherited disease. *Hum Mol Genet*; 12 Spec No 2: pp R159-165

Cheadle JP and Sampson JR. (2007). MUTYH-associated polyposis - from defect in base exicision repair to clinical genetic testing. *DNA Repair (Amst)*; 6(3): pp 274-279.

Chen L *et al.* (2000). Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the DNA-dependent protein kinase. *J Biol Chem*; 275(34): pp 26196-26205.

Cheng WH *et al.* (2006). Collaboration of Werner syndrome protein and BRCA1 in cellular responses to DNA interstrand cross-links. *Nucleic Acids Res*; 34(9): pp 2751-2760.

Choi M *et al.* (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A*; 106(45): pp 19096-19101.

Chung CH *et al.* (2008). Cetuximab-induced anaphylaxis and IgE specific for galactose-alpha-1,3-galactose. *N Engl J Med*; 358(11): pp 1109-1117.

Cipak L *et al*. (2006). The role of BRCA2 in replication-coupled DNA interstrand cross-link repair in vitro. *Nat Struct Mol Bio*; 13(8): pp 729-733.

Citterio E *et al*. (2000). ATP-dependent chromatin remodeling by the Cockayne syndrome B DNA repair-transcription-coupling factor. *Mol Cell Biol*; 20(20): pp 7643-7653.

Cleaver JE *et al.* (2001). Nucleotide excision repair "a legacy of creativity". *Mutat Res*; 485(1): pp 23-36.

Cohen JC *et al*. (2004). Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science*; 305(5685): pp 869-872.

Coin F *et al.* (2007). Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. *Mol Cell*; 26(2): pp 245-256.

Collie-Duguid ES *et al.* (2000). Known variant DPYD alleles do not explain DPD deficiency in cancer patients. *Pharmacogenetics*; 10(3): pp 217-223.

Constantinou A *et al.* (2000). Werner's syndrome protein (WRN) migrates Holliday junctions and co-localizes with RPA upon replication arrest. *EMBO Rep*; 1(1): pp 80-84.

Cooper GM and Shendure J. (2011). Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat Rev Genet*; 12(9): pp 628-640.

Corradetti MN *et al.* (2004). Regulation of the TSC pathway by LKB1: evidence of a molecular link between tuber sclerosis complex and Peutz-Jeghers syndrome. *Genes Dev*; 18(13): pp1533-1538.

Culy CR *et al.* (2000) Oxaliplatin. A review of its pharmacological properties and clinical efficacy in metastatic colorectal cancer and its potential in other malignancies. *Drugs;* 60(4): pp 895-924.

Cunningham D *et al.* (1998). Randomised trial of irinotecan plus supportive care versus supportive care alone after fluorouracil failure for patients with metastatic colorectal cancer. *Lancet*; 352(9138): pp 1413-1418.

Cunningham D *et al*. (2004). Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N Engl J Med*; 351(4): pp 337-345.

Curado MP *et al*. (2007). Cancer incidence in five continents. Vol IX. Lyon: *IARC Scientific Publications.*

Curtin K *et al*. (2009). Meta association of colorectal cancer confirms risk alleles at 8q24 and 18q21. *Cancer Epidemiol Biomarkers Prev*; 18(2): pp 616-621.

Daly AK. (2010). Genome-wide association studies in pharmacogenomics. *Nat Genet*; 11: pp 241-246.

Dan HC *et al.* (2004). Akt phosphorylation and stabilization of X-linked inhibitor of apoptosis protein (XIAP). *J Biol Chem*; 279(7): pp 5405-5412.

David SS *et al.* (2007). Base-excision repair of oxidative DNA damage. *Nature*; 447(7147): pp 941-950.

Davies H *et al*. (2002). Mutations of the BRAF gene in human cancer. *Nature*; 417(6892): pp 949-954

Deans AJ and West SC. (2011). DNA interstrand crosslink repair and cancer. *Nat Rev Cancer*; 11(7): pp 467-480.

Debeljak M *et al.* (2009). A patient with Baller-Gerold syndrome and midline NK/T lymphoma. *Am J Med Genet A*; 149A(4): pp 755-759.

de Boer J and Hoeijmaker JHJ. (2000). Nucleotide excision repair and human syndromes. *Carcinogenesis;* 21(3): pp 453-460.

de Gramont A *et al*. (2000). Leucovorin and fluorouracil with or without oxaliplatin as first-line treatment in advanced colorectal cancer. *J Clin Oncol*; 18(16): pp 2938-2947.

de Gramont A *et al.* (2004). OPTIMOX study: FOLFOX7/LV5FU2 compared to FOLFOX4 in patients with advanced colorectal cancer. *J Clin Oncol*; Meeting Abstracts vol. 22 no. 14_suppl 3525.

de Laat WL *et al*. (1998). Mapping of interaction domains between human repair proteins ERCC1 and XPF. *Nucleic Acids Res*; 26(18): pp 4146-4152.

De Roock W *et al.* (2008). KRAS wild-type state predicts survival and is associated to early radiological response in metastatic colorectal cancer treated with cetuximab. *Ann Oncol*; 19(3): pp 508-515.

De Roock W *et al.* (2010). Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. *Lancet Oncol*; 11(8): pp 753-762.

Desoize B and Madoulet C. (2002). Particular aspects of platinum compound used at present in cancer treatment. *Crit Rev Oncol Hematol*; 42(3): pp 317-325.

Dianov GL *et al.* (2003). Repair of abasic sites in DNA. *Mutat Res*; 531(1-2): pp 157-163.

Dickson SP *et al.* (2010). Rare variants create synthetic genome-wide association. *PLoS Biol*.

Diderich K *et al.* (2011). Premature aging and cancer in nucleotide excision repair-disorders. *DNA Repair (Amst)*; 10(7): pp 772-780.

Di Nicolantonio F *et al*. (2008). Wild-type BRAF is required for response to panitumumab or cetuximab in metastatic colorectal cancer. *J Clin Oncol*; 26(35): pp 5705-5712.

Donzelli E *et al*. (2004). Neurotoxicity of platinum compounds: comparison of the effects of cisplatin and oxaliplatin on the human neuroblastoma cell line SH-SY5Y. *J Neurooncol*; 67(1-2): pp 65-73.

Doré AS *et al.* (2009). Crystal structure of the rad9-rad1-hus1 DNA damage checkpoint complex--implications for clamp loading and regulation. *Mol Cell*; 34(6): pp 735-745.

Douillard JY *et al*. (2010). Randomized, phase III trial of panitumumab with infusional fluorouracil, leucovorin, and oxaliplatin (FOLFOX4) versus FOLFOX alone as first-line treatment in patients with previously untreated metastatic colorectal cancer: the PRIME study. *J Clin Oncol*; 28(31): pp 4697-4705.

Dronkert ML and Kanaar R. (2001). Repair of DNA interstrand crosslinks. *Mutat Res;* 486(4): pp 217-247.

Ducreux M *et al.* (2011). Capecitabine plus oxaliplatin (XELOX) versus 5-fluorouracil/leucovorin plus oxaliplatin (FOLFOX-6) as first-line treatment for metastatic colorectal cancer. *Int J Cancer*; 128(3): pp 682-690.

Dueland S *et al.* (2003). Epidermal growth factor receptor inhibition induces trichomegaly. *Acta Oncol*; 42(4): pp 345-346.

Dulak AM *et al.* (2013). Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet*; 45(5): pp 478-486.

Dunlop MG *et al.* (1997). Cancer risk associated with germline DNA mismatch repair gene mutations. *Hum Mol Genet;* 6(1): pp 105-110.

Dunlop MG *et al*. (2012a). Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nat Genet*; 44(7): pp 770-776.

Dunlop MG *et al.* (2012b). Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42 103 individuals. *Gut*; 62(6): pp 871-881.

Dutta A *et al*. (1993). Inhibition of DNA replication factor RPA by p53. *Nature*; 365(6441): pp 79-82.

Duval A and Hamelin R. (2002). Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability. *Cancer Res*; 62(9): pp 2447-2454.

Dzagnidze A *et al*. (2007). Repair capacity for platinum-DNA adducts determines the severity of cisplatin-induced peripheral neuropathy. *J Neurosci*; 27(35): pp 9451-9457.

Eastman A. (1987). The formation, isolation and characterisation of DNA adducts produced by anticancer platinum complexes. *Pharmacol Ther*; 34(2): pp 155-166.

Ebisawa T *et al*. (2001). Smurf1 interacts with transforming growth factor-beta type I receptor through Smad7 and induces receptor degradation. *J Biol Chem*; 276(16): pp 12477-12480.

Egel R. (2004). The molecular biology of Schizosaccharomyces pombe. *Springer.*

Eichler HG *et al.* (2011). Bridging the efficacy-effectiveness gap: a regulator's perspective on addressing variability of drug response. *Nat Rev Drug Discov*; 10(7): pp 495-506.

Eisen JA and Hanawalt PC. (1999). A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat Res*; 435(3): pp 171-213

Ellis NA *et al.* (1995). The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell*; 83(4): pp 655-666.

Enzlin JH and Schärer OD. (2002). The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. *EMBO J*; 21(8): pp 2045-2053.

Etienne-Grimaldi MC *et al*. (2010). Methylenetetrahydrofolate reductase (MTHFR) gene polymorphisms and FOLFOX response in colorectal cancer patients. *Br J Clin Pharmacol*; 69(1): pp 58-66.

Extra JM *et al*. (1998). Pharmacokinetics and safety profile of oxaliplatin. *Semin Oncol*; 25(2 Suppl 5): pp 13-22.

Faivre L *et al.* (2000). Association of complementation group and mutation type with clinical outcome in fanconi anaemia. European Fanconi Anemia Research Group. *Blood*; 96(13): pp 4064-4070.

Faivre S *et al*. (2003). DNA strand breaks and apoptosis induced by oxaliplatin in cancer cells. *Biochem Pharmacol*; 66(2): pp 225-237.

Fan W and Luo J. (2008). RecQ4 facilitates UV light-induced DNA damage repair through interaction with nucleotide excision repair factor xeroderma pigmentosum group A (XPA). *J Biol Chem*; 283(43): pp 29037-29044.

Fariña Sarasqueta A *et al.* (2011). Pharmacogenetics of oxaliplatin as adjuvant treatment in colon carcinoma: are single nucleotide polymorphisms in GSTP1, ERCC1, and ERCC2 good predictive markers? *Mol Diagn Ther*; 15(5): pp 277-283.

Farrington SM *et al.* (2005). Germline susceptibility to colorectal cancer due to base-excision repair gene defects. *Am J Hum Genet*; 77(1): pp 112–119.

Fearnhead NS *et al*. (2001). The ABC of APC. *Hum. Mol. Genet*; 10(7): pp 721-733.

Fearnhead NS *et al*. (2004). Multiple rare variants in different genes account for multifactorial inherited susceptibility to colorectal adenomas. *Proc Natl Acad Sci USA*; 101(45): pp 15992–15997.

Fearnhead NS *et al.* (2005). Rare variant hypothesis for multifactorial inheritance: susceptibility to colorectal adenomas as a model. *Cell Cycle*; 4(4): pp 521-525.

Fernandez-Rozadilla C *et al.* (2013). Pharmacogenomics in colorectal cancer: a genome-wide association study to predict toxicity after 5-fluorouracil or FOLFOX administration. *Pharmacogenomics J*; 13(3): pp 209-217.

Ferrara N *et al*. (2004). Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. *Nat Rev Drug Discov*; 3(5): pp 391-400.

Fink D *et al*. (1996). The role of DNA mismatch repair in platinum drug resistance. *Cancer Res*; 56(21): pp 4881-4886.

Fink D *et al.* (1998). Enrichment for DNA mismatch repair-deficient cells during treatment with cisplatin. *Int J Cancer*; 77(5): pp 741-746.

Fleck O *et al*. (1999). Involvement of nucleotide excision repair in msh2pms1-independent mismatch repair. *Nat Genet*; 21(3): pp 314-317.

Flicek P and Birney E. (2009). Sense from sequence reads: methods for alignment and assembly. *Nat Methods*; (11 Suppl):S6-S12.

Flicek P *et al.* (2013). Ensembl 2013. *Nucleic Acids Res*; 41(Database issue): pp D48-55.

Ford D *et al.* (1994). Risks of cancer in *BRCA1*-mutation carriers. Breast Cancer Linkage Consortium. *Lancet*; 343(8899): pp 692-695.

Ford D *et al.* (1998). Genetic heterogeneity and penetrance analysis of the *BRCA1* and *BRCA2* genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am J Hum Genet*; 62(3): pp 676-689.

Fousteri M *et al*. (2006). Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II in vivo. *Mol Cell*; 23(4): pp 471-482.

Fousteri M, Mullenders LH (2008). Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. *Cell Res*; 18(1): pp 73-84.

Frank KM *et al.* (2000). DNA ligase IV deficiency in mice leads to defective neurogenesis and embryonic lethality via the p53 pathway. *Mol Cell*; 5(6): pp 993-1002.

Frattini IM *et al*. (2007). PTEN loss of expression predicts cetuximab efficacy in metastatic colorectal cancer patients. *Br J Cancer*; 97(8): pp 1139-1145.

Frayling IM *et al*. (1998). The *APC* variants I1307K and E1317Q are associated with colorectal tumors, but not always with a family history. *Proc Natl Acad Sci USA*; 95(18): pp 10722-10727.

Freedman ML *et al*. (2004). Assessing the impact of population stratification on genetic association studies. *Nat Genet*; 36(4): pp 388-393.

Freeman DJ *et al*. (2008). Association of K-ras mutational status and clinical outcomes in patients with metastatic colorectal cancer receiving panitumumab alone. *Clin Colorectal Cancer*; 7(3): pp 184-190.

Freedman ML *et al.* (2011). Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet*; 43(6): pp 513-518.

Friedrich-Heineken E *et al*. (2005). The two DNA clamps Rad9/Rad1/Hus1 complex and proliferating cell nuclear antigen differentially regulate flap endonuclease 1 activity. *J Mol Biol*; 353(5): pp 980-989.

Fromme JC *et al*. (2004). Structural basis for removal of adenine mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. *Nature*; 427(6975): pp 652-656.

Fuchs CS *et al.* (2003). Phase III comparison of two irinotecan dosing regimens in second-line therapy of metastatic colorectal cancer. *J Clin Oncol*; 21(5): pp 807-814.

Fukuchi K *et al.* (1989). Mutator phenotype of Werner syndrome is characterized by extensive deletions. *Proc Natl Acad Sci USA*; 86(15): pp 5893-5897.

Galiatsatos P *et al*. (2006). Familial adenomatous polyposis. *Am J Gastroenterol*; 101: pp 385-398.

Gamelin L *et al*. (2007). Predictive factors of oxaliplatin neurotoxicity: the involvement of the oxalate outcome pathway. *Clin Cancer Res*; 13(21): pp 6359-6368.

Gammon A *et al*. (2009). Hamartomatous polyposis syndromes. *Best Pract Res Clin Gastroenterol*; 23(2): pp 219-231.

Gayther SA *et al.* (2000). The frequency of germ-line mutations in the breast cancer predisposition genes *BRCA1* and *BRCA2* in familial prostate cancer. The Cancer Research Campaign/British Prostate Group United Kingdom Familial Prostate Cancer Study Collaborators. *Cancer Res*; 60(16): pp 4513-4518.

German J. (1997). Bloom syndrome. XX. The first 100 cancer. *Cancer Genet Cytogenet*; 93(1): pp 100-106.

German J *et al.* (2007). Syndrome-causing mutations of the BLM gene in persons in the Bloom's Syndrome Registry. *Hum Mutat;* 28(8): pp 743-753.

Giantomio BJ *et al*. (2007). Bevacizumab in combination with oxaliplatin, fluorouracil, and leucovorin (FOLFOX4) for previously treated metastatic colorectal cancer: results from the Eastern cooperative oncology group study E3200. *JCO*; 25(12): pp 1539-1544.

Giardiello FM *et al*. (2000). Very high risk of cancer in familial Peutz-Jeghers syndrome. *Gastroenterology*; 119(6): pp 1447-1453.

Gilissen C *et al.* (2012). Disease gene identification strategies for exome sequencing. *Eur J Hum Genet*; 20(5): pp 490-497.

Gillet LC and Schärer OD. (2006). Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem Rev*; 106(2): pp 253-276.

Giovannini M *et al.* (2009). Clinical significance of skin toxicity due to EGFR-targeted therapies. *J Oncol.*

Giovannucci E. (2004). Alcohol, one carbon metabolism, and colorectal cancer: recent insights from molecular studies. *J Nutr*; 134(9): pp 2475-2481.

Girard PM *et al* (2004). Analysis of DNA ligase IV mutations found in LIG4 syndrome patients: the impact of two linked polymorphisms. *Hum Mol Genet*; 13(20): pp 2369-2376

Gismondi V *et al*. (2002). Prevalence of the E1317Q variant of the APC gene in Italian patients with colorectal adenomas. *Genet Test*; 6(4): pp 313-317.

Glazier AM *et al.* (2002). Finding genes that underlie complex traits. *Science*; 298(5602): pp 2345-2349.

Glazov EA *et al.* (2011). Whole-exome re-sequencing in a family quartet identifies POP1 mutations as the cause of a novel skeletal dysplasia. *PLoS Genet*; 7(3): e1002027.

Glenn TC. (2011). Field guide to next generation DNA sequencing. *Mol Ecol Resour*; 11(5): pp 759-769.

Goldberg RM *et al*. (2004). A randomised controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer. *J Clin Oncol*; 22(1): pp 23-30.

Goldmacher VS *et al*. (1986). Isolation and partial characterization of human cell mutants differing in sensitivity to killing and mutation by methylnitrosourea and N-methyl-N'-nitro-N-nitrosoguanidine. *J Biol Chem*; 261(27): pp 12462-12471.

Goldstein DB. (2009). Common genetic variation and human traits. *N Engl J Med*; 360(17): pp 1696-1698.

Gonzalez-Angulo AM *et al.* (2010). Future of personalized medicine in oncology: a systems biology approach. *J Clin Oncol*; 28(16): pp 2777-2783.

Goss JR *et al*. (2011). Premature aging-related peripheral neuropathy in a mouse model of progeria. *Mech Ageing Dev*; 132(8-9): pp 437-442.

Goto M *et al.* (1996). Excess of rare cancers in Werner syndrome (adult progeria). *Cancer epidimiology, biomarkers and prevention*; 5(4): pp 239-246.

Goujon M *et al*. (2010). A new bioinformatics analysis tools framework at EMBL-EBI *Nucleic Acids Res*; 38.

Graham MA *et al*. (2000). Clinical pharmacokinetics of oxaliplatin: A critical review. *Clin Cancer Res*; 6(4): pp 1205-1218.

Granger JP and Alexander BT. (2000). Abnormal pressure-natriuresis in hypertension: role of nitric oxide. *Acta Physiol Scand*; 168(1): pp 161-168.

Gray MD *et al.* (1997). The Werner syndrome protein is a DNA helicase. *Nat Genet*; 17(1): pp 100-103.

Gregg SQ *et al*. (2011). Physiological consequences of defects in ERCC1-XPF DNA repair endonuclease. *DNA Repair (Amst)*; 10(7): pp 781-791.

Gretarsdottir S *et al.* (1998). BRCA2 and p53 mutations in primary breast cancer in relation to genetic instability. *Cancer Res*; 58(5): pp 859-862.

Groen EJ *et al*. (2008). Extra-intestinal manifestations of familial adenomatous polyposis. *Ann Surg Oncol*; 15(9): pp 2439-2450.

Grolleau F *et al*. (2001). A possible explanation for a neurotoxic effect of the anticancer agent oxaliplatin on neuronal voltage-gated sodium channels. *J Neurophysiol*; 85(5): pp 2293-2297.

Grothey A. (2003). Oxaliplatin-safety profile: neurotoxicity. *Semin Oncol*; 30 (4 Suppl 15): pp 5-13.

Grothey A *et al.* (2004). Capecitabine/irinotecan (CapIri) and capecitabine/oxaliplatin (CapOx) are active second line protocols in patients with advanced colorectal cancer (aCRC) after failure of first line combination therapy: Results of a randomised phase II study. *J Clin Oncol* (Meeting Abstracts); vol. 22 no. 14_suppl 3534

Grothey A *et al*. (2005). Glutathione S-transferase P1 I105V (GSTP1 I105V) polymorphism is associated with early onset of oxaliplatin-induced neurotoxicity. *J Clin Oncol*; 23(16): suppl 3509.

Gu J *et al.* (2007). Single-stranded DNA ligation and XLF-stimulated incompatible DNA end ligation by the XRCC4-DNA ligase IV complex: influence of terminal DNA sequence. *Nucleic Acids Res*; 35(17): pp 5755-5762.

Guan X *et al*. (2007a). The human checkpoint sensor Rad9-Rad1-Hus1 interacts with and stimulates DNA repair enzyme TDG glycosylase. *Nucleic Acids Res*; 35(18): pp 6207-6218.

Guan X *et al.* (2007b). The human checkpoint sensor Rad9-Rad1-Hus1 interacts with and stimulates NEIL1 glycosylase. *Nucleic Acids Res*; 35(8): pp 2463-2472.

Gumy-Pause F *et al.* (2004). *ATM* gene and lymphoid malignancies. *Leukaemia*; 18(2): pp 238-242.

Gupta E *et al*. (1997). Modulation of glucuronidation of SN-38, the active metabolite of irinotecan, by valproic acid and phenobarbitone. *Cancer Chemother Pharmacol*; 39(5): pp 440-444.

Gurwitz D and McLeod HL. (2013). Genome-wide studies in pharmacogenomics: harnessing the power of extreme phenotypes. *Pharmacogenomics*; 14(4): pp 337-339. .

Hahnloser D *et al*. (2003). The APC E1317Q variant in adenomatous polyps and colorectal cancers. *Cancer Epidemiol Biomarkers Prev*; 12(10): pp 1023-1028.

Half E *et al*. (2009). Familial adenomatous polyposis. *Orphanet Journal of Rare Diseases*; 4(22): pp 1-23.

Hall JM *et al.* (1990). Linkage of early-onset familial breast cancer to chromosome 17q21. *Science*; 250(4988): pp 1684-1689.

Hall NR *et al.* (1994). Genetic linkage in Muir-Torre syndrome to the same chromosomal region as cancer family syndrome. *Eur J Cancer*; 30A(2): pp 180-182.

Hamilton SR *et al.* (1995). The molecular basis of Turcot's syndrome. *N Engl J Med*; 332(13): pp 839-847.

Han L *et al.* (2010). Mouse Rad1 deletion enhances susceptibility for skin tumor development. *Mol Cancer*; 9: pp 67.

Hanawalt PC. (2002). Subpathways of nucleotide excision repair and their regulation. *Oncogene*; 21(58): pp 8949-8956.

Handra-Luca A *et al*. (2005). Vessels' morphology in SMAD4 and BMPR1A-related juvenile polyposis. *Am J Med Genet A*; 138A(2): pp 113-117.

Hansen RM *et al.* (1996). Phase III study of bolus versus infusion fluorouracil with or without cisplatin in advanced colorectal cancer. *J Natl Cancer Inst*; 88(10): pp 668-674.

Hardy GH. (1908). Mendelian proportions in a mixed population. *Science*; 28(706): pp 49-50.

Harfe BD and Jinks-Robertson S (2000). DNA mismatch repair and genetic instability. *Annu Rev Genet*; 34: pp 359-399.

Hayashi T *et al.* (2004). Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. *Cell*; 118(6): pp 715-729.

Hecht JR, (1998). Gastrointestinal toxicity or irinotecan. *Oncology*; 12(8 Suppl 6): pp 72-78.

Hecht JR *et al*. (2007). Panitumumab monotherapy in patients with previously treated metastatic colorectal cancer. *Cancer*; 110(5): pp 980-988.

Heinemann V *et al*. (2009). Clinical relevance of EGFR- and KRAS-status in colorectal cancer patients treated with monoclonal antibodies directed against the EGFR. *Cancer Treat Rev*; 35(3): pp 262-271.

Hendriks YM *et al.* (2004). Cancer risk in hereditary nonpolyposis colorectal cancer due to *MSH6* mutations: impact on counselling and surveillance. *Gastroenterology*; 127(1): pp 17-25.

Henson J *et al.* (2012). Next-generation sequencing and large genome assemblies. *Pharmacogenomics*; 13(8): pp 901-915.

Hindorff LA *et al.* (2013, July 5). A Catalog of Published Genome-Wide Association Studies. Retrieved July 9, 2013, from www.genome.gov/gwastudies

Hirschhorn JN *et al*. (2002). A comprehensive review of genetic association studies. *Genet Med*; 4(2): pp 45-61.

Hoeijmakers JHJ. (1993a). Nucleotide excision repair I: from *E. coli* to yeast. *Trends Genet*; 9(5): pp 173-177.

Hoeijmakers JHJ. (1993b). Nucleotide excision repair II: from yeast to humans. *Trends Genet*; 9(6): pp 211-217.

Hoeijmakers JHJ. (2001). Genome maintenance mechanisms for prevention of cancer. *Nature*; 411(6835): pp 366-374.

Hoeijmakers JHJ. (2009). DNA Damage, Aging and Cancer. *N Engl J Med*; 361: pp 1475-1485.

Hoess RH *et al*. (1986). The role of the loxP spacer region in P-1 site-specific recombination. *Nucleic Acids Res*; 14(5): pp 2287-2300.

Hoischen A *et al.* (2010). De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet*; 42(6): pp 483-485.

Holliday R. (1964). A mechanism for gene conversion in fungi. *Genet Res*; 89(5-6): pp 285-307.

Holmes RP and Assimos DG. (1998). Glyoxylate synthesis, and its modulation and influence on oxalate synthesis. *J Urol*; 160(5): pp 1617-1624.

Holzer AK *et al*. (2006). Contribution of the major copper influx transporter CTR1 to the cellular accumulation of cisplatin, carboplatin and oxaliplatin. *Mol Pharmacol*; 70(4): pp 1390-1394.

Horie N *et al.* (1995). Functional analysis and DNA polymorphism of the tandemly repeated sequences in the 5'-terminal regulatory region of the human gene for thymidylate synthase. *Cell Struct Funct*.; 20(3): pp 191-197.

Hoskins JM *et al.* (2007). UGT1A1*28 genotype and irinotecan-induced neutropenia: dose matters. *J Natl Cancer Inst*; 99(17): pp 1290-1295

Houlston RS *et al.* (1998). Mutations in DPC4 (SMAD4) cause juvenile polyposis syndrome, but only account for a minority of cases. *Hum Mol Genet*; 7(12): pp 1907-1912.

Houlston RS *et al*. (2008). Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet*; 40(12): pp 1426-1435

Houlston RS *et al.* (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(11): pp 973-977

Howe JR *et al.* (1998). Mutations in the SMAD4/DPC4 Gene in Juvenile Polyposis. *Science*; 280(5366): pp 1086-1088.

Howe JR *et al.* (2001). Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis. *Nat Genet*; 28(2): pp 184-187.

Howe JR *et al.* (2002). Common deletion of SMAD4 in juvenile polyposis is a mutational hotspot. *Am J Hum Genet*; 70(5): pp 1357-1362.

Howe JR *et al.* (2004). The prevalence of MADH4 and BMPR1A mutations in juvenille polyposis and absence of BMPR2, BMPR1B and ACVR1 mutations. *J Med Genet*; 41(7): pp 484-491.

Howe JR *et al.* (2007). ENG mutations in MADH4/BMPR1A mutation negative patients with juvenile polyposis. *Clin Genet*; 71(1): pp 91-92.

Hsaing YH *et al.* (1985). Camptothecin induces protein-linked DNA breaks via mammalian DNA topisomerase I. *J Biol Chem*; 260(27): pp 14873-14878.

Huang JC *et al*. (1994). HMG-domain proteins specifically inhibit the repair of the major DNA adduct of the anticancer drug cisplatin by human excision nuclease. *Proc Natl Acad Sci USA*; 91(22): pp 10394-10398.

Huang S *et al.* (1998). The premature ageing syndrome protein, WRN, is a 3'-->5' exonuclease. *Nat Genet*; 20(2): pp 114-116.

Huang SM *et al.* (1999). Epidermal growth factor receptor blockade with C225 modulates proliferation, apoptosis, and radiosensitivity in squamous cell carcinomas of the head and neck. *Cancer Res*; 59(8): pp 1935-1940.

Huelsken J and Behrens J. (2002). The WNT signalling pathway. *J Cell Sci*; 115(Pt 21): pp 3977-3978.

Hurwitz HI *et al.* (2004). Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N Engl J Med*; 350(23): pp 2335-2342.

Hurwitz HI *et al.* (2005). Bevacizumab in combination with fluorouracil and leucovorin: an active regimen for first-line metastatic colorectal cancer. *J Clin Oncol*; 23(15): pp 3502-3508.

Iannello A and Ahmad A. (2005). Role of antibody-dependent cell-mediated cytotoxicity in the efficacy of therapeutic anti-cancer monoclonal antibodies. *Cancer Metastasis Rev*; 24(4): pp 487-499.

Inada M *et al*. (2010). Associations between oxaliplatin-induced peripheral neuropathy and polymorphisms of the ERCC1 and GSTP1 genes. *Int J Clin Pharmacol Ther*; 48(11): pp 729-734.

Ioannidis JP *et al*. (2001). Replication validity of genetic association studies. *Nat Genet*; 29(3): pp 306-309.

Ionov Y *et al.* (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colorectal carcinogenesis. *Nature*; 363(6429): pp 558-561.

International HapMap Consortium. (2003). The International HapMap Project. *Nature*; 426(6968): pp 789-796.

Ishikawa T *et al*. (1998). Positive correlation between the efficacy of capecitabine and doxifluridine and the ratio of thymidine phosphorylase to dihydropyrimidine dehydrogenase activities in tumors in human cancer xenografts. *Cancer Res*; 58(4): pp 685-690.

Ito H *et al*. (1983). Transformation of intact yeast cells treated with alkali cations. *J Bacteriol*; 153(1): pp 163-168.

Itzkowitz SH and Hapraz N. (2004). Diagnosis and management of dysplasia in patients with inflammatory bowel disease. *Gastroenterology*; 126(6): pp 1634-1648.

Jaarsma D *et al*. (2011). Age-related neuronal degeneration: complementary roles of nucleotide excision repair and transcription coupled repair in preventing neuropathology. *PLoS genetics*.

Jaeger E *et al.* (2003). An ancestral Ashkenazi haplotype at the HMPS/CRAC1 locus on 15q13-q14 is associated with hereditary mixed polyposis syndrome. *Am J Hum Genet*; 72(5): pp 1261-1267.

Jaeger E *et al.* (2008). Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat Genet*; 40(1): pp 26-28.

Jaeger E *et al.* (2012). Hereditary mixed polyposis syndrome is caused by a 40kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1. *Nat Genet*; 44(6): pp 699-703.

Jang E and Chung DC. (2010). Hereditary colon cancer: Lynch Syndrome. *Gut Liver*; 4(2): pp 151–160.

Jascur T and Boland CR. (2006). Structure and function of the components of the human DNA mismatch repair system. *Int. J. Cancer*; 119: pp 2030–2035.

Jasin M. (2002). Homologous repair of DNA damage and tumourigenesis: the BRCA connection. *Oncogene*; 21(58): pp 8981-8993.

Jaspers NG *et al*. (2007). First reported patient with human ERCC1 deficiency has cerebro-oculo-facio-skeletal syndrome with a mild defect in nucleotide excision repair and severe developmental failure. *Am J Hum Genet*; 80(3): pp 457-466.

Jasperson KW *et al.* (2010). Hereditary and familial colon cancer. *Gastroenterology*; 138(6): pp 2044-2058.

Jayaraman L *et al*. (1998). High mobility group protein-1 (HMG-1) is a unique activator of p53. *Genes Dev*; 12(4): pp 462-472.

Johns LE and Houlston RS. (2001). A systematic review and meta-analysis of familial colorectal cancer risk. *Am J Gastroenterol*; 96(10): pp 2992-3003.

Jones S *et al.* (2002). Biallelic germline mutations in MYH predispose to multiple colorectal adenoma and somatic G:C → T:A mutations. *Hum Mol Genet*; 11(23): pp 2961-2967.

Jonker DJ *et al.* (2008). Cetuximab for the treatment of colorectal cancer. *N Engl J Med*; 357(20): pp 2040-2048.

Kabbinavar FF *et al.* (2005). Combined analysis of efficacy: the addition of bevacizumab to fluorouracil/leucovorin improves survival for patients with metastatic colorectal cancer. *J Clin Oncol*; 23(16): pp 3706-3712.

Kanamitsu K and Ikeda S. (2011). Fission yeast homologs of human XPC and CSB, rhp41 and rhp26, are involved in transcription-coupled repair of methyl methanesulfonate-induced DNA damage. *Genes Genet Syst*; 86(2): pp 83-91.

Kanda T *et al.* (1990). Peripheral neuropathy in xeroderma pigmentosum. *Brain*; 113 (Pt 4): pp 1025-1044.

Kane MF *et al.* (1997). Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumours and mismatch repair defective cell lines. *Cancer Res*; 57(5): pp 808-811.

Karapetis CS *et al.* (2008). Kras mutations and benefit from cetuximab in advanced colorectal cancer patients. *N Engl J Med*; 359(17): pp 1757-1765.

Kashiyama K *et al.* (2013). Malfunction of nuclease ERCC1-XPF results in diverse clinical manifestations and causes Cockayne syndrome, xeroderma pigmentosum, and Fanconi anemia. *Am J Hum Genet*; 92(5): pp 807-819.

Katano K *et al*. (2002). Acquisition of resistance to cisplatin is accompanied by changes in the cellular pharmacology of copper. *Cancer Res*; 62(22): pp 6559-6565.

Kawaguchi Y *et al*. (2007). Cetuximab induce antibody-dependent cellular cytotoxicity against EGFR-expressing oesophageal squamous cell carcinoma. *Int J Cancer*; 120(4): pp 781-787.

Kawato Y *et al.* (1991). Intracellular roles of SN-38, a metabolite of the camptothecin derivative CPT-11, in the antitumor effect of CPT-11. *Cancer Res*; 51(16): pp 4187-4191.

Kawato Y *et al.* (1993). Inhibitory activity of camptothecin derivatives against acetylcholinesterase in dogs and their binding activity to acetylcholine receptors in rats. *J Pharm Pharmacol*; 45(5): pp 444-448.

Keightley PD *et al.* (2009). Analysis of the genome sequences of three Drosophila melanogaster spontaneous mutation accumulation lines. *Genome Res*; 19(7): pp 1195-1201.

Kemeny N *et al*. (1990). Randomized study of continuous infusion fluorouracil versus fluorouracil plus cisplatin in patients with metastatic colorectal cancer. *J Clin Oncol*; 8(2): pp 313-318.

Kern W *et al.* (1999). Oxaliplatin pharmacokinetics during a four hour infusion. *Clin Cancer Res*; 5(4): pp 761-765.

Kidani Y *et al*. (1978). Antitumor activity of 1,2-diaminocyclohexane-platinum complexes against sarcoma-180 ascites form. *J. Med. Chem;* 21(12): pp 1315-1318.

Kiezun A *et al.* (2012). Exome sequencing and the genetic basis of complex traits. *Nat Genet*; 44(6): pp 623-630.

Kikuchi K *et al*. (2013). Structure-specific endonucleases Xpf and Mus81 play overlapping but essential roles in DNA repair by homologous recombination. *Cancer Res*; 73(14): pp 4362-4371.

Kilpivaara O and Aaltonen LA. (2013). Diagnostic cancer genome sequencing and the contribution of germline variants. *Science*; 339(6127): pp 1559-1562.

Kim KJ *et al*. (1993). Inhibition of vascular endothelial growth factor-induced angiogenesis suppresses tumour growth in vivo. *Nature*; 362(6423): pp 841-844.

Kim YS *et al*. (2007). Dietary modulation of colon cancer risk. *J Nutr*; 137(11 Suppl): pp 2576-2579.

Kitao S *et al.* (1999). Mutations in *RECQL4* cause a subset of cases of Rothmund-Thomson syndrome. *Nat Genet*; 22(1): pp 82-84.

Klaus A *et al*. (2008). Wnt signalling and its impact on development and cancer. *Nat Rev Cancer*; 8(5): pp 387-398.

Knudsen AL *et al*. (2003). Attenuated familial adenomatous polyposis (AFAP). A review of the literature. *Fam Cancer*; 2(1): pp 43-55.

Knudson AG. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA*; 68(4): pp 820-823.

Knudson AG. (1985). Hereditary cancer, oncogenes, and anti-oncogenes. *Cancer Res*; 45(4): pp 1437-1443.

Kobayashi J *et al.* (2004). NBS1 and its functional role in the DNA damage response. *DNA Repair (Amst);* 3(8-9): pp 855-861.

Koç A *et al*. (2004). Hydroxyurea arrests DNA replication by a mechanism that preserves basal dNTP pools. *J Biol Chem*; 279(1): pp 223-230.

Kolodner R. (1996). Biochemistry and genetics of eukaryotic mismatch repair. *Genes Dev*; 10(12): pp 1433-1442.

Korinek V *et al*. (1997). Constitutive transcriptional activation by a β-catenin-Tcf complex in APC-/- colon carcinoma. *Science*; 275(5307): pp 1784-1787.

Kovacs ME *et al*. (2009). Deletions removing the last exon of *TACSTD1* constitute a distinct class of mutations predisposing to Lynch syndrome. *Hum Mutat*; 30(2): pp 197-203.

Kraemer KH *et al.* (1984). DNA repair protects against cutaneous and internal neoplasia: evidence from xeroderma pigmentosum. *Carcinogenesis*; 5(4): pp 511-514.

Kraemer KH *et al.* (1987). Xeroderma pigmentosum. Cutaneous, ocular, and neurologic abnormalities in 830 published cases*. Arch Dermatol*; (2): pp 241-250.

Kraemer KH. (1997). Sunlight and skin cancer: another link revealed. *Proc Natl Acad Sci USA*; 94(1): pp 11-14.

Krasinskas AM. (2011). EGFR signalling in colorectal cancer. *Pathology Research International*. Article ID 932932.

Krawchuk MD and Wahls WP. (1999). High-efficiency gene targeting in Schizosaccharomyces pombe using a modular, PCR-based approach with long tracts of flanking homology. *Yeast*; 15(13): pp 1419-1427.

Krishnan AV *et al*. (2006). Oxaliplatin-induced neurotoxicity and the development of neuropathy. *Muscle Nerve*; 32(1): pp 51-60.

Krogh BO and Symington LS. (2004). Recombination in yeast. *Annu Rev Genet*; 38: pp 233-271.

Kunkel TA. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc Natl Acad Sci USA*; 82(2): pp 488-492.

Kunkel TA. (2004). DNA replication fidelity. *J Biol Chem*; 279(17): pp 16895-16898.

Kunkel TA *et al*. (2005). DNA mismatch repair. *Annu. Rev. Biochem*; 74: pp 681-710.

Kuraoka I *et al*. (2000). Repair of an interstrand DNA cross-link initiated by ERCC1-XPF repair/recombination nuclease. *J Biol Chem*; 275(34): pp 26632-26636.

Kweekel DM *et al*. (2005). Pharmacology of oxaliplatin and the use of pharmacogenomics to individualize therapy. *Cancer Treat Rev*; 31(2): pp 90-105.

Lahue RS *et al.* (1989). DNA mismatch correction in a defined system. *Science*; 245(4914): pp 160-164.

Laken SJ *et al.* (1997). Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. *Nat Genet*; 17(1): pp 79-83.

Lamberti C *et al.* (1999). Microsatellite instability-a useful diagnostic tool to select patients at high risk for hereditary non-polyposis colorectal cancer: a study in different groups of patients with colorectal cancer. *Gut*; 44(6): pp 839-843.

Lamlum H *et al*. (1999). The type of somatic mutation at *APC* in familial adenomatous polyposis is determined by the site of the germline mutation: a new facet to Knudson's 'two hit' hypothesis. *Nat Med*; 5(9): pp 1071-1075.

Lamlum H *et al*. (2000). Germline APC variants in patients with multiple colorectal adenomas, with evidence for the particular importance of E1317Q. *Hum Mol Genet*; 9(15): pp 2215-2221.

Lander ES and Schork NJ. (1994). Genetic dissection of complex traits. *Science*; 265(5181): pp 2037-2048.

Langer SJ *et al*. (2002). A genetic screen identifies novel non-compatible loxP sites. *Nucleic Acids Res*; 30(14): pp 3067-3077.

Lankisch TO *et al.* (2008). Gilbert's Syndrome and irinotecan toxicity: combination with UDP-glucuronosyltransferase 1A7 variants increases risk. *Cancer Epidemiol Biomarkers Prev*; 17(3): pp 695-701.

Lans H and Vermeulen W. (2011). Nucleotide excision repair in *Caenorhabditis elegans*. *Molecular biology international.*

Larsson SC *et al*. (2006). Meat consumption and risk of colorectal cancer: A meta-analysis of prospective studies. *Int J Cancer*; 119(11): pp 2657-2664.

Laurent-Puig P *et al*. (2009). Analysis of PTEN, BRAF, and EGFR status in determining benefit from cetuximab therapy in wild-type KRAS metastatic colon cancer. *J Clin Oncol*; 27(35): pp 5924-5930.

Lecomte T *et al*. (2004). Thymidylate synthase gene polymorphism predicts toxicity in colorectal cancer patients receiving 5-fluorouracil-based chemotherapy. *Clin Cancer Res*; 10(17): pp 5880-5888.

Lecomte T *et al*. (2006). Glutathione S-transferase P1 polymorphism (Ile105Val) predicts cumulative neuropathy in patients receiving oxaliplatin-based chemotherapy. *Clin Cancer Res*; 12(10): pp 3050-3056.

Lee CG *et al*. (2000). Anti-vascular endothelial growth factor treatment augments tumor radiation response under normoxic or hypoxic conditions. *Cancer Res*; 60(19): pp 5565-5570.

Lee S *et al*. (2010). Genetic polymorphism associated with chronic neurotoxicity and recurrence in curatively-resected colon cancer patients receiving oxaliplatin-based adjuvant chemotherapy. *J Clin Oncol*; 28(15): suppl 358.

Lehmann AR. (1996). Molecular biology of DNA repair in the fission yeast *Schizosaccharomyces pombe*. *Mutat Res*; 363(3): pp 147-161.

Lehmann AR *et al*. (2011). Xeroderma Pigmentosum. *Orphanet Journal of Rare Diseases*; 6(70).

Leibeling D *et al.* (2006). Nucleotide excision repair and cancer. *J Mol Histol*; 37(5-7): pp 225-238.

Lemmens BB and Tijsterman M, (2011). DNA double strand break in Caenorhabditis elegans. *Chromosoma*; 120(1): pp 1-21

Leung MC *et al*. (2008). Caenorhabditis elegans: an emerging model in biomedical and environmental toxicology. *Toxicol Sci*; 106(1): pp 5-28.

Levi F *et al*. (2000). Oxaliplatin: Pharmacokinetics and chronopharmacological aspects. *Clin Pharmacokinet*; 38(1): pp 1-21.

Levy S *et al.* (2007). The diploid genome sequence of an individual human. *PLoS Biol*; 5(10): e254.

Ley TJ *et al.* (2008). DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature; 456(7218): pp 66-72.

Li D *et al.* (2011). Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. *Genet Epidemiol*; 35(8): pp 790-799

Li GM. (2008). Mechanisms and functions of DNA mismatch repair. *Cell Res*; 18(1): pp 85-98.

Li H and Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*; 25(14): pp 1754-1760.

Li X and Heyer WD. (2008). Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res*; 18(1): pp 99-113.

Liang PS *et al*. (2009). Cigarette smoking and colorectal cancer incidence and mortality: systematic review and meta-analysis. *Int J Cancer*; 124(10): pp 2406-2415.

Lichtenstein P *et al*. (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(2): pp 78-85.

Lieber MR *et al.* (2003). Mechanism and regulation of human non-homologous DNA end-joining. *Nat Rev Mol Cell Biol*; 4(9): pp 712-720.

Liedart B *et al*. (2003). Overexpression of cMOAT (MRP2/ABCC2) is associated with decreased formation of platinum-DNA adducts and decreased G"-arrest in melanoma cells resistant to cisplatin. *J Invest Dermatol*; 121(1): pp 172-176.

Lièvre A *et al*. (2008). KRAS mutations as an independent prognostic factor in patients with advanced colorectal cancer treated with cetuximab. *J Clin Oncol*; 26(3): pp 374-379.

Ligtenberg MJ *et al.* (2008). Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. *Nat Genet*; 41(1): pp 112-117.

Linardou H *et al*. (2008). Assessment of somatic k-RAS mutations as a mechanism associated with resistance to EGFR-targeted agents: a systematic review and meta-analysis of studies in advanced non-small-cell lung cancer and metastatic colorectal cancer. *Lancet Oncol*; 9(10): pp 962-972.

Lindahl T. (1993). Instability and decay of the primary structure of DNA. *Nature*; 362(6422): pp 709-715.

Lindahl T and Wood RD. (1999). Quality control by DNA repair. *Science*; 286(5446): pp 1897-1905.

Ling H *et al*. (2013). CCAT2, a novel non-coding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. *Genome Res.*

Lisby M *et al.* (2004). Choreography of the DNA damage response: spatiotemporal relationships among checkpoint and repair proteins. *Cell*; 118(6): pp 699-713.

Little J *et al*. (2003). Colon cancer and genetic variation in folate metabolism: The clinical bottom line. *J Nutr*; 133(11 Suppl 1): pp 3758-3766.

Liu HX *et al.* (2003). The role of *hMLH3* in familial colorectal cancer. *Cancer Res*; 63(8): pp 1894-1899.

Liu L *et al.* (2012). Comparison of next-generation sequencing systems. *J Biomed Biotechnol.*

Liu W *et al*. (2000). Mutations in AXIN2 cause colorectal cancer with defective mismatch repair by activating β-catenin/TCF signalling. *Nat Genet*; 26(2): pp 146-147.

Liu Y *et al.* (2005). DNA polymerase beta and flap endonuclease 1 enzymatic specificities sustain DNA synthesis for long patch base excision repair. *J Biol Chem*; 280(5): pp 3665-3674.

Lodish H *et al.* (2000). Molecular Cell Biology, 4th edition. New York: *W.H. Freeman.*

Loeb LA *et al*. (2003). Multiple mutations and cancer. *PNAS*; 100(3): pp 776-781.

Loehrer PJ Sr *et al*. (1988). A prospective randomized trial of fluorouracil versus fluorouracil plus cisplatin in the treatment of metastatic colorectal cancer: a Hoosier Oncology Group trial. *J Clin Oncol*; 6(4): pp 642-648.

Lokich JJ *et al*. (1989). A prospective randomized comparison of continuous infusion fluorouracil with a conventional bolus schedule in metastatic colorectal carcinoma: a Mid-Atlantic Oncology Program Study. *J Clin Oncol*; 7(4): pp 425-432.

Lou FR *et al*. (1999). High performance liquid chromatographic separation of the biotransformation products of oxaliplatin. *J Chromatogr B Biomed Sci Appl*; 724(2): pp 345-356.

Loupakis F *et al*. (2009a). KRAS codon 61, 146 and BRAF mutations predict resistance to cetuximab plus irinotecan in KRAS codon 12 and 13 wild-type metastatic colorectal cancer. *Br J Cancer*; 101(4): pp 715-721.

Loupakis F *et al*. (2009b). PTEN expression and KRAS mutations on primary tumors and metastases in the prediction of benefit from cetuximab plus irinotecan for patients with metastatic colorectal cancer. *J Clin Oncol*; 27(16): pp 2622-2629.

Lu SL *et al*. (1995). Mutations of the transforming growth factor-β type II receptor gene and genomic instability in hereditary nonpolyposis colorectal cancer. *Biochemical and Biophysical Research Communications*; 216(2): pp 452-457.

Lurje G *et al*. (2009). Thymidylate synthase gene variations: predictive and prognostic markers. *Mol Cancer Ther*; 8(5): pp 1000-1007.

Lv H *et al.* (2013). Genetic polymorphism of *XRCC1* correlated with response to oxaliplatin-based chemotherapy in advanced colorectal cancer. *Cancer Invest*; 31(1): pp 24-28.

Lynch HT and Lynch PM. (1979). The cancer-family syndrome: a pragmatic basis for syndrome identification. *Dis Colon Rectum*; 22(2): pp 106-110.

Lynch HT and de la Chapelle A. (1999). Genetic susceptibility to nonpolyposis colorectal cancer. *J Med Genet*; 36(11): pp 801-818.

Lynch HT and de la Chapelle A. (2003). Hereditary colorectal cancer. *N Engl J Med*; 348(10): pp 919-932.

Lynch HT *et al*. (2006). Phenotypic and genotypic heterogeneity in the Lynch syndrome: diagnostic, surveillance and management implications. *Eur J Hum Genet*; 14(4): pp 390-402.

Lynch HT *et al*. (2007). Who Should Be Sent For Genetic Testing in Hereditary Colorectal Cancer Syndromes? *J Clin Oncol*; 25(23): pp 3534-3542.

Ma Y *et al.* (2002). Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in non-homologous end joining and V(D)J recombination. *Cell*; 108(6): pp 781-794.

Mabuchi S *et al*. (2004). Inhibition of NFκB increases the efficacy of cisplatin in in vitro and in vivo ovarian cancer models. *J Biol Chem*; 279(22): pp 23477-23485.

Majewski J *et al.* (2011). What can exome sequencing do for you? *J Med Genet*; 48(9): pp 580-589.

Malkhoysen S *et al*. (1996). Frameshift mutator mutations. *Nature*; 382: pp 499-500.

Mallery DL *et al*. (1998). Molecular analysis of mutations in the CSB (ERCC6) gene in patients with Cockayne syndrome. *Am J Hum Genet*; 62(1): pp 77-85.

Mandola MV *et al*. (2003). A novel single nucleotide polymorphism within the 5' tandem repeat polymorphism of the thymidylate synthase gene abolishes USF-1 binding and alters transcriptional activity. *Cancer Res*; 63(11): pp 2898-2904.

Mandola MV *et al*. (2004). A 6 bp polymorphism in the thymidylate synthase gene causes message instability and is associated with decreased intratumoral TS mRNA levels. *Pharmacogenetics*; 14(5): pp 319-327.

Mann B *et al*. (1999). Target genes of β-catenin-T cell-factor/lymphoid-enhancer-factor signalling in human colorectal carcinomas. *Proc Natl Acad Sci USA*; 96(4): pp 1603-1608.

Mannsfeldt AG *et al*. (1999). Stomatin, a MEC-2 like protein, is expressed by mammalian sensory neurons. *Mol Cell Neurosci*; 13(6): pp 391-404.

Manolio TA *et al.* (2009). Finding the missing heritability of complex diseases. *Nature*; 461(7265): pp 747-753.

Mao Z *et al.* (2008). DNA repair by non-homologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle*; 7(18): pp 2902-2906.

Mareike Voigt *et al*. (2012). Functional Dissection of the Epidermal Growth Factor Receptor Epitopes Targeted by Panitumumab and Cetuximab. *Neoplasia*; 14(11): pp 1023-1031.

Martin LP *et al*. (2008). Platinum resistance: the role of DNA repair pathways. *Clin Cancer Res*; 14(5): pp 1291-1295.

Martinez ME *et al*. (2007). Meat intake, preparation methods, mutagens and colorectal adenoma recurrence. *Carcinogenesis*; 28(9): pp 2019-2027.

Masutani C *et al.* (1999). The *XPV* (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta. *Nature*; 399(6737): pp 700-704.

Matsumoto Y and Kim K. (1995). Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair. *Science*; 269(5224): pp 699-702.

Matsumura Y *et al.* (1998). Characterization of molecular defects in xeroderma pigmentosum group F in relation to its clinically mild symptoms. *Hum Mol Genet*; 7(6): pp 969-974.

Matsuoka S *et al.* (1998). Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. *Science*; 282(5395): pp 1893-1897.

Maughn TS *et al*. (2011). Addition of cetuximab to oxaliplatin-based first line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. *Lancet*; 377(9783): pp 2103-2114.

McCarthy MI and Hirschhorn JN. (2008). Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet*; 17(R2): pp R156-65.

McCready S *et al*. (1993). Repair of cyclobutane pyrimidine dimers and 6-4 photoproducts in the fission yeast Schizosaccharomyces pombe. *Mol Microbiol*; 10(4): pp 885-890.

McDonell LM *et al.* (2013). Mutations in STAMBP, encoding a deubiquitinating enzyme, cause microcephaly-capillary malformation syndrome. *Nat Genet*; 45(5): pp 556-562.

McKay BC *et al.* (1999). Potential roles for p53 in nucleotide excision repair. *Carcinogenesis*; 20(8): pp 1389-1396.

McKenna A *et al*. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*; 20(9): pp 1297-1303.

McMurray M *et al*. (2005). To die or not to die; DNA repair in neurons. *Mutat Res*; 577(1-2): pp 260-274.

McNeil EW and Melton DW. (2012). DNA repair endonuclease ERCC1-XPF as a novel therapeutic target to overcome chemoresistance in cancer therapy. *Nucleic Acids Res*; 40(20): pp 9990-10004.

McVey M. (2010). Strategies for DNA interstrand cross link repair: Insights from worms, flies, frogs, and slime mould. *Environ Mol Mutagen*; 51(6): pp 646-658.

Medeiros R *et al*. (2003). Platinum/paclitaxel-based chemotherapy in advanced ovarian carcinoma: glutathione S-transferase genetic polymorphisms as predictive biomarkers of disease outcome. *Int J Clin Oncol*; 8(3): pp 156-161.

Mehlen P and Fearon ER. (2004). Role of the dependence receptor DCC in colorectal cancer pathogenesis. *J Clin Oncol*; 22(16): pp 3420-3428.

Mendelsohn J and Baselga J. (2003). Status of the epidermal growth factor receptor antagonists in the biology and the treatment of cancer. *J Clin Oncol*; 21(14): pp 2787-2799.

Metzker ML. (2010). Sequencing technologies - the next generation. *Nat Rev Genet*; 11(1): pp 31-46.

Meyn MS. (1995). Ataxia-telangiectasia and cellular responses to DNA damage. *Cancer Res*; 55(24): pp 5991-6001.

Milanowska K *et al.* (2011). REPAIRtoire - a database of DNA repair pathways. *Nucleic Acids Res*; (Database issue): pp D788-792.

Miki Y *et al.* (1994). A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1. Science*; 266(5182): pp 66-71.

Mimitou EP and Symington LS. (2009). Nucleases and helicases take centre stage in homologous recombination. *Trends Biochem Sci*; 34(5): pp 264-272.

Miwa M *et al*. (1998). Design of a novel oral fluoropyrimidine carbamate, capecitabine, which generates 5-fluorouracil selectively in tumours by enzymes concentrated in human liver and cancer tissue. *Eur J Cancer*; 34(8): pp 1274-1281.

Miyoshi Y *et al*. (1992a). Germ-line mutations of the *APC* gene in 53 familial adenomatous polyposis patients. *Proc Natl Acad Sci USA*; 89(10): pp 4452-4456.

Miyoshi Y *et al*. (1992b). Somatic mutations of the *APC* gene in colorectal tumors: mutation cluster region of the *APC* genes. *Hum Mol Genet*; 1(4): pp 229-233.

Modesti M and Kanaar R. (2001). Homologous recombination: from model organisms to human disease. *Genome Biol;* 2(5).

Montagut C *et al*. (2012). Identification of a mutation in the extracellular domain of the Epidermal Growth Factor Receptor conferring cetuximab resistance in colorectal cancer. *Nat Med*; 18(2): pp 221-223.

Morel A *et al*. (2006). Clinical relevance of different dihydropyrimidine dehydrogenase gene single nucleotide polymorphisms on 5-fluorouracil tolerance. *Mol Cancer Ther;* 5(11): pp 2895-2904.

Morin PJ *et al*. (1997). Activation of β-catenin-Tcf signaling in colon cancer by mutations in *β-catenin* or *APC*. *Science*; 275(5307): pp 1787-1790.

Moroni M *et al*. (2005). Gene copy number for epidermal growth factor receptor (EGFR) and clinical response to antiEGFR treatment in colorectal cancer: a cohort study. *Lancet Oncol*; 6(5): pp 279-286.

Morrison C *et al*. (2000). The controlling role of ATM in homologous recombinational repair of DNA damage. *EMBO J*; 19(3): pp 463-471.

Moser J *et al.* (2007). Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III alpha in a cell-cycle-specific manner. *Mol Cell*; 27(2): pp 311-323.

Motl S. (2005). Bevacizumab in combination chemotherapy for colorectal and other cancers. *Am J Health Syst Pharm*; 62(10): pp 1021-1032.

Mourad JJ *et al.*(2008). Blood pressure rise following angiogenesis inhibition by bevacizumab. A crucial role for microcirculation. *Ann Oncol*; 19(5): pp 927-934.

Moynahan ME *et al.* (2001). Homology directed DNA repair, mitomycin-C resistance, and chromosome stability is restored with correction of a *Brca1* mutation. *Cancer Res*; 61(12): pp 4842-4850.

Mullis K *et al*. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol*; 51(Pt 1): pp 263-273.

Nagengast FM *et al.* (2002). Surveillance for hereditary nonpolyposis colorectal cancer: a long-term study on 114 families. *Dis Colon Rectum*; 45(12): pp 1588-1594.

Nagy R *et al*. (2004). Highly penetrant hereditary cancer syndromes. *Oncogene*; 23(38): pp 6445-6470.

Nance MA and Berry SA. (1992). Cockayne syndrome: review of 140 cases. *Am J Med Genet*; 42(1): pp 68-84.

NCBI Resource Coordinators. (2013). Database resources of the National Center for Biotechnology Information. *Nucl. Acids Res*; 41 (D1): D8-D20.

Neeley WL and Essigmann JM. (2006). Mechanisms of formation, genotoxicity and mutation of guanine oxidation products. *Chem Res Toxicol*; 19(4): pp 491-505.

Negrini S *et al*. (2010). Genomic instability - an evolving hallmark of cancer. *Nat Rev Mol Cell Biol*; 11(3): pp 220-228.

Nehmé A *et al*. (1999). Induction of JNK and c-Abl signalling by cisplatin and oxaliplatin in mismatch repair-proficient and -deficient cells. *Br J Cancer*; 79(7-8): pp 1104-1110.

Nejentsev S *et al.* (2009). Rare variants of *IFIH1*, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science*; 324(5925): pp 387-389.

Nelson DL and Warren ST. (1993). Trinucleotide repeat instability: when and where? *Nat Genet*; 4(2): pp 107-108.

Neumann E *et al*. (1982). Gene transfer into mouse lyoma cells by electroporation in high electric fields. *EMBO J*; 1(7): pp 841-845.

Newton KF *et al*. (2012). Review of biomarkers in colorectal cancer. *Colorectal Dis*; 14(1): pp 3-17.

Ng PC and Henikoff S. (2001). Predicting deleterious amino acid substitutions. *Genome Res*; 11(5): pp 863-874.

Ng SB *et al.* (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*; 461(7261): pp 272-276.

Ng SB *et al.* (2010). Exome sequencing identifies the cause of a Mendelian disorder. *Nature*; 42(1): pp 30-35.

Nicolaides NC *et al.* (1994). Mutations of two *PMS* homologues in hereditary nonpolyposis colon cancer. *Nature*; 371(6492): pp 75-80.

Nicum S *et al*. (2000). Chemotherapy for colorectal cancer. *J R Soc Med*; 93(8): pp 416–419.

Niedernhofer LJ *et al.* (2003). Malondialdehyde, a product of lipid peroxidation, is mutagenic in human cells. *J Biol Chem*; 278(33): pp 31426-31433.

Niedernhofer LJ *et al*. (2004). The structure-specific endonuclease Ercc1-Xpf is required to resolve DNA interstrand cross-link-induced double-strand breaks. *Mol Cell Biol*; 24(13): pp 5776-5787.

Niedernhofer LJ *et al*. (2006). A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. *Nature*; 444(7122): pp 1038-1043.

Ning Y *et al*. (2010). A quantitative analysis of body mass index and colorectal cancer: findings from 56 observational studies. *Obes Rev*; 11(1): pp 19-30.

Noda E *et al*. (2012). Predictive value of expression of ERCC1 and GST-p for 5-fluorouracil/oxaliplatin chemotherapy in advanced colorectal cancer. *Hepatogastroenterology*; 59(113): pp 130-133.

O'Dwyer *et al.* (2000). Clinical pharmacokinetics and administration of established platinum drugs. *Drugs.* 59 (4); pp. 19-27.

Ogi T and Lehmann AR. (2006). The Y-family DNA polymerase kappa (POLκ) functions in mammalian nucleotide-excision repair. *Nat Cell Biol*; 8(6): pp 640-642.

Oguri T *et al*. (2000). Increased expression of the MRP5 gene is associated with exposure to platinum drugs in lung cancer. *Int J Cancer*; 86(1): pp 95-100.

Oguri T *et al*. (2013). Genetic polymorphisms associated with oxaliplatin-induced peripheral neurotoxicity in Japanese patients with colorectal cancer. *Int J Clin Pharmacol Ther*; 51(6): pp 475-481.

Okou DT *et al.* (2007). Microarray-based genomic selection for high-throughput resequencing. *Nat Methods*; 4(11): pp 907-909.

Olsson AK *et al*. (2006). VEGF receptor signalling - in control of vascular function. *Nat Rev Mol Cell Biol*; 7(5): pp 359-371.

Ouyang KJ *et al.* (2008). Homologous recombination and maintenance of genome integrity: Cancer and aging through the prism of human RecQ helicases. *Mech Ageing Dev*; 129(7-8): pp 425-440.

Ouyang KJ *et al.* (2009). SUMO modification regulates BLM and RAD51 interaction at damaged replication forks. *PLoS Bio.*

Pabinger S *et al.* (2013). A survey of tools for variant analysis of next-generation genome sequencing data. *Brief Bioinform.*

Palles C *et al*. (2013). Germline mutations affecting the catalytic domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nature Genetics*; 45 (2).

Pande AU *et al.* (2006). Bevacizumab (BV) induced hypertension (HT): A manageable toxicity. *J Clin Oncol*; (Meeting abstract), vol. 24 no. 18_suppl 13539.

Pandey P *et al*. (1996). Activation of p38 mitogen-activated protein kinase by c-Abl-dependent and -independent mechanisms. *J Biol Chem*; 271(39): pp 23775-23779.

Park CH *et al.* (1995). Purification and characterization of the XPF-ERCC1 complex of human DNA repair excision nuclease. *J Biol Chem*; 270(39): pp 22657-22660.

Park DJ *et al*. (2003). ERCC1 gene polymorphism as a predictor for clinical outcome in advanced colorectal cancer patients treated with platinum-based chemotherapy. *Clin Adv Hematol Oncol*; 1(3): pp 162-166.

Parla JS *et al*. (2011). A comparative analysis of exome capture. *Genome Biol*; 12(9): pp R97.

Parrilla-Castellar ER *et al*. (2004). Dial 9-1-1 for DNA damage: the Rad9-Hus1-Rad1 (9-1-1) clamp complex. *DNA Repair (Amst)*; 3(8-9): pp 1009-1014.

Parsons R *et al*. (1993). Hyper mutability and mismatch repair deficiency in RER+ tumor cells. *Cell*; 75(6): pp 1227-1236.

Patel SG and Ahnen DJ. (2012). Familial colon cancer syndromes: an update of a rapidly evolving field. *Curr Gastroenterol Rep*; 14(5): pp 428-438.

Pearson SJ *et al*. (2006). A novel DNA damage recognition protein in Schizosaccharomyces pombe. *Nucleic Acids Res*; 34(8): pp 2347-2354.

Peeters M *et al*. (2009). Association of progression-free survival, overall survival, and patient-reported outcomes by skin toxicity and KRAS status in patients receiving panitumumab monotherapy. *Cancer*; 115(7): pp 1544-1554.

Peeters M *et al*. (2010). Randomized phase III study of panitumumab with fluorouracil, leucovorin, and irinotecan (FOLFIRI) compared with FOLFIRI alone as second-line treatment in patients with metastatic colorectal cancer. *J Clin Oncol*; 28(31): pp 4706-4713.

Pegg AE. (2000). Repair of O(6)-alkylguanine by alkyltransferases. *Mutat Res*; 462(2-3): pp 83-100.

Peltomäki P and Vasen HF. (1998). Mutations predisposing to hereditary nonpolyposis colorectal cancer: database and results of a collaborative study. The International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer. Gastroenterology; 113(4): pp 1146-1158.

Peltomäki P. (2001). DNA mismatch repair and cancer. *Mutat Res*; 488(1): pp 77-85.

Peltomäki P. (2005). Lynch syndrome genes. *Fam Cancer*; 4(3): pp 227-232.

Pendyala L *et al*. (1993). In vitro cytotoxicity, protein binding, red blood cell partitioning and biotransformation of oxaliplatin. *Cancer Res*; 53(24): pp 5970-5976.

Peng Z *et al*. (2013). Association between GSTP1 Ile105Val polymorphism and oxaliplatin-induced neuropathy: a systematic review and meta-analysis. *Cancer Chemother Pharmacol*; 72(2): pp 305-314.

Perrone F *et al*. (2009). PI3KCA/PTEN deregulation contributes to impaired responses to cetuximab in metastatic colorectal cancer patients. *Ann Oncol*; 20(1): pp 84-90.

Petermann E *et al*. (2010). Hydroxyurea-stalled replication forks become progressively inactivated and require two different RAD51-mediated pathways for restart and repair. *Mol Cell*; 37(4): pp 492-502.

Petkovic M *et al.* (2005). The human Rothmund-Thomson syndrome gene product, RECQL4, localizes to distinct nuclear foci that coincide with proteins involved in the maintenance of genome stability. *J Cell Sci*; 118(Pt 18): pp 4261-4269.

Pham AN *et al*. (2012). Pharmacogenomic analysis identifies increased efficacy of oxaliplatin in ABCB1 overexpressing tumours. *Pharmacogenomics and Pharmacoproteomics.*

Pichierri P *et al*. (2012). The RAD9-RAD1-HUS1 (9.1.1) complex interacts with WRN and is crucial to regulate its response to replication fork stalling. *Oncogene*; 31(23): pp 2809-2823.

Pitot HC *et al*. (2000). Phase I dose-finding and pharmacokinetic trial of irinotecan hydrochloride (CPT-11) using a once-every-three-week dosing schedule for patients with advanced solid tumor malignancy. *Clin Cancer Res*; 6(6): pp 2236-2244.

Pittman AM *et al*. (2009). The colorectal cancer risk at 18q21 is caused by a novel variant altering SMAD7 expression. *Genome Res*; 19(6): pp 987-993.

Pittman AM *et al*. (2010). Allelic variation at the 8q23.3 colorectal cancer risk locus functions as a cis-acting regulator of EIF3H. *PLoS Genet*; 6(9): e1001126.

Plasencia C *et al*. (2006). Expression analysis of genes involved in oxaliplatin response and development of oxaliplatin-resistant HT29 colon cancer cells. *Int J Oncol*; 29(1): pp 225-235.

Pomerantz M *et al*. (2009). The 8q24 cancer risk variant shows long range interaction with MYC in colorectal cancer. *Nature Genetics;* **41**: pp 882-884.

Pommier Y. (2013). Drugging topoisomerases: lessons and challenges. *ACS Chem Biol*; 8(1): pp 82-95.

Popat S *et al*. (2000). Prevalence of the APC E1317Q variant in colorectal cancer patients. *Cancer Lett*; 149(1-2): pp 203-206.

Porreca GJ *et al.* (2007). Multiplex amplification of large sets of human exons. *Nat Methods*; pp 931-936.

Prakash S *et al.* (1993). DNA repair genes and proteins of *Saccharomyces cerevisiae*. *Annu Rev Genet*; 27: pp 33-70.

Prakash S and Prakash L. (2000). Nucleotide excision repair in yeast. *Mutat Res*; 451(1-2): pp 13-24

Pritchard JK. (2001). Are rare variants responsible for susceptibility to complex disease? *Am J Hum Genet*; 69(1): pp 124-137.

Prudden J *et al*. (2003). Pathway utilization in response to a site-specific DNA double-strand break in fission yeast. *EMBO J*; 22(6): pp 1419-1430.

Pruitt KD *et al*. (2009). The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res*; 19(7): pp 1316-1323.

Pullarkat ST *et al*. (2001). Thymidylate synthase gene polymorphism determines response and toxicity of 5-FU chemotherapy. *Pharmacogenomics J*; 1(1): pp 65-70.

Purcell S *et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*; 81(3): pp 559-575.

Rabik CA and Dolan ME. (2007). Molecular mechanism of resistance and toxicity associated with platinating agents. *Cancer Treat Rev*; 33(1): pp 9-23.

Radicela JP *et al.* (1997). Cloning and characterisation of hOGG1, a human homolog of the OGG1 gene of Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. USA*; (94): pp 8010-8015.

Radparvar S *et al*. (1989). Effect of polyglutamylation of 5,10-methylenetetrahydrofolate on the binding of 5-fluoro-2'-deoxyuridylate to thymidylate synthase purified from a human colon adenocarcinoma xenograft. *Biochem Pharmacol*; 38(2): pp 335-342.

Raida M *et al*. (2001). Prevalence of a common point mutation in the dihydropyrimidine dehydrogenase (DPD) gene within the 5'-splice donor site of intron

14 in patients with severe 5-fluorouracil (5-FU)- related toxicity compared with controls. *Clin Cancer Res*; 7(9): pp 2832-2839.

Rajagopalan H *et al*. (2002). Tumorigenesis: RAF/RAS oncogenes and mismatch-repair status. *Nature*; 418(6901): pp 934.

Raji H and Hartsuiker E. (2006). Double strand break and homologous recombination in Schizosaccharomyces pombe. *Yeast*; 23(13): pp 963-976.

Rampino N *et al*. (1997). Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. *Science*; 275(5302): pp 967-969.

Rang HP *et al*. (2007). Pharmacology. Philadelphia: Churchill Livingstone; *Elsevier.*

Ranpura V *et al.* (2010). Increased risk of high-grade hypertension with bevacizumab in cancer patients: a meta-analysis. *Am J Hypertens*; 23(5): pp 460-468.

Ranson M. (2003). Technology evaluation: ABX-EGF, Abgenix/Amgen. *Curr Opin Mol Ther*; 5(5): pp 541-546.

Räschle M *et al.* (2008). Mechanism of replication-coupled DNA interstrand crosslink repair. *Cell*; 134(6): pp 969-980.

Raymond E *et al*. (1998). Oxaliplatin: mechanism of action and antineoplastic activity. *Semin Oncol*; 25(2 Suppl 5): pp 4-12.

Reardon JT *et al.* (1999). Efficient nucleotide excision repair of cisplatin, oxaliplatin, and Bis-aceto-ammine-dichloro-cyclohexylamine-platinum(IV) (JM216) platinum intrastrand DNA diadducts. *Cancer Res*; 59(16): pp 3968-3971.

Reilly MM *et al*. (2011). Charcot-Marie-Tooth disease. *JPNS*; 16(1): pp 1-14.

Reix P *et al.* (2007). Bronchiectasis in two paediatric patients with Rothmund-Thomson syndrome. *Pediatr Int;* 49(1): pp 118-120.

Rixe O *et al*. (1996). Oxaliplatin, tetraplatin, cisplatin and carboplatin: Spectrum of activity in drug resistant cell lines of the national cancer institute's anticancer drug screen panel. *Biochem Pharmacol*; 52(12): pp 1855-1865.

Robbins JH *et al.* (2002). Adult-onset xeroderma pigmentosum neurological disease--observations in an autopsy case. *Clin Neuropathol*; 21(1): pp 18-23.

Robertson AB *et al.* (2009). Base excision repair: the long and short of it. *Cell. Mol. Life Sci*; (66): pp 981-993.

Robinson PN *et al.* (2011). Strategies for exome and genome sequence data analysis in disease-gene discovery projects. *Clin Genet*; 80(2): pp 127-132.

Roffers-Agarwal J and Gammill LS. (2009). Neuropilin receptors guide distinct phases of sensory and motor neuronal segmentation. *Development*; 136(11): pp 1879-1888.

Rossor AM *et al.* (2013). Clinical implications of genetic advances in Charcot-Marie-Tooth disease. *Nat Rev Neurol*; 9(10): pp 562-571.

Rothenberg ML. (2001). Irinotecan (CPT-11): Recent developments and future directions-Colorectal cancer and beyond. *Oncologist*; 6(1): pp 66-80.

Rothenberg ML *et al*. (2003). Superiority of oxaliplatin and fluorouracil-leucovorin compared with either therapy alone in patients with progressive colorectal cancer after irinotecan and fluorouracil-leucovorin: interim results of a phase III trial. *J Clin Oncol*; 21(11): pp 2059-2069.

Roy R *et al*. (2011). BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat Rev Cancer*; 12(1): pp 68-78.

Rozen S and Skaletsky H. (2000). Primer3 on the WWW for the general users and biologists programmers. *Methods Mol Biol*; 132: pp 365-386.

Ruzzo A *et al*. (2007). Pharmacogenetic profiling in patients with advanced colorectal cancer treated with first-line FOLFOX-4 chemotherapy. *J Clin Oncol*; 25(10): pp 1247-1254.

Safaei R *et al*. (2004). The role of copper transporters in the development of resistance to Pt drugs. *J Inorg Biochem*; 98(10): pp 1607-1613.

Saif MW. (2009). Managing bevacizumab-related toxicities in patients with colorectal cancer. *J Support Oncol*; 7(6): pp 245-251.

Saltz LB *et al*. (2000). Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. Irinotecan Study Group. *N Engl J Med*; 343(13): pp 905-914.

Saltz LB *et al.* (2004). Phase II trial of cetuximab in patients with refractory colorectal cancer that expresses the epidermal growth factor receptor *J Clin Oncol*; 22(7): pp 1201-1208.

Saltz LB *et al*. (2008). Bevacizumab in combination with oxaliplatin-based chemotherapy as first-line therapy in metastatic colorectal cancer: a randomized phase III study. *J Clin Oncol*; 26(12): pp 2013-2019.

Sambrook J *et al.* (1989). Molecular cloning: A laboratory manual. New York; *Cold spring harbour laboratory press*.

Sampson JR *et al*. (2003). Autosomal recessive colorectal adenomatous polyposis due to inherited mutations of MYH. *Lancet*; 362(9377): pp 39-41.

Sanger F *et al.* (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA*; 74(12): pp 5463-5467.

Sargent RG *et al*. (1997). Recombination-dependent deletion formation in mammalian cells deficient in the nucleotide excision repair gene ERCC1. *Proc Natl Acad Sci USA*; 94(24): pp 13122-13127.

Sargent RG *et al*. (2000). Role of the nucleotide excision repair gene ERCC1 in formation of recombination-dependent rearrangements in mammalian cells. *Nucleic Acids Res*; 28(19): pp 3771-3778.

Saris CP *et al*. (1996). In vitro formation of DNA adducts by cisplatin, lobaplatin and oxaliplatin in calf thymus DNA in solution and in cultured human cells. *Carcinogenesis*; 17(12): pp 2763-2769.

Sartore-Bianchi A *et al*. (2007). Epidermal growth factor receptor gene copy number and clinical outcome of metastatic colorectal cancer treated with panitumumab. *J Clin Oncol*; 25(22): pp 3238-3245.

Sartore-Bianchi A *et al*. (2009). PIK3CA mutations in colorectal cancer are associated with clinical resistance to EGFR-targeted monoclonal antibodies. *Cancer Res*; 69(5): pp 1851-1857.

Sartori AA *et al.* (2007). Human CtIP promotes DNA end resection. *Nature*; 450(7169): pp 509-514.

Saunders and Iveson T. (2006). Management of advanced colorectal cancer: state of the art. *Br J Cancer*; 95(2): pp 131-138.

Savinainen KJ *et al*. (2006). Overexpression of EIF3S3 promotes cancer cell growth. *Prostate*; 66(11): pp 1144-1150.

Schlötterer C and Harr H. (2001). Microsatellite instability. *eLS.*

Schmiegel W *et al*. (2013). Capecitabine/irinotecan or capecitabine/oxaliplatin in combination with bevacizumab is effective and safe as first-line therapy for metastatic colorectal cancer: a randomized phase II study of the AIO colorectal study group. *Ann Oncol*; 24(6): pp 1580-1587.

Schmoll HJ *et al*. (2007). Phase III trial of capecitabine plus oxaliplatin as adjuvant therapy for stage III colon cancer: a planned safety analysis in 1,864 patients. *J Clin Oncol*; 25(1): pp 102-109.

Schneikert J *et al.* (2007). The canonical Wnt signalling pathway and its APC partner in colon cancer development. *Gut*; 56(3): pp 417-425.

Schurman SH *et al.* (2009). Direct and indirect roles of RECQL4 in modulating base excision repair capacity. *Hum Mol Genet*; 18(18): pp 3470-3483.

Schüller J *et al.* (2000). Preferential activation of capecitabine in tumor following oral administration to colorectal cancer patients. *Cancer Chemother Pharmacol*; 45(4): pp 291-297.

Schwarz Q *et al*. (2009). Neuropilin-mediated neural crest cell guidance is essential to organise sensory neurons into segmented dorsal root ganglia. *Development*; 136(11): pp 1785-1789.

Scully R *et al.* (1997). Association of BRCA1 with Rad51 in mitotic and meiotic cells. *Cell*; 88(2): pp 265-275.

Seal S *et al.* (2006). Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet*; 38(11): pp 1239-1241.

Seetharam RN *et al*. (2010). Oxaliplatin resistance induced by ERCC1 up-regulation is abrogated by siRNA-mediated gene silencing in human colorectal cancer cells. *Anticancer Res*; 30(7): pp 2531-2538.

Segaert S and Van Cutsem E. (2005). Clinical signs, pathophysiology and management of skin toxicity during therapy with epidermal growth factor receptor inhibitors. *Ann Oncol*; 16(9): pp 1425-1433.

Sekelsky JJ *et al.* (2000). DNA repair in Drosophila: Insights from the Drosophila genome sequence. *J Cell Biol*; 150(2): pp F31-6.

Serra R. (2002). Transforming Growth Factor beta (TGFβ). *Encyclopedia of Life Sciences*, pp 1-7.

Shi G *et al*. (2006). Physical and functional interactions between MutY glycosylase homologue (MYH) and checkpoint proteins Rad9-Rad1-Hus1. *Biochem J*; 400(1): pp 53-62.

Sieber OM *et al.* (2003). Multiple colorectal adenomas, classic adenomatous polyposis and germline mutations in MYH. *N Engl J Med*; 348(9): pp 791-799.

Siegsmund MJ *et al*. (1999). Cisplatin-resistant bladder carcinoma cells: enhanced expression of metallothioneins. *Urol Res*; 27(3): pp 157-163.

Siitonen HA *et al*. (2009). The mutation spectrum in RECQL4 diseases. *Eur J Hum Genet*; (17): pp 151-158.

Sijbers AM *et al*. (1998). Homozygous R788W point mutation in the XPF gene of a patient with xeroderma pigmentosum and late-onset neurologic disease. *J Invest Dermatol*; 110(5): pp 832-836.

Sinha RP and Häder DP. (2002). UV-induced DNA damage and repair: a review. *Photochem Photobiol Sci*; 1(4): pp 225-236.

Shaul Y. (2000). c-Abl: activation and nuclear targets. *Cell Death Differ*; 7(1): pp 10-16.

Shaw RJ. (2006). Glucose metabolism and cancer. *Curr Opin Cell Biol*; 18(6): pp 598-608.

Shendure J and Ji H. (2008). Next-generation DNA sequencing. *Nat Biotechnol*; 26(10): pp 1135-1145.

Shi Y and Massagué J. (2003). Mechanisms of TGF-β signalling from cell membrane to the nucleus. *Cell*; 113(6): pp 685-700.

Shirota Y *et al*. (2001). ERCC1 and thymidylate synthetase mRNA levels predict survival for colorectal cancer patients receiving combination oxaliplatin and fluorouracil chemotherapy. *J Clin Oncol*; 19(23): pp 4298-4304.

Shitoh K *et al*. (2001). Frequent activation of the β-catenin-Tcf signaling pathway in nonfamilial colorectal carcinomas with microsatellite instability. *Genes Chromosomes Cancer*; 30(1): pp 32-37.

Shord SS *et al.* (2009). Understanding and managing the possible adverse effects associated with bevacizumab. *Am J Health Syst Pharm*; 66(11): pp 999-1013.

Singh DK *et al.* (2010). The involvement of human RECQL4 in DNA double-strand break repair. *Aging Cell*; 9(3): pp 358-371.

Slattery ML *et al.* (2010). Increased risk of colon cancer associated with a genetic polymorphism of SMAD7. *Cancer Res*; 70(4): pp 1479-1485.

Smith CG *et al.* (2013). Role of the Oxidative DNA Damage Repair Gene OGG1 in Colorectal Tumorigenesis. *J Natl Cancer Inst*; 105(16): pp 1249-1253.

Sobrero A *et al.* (2009). Phase IV study of bevacizumab in combination with infusional fluorouracil, leucovorin and irinotecan (FOLFIRI) in first-line metastatic colorectal cancer. *Oncology*; 77(2): pp 113-119.

Song IM *et al.* (2004). Role of human copper transporter Ctr1 in the transport of platinum-based antitumor agents in cisplatin-sensitive and cisplatin-resistant cells. *Mol Cancer Ther;* 3(12): pp 1543-1549.

Song W *et al*. (2009). The DNA binding domain of human DNA ligase I interacts with both nicked DNA and the DNA sliding clamps, PCNA and hRad9-hRad1-hHus1. *DNA Repair (Amst)*; 8(8): pp 912-919.

Sood A *et al.* (2012). PTEN gene expression and mutations in the PIK3CA gene as predictors of clinical benefit to anti-epidermal growth factor receptor antibody therapy in patients with KRAS wild-type metastatic colorectal cancer. *Clin Colorectal Cancer*; 11(2): pp 143-150.

Soravia C *et al.* (1998). Genotype-phenotype correlations in attenuated adenomatous polyposis coli. *Am J Hum Genet*; 62(6): pp 1290-1301.

Spencer CC *et al.* (2009). Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet*; 5(5): e1000477.

Spirio L *et al.* (1993). Alleles of the APC gene: an attenuated form of familial polyposis. *Cell*; 75(5): pp 951-957.

Sprowl JA *et al*. (2013). Oxaliplatin-induced neurotoxicity is dependent on the organic cation transporter OCT2. *Proc Natl Acad Sci USA*; 110(27): pp 11199-11204.

Stangl R *et al*. (1994). Factors influencing the natural history of colorectal liver metastases. *Lancet*; 343(8910): pp 1405-1410.

Staresincic L *et al.* (2009). Coordination of dual incision and repair synthesis in human nucleotide excision repair. *EMBO J*; 28(8): pp 1111-1120.

Stoehlmacher J *et al*. (2001). A polymorphism of the XRCC1 gene predicts for response to platinum based treatment in advanced colorectal cancer. *Anticancer Res*; 21(4B): pp 3075-3079.

Stoehlmacher J *et al.* (2002). Association between glutathione-S-transferase P1, T1 and M1 genetic polymorphism and survival of patients with metastatic colorectal cancer. *J Natl Cancer Inst*; 94(12): pp 936-942.

Sung P and Klein H. (2006). Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nat Rev Mol Cell Biol*; (7): pp 739-750.

Sweet K *et al.* (2005). Molecular classification of patients with unexplained hamartomatous and hyperplastic polyposis. *JAMA*; 294(19): pp 2465-2473.

Ta LE *et al.* (2006). Neurotoxicity of oxaliplatin and cisplatin for dorsal root ganglion neurons correlates with platinum-DNA binding. *Neurotoxicology*; 27(6): pp 992-1002.

Takata M *et al.* (1998). Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO J*; 17(18): pp 5497-5508.

Tang JY *et al*. (2000). Xeroderma pigmentosum p48 gene enhances global genomic repair and suppresses UV-induced mutagenesis. *Mol Cell*; 5(4): pp 737-744.

Tavtigian SV *et al.* (2006). Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet;* 43(4): pp 295-305.

Taylor WB. (1957). Rothmund's syndrome-Thomson's syndrome. *JAMA Arch Derm*; 75(2): pp 236-244.

Tenesa A *et al.* (2008). Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet*; 40(5): pp 631-637.

Tenesa A and Dunlop MG. (2009). New insights into the aetiology of colorectal cancer from genome wide association studies. *Nat Rev Genet*; 10: pp 353-358.

Tentori L *et al*. (2013). MSH3 expression does not influence the sensitivity of colon cancer HCT116 cell line to oxaliplatin and poly (ADP-ribose) polymerase (PARP) inhibitor as monotherapy or in combination. *Cancer Chemother Pharmacol*; 72(1): pp 117-125.

Theile D *et al.* (2009). Involvement of drug transporters in the synergistic action of FOLFOX combination chemotherapy. *Biochem Pharmacol*; 78(11): pp 1366-1373.

Thomson JP. (1990). Familial adenomatous polyposis: the large bowel. *Ann R Coll Surg Engl*; 72(3): pp 177–180.

Thrush DC *et al.* (1974). Neurological manifestations of xeroderma pigmentosum in two siblings. *J Neurol Sci*; 22(1): pp 91-104.

Tischkowitz M and Eeles R. (2003). Mutations in *BRCA1* and *BRCA2* and predisposition to prostate cancer. *Lancet*; 362(9377): pp 80

Tobin P *et al.* (2004). Inhibition of acetylcholinesterase in patients receiving irinotecan (camptothecin-11). *Clin Pharmacol Ther*; 76(5): pp 505-506.

Tofthagen C *et al.* (2013). Oxaliplatin-induced peripheral neuropathy's effects on health-related quality of life of colorectal cancer survivors. *Support Care Cancer*.

Tomlinson IP *et al.* (1999). Inherited susceptibility to colorectal adenomas and carcinomas: evidence for a new predisposition gene on 15q14-q22. *Gastroenterology*; 116(4): pp 789-795.

Tomlinson IP *et al.* (2007). A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet*; 39(8): pp 984-988.

Tomlinson IP *et al.* (2008). A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet*; 40(5): pp 623-630.

Tomlinson IP *et al.* (2011). Multiple common susceptibility variants near the BMP pathway loci GREM1, BMP4 and BMP2 explain part of the missing heritability of colorectal cancer. *PLoS Genet*; 7(6): e1002105.

Tournigand C *et al.* (2006). OPTIMOX1: a randomized study of FOLFOX4 or FOLFOX7 with oxaliplatin in a stop-and-go fashion in advanced colorectal cancer - a GERCOR study. *J Clin Oncol*; 24(3): pp 394-400.

Toyoda H *et al.* (2000). HeLa cell transformants overproducing mouse metallothionein show in vivo resistant to cis-platinum in nude mice. *Jpn J Cancer Res*; 91(1): pp 91-98.

Toyota M *et al.* (1999). CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci USA*; 96(15): pp 8681-8686.

Tripsianes K *et al.* (2005). The structure of the human ERCC1/XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair. *Structure*; 13(12): pp 1849-1858.

Troelstra C *et al*. (1992). ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. *Cell*; 71(6): pp 939-953.

Truglio JJ *et al.* (2006). Prokaryotic nucleotide excision repair: the UvrABC system. *Chem Rev*; 106(2): pp 233-252.

Tsalic M *et al.* (2003). Severe toxicity related to the 5-fluorouracil/leucovorin combination (the Mayo Clinic regimen): a prospective study in colorectal cancer patients. *Am J Clin Oncol*; 26(1): pp 103-106.

Tsodikov OV *et al*. (2005). Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. *Proc Natl Acad Sci USA*; 102(32): pp 11236-11241.

Tsutsumi S *et al.* (2012). Phase II trial of chemotherapy plus bevacizumab as second-line therapy for patients with metastatic colorectal cancer that progressed on bevacizumab with chemotherapy: the Gunma Clinical Oncology Group (GCOG) trial 001 SILK study. *Oncology*; 83(3): pp 151-157.

Tubbs JL *et al*. (2009). Flipping of alkylated DNA damage bridges base and nucleotide excision repair. *Nature*; 459(7248): pp 808-813.

Turco E *et al.* (2013). Understanding the role of the Q338H MUTYH variant in oxidative damage repair. *Nucleic Acids Res*; 41(7): pp 4093-4103.

Tuupanen S *et al.* (2008). Allelic imbalance at rs6983267 suggests selection of the risk allele in somatic colorectal tumour evolution. *Cancer Res*; 68(1): pp 14-17.

Tuupanen S *et al.* (2009). The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signalling. *Nat Genet*; 41(8): pp 885-890.

Tveit KM *et al.* (2012). Phase III trial of cetuximab with continuous or intermittent fluorouracil, leucovorin, and oxaliplatin (Nordic FLOX) versus FLOX alone in first-line

treatment of metastatic colorectal cancer: the NORDIC-VII study. *J Clin Oncol*; 30(15): pp 1755-1762.

Twelves C *et al.* (2005). Capecitabine as adjuvant treatment for stage III colon cancer. *N Engl J Med*; 352(26): pp 2696-2704.

Umar A *et al.* (2004). Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst*; 96(4): pp 261-268.

Van Cutsem E *et al*. (1999). Clinical activity and benefit of irinotecan (CPT-11) in patients with colorectal cancer truly resistant to 5-fluorouracil (5-FU). *Eur J Cancer*; 35(1): pp 54-59.

Van Cutsem E *et al.* (2000). Capecitabine, an oral fluoropyrimidine carbamate with substantial activity in advanced colorectal cancer: results of a randomized phase II study. *J Clin Oncol;* 18(6): pp 1337-1345.

Van Cutsem E *et al.* (2004). Oral capecitabine vs intravenous 5-fluorouracil and leucovorin: integrated efficacy data and novel analyses from two large, randomised, phase III trials. *Br J Cancer*; 90(6): pp1190–1197.

Van Cutsem E *et al.* (2007). Open-label phase III trial of panitumumab plus best supportive care compared with best supportive care alone in patients with chemotherapy-refractory metastatic colorectal cancer. *J Clin Oncol*; 25(13):1658-1664.

Van Cutsem E and Oliveira J (2009a). Advanced colorectal cancer: ESMO clinical recommendations for diagnosis, treatment and follow-up. *Ann Oncol;* Suppl 4:61-3.

Van Cutsem E *et al.* (2009b). Cetuximab and chemotherapy as initial treatment for metastatic colorectal cancer. *N Engl J Med*; 360(14): pp 1408-1417.

Van Cutsem E *et al.* (2009c). Safety and efficacy of first-line bevacizumab with FOLFOX, XELOX, FOLFIRI and fluoropyrimidines in metastatic colorectal cancer: the BEAT study. *Ann Oncol*; 20(11): pp 1842-1847.

van der Burgt I *et al.* (2005). Nijmegen breakage syndrome. *J Med Genet*; 33(2): pp 153-156.

van der Luijt RB *et al*. (1995). APC mutation in the alternatively spliced region of exon 9 associated with late onset familial adenomatous polyposis. *Hum Genet*; 96(6): pp 705-710.

van Gent DC *et al*. (1997). Stimulation of V(D)J cleavage by high mobility group proteins. *EMBO J*; 16(10): pp 2665-2670.

van Kuilenburg AB *et al*. (2000). Clinical implications of dihydropyrimidine dehydrogenase (DPD) deficiency in patients with severe 5-fluorouracil-associated toxicity: identification of new mutations in the DPD gene. *Clin Cancer Res*; 6(12): pp 4705-4712.

van Vuuren AJ *et al*. (1993). Evidence for a repair enzyme complex involving ERCC1 and complementing activities of ERCC4, ERCC11 and xeroderma pigmentosum group F. *EMBO J*; 12(9): pp 3693-3701.

Varon R *et al*. (1998). Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. *Cell*; 93(3): pp 467-476.

Vasen HF *et al*. (1999). New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology*; 116(6): pp 1453-1456.

Vidal AE *et al.* (2001). XRCC1 coordinates the initial and late stages of DNA abasic site repair through protein-protein interactions. *EMBO J*; 20(22): pp 6530-6539.

Vilar E and Grubber SB. (2010). Microsatellite instability in colorectal cancer - the stable evidence. *Nat Rev Clin Oncol*; 7(3): pp 153-162.

Walker JR *et al.* (2001). Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature*; 412(6847): pp 607-614.

Wallace SS *et al*. (2012). Base excision repair and cancer. *Cancer Lett*; 327(1): pp 73-89.

Wang D and Lippard SJ. (2005). Cellular processing of platinum anticancer drugs. *Nat Rev Drug Discov*; 4(4): pp 307-320.

Wang K *et al.* (2010). ANNOVAR: functional annotation of genetic variants from high throughput sequencing data. *Nucleic Acids Res*; 38(16): pp e164.

Wang LL *et al.* (2001). Clinical manifestations in a cohort of 41 Rothmund-Thomson syndrome patients. *Am J Med Genet*; 102(1): pp 11-17.

Wang XW *et al*. (1995). p53 modulation of TFIIH-associated nucleotide excision repair activity. *Nat Genet*; 10(2): pp 188-195.

Wasserman E *et al*. (1997). Severe CPT-11 toxicity in patients with Gilbert's syndrome: two case reports. *Ann Oncol*; 8(10): pp 1049-1051.

Watson AT *et al.* (2008). Gene tagging and gene replacement using recombinase-mediated cassette exchange in Schizosaccharomyces pombe. *Gene*; 407(1-2): pp 63-74.

Webb EL *et al*. (2006). Search for low penetrance alleles for colorectal cancer through a scan of 1467 non-synonymous SNPs in 2575 cases and 2707 controls with validation by kin-cohort analysis of 14 704 first-degree relatives. *Hum Mol Genet*; 15(21): pp 3263-3271.

Weemaes CM *et al.* (1981). A new chromosomal instability disorder: the Nijmegen breakage syndrome. *Acta Paediatr Scand*; 70(4): pp 557-564.

Wei X *et al*. (1996). Molecular basis of the human dihydropyrimidine dehydrogenase deficiency and 5-fluorouracil toxicity. *J Clin Invest*; 98(3): pp 610-615.

Weickhardt A *et al*. (2011). Oxaliplatin induced neuropathy in colorectal cancer. *J Oncol*; 201593.

Welsh MJ and Smith AE. (1993). Molecular mechanisms of CFTR chloride channel dysfunction in cystic fibrosis. *Cell*; 73(7): pp 1251-1254.

Wetzel C *et al.* (2007). A stomatin-domain protein essential for touch sensation in the mouse. *Nature*; 445(7124): pp 206-209.

Wheeler DA *et al.* (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature*; 452(7189): pp 872-876.

Whinney SR *et al*. (2009). Platinum neurotoxicity pharmacogenetics. *Mol Cancer Ther*; 8(1): pp 10-16.

Wildiers H *et al*. (2003). Effect of antivascular endothelial growth factor treatment on the intratumoral uptake of CPT-11. *Br J Cancer*; 88(12): pp 1979-1986.

Williams JS *et al.* (2009). Fission yeast Scm3 mediates stable assembly of Cnp1/CENP-A into centromeric chromatin. *Mol Cell*; 33(3): pp 287-298.

Wolf S *et al*. (2008). Chemotherapy-induced peripheral neuropathy: Prevention and treatment strategies. *Eur J Cancer*; 44(11): pp 1507-1515.

Wolin KY *et al*. (2009). Physical activity and colon cancer prevention: a meta-analysis. *Br J Cancer*; 100(4): pp 611-616.

Won HH *et al.* (2012). Polymorphic markers associated with severe oxaliplatin-induced, chronic peripheral neuropathy in colon cancer patients. *Cancer*; 118(11): pp 2828-2836.

Woo LL *et al.* (2006). The Rothmund-Thomson gene product RECQL4 localizes to the nucleolus in response to oxidative stress. *Exp Cell Res*; 312(17): pp 3443-3457.

Wood RD. (1997). Nucleotide excision repair in mammalian cells. *J Biol Chem*; 272(38): pp 23465-23458.

Wood RD *et al.* (2005). Human DNA repair genes, 2005. *Mutat Res*; 577(1-2): pp 275-283.

Woodcock TM *et al*. (1980). Combination clinical trials with thymidine and fluorouracil: A phase I and clinical pharmacological intervention. *Cancer*; 45(5 Suppl): pp 1135-1143.

Wooster R *et al.* (1995). Identification of the breast cancer susceptibility gene BRCA2. *Nature*; 378(6559): pp 789-792.

Woynarowski JM *et al*. (1998). Sequence- and region-specificity of oxaliplatin adducts in naked and cellular DNA. *Mol Pharmacol*; 54(5): pp 770-777.

Woynarowski JM *et al*. (2000). Oxaliplatin-induced damage of cellular DNA. *Mol Pharmacol*; 58(5): pp 920-927.

Wozniak K and Blasiak J. (2002). Recognition and repair of DNA-cisplatin adducts. *Acta Biochem Pol*; 49(3): pp 583-596.

Wrana JL *et al*. (1994). Mechanism of activation of the TGFβ receptor. *Nature*; 370(6488): pp 341-347.

Wright JB *et al*. (2012). Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single nucleotide polymorhpism in colorectal cancer cells. *Mol Cell Biol*; 30(6): pp 1411-1420.

Wu Y *et al.* (2001). A role for MLH3 in hereditary nonpolyposis colorectal cancer. *Nat Genet*; 29(2): pp 137-138.

Wyman C and Kanaar R. (2006). DNA double-strand break repairs: all's well that ends well. *Annu Rev Genet*; 40: pp 363-383.

Xu G *et al*. (2008). Base excision repair, aging and health span. *Mech Ageing Dev*; 129(7-8): pp 366-382.

Yandell M *et al.* (2011). A probabilistic disease-gene finder for personal genomes. *Genome Res*; 21(9): pp 1529-1542.

Yarnell AT *et al*. (2001). Interaction of FACT, SSRP1, and the high mobility group (HMG) domain of SSRP1 with DNA damaged by the anticancer drug cisplatin. *J Biol Chem*; 276(28): pp 25736-25741.

Yeager M *et al*. (2008). Comprehensive resequence analysis of a 136 kb region of human chromosome 8q24 associated with prostate and colon cancers. *Hum Genet*; 124(2): pp 161-170.

Yonemasu R *et al*. (1997). Characterization of the alternative excision repair pathway of UV-damaged DNA in Schizosaccharomyces pombe. *Nucleic Acids Res*; 25(8): pp 1553-1558.

Youds JL *et al*. (2009). C. elegans: a model of Fanconi anemia and ICL repair. *Mutat Res*; 668(1-2): pp 103-116.

Yu CE *et al*. (1996). Positional cloning of the Werner's syndrome gene. *Science*; 272(5259): pp 258-262.

Yu TW *et al.* (2013). Using whole-exome sequencing to identify inherited causes of autism. *Neuron*; 77(2): pp 259-273.

Yuan F *et al*. (2004). Evidence for involvement of HMGB1 protein in human DNA mismatch repair. *Biol Chem*; 279(20): pp 20935-20940.

Zanke BW *et al*. (2007). Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet*; 39(8): pp 989-994.

Zbuk KM and Eng C. (2007). Hamartomatous polyposis syndromes. *Nature*; 4(9): pp 492-502.

Zhang S *et al*. (2006). Organic cation transporters are determinants of oxaliplatin cytotoxicity. *Cancer Res*; 66(17): pp 8847-8857.

Zhang W *et al*. (2007). FCGR2A and FCGR3A polymorphisms associated with clinical outcome of epidermal growth factor receptor expressing metastatic colorectal cancer patients treated with single-agent cetuximab. *J Clin Oncol*; 25(24): pp 3712-3718.

Zöllner S and Pritchard JK. (2005). Coalescent-based association mapping and fine mapping of complex trait loci. *Genetics*; 169(2): pp 1071-1092.

Zöllner S and Pritchard JK. (2007). Overcoming the winner's curse: estimating penetrance parameters from case-control data. *Am J Hum Genet*; 80(4): pp 605-615.

Zwelling LA *et al*. (1979). DNA-Protein and DNA interstrand cross-linking by cis and trans-platinum (II) diamminedichloride in L1210 mouse leukaemia cells and relation to cytotoxicity. *Cancer Res*; 39(2 Pt 1): pp 365-369.

1000 Genomes Project Consortium; Abecasis GR *et al*. (2010). A map of human genome variation from population-scale sequencing. *Nature*; 28; 467(7319): pp 1061-1073.