Investigation of the Pharmacogenetics of Colorectal Cancer

Submitted for the degree of Doctor of Philosophy at Cardiff University

Michelle Coffey

2015

DECLARATION

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

Signed - Date –

STATEMENT 1

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

Signed - Date –

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 - Date –

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 - Date -

Summary

In this thesis, we aimed to elucidate the functional effects of three rare nonsynonymous variants in *ERCC4* on cell viability, DNA repair and localisation of the ERCC1-XPF repair complex. We also aimed to identify alleles that contribute to extreme adverse reactions to colorectal cancer (CRC) treatment. Previous work identified three rare nonsynonymous variants (Pro379Ser, Arg576Thr and Glu875Gly) in the *ERCC4* gene of patients with severe peripheral neuropathy associated with oxaliplatin (PNAO). All variants were predicted to affect protein function, whilst two collectively contributed to the risk of the phenotype (11.11% of patients with PNAO vs. 4.88% of patients without PNAO; $P = 0.03$).

We genotyped 480 EBV-transformed lymphoblastoid cell lines from healthy humans for *ERCC1* and *ERCC4* polymorphisms. From this data, we selected twelve cell lines for functional analysis (three wild type and three cell lines per *ERCC4* variant). Following treatment with UV or oxaliplatin, we observed negligible differences in survival between wild type and variant cell lines. We identified negligible differences in localisation of the ERCC1-XPF repair complex between wild type and variant cell lines after oxaliplatin treatment. Using the haploid fission yeast, *Schizosaccharomyces pombe*, similarly we observed minimal differences in rad16 localisation between wild type and variant strains.

We observed similar repair rates following oxaliplatin treatment in wild type and variant cell lines. However, following UV treatment, all cell lines carrying Pro379Ser exhibited retarded repair (defined as virtually no repair for the duration of the experiment) whilst all cell lines carrying Arg576Thr and Glu875Gly displayed delayed repair commencing 24 hours after damage. This suggested a functional defect of XPF in the repair of UV damage in cells carrying an *ERCC4* variant.

We used whole exome resequencing data from 10 patients with PNAO and 45 patients without PNAO, to identify causal alleles for the phenotype. We validated one variant in *PAPLN*, suggesting that this gene may cause PNAO. We identified a high rate of false-positive indel calls (98.1%), of which 96.1% contained homopolymers or short tandem repeats (STRs) in their flanking regions. Additionally, we used five filtering strategies to identify causal variants and manually excluded indels that were likely false-positives. In addition to *PAPLN*, this identified a further seven genes (*ASGR2*, *ANXA7*, *OR2T35*, *SLC17A5*, *SMOC2*, *SPERT* and *TAF1D*) that may cause PNAO.

Acknowledgments

I would like to thank the following:

My supervisors, Prof. Jeremy Cheadle and Dr. Richard Adams, for their supervision and guidance throughout this project.

Tenovus for funding this work.

Chris George for his extensive help with the immunofluorescence assay, in terms of training on the confocal microscope and assistance with troubleshooting.

Catherine Naseriyan for help with FACS and technical advice.

Oliver Fleck for his contribution to *Schizosaccharomyces pombe* strain construction and phenotype testing, as well as help with culturing and freezing down strains.

David Fisher for providing information about the patients with severe peripheral neuropathy for the sequencing study.

All of the patients in the MRC COIN trial who generously donated DNA, without whom the project would not have been possible.

Simon Reed and Richard Webster for their advice.

Special thanks to Hannah West for her advice and help with various techniques in the lab, guidance for writing up and for all the background information for this project.

Thanks to Chris Smith and James Colley for help and advice with writing up.

Special thanks to Marc Naven for help with bioinformatic analysis and information during writing up.

Shelley for help and advice regarding technical issues. Laura Thomas for advice about cell culture and for providing a kind ear when needed.

Pete and Julie for technical help and providing a nice atmosphere in the lab. Rebecca Harris for help with sequencing in the early days.

The administration team, Linda, Sherrie, Sathiya, Sally, Alma, Beth and Mark, for helping out with technical things.

Dave Millar and Andy Tee for advice about cell culture.

Marc, Matt, Laura, Ellie, Sara, Kayleigh, Elaine and my friends back home for making the PhD enjoyable.

Carl, for his love and continuous support throughout this process.

Mum, Dad, Elaine and Paul, for their love and support.

Abbreviations

ABC	ATP-binding cassette
aCRC	Advanced colorectal cancer
ADCC	Antibody-dependent cell-mediated cytotoxicity
ADME	Absorption, distribution, metabolism and excretion
AGXT	Alanine glycoxylate transferase
Align-GVGD	Align-Grantham Variation/Grantham Deviation
AOA1	Ataxia with oculomotor apraxia 1
APTX	Aprataxin
AT	Ataxia telangiectasia
ATM	Ataxia telangiectasia, mutated
ATP	Adenosine triphosphate
AU	Arbitrary units
BER	Base excision repair
BGI	Beijing Genomics Institute
BHRF1	BamHI-H right reading frame-1
BLM	Bloom syndrome
BPDE	Benzo(a)pyrene diolepoxide
BRAF	V-raf murine sarcoma viral oncogene homolog B1
BRCA1/2	Breast cancer 1/2, early onset
BSA	Bovine serum albumin
BWA	Burrows-Wheeler Alignment
CBS	Central Biotechnology Services
CCNH	Cyclin H
CCS	Copper chaperone for superoxide dismutase
CEA	Carcinoembryonic antigen
CFTR	Cystic fibrosis transmembrane conductance regulator
ChIP	Chromatin immunoprecipitation
ChIP-chip	Chromatin immunoprecipitation combined with microarray
ChIP-seq	Chromatin immunoprecipitation combined with sequencing
CMT	Charcot-Marie-Tooth disease
CMT1A	Charcot-Marie-Tooth disease type 1A
CMT2A	Charcot-Marie-Tooth disease type 2A

COIN	Continuous vs. intermittent therapy
Co-IP	Co-immunoprecipitation
COX17	Cytochrome C oxidase 17
CPB	Capecitabine
CPD	Cyclobutane pyrimidine dimer
CRA	Colorectal adenoma
CRC	Colorectal cancer
CS	Cockayne syndrome
CSA/B	Cockayne syndrome, complementation group A/B
CTR1	Copper transporter protein 1
CYP	Cytochrome P
DABCO	1,4-diazabicyclo[2.2.2]octane
DACH	1,2-diaminocyclohexane
DAPI	4’,6-diamidino-2-Phenylindole, dihydrochloride
DDB1/2	DNA damage binding protein 1/2
ddNTPs	Dideoxyribonucleotide triphosphates
DFUR	5’-deoxy-5-fluorouridine
dH_2O	Distilled water
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphates
DPN	Diabetic peripheral neuropathy
DPYD	Dihydropyrimidine dehydrogenase
DRG	Dorsal root ganglia
DSB	Double strand break
DSBR	Double strand break repair
EBNA1-3C	Epstein-Barr nuclear antigen 1-3C
EBNA-LP	Epstein-Barr nuclear antigen-leader protein
EBV	Epstein-Barr virus
ECACC	European Collection of Cell Cultures
EGFR	Epidermal growth factor receptor
EGTA	Ethylene-bis(oxyethylenenitrilo)tetraacetic acid
EL	Ectopia lentis
ELISA	Enzyme-linked immunosorbent assay
ER	Oestrogen receptor

ERCC(1-6)	Excision repair cross complementation group 1-6
Exo	Exonuclease
FA	Fanconi anaemia
FACS	Fluorescence activated cell sorting
FAD	Flavin adenine dinucleotide
FAP	Familial adenomatous polyposis
FBS	Foetal bovine serum
FDA	US food and drug administration
fdUMP	Fluorodeoxyuridine monophosphate
fdUTP	Fluorodeoxyuridine triphosphate
FOLFIRI	5-Fluorouracil, leucovorin and irinotecan
FOLFOX	5-Fluorouracil, leucovorin and oxaliplatin
FSC-A	Forward scatter according to area
FSC-H	Forward scatter according to height
FUTP	Fluorouridine triphosphate
GATK	Genome Analysis Toolkit
GBM	Glioblastoma multiforme
GC	Guanine-cytosine base pair
GD	Geleophysic dysplasia
GG-NER	Global genomic nucleotide excision repair
GI	Gastrointestinal
GRHPR	Glycoxylate reductase-hydroxypyruvate reductase
GST	Glutathione S-transferase
GWAS	Genome-wide association study
HAH1	Human antioxidant homologue 1
HNPCC	Hereditary nonpolyposis colorectal cancer
HPA	Health Protection Agency
HR	Homologous recombination
HRP	Horseradish peroxidase
ICL	Interstrand crosslink
ICLR	Interstrand crosslink repair
Indel	Insertion or deletion
kb	Kilobase
kDa	Kilo-Dalton
KRAS	Kirsten rat sarcoma viral oncogene homolog

LD	Linkage disequilibrium
LDH	Lactate dehydrogenase
LIG4	DNA Ligase IV
LM	Ligation mediated
LMP1-2B	Latent membrane protein1-2B
MAF	Minor allele frequency
MAP	MUTYH associated polyposis
Mb	Megabase
MCL	Metachromatic leucodystrophy
mCRC	Metastatic colorectal cancer
MCV	Motor conduction velocity
MFN2	Mitofusin-2
MLPR	Multiplex ligation-dependent probe amplification
MMC	Mitomycin C
MMR	Mismatch repair
MP	Maximum projection
MRC	Medical Research Council
mRNA	Messenger ribonucleic acid
MSC	Mesenchymal stem cell
MSI	Microsatellite instability
MT	Metallothionein
MTHF	5,10-methylenetetrahydrofolate
MTHFR	Methylenetetrahydrofolate reductase
mTOR	Mammalian target of rapamycin
N/C F	Nuclear/cytoplasmic fluorescence
NCBI	National Center for Biotechnology Information
NER	Nucleotide excision repair
NGS	Next-generation DNA sequencing
NHEJ	Non-homologous end-joining
NRAS	Neuroblastoma ras viral oncogene homolog
NSCLC	Non-small cell lung carcinoma
NT	No treatment
OCT1/2	Organic cation transporter 1/2
ORF	Open reading frame
OXA	Oxaliplatin

PALB1	Partner and localiser of BRCA1
PBMC	Peripheral blood mononuclear cells
PBS	Phosphate buffered saline
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PFS	Progression free survival
PHE	Public Health England
PI	Propidium iodide
PI3KCA	Phosphatidylinositol-4,5-biphosphate 3-kinase
PIPES	1,4-Piperazinediethanesulfonic acid
PN	Peripheral neuropathy
PNAO	Peripheral neuropathy associated with oxaliplatin treatment
POLD	Polymerase (DNA directed), delta
POLE	Polymerase (DNA directed), epsilon
PR	Progesterone receptor
q-PCR	Real time, quantitative polymerase chain reaction
RECQ4	RecQ like helicase 4
RFI	Relative-fold increase
RNA	Ribonucleic acid
ROI	Region of interest
ROS	Reactive oxygen species
RPA	Replication protein A
RPM	Revolutions per minute
RPMI-1640	Roswell Park Memorial Institute 1640
RTS	Rothmund Thomson syndrome
SAP	Shrimp alkaline phosphatase
SB	Slot blot
SCAN1	Spinocerebellar ataxia with axonal neuropathy
SF	Serum-free
siRNA	Small interfering RNA
SNP	Single nucleotide polymorphism
S. cerevisiae	*Saccharomyces cerevisiae*
S. pombe	*Schizosaccharomyces pombe*
SSB	Single strand break
SSC	Squamous cell carcinoma

SSC-A	Side scatter according to area
ssDNA	Single stranded DNA
SSRI	Selective serotonin reuptake inhibitor
STR	Short tandem repeat
SV	Structural variant
TA	Thymine-adenine
TBP	TATA box binding protein
TCGA	The Cancer Genome Atlas
TC-NER	Transcription-coupled NER
TDP1	Tyrosyl-DNA phosphodiesterase-1
TFIIH	Transcription factor II human
TNM	Tumour, node, metastasis
TP	Thymidylate phosphorylase
TPMT	Thiopurine methyltransferase
TS	Thymidylate synthase
TSC	Tuberous sclerosis
TTD	Trichothiodystrophy
TTD-A	Trichothiodystrophy, complementation group A
TTP	Thrombocytopenic purpura
UGT	Uridine diphosphate glucouronosyl-transferase
UV	Ultraviolet
UV-C	Ultraviolet C wavelength radiation
UVDE	UV-damaged DNA endonuclease
UVER	UVDE-dependent excision repair
VEGF	Vascular endothelial growth factor
WB	Western blotting
WES	Whole exome sequencing
WGS	Whole genome sequencing
WS	Werner syndrome
WT	Wild type
XELIRI	Capecitabine and irinotecan
XELOX	Capecitabine and oxaliplatin
XFE	XPF-ERCC1 progeroid syndrome
XP	Xeroderma pigmentosum
XP(A-G)	Xeroderma pigmentosum, complementation group A-G

XRCC1	X-ray repair cross-complementing protein 1
YEA	Yeast extract agar
YEL	Yeast extract liquid
3’-UTR	3’ untranslated region
5-FU	5-fluorouracil
5’-UTR	5’ untranslated region
6-4 PPs	6-4 photoproducts

Contents

List of figures

Chapter One – Introduction

Chapter Three – Viability

Chapter Four – Immunofluorescence

Chapter Five – DNA Damage ELISA

Chapter Six – PNAO study

List of tables

Chapter One – Introduction

Chapter Three – Survival

Chapter Four – Immunofluorescence

Chapter Five – DNA Damage ELISA

Chapter Six – PNAO study

Chapter Seven – General discussion

Chapter One – Introduction

1.1 Pharmacogenetics of cancer

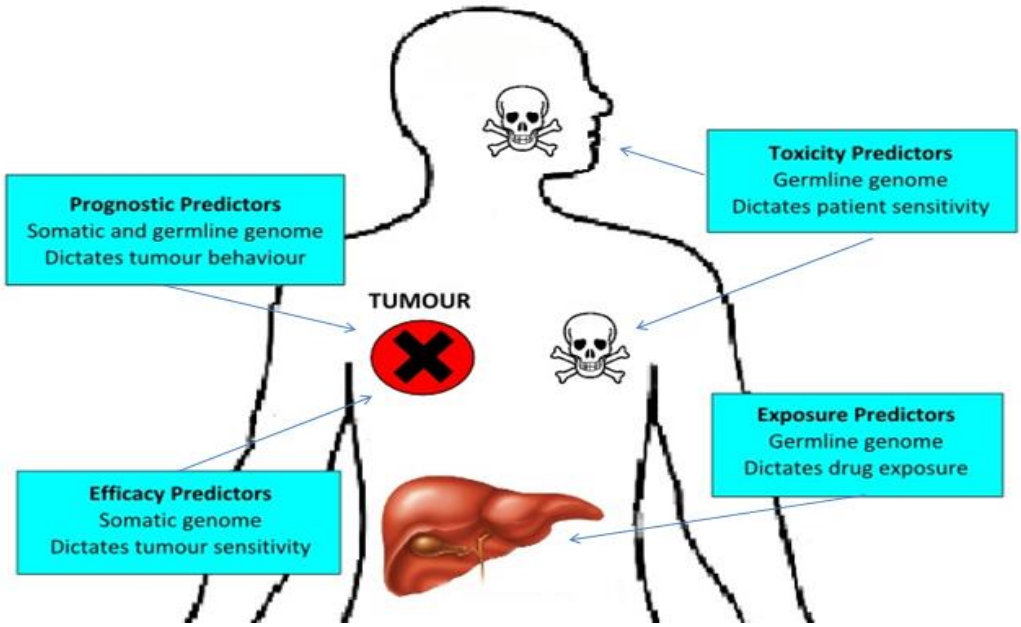
Cancer is a leading cause of mortality and morbidity in the developed world, with failure to respond to drug therapy posing a life threatening risk. The emerging field of pharmacogenetics aims to understand how genetic variation between individuals affects drug efficacy and toxicity (Wheeler *et al*. 2013). For any drug used to treat a patient population, there is significant variability in response to that drug, resulting in ineffective treatment for a proportion of that population. This is disadvantageous for a number of reasons; firstly, there is a reduction in the patient's quality of life resulting from treatment-related toxicity, which is only acceptable while the tumour is responding. Secondly, in cases where a patient's tumour fails to respond to therapy, this allows more opportunity for the tumour to grow and potentially become harder to treat. In some cases, therapy with one drug can cause resistance, resulting in a tumour that is resistant to other lines of therapy (Dean *et al*. 2005). Finally, ineffective therapy incurs a financial cost. All of these issues provide a rationale for developing strategies to select optimal treatment options for patients on an individual basis (Hertz and McLeod, 2013).

When devising a cancer treatment plan, there are two relevant genomes to consider: the tumour's somatic genome and the patient's germline genome. Tumour prognosis is at least partially dependent on variation in these genomes, which affect tumour behaviour irrespective of treatment. The somatic genome dictates tumour sensitivity to treatment, which can be useful for efficacy prediction, whilst the germline genome primarily dictates drug pharmacokinetics, which can indirectly affect efficacy and toxicity. Additionally, variants in the germline genome can directly influence patient sensitivity to a drug's side effects (McLeod, 2013). A summary of pharmacogenetic predictive factors of cancer treatment outcomes is depicted in Figure 1.1.

1.1.1 Germline genetics

The germline genome is responsible for encoding proteins involved in all functioning processes throughout the body; therefore, variation in this genome

(A)



(B)

Outcome Prediction	Genome	Genes	Treatment Decision	Example
Prognosis	Somatic/ Germline	Oncogenes; tumour suppressors; DNA repair genes; metastasis genes	Dose	*BRCA2* gene expression in breast and ovarian cancers
Exposure	Germline	Transporter enzymes	Dose	*UGT1A1* in irinotecan and SN-38 exposure
Efficacy	Somatic	Drug targets; tumour drivers	Dose	*BRAF* V600E and *TSC1* in vemurafenib and everolimus, respectively
Toxicity	Germline	Drug targets; cellular response genes; genes that affect toxicity mechanisms	Dose	*TUBB2A*, *FGD4* and *FANCD2* in paclitaxel-induced peripheral neuropathy

Figure 1.1 Pharmacogenetic predictors of cancer treatment outcomes. (A) Tumour prognosis is at least partially influenced by variation in the germline and somatic genomes, which control tumour behaviour independent of treatment. The somatic genome dictates tumour sensitivity to therapy and therefore allows prediction of efficacy, whilst the germline genome governs drug pharmacokinetics, which may indirectly influence efficacy and toxicity. (B) All outcomes are affected by the drug dose selected. Different gene groups have variable effects on outcomes; therefore mutations in genes encoding drug targets affect drug pharmacodynamics, resulting in changes in efficacy and toxicity. Alternatively, variants in oncogenes and tumour suppressor genes may adversely affect prognosis.

may have implications for any organ system. Inherited variants in genes responsible for maintenance of cellular or DNA replication can predispose a patient to specific cancer types (Li, 1995). Germline variants present in genes encoding enzymes or transporters can adversely affect a patient's exposure to treatment or the severity of side effects they experience (McLeod and Evans, 2001). A number of germline genetic variants have been identified across multiple cancer types through candidate gene and genome wide association studies (GWAS) (Table 1.1).

1.1.1.1 Prognostic predictors

It is widely accepted that germline genetic variants can contribute to susceptibility for cancer (Theodoratou *et al.* 2012; Siddiq *et al.* 2012), although less is known about whether these variants have consequences for prognosis. The Cancer Genome Atlas (TCGA) is a coordinated project designed to compile information about somatic mutations, copy number, variant expression and epigenetic modification in specific tumour types. TCGA has also collected germline DNA and outcome data in an attempt to characterise the effects of germline variants on cancer prognosis (Collins and Barker, 2007).

Initial results have been published using patient cohorts with glioblastoma, ovarian, breast, lung and colorectal cancers. These have shown a slightly higher survival rate in patients with colorectal tumours displaying high microsatellite instability (MSI), suggesting that mutation status of mismatch repair (MMR) genes could be useful in predicting prognosis (Cancer Genome Atlas Network, 2012). Interestingly, it was reported that patients with ovarian cancer that were carrying germline founder variants in either the *BRCA1* or *BRCA2* genes demonstrated superior survival than patients without variants in either gene (E23fs deletion and Q1756 insertion in *BRCA1*; S1982 deletion in *BRCA2*) (Cancer Genome Atlas Research Network, 2011). Further analysis of this data confirmed that patients carrying a S1982 deletion in the RAD51 binding domain of *BRCA2* had longer overall and progression-free survival (PFS) compared with other patients (Yang *et al.* 2011). Bolton *et al*. (2012) performed a large pooled analysis, which reported that patients with this *BRCA2*

Drug	Mechanism of action	Cancer type(s)	Germline markers	Variants (Nucleotide changes)	Phenotype	References
Irinotecan	Inhibition of topoisomerase I (involved in DNA replication and transcription)	Colorectal, lung	*UGT1A1*	rs8175347 (UGT1A1*28 and 3156G>A)	Neutropenia, diarrhoea	Innocenti *et al.* 2004; Iyer *et al.* 2002; Hoskins *et al.* 2007
Tamoxifen	Inhibition of the oestrogen receptor	Hormone-receptor-positive breast	*CYP2D6*	rs16947 (2850C>T) rs1065852 (100C>T) rs28371706 (1023C>T) rs28371725 (2988G>A) rs35742686 (2549delA) rs3892097 (1846G>A) rs5030655 (1707delT) rs5030656 (CYP2D6*9) rs59421388 (3183G>A)	Tamoxifen metabolism, progression-free and overall survival	Schroth *et al.* 2007; Kiyotani *et al.* 2010; Schroth *et al.* 2009; Regan *et al.* 2012; Goetz *et al.* 2011; Nowell *et al.* 2005; Okishiro *et al.* 2009
Mercaptopurine	Inhibition of purine nucleotide synthesis (required for DNA replication)	Paediatric acute lymphoblastic leukaemia	*TPMT*	rs1142345 (TPMT*2) rs1800460 (TPMT*3A) rs1800462 (TPMT*3C)	Myelosuppresion	Relling *et al.* 1999; Relling *et al.* 2011
Methotrexate	Inhibition of folic acid metabolism (required for DNA replication)	Paediatric acute lymphoblastic leukaemia	*SLCO1B1*	rs11045879 (3886C>T)	Methotrexate clearance, GI toxicity	Trevino *et al.* 2009; Lopez-Lopez *et al.* 2011; Ramsey *et al.* 2012

Table 1.1 A selection of germline genetic variants associated with cancer drug-induced phenotypes.

variant showed superior 5-year survival (52%) compared with *BRCA1* carriers (44%) and non-carriers (36%).

Several studies have identified several germline genes that may affect metastatic potential in multiple breast cancer subtypes; *SIPA1*, *BRD4* and *RRP1B* have all been implicated in breast cancer metastasis (Park *et al.* 2005; Crawford *et al.* 2006; Pei *et al.* 2013; Crawford *et al.* 2008; Hsieh *et al.* 2009) and are associated with regulation of the Rap GTPase-activating protein SPA-1. It has been hypothesised that these germline genetic variants could influence tumour growth and metastasis by affecting the tumour microenvironment (Shimizu *et al.* 2011), although the mechanism for this is currently unknown.

1.1.1.2 Exposure predictors

Current clinical practice in chemotherapy treatment involves administration of a drug dose intended to maximise drug exposure with limited risk of toxicity to the patient, typically based on patient weight and body surface area (Felici *et al.* 2002). Most drugs demonstrate substantial interpatient pharmacokinetic variability, some of which is explained by patient size (McLeay *et al.* 2012); however the remaining variability is a consequence of variation in genes associated with absorption, distribution, metabolism, or excretion (ADME) of the drug and its intermediates.

All drugs require a series of metabolising enzymes and drug transporters to perform the ADME processes, which are encoded by germline genes. Variation in exons of these genes may affect protein function, whereas intronic germline variants are more likely to influence protein expression. Several pharmacogenetic studies have identified germline variants that affect patient exposure to anti-cancer drugs. Phase I enzymes including cytochrome P450 (CYP) can directly activate or inactivate the administered drug (Bosch *et al.* 2006). Genetic polymorphisms have been reported in *CYP2C8*, *CYP2C9*, *CYP2C19*, *CYP2D6* and *CYP3A4* isoforms, which are associated with metabolism of taxane drugs including tamoxifen for treatment of oestrogen receptor-positive breast cancer. Tamoxifen is metabolised by the CYP system to form the active metabolites N-desmethyltamoxifen, formed by CYP3A4, and 4-hydroxytamoxifen and endoxifen, both formed by CYP2D6 (Scripture and

Figg, 2006). Stearns *et al*. (2002) assessed plasma endoxifen levels in patients with breast cancer that were prescribed tamoxifen as adjuvant therapy in addition to paroxetine, a selective serotonin reuptake inhibitor (SSRI) metabolised by CYP2D6 and prescribed for management of hot flushes. They reported lower plasma endoxifen concentrations in women treated with both drugs, and interestingly the endoxifen concentrations decreased more rapidly in patients that were wild-type for *CYP2D6*, compared with homozygous non-functional genotype carriers (reduction of 64% and 24% respectively). This was confirmed by Jin *et al*. (2005), suggesting that tamoxifen metabolite concentrations are directly affected by SSRIs and polymorphisms in *CYP2D6*.

Phase II enzymes including uridine diphosphate glucouronosyl-transferases (UGTs) activate the active form of the drug or its metabolites via conjugation (Bosch *et al.* 2006). Variants in genes encoding various isoforms of UGTs have been implicated in varying drug exposures to irinotecan, a compound used for treatment of metastatic colorectal cancer (CRC). Irinotecan is metabolised to form SN-38, an active compound responsible for both efficacy and toxicity associated with the drug. SN-38 is inactivated via glucuronidation via the isoform UGT1A1 and excreted through biliary elimination. Transcription of UGT1A1 is variable in patients due to germline variation in the gene's promoter region; typically six thymine-adenine (TA) repeats form in the TATA box, whilst in a proportion of patients a seventh TA forms, known as UGT1A1*28. This isoform has been shown to diminish glucuronidation (Iyer *et al.* 1999), resulting in increased SN-38 levels and slower clearance of the drug in patients carrying a homozygous *28 allele (Ando *et al.* 1998). Homozygous and heterozygous variants in genes encoding UGTs have been associated with greater toxicity levels resulting from prolonged exposure to SN-38 in patients with CRC in some clinical studies (Paoluzzi *et al.* 2004; Saito *et al.* 2009), although others have suggested the association is less clinically important (Toffoli *et al.* 2006). Irinotecan is discussed in more detail in section 1.3.2.

The association between variants in pharmacokinetic genes and clinical outcome is an indirect one caused by a direct influence of the variant on drug exposure. Clinical interpretation of such pharmacokinetic studies would result in

reduction of dose, altered administration or discontinuation of the drug (Bosch *et al.* 2006).

1.1.1.3 Toxicity predictors

Clinical factors such as drug interactions or impaired liver function can affect a patient's likelihood of developing side effects by influencing drug exposure, similar to mechanisms outlined in section 1.1.1.2. A potent example is the increased toxicity seen in patients carrying polymorphisms in genes encoding proteins involved in transport and excretion undergoing taxane treatment for various cancers; reduced clearance of the drug arising from truncated excretion proteins results in drug retention and severe toxicity (Jabir *et al.* 2012). Alternatively, individuals receiving equivalent drug exposure can develop toxicities of variable severity due to several biological mechanisms; these include high endogenous expression of the drug's target protein, compromised response to a drug's cytotoxic mechanism or an underlying pathology (Hertz and McLeod, 2013).

Cells exhibiting high expression of a drug target may be highly sensitive to the drug in question; β-tubulin, the drug target of paclitaxel, exists as several isoforms that are differentially expressed throughout the peripheral nervous system. A common side effect of paclitaxel treatment is peripheral neuropathy. An analysis of germline variants in the *TUBB2A* promoter, which encodes the β-tubulin class IIa isoform located primarily in neurons, identified a single nucleotide polymorphism (SNP) in the promoter which was associated with increased gene transcription and reduced peripheral neuropathy (Leandro-Garcia *et al.* 2012).

Additionally, cells may carry germline polymorphisms that render them unable to respond to a stimulus, such as compromised DNA repair in response to a DNA damaging chemotherapeutic. Sucheston *et al*. (2011) reported a significant association between SNPs and haplotypes present in the DNA repair gene *FANCD2* with neuropathy risk with paclitaxel treatment.

Finally, where a patient has an underlying condition which presents with similar symptoms to the drug's toxicity, this condition may be aggravated during

treatment. Charcot-Marie-Tooth syndrome, a familial hereditary peripheral neuropathy (PN) disorder, has many subtypes caused by different germline variants; Charcot-Marie-Tooth disease type 4H is caused by a variant in *FGD4*, encoding FGD1-related F-actin-binding protein (Delague *et al.* 2007). Patients carrying germline variants in *FGD4*, or other genes encoding proteins that interact with FGD4, are therefore at increased risk of suffering extreme sensitivity to drug-induced PN associated with paclitaxel. Baldwin *et al*. (2012) confirmed an association between a novel *FGD4* polymorphism and increased incidence of sensory PN in patients with breast cancer during paclitaxel treatment. It is likely that variants in other genes associated with familial PN syndromes will give rise to enhanced sensitivity to chemotherapy-induced PN (Hertz and McLeod, 2013).

1.1.2 Somatic genetics

During tumourigenesis, healthy cells develop genetic aberrations that result in an inability to regulate their replication. These malignant cells carry a somatic genome, which acquires additional genetic abnormalities through repeated cellular replication (Gerlinger *et al.* 2013). Some of these additional acquired variants may enhance the tumour's pathogenic qualities, particularly variants found in genes that regulate DNA replication or control metastatic potential (Ganem and Pellman, 2012). Both initial and acquired variants affect primary tumour behaviour, including its sensitivity to certain drugs and its potential for metastasis. Therefore, an understanding of the tumour's underlying somatic genetics may help to predict the tumour prognosis and response to therapy. A number of somatic variants have been identified that can influence prognosis and efficacy in multiple tumour types (Table 1.2).

According to Knudson's two-hit hypothesis of tumour suppressor genes, an initial inherited mutation increases the probability of disease due to an increased chance of loss of the second allele within somatic cells. Alternatively, sporadic disease of the same type requires a somatic mutation to occur on both alleles (Knudson, 1996; Figure 1.2).

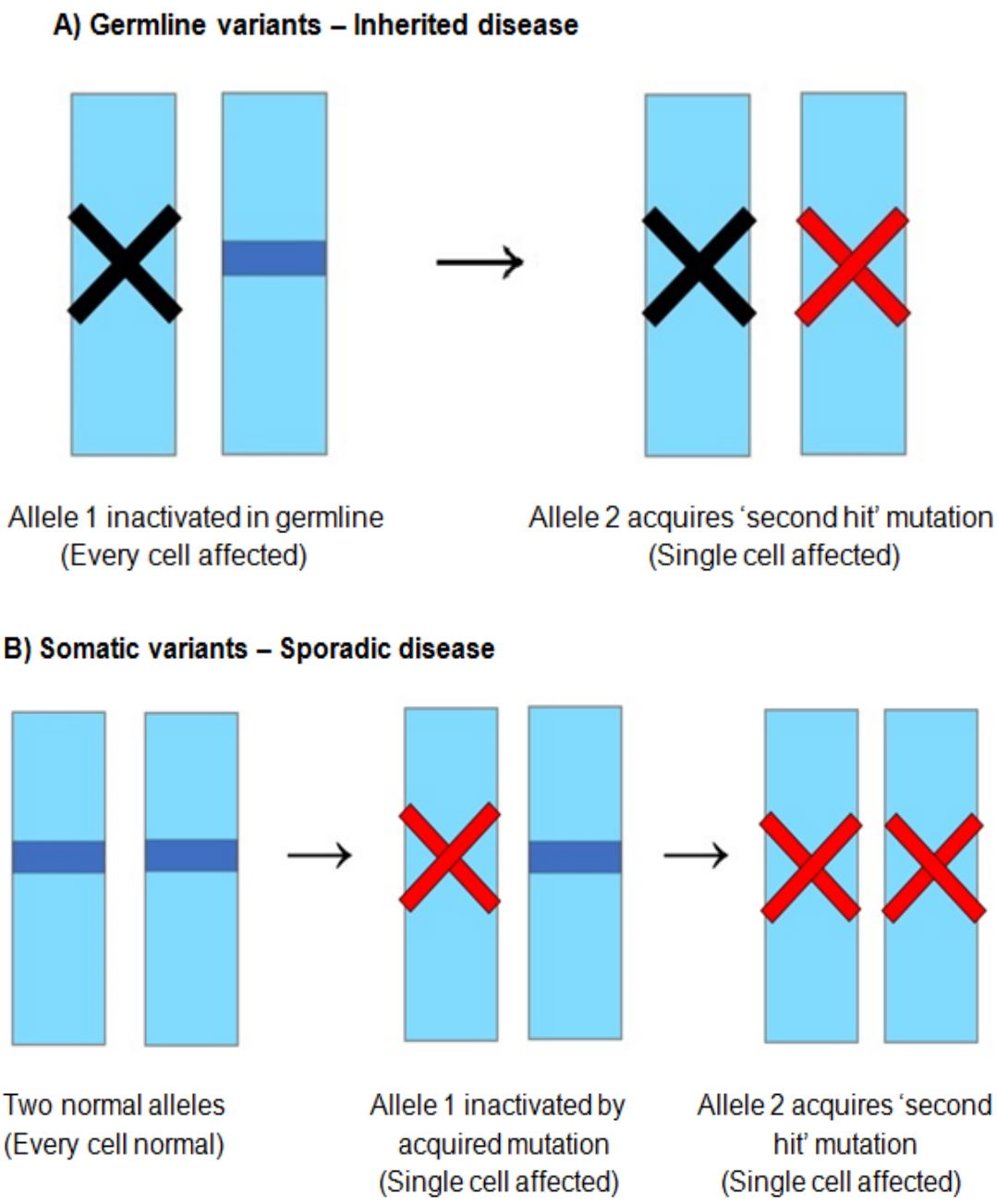


Figure 1.2 Knudson's 'two-hit' hypothesis of tumourigenesis in inherited (A) and sporadic (B) forms of cancer. (A) In inherited disease, a variant of one allele is inherited in every germline cell, whereas the second allele variant is acquired in a single cell. (B) In sporadic disease, two normal alleles are inherited in every cell. One allele becomes inactivated by sporadic mutation, followed by a second sporadic mutation in the second allele. Inactivation of tumour suppressor genes in this way can contribute to cellular growth advantages and tumourigenesis (Knudson, 1985).

Drug	Drug target	Cancer type(s)	Somatic markers	Variants (Amino acid changes)	Phenotype
Cetuximab	EGFR	Colorectal, head and neck	*EGFR* and *KRAS*	*KRAS* variants (Gly12Asp, Gly12Ala, Gly12Val, Gly13Asp) *EGFR* (EGFRvIII)	*KRAS* mutations associated with resistance to cetuximab (Linardou *et al.* 2008); mutant EGFR contributes to increased tumour growth and drug resistance (Sok *et al.* 2006)
Panitumumab	EGFR	Colorectal	*EGFR* and *KRAS*	*KRAS* variants (Gly12Asp, Gly12Ala, Gly12Val, Gly12Ser, Gly12Arg, Gly12Cys, Gly13Asp)	Associated with panitumumab resistance and reduced overall survival (Amado *et al.* 2008)
Lapatinib	ERBB2/ HER2 receptor	Breast	*ERBB2/HER2*	*HER2* (L755S, P780insertion)	Associated with lapatinib resistance in HER2-negative breast cancer patients (Bose *et al.* 2013)
Trastuzumab (Herceptin®)	ERBB2/ HER2 receptor	Breast, lung	*ERBB2/HER2*	*HER2*; in-frame insertions in exon 20 (G776V)	HER2 variants associated with increased risk of lung and breast cancer development in non smokers (Shigematsu *et al.* 2005)
Gefitinib and Erlotinib	EGFR	Lung	*EGFR*	*EGFR*; exon 21 (L858R); present in the activation loop of the tyrosine kinase domain	Increased response to gefitinib and erlotinib, followed by acquired resistance in the presence of this variant (Pao *et al.* 2005)
Imatinib	BCR-ABL, KIT and PDGFRα tyrosine kinases	Chronic myeloid leukaemia, GI cancers	Philadelphia chromosome, *KIT* and *PDGFRA*	*ABL* (Thr315Ile) *KIT* (V654A, T670I, T670E, S709F) *PDGFRA* (Val561Asp)	BCR-ABL variants associated with imatinib resistance and accelerated progression of disease (Roche-Lestienne *et al.* 2002); Activating mutations in KIT and PDGFRA increase progression of stromal GI tumours and cause secondary resistance (Hirota *et al.* 2003)

Table 1.2 A selection of somatic mutations in cancer pharmacogenetics.

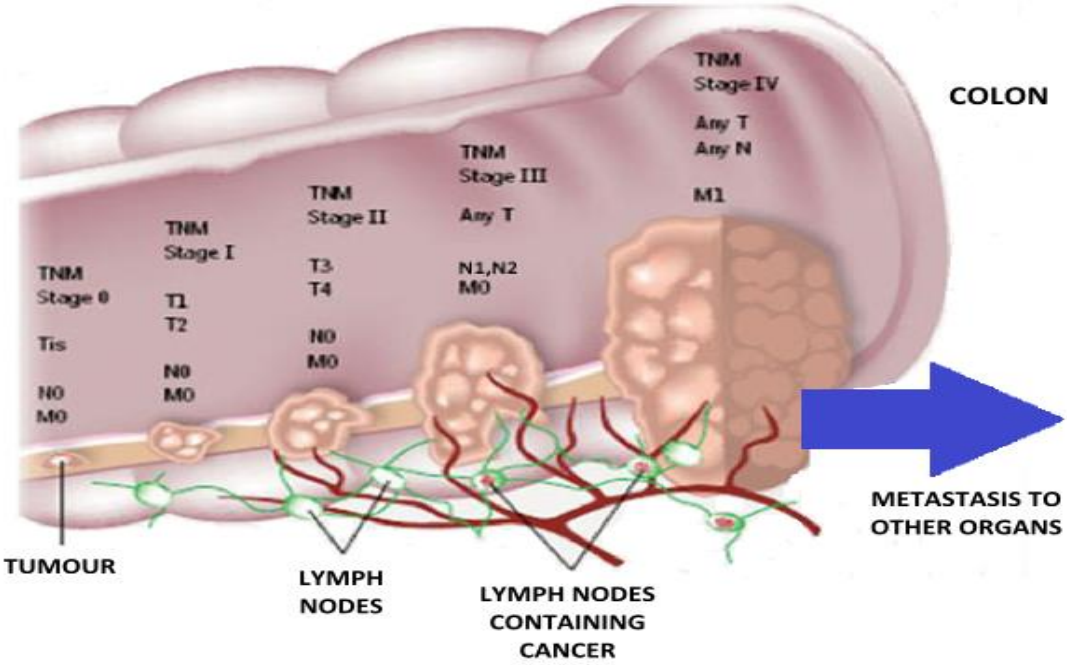
1.1.2.1 Prognostic predictors

Although germline mutations may affect prognosis in certain cancer types (see section 1.1.1.1), the majority of genetic determination of tumour prognosis is dependent on the tumour genome. Prognosis depends on local tumour growth and distant tumour spread, which can be measured using the tumour, node and metastasis (TNM) grading system (Figure 1.3). Prognostic grading systems exist for numerous tumour types and are essential for selecting a therapy regimen in the clinic. However, somatic genetics have a large impact on the aggressiveness of a tumour irrespective of treatment choice, and therefore the addition of genetic prognostic determinants to these staging systems would be hugely beneficial to clinicians with regard to making appropriate treatment decisions for patients on an individual basis (Greene and Sobin, 2008).

Somatic genetic screening for cancer is most commonly used in the treatment of breast cancer. Breast tumours are classified according to expression of several receptors with chemotherapeutics targeted at each receptor specifically; oestrogen (ER), progesterone (PR) and human epidermal growth factor receptor-2 (HER2)-positive tumours are typically treated with tamoxifen- and trastuzumab-based regimens, for ER/PR and HER2-positive tumours respectively (Arpino *et al.* 2005; Romond *et al.* 2005; Gianni *et al.* 2010). Sorlie *et al* (2001) developed an alternative system that classifies tumours according to their 'intrinsic subtype' based on gene expression, which is largely concordant with receptor-based classification. The intrinsic subtypes have been shown to correlate with disease recurrence (Nielsen *et al.* 2010), but receptor-based classification remains the current clinical approach to treatment (Goldhirsch *et al.* 2009).

Olivier *et al*. (2006) demonstrated an association between somatic *TP53* mutations in patients with primary breast cancer and higher incidence of breast cancer-related death, compared with patients without *TP53* variants, irrespective of tumour size, node status or hormone receptor expression. They also reported that patients carrying the R248W variant in *TP53* in the absence of the progesterone receptor demonstrated the poorest prognosis.

(A)



(B)

TNM Staging				
Stage	**Tumour size (T)**	**Lymph nodes (N)**	**Metastasis (M)**	**Description**
0	Tis	N0	M0	(Tis) Cancer *in situ* – tumour confined to mucosa (N0) No nodes affected (M0) No metastasis
I	T1 T2	N0 N0	M0 M0	(T1) Tumour invades submucosa (T2) Tumour invades muscle layer
II	T3 T4	N0 N0	M0 M0	(T3) Tumour invades subserosa or beyond (T4) Tumour invades adjacent organs
III	T1-2 T3-4 Any	N1 N1 N2	M0 M0 M0	(N1) Metastasis to 1-3 lymph nodes (T1-2) Tumour spread to either submucosa or muscle layer (T3-4) Tumour spread to either subserosa or adjacent organs (N2) Metastasis to 4 or more lymph nodes
IV	Any	Any	M1	(M1) Distant metastasis

Figure 1.3 TNM Staging of cancer. (A) Schematic of a colonic tumour at each TNM stage. (B) Number stages and corresponding TNM staging of CRC with description of tumour growth at each stage (adapted from Greene and Sobin, 2008). Tis = Tumour *in situ*.

1.1.2.2 Efficacy predictors

Tumours are particularly sensitive to drugs that specifically target the driver of tumour progression. As a result of advances in genetic screening technologies, researchers can now perform somatic gene profiling of tumours, which allows researchers and clinicians the opportunity to target deficiencies in the tumour's growth or metastatic potential (Albertson *et al.* 2003). In order to develop this translational work, drugs need to be developed to target specific biomarkers that can be screened for in a tumour prior to commencing treatment. Personalised treatment using this approach is being piloted currently (Tsimberidou *et al.* 2012; Roychowdhury *et al.* 2011); however, prospective validation of this approach is challenging (Mandrekar and Sargent, 2009). This approach has the potential to transform cancer treatment, by identifying patients with a tumour carrying a biomarker that increases efficacy whilst sparing non-responders from unnecessary treatment.

There are two strategies required for increasing access to personalised cancer treatment; firstly, identification of genetic biomarkers related to tumour sensitivity to drugs currently in use, and secondly, the design of new drugs to target oncogenic mutations found in tumour types. Everolimus, an inhibitor of the mammalian target of rapamycin (mTOR), was approved for treatment of advanced kidney cancer in 2009, followed by approval for other tumour types in more recent years. The drug was originally approved for the treatment of tuberous sclerosis (TS), a disease characterised by formation of benign tumours throughout the body which is caused by a mutation in *TSC1* (Meikle *et al.* 2008; Franz, 2011). Following complete response to everolimus lasting >2 years in a single patient with metastatic bladder cancer, the tumour was genotyped and a deletion in *TSC1* was detected. Based on this finding, Iyer *et al* (2012) genotyped a further 96 bladder tumours and identified six patients out of thirteen with deletions in *TSC1* that demonstrated impressive time to response (4.1 vs. 1.8 months) and longer duration of treatment (7.7 vs. 2.0 months). This is one case of an existing drug that has been shown to potentially target a biomarker present in bladder tumours, which now requires prospective validation and incorporation into clinical practice (Hertz and McLeod, 2013).

1.1.2.3 Overlap between somatic and germline genetics

Variation in germline DNA may influence which somatic mutations a tumour is most likely to acquire; a study in squamous cell carcinomas (SCC) of the skin revealed that chromosomal aberrations in these tumours were more similar within than between individuals, consistent with the hypothesis that background variation influences the pattern of somatic mutations within an individual (Dworkin *et al.* 2010). As a result, recent studies have used somatic mutations as endophenotypes to investigate germline genetic variants that may confer risk for acquiring specific somatic mutations (Jonsson *et al.* 2005; Kiemeney *et al.* 2010; Kilpivaara *et al.* 2009). Such studies have associated germline variants in *MC1R* and *EGFR* (in particular exon 19 microdeletions) with increased likelihood of somatic variants that increase risk of BRAF-mutant melanoma and non-small cell lung carcinoma (NSCLC) development, respectively (Landi *et al.* 2006; Liu *et al.* 2011).

1.2 Colorectal Cancer

CRC is the fourth most common cancer in the UK, accounting for approximately 13% of all new diagnoses. Over 40,000 new cases are diagnosed each year, with an overall lifetime risk of developing the disease estimated at 5%. Advances in treatment and early screening programmes for CRC have reduced mortality rates by approximately 50% in the last four decades; however, despite this improvement, there are currently 16,000 deaths due to CRC each year in the UK (Cancer Research UK, Bowel cancer statistics, 2010).

Age is an important risk factor governing CRC development; 85% of new cases are diagnosed in patients over the age of sixty (Cancer Research UK, Bowel cancer statistics, 2010). Other non-modifiable risks include a personal history of adenomatous polyps or inflammatory bowel disease, or a family history of CRC or adenomatous polyps, which are considered a precursor lesion of CRC (Haggar and Boushey, 2009). Deaths due to CRC development in patients with ulcerative colitis and Crohn's disease account for one third of CRC-related deaths annually (Itzkowitz and Yio, 2004).

Environmental factors associated with CRC onset include poor diet, obesity, increased alcohol consumption and cigarette smoking; these lifestyle habits are often associated with Westernised cultures, which may account for the disproportionately high incidence of CRC in Westernised continents (Europe and North America) compared with Africa and Asia (>10-fold increase in CRC incidence in Westernised continents) (Ferrari *et al.* 2007; Jemal *et al.* 2010).

1.2.1 Heritable colorectal cancer

There is a strong heritable component associated with CRC, which is indicated by the identification of a number of hereditary syndromes in recent decades. Advances in genetics have allowed elucidation of the underlying molecular mechanism of CRC and provided better treatment options and screening programmes for both patients and their relatives (Lynch *et al.* 2007).

Hereditary CRC is defined as a tumour caused by germline mutations that give rise to a CRC 'syndrome'. With regard to CRC specifically, these 'syndromes' are a collection of signs and symptoms that characterise tumour progression in the presence of high penetrance germline mutations in the mismatch repair (MMR) pathway or the *APC* gene (Lynch and De la Chapelle, 2003). Alternatively, so-called 'sporadic' cases of CRC arise due to low penetrance germline variants and somatic mutations that are acquired over time, without the presence of a germline mutation or signs of a hereditary CRC syndrome (Valle, 2014).

Approximately 75% of all CRC cases are sporadic in nature, while the remaining 25% are comprised of several heritable CRC syndromes (Amersi *et al.* 2005). Epidemiological studies have estimated that, of these inherited CRC cases, at least 15% occur in dominantly inherited patterns, of which the two most well defined types are familial adenomatous polyposis (FAP) and hereditary nonpolyposis colorectal cancer (HNPCC) (Kinzler and Vogelstein, 1996).

Of the 25% of all CRC cases that are inherited, the underlying genetic cause of tumour progression has been identified in approximately 12% cases, whilst the remaining 13% are considered 'familial' due to a strong family history (Amersi *et al.* 2005). The latter are cases where the tumour's morphology and

patient's symptoms are similar to those of hereditary CRC syndromes, but where no germline causal mutation has been identified. In these cases, there is typically evidence of a family history of CRC in previous generations; therefore it becomes difficult to ascertain whether familial occurrence is the result of predisposition or chance. Similarly, seemingly 'sporadic' cases may have a familial component that is not immediately apparent, either due to chance, an increased risk of CRC that is inherited, a small family or incomplete diagnostics, therefore making the distinction between sporadic and familial CRC cases blurred (De la Chappelle, 2004; Figure 1.4).

1.2.1.1 High penetrance alleles

An allele is attributed high penetrance if its presence dramatically increases the likelihood of a specific phenotype. These characteristics are usually highly heritable and therefore relatively easy to track and identify. Approximately 6% of all CRC cases are attributable to high penetrance mutations (Jasperson *et al.* 2010; Patel and Ahnen, 2012). A substantial proportion of high penetrance, heritable CRC syndromes are directly caused by inactivating mutations in tumour suppressor genes. Most loss of function mutations are recessive, therefore requiring loss of both alleles in order to allow a cell to become cancerous.

1.2.1.2 HNPCC

Hereditary non-polyposis colorectal cancer (HNPCC; OMIM #120435), or Lynch syndrome, is an autosomal dominant disease characterised by formation of CRC and an increased risk of several extra-colonic malignancies, including ovary, endometrium, stomach, hepatobiliary tract, urinary tract, brain and skin cancers (Aarnio *et al.* 2006). It is the most common hereditary CRC syndrome, accounting for 2-6% of all CRC cases (Lynch *et al.* 2006). CRC is the most common malignancy associated with HNPCC, with an 80% lifetime risk and an average age of onset of 45 years. The syndrome arises as a result of inherited defects in key genes in the mismatch repair (MMR) pathway, which have been implicated in HNPCC disease etiology; *MLH1* and *MSH2* germline mutations, *MSH6* mutations and *PMS2* mutations are reported to account for 90%, 7-10%

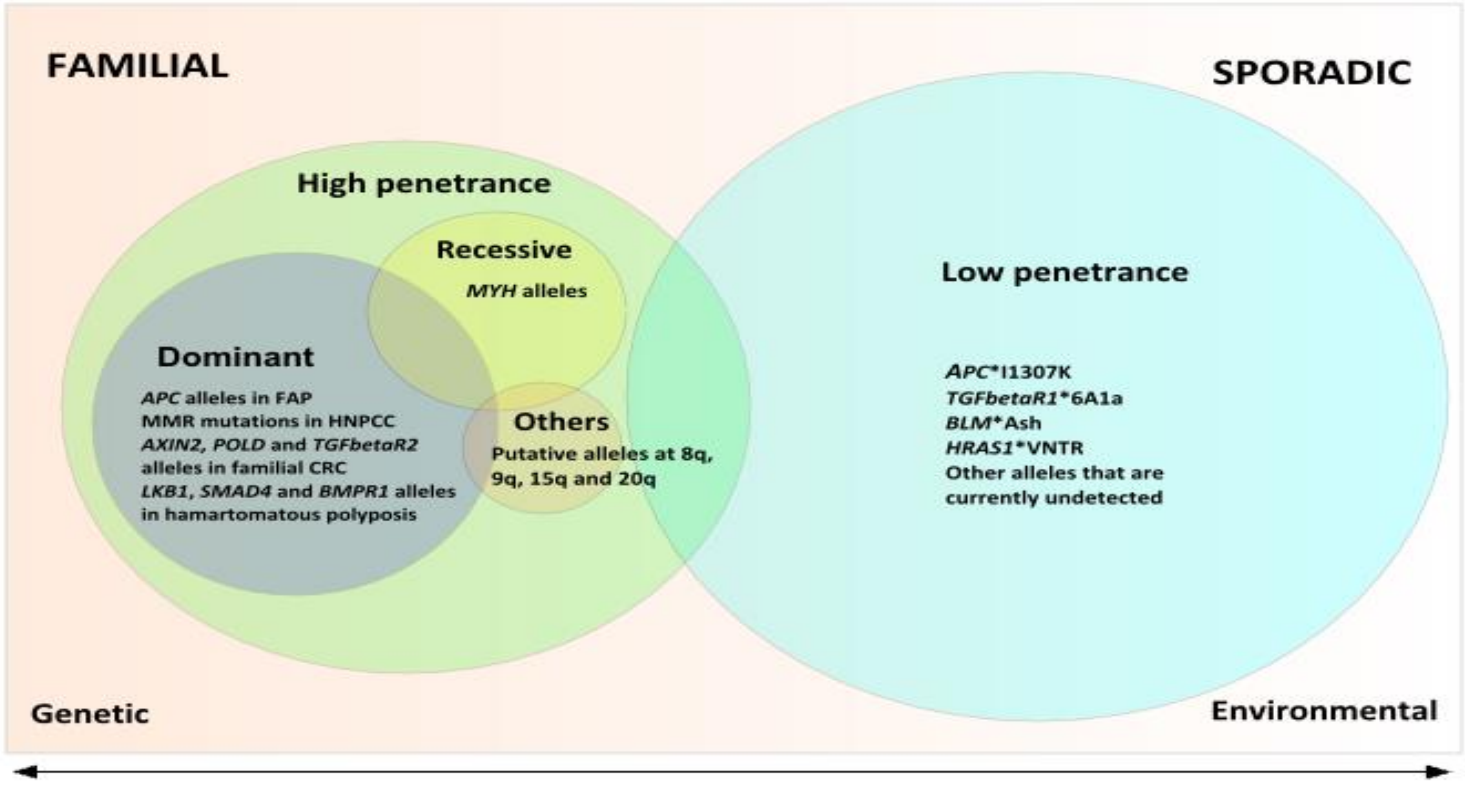


Figure 1.4 Schematic of the continuum of familial and sporadic hereditary CRC (adapted from De La Chapelle, 2004). Highly penetrant causative mutations in FAP, HNPCC and hamartomatous polyposis syndromes account for a small proportion of hereditary CRC cases, with little environmental influence. However, several other low penetrance alleles contribute to hereditary CRC incidence in a cumulative way, due to environmental influences and interactions between genes. These alleles are also believed to contribute to sporadic CRC cases. Additionally, modifier genes are believed to affect the genetic and environmental factors that contribute to CRC; therefore, the distinction between familial and sporadic CRC cases and between genetic and environmental predisposition factors has become increasingly blurred.

and 5% of mutations in families with HNPCC, respectively (Miyaki *et al.* 1997; Berends *et al.* 2002; Peltomaki, 2003; Senter *et al.* 2008). Deficiencies in the MMR pathway result in an increased mutation rate and secondary somatic mutations that ultimately give rise to the cancer phenotype in HNPCC (Burt and Neklason, 2005; Kinzler and Vogelstein, 1996).

1.2.1.3 MAP

Mut-YH associated polyposis (MAP; OMIM #604933) is an autosomal recessive disease characterised by formation of tens to a few hundred colonic adenomatous polyps, which contributes to a significantly increased lifetime risk of CRC (Jasperson *et al.* 2010). MAP was identified by somatic analysis of *APC* in patients presenting with multiple CRAs without a germline *APC* mutation. This revealed a disproportionately high expression of G:C→T:A transversions, resulting in excessive truncation mutations in tumours resulting from acquisition of somatic variants (Jones *et al.* 2002). These transversion mutations are formed during oxidative damage to guanine bases, resulting in formation of 8-oxo-7, 8dihydro-2'-deoxyguanosine (8-oxoG), a highly mutagenic lesion that results in G:C→T:A transversion during subsequent replication cycles (Shibutani *et al.* 1991). The 8-oxoG lesions are removed during base excision repair (BER), during which the DNA glycosylases OGG1 and MUTYH remove 8-oxoG and the mispaired adenine, respectively (Shinmura *et al.* 2012). An additional common transversion of this type in the *KRAS* gene (c.34G>T in codon 12) is present in 64% of patients with MAP CRC, which has led to recommendations for somatic *KRAS* screening of CRC tumours to identify patients eligible for *MUTYH* germline molecular genetic testing (Lipton *et al.* 2003; Nielsen *et al.* 2009).

1.2.2 Low penetrance alleles

The majority of inherited CRC cases occur without any identifiable genetic causes, prompting the hypothesis that the remaining heritable CRC component could be attributed to common or rare, low penetrance alleles (Fearnhead *et al.* 2005). An allele is considered low penetrance if its effect on phenotype is small, although it may contribute significantly to disease incidence due to its relative frequency within the population (Valle, 2014).

1.2.2.1 'Common disease, common variant' hypothesis

According to the 'common disease, common variant' hypothesis, an individual variant risk of contributing to disease onset is relatively small (Bodmer and Bonilla, 2008). However, the relatively common incidence of such alleles in the population, and their ability to interact with one another polygenically, significantly increase the likelihood that they will affect disease incidence. This hypothesis helps to explain the variation witnessed in complex diseases such as CRC. In recent years, a number of susceptibility variants for CRC have been identified using GWAS.

1.2.2.2 GWAS of CRC variants

Several factors have allowed use of GWAS; firstly, the completion of the HapMap project has enabled knowledge of linkage disequilibrium (LD) that captures variation across the genome to be readily accessed (International HapMap Consortium, 2003). Secondly, the collection of large case and control cohorts have allowed sample numbers large enough to overcome statistical power issues often faced during statistical analyses of large populations. Finally, technological advances in platforms for high throughput genotyping have allowed screening of large numbers of variants across thousands of samples at a reasonable cost with rapid turnaround.

Several GWAS studies in Caucasian populations of European ancestry have been performed and identified >20 low penetrance susceptibility alleles for CRC risk. Although individually all variants show a modest contribution to overall risk, it is estimated that they could collectively account for approximately 7% of CRC familial risk (Dunlop *et al.* 2012).

Despite the impact of GWAS in advancing current knowledge of CRC risk, there are limitations in their ability to detect novel variants. These include difficulty in acquiring a large enough sample number to supply enough power for detection of modest effect sizes, as well as a high rate of false positive readings. This requires validation of variants, which can be costly and labour intensive. Additionally, the importance of ruling out other ethnic populations to avoid population stratification can further restrict the sample sizes for research.

1.2.2.3 'Common disease, rare variant' hypothesis

Detection of rare variants usually involves candidate searches of genes associated with disease aetiology, and rare variants have been shown to contribute to the phenotype of complex diseases (Pritchard, 2001). The 'common disease, rare variant' hypothesis proposes that inherited susceptibility may be an additive effect of low frequency dominantly and independently acting variants in various genes. In CRC, it is thought that 20-30% of cases are a result of multifactorial inherited susceptibility arising from rare, low penetrance alleles without a clear cut familial inheritance pattern (Bodmer and Bonilla, 2008).

Research by Fearnhead *et al*. (2004) has collectively associated rare variants in the Wnt signalling genes *AXIN1* and *CTNNB1*, in addition to MMR genes *MLH1* and *MSH2*, with a slightly increased risk of colorectal adenoma (CRA) onset. Additionally, non-synonymous variants Ile1307Lys and Glu1317Gly in *APC* have been shown to predispose to CRCs and CRAs in a number of studies (Frayling *et al.* 1998; Lamlum *et al.* 2000; Azzopardi *et al.* 2008). Additional rare variants in *APC* have also been reported, suggesting a potential low penetrance effect of these rare alleles (Azzopardi *et al.* 2008).

1.3 Treatments for CRC

Treatment for CRC may involve surgery, radiotherapy, chemotherapy or targeted therapies depending on the stage of disease at diagnosis (Carethers, 2008). The most significant prognostic factor for CRC progression is tumour staging, and treatment options rely heavily on this (Section 1.1.2.1; Figure 1.2). The five year relative survival rates drop to approximately 11% and 12% for patients with stage IV colon and rectal cancers respectively, compared with 92% and 87% respectively, when diagnosed with stage I cancer (National Cancer Institute SEER database, 2014). Approximately 80% of patients diagnosed with stage I-III CRC undergo surgical resection of the tumour, resulting in curative treatment. Additionally, adjuvant treatment with radiotherapy or chemotherapy is common (Winawer *et al.* 2003). However, as 25% of patients present with metastatic CRC and approximately 50% progress to metastases, further treatments are necessary and treatment at this stage is

often challenging (Xu *et al.* 2011). Only 20% of patients with hepatic metastases are candidates for curative surgery, meaning that currently chemotherapy remains the treatment of choice for advanced CRC (aCRC) (Stangl *et al.* 1994).

A number of drugs used for the treatment of aCRC have been identified and reported to improve survival and response via clinical trials (Table 1.3).

1.3.1 Fluoropyrimidines

The fluoropyrimidine class of drugs is widely used for aCRC treatment. Fluorouracil (5-FU, Efudex) has been administered for CRC treatment for more than fifty years. It is an analogue of uracil and therefore uses the same cellular transport systems to enter the cell following parenteral administration. 5-FU can be considered a 'fraudulent' nucleotide; upon cell entry, the drug is converted to several active metabolites including fluorodeoxyuridine monophosphate (fdUMP), fluorodeoxyuridine triphosphate (fdUTP) and fluorouridine triphosphate (FUTP). These metabolites firstly insert themselves into newly synthesised DNA strands, and secondly inhibit thymidylate synthetase (TS) activity in the production of deoxythymine monophosphate, resulting in inhibition of DNA synthesis (Longley *et al.* 2003). 5-FU is most commonly administered in addition to the folate supplement leucovorin (5'-formyltetrahydrofolate, folonic acid). Leucovorin is metabolised to the methyl donor 5,10-methylenetetrahydrofolate (MTHF), which interacts with fdUMP. Leucovorin has been shown to increase cellular levels of the donor and to stabilise the TS-fdUMP complex, which is essential for 5-FU metabolism (Radparvar *et al.* 1989). Studies have revealed that administration of 5-FU alongside leucovorin results in clinical synergism and doubles the overall response rate in aCRC (Advanced Colorectal Cancer Meta-Analysis Project, 1992).

Capecitabine (CPB, Xeloda) is an oral fluoropyrimidine drug that is absorbed via the gut wall (Lamont and Schilsky, 1999). Capecitabine is converted to 5'-deoxy-5-fluorouridine (DFUR) in the liver by the enzymes carboxylesterase and cytidine deaminase. DFUR is then transported to tumour cells and converted to 5-FU by thymidylate phosphorylase (TP). TP is expressed preferentially in tumour cells compared with normal tissue, which allows metabolism of 5-FU at a higher rate in tumour cells (Kosuri *et al.* 2010).

Regimen (Clinical Trial)	Effects seen/Outcome of trial
5-FU, leucovorin and oxaliplatin (FOLFOX); (National Cancer Institute trial; led by the North Central Cancer Treatment Group)	Significantly improved response rate and median survival in patients with previously untreated aCRC (Goldberg *et al.* 2004)
CPB and oxaliplatin (XELOX); (Large phase II, open label study; involved 13 centres in Europe, North America and Israel)	Similar response rate and overall survival to FOLFOX regimen; additional advantage of reducing prolonged infusion times with 5-FU due to substitution of oral CPB (Cassidy *et al.* 2004)
FOLFOX plus irinotecan (FOLFIRI); (Prospective phase II study at the Institute of Oncology Ljubljana, Slovenia)	Significantly improved response to treatment (Patt *et al.* 2007)
XELOX plus irinotecan (XELIRI); (Prospective phase II study at the Institute of Oncology Ljubljana, Slovenia)	Increased efficacy, improved safety and improved convenience for patients with aCRC (Patt *et al.* 2007); more acceptable toxicity levels than FOLFIRI, although patients displayed similar overall survival and PFS compared with those treated with FOLFIRI (Skof *et al.* 2009)
Irinotecan; (A series of phase II studies in Japan, North America and Europe)	Efficacy in mCRC when used alone as first line treatment (Cunningham *et al.* 2001) and as second line treatment after 5-FU failure (Guglielmi and Sobrero, 2007)
Irinotecan in combination with 5-FU and leucovorin; (A series of phase II studies in Japan, North America and Europe)	Improved overall survival, PFS and response rates when used as a first-line therapy for metastatic CRC (Saltz *et al.* 2000; Douillard *et al.* 2000)
Cetuximab (monotherapy or in combination with irinotecan regimens); (Phase II, non-randomised, multicentre trial)	Improved prognosis in patients with aCRC with chemotherapy-refractory EGFR-expressing tumours (Saltz *et al.* 2004)
Panitumumab; (Phase III, open label, multicentre trial; treatment regimens were stratified according to centres based in Western Europe, Canada and Australia vs. rest of world)	Increased PFS as a first-line treatment alongside FOLFOX in mCRC patients (Douillard *et al.* 2010); panitumumab efficacy relies on wild type KRAS and BRAF status (Amado *et al.* 2008; Di Nicolantonio *et al.* 2008)
Bevacizumab; (Eastern Cooperative Oncology Group Study E2300; Phase III randomised, open label, multi-institutional trial)	Statistically significant and clinically meaningful improvement in overall survival and PFS in patients with mCRC, when used in combination with fluoropyrimidine based regimens (Kabbinavar *et al.* 2003; Hurwitz *et al.* 2004; Giantonio *et al.* 2007; Saltz *et al.* 2008)

Table 1.3 Summary of clinical trials investigating treatments for CRC.

As a result, addition of capecitabine to fluorouracil regimens has been shown to reduce systemic exposure of 5-FU and therefore any associated toxicity (Van Cutsem *et al.* 2001). 5-FU is administered alongside leucovorin and oxaliplatin as part of the FOLFOX regimen, whilst CPB is given in combination with oxaliplatin as part of the XELOX regimen for either first or second line treatments for metastatic CRC (Rothenberg *et al.* 2008). Alternatively, the two regimens may be administered in addition to irinotecan as part of the FOLFIRI and XELIRI regimens, for use as first and second line treatments for aCRC, although not as adjuvant therapy (Patt *et al.* 2007).

1.3.2 Irinotecan

Irinotecan (Camptosar) is a plant alkaloid isolated from the *Camptotheca acuminata* tree, which functions as a topoisomerase I inhibitor. Topoisomerase I is a nuclear enzyme involved in relaxation of super-coiled DNA during replication (Saltz *et al.* 2000). The enzyme introduces a single transient nick into double-stranded DNA during repair and replication, allowing strand rotation and subsequent relief of the torsional strain created by supercoiling (Rivory *et al.* 2002). Irinotecan is processed by intestinal and hepatic carboxylesterases to form the active metabolite SN38, which has been reported to demonstrate a 1000-fold higher anti-tumour activity than irinotecan (Iyer *et al.* 1998). SN38 functions to stabilise the topoisomerase-DNA complex, preventing reannealing of the DNA strand after nicking, which ultimately leads to replication fork stalling and apoptosis (Pommier, 2013; Rudolf *et al.* 2013).

1.3.3 Oxaliplatin

Oxaliplatin (Eloxatin) is a third generation platinum compound developed in 2002 and approved, in combination with 5-fluorouracil, for treatment of metastatic CRC. Oxaliplatin molecules consist of 1,2-diaminocyclohexane (DACH) carrier group and a bidentate oxalate group, which becomes displaced non-enzymatically following drug absorption, resulting in formation of DACH intermediates (Kidani *et al.* 1978). These active intermediates react with DNA, in particular with adenine and guanine bases, to form multiple crosslinks (Woynarowski *et al.* 2000). Oxaliplatin treatment results in formation of three types of lesions: DNA-intrastrand crosslinks, DNA-interstrand crosslinks and

DNA-protein crosslinks. Approximately 90% of lesions are intrastrand adducts between two adjacent guanine bases or, less commonly, between adjacent adenine and guanine bases (Chaney *et al.* 2005); the remaining 10% are made up of interstrand crosslinks and DNA-protein crosslinks (Woynarowski *et al.* 2000).

During intrastrand adduct formation, the DNA strand becomes distorted the DACH moiety bends approximately 30 degrees towards the major groove, resulting in DNA distortion. Formation of intrastrand crosslinks, in addition to secondary lesions that form due to an accumulation of damage, eventually result in cell cycle arrest, apoptosis and initiation of immunologic reactions (Springler *et al.* 2003; Woynarowski *et al.* 2000; Alcindor and Beauger, 2011). Prior to the development of oxaliplatin, CRC was thought to have intrinsic resistance to other platinum agents (Rixe *et al.* 1996).

1.3.4 Targeted therapies

Stratified cancer treatment has led to the development of therapies to target specific weaknesses or advantages displayed by tumour cells. A number of monoclonal antibodies have been developed to seek out antigens present in cancer cells, in an attempt to increase efficacy and reduce the severe toxicity that arises as a result of chemotherapy. However, this treatment option is relatively expensive and these drugs also give rise to side effects, meaning that investigation of pharmacogenetic reasons to explain different responses between individuals could be crucial for effective use of these drugs.

1.3.4.1 Cetuximab

Cetuximab (Erbitux) is a recombinant chimeric monoclonal antibody that was approved for CRC treatment in 2004 after it was shown to improve prognosis in aCRC either as a monotherapy or in combination with irinotecan in chemotherapy-refractory EGFR-expressing tumours (Van Cutsem *et al.* 2011; Cunningham *et al.* 2004). However, it was shown that cetuximab was ineffective in aCRC treatment when combined with oxaliplatin-based regimens (Maughan *et al.* 2011).

The human epidermal growth factor receptor (EGFR) is involved in ligand binding and subsequent regulation of downstream cellular signalling cascades including the Ras/Raf/MEK/MAPK and PI3K-Akt pathways, which monitor cell survival and growth (Friday and Adjei, 2008). Overexpression of EGFR is correlated with poor prognosis in a number of cancers including CRC, non-small cell lung cancer (NSCLC), pancreatic ductal adenocarcinoma and cervical carcinoma (Spano *et al.* 2005; Selvaggi *et al.* 2004; Ueda *et al.* 2004; Kersemaekers *et al.* 1999). Cetuximab targets the extracellular ligand-binding domain of the EGFR and blocks receptor activation by endogenous ligands, effectively preventing the receptor-dependent transduction pathway. This triggers a number of anti-tumour effects including cell cycle arrest in G1 leading to induction of apoptosis (Huang *et al.* 1999), in addition to inhibition of angiogenesis and metastasis (Bou-Assaly and Mukherji, 2010). There is also evidence that cetuximab, as an IgG1 antibody, stimulates antibody-dependent cell-mediated cytotoxicity (ADCC) in which the Fc region of the antibody is recognised as an antigen. This results in an immune reaction in which tumour cells expressing EGFR on their cell surface are targeted and destroyed, suggesting that this is an important mechanism of the drug's action (Kimura *et al.* 2007).

1.3.4.2 Panitumumab

Similar to cetuximab, panitumumab (Vectibix) is another EGFR inhibitor that selectively targets the extracellular binding domain and blocks downstream signal transduction. Panitumumab is an entirely humanised monoclonal antibody that appears to act through a slightly different epitope, as shown by studies of cetuximab-resistant but panitumumab-sensitive cell lines (Montagut *et al.* 2012). The drug has been approved for treatment of aCRC since 2006 and is frequently used alongside standard chemotherapy regimens (Van Cutsem *et al.* 2007).

1.3.4.3 Bevacizumab

Bevacizumab (Avastin) is another humanised monoclonal antibody developed specifically to target the vascular endothelial growth factor (VEGF)-A ligand to prevent it binding to the VEGF receptor. The VEGF system is involved

primarily in promotion of angiogenesis and regulation of endothelial cell proliferation, both of which are crucial for tumour development. Where tumour vasculature forms to enhance tumour growth, often the vasculature structure becomes structurally and functionally imbalanced, resulting in elevated interstitial pressure and subsequent non-uniform perfusion throughout the tumour (Ellis, 2006). This can lead to hypoxia and upregulation of VEGF, which drives selection of highly aggressive tumour cells; therefore VEGF levels correlate not only with extent of angiogenesis but also with prognosis (Goel and Mercurio, 2013; Hockel, 2001; Dubois and Demetri, 2006).

1.3.5 Toxicities associated with therapy

Chemotherapeutic drugs and targeted treatments cause toxicities of varying degrees, with significant interpatient variation. Management of side effects is crucial for maintaining quality of life whilst maximising drug efficacy, and the spectrum of drug related toxicities can vary from mild to dose-limiting in nature, resulting in withdrawal from treatment in the most severe cases (Cassidy and Misset, 2002). In addition to general side effects resulting from destruction of rapidly dividing cells, toxicities may be more specific to a drug's mechanism of action (Table 1.4).

1.3.5.1 Fluoropyrimidines

The main toxicities associated with 5-FU with leucovorin treatment are gastrointestinal (GI) epithelial damage resulting in diarrhoea, nausea, vomiting, stomatitis, neutropaenia and hand-foot syndrome (De Gramont *et al.* 2000). The severity of toxicity varies according to different regimens, although infusion of 5-FU is typically tolerated better than bolus administration, due to fewer peaks in exposure to the drug with the latter option (Spicer *et al.* 1988). Similar side effects are experienced with CPB treatment, though at a reduced frequency (Cassidy *et al.* 2002). However, hand-foot syndrome occurs in 50% of patients receiving CPB, a considerably higher incidence than those receiving 5-FU with leucovorin. This is commonly characterised by dysthesia, erythema and, in more severe cases, ulceration and blistering of the skin, in particular the hands and feet (Gressett *et al.* 2006), which is believed to be the result of elevated levels of thymidine phosphorylase, the CPB metabolising enzyme, within the

Year	Therapy	Advance	Toxicity
1962	5-Fluorouracil	FDA approved 5-FU for treatment of aCRC	GI epithelial damage; hand foot syndrome; neutropenia
1990	Adjuvant therapy	Chemotherapy becomes most common treatment following surgery; improved survival after surgical resection by 40%	Higher risk of infection; fatigue; anaemia; contusions; nausea; diarrhoea; alopecia
1996-98	Irinotecan	FDA and EMA approve use of irinotecan alongside 5-FU and leucovorin (FOLFIRI) in first line treatment and as second line monotherapy for aCRC	Hyper-stimulation of cholinergic system; neutropenia
1996-99	Oxaliplatin	EMA approved use of oxaliplatin in combination with 5-FU and leucovorin (FOLFOX) for second line treatment of aCRC	Acute and chronic peripheral neuropathy
2001-4	Capecitabine	FDA and EMA approved use of capecitabine in treatment of aCRC in combination with irinotecan and oxaliplatin in the XELIRI and XELOX regimens, respectively	Anaemia; nausea; diarrhoea; fatigue
2004-5	Bevacizumab	FDA and EMA approved addition of bevacizumab to the FOLFIRI and XELIRI regimens for aCRC	Hypertension
2004 & 2008	Cetuximab	FDA and EMA approved cetuximab use for aCRC as monotherapy or in combination with irinotecan (2004); mutations in codons 12 and 13 in the EGFR pathway gene *KRAS* found to result in ineffective treatment (2008)	Acneiform skin rash; alopecia; trichomegaly (Karapetis *et al.* 2008)
2006-7	Panitumumab	FDA and EMA approved panitumumab use for aCRC as monotherapy, as first line treatment alongside FOLFOX and as second line treatment alongside FOLFIRI	Acneiform skin rash; alopecia; trichomegaly
2009-10	Cetuximab and Panitumumab	FDA and EMA altered guidelines for use of EGFR inhibitors to account for mutations in codons 12 and 13 in *KRAS*, which result in ineffective therapy	As above
2012	Regorafenib	FDA approved use of regorafenib for aCRC refractory to other chemotherapeutics	Hand-foot syndrome; fatigue; diarrhoea; hypertension (Grothey *et al.* 2013)
2013	Aflibercept	EMA approved use of aflibercept for treatment of aCRC refractory to oxaliplatin-based regimens	Higher risk of infection; anaemia; petechia; fatigue; diarrhoea; fistula development (Van Cutsem *et al.* 2012)

Table 1.4 Major therapeutic advances in the treatment of aCRC and their respective toxicities.

epidermis (Milano *et al.* 2008). In severe cases, this side effect may lead to treatment discontinuation due to a reduced quality of life (Surjushe *et al.* 2008).

1.3.5.2 Irinotecan

Irinotecan causes acute dose-limiting toxicities associated with hyperstimulation of the cholinergic system, resulting in abdominal cramps, diarrhoea, emesis, hypotension and bradycardia (Blandizzi *et al.* 2001). Experiments using animal models have shown that irinotecan effectively inhibits acetycholinesterases and stimulates muscarinic receptors (Zambrowicz and Sands, 2003). Additional toxicities include delayed onset of diarrhoea resulting from circulation of high concentrations of SN38 in the intestine following hepatic elimination. This side effect is severe in approximately 40% of patients (Yamamoto *et al.* 2008). The anti-cholinergic drug, atropine, has been shown to alleviate the acute cholinergic symptoms effectively (Cheng *et al.* 2015), whilst the delayed onset diarrhoea responds well to high doses of loperamide (Takasuna *et al.* 2006).

1.3.5.3 Oxaliplatin

Peripheral neuropathy is the main dose limiting side effect associated with oxaliplatin treatment and as such presents a major problem for the treatment of CRC. It is more common for patients withdraw from therapy due to the onset of peripheral neuropathy associated with oxaliplatin (PNAO) than to die as a result of disease progression, and as PNAO is not correlated with drug response it is considered an avoidable toxicity (Whinney *et al.* 2009). There are currently no existing treatment options to reduce PNAO symptoms (Wolf *et al.* 2008), although ongoing clinical trials are investigating whether vitamin D supplements during oxaliplatin-based therapy can alleviate side effects (Higa, 2011).

PNAO occurs in two forms; acute, transient neuropathy and chronic, cumulative sensory neuropathy. The acute, transient form occurs in most patients and has a rapid onset, occurring during or within hours of infusion. Patients may experience distal sensory and motor toxicities that are particularly aggravated by the cold. Sensory toxicity includes paresthesia and dysesthesia in the extremities and jaw. Additionally, 2% of patients report a transient cold-

induced pharyngolaryngeal dysesthesia, resulting in a feeling of difficulty breathing. Occasionally patients will also develop transient motor toxicities such as involuntary muscle spasms, although the incidence of this is much less frequent than sensory neuropathy (Waisif Saif and Reardon, 2005).

Chronic, cumulative sensory neuropathy is present in 10-15% of patients once a cumulative dose of 780-850mg/m^2 has been reached (De Gramont *et al.* 2000; Grothey *et al.* 2003). Symptoms consist of temperature-independent dysesthesia and paresthesia of extremities, similar to the acute form. However, symptoms occur in cycles and increase in intensity with cumulative doses, resulting in impaired sensation and coordination over time. Manifestation of chronic, cumulative neuropathy may be so severe that patients become unable to perform normal daily activities (Waisif Saif and Reardon, 2005). These symptoms are usually reversible; the majority of patients recover from Grade 3 neurotoxicity to grade 1 or lower within 12 months of discontinuing therapy (Grothey, 2003).

Acute PNAO has been associated with interference of voltage gated sodium channels indirectly via chelation of calcium ions by oxalate, an oxaliplatin metabolite (Grolleau *et al.* 2001), whilst the chronic form has been associated with accumulation of platinum in the dorsal root ganglia, resulting in direct toxicity to nerve cells (Argyriou *et al.* 2008; Ta *et al.* 2006). PNAO manifests as several separate symptoms and, as such, has been classified to allow oncologists to monitor dosage according to the type and severity of side effects experienced by patients (Wasif Saif and Reardon, 2005; Table 1.5).

Currently there is little knowledge of genes that confer predisposition to PNAO. An association between chronic PNAO and a coding variant in the *GSTP1* gene, causing an isoleucine to valine substitution at codon 105, has been reported (Grothey *et al.* 2005; Ruzzo *et al.* 2007), although the impact of the risk allele has been questioned (Gamelin *et al.* 2007; Inada *et al.* 2010). Specific haplotypes of the *AGXT* gene have also been linked with predisposition to both acute and chronic PNAO (Gamelin *et al.* 2007; Table 1.6).

Peripheral neuropathy grade	National Cancer Institute description of peripheral sensory neuropathy
1	Mild paresthesia that does not interfere with function; Loss of deep tendon reflexes
2	Moderate paresthesia affecting function but not daily activities
3	Moderate paresthesia affecting daily activities
4	Disability; Permanent sensory loss
5	Death

Table 1.5 Grading criteria of symptoms associated with PNAO, according to the National Cancer Institute guidelines.

Drug	Polymorphism (Gene)	Effect on response/toxicity
5-FU	Truncating splice variant IVS14+1G>A and 2846A>T in 60% of patients (*DYPD*)	Reduced activity of the DPYD enzyme is highly correlated with greater severity and more rapid onset of 5-FU toxicity (Morel *et al.* 2006)
5-FU	G>A point mutation in 5'-splice recognition site of intron 14 leads to removal of entire exon 14 (*DYPD*)	Greater severity of toxicity is associated with a splice site mutation resulting in a 165bp deletion (Wei *et al.* 1996); up to 25% of patients carrying at least one copy of this allele display Grade >3 toxicity, despite a low prevalence of the variant (MAF <1% in Caucasian population; Raida *et al.* 2001)
5-FU	Asp949Val in exon 22 and intronic Ser534Asn associated with reduced enzyme activity (*DYPD*)	Greater severity of toxicity associated with a rare nonsynonymous variant at position 949, causing a substitution of a valine for an aspartic acid residue (similar toxicity as that seen with the 165bp deletion; Morel *et al.* 2006; Seck *et al.* 2005)
5-FU	Ala222Val and Glu429Ala (*MTHFR*)	Two common polymorphisms associated with improved response to 5-FU treatment in patients carrying with aCRC (Toffoli and De Mattia, 2008)
Irinotecan	TA indel in promoter region and two exon 1 variants; 211 G>A and 686C>A (*UGT1A1*)	Strong correlation between presence of variants and increased severity of irinotecan-induced neutropaenia; most severe (Grade 4) toxicity associated with the TA indel genotype (Innocenti *et al.* 2004)
Oxaliplatin	Asp118Asn (C>T silent polymorphism) (*ERCC1*)	Correlated with an increased rate of PNAO onset in a Japanese population (Oguri *et al.* 2013)
Oxaliplatin	Asp118Asn (C>T silent polymorphism) (*ERCC1*)	Associated with clinical outcome following oxaliplatin treatment; individuals homozygous for the C allele respond well (Park *et al.* 2003), whilst individuals carrying a heterozygous T allele exhibit increased mRNA levels and drug resistance (Ruzzo *et al.* 2007)
Oxaliplatin	154C>T substitution on exon 1 and 1142A>G substitution on exon 10 (*AGXT*)	Minor allele haplotype associated with increased predisposition towards oxaliplatin-induced toxicity (acute and chronic forms) (Gamelin *et al.* 2007)
Oxaliplatin	Ile105Val (*GSTP1*)	Contributes to increased risk of chronic PNAO (Grothey *et al.* 2005; Ruzzo *et al.* 2007)

Drug	Polymorphism (Gene)	Effect on response/toxicity
Oxaliplatin	Leu1092Pro (*SCN10A*) and an intronic variant, rs2302237 (*SNC4A*)	Associated with increased risk of acute PNAO, whilst also increasing severity of toxicity (Argyriou *et al.* 2013)
Cetuximab and panitumumab	Activating mutations in codons 12 and 13; rarer activating mutations in codons 61 and 146 (*KRAS*)	Significantly reduced response rates to cetuximab treatment; from 13% to 1.2% (Karapetis *et al.* 2008); Variants in codons 61 and 146 associated with a similar lack of clinical response to treatment (Loupakis *et al.* 2009)
Cetuximab and panitumumab	Mutations in exon 2 (*KRAS*); population wild-type for *KRAS* carried mutations in exon 20 (*PI3KCA*) and multiple nonsynonymous variants in *BRAF* and *NRAS*	Variants in *KRAS* associated with resistance to cetuximab (Karapetis *et al.* 2008; Lievre *et al.* 2006; Di Fiore *et al.* 2007); variants in *BRAF*, *NRAS* and *PIK3CA* exon 20 associated with reduction in response rate in a *KRAS*- wild type population (De Roock *et al.* 2010)
Cetuximab and panitumumab	V600E (*BRAF*)	Associated with a reduction in drug efficacy; present in approx. 10% of all aCRC cases (Di Nicolantonio *et al.* 2008; Benvenuti *et al.* 2007)
Cetuximab and panitumumab	Variants in codon 61 (*NRAS*)	Strongly correlated with reduced response rate (reduced by over 30% in carriers; Peeters *et al.* 2013)
Cetuximab and panitumumab	Oncogenic mutations and loss of expression (*PI3KCA* and *PTEN*, respectively)	Strongly associated with EGFR inhibitor treatment failure (Frattini *et al.* 2007; Perrone *et al.* 2009; Loupakis *et al.* 2009); variants in both genes reported to co-occur with other variants in EGFR-related genes (Sartore-Bianchi *et al.* 2009)

Table 1.6 Summary of pharmacogenetic factors that affect response and toxicity in the treatment of CRC.

1.3.5.4 Cetuximab

The most common dose-limiting toxicity associated with cetuximab is the development of an acneiform skin rash on the face and upper body resulting from keratinocyte alterations, in addition to hair follicle proliferation and maturation, which occurs in approximately 50% of patients receiving treatment (Agero *et al.* 2006). Clinical manifestation of the skin rash is similar to acne vulgaris although histological examination has revealed it more closely resembles infectious folliculitis (Lenz, 2006). The skin rash peaks in severity during the first two weeks of therapy and usually stabilises in subsequent weeks (Busam *et al.* 2001).The mechanism of rash formation remains poorly elucidated, although interference of follicular epidermal growth signalling is considered crucial (Lynch *et al.* 2007). The EGFR, located within the epidermis, stimulates epidermal growth, inhibits differentiation, protects against UV damage and inflammation, and accelerates wound healing (Jost *et al.* 2000). EGFR is ubiquitously expressed in epidermal keratinocytes, sebaceous and eccrine glands, and hair follicle epithelial cells (Nanney *et al.* 1990; Fox, 2006). Cetuximab-induced inhibition of EGFR alters keratinocyte differentiation, proliferation, migration and attachment, which most likely contributes to acneiform skin rash presentation (Peus *et al.* 1997; Woodworth *et al.* 2005).

There is evidence to suggest that skin rash presentation correlates with cetuximab efficacy, indicated by increased survival in patients across multiple tumour types (Bruno *et al.* 2003; Saltz *et al.* 2003, Orditura *et al.* 2009). Given that the skin rash acts as a surrogate marker for efficacy, clinicians manage toxicity in patients using topical antibiotic and steroid treatments to prevent treatment discontinuation (Pinto *et al.* 2011).

1.3.5.5 Panitumumab

As an EGFR inhibitor, panitumumab causes similar toxicities to those seen with cetuximab, with dermatological side effects accounting for 90% of reported toxicity. Additionally, patients commonly experience fatigue, diarrhoea, nausea, neutropaenia and hypomagnesaemia (Van Cutsem *et al.* 2007).

1.3.5.6 Bevacizumab

The most common toxicity associated with bevacizumab therapy is severe hypertension; 23% of patients experience hypertension, with 8% of these considered severe cases (Scartozzi *et al.* 2009). Inhibition of vascular endothelial growth factor (VEGF) is believed to result in hypertension as a result of two mechanisms; firstly, a reduction in vasodilator production lowers normal physiological levels, causing vasoconstriction (Isenberg *et al.* 2009). Secondly, reduced nitric oxide levels cause a reduction in sodium excretion, leading to hypertension due to water retention in the bloodstream (Beevers *et al.* 2001). Additional toxicities include proteinuria, bleeding, increased risk of arterial and venous embolisms, and, in the most severe cases, poor wound healing and perforation of the GI wall (Scartozzi *et al.* 2009).

Hypertension can be treated with administration of anti-hypertensive drugs, including angiotensin-converting enzyme (ACE) inhibitors, calcium channel blockers or beta-blockers (Messerli, 1999). As with most drugs, a proportion of patients remain unresponsive to drugs to alleviate side effects, in which case dose modification can reduce severity of side effects (Scartozzi *et al.* 2009).

1.3.6 Pharmacogenetics of CRC treatments

Multiple genetic variants have been associated with differing degrees of response to and side effects from drugs used in the treatment of CRC, as shown in Table 1.6.

1.3.6.1 Fluoropyrimidines

Several genetic variants have been associated with varying response to the fluoropyrimidine class of drugs in CRC treatment. The primary mechanism of 5-FU metabolism is via dihydropyrimidine dehydrogenase (DPYD) activity; approximately 80% of administered drug is metabolised by the enzyme (Amstutz *et al.* 2011). At least 15 polymorphisms in the *DPYD* gene have been reported to alter enzyme activity, typically with reduced activity associated with greater severity and more rapid onset of 5-FU-induced toxicity (Schwab *et al.* 2008; Morel *et al.* 2006; Ciccolini *et al.* 2010; Raida *et al.* 2001; Table 1.6).

Additional polymorphisms have been shown to influence response to the fluoropyrimidine drugs, including common polymorphisms in the gene encoding the methylenetetrahydrofolate reductase (MTHFR) enzyme. MTHFR is critical for reduced folate production, the presence of which is essential for the drug's action (Etienne-Grimaldi *et al.* 2010).

Variants in TS have also been correlated with altered protein expression, with increased expression typically inversely linked to clinical outcome. One particular variant consists of a 28 base pair repeat sequence in the 5' untranslated region (5'UTR) of *TS*, with markedly increased expression of the protein when associated with three 5'UTR repeats as opposed to two (Lecomte *et al.* 2004). Alternatively, a six base pair deletion in the 3'UTR has been associated with significantly reduced mRNA stability, which subsequently affects protein expression (Mandola *et al.* 2004).

1.3.6.2 Irinotecan

As mentioned in section 1.1.1.2, polymorphisms in *UGT1A1* have been associated with increased severity of irinotecan-induced toxicity, owing to the enzyme's role in the deactivation of irinotecan's active metabolite, SN38 (Iyer *et al.* 1999; Ando *et al.* 1998). Additionally, patients exhibiting elevated levels of bilirubin, another substrate of UGT1A1, or with an inherited deficiency in the enzyme such as Gilbert's syndrome (OMIM #143500) have also been reported to have an increased risk of irinotecan-induced toxicity (Sai *et al.* 2004; Lankisch *et al.* 2008).

1.3.6.3 Oxaliplatin

The efficacy of oxaliplatin treatment for individuals with aCRC is typically affected by variants in genes involved in the pharmacokinetic and cellular response pathways. Modified expression of *ERCC1*, a gene essential for nucleotide excision repair (NER) of platinum adducts, has been reported to affect response to the drug; typically increased expression is associated with resistance to oxaliplatin (Kwon *et al.* 2007). Other DNA repair genes also have shown an association with clinical outcome of oxaliplatin treatment, including

the base excision repair (BER) gene *XRCC1*. Patients carrying an Arg399Gln variant in this gene typically respond better to treatment (Lv *et al.* 2013).

Several variants have also been reported to affect oxaliplatin-induced toxicity. Certain haplotypes of alanine glycoxylate transferase (AGXT), required for oxalate metabolism, have been reported to increase predisposition towards the acute and chronic forms of peripheral neuropathy (Gamelin *et al.* 2007). Interestingly, the synonymous variant in *ERCC1* encoding Asn118 was correlated with an increased rate of onset of peripheral neuropathy in a Japanese population following oxaliplatin treatment (Inada *et al.* 2010; Oguri *et al.* 2013). Oguri *et al*. also published an association between rs10486003 in tachykinin (*TAC1*) and rs17140129 in phenylalanyl-tRNA synthetase 2 (*FARS2*) with the rate of onset of peripheral neuropathy and its severity, respectively.

1.3.6.4 Cetuximab and panitumumab

Variants in a downstream effector of the EGFR pathway, Kirsten rat sarcoma viral oncogene homolog (KRAS), have been reported to influence resistance to EGFR inhibitors. A lack of response has been reported in patients carrying *KRAS* mutations undergoing either monotherapy or combination therapies for both EGFR inhibitors (De Roock *et al.* 2008; Amado *et al.* 2008; Lievre *et al.* 2008; Bokemeyer *et al.* 2009). Tumours with inactivating mutations in KRAS at codons 12 and 13 demonstrated significantly reduced response to cetuximab therapy, down from 13% to 1.2% (Karapetis *et al.* 2008). Despite these findings, up to 60% of *KRAS*-wild type tumours display a lack of response, which prompted investigation of other components of the EGFR pathway for a role in lack of response to EGFR inhibitors (Linardou *et al.* 2008). In addition to *KRAS* mutations, variants in activating v-raf murine sarcoma viral oncogene homolog B1 (*BRAF*), neuroblastoma ras viral oncogene homolog (*NRAS*) and phosphatidylinositol-4,5-biphosphate 3-kinase (*PI3KCA*), as well as loss of expression of the PI3K pathway inhibitor and tumour suppressor, *PTEN*, are implicated in EGFR inhibitor treatment failure (Di Nicolantonio *et al.* 2008; De Roock *et al.* 2010; Laurent-Puig *et al.* 2009; Frattini *et al.* 2007; Perrone *et al.* 2009; Table 1.6).

1.4 Metabolism of oxaliplatin

1.4.1 Pharmacokinetics

Oxaliplatin is administered intravenously; first line treatment requires a dose of 85mg/m^2 once every two weeks (De Gramont *et al.* 2000) and for second line treatment a dose of 130mg/m^2 is given every three weeks when combined with fluoropyrimidines over 2-6 hours (Vyzula *et al.* 2006).

1.4.2 Absorption and distribution

Immediately after absorption, the initial oxaliplatin molecule is hydrolysed by displacement of the oxalate group, forming reactive intermediates that readily bind to amino groups in DNA, RNA and proteins (Desoize and Madoulet, 2002). Additional interactions include binding of oxalate intermediates to sulphur atoms on cysteine and methionine residues (Raymond *et al.* 2002). After 2 hours of drug infusion, at least 70% of these molecules bind reversibly to plasma proteins, facilitating distribution of the drug (Culy *et al.* 2000). From the plasma the DACH ligand of oxaliplatin is easily distributed throughout the body due to its highly lipophilic nature, a process that is further aided by its ability to bind to proteins, DNA and other large molecules (Graham *et al.* 2000).

1.4.3 Metabolism and excretion

Oxalate is the major metabolite of oxaliplatin produced following non-enzymatic displacement of H_2O and chlorine ions. Oxalate is metabolised in a similar manner to that of glycoxylate, a by-product of amino acid metabolism, which involves detoxification and metabolism by AGXT and glycoxylate reductase-hydroxypyruvate reductase (GRHPR), respectively (Holmes and Assimos, 1998). Since oxalate is a chelator of calcium ions, there is a high possibility that oxalate production may cause the acute neuropathy seen with oxaliplatin therapy (Grolleau *et al.* 2001).

Renal elimination accounts for approximately 50% of unbound oxaliplatin molecules. This has been reported to occur at a rate of 121ml/min (Kern *et al.* 1999), whilst the remaining oxaliplatin bound to erythrocytes is excreted at a rate in concordance with the cells' half-life (Levi *et al.* 2000).

1.4.4 Cellular processing

1.4.4.1 Cellular influx

Although the primary uptake mechanism of oxaliplatin is passive diffusion, multiple drug transporters have also been implicated. These include the copper transporter protein 1 (CTR1) and both organic cation transporters (OCT1 and OCT2), which increase cellular levels of oxaliplatin (Holzer *et al.* 2006; Zhang *et al.* 2006). Studies in mice have reported that *OCT2* knockdown is associated with an elevated rate of oxaliplatin-induced neuropathy (Sprowl *et al.* 2013).

1.4.4.2 Cellular localisation and trafficking

Additional members of the copper transport system have been reported to control cellular localisation of platinum compounds by binding to platinum molecules and distributing these throughout the cell (Chen and Kuo, 2010). These include cytochrome C oxidase 17 (C0X17), copper chaperone for superoxide dismutase (CCS) and human antioxidant homologue 1 (HAH1), which are involved in trafficking of platinum compounds to the mitochondria for metabolism and to the Golgi apparatus, in preparation for efflux from the cell (Gupta and Lutsenko, 2010; Nejdl *et al.* 2015; Figure 1.5).

1.4.4.3 Detoxification

Detoxification of platinum compounds directly influences the amount of active circulating platinum that is free to interact with DNA. Biotransformation occurs when free platinum drug molecules form a complex with reducing agents rich in thiol groups, including L-methionine, L-glutathione and L-cysteine, which yields an unreactive species (Luo *et al.* 1999; Levi *et al.* 2000). Conjugation following detoxification causes efflux of platinum from cells, therefore protecting DNA from damage (Mackay *et al.* 2007). Glutathione conjugation is catalysed by glutathione S-transferase (GST), a metabolic enzyme of which multiple subclasses exist. However, only a few have been reported to mediate platinum detoxification, including GSTP1, GST-τ (GSTT1) and GST-μ (GSTM1) (Peklak-Scott *et al.* 2008; Lecomte *et al.* 2006).

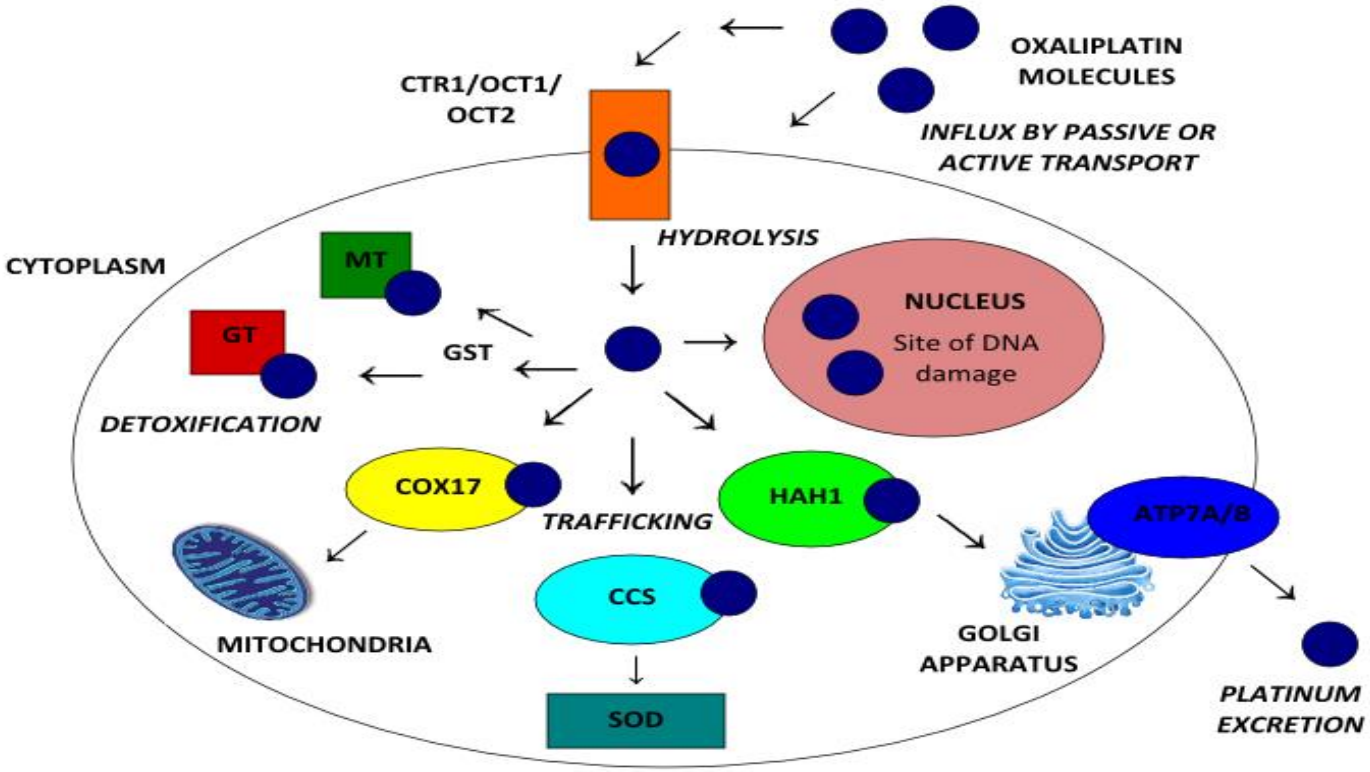


Figure 1.5 Cellular processing of oxaliplatin. Oxaliplatin enters the cell via passive or active transport using the copper transporter protein 1 (CTR1) and both organic cation transporters (OCT1 and OCT2). Cytochrome C oxidase 17 (COX17) and copper chaperone for superoxide dismutase (CCS) facilitate trafficking of platinum compounds to the mitochondria and cytoplasmic superoxide dismutase (SOD), respectively (Chen and Kuo, 2010). The human antioxidant homologue 1 (HAH1) escorts platinum compounds to the copper transporting P-type adenosine triphosphatase 7A and 7B (ATP7A and ATP7B) located in the Golgi apparatus (Gupta and Lutsenko, 2010). Trafficking of these proteins to the plasma membrane is associated with a role in platinum efflux from the cell (Kim *et al.* 2009). GST = Glutathione S-transferase; GT = Glutathione; MT = Metallothionein.

Metallothioneins (MT), low molecular weight proteins rich in cysteine resides, also mediate platinum detoxification; MTs are thought to be important in the control of exposure to heavy metals and copper (Mejare and Bulow, 2001). Tumours displaying high expression of MT1A and MT2A have been reported to show reduced response to platinum agents (Marsh *et al.* 2009).

1.4.4.4 Cellular Efflux

A key mechanism involved in cellular efflux of platinum agents is via the ATP7A and ATP7B copper transport proteins, although additional reports have implicated the ATP-binding cassette (ABC) family of proteins in platinum efflux, in particular subfamilies ABCB1, ABCC1, ABCC2, ABCC3, ABCC5 and ABCG2 (Sprowl *et al.* 2013). Overexpression of several of these is reportedly associated with response to platinum therapy (Han *et al.* 2011; Tian *et al.* 2012; Marsh *et al.* 2007; Kim *et al.* 2009).

1.4.5 Pharmacodynamics

The major neoplastic property of platinum drugs is the ability to form platinum-DNA adducts (Figure 1.6). Crosslink formation results in stalled DNA synthesis (Raymond *et al.* 1998), which subsequently impairs replication and transcription, resulting in apoptosis (Faivre et al. 2003).

Oxaliplatin and cisplatin demonstrate similar sequence-specific and regional localisation of DNA damage (Woynarowski *et al.* 1998), despite differing in other aspects. It is believed that oxaliplatin forms fewer lesions than cisplatin at an equimolar concentration, despite a higher efficiency of DNA chain elongation inhibition. This is partly due to the ability of the bulky DACH group to modify the helical structure of DNA (Woynarowski *et al.* 2000).

1.5 DNA repair

The cell's ability to recognise and repair DNA damage is crucial for maintenance of genetic stability and ultimately survival. An individual cell is subjected to up to one million genotoxic events per day; some occur as a by-product of normal cell metabolism whilst others are induced by radiation and environmental stressors (Jackson and Bartek, 2009; Ciccia and Elledge, 2010).

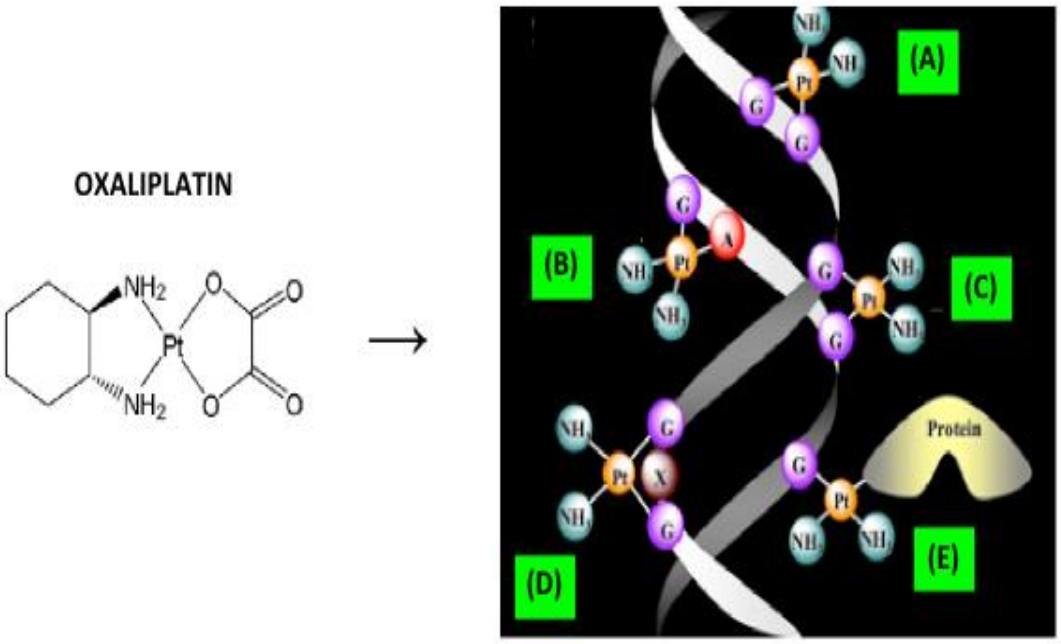


Figure 1.6 Oxaliplatin adduct formation. Following oxaliplatin treatment, initially monoadducts form, although these are considered non-toxic (Alcindor and Beauger, 2011). The drug's cytotoxic action is only apparent following formation of biadducts; of these lesions, approximately 60-65% constitute intrastrand crosslinks between adjacent guanine residues (A) and 20-25% constitute crosslinks between guanine and adenine residues (B). The remaining lesions include interstrand crosslinks (ICLs) between consecutive (C) and more distant guanine residues (D) (Woynarowski et al. 2000) and DNA-protein interactions (E) (Raymond et al. 2002). Figure adapted from Boulikas, 2007.

Inability to repair damage can result in chromosomal instability and modifications to coding DNA, resulting in genomic aberrations that can trigger activation of oncogenes and inactivation of tumour suppressor genes, which promote tumour development (Hoeijmakers, 2001).

Currently there are 168 proteins that are implicated in the various DNA repair pathways; each pathway contains numerous steps that are performed by specialised proteins required for repair of different forms of damage, although significant overlap of different pathways has been reported (Wood, 2005; Bernstein *et al.* 2002). An association between mutations in DNA repair genes and cancer-predisposing syndromes has been demonstrated in a number of studies (Muller and Fischel, 2002; Kinzler and Vogelstein, 1996).

1.5.1 NER Pathway

Nucleotide excision repair (NER) is responsible for the removal of bulky adducts that cause distortion of the DNA helix, resulting in impaired replication and transcription. It is primarily involved in the repair of ultraviolet (UV)-induced photoproducts, which consist of 6-4 photoproducts (6-4 PP) and cyclobutane pyrimidine dimers (CPD) (De Laat *et al.* 1999), although it can additionally repair other types of lesions resulting from exposure to a range of chemical and environmental stressors (Fousteri and Mullenders, 2008).

The NER pathway consists of two distinct sub-pathways, which differ in their damage recognition processes. Transcription coupled (TC-NER) repair takes place when RNA polymerase II (RNA pol II) stalls at damaged DNA within a transcription site. Cockayne syndrome group A and B (CSA and CSB) are recruited to stalled RNA pol II and commence the damage recognition process (Svejstrup, 2002). Alternatively, where damage occurs at non-transcribed regions, global genomic (GG-NER) repair occurs via damage recognition by the xeroderma pigmentosum, group C (XPC) – HR23B complex and the DNA damage binding proteins, 1 and 2 (DDB1 and DDB2). The precise proteins involved depend on the extent of distortion of the DNA double helix and the specific lesions induced (Sugasawa *et al.* 2001; Kusumoto *et al.* 2001). CSB and XPC-HR23B are involved in the recruitment of the ten sub-unit basal transcription factor complex (TFIIH) to the damage site; TFIIH consists of two

helicases, XPB and XPD, which mediate DNA strand unwinding in the presence of ATP (Laine *et al.* 2006). This process enables binding of XPA, XPG and replication protein A (RPA). XPA confirms the site is damaged and then acts in combination with RPA to protect the undamaged single stranded DNA and recruits XPF and excision repair cross complementing, group 1 (ERCC1) complex. The 5' ERCC1-XPF complex incises and releases the damaged strand, in conjunction with the 3' incision by XPG (Tsodikov *et al.* 2007). This is followed by recruitment of repair machinery to facilitate repair and ligation using the undamaged strand as a template (Ogi *et al.* 2010; Figure 1.7). Multiple NER genes have been associated with disease, primarily disorders of NER and transcription (Table 1.7).

1.5.2 Disorders of repair and transcription

Xeroderma pigmentosum (XP) is an autosomal recessive disorder characterised by severe sensitivity to UV light and a 1000-fold increased incidence of skin cancers resulting from impaired NER and accumulation of UV-induced lesions (Kraemar *et al.* 1994). There are currently eight complementation groups of XP, which all display similar phenotypes but with varying degrees of severity. Some complementation groups additionally exhibit neurological degenerative symptoms; approximately 20-30% of all XP patients present with neurological symptoms (De Boer and Hoeijmakers, 2000; Anttinen *et al.* 2008).

Seven of the complementation groups of XP (XPA-G) arise due to function impairing mutations in NER pathways genes, whilst the eighth complementation group, XPV, is the result of variants in the replicative bypass polymerase η (POLH). Patients with XPV have fully functioning NER; however the cells' inability to perform DNA replication past regions with UV lesions gives rise to the traditional XP phenotype (Masutani *et al.* 1999).

Additional rare recessive disorders of repair and transcription include Cockayne syndrome (CS) and trichothiodystrophy (TTD). CS is characterised by cutaneous photosensitivity, despite no increased risk of skin cancer. Two known complementation groups exist, CSA and CSB, which arise due to functional mutations in *ERCC8* and *ERCC6*, respectively (Wood, 1997).

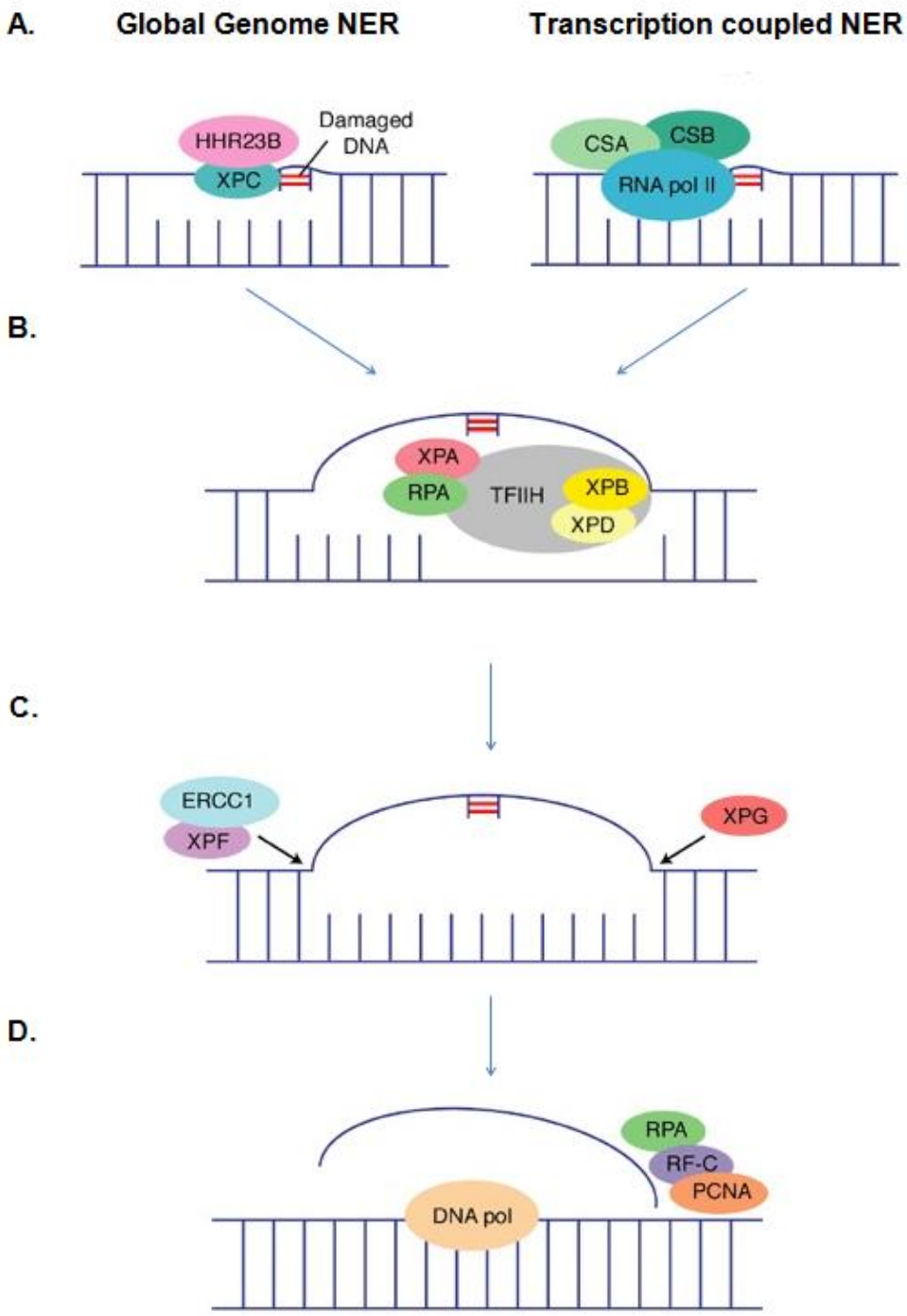


Figure 1.7 The two sub-pathways of the NER pathway. The two branches of NER, showing A) DNA damage recognition, B) DNA unwinding, C) incision of the damaged strand and D) excision of damaged oligonucleotides followed by *de novo* synthesis. NER functions to remove helix distorting lesions. In coding regions of DNA, TC-NER is initiated via RNA pol II stalling, whilst lesions in non-transcribed DNA regions are repaired by GG-NER, whereby lesions are recognised by XPC-HR23B, DDB1 and DDB2. The damaged strand is excised and repair machinery is recruited, including POLD, POLE or polymerase $_{K}$ (POLK; Gillet and Scharer, 2006). Additionally XRCC1-LIG1 or LIG3 may be recruited, depending on the stage of the cell cycle the cell in question is in (Moser *et al.* 2007).

Human Gene (Protein)	Sub-pathway	Function in NER	Disease phenotype
CCNH (Cyclin H)	Both	CDK Activator Kinase (CAK) subunit	None
CDK7 (Cyclin Dependent Kinase (CDK) 7)	Both	CAK subunit	Elevated protein expression in neurons associated with Alzheimer's disease (Zhu *et al.* 2000)
CETN2 (Centrin-2)	GG-NER	Damage recognition, forms complex with XPC	None
DDB1 (DDB1)	GG-NER	Damage recognition; forms complex with DDB2	Involved in pathogenesis of Best's disease (Best's vitelliform macular dystrophy; Stohr *et al.* 1998)
DDB2 (DDB2)	GG-NER	Damage recognition; recruits XPC	Absent DDB2 activity in patients with XPE, a hereditary photosensitive disease with high incidence of skin malignancy (Itoh *et al.* 2001)
ERCC1 (ERCC1)	Both	Incises on 3' side of damage; forms a complex with XPF	Certain polymorphisms associated with increased susceptibility to certain cancers; 19007 C allele (rs11615) contributes to skin cancer development; 17677 A>C and 19007 T>C polymorphisms are low penetrance alleles for cancer susceptibility (Zhang *et al.* 2011)
ERCC2 (XPD)	Both	ATPase and helicase activity; transcription factor II H (TFIIH) subunit	XPD characterised by a 1000-fold increased risk of cancer (Kraemer *et al.* 1994); *ERCC2* variants also cause trichothiodystrophy (TTD), although TTD patients do not exhibit increased cancer risk (Stefanini, 2000); Lys751Gln homozygous variant associated with higher risk of head and neck cancer (Sturgis *et al.* 2000)
ERCC3 (XPB)	Both	ATPase and helicase activity; transcription factor II H (TFIIH) subunit	XPB characterised by relatively mild XP symptoms, also causes an XP/CS phenotype with more severe symptoms and cancer predisposition (Oh *et al.* 2006); Variants in *ERCC3* cause a rarer TTD phenotype, which again is cancer-free (Stefanini, 2000)
ERCC4 (XPF)	Both	Incises on 3' side of damage; forms a complex with ERCC1	XPF characterised by acute sun sensitivity and neurologic symptoms and cancer predisposition (Sijbers *et al.* 1998); Fanconi anaemia caused by Leu238Pro and Arg689Ser (Bogliolo *et al.* 2013); variants in ERCC4 associated with increased cancer predisposition (e.g. 357A>C variant predicts risk and recurrence of bladder cancer; Wang *et al.* 2010b)

Human Gene (Protein)	Sub-pathway	Function in NER	Disease phenotype
ERCC5 (XPG)	Both	Incises on 5' side of damage; stabilises TFIIH	XPG characterised by mild to severe XP symptoms, occasionally coupled with symptoms of Cockayne syndrome (CS; O'Donovan *et al.* 1994); variants increase risk of certain cancers (e.g. prostate cancer; Berhane *et al.* 2012)
ERCC6 (CSB)	TC-NER	Transcription elongation factor; involved in transcription coupling and chromatin remodelling	Cockayne syndrome; certain variants associated with significantly increased predisposition to lung cancer (rs2228526, rs4253160, rs12571445 and rs3793784; Ma *et al.* 2009)
ERCC8 (CSA)	TC-NER	Ubiquitin ligase complex; interacts with CSB and p44 of TFIIH	Cockayne syndrome; predisposition to skin cancers (Laugel, 2013)
LIG1 (DNA Ligase I)	Both	Final ligation	Disruption of LIG1 led to a psoriasis phenotype in mice (Suzuki, 2002)
MNAT1 (MNAT1)	Both	Stabilises CAK complex	None
MMS19 (MMS19)	Both	Interacts with XPD and XPB subunits of TFIIH helicases	None
RAD23A (RAD23A)	GG-NER	Damage recognition; forms complex with XPC	None
RAD23B (RAD23B)	GG-NER	Damage recognition; forms complex with XPC	Ala249Val variant in *RAD23B* associated with significantly increased risk of laryngeal cancer (Abbasi *et al.* 2009)
RPA1 (RPA1)	Both	Subunit of RFA complex	None
RPA2 (RPA2)	Both	Subunit of RFA complex	None
TFIIH (Transcription factor II H)	Both	Forms a complex around lesion prior to incision	Variants in *p44* gene, encoding a subunit of TFIIH, causes large-scale deletions that are implicated in Werdnig-Hoffmann disease (Burglen *et al.* 1997; Another subunit of TFIIH, GTF2H5, causes TTD group A (Giglia-Mari *et al.* 2004)
XAB2 (XAB2)	TC-NER	Damage recognition; interacts with XPA, CSA and CSB	None
XPA (XPA)	Both	Damage recognition	None
XPC (XPC)	GG-NER	Damage recognition	Increased predisposition to skin cancers demonstrated in mice (Sands *et al.* 1995)

Table 1.7 Summary of NER genes with their protein functions and sub-pathways.

Alternatively, TTD consists of three known complementation groups, caused by mutations in genes encoding XPB, XPD and TTD-A, which cause up to a 70% reduction in cellular expression of TFIIH, regardless of zygosity (Botta *et al.* 2002). As a result, TTD is characterised by short stature, immature sexual development and brittle hair (Kraemer *et al.* 2007). Unlike XP, all known complementation groups of CS and TTD display cancer-free phenotypes (Berneburg *et al.* 2000).

In contrast to XP, CS is widely regarded a disorder of transcription, whereas TTD is considered a disorder of both repair and transcription. The similar clinical phenotypes exhibited by patients with these syndromes can be attributed to defects in common genes. This provides evidence for crosstalk between the NER and transcription pathways. For instance, overlap occurs between TTD and XP disorders as a result of shared mutations in XPB and XPD, both of which encode subunits of TFIIH responsible for unwinding the DNA helix prior to lesion removal. However, despite the similarities in the genetics between each disease, TTD patients do not exhibit a cancer-prone phenotype whilst XP sufferers exhibit massively increased predisposition to skin cancers (Nishiwaki *et al.* 2004; Berneburg *et al.* 2000). Additional overlap occurs due to mutations in CSA and CSB, both of which encode proteins involved in transcription coupled (TC)-NER, enabling removal of transcription-blocking lesions from the transcribed strands of actively transcribed genes (Nardo *et al.* 2009). Similarly, mutations in *ERCC5* encoding XPG, the endonuclease required for 3' incision during NER, cause XP and, in rare cases, a phenotype that resembles that of CS. Cells from CS patients have demonstrated photosensitivity and an inability to recover from UV damage-induced inhibition of RNA synthesis (Rockx *et al.* 2000). A summary of these overlapping disorders is given in Table 1.8.

1.5.3 Mechanisms of peripheral neuropathy

1.5.3.1 Inherited disorders of peripheral neuropathy

Inherited peripheral neuropathy disorders are primarily caused by mutations in the peripheral myelin protein 22 (*PMP22*) gene. One of the most common PMP22-related neuropathies is Charcot-Marie-Tooth disease (CMT).

Disease	Genes affected	Phenotype	Defect in repair or transcription
Cockayne syndrome (CS)	*CSA, CSB*	Cancer-free multisystem disorder; photosensitivity; physical and mental retardation; severe early onset neurological degeneration; pigmentary retinopathy; cataracts	Transcription
Trichothiodystrophy (TTD)	*XPB, XPD, TTDA*	Cancer-free multisystem disorder; photosensitivity in TTD-A only; physical and mental retardation; sulphur-deficient brittle hair; ichthyosis; decreased fertility; short stature	Both
Xeroderma pigmentosum (XP)	*XPA, XPB, XPC, XPD, XPE, XPF, XPG, XPV*	Highly cancer-prone (skin cancer); photosensitivity; pigmentation abnormalities; progressive neurological degeneration; premature skin ageing; cataracts	Repair
XP/CS	*XPB, XPD*	Progressive neurological degeneration; impaired intellectual and sexual development	Both
XP/TTD	*XPB, XPD, XPG*	Features of TTD in addition to cancer predisposition seen in XP	Both

Table 1.8 Disorders of repair and transcription with their associated mutations and phenotype.

CMT encompasses a heterogeneous group of neuropathy disorders characterised by sensory loss, distal muscle weakness and atrophy. These are classified according to their genetic cause and neurophysiologic signs; the most common form, CMT type 1A (CMT1A) arises as a result of a 1.5 megabase (Mb) duplication on chromosome 17p11.2, resulting in three copies of the *PMP22* gene (Van Paassen *et al.* 2014). The functional consequence of this is overproduction of PMP22 RNA and protein within Schwann cells, resulting in impaired neuron function. Although the exact mechanism for this is still under investigation, it is known that a defining factor of CMT1A is a motor conduction velocity (MCV) of less than 38 metres per second (m/s), compared with an MCV of up to 120m/s in physiologically normal peripheral neurons (Gasser and Erlanger, 1927; Manzano *et al*. 2008). CMT1A is a demyelinating neuropathy, meaning that the disease is due to degradation of either the protective myelin sheath that covers the long axons of neurons or the Schwann cells that produce and secrete myelin. Alternatively, CMT type 2A (CMT2A) is an axonal peripheral sensorimotor neuropathy, which means that the disease involves direct destruction of peripheral nervous system neurons. In particular, the axons that relay action potentials and transfer protein signals between adjacent neurons are degraded, resulting in a combination of sensory and motor symptoms. Unlike CMT1A, it is caused by autosomal dominant mutations in the mitofusin-2 (*MFN2*) gene, and is defined by a CMV of greater than 38m/s (Harding and Thomas, 1980).

Rarer genetic disorders include leucodystrophy, a disorder of the white matter in which myelin may be decreased or absent, unstable and/or have an abnormal structure, resulting in symptoms of peripheral neuropathy. A number of leucodystrophy disorders exist with a number of various genetic causes, such as a deficiency of aryl sulphatase A, which causes metachromatic leucodystrophy (MCL). MCL exhibits symptoms of both peripheral and central nervous system neuropathy, of which the severity differs greatly between patients (Bindu *et al*. 2005).

1.5.3.2 Diabetic peripheral neuropathy

Peripheral neuropathy has also been recognised as a common complication of diabetes mellitus and is associated with debilitating pain in these patients. Typically these complications arise in patients with a higher average hyperglycaemic index; however, aside from this observation, little is known about the aetiology of diabetic peripheral neuropathy. Using animal models, several studies have suggested that diabetic peripheral neuropathy (DPN) can be attributed to demyelination (removal of the protective sheath covering the axons, which communicate electrical action potentials between adjacent neurons) of neurons and axonal atrophy, although the association between these clinical signs and excess blood glucose has not yet been identified (Dyck and Giannini, 1996; Thomas, 1999). Additionally, patients with advanced DPN show evidence of axonal and glial degeneration, which supports the evidence in animal models (Britland *et al*. 1990). This could be due to an accumulation of injuries at various regions of an axon, resulting in impaired axonal protein transport and inefficient transfer of the action potential to its intended neuron (Thomas, 1999). Similarly, axonal injury may impair protein synthesis, in particular synthesis of neurotrophic proteins required to facilitate neuron survival and development. This would therefore contribute to reduced regeneration and recovery from damage (Zochodne, 1996). Further research is required to elucidate the molecular impact of excess blood glucose on the development and progression of DPN and myelin regeneration.

1.5.3.3 Chemotherapy-induced peripheral neuropathy

Peripheral neuropathy is a well-recognised toxicity of chemotherapy regimens using certain drugs, such as oxaliplatin, paclitaxel and bortezomib. Platinum drugs including cisplatin exert their cytotoxic effects via the formation of platinum-DNA lesions. It is believed that excess platinum molecules accumulate within the dorsal root ganglia (DRG), resulting in peripheral neuropathy as the primary associated toxicity (Dzagnidze *et al.* 2007; Ta *et al.* 2006).

Taxane drugs such as paclitaxel and docetexel for the treatment of breast cancer also cause peripheral neuropathy, primarily in the form of minor

or moderate axonal sensory polyneuropathy (Osmani *et al*. 2012). Taxane drugs promote microtubule formation within the cytoplasm of affected cells, which interferes with normal cell physiology and inhibits cell proliferation (Scripture *et al*. 2006). Microtubules perform vital maintenance and developmental functions for neurons; microtubule elongation facilitates the growth of precursor neurite cells during neuron development (Kobayashi and Mundel, 1998). Additionally, microtubules mediate axonal transport of proteins and provide structural support to the neuron as a whole (Osmani *et al.* 2012). *In vitro* studies and studies using animal models involving injection of the drug into the sciatic nerve demonstrated degeneration of the DRG, demyelination and reduced nerve conduction velocities following paclitaxel treatment (Cavaletti *et al*. 1995; Cliffer *et al*. 1998). Therefore, the ability of taxanes to disrupt microtubule assembly suggests that consequent structural and functional changes within neuron cells result in peripheral neuropathy as a result of drug treatment.

Additionally, drugs such as bortezomib used in the treatment of multiple myeloma have been shown to trigger peripheral neuropathy. This is likely due to accumulation of unbound drug in the DRG, coupled with dysregulation of neurotrophic proteins in treated individuals, although the exact mechanism underlying peripheral neuropathy associated with bortezomib has not yet been elucidated (Argyriou *et al*. 2008).

1.5.3.4 DNA repair defects and neurological disease

The nervous system is particularly sensitive to DNA damaging effects, compared with other non-replicating cell types. This is likely due to the brain's ability to metabolise 20% of oxygen consumed by an individual, coupled with a lower capacity to neutralise reactive oxygen species (ROS) compared with other organ systems. Additionally, neurons are known to be highly susceptible to oxidative stress (Barzilai, 2007). The net result is a high oxygen load causing free radical generation, which increases DNA damage, in particular, the formation of single-stranded breaks (SSB), within mature neurons (Saxowsky and Doetsch, 2006; Cleaver, 2005; Ljungman and Lane, 2004). Defects in the repair of SSBs are associated with spinocerebellar ataxia with axonal

neuropathy (SCAN1) and ataxia with oculomotor apraxia 1 (AOA1). These are two neurodegenerative disorders that typically present with similar symptoms, including oculomotor apraxia (absence of control of voluntary eye movement), severe sensorimotor neuropathy, cognitive impairment and cerebellar atrophy in late childhood (Shawan *et al*. 2006; El-Khamisy and Caldecott, 2007). SCAN1 and AOA1 are caused by mutations in the tyrosyl-DNA phosphodiesterase-1 (TDP1) and aprataxin (APTX) genes, respectively. TDP1 repairs 3' DNA ends that have been modified by oxidative damage, whilst APTX possesses AMP-lysine hydrolase activity that is required for removal of 5'-AMP intermediates following unsuccessful DNA ligation (Rass *et al*. 2007; Ahel *et al.* 2006; Seidle *et al*. 2005).

Deficiencies in double-stranded break (DSB) repair processes also give rise to neurological disease; one of the earliest diagnosed DNA repair defects associated with neurodegeneration was ataxia-telangiectasia (AT), which develops during childhood and is characterised by severe sensitivity to ionising radiation and rapid neurodegeneration (Frappart and McKinnon, 2006; McKinnon, 2004; Perlman *et al*. 2003). AT is caused by mutations of the ataxia telangiectasia, mutated (ATM) gene, which encodes a protein kinase involved in the regulation of the DSB repair pathways (Shiloh, 2003; Helleday *et al*. 2007). The major DSB pathways are non-homologous end-joining (NHEJ) and homologous recombination (HR). Although HR is typically error-free, errors arise in NHEJ, which is primarily involved in DSB repair in non-replicating cell types including those of the nervous system (Lieber *et al*. 2003). During NHEJ, both ends of the DNA strand undergoing repair are modified to enable direct ligation by DNA ligase IV (LIG4). Mutations in LIG4 that result in attenuated LIG4 activity are associated with LIG4-syndrome, a disorder that presents with immunodeficiency, microcephaly and severe radio-sensitivity (Buck *et al*. 2006).

Mutations in several components within the NER pathway result in the human disorders XP, TTD and CS (Section 1.5.2). Although these are clinically distinct and have different aetiologies, there is overlap between the signs that patients present with. XP patients from most complementation groups (A-G) present with a range of neurological symptoms, with 30% of all XP patients showing signs of neurodegeneration and global brain atrophy (Kraemar *et al*.

2007; Mimaki *et al*. 1986). Additionally sensorineural hearing loss and progressive peripheral neuropathy develops in the second decade of disease (Anttinen *et al*. 2008; Robbins *et al.* 2002). Similarly, patients with CS show signs of progressive and severe neuropathology, including demyelination and calcification of the basal ganglia, although this is absent from XP (Brooks *et al.* 2008; Itoh *et al*. 1999). TTD demonstrates similar neurological symptoms to CS, including demyelination and motor dysfunction (Brooks *et al.* 2008; Nance and Berry, 1992).

A number of other human disorders are caused by defects in DNA repair genes, including Fanconi anaemia (FA), Bloom syndrome (BLM), Werner syndrome (WS) and Rothmund Thomson syndrome (RTS). FA arises due to disruption of any one of many DNA repair proteins needed for removal of DNA crosslinks, including BRCA1 and its interaction partner, partner and localiser of BRCA1 (PALB1). Currently there are twelve known complementation groups of FA, many of which present with microcephaly and cognitive impairment, in addition to several non-neurological symptoms (Reid *et al*. 2007; Mirchandani and D'Andrea, 2006; D'Andrea and Grompe, 2003). Alternatively, BLM, WS and RTS are all associated with defects in the RecQ like helicase 4 (RECQ4) function (Hunter, 2008; Hickson, 2003; Harrigan and Bohr, 2003; Van Brabant *et al*. 2000). The RECQ4 helicase functions to unwind double-stranded DNA to facilitate repair of lesions on either strand (Sangrithi *et al.* 2005). Despite this essential role in DNA repair processes, the resulting neuropathology of these three syndromes has not been thoroughly characterised and remains poorly understood (McKinnon, 2009).

1.6 Models of functional analysis

Model organisms have played a significant role in the history of research in the field of human genetic disease, both in terms of identifying causal genes and characterising normal and abnormal gene functions (Hedges, 2002). This has been particular important for drug discovery and development for targeting specific protein biomarkers (Rual *et al.* 2005). However, the emergence of high throughput technologies such as GWAS and genome sequencing has revolutionised this field, making discovery of disease genes in humans far

simpler and more accessible. Additionally, the genetic architecture identified via sequencing technologies is often complex, therefore making it challenging to replicate this variation using model organisms other than human cell lines (Aitman *et al.* 2011). This has led to the question of whether functional assays will be applicable to future avenues of human genetic research, owing to the limitations of certain organisms to replicate human cell physiology (Green *et al.* 2011). A number of model organisms exist for investigation of genetic variants and their respective protein function, with different advantages and limitations (Table 1.9).

1.7 Next generation sequencing (NGS)

Advances in next generation sequencing (NGS) in recent years have transformed our understanding of genomics, which can influence approaches to diagnosis and treatment in the clinic. Use of NGS in research has led to the identification of multiple causal alleles in different diseases, which would not have been possible using earlier technology (Schweiger *et al.* 2011; Table 1.10). NGS employs the use of massively parallel sequencing to amplify and sequence the genome accurately and at a reasonable cost; sequencing of the first human genome on an NGS platform was considerably cheaper than preceding platforms (Sboner *et al.* 2011).

1.7.1 General workflow

NGS protocols consist of three stages: initial sample preparation, massively parallel sequencing and imaging of sequence data, and data analysis. Several NGS platforms exist, each with their own variations in amplification methods and applications, and different advantages and disadvantages.

Approximately 85% of disease-causing variants are located within protein-encoding regions of genes, yet only 1% of an entire genome makes up the exome (Botstein and Risch, 2003). This, coupled with the considerably reduced cost of whole exome sequencing (WES) in comparison to whole genome sequencing (WGS), makes WES a promising alternative for identifying

Model (Species)	Analysis	Advantages	Disadvantages	Protein function assays
Mammalian cell lines	*In vitro*	Representative of a multicellular organism; well-studied; efficient recognition of signals for synthesis and secretion of proteins; relatively easy to genetically manipulate (although variation between cell lines/species)	Expensive to culture over long periods; considerable variation in transfection rates; diploid genome which can it difficult to study effects of heterozygous mutations; some differences in pathways depending on species the cell line was isolated from	Transfection of plasmids containing gene of interest with fluorescent tag, immunofluorescence, flow cytometry, ELISA, colorimetric assays, etc.
Fission yeast (*Schizo-saccharomyces pombe; S. pombe*)	*In vitro*	Easy to genetically manipulate; well-annotated genome; haploid organism which makes it useful for studying effects of recessive mutations; excises mammalian introns (unlike *S. cerevisiae*)	Not a mammal; not representative of a multicellular organism; some differences in pathways compared with humans; additional pathway for repair of UV light	Similar to mammalian cell line assays
Budding yeast (*Saccharomyces cerevisiae*; *S. cerevisiae*)	*In vitro*	Easy to genetically manipulate; well-annotated genome; haploid organism which makes it useful for studying effects of recessive mutations;	Not a mammal; not representative of a multicellular organism; some differences in pathways compared with humans	Similar to mammalian cell line assays
Bacteria (*Escherichia coli*; *E. coli*)	*In vitro*	Easy to genetically manipulate; cheap; well-annotated genome	Prokaryote; large differences in pathways compared with humans; not representative of a multicellular organism	Transformation; transfection of plasmids with gene of interest; microarrays etc.
Mice (*Mus muscularis*)	*In vivo*	Mammal; well-annotated genome; easy to genetically manipulate; high proportion of genome (>80%) is homologous with human genome	Some differences in pathways compared with humans	Similar to mammalian cell line assays; tissues isolated from animal after killing and processed for assay
Fruit fly (*Drosophila melanogaster*)	*In vivo*	Well-annotated genome; easy and cheap for laboratory use; representative of a multicellular organism	Not a mammal; some differences in pathways compared with humans	Primarily behavioural assays; high throughput alternatives available
Systems biology model	*In silico*	High-throughput nature; utilises DNA/RNA isolated from humans, allowing easy comparison of ‘control’ and ‘disease’ phenotypes; rapid	Expensive to run depending on assay and platform used; expensive start-up costs for NGS/genome-wide sequencing	qPCR; genome-wide sequencing; NGS; ChIP-chip; ChIP-seq, etc.

Table 1.9 Summary of models used to assess protein or cell function after genetic manipulation. Model organisms used for assessment of cell function are listed with their respective advantages, limitations and potential assays.

Discovery using NGS technology	Reference
First genome sequenced by NGS	Wheeler *et al.* 2008
First cancer genome sequenced using WGS in a patient with acute myeloid leukaemia	Ley *et al.* 2008
Variants in *IDH1* associated with glioblastoma multiforme (GBM) in younger patients; secondary somatic variants in the gene associated with increased overall survival	Parsons *et al.* 2008
First twelve human exomes sequenced using targeted capture sequencing technology; demonstrated a use for WES in identifying Mendelian disorders by investigating four individuals with Freeman-Sheldon syndrome (OMIM #193700)	Ng *et al.* 2009
WES used to identify somatic mutations in the intracellular kinase domain of bone morphogenetic protein receptor 1A (*BMPR1A*) gene associated with microsatellite-stable and -instable colon cancers	Timmermann *et al.* 2010
Inherited mutations in 21 genes including *BRCA1* and *BRCA2* associated with predisposition to breast and ovarian cancers; associations confirmed via WGS	Walsh *et al.* 2010
Variants in *POLD1* and *POLE* (OMIM #615083, #612591, respectively) associated with increased predisposition to numerous CRA and CRCs, identified via WES and linkage data	Palles *et al.* 2013
ERCC4 identified as a candidate gene for Fanconi anaemia (FA; OMIM #615272) in a single patient.	Bogliolo *et al.* 2013
NGS technology used to identify pathways and driver mutations implicated in oesophageal adenocarcinoma development	Dulak *et al.* 2013
Targeted NGS technology used to identify somatic mutations in *BRAF*, *RAS*, *PIK3CA*, *PTEN*, *GNAS*, *CTNNB1* and *RET* associated with papillary and anaplastic thyroid carcinomas	Nikiforova *et al.* 2013
NGS technology used to screen for mutational 'hotspots' in 46 cancer-related genes; clinical validation performed to demonstrate the ease by which these genes could be routinely tested for at the point of diagnosis	Singh *et al.* 2013

Table 1.10 Summary of discoveries using NGS technology in recent years.

variants associated with a specific phenotype. During sample preparation for WES, an additional ‘target capture’ step is carried out to ensure only protein coding regions of DNA are selected. DNA is sheered to produce small fragments, followed by adapter ligation to fragment flanking regions and hybridisation assays to isolate coding sequences of DNA (Pruitt *et al.* 2009). Enrichment of DNA fragments is carried out using either microarray-based or solution-based assays (Albert *et al.* 2007; Porreca *et al.* 2007).

Massively parallel sequencing involves coupling of DNA synthesis and detection stages, followed by simultaneous sequencing of multiple reactions; for most NGS platforms the most common cause of sequencing errors and short reads is desynchronisation of reads during the sequencing and detection cycle (McNally *et al.* 2012). Following generation of sequencing reads, quality control is performed to remove errors that occur during sequencing. Reads are then aligned with a reference sequence using one of several available mapping algorithms; selection of the most appropriate mapping tool will depend on the sequencing platform used and the downstream applications required (Bao *et al.* 2011; Ku *et al.* 2012). Annotation of variants can be aided by multiple available tools (McKenna *et al.* 2010; Wang *et al.* 2010; Yandell *et al.* 2011).

1.8 Strategies for gene discovery

NGS yields vast amounts of genomic data that is hugely challenging to sift through to identify disease-causing variants; approximately 20-50,000 variants are identified in a single human exome (Gilissen *et al.* 2012). This number increases significantly when variation in the whole genome is included. Therefore, strategies are required to aid the identification of disease-causing alleles (Cooper and Shendure, 2011).

1.8.1 Candidate gene and pathway analysis

The vast majority of known Mendelian disease-causing mutations are located in protein coding regions; however, approximately 90% of coding variants are known polymorphisms (Robinson *et al.* 2011), suggesting that they are unlikely to contribute to disease. Until recent advances in NGS technology,

the traditional method of disease gene identification involved candidate pathway or gene analyses; this involved selection of genes associated with other diseases or where the predicted protein function was relevant to the disease physiology. Alternatively, a positional mapping approach such as karyotyping, linkage analysis or homozygosity mapping may have highlighted genes in specific genomic regions (Botstein and Risch, 2003; Kerem *et al.* 1989; Lander *et al.* 1987). Several limitations of the candidate approach exist, including difficulties in using genetic mapping technologies to predict whether a disease is caused by structural genetic variation or a single SNP. Additionally, these mapping approaches rarely reduce the number of candidate genes sufficiently to allow validation by Sanger sequencing, particularly when the disease locus is large (Tabor *et al.* 2002). Candidate approaches also cannot be used for rare genetic conditions caused by sporadic mutations, since neither a genetic mapping nor family-based data approach can be used (Gilissen *et al.* 2012).

1.8.2 'No prior hypothesis' approach

Manipulation of NGS data using an approach with no prior hypothesis allows for selection of novel or rare variants that most likely would not be detected via a candidate gene or pathway approach. According to the 'common disease, rare variant' hypothesis (Section 1.2.2.3), rare variants could dramatically affect overall risk (Maniolo *et al.* 2009). However, rare variants are rarely included on large-scale genotyping arrays used for GWAS. Additionally, the low frequency of such variants means that often GWAS are not powerful enough to detect linkage with this variation (Wray *et al.* 2011), making NGS a more useful tool for rare causal variant discovery. Novelty and rarity status can be assessed using online databases, including dbSNP (NCBI Resource Coordinators, 2013), the 1000 Genome Project (1000 Genomes Project Consortium *et al.* 2010) and Ensembl (Flicek *et al.* 2013). Filtering for novelty status enables reduction of data output whilst maintaining sufficient power to detect causal variants (Mwenifumbo and Marra, 2013).

An alternative filtering strategy is to investigate the likelihood of a variant's effect on protein function, since it is often assumed that synonymous variants are unlikely to influence pathogenesis (Gilissen *et al.* 2012). Typically

missense variants including SNPs and indels are considered for analysis. Indels are the second most common form of genomic variant and the most common form of structural variant (Mullaney *et al.* 2010), accounting for approximately 1.6 million collective indel polymorphisms within the human population (Mills *et al.* 2006). The presence of indels can cause disease when they adversely affect either amino acid coding sequences or regulatory functions of non-coding sequences (Barcena *et al.* 2014; Schutte *et al.* 2003; Zhang *et al.* 2014).

Additional filtering strategies may be applied, depending on the nature of the disease in question. For instance, where a monogenic inherited disorder is being investigated, a linkage strategy may be employed whereby DNA from multiple affected family members is sequenced to identify common variation (Genin *et al.* 2008). Alternatively, in cases where a single patient is available without DNA from parents, and where the disease is suspected to be recessive in nature, a 'double-hit' strategy may be used. This involves exome sequencing of the patient's DNA to identify homozygous or compound heterozygous variants, as private variants of this nature exist in relatively low abundance in unaffected individuals (Oetting, 2012). Investigation of highly heterogeneous disease is more complex, with a low likelihood of identifying mutations in the same gene in two affected individuals. For *de novo* variant identification, most commonly a family-based exome sequencing approach is required, whereby *de novo* candidate mutations can be targeted by filtering out all inherited variants present in the parents' exomes. This results in a small yield of potentially pathogenic variants, since the average exome contains only 0-4 *de novo* variants (Vissers *et al.* 2010; Roach *et al.* 2010; Durbin *et al.* 2010). Multiple additional filtering strategies exist for sifting through vast amounts of NGS data with different criteria depending on the nature of the disease under investigation (Gilissen *et al.* 2012).

1.9 Background to this project

1.9.1 MRC COIN trial

The Medical Research Council (MRC) COIN trial was a randomised trial investigating patient outcome with first-line oxaliplatin and fluoropyrimidine chemotherapy in 2,445 patients with advanced CRC. Participants were

randomly assigned to continuous oxaliplatin and fluoropyrimidine chemotherapy (Arm A), the same combination with the addition of cetuximab (Arm B), or intermittent chemotherapy (Arm C) (Maughan *et al*. 2011). The aim of the trial was to investigate response to therapy, survival and toxicity outcomes in patients with aCRC.

The trial revealed excess toxicity in the oxaliplatin-, CPB-, cetuximab-treated patients compared with continuous chemotherapy alone. This led to the conclusion that a CPB dose adjustment would be required to maintain safe toxicity levels during use of this regimen (Adams *et al.* 2009). Comparison of intermittent and continuous oxaliplatin-based chemotherapy plus cetuximab revealed comparable survival statistics with improved toxicity in patients undergoing intermittent therapy, suggesting that intermittent cetuximab-based chemotherapy could be promising for palliative care (Wasan *et al.* 2014).

With regard to survival, the trial was unable to confirm whether addition of cetuximab to oxaliplatin-based chemotherapy was beneficial in terms of PFS or overall survival. Interestingly, response to chemotherapy was strongly correlated with somatic mutant *KRAS* or *BRAF* status, irrespective of treatment. Cetuximab increased response rate from 57% to 64%; however, this was coupled with increased severity of GI and skin-related toxicity, suggesting that addition of cetuximab for treatment of aCRC could not be recommended (Maughan *et al*. 2011).

1.9.2 Patient selection

The sequencing studies and sequencing data analysis outlined in sections 1.9.2 – 1.9.6 were organised and performed by Dr Hannah West. Where appropriate, data from experiments that were outsourced or performed by other staff are referenced in the relevant sections. Full details of this research, including outsourced data analyses, are outlined in Dr West's PhD thesis (West, 2013).

PNAO grade 3 or greater was recorded in 23% and 16% of patients undergoing fluorouracil-based and capecitabine-based regimens throughout the duration of the trial, respectively (Maughan *et al.* 2011). From commencement

of treatment, an assessment of PNAO was performed every six weeks by a consultant and clinical nurse using the Common Terminology Criteria for Adverse Effects v3.0 (CTCAE; Wasif Saif and Reardon, 2005). Patients with recorded PNAO grade 3 or greater carried out a Quality of life questionnaire (QLQ C30), which provided supporting evidence of severe PNAO. Following review of COIN toxicity data, a panel of ten patients with severe and dose limiting PNAO, requiring withdrawal from therapy within the first seven weeks, were selected for exome resequencing and analysis.

Exome resequencing, read alignment and variant calling was carried out by Dr James Colley. Library fragments of exomic DNA from the 10 patients with severe PNAO were prepared using the Roche Nimblegen SeqCap EZ Exome Library v2.0 solution-based method. Massively parallel sequencing was carried out using the Illumina Genome Analyser at the University of North Carolina. Fastq files were resolved by a sequence analysis pipeline using the Burrows-Wheeler Alignment (BWA) tool (Li and Durbin, 2009) for sequence alignment. Modules from the Broad Institute's Genome analysis Toolkit (GATK) (McKenna *et al.* 2010) were used to ensure recalibration of quality scores and generation of GC metrics. Additionally, GATK modules were used to refine alignments around potential insertions and deletions (indels), remove duplicate reads, call indel and SNP genotypes and apply quality filters to the genotype calls. SNP calls were annotated using the ANNOVAR analysis package (Wang *et al.* 2010).

1.9.3 Exclusion of known neuropathies

Exclusion of known neuropathies in the ten patients selected for exome resequencing was performed at Bristol Genetics Laboratory, using multiplex ligation-dependent probe amplification (MLPA) of *PMP22*, which encodes a peripheral nervous system protein called peripheral myelin protein 22, and all other genes associated with rare inherited neuropathies. No stop gain or truncating indel mutations were detected in any of these genes in the ten patients with PNAO. Several nonsynonymous variants were detected in *IGHMBP2*, which encodes DNA-binding protein SMUBP-2 helicase, although these were recorded in dbSNP at comparable frequencies, suggesting they are

benign polymorphisms. Therefore, all known neuropathies were excluded from the ten patients selected for exome resequencing (West, 2013).

1.9.4 ERCC4 and Patient 8

A novel stop gain (S613X) mutation in *ERCC4* was identified in one patient with severe PNAO, which required withdrawal from treatment within seven weeks of commencement. This patient was a 79 year old female who was diagnosed with metastatic CRC in March 2006 after an ultrasound scan on her liver. The patient's medical records showed no indication of xeroderma pigmentosum (XP). The second *ERCC4* allele was assayed for biallelic mutation by Sanger sequencing using amplified sequences of the entire open reading frame (ORF) and flanking regions of *ERCC4*, which revealed no other coding variants. This indicated that the patient did not carry a second mutant allele (West, 2013).

1.9.5 Variants identified in ERCC4

Following identification of the novel stop gain in *ERCC4* (patient 8), this gene was sequenced for other protein-coding variants in a panel of patients (n=63) with PNAO; 54 of these patients had severe and dose limiting PNAO, which meant they were withdrawn from treatment within the first twelve weeks, and 9 patients were withdrawn from therapy within seven weeks. DNA from one patient failed sequencing.

Screening of the entire ORF, flanking regions and 5'UTR of *ERCC4* in this panel identified five nonsynonymous variants in patients with PNAO. Pro379Ser was present in 3 patients (MAF=4.69%) and was previously documented in dbSNP (rs1799802); Arg576Thr in a single patient and in dbSNP (rs1800068); Glu875Gly in 4 patients (MAF=6.25%) and in dbSNP (rs1800124); Arg415Gln in 9 patients (MAF=14.1%) and in dbSNP (rs1800067) and His466Gln in a single patient and not in dbSNP (Table 1.11).

All variants were assessed for their impact on protein function in Align-GVGD, which revealed that Pro379Ser, Arg576Thr and Glu875Gly were likely to affect function, whereas Arg415Gln and His466Gln were predicted to have little or no effect on protein function. All non-synonymous variants were

sequenced in patients without PNAO and their incidences and minor allele frequencies were analysed statistically, with respect to their Align-GVGD score. Combined analysis of two variants (Pro379Ser and Glu875Gly) revealed that they were present in a significantly higher proportion of patients with PNAO than those without (p=0.037). This suggested that, while individually rare and predicted to be functionally damaging, they contribute to PNAO only when analysed collectively (Table 1.11). The Arg576Thr private variant was also predicted to interfere with protein function during *in silico* analysis, but could not be incorporated into the combined analysis of functional variants because it was only present in one patient (West, 2013).

1.9.6 Selected variants for functional analysis

Clustal Omega was used to carry out alignment of all mammalian sequences available on NCBI, which identified a high level of conservation of XPF across several species (West, 2013). This revealed conservation in all species (n=34) analysed for Pro379, Arg576, Glu875, Arg415 and Ser613. However, His466 was not well conserved. Since Pro379, Arg576 and Glu875 were considered most likely to alter protein function and were well conserved in mammalian species, we therefore selected these variants for functional analyses *in vitro* (Table 1.11).

	Variant	rs number	Frequency in patients (%)		*P*
			+ PNAO	- PNAO	
Predicted to affect function	**Pro379Ser**	**rs1799802**	**3/63 (4.76%)**	**27/1,763 (1.53%)**	**0.08**
	Arg576Thr	rs1800068	1/63 (1.59%)	4/1,763 (0.22%)	0.16
	Ser613X	Novel	1/63 (1.59%)	-	
	Glu875Gly	**rs1800124**	**4/63 (6.35%)**	**60/1,763 (3.41%)**	**0.28**
	Total (No. private variants)		**7/63 (11.11%)**	**86/1,763 (4.88%)**	**0.03**
Less likely to affect function	Arg415Gln	rs1800067	9/63 (14.1%)	260/1,754 (14.8%)	0.91
	His466Gln	Novel	1/63 (1.59%)	0/1,677 (0%)	0.04

Table 1.11 Incidence and minor allele frequencies (MAF) of nonsynonymous and stop gain variants found in *ERCC4* in germline DNA from patients with and without PNAO, analysed with respect to their likelihood of affecting function. Variants identified in more than one PNAO patient (shown in bold) were analysed using a combined analysis (total). The private variant in Arg576Thr was not included in the combined analysis due to its potential to skew the data, and Ser613X was not included since it was assayed for only in samples from patients with PNAO (West, 2013). Values in the total column represent the number of patients genotyped with respective p values (*P*).

1.10 Aims of this thesis

1. To investigate the functional consequences of three variants in *ERCC4* associated with PNAO to provide evidence of a causal relationship, by studying their effects on:

   (i) Viability, assayed by Trypan blue exclusion (Chapter 3)

   (ii) Localisation of ERCC1-XPF, assayed by immunofluorescence (Chapter 4)

   (iii) Localisation of Rad16 in *Schizosaccharomyces pombe*, assayed by immunofluorescence (Chapter 4)

   (iv) Rate of DNA repair, assayed by ELISA (Chapter 5)

2. To identify candidate genes that may cause PNAO for future sequencing studies (Chapter 6)

Chapter Two – Materials and Methods

2.1 List of suppliers

Materials and equipment were purchased from the following companies:

Abcam (Cambridge UK)

ABgene Ltd (Surrey, UK)

Agilent Technologies (California, USA)

Amersham Life Science (See GE Healthcare)

Anachem Ltd (Bedfordshire, UK)

Applied Biosystems (Cheshire, UK)

BD Biosciences (New Jersey, USA)

Beijing Genomics Institute (BGI; Shenzhen, China)

Bibby Sterilin (see Thermo Fisher Scientific)

Bioquell UK Ltd (Hampshire, UK)

Bioquote (York, UK)

Biorad (Hertfordshire, UK)

Biotium (California, USA)

Cambridge Biosciences (Cambridge, UK)

Core Life Sciences (California, USA)

Corning Incorporated (Flintshire, UK)

DJB Labcare (Buckinghamshire, UK)

Eurogentec (Hampshire, UK)

Fisher Scientific (Leicestershire, UK)

Formedium (Norfolk, UK)

GE Healthcare (Buckinghamshire, UK)

Helena Biosciences (Tyne & Wear, UK)

Illumina (California, USA)

Invitrogen/Life Technologies (Strathclyde, UK)

Jencon (West Sussex, UK)

KBiosciences (Teddington, UK)

Labtech International (Florida, USA)

Leica Microsystems (Wetzlar, Germany)

The Linde Group (Munich, Germany)

Marienfield (Lauda-Konigshofen, Germany)

Measuring and Scientific Equipment (London, UK)

MJ Research (Massachusetts, USA)

New England Biolabs (Hertfordshire, UK)

Public Health England – HPA Culture Collection (Wiltshire, UK)

Qiagen (West Sussex, UK)

R&D Systems (Oxford, UK)

Roche (West Sussex, UK)

Sigma-Aldrich Ltd (Dorset, UK)

Tecan (Reading, UK)

Thermo Fisher Scientific (Massachusetts, USA)

VWR International (Leicestershire, UK)

2.2 Materials

2.2.1 Chemicals

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

2.2.2 Mammalian tissue culture and reagents

All mammalian cell lines were purchased from Public Health England/HPA Culture Collection.

2.2.2.1 Cell culture

RPMI 1640 culture media supplemented with L-glutamine was purchased from VWR International. This was supplemented with foetal bovine serum (FBS) and penicillin/streptomycin purchased from Invitrogen/Life Technologies. Dimethyl sulfoxide (DMSO) for long-term storage was purchased from Sigma Aldrich.

2.2.2.2 Measurement of cell viability

Trypan blue for viability measurement was purchased from VWR International.

2.2.2.3 Antibodies for immunofluorescence

A primary rabbit polyclonal anti-ERCC1 antibody was purchased from Insight Biotechnology and a primary mouse monoclonal anti-XPF (3F2/3) antibody was purchased from Abcam. A secondary goat anti-rabbit Alexa 488 antibody and a secondary donkey anti-mouse Alexa 546 antibody were both purchased from Life Technologies.

An additional primary goat polyclonal XPF (M-16) antibody was purchased from Santa Cruz Biotechnology and an additional secondary donkey anti-goat Alexa 546 antibody was purchased from Life Technologies for the immunofluorescence assay with oxaliplatin treatment.

2.2.2.4 Reagents for immunofluorescence

The following reagents were all purchased from Sigma Aldrich: Paraformaldehyde, Triton X-100, Bovine serum albumin (BSA), DAPI nuclear stain, Mowiol-488 and 1,4-diazabicyclo[2.2.2]octane (DABCO). Phosphate buffered saline (PBS) without magnesium and calcium was purchased from Life Technologies.

2.2.2.5 Extraction of mammalian cell line genomic DNA

The QIAmp DNA mini kit for genomic DNA extraction was purchased from Qiagen.

2.2.2.6 Fluorescence-activated cell sorting (FACS)

FACS was carried out by Catherine Naseriyan in Central Biotechnology Services (CBS). Propidium iodide stain for permeation of non-viable cells was purchased from Sigma Aldrich. All other buffers were provided by CBS.

2.2.2.7 Enzyme-linked immunosorbent assay (ELISA)

The OxiSelect DNA Damage ELISA for detection of cyclobutane pyrimidine dimers (CPDs) was purchased from Cambridge Biosciences. A rat anti-cisplatin modified DNA antibody and a rabbit polyclonal anti-rat HRP-conjugated secondary antibody were both purchased from Abcam.

2.2.3 *Schizosaccharomyces pombe* (*S. pombe*) culture and reagents

All solutions were made up using dH_2O water and sterilised by autoclaving on the liquid cycle at 15Ib/sq.in at 121°C for 20 minutes.

2.2.3.1 Yeast strains

All *S. pombe* strains were kindly constructed and provided by Oliver Fleck (Bangor University).

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

To make YEL, 0.5% w/v yeast extract and 3% w/v glucose were made up to 1L in dH_2O. This was supplemented with 100mg/L of adenine, histidine, uracil, lysine and arginine (Formedium). To make YEA, 1.6% w/v Bacto agar was added to YEL.

2.2.3.3 Glycerol for long-term storage

Glycerol was purchased from Applied Biosystems. A solution of 50% glycerol for long term storage of *S. pombe* cultures was made by diluting 100ml of 100% glycerol with 100ml dH_2O.

2.2.3.4 Antibodies for immunofluorescence

A primary rat monoclonal anti-cisplatin modified DNA antibody was purchased from Abcam, which recognises the same epitope as that formed by

oxaliplatin adducts. A secondary goat anti-rat Alexa 488 antibody was purchased from Life Technologies.

2.2.3.5 Reagents for immunofluorescence

The following reagents were all purchased from Sigma Aldrich: 1,4-Piperazinediethanesulfonic acid (PIPES), Ethylene-bis(oxyethylenenitrilo)tetraacetic acid (EGTA), magnesium sulphate ($MgSO_4$), sorbitol, Triton X-100, BSA and gelatin. PBS without magnesium and calcium was purchased from Life Technologies. Lysine was purchased from Formedium.

2.2.4 Drugs for mammalian cell line and *S. pombe* treatment

Oxaliplatin was purchased from R&D Systems and dissolved in H_2O.

2.2.5 Clinical material

All blood samples from COIN (MRC Clinical Trials Unit [CTU] No. CR10) and COIN-B trial patients were obtained with patient consent and ethical approval for bowel cancer research.

2.2.6 Polymerase chain reaction (PCR)

AmpliTaq Gold DNA polymerase with appropriate buffer and $MgCl_2$ were purchased from Applied Biosystems. Deoxyribonucleotide triphosphates (dNTPs) were purchased from GE Healthcare. All primers were purchased from Eurogentec. Dimethyl sulfoxide (DMSO) was purchased from Sigma Aldrich.

2.2.7 PCR purification

Exonuclease I (Exo) was purchased from New England Biolabs. Shrimp alkaline phosphatase (SAP) was purchased from GE Healthcare.

2.2.8 Electrophoresis

Agarose was purchased from Eurogentec. GelRed was supplied by Biotium. 100bp DNA ladder and 1kb Plus DNA ladder were purchased from New England Biolabs and Life Technologies, respectively.

2.2.9 Sanger sequencing

BigDye Terminator cycle sequencing kit version 3.1 and POP6 polymer were purchased from Applied Biosystems. Capillary electrophoresis buffers were purchased from Sigma Aldrich.

2.2.10 Sanger sequencing clean up

Isopropanol was purchased from Fisher Scientific and HiDi formamide was purchased from Applied Biosystems.

2.3 Equipment

2.3.1 Plastics and glassware

Sterile pipette tips and tips for multi-channel pipettes were purchased from Anachem. Sterile stripettes were purchased from Corning Incorporated. Plastic Eppendorf tubes (0.65ml and 2ml) were purchased from Bioquote, whereas 1.5ml Eppendorf tubes were purchased from Sigma Aldrich. Sterile universal tubes were purchased from Bibby Sterilin. Thermo-Fast PCR reaction plates were supplied by Fisher Scientific, whilst 4titude adhesive PCR sealing sheets and 0.2ml plastic strip tubes were supplied by ABgene. Sterile falcons (15ml and 50ml) were purchased from Fisher Scientific. Sterile tissue culture flasks were purchased from Helena Biosciences. Sterile six-well multidishes were purchased from Fisher Scientific. Sterile 150mm dishes and 100mm dishes were purchased from Sigma Aldrich. Sterile cryovial tubes were supplied by Corning Inc. Glass flasks and beakers were supplied by Jencon or Fisher Scientific.

2.3.2 Mammalian tissue culture

Cell lines were cultured in a Microflow Peroxide Class II advanced biological safety cabinet (Bioquell Ltd). Flasks were incubated in a Heraeus HERAcell 150 incubator (DJB Labcare) connected to a CO_2 cylinder (The Linde Group).

2.3.2.1 UV treatments

Mammalian cells were treated with UV light using a UV-C lamp (Amersham Life Science) during optimisation and with a stratalinker (Amersham Life Science) during data acquisition.

2.3.2.2 Calculating percentage viability of cells

For measurement of cell viability, cells were counted using a haemocytometer (Marienfield). Viable cells were transparent upon viewing through a light microscope, whereas non-viable cells were stained blue.

2.3.2.3 Immunofluorescence

Microscope slides, Shandon filter cards and Shandon Cytofunnels were supplied by Fisher Scientific. Cytospinning was carried out using a Shandon Cytospin 4.0 cytocentrifuge (Thermo Scientific). Cells were visualised using a Leica SP5 confocal laser scanning microscope (Leica; Central Biotechnology Services).

2.3.2.4 Preparation of cells for FACS

For increasing cell density, cell lines were centrifuged using a Centurion Scientific C2000 bench top centrifuge (Core Life Sciences). Cells were sorted according to viability using a FACS Aria III (BD Biosciences).

2.3.2.5 Extraction of genomic DNA from cell lines

For extraction of DNA from cell lines, samples were centrifuged using a Micro Centaur Compact centrifuge (Measuring and Scientific Equipment).

2.3.2.6 Quantification of nucleic acids

For measurement of DNA concentration, a UV spectrophotometer (NanoDrop ND-800, Labtech International) with appropriate buffers was used.

2.3.2.7 Quantification of ELISA absorbance

Absorbance of ELISA plates was quantified using a standard microplate reader (Tecan) at 450nm, with a reference wavelength of 620nm.

2.3.3 Thermocycling

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

2.3.4 Electrophoresis

Electrophoresis was carried out in an ABgene AB0708 100V gel tank attached to a BioRad 200/2.0 power pack. Visualisation of GelRed stained gels was carried out using a BioRad GelDoc XR transluminator.

2.3.5 Sanger sequencing

Sanger sequencing was achieved using an ABI 3100 Genetic Analyser (Applied Biosystems). All data was analysed using Sequencher v.5.3. Reference sequences were obtained from online databases including NCBI (http://www.ncbi.nlm.nih.gov/gene/) and Ensembl (http://www.ensembl.org/index.html).

2.3.6 Whole exome sequencing

Whole exome sequencing was carried out on germline DNA samples from 50 COIN and COIN-B patients with advanced CRC. All DNA samples were processed at Beijing Genomics Institute (BGI). Thirty exomes were captured using a SureSelectAll Exon V2 enrichment kit (Agilent) and sequenced using a HiSeq 2000 (Illumina) to a depth of >30x coverage. A further ten exomes were captured using a SureSelectAll Exon 50Mb enrichment kit (Agilent) and sequenced using a HiSeq 2000 to >50x coverage. The remaining exomes were captured using a SeqCap EZ Exome capture kit V2 (Roche Nimblegen) and sequenced on a Genome Analyzer IIx (Illumina) to a depth of approximately 30x coverage.

2.4 Imaging, statistical and bioinformatics software

Images acquired by confocal microscopy were quantified using Leica LAS-AF Lite software, which was kindly provided by Dr. Christopher George (Cardiff University). Statistical analysis was carried out using SPSS v20.0. Microplate absorbance was read using XFluor software (supplied by Tecan).

Variant calling was carried out by Marc Naven using a bioinformatics pipeline that was established using BWA v.0.5.9 (Li and Durbin, 2009), SAMtools v.0.1.18 (Li *et al.* 2009), GATK v.1.3-145 (McKenna *et al.* 2010; DePristo *et al.* 2011) and SnpEff v.2.04 (Cingolani *et al.* 2012). A custom script written in Perl was established to analyse pipeline output. All software was run using the UNIX command line with default parameters. All bioinformatic analysis was carried out on a workstation running the Ubuntu operating system v.11.10.

2.5 Methods

2.5.1 General reagents

10 x TAE buffer: 400mM Tris, 200mM Acetic acid, 10mM EDTA to pH 8.0

PEM buffer: 100mM PIPES, 1mM EGTA, 1mM $MgSO_4$, pH 6.9

PEMS buffer: PEM supplemented with 1M sorbitol

PEMST buffer: PEMS supplemented with 1% Triton X-100

PEMBALG buffer: PEM pH 6.9 supplemented with 1% BSA, 100mM L-lysine, 1% BSA and 1% gelatin

2.5.2 Genotyping

Genotyping of Pro379Ser (rs1799802), Arg576Thr (rs1800068) and Glu875Gly (rs1800124) in *ERCC4* in 480 human EBV-transformed lymphoblastoid cell lines was carried out by KBiosciences using their Kaspar technology. Genotyping of other known polymorphisms in *ERCC4* and other NER genes in these cell lines was carried out by Illumina using their GoldenGate technology.

2.5.3 Mammalian tissue culture techniques

2.5.3.1 Growth of mammalian cell lines

Human EBV-transformed lymphoblastoid cell lines were maintained in RPMI-1640 with L-glutamine supplemented with 10% foetal bovine serum and penicillin/streptomycin and incubated at 37°C with 5.0% CO_2.

2.5.3.2 Long-term storage of cell lines

Cell lines were cooled gradually to -80°C in a solution of 90% FBS and 10% DMSO at a density of 1 x 10^7 cells/ml for 24 hours, followed by transfer to liquid nitrogen for long term storage.

2.5.3.3 Viability measurements

Viability was measured by trypan blue exclusion, whereby cell suspension was mixed with trypan blue in a 1:1 ratio and examined visually. Viable, intact cells appeared clear whereas non-viable cells appeared blue, particularly in the cytoplasm. Percentage viability was calculated at each time point by calculating the ratio of viable cells to total cell number.

2.5.3.4 Immunofluorescence

Mammalian cells were cytospun onto glass microscope slides at 1,000 RPM for four minutes at medium acceleration without prior serum starving (approx. 5 x 10^5 cells per slide). Cells were fixed in 4% paraformaldehyde, permeabilised in 0.2% Triton-X100, incubated in sodium borohydride (0.5mg/ml) to quench aldehyde groups and blocked in 1% bovine serum albumin (BSA). Cells were incubated with primary antibody overnight at 4°C and with secondary antibody for 1 hour at room temperature, with three wash steps in PBS in between. Cells were incubated with DAPI nuclear stain (0.1mg/ml) for 5 minutes at room temperature, prior to mounting microscope slides with Mowiol-488 containing DABCO for anti-fade. Slide preparation for confocal imaging of untreated and oxaliplatin-treated cells was identical, with the exception of antibodies used.

2.5.3.5 FACS

Aliquots of cell suspension were centrifuged at 1,200 RPM for five minutes and suspended at a density of 1 x 10^7 cells/ml in normal tissue culture media with FBS and antibiotics. Propidium iodide (PI; 50µg/ml) stain was added for permeation of non-viable cells, which were removed by FACS using a BD Aria III (BD Biosciences) via an 85 micron nozzle, to prevent masking of DNA repair due to the presence of non-viable cells. Cells were collected in normal tissue culture media containing FBS and antibiotics prior to DNA extraction.

2.5.3.6 Extraction of genomic DNA

Genomic DNA was extracted using a commercial extraction kit according to the manufacturer's instructions (Qiagen). Briefly, cell suspensions were centrifuged at 1,200RPM for eight minutes using a Centurion Scientific C2000 bench top centrifuge (Core Life Sciences) and cell pellets were suspended in PBS containing proteinase K and lysis buffer. Cells were incubated at 56°C for 15 minutes, followed by ethanol precipitation and DNA purification by wash buffers specific to the Qiagen QIAamp DNA Mini Kit (Cat no. 51306). DNA was eluted in 20µl elution buffer and stored at -20°C for long-term storage.

2.5.4 *S. pombe* techniques

2.5.4.1 Growth of S. pombe

All cultures were incubated at 30°C. All glassware and reagents were sterilised by autoclaving before use.

2.5.4.2 Preparation of YEA and YEL

YEA and YEL were prepared as outlined in section 2.2.4.2 and autoclaved on a liquid cycle. YEA was cooled to an appropriate temperature and subsequently poured into sterile petri dishes (25ml per dish) and allowed to solidify. Plates were stored at room temperature until use.

2.5.4.3 Starter cultures

For each culture, 10mls of YEL was added to sterile glass flasks. Using a sterile loop, a single colony was transferred from growing plates to the glass flask. Cultures were incubated at 30°C on an orbital shaker for up to 24 hours at 200RPM. The following six strains of *S. pombe* were cultured: $rad16^{WT}$, $rad16^{Pro361Ser}$, $rad16^{Arg399Gln}$, $rad16^{Arg548Thr}$, $rad16^{Ser585X}$ and $rad16^{Glu844Gly}$.

2.5.4.4 Long-term storage of S. pombe

Cultures were prepared for long-term storage by mixing 600ul of glycerol with 400ul of overnight culture in YEL and freezing to -80°C. Cultures were restored by thawing on ice, vortexing and streaking onto YEA plates, prior to incubation at 30°C for 3-5 days.

2.5.4.5 Oxaliplatin treatment

Overnight cultures (1 x 10^7 cells in YEL) were incubated with 1mM oxaliplatin (R&D Systems) for 18 hours at 30°C. Following incubation, cells were centrifuged at 2000RPM for 3 minutes and the supernatant containing oxaliplatin was removed. The cell pellet was re-suspended in YEL in preparation for immunofluorescence.

2.5.4.6 Immunofluorescence

Cultured cells were fixed in 3.7% formaldehyde and 0.25% glutaraldehyde for 1 hour at 30°C on an orbital shaker at 200RPM, followed by centrifuging at 2000RPM for 3 minutes to remove YEL containing fixative. Cell wall digestion was achieved by incubation with 0.1mg/ml zymolyase suspended in PEM at 37°C for 90 minutes. The cell membrane was permeabilised with 4% Triton X-100 for two minutes at room temperature, followed by blocking with PEMBALG for 1 hour at room temperature, with multiple washes in PEM between each step. Cells were incubated with primary antibody overnight and with secondary antibody for 1 hour, both at room temperature, with three wash steps in PEM in between. Cells were washed in PEM, incubated with DAPI nuclear stain (0.2µg/ml) for 5 minutes at room temperature, prior to cytospinning onto sterile glass microscope slides at 1,000 RPM for 4 minutes at medium acceleration. Cells were mounted with Mowiol-488 containing DABCO for antifade.

2.5.5 Quantification of nucleic acids

To measure DNA concentration, a UV spectrophotometer was used at wavelengths of 260nm and 280nm. An absorbance ratio of 1.8 at these wavelengths indicated high purity in a given sample.

2.5.6 Variant prediction using whole exome resequencing data

Using the GATK UnifedGenotyper tool, single nucleotide polymorphisms (SNPs) and insertions/deletions (indels) were called. Addition of rsIDs to called variants was performed using dbSNP v.132. Variant quality score recalibration was performed according to GATK guidelines; SNPs and indels were firstly separated since variant quality recalibration only applied to SNPs. Training sites

from HapMap release 3.3 (http://hapmap.ncbi.nlm.nih.gov) and Illumina's Omni2.5M SNP microarray were used to recalibrate scores, to ascertain whether called variants were genuine or artefacts.

Annotation of variants was applied using the SnpEff program and its database GRCh37.64. Annotations were assigned to variants called by the pipeline using GATK's VariantAnnotator, producing files for each exome containing recalibrated quality scores.

2.5.7 Primer design

All primers were designed either using Primer3, version 0.4.0 (http://frodo.wi.mit.edu/primer3/), or alternatively using a Perl script produced by Marc Naven (Naven, 2015; Chapter 2). Where possible, primers were designed to be 18-25 nucleotides in length, with annealing temperatures within 2°C of their partners, and demonstrated low predicted dimerisation and secondary structure formation. The locus specificity of all primers was checked using the Primer-Blast software (http://www.ncbi.nlm.nih.gov/tools/primer-blast/).

2.5.8 PCR

Unless otherwise mentioned, standard PCR reaction mixtures comprised of 10pmols forward and reverse primer, 0.2mM dNTPs, GeneAmp 10x buffer (added to a final concentration of 10mM Tris-HCl, 50mM KCl, 1.5mM $MgCl_2$, 0.01% (w/v) gelatine, pH 8.3), 1U AmpliTaq Gold DNA polymerase, 5% DMSO and 40ng of DNA (final volume of 25µl). Cycling conditions involved an initial denaturisation step of 95°C for 2 minutes, followed by 35 cycles of 95°C for 30 seconds, annealing temperature of between 50-60°C for 30 seconds and an elongation step of 72°C for 30 seconds. A final elongation step of 72°C for 10 minutes followed these cycles.

2.5.9 Agarose gel electrophoresis

Agarose gels were formed using 1.5g agarose dissolved in 100mls 1X TAE buffer (depending on fragment sizes that required separation). Conical flasks were heated to melt the agarose, cooled slightly and 0.05µg/ml of either ethidium bromide or GelRed stain was added. This was poured into a gel tank between dams and allowed to cool. Once set, the gel was submerged in 1x TAE

buffer in an AB0708 100V gel tank. 8µl of PCR product was added to 2µl of loading dye (15% w/v ficol, 10mM Tris pH 8, 1mM EDTA and 0.2% orange G), mixed thoroughly and the entire volume was loaded onto the gel. Gels were run at 100V for 30-40 minutes with a 100bp or 1kb DNA ladder to visualise fragment separation. After separation, UV visualisation was carried out and photographed on the BioRad XR system. As a safety precaution, ethidium bromide destaining bags were added to running buffer for at least 24 hours prior to disposal to remove the dye. Where GelRed was used this practice was not necessary.

2.5.10 ExoSap PCR purification

ExoSap degrades excess primers, ssDNA and phosphate groups. Exo is a 3'-5' exonuclease that degrades single stranded oligonucleotides from solutions that contain a double stranded product. SAP is an alkaline phosphatase responsible for removal of 5'-phosphates from the PCR product. 1µl (3.2pmol) of ExoSap was added directly to the PCR product after electrophoresis and the samples were incubated at 37°C for 60 minutes, followed by an enzyme denaturising step of 80°C for 15 minutes.

2.5.11 Sanger sequencing

The BigDye Terminator version 3.1 Cycle Sequencing kit was used to sequence the ExoSap treated PCR products. The reaction mixture used was based on the manufacturer's instructions; 5µl purified PCR product was added to 0.2% BigDye version 3.1, 10pmol desired primer (forward or reverse) and 1x BigDye sequencing buffer, all of which was made up to 10µl with dH_2O.

Cycling conditions involved 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 3 minutes and 30 seconds. The products of the BigDye termination sequencing reactions were cleaned of excess unincorporated nucleotides and dyes by the isopropanol method, as outlined below.

2.5.12 Isopropanol clean up method

For isopropanol clean up, 40µl of 75% isopropanol was added to the 10µl BigDye reaction mixture and incubated for 30 minutes at room temperature. Samples were centrifuged at 4000RPM for 45 minutes to pellet the DNA, and

then inverted onto absorbent paper to remove the isopropanol supernatant. Samples were then centrifuged inverted for 500RPM for 30 seconds and air dried in the dark for 10 minutes to allow evaporation of remaining liquid. The DNA pellet was resuspended in 10µl of HiDi formamide, prior to Sanger sequencing.

Chapter 3 – Investigating viability of wild type and variant *ERCC4* cell lines after DNA damage

3.1 Introduction

Mutations in XP genes have been associated with decreased survival in model organisms and clinical samples, resulting from defective DNA repair in numerous pathways, primarily NER and DSB repair. Studies looking at overall survival have been carried out in clinical trial patients undergoing chemotherapy (Grothey *et al.* 2004; Grothey *et al.* 2005; Andre *et al.* 2009), whilst *in vitro* models have primarily looked at biallelic cell lines carrying XP variants. A summary of studies investigating survival using *in vitro* and *in vivo* NER-deficient models is shown in Table 3.1. We hoped to further this work by examining the *in vitro* consequences of potentially functional heterozygous variants in *ERCC4*.

Platinum drugs including oxaliplatin have been shown to bind irreversibly to plasma proteins, in particular serum albumin (Graham *et al.* 2000; Yue *et al.* 2009), and as a result cells were treated with oxaliplatin in serum-free (SF) media throughout optimisation stages to prevent this drug-serum interaction. Additionally, UV irradiation can be absorbed by serum molecules, which could prevent damage from occurring (Polet and Steinhardt, 1968). Therefore, cell lines were treated with UV in SF media to prevent absorption of UV light and potential interference with the viability phenotype.

Trypan blue is an effective exclusion dye used for viability measurement based on its ability to enter non-viable cells that have lost membrane integrity (Krause *et al.* 1984). This assay is also advantageous because it does not require expensive, specialist equipment for viability measurement.

We hypothesised that functional variants in *ERCC4* may contribute to reduced repair capacity, which may directly correlate with reduced viability. Here we aimed to elucidate the functional consequences of variants in *ERCC4* on viability after DNA damage by oxaliplatin and UV light.

Gene (Protein)	Model organism	Effect seen	References
ERCC1 (ERCC1)	Human colorectal cancer cell lines wild type and *KRAS*-mutant (biallelic)	Significantly higher sensitivity to oxaliplatin in *KRAS*-mutant cell lines, silencing of ERCC1 significantly reduced survival following oxaliplatin, irrespective of *KRAS* status	Orlandi *et al.* 2015
ERCC2 (XPD)	Human lymphoblastoid cell lines with transfected constructs to induce heterozygous and homozygous *ERCC2*-deficient phenotypes	Reduction in survival following UV treatment in heterozygous cell lines; significantly reduced survival with homozygous variants (2.5-fold increase in sensitivity to UV)	Seker *et al.* 2001
ERCC3 (XPB)	Human heterozygous fibroblast cell lines	Significantly reduced survival in XPB-deficient cell lines compared with wild type and cell lines isolated from patients with TTD	Weeda *et al.* 1997
ERCC4 (XPF)	Biallelic fibroblast cell lines isolated from patients with Fanconi anaemia (FA)	Reduced survival in mutant-ERCC4 patients after treatment with UV-C and mitomycin C (MMC), compared with wild type cells; Suggests a multifunctional role of XPF	Bogliolo *et al.* 2013
ERCC5 (XPG)	Fibroblasts isolated from XPG-deficient patients (one homozygous and two compound heterozygous cell lines)	Significantly reduced survival after UV-C irradiation in XPG-deficient fibroblasts compared with wild type cells; effect seen at varying UV doses (5-30J)	Schafer *et al.* 2013
XPA (XPA)	Human heterozygous XPA-mutant fibroblast cell lines, which were transfected with constructs to induce a homozygous phenotype	Increased sensitivity to UV treatment in mutant XPA cell lines compared with wild type	Wu *et al.* 2006
XPC (XPC)	HeLa cells with silenced *XPC* (using constructs to induce a homozygous phenotype)	Reduced survival following UV treatment in *XPC*-silenced cells; Reduced cell division rate following UV treatment, suggesting that XPC is essential for survival and also regulates cell division following DNA damage	Renaud *et al.* 2011

Table 3.1 Summary of survival studies after treatment with DNA damaging agents in NER-deficient model organisms.

3.2 Materials and Methods

3.2.1 Cell line selection and genotyping

Four hundred and eighty human EBV-transformed lymphoblastoid cell lines from healthy Caucasian blood donors were genotyped for Pro379Ser, Arg576Thr and Glu875Gly in *ERCC4* using KBioscience® KASPar genotyping. From this analysis, twelve cell lines were selected for functional analysis; three wild type cell lines and three cell lines for Pro379Ser, Arg576Thr and Glu875Gly (all found in a heterozygous state). Cell lines were purchased from ECACC (Public Health England).

3.2.2 Oxaliplatin treatment

During optimisation, wild type cells were treated with oxaliplatin in SF media as an adherent monolayer on poly-L-lysine coated dishes and in their natural suspension state. A range of suspension media (serum-free [SF], 1% serum and phosphate buffered saline [PBS]) and doses (125-600μM) in SF media were tested over 48 hours in wild type cells and cells carrying Pro379Ser, followed by smaller dose ranges (100-200μM) and blind treatment of the selected dose (125μM) in SF media over 48 hours. A schematic of the stages of optimisation is given in Figure 3.1.

Optimised conditions were as follows: cells were centrifuged at 1,200 RPM for five minutes before removal of supernatant and re-suspension in SF media. Cells were incubated in SF media for 1 hour before treatment with oxaliplatin (125μm) by adding the drug directly to the cell suspension. The drug was incubated with cells for the duration of the experiment.

3.2.3 UV irradiation

During optimisation, wild type cells were irradiated with a lower dose (20-60J) and high dose (60-100J) UV-C over 3 and 48 hours. Wild type cells were suspended in SF media and PBS prior to UV (20-60J) treatment and viability was measured over 48 hours.

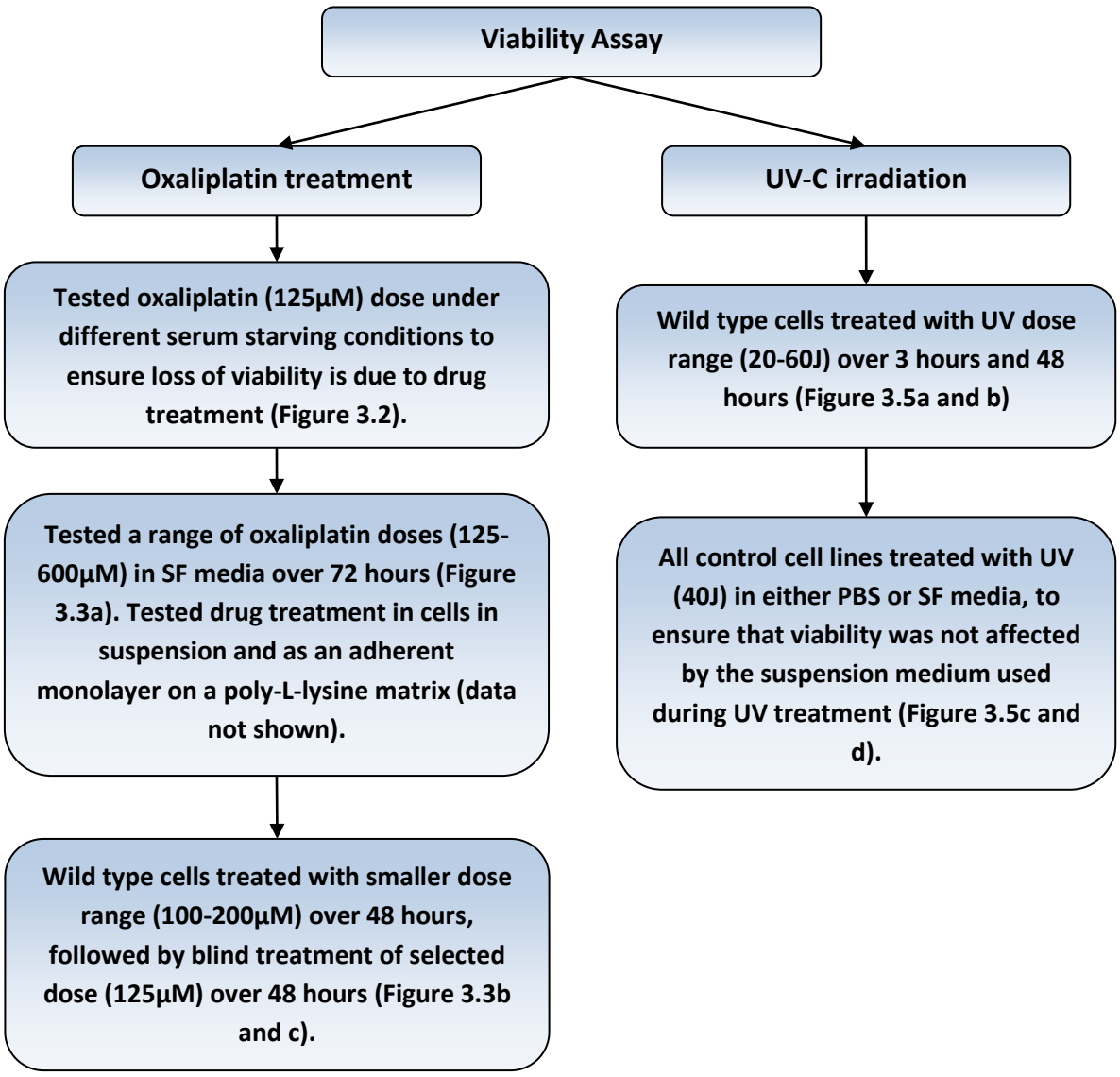


Figure 3.1 Schematic diagram of optimisation of oxaliplatin and UV viability assays. Wild type cells were treated with oxaliplatin in suspension and as an adherent monolayer on a poly-L-lysine surface. Different serum starving conditions and a range of doses (125-600µM) were tested, which led to a narrower dose range (100-200µM) being tested over 48 hours. A final dose of 125µM was selected on the basis of these experiments and published data, which was tested blindly across all wild type cell lines to confirm adequate loss of viability over 48 hours at this dose.

Wild type cells were treated with UV (20-60J) using a germicidal lamp and viability was measured over 3 hours and 48 hours. UV irradiation in phosphate buffered saline (PBS) and serum-free (SF) media were tested to rule out the possibility of UV-C radiation being absorbed by other molecules including water molecules in the suspension medium.

Optimised conditions were as follows: cells were suspended in SF media as outlined in 3.2.2 and irradiated with UV (40J) using a germicidal lamp. After UV treatment, cells were resuspended in normal media containing foetal bovine serum (FBS) and antibiotics.

3.2.4 Viability measurement

Percentage viability was measured by trypan blue exclusion as outlined in section 2.5.3.3 at 0, 4, 24 and 48 hours after treatment. Each experiment was carried out four times under identical conditions, and percentage viability for each cell line at each time point was averaged. Initial percentage viability across all cell lines varied considerably (61.5-90.9%; data not shown), so viability values were normalised to produce a starting viability of 100%, prior to averaging. Average percentage viability (n=4) for all *ERCC4* variants was plotted against time (hr) with standard error bars.

3.2.5 Statistical analysis

Statistical analysis was performed in SPSS v.20.0 using a two-way analysis of variance (ANOVA), using mutation status (wild type, Pro379Ser, Arg576Thr and Glu875Gly) and treatment (no treatment, oxaliplatin and UV treatment) as independent variables. Percentage viability at each time point was used as the dependent variable. Percentage viability values were averaged for cell lines carrying the same *ERCC4* variant undergoing the same treatment and an ANOVA was performed to allow comparison of percentage viability between cell groups. We were unable to apply statistics to the viability at zero hours due to no standard deviations (counted as 100% starting viability). Statistics were applied to the viability values measured at 2, 4, 24 and 48 hours across all cell groups for each treatment type.

3.3 Results

3.3.1 Genotyping of human cell lines for variants in NER genes

Of the 480 lymphoblastoid cell lines that were genotyped, nine (1.88%) carried Pro379Ser, four (0.83%) carried Arg576Thr and seventeen (3.54%) carried Glu875Gly. We selected three of each of these cell lines together with three wild type cell lines for subsequent analyses.

3.3.2 Optimisation of oxaliplatin treatment

A summary of optimisation experiments is shown in Figure 3.1. Wild type cells were treated with oxaliplatin (125µM) over 4 hours in SF media and viability was measured every 30 minutes, which demonstrated no noticeable difference between treated and untreated cells (data not shown).

The cell lines in question exist naturally in suspension. We tested the effect of treating cells as an adherent monolayer using poly-L-lysine as an adhesion matrix. Wild type cells were treated with oxaliplatin (125µM) in SF media over 4 hours and removed either by mechanical scraping or using trypsin at hourly intervals and viability was measured. This revealed a lower initial viability after trypsin treatment or scraping compared with initial viability when cells are treated in suspension (approximately 45-65% compared with 85% when treated in suspension; data not shown). This indicated that these cell lines are healthier when treated in their natural suspension state.

We tested different serum conditions with oxaliplatin treatment (125µM) over 48 hours, by comparing 1% serum and serum-free media with 1-hour and overnight starvation in wild type cells and cells carrying Pro379Ser (Figure 3.2). Viability measurements over 48 hours revealed that neither duration of serum starving nor the presence of 1% serum in the tissue culture media had any detrimental effect on the viability of cells.

Wild type cell lines were treated with range of concentrations (0, 125, 250, 375, 500 and 600µM) over 48 hours, showing considerable reduction in viability in treated cells across all doses tested at 24, 48 and 72 hours (Figure 3.3a). As a result, the concentration range was reduced (100, 125, 150, 175 and 200µM) and tested over 48 hours in wild type cells, revealing similar viabilities at each concentration (Figure 3.3b). A final concentration of 125µM was used to treat control cells blindly, which effectively demonstrated a marked reduction in viability between treated and untreated cells at 24 and 48 hours (Figure 3.3c).

3.3.3 Effect of oxaliplatin on viability of variant cell lines

We observed very little reduction in viability over 48 hours across all cell lines without drug treatment (approx. 10% reduction; Figure 3.4). Oxaliplatin

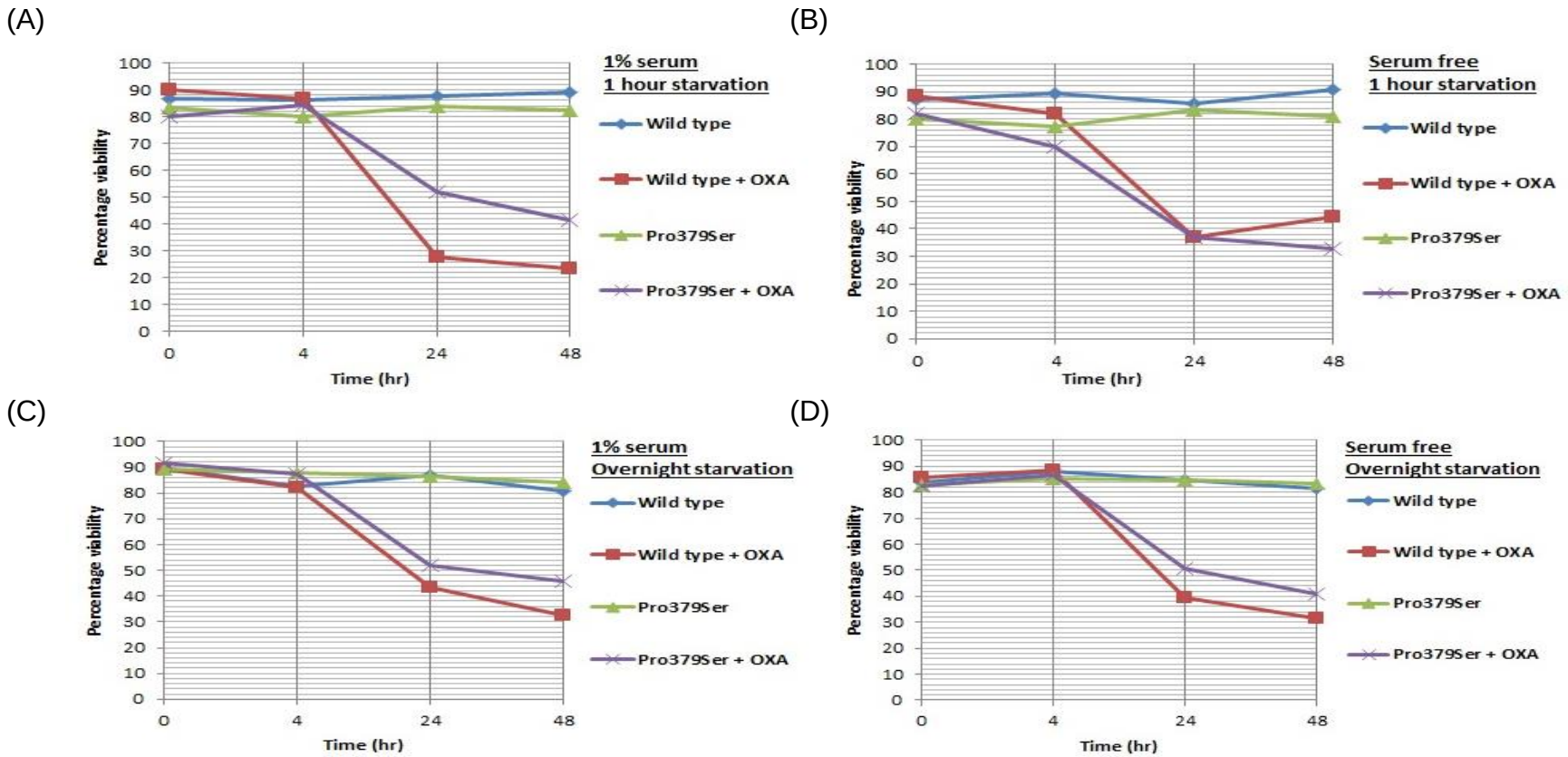


Figure 3.2 Effects of serum starvation on oxaliplatin treatment in wild type cells and cells carrying Pro379Ser. Wild type cells and cells carrying Pro379Ser were treated with oxaliplatin (125µM) over 48 hours under different serum starving conditions. Plots show cells treated in (A) 1% serum with 1 hour starvation prior to drug treatment, (B) serum-free media with 1 hour starvation, (C) 1% serum with overnight starvation and (D) serum-free media with overnight starvation. Different serum starving conditions had very little effect on viability in either wild type or variant cells during optimisation.

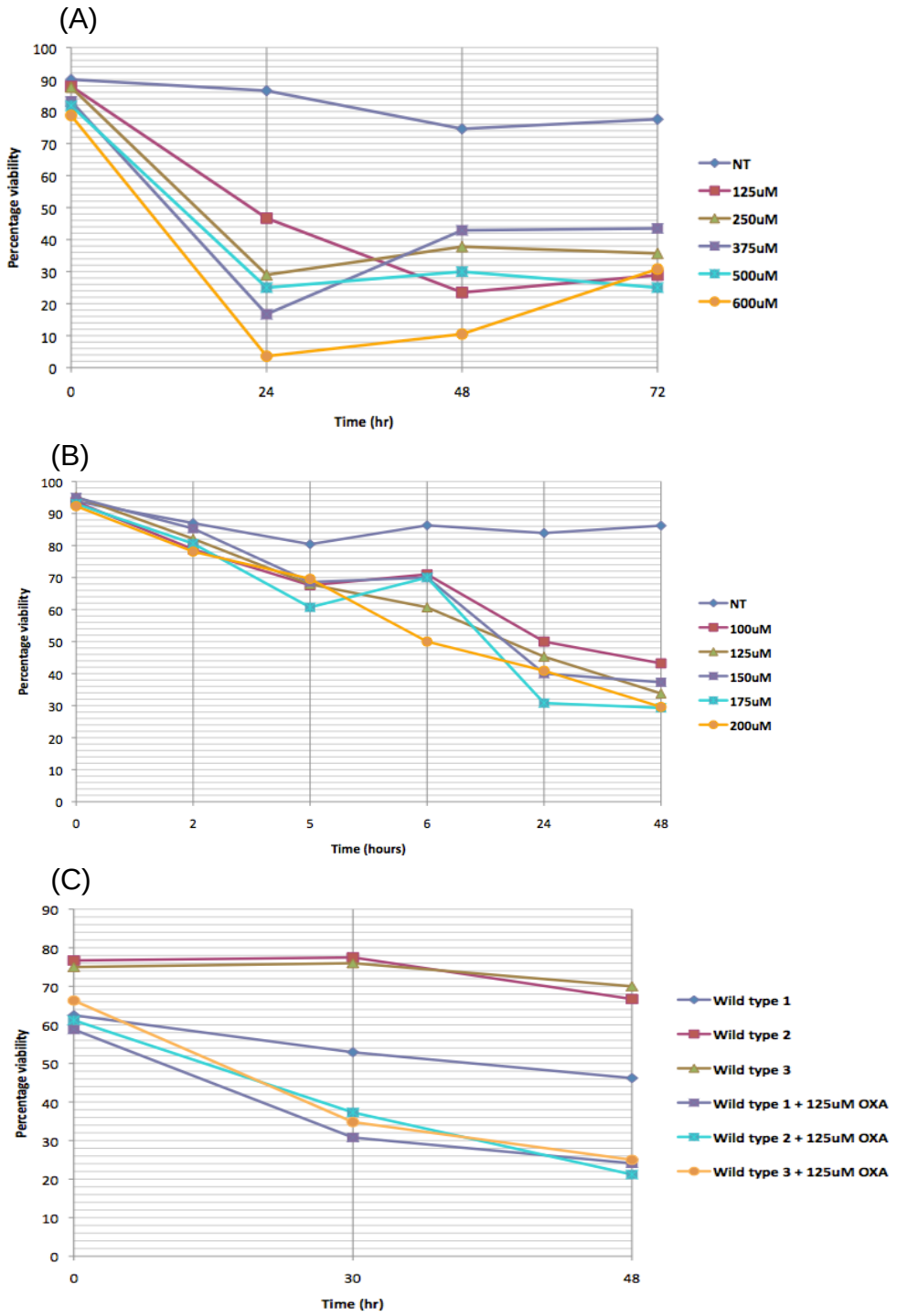


Figure 3.3 Optimisation of oxaliplatin dose in wild type cell lines. (A) Cells from wild type cell line 1 were suspended in SF media containing a range of oxaliplatin doses (125-600uM) and viability was measured at 0, 24, 48 and 72 hours. (B) From this data, the dose range was narrowed down to 100-200µM, which was tested in wild type cells suspended in SF media over 48 hours. Oxaliplatin remained in the media for the duration of the experiment. (C) All wild type cell lines were suspended in SF media and treated with oxaliplatin (125uM). Viability was measured at 0, 30 and 48 hours. Drug treatment over 48 hours resulted in a noticeable reduction in viability in drug treated cells compared with untreated cells. NT = No treatment.

treated cell lines demonstrated a reduction in viability at 24 and 48 hours compared with untreated cell lines, with very little variation between wild type and variant cell lines at these time points (Figure 3.4). This was true for all *ERCC4* variants in question.

Comparison of each cell group based on *ERCC4* mutation with and without oxaliplatin treatment by ANOVA revealed no statistically significant association between mutation status and treatment type after 24 and 48 hours (P = 0.941 and 0.975, respectively). Similarly no significant difference between different cell groups was identified at 24 and 48 hours (P= 0.149 and 0.108, respectively). A significant difference between treatment types was identified, as would be expected when comparing untreated and DNA damaged cell lines (P=0.000 at 24 and 48 hours; Table 3.2).

3.3.4 Optimisation of UV treatment

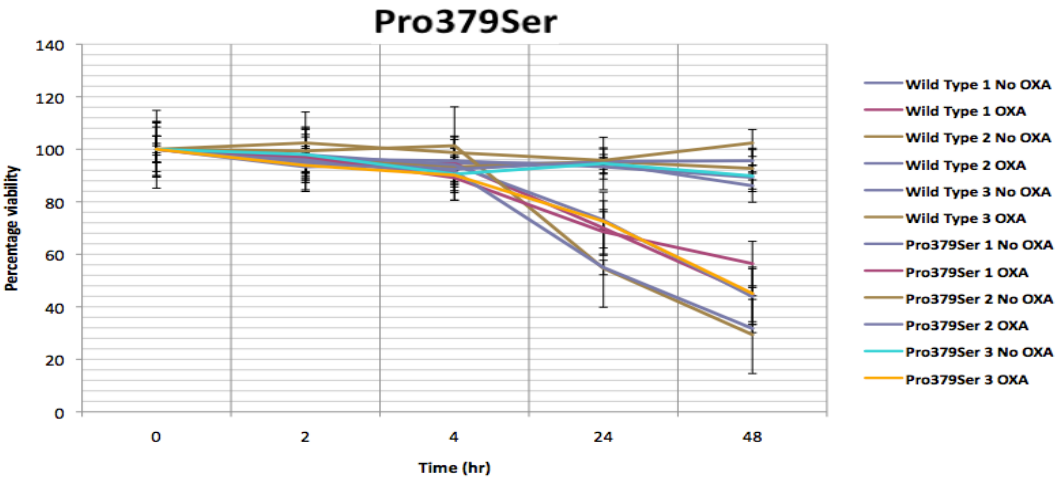
Wild type cells were treated with UV-C (20–60J) irradiation in SF media over 3 hours (Figure 3.5a) and over 48 hours (Figure 3.5b), demonstrating very little cell death after 3 hours and a marked reduction in viability at 24 and 48 hours. This was repeated in the other two wild type cell lines, with similar results (data not shown). Wild type cells were treated with a higher UV dose range (60–100J) in SF media over 3 hours and 48 hours, which showed similar results to the lower dose range (data not shown). We tested UV (20–60J) treatment in different suspension media in all wild type cell lines over 48 hours, to compare viability after DNA damage in SF media and in PBS (Figure 3.5 c and d). This showed a reduction in viability in UV irradiated cells at 24 and 48 hours compared to untreated cells, with very little difference in viability between cells treated in either suspension medium.

3.3.5 Effect of UV on viability of variant cell lines

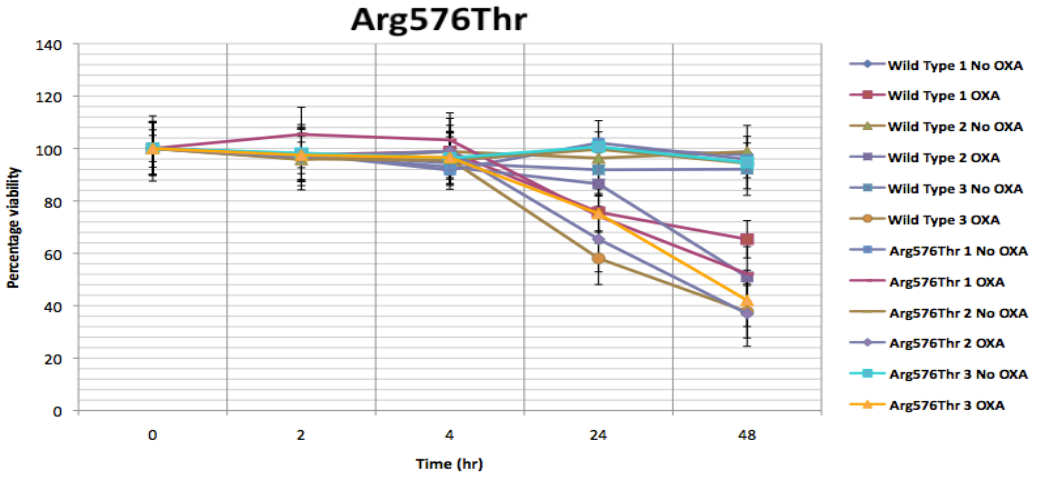
We observed consistently high viability over 48 hours across cell lines without UV treatment (Figure 3.6). Percentage viability in UV treated cell lines reduced considerably 24 hours after irradiation and remained very similar across all cell lines at all time points, irrespective of *ERCC4* variant (Figure 3.6).

Comparison of percentage viability values between cell groups by ANOVA identified no significant association between mutation status and

(A)



(B)



(C)

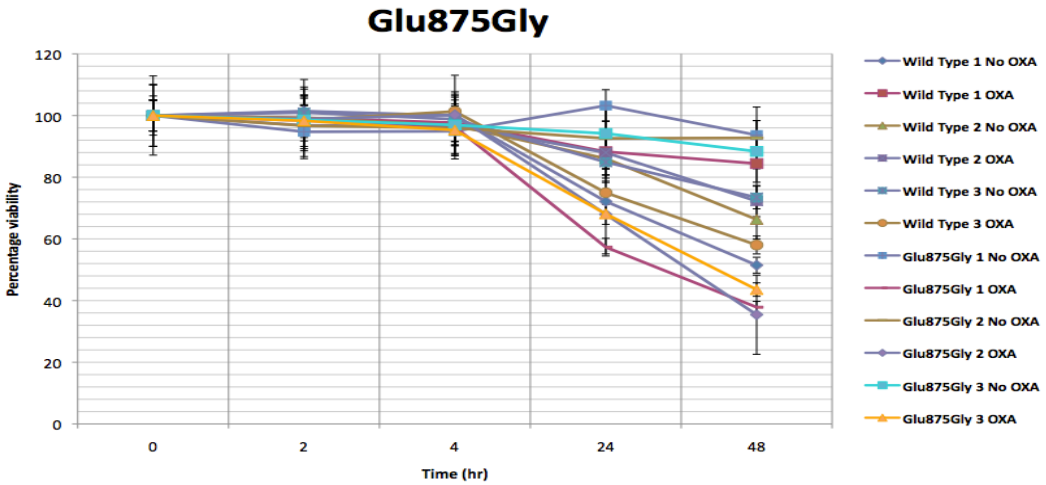


Figure 3.4 Effect of oxaliplatin treatment on viability in wild type and variant cell lines. Wild type cell lines and cell lines carrying *ERCC4* (A) Pro379Ser, (B) Arg576Thr and (C) Glu875Gly with and without oxaliplatin (125µM) treatment over 48 hours in SF media with standard error bars. Each plot represents four biological repeats of the same experiment performed on separate days using identical conditions. V = variant cell line; C = wild type cell line; 1-3 = three cell lines carrying the same *ERCC4* variant.

A.

		24hr		48hr	
Mutation	**Treatment**	**Mean**	**SD**	**Mean**	**SD**
Wild Type	NT	94.4417	5.08946	89.3667	7.15139
	OXA	65.8833	14.94839	39.3333	9.33004
Pro379Ser	NT	95.2250	12.74164	95.9500	10.09874
	OXA	65.3750	11.02519	44.4250	14.25821
Arg576Thr	NT	100.8167	5.96731	94.9583	8.10067
	OXA	71.6000	10.28264	43.7500	11.53095
Glu875Gly	NT	96.6667	7.31478	91.6000	7.66420
	OXA	64.4750	13.45140	38.9500	8.74326

B.

	24hr	48hr
Mutation	0.149	0.108
Treatment	0.000	0.000
Mutation*Treatment	0.941	0.975

Table 3.2 Statistical analyses of cell groups after oxaliplatin treatment.
Percentage viability values for cell lines carrying the same *ERCC4* variant with and without oxaliplatin (125µM) treatment were averaged and compared using a two-way ANOVA. A) Mean and standard deviation values for each cell group without treatment and following oxaliplatin (OXA) treatment after 24 and 48 hours are shown. B) P values obtained by ANOVA are given for percentage viability values after 24 and 48 hours for the variables mutation, treatment and the association between mutation and treatment (mutation*treatment). Highlighted values were statistically significant ($P<0.05$). NT = no treatment; SD = Standard deviation.

treatment type at either 24 or 48 hours (P = 0.930 and 0.088, respectively; Table 3.3). A significant difference between cell groups undergoing different treatment was identified (P = 0.000 at 24 and 48 hours) was identified, as would be expected. However, a significant difference between percentage viability after 48 hours, but not 24 hours, based on cell group was identified (P = 0.004 at 48 hours; P = 0.100 at 24 hours). Further investigation using a post hoc Tukey test identified a significant difference in viability between wild type cells and cells carrying Pro379Ser (P = 0.047), and between cells carrying Pro379Ser and Glu875Gly (P = 0.002; Table 3.3).

3.4 Discussion

3.4.1 Selection of cell lines for functional analysis

It is likely that each of the cell lines analysed carried other coding region variants in *ERCC4*, *ERCC1* or other NER-related pathway genes. A search of dbSNP on 3rd August 2015, revealed two rare (MAF <1%) and sixteen common (MAF >1%) nonsynonymous variants in *ERCC4*, one rare (MAF<1%) and four common (MAF >1%) nonsynonymous variants in *ERCC1*, and, seven rare (MAF <1%) and six common (MAF >1%) nonsynonymous variants in other core NER-related pathway genes (*ERCC2*; *ERCC3*; *ERCC5*; *RPA*; *TFIIH*; *XPA*). It would therefore be impossible to account for all of these genetic variations and to find cell lines that were wild type at all of these mutant loci. Therefore, we selected three cell lines with each mutation of interest in an attempt to allow for possible differences in their genetic backgrounds. The fact that almost all of the individual mutant cell lines showed similar effects in the data presented in this chapter (and elsewhere in this thesis) suggests that other un-assayed variants did not affect the results or their reproducibility.

3.4.2 Optimisation of assay

We tested different serum conditions comparing 1% serum and SF media with 1 hour and overnight starvation prior to drug treatment, to ensure that loss of viability was a consequence of drug treatment, and not due to removal of serum. It was necessary to remove serum prior to oxaliplatin treatment due to the serum-drug interaction that can occur and prevent damage

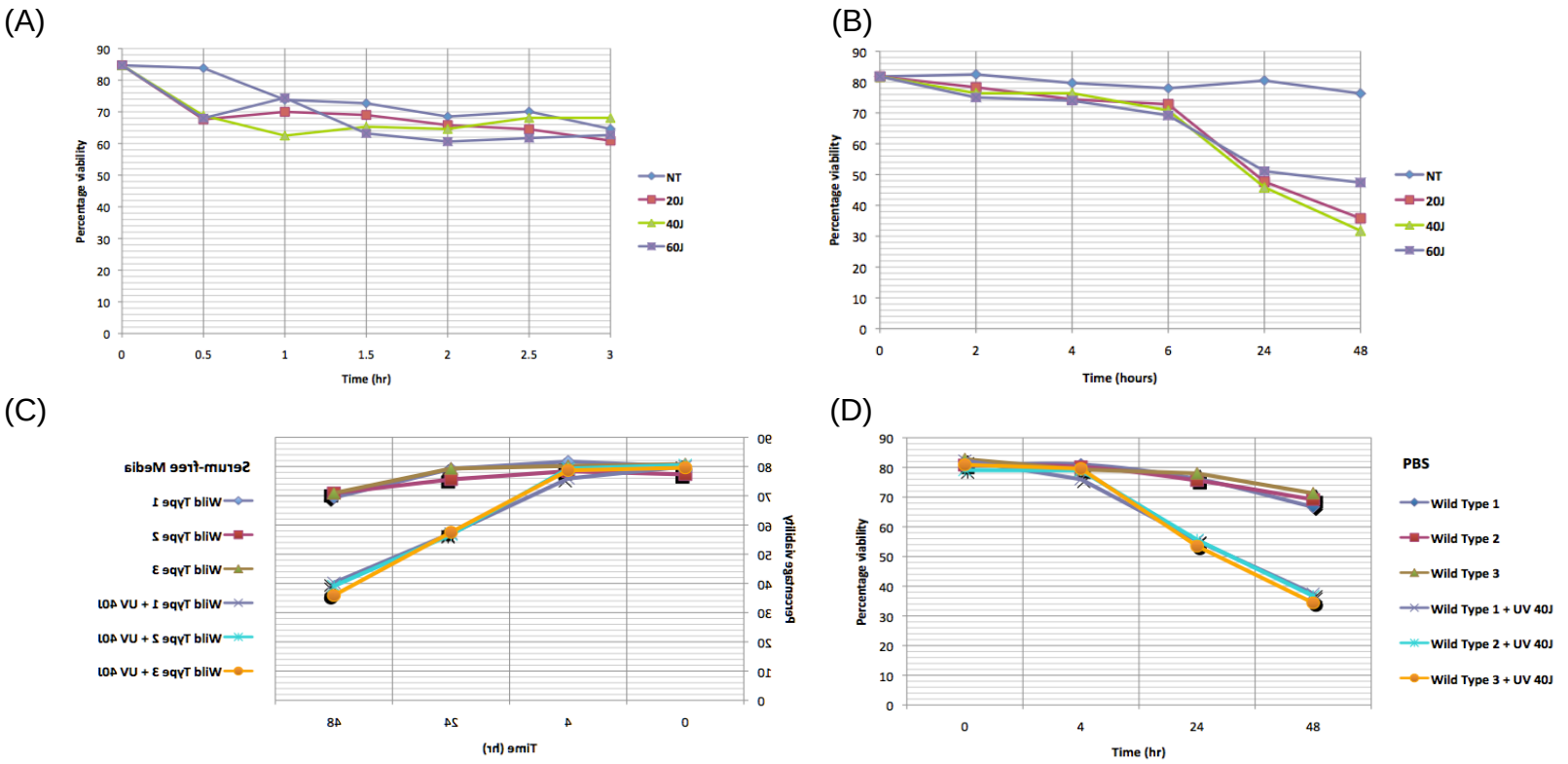


Figure 3.5 Optimisation of UV dose and suspension medium. Wild type cell lines were treated with UV (20-60J), followed by re-suspension of cells for 48 hours in different suspension media. Cells were irradiated with UV (20-60J) using a UV-C germicidal lamp in SF media, prior to re-suspension in normal media for (A) 3 hours and (B) 48 hours. All wild type cell lines were suspended in (C) SF media or (D) PBS prior to UV (40J) using a UV-C germicidal lamp. Cells were incubated for 48 hours and viability was measured at 0, 4, 24 and 48 hours. The suspension medium made very little difference to the viability of cells at each time point, regardless of whether cells were UV treated or untreated. A reduction in viability was only noticeable when cells were incubated for more than 4 hours. NT = No treatment.

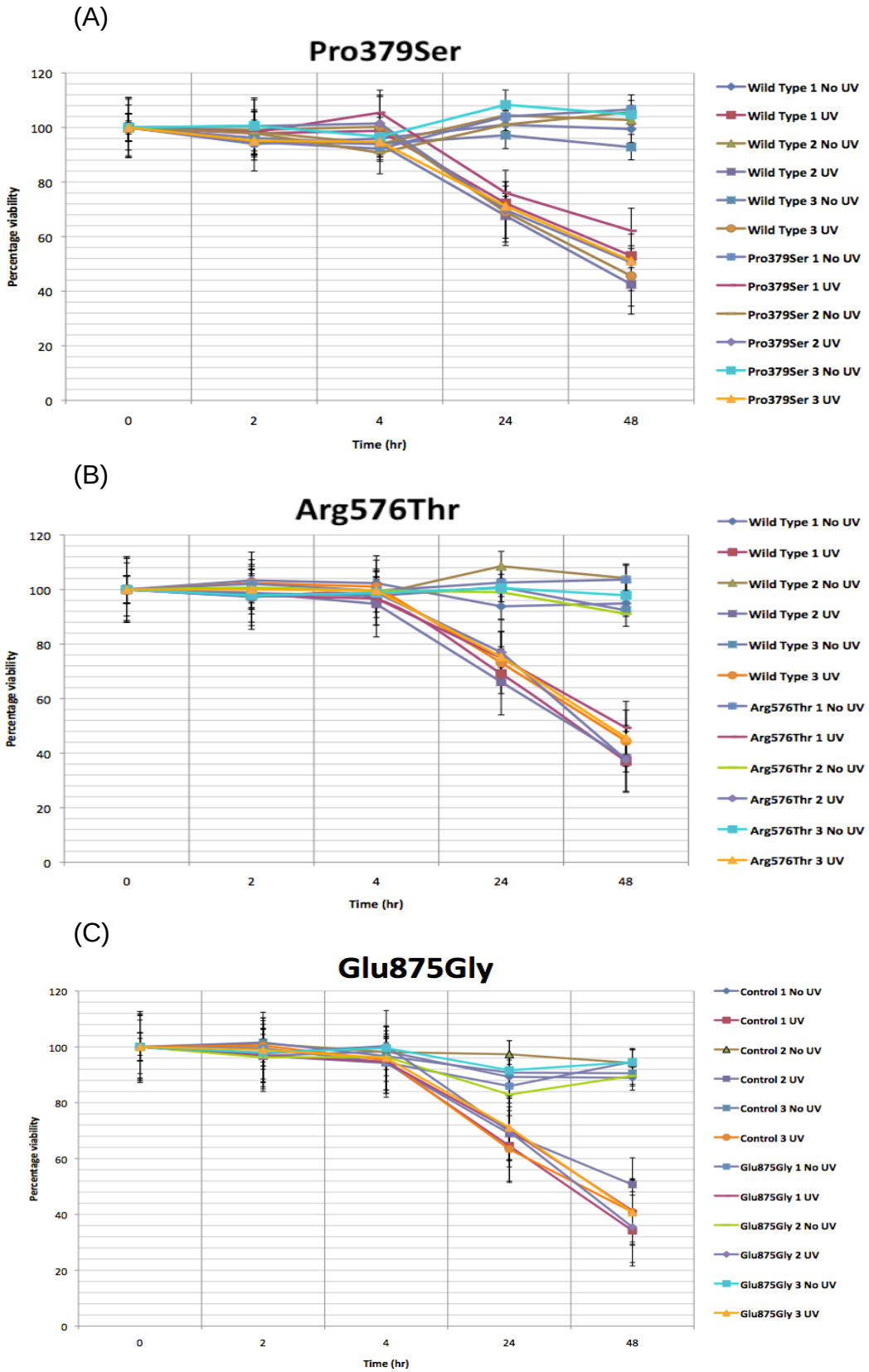


Figure 3.6 Effect of UV irradiation on viability in wild type and variant cell lines. Wild type cell lines and cell lines carrying a (A) Pro379Ser, (B) Arg576Thr and (C) Glu875Gly variant with and without UV-C (40J) irradiation over 48 hours in SF media with standard error bars. Each plot represents four biological repeats of the same experiment performed on separate days using identical conditions. V = variant cell line; C = wild type cell line; 1-3 = three cell lines carrying the same *ERCC4* variant.

A.

		24hr		48hr	
Mutation	**Treatment**	**Mean**	**SD**	**Mean**	**SD**
Wild Type	NT	94.4417	5.08946	89.3667	7.15139
	UV	69.6500	9.01115	47.0083	9.61225
Pro379Ser	NT	95.2250	12.74164	95.9500	10.09874
	UV	72.3500	8.30723	54.6167	9.60339
Arg576Thr	NT	100.8167	5.96731	94.9583	8.10067
	UV	75.7167	6.48352	44.1667	10.90607
Glu875Gly	NT	96.6667	7.31478	91.6000	7.66420
	UV	70.3833	13.26272	39.1750	10.51347

B.

	24hr	**48hr**
Mutation	0.100	0.004
Treatment	0.000	0.000
Mutation*Treatment	0.930	0.088

C.

Mutation		**P value (48hr)**
Wild Type	**Pro379Ser**	0.047
	Arg576Thr	0.956
	Glu875Gly	0.725
Pro379Ser	**Wild Type**	0.047
	Arg576Thr	0.151
	Glu875Gly	0.002
Arg576Thr	**Wild Type**	0.956
	Pro379Ser	0.151
	Glu875Gly	0.409
Glu875Gly	**Wild Type**	0.725
	Pro379Ser	0.002
	Arg576Thr	0.409

Table 3.3 Statistical analyses of cell groups after UV irradiation. Percentage viability values for cell lines carrying the same *ERCC4* variant with and without UV (40J) treatment were averaged and compared using a two-way ANOVA. A) Mean and standard deviation values for each cell group without treatment and following UV treatment after 24 and 48 hours are shown. B) P values obtained by ANOVA are given for percentage viability values after 24 and 48 hours for the variables mutation, treatment and the association between mutation and treatment (mutation*treatment). Highlighted values were statistically significant (P<0.05). C) A post hoc Tukey test was performed to identify which groups significantly differed from one another after 48 hours, with significant p values (<0.05) highlighted. NT = no treatment; SD = Standard deviation.

from taking place (Yue *et al.* 2009). However, we were unsure whether complete serum removal would create an unfavourable environment for cell growth and replication. Overnight starvation was intended to induce cell cycle arrest, although this results in changes in gene expression (Collier *et al.* 2006), and we were unsure of the effects this could have on viability after damage. It was also unknown whether this overnight starvation, followed by drug treatment for 48 hours in the absence of serum, would be either too detrimental to cells to promote proliferation, or prevent cell cycle arrest from occurring entirely. There is evidence to suggest that mitochondrial electron transport chain activity, but not ATP synthesis, is required to drive certain cell types through apoptosis (Jia *et al.* 1997), although it is unclear whether this mitochondrial activity can take place in non-dividing cells. However, the consistently high viability in untreated cells and similar patterns of cell death after drug treatment, irrespective of serum conditions, suggested that loss of viability was directly due to oxaliplatin treatment.

A range of oxaliplatin doses were tested to ensure that loss of viability in treated cells was apparent enough to distinguish between untreated cell lines. The larger dose range showed similar cell death across all doses, with viability that was substantially lower than in untreated cells. The dose range was narrowed down, followed by eventual selection of 125µM as the dose for oxaliplatin treatment. During other studies, researchers have previously used similar doses of oxaliplatin for treatment of colorectal cancer cell lines for assessment of ERCC1 expression (Orlandi *et al.* 2015) and to induce cell cycle arrest and apoptosis (Flis *et al.* 2009).

For UV treatment, we tested a range of UV doses to select a dose that allowed for a distinct reduction in viability compared to untreated cell lines. A number of protocols recommended UV treatment in PBS (Bruins *et al.* 2004; Stiff *et al.* 2006). We therefore compared UV treatment in PBS and SF media, to rule out the possibility of UV absorption by other molecules in tissue culture media, in the same way that UV light can be absorbed by serum and water molecules (Polet and Steinhardt, 1968). We observed comparable cell death in cells after UV treatment in PBS and SF media, confirming that our protocol for UV irradiation was appropriate for this viability assay.

3.4.3 Effect of *ERCC4* variants on viability after DNA damage

Comparison of percentage viability in wild type cell lines after both oxaliplatin and UV treatment was very similar, suggesting there was little variation between individual wild type cell lines. As a result, the decision was made to analyse the percentage viability values for cell lines carrying the same *ERCC4* variant collectively. For the purpose of this thesis, cell lines carrying the same variant will be referred to as 'cell groups'.

Percentage viability of wild type and variant cell groups showed very little apparent difference in cell death after oxaliplatin treatment, regardless of *ERCC4* variant. UV treatment in wild type and variant cell groups showed very similar results. Statistical comparison of percentage viability for each cell group following oxaliplatin treatment revealed no significant differences. However, comparison of percentage viability values by ANOVA in cell groups following UV treatment identified statistical differences between wild type cell lines and cell lines carrying Pro379Ser, and between cells carrying Pro379Ser and Glu875Gly. It is possible that this could have been an artefact of statistical analysis, due to analysis of percentage viability data from multiple biological repeat experiments.

Despite the statistically significant difference identified between three cell groups following UV treatment only, the average percentage viability values between cell groups following oxaliplatin and UV treatment remained small. This suggested that the variants did not appear to affect XPF function severely enough to impact upon viability. This could be due to the heterozygous nature of the cell lines, meaning that there may not be enough mutated protein present in variant cell lines to impair DNA repair efficiently enough, in order to see a difference in viability following DNA damage.

Previous work in our lab has shown a statistically significant decrease in viability (<20% compared with 100% in rad16^{WT} cells) after oxaliplatin treatment in the fission yeast *Schizosaccharomyces pombe* (*S. pombe*) carrying a novel stop gain in the *S. pombe* homolog of *ERCC4*, rad16. All strains carrying nonsynonymous rad16 variants, including the *S. pombe* homologs of Pro379Ser, Arg576Thr and Glu875Gly, displayed <60% viability after oxaliplatin treatment. Percentage viability in all strains after UV treatment showed a very

similar trend (West, 2013, Chapter 7). This clear pattern of reduced viability after UV and oxaliplatin treatment in *S. pombe* strains suggested that a haploid model for functional analysis of viability might be more ideal. It is possible that the variant cell lines are haploinsufficient due to the presence of one wild type allele, and that consequently the *ERCC4* variants affect function in a subtle way. One indication to support this hypothesis is the fact that their predicted effect on function using Align-GVGD was only statistically significant once all three variants were combined for statistical analysis (Table 1.11).

Lymphoblastoid cell lines were selected as a model for functional analyses on the basis that they were commercially available, relatively inexpensive and the reagents and consumables required for culturing and maintenance of these cell lines were widely available at a reasonable cost. Additionally they proliferate rapidly, with doubling times of 3-4 days, allowing growth of large enough cell numbers to perform standard cell-based assays such as immunofluorescence and ELISA. A particular advantage was the presence of their genotyping data, which facilitated selection of cells based on their *ERCC4* mutation status. A potential mechanism by which oxaliplatin is thought to exert its cytotoxic effects is via platinum accumulation in dorsal root ganglia (DRG) of treated individuals, which has been demonstrated in mice following cisplatin and oxaliplatin treatment (Dzagnidze *et al.* 2007; Ta *et al.* 2006). Therefore, with regard to investigation of PNAO *in vitro*, despite their numerous advantages, lymphoblastoid cells may not be the most ideal model organism due to their high replication rates, which are not representative of fully differentiated neuronal cells with low replication potential. Additionally, we ideally hoped to assess the functional effects of *ERCC4* variants in human cell lines; we were unable to find commercially available human neuronal cell lines for *in vitro* culture. Even if this was a possibility, there were additional concerns that the use of differentiated neuronal cell lines as a model organism would not reflect the properties of rapidly dividing cells in the body, including cells of the bone marrow, gastrointestinal tract and hair follicles, which are typically targeted by chemotherapeutic drugs. Therefore, this may not have yielded results that were representative of the effects of oxaliplatin-mediated toxicity in patients.

Viral transformation of primary cell lines involves mutation or deletion of the cell's genetic material as a consequence of integration of the viral genome

into the host cell. Typically this adversely impacts the cell's replicative cycle, resulting in stimulation of continuous proliferation and/or inhibition of apoptosis (Klein and Ernberg, 2007). Although these properties are useful for culturing cell lines *in vitro*, this suggests that virally transformed cell lines are not ideal for assessing the functional effects of genetic variants on survival, as the transformed cells are likely to demonstrate reduced apoptosis, therefore resulting in biased results. Similarly, it is possible that transformed cell lines may have a selective advantage with regard to proliferation following DNA damage compared with primary cell lines. This enhanced proliferative potential could result in apparently larger cell numbers during DNA repair, therefore skewing the amount of DNA damage that is quantified by standard assays, than would be seen using primary cell lines.

EBV transformation of primary resting B cells results in several genetic changes within cells; firstly the viral Wp promoter is activated and several protein-coding genes, including Epstein-Barr nuclear antigen-2 and –leader protein (EBNA-2 and –LP) and latent BamHI-H right reading frame-1 (BHRF1) are expressed. These proteins are critical for cellular transformation. This results in activation of a second promoter Cp, followed by subsequent expression of all six Epstein-Barr nuclear antigens (EBNA; EBNA1, 2, 3A, 3B, 3C and LP) and three latent membrane proteins LMP1, 2A and 2B (Tierney *et al.* 2015). The coordinated expression of these proteins in newly infected B cells triggers cell growth, resulting in B cell transformation. In particular the LMP proteins are known to modulate signalling pathways within cells undergoing transformation; LMP1 confers a survival advantage in EBV-infected B cells via LMP1-induced disruption of p53-mediated apoptosis and upregulation of the anti-apoptotic protein Bcl-2 (Henderson *et al.* 1991). LMP2A inhibits the activation of EBV replication via cell-surface-mediated signal transduction, which contributes to the latent phase of EBV infection of B cells (Miller *et al.* 1994). This series of latent gene expression events is known as Latency III and results in the establishment of lymphoblastoid cell lines with infinite growth potential (Tierney *et al*. 2000). Therefore, it is possible that the effects of EBV transformation on our *ERCC4* variant cell lines have resulted in enhanced proliferation, which could have increased the percentage viability at each time

point, therefore resulting in unreliable viability measurements throughout the assay.

The viability assays were repeated four times for each variant and treatment type, with very similar results acquired each time, suggesting that the assay was very consistent. However, it is possible that the trypan blue exclusion method of quantifying viability may not be the most sensitive or reliable method. Trypan blue exclusion can only detect cells that are no longer viable due to influx of the stain into cells via their damaged membranes. However, this offers no information on whether the cells are apoptotic due to DNA damage or whether they have undergone necrosis or autophagy, since multiple cell death pathways may be in progress simultaneously (Lenardo *et al.* 2009). Therefore, it is possible that a more sensitive method of viability measurement may be required to elucidate any true differences in viability between wild type and variant cell lines.

3.4.4 Alternative assays for viability studies

Alternative assays may be useful for viability studies in mammalian cell lines; one potential alternative is the enzyme-linked immunosorbent assay (ELISA) for detection of caspases, a group of cysteine-proteases that hydrolyse target sequences and become activated during apoptosis (Chen and Wang, 2002). This assay allows absolute quantitation of fluorescent substrates that indicate caspase activation, which can be quantified by changes in optical density (Lenardo *et al.* 2009). This is considered a relatively specific assay, since caspase activation is considered a hallmark of apoptosis, and is not thought to stimulate necrosis (Elmore, 2007).

Similar to caspase ELISA, a lactate dehydrogenase (LDH)-cytotoxicity assay can be used to quantify LDH, a soluble cytosolic enzyme that leaks from cells upon death due to increased cell membrane permeability. The assay is also quantified by changes in optical density similar to ELISA, which quantifies amounts of LDH relative to the negative and positive controls. For this reason, it is considered less sensitive than ELISA, which can quantify absolute concentrations of proteins, but more sensitive than trypan blue exclusion. However, LDH-cytotoxicity assays lack specificity, since LDH is released from cells upon death by necrosis as well as apoptosis (Lenardo *et al.* 2009).

Flow cytometric methods for viability measurement require exclusion of non-viable cells by propidium iodide (PI) staining, which detects loss of membrane integrity in the same way as trypan blue. PI stain penetrates non-viable cells and causes a strong emission at 620nm, which can be detected in the FL-3 channel, allowing quantitation of survival (Lenardo *et al.* 2009). This is a rapid and more sensitive method of cell viability detection than trypan blue exclusion, although it requires expensive, specialist equipment. Additionally its specificity for apoptotic cells is low, since PI stain will penetrate cells with semi-permeable membranes resulting from other cell death pathways (Ross *et al.* 1989).

A major limitation of viability measurement by trypan blue exclusion is its inability to distinguish cells that have lost their membrane integrity due to induction of cell death pathways other than apoptosis. A slightly more sophisticated method is the standard comet assay, which is typically used for measurement of DNA strand breaks by gel electrophoresis of single cells embedded in agarose on microscope slides (Fairbairn *et al.* 1995). It is a rapid, relatively inexpensive assay that provides greater sensitivity than that associated with more standard forms of gel electrophoresis for DNA separation (in the region of 10^9 Da compared with kDa respectively) (Collins *et al.* 1997; Collins, 2004). However, one limitation experienced previously was a tendency towards false positive results, owing to the standard assay's inability to discriminate between apoptotic and necrotic cells (Collins *et al.* 1997). More recently, a modified version of the comet assay was developed that can distinguish viable, apoptotic and necrotic single cells. This modified assay incorporates simultaneous staining with ethidium-homodimer exclusion dye, calcein-blue-AM and Annexin-V, for assessment of membrane integrity, cellular esterase activity and translocation of phosphadidyl-serine respectively, to quantify cell numbers at different stages of cell death. This appeared to resolve the issue of false positive results seen previously with the standard comet assay (Morley *et al.* 2006).

Chapter Four – Localisation of ERCC1-XPF in wild type and variant *ERCC4* cell lines after DNA damage

4.1 Introduction

Several studies in mammalian cell lines have suggested that ERCC1 and XPF are unstable in the absence of their protein partner. Research by Yagi *et al.* (1997) using fibroblasts isolated from XP-F patients revealed low levels of ERCC1 and the ERCC1-XPF complex overall. In addition, the same has been identified for XPF, identifying low XPF levels in the absence of ERCC1 (Gaillard and Wood, 2001). Other research has shown that ERCC1 and XPF mutant cell extracts do not complement one another when mixed together *in vitro*, suggesting that unstable unconjugated proteins most likely mis-fold and become marked for degradation (Biggerstaff *et al.* 1993; Van Vuuren *et al.* 1993). For our assay, each protein was probed for using separate antibodies and any ERCC1 or XPF detected by immunofluorescence was therefore assumed to be conjugated protein as part of the ERCC1-XPF repair complex.

Ahmad *et al.* (2010) previously showed mislocalisation of both ERCC1 and XPF to the cytoplasm in XPF mutant cell lines, including cells carrying Pro379Ser, without treatment. This was in comparison to exclusively nuclear ERCC1 and XPF localisation in wild type fibroblast cells. We hoped to further this work by investigating the effect of DNA damage on localisation of the repair complex in human Epstein Barr virus (EBV)-transformed lymphoblastoid cell lines and various strains of *S. pombe*. Human EBV-transformed lymphoblastoid cells are spherical and consist of a large, dense nucleus (approx. 90% of cell volume) and a narrow cytoplasm. Typically these cells range from 10-30μm in diameter (Thompson *et al.* 1984; Ryan *et al.* 2006). In contrast, *S. pombe* cells are cylindrical in shape due to growth by extension of cell ends, with a small spherical nucleus at the centre of each cell and a large cytoplasm. Typically cells range in length from 4-10μm; cell division usually occurs once cells reach approx. 9.5μm (Mitchison and Nurse, 1985).

Here, we sought to identify the functional consequences of three *ERCC4* variants on cellular localisation of ERCC1-XPF and Rad16 before and after damage with oxaliplatin treatment by indirect immunofluorescence.

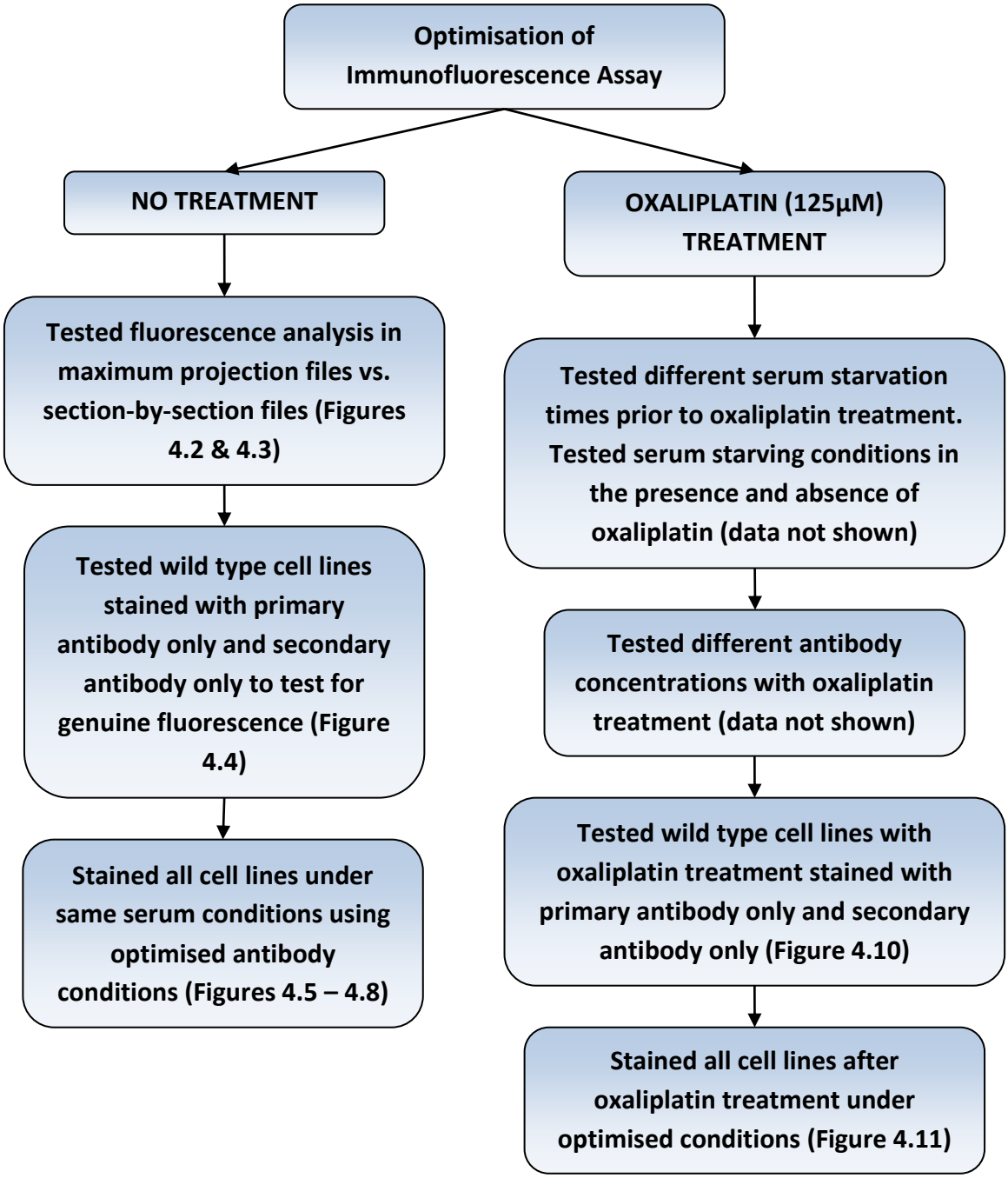


Figure 4.1 Schematic diagram of optimisation of immunofluorescence assay in human cell lines with and without oxaliplatin treatment. Wild type cells were imaged as a combination of maximum projection and section-by-section files, and fluorescence was quantified to identify the best approach. Different serum starving times were tested and cells were stained with and without oxaliplatin to rule out the drug as an inhibitor of fluorescent activity. Wild type cells were stained with primary antibody only and secondary antibody only as a negative control. This identified optimum antibody conditions for immunofluorescence of all cell lines without and with oxaliplatin treatment.

4.2 Materials and Methods

4.2.1 Immunofluorescence of cell lines without treatment

A schematic diagram summarising the stages of optimisation is shown in Figure 4.1. During optimisation, wild type cells were adhered to microscope slides using various adhesion matrices and by cytospinning, prior to staining. Wild type cells were imaged as maximum projection (MP) and section-by-section files and total fluorescence quantified for each file type were compared. Different concentrations of permeabilising agent and antibodies were tested. Wild type cells were stained with primary antibodies only and secondary antibodies only as negative controls.

Optimised conditions are described in section 2.5.3.3; antibody species and concentrations are shown in Table 4.1a. Cell lines were imaged using a Leica SP5 confocal laser scanning microscope using optimised imaging parameters (Table 4.1b).

4.2.2 Immunofluorescence of cell lines with oxaliplatin treatment

During optimisation, wild type cells were treated with and without oxaliplatin in normal and serum-free (SF) media over 16 hours. Higher concentrations of secondary antibodies were tested and both primary and secondary antibodies for XPF detection were replaced. Negative control wild type cells were stained with primary antibodies and secondary antibodies only after oxaliplatin treatment.

Optimised conditions were as follows: cells were suspended in SF media and treated with oxaliplatin (125μM) for 1 hour without prior serum starvation. Cells were resuspended in normal media containing FBS and antibiotics for two hours, prior to cytospinning, fixation, permeabilising and blocking as outlined in section 2.5.3.3. Cells were incubated with antibodies and DAPI nuclear stain in the same way as untreated cells, with the exception of different antibodies for XPF detection after oxaliplatin treatment (antibody names and concentrations are shown in Table 4.1a). Cells were mounted and imaged as in section 4.2.1 (imaging parameters in Table 4.1b).

4.2.3 Quantifying localisation in cell lines

Cellular fluorescence was quantified using LAS-AF Lite software by manually encircling the nucleus (region of interest 1; ROI1) and entire cell

A.

Antibody	*Primary/ Secondary*	*Species*	*Concentration*	*Dilution*	*Staining conditions*
ERCC1	Primary	Rabbit polyclonal	4µg/mL	1:50	Overnight at 4°C Diluted in 1% BSA in PBS
Alexa 488	Secondary	Goat anti-rabbit IgG	10µg/mL	1:200	1 hour at RT Diluted in 1% BSA in PBS
XPF (3F2/3)	Primary	Mouse monoclonal	11µg/mL	1:100	Overnight at 4°C
XPF (M-16)*		Goat polyclonal	11µg/ml	1:100	Diluted in 1% BSA in PBS
Alexa 546	Secondary	Goat anti-mouse IgG	1µg/mL	1:2000	1 hour at RT Diluted in 1%
Alexa 546*		Donkey anti-goat IgG	10µ/ml	1:200	BSA in PBS

B.

Fluorescent stain	*DAPI*	*ERCC1 - Alexa 488*	*XPF - Alexa 546*
Laser	405 Diode	Argon	HeNe 543
Overall laser power	30%	30%	30%
Individual laser powers	10% 30%*	20%	90%
Format	512x512	512x512	512x512
Speed	8000 Hz	8000 Hz	8000 Hz
Zoom factor	1.7	1.7	1.7
Image size	144.72 x 144.72µm	144.72 x 144.72µm	144.72 x 144.72µm
Pixel size	283.22 x 283.22 x 42.00nm	283.22 x 283.22 x 42.00nm	283.22 x 283.22 x 42.00nm
Section thickness	0.772µm	0.772µm	0.772µm
Line/Frame average	4/4	4/4	4/4
Rotation	0.00	0.00	0.00
Voltage of channel	570	530	780
Section thickness of Z-stacks	0.49µm	0.49µm	0.49µm

Table 4.1 Summary of antibody concentrations and imaging parameters for immunofluorescence using human cell lines. A) Summary of staining conditions after optimisation of the immunofluorescence assay showing conditions with* and without oxaliplatin (125µM) treatment. B) Summary of imaging parameters used to image cells on the Leica SP5 confocal microscope using the 63x oil objective, showing conditions with* and without oxaliplatin (125µM) treatment.

	Number of pixels		***Fluorescent count***	
Intensity	**ROI1**	**ROI2**	**ROI1**	**ROI2**
0	a^1	a^2	$a^1 \times 0$	$a^2 \times 0$
1	b^1	b^2	$b^1 \times 1$	$b^2 \times 1$
2	c^1	c^2	$c^1 \times 2$	$c^2 \times 2$
⬇	⬇	⬇	⬇	⬇
⬇	⬇	⬇	⬇	⬇
253	x^1	x^2	$x^1 \times 253$	$x^2 \times 253$
254	y^1	y^2	$y^1 \times 254$	$y^2 \times 254$
255	z^1	z^2	$z^1 \times 255$	$z^2 \times 255$
TOTAL FLUORESCENT COUNT			**Sum of ROI1 values**	**Sum of ROI2 values**

Table 4.2 Schematic of template table for calculation of total fluorescent count. Fluorescent counts and total fluorescent count of ERCC1 and XPF in each cell during fluorescence quantification were calculated using this table. Imaging files were acquired in an 8-bit format with a maximum of 256 different intensities (minimum intensity of 0; maximum intensity of 255). The following equations were used:

Fluorescent Count = Pixel Number x Pixel Intensity

Total fluorescent count (F) = Sum of all Fluorescent counts

perimeter (ROI2). For each pixel intensity ranging from 0 to 255, fluorescent counts were calculated for ROI1 and ROI2 independently as outlined in Table 4.2. During optimisation, imaging of the same wild type cells was performed as MP and as section-by-section files, where MP files were an overlay of all individual sections throughout the depth of the cell (Figure 4.2). This involved manually encircling all ROI1 and ROI2 regions of cells for every individual section in a section-by-section file, compared with only one manual encircling per MP file. For section-by-section files, the total fluorescent counts of each section were added together.

Optimised conditions involved imaging all cell lines as MP files, which required only a single fluorescence calculation per cell. Total fluorescent count was calculated for ROI1 and ROI2 as the sum of every fluorescent count across all intensities (0-255). Total fluorescent count in ROI1 and ROI2 for both ERCC1 and XPF was calculated for 50 and 20 cells per cell line for untreated and oxaliplatin treated cells respectively. These values were converted to percentage localisation values for nuclear and cytoplasmic ERCC1 and XPF, before averaging.

Relative-fold increase (RFI) of nuclear ERCC1 and XPF localisation was calculated by the following equation:

$$\text{RFI} = \frac{\%\ \text{localisation after oxaliplatin treatment}}{\%\ \text{localisation without treatment}}$$

RFI values were calculated in 20 cells per cell line for both ERCC1 and XPF localisation, before averaging.

4.2.4 Statistical analysis

Statistical analysis was performed in SPSS v.20.0 using a two-way ANOVA, with mutation status (wild type, Pro379Ser, Arg576Thr and Glu875Gly) and treatment (no treatment and oxaliplatin treatment) as independent variables. Percentage localisation of ERCC1 and XPF was used as the dependent variable. Percentage localisation values for ERCC1 and XPF for each cell line with and without oxaliplatin treatment were averaged and the mean values for cell lines carrying the same *ERCC4* variant were compared by ANOVA. This allowed comparison of localisation for cell groups. Statistical comparison of RFI values for nuclear ERCC1 and XPF following oxaliplatin

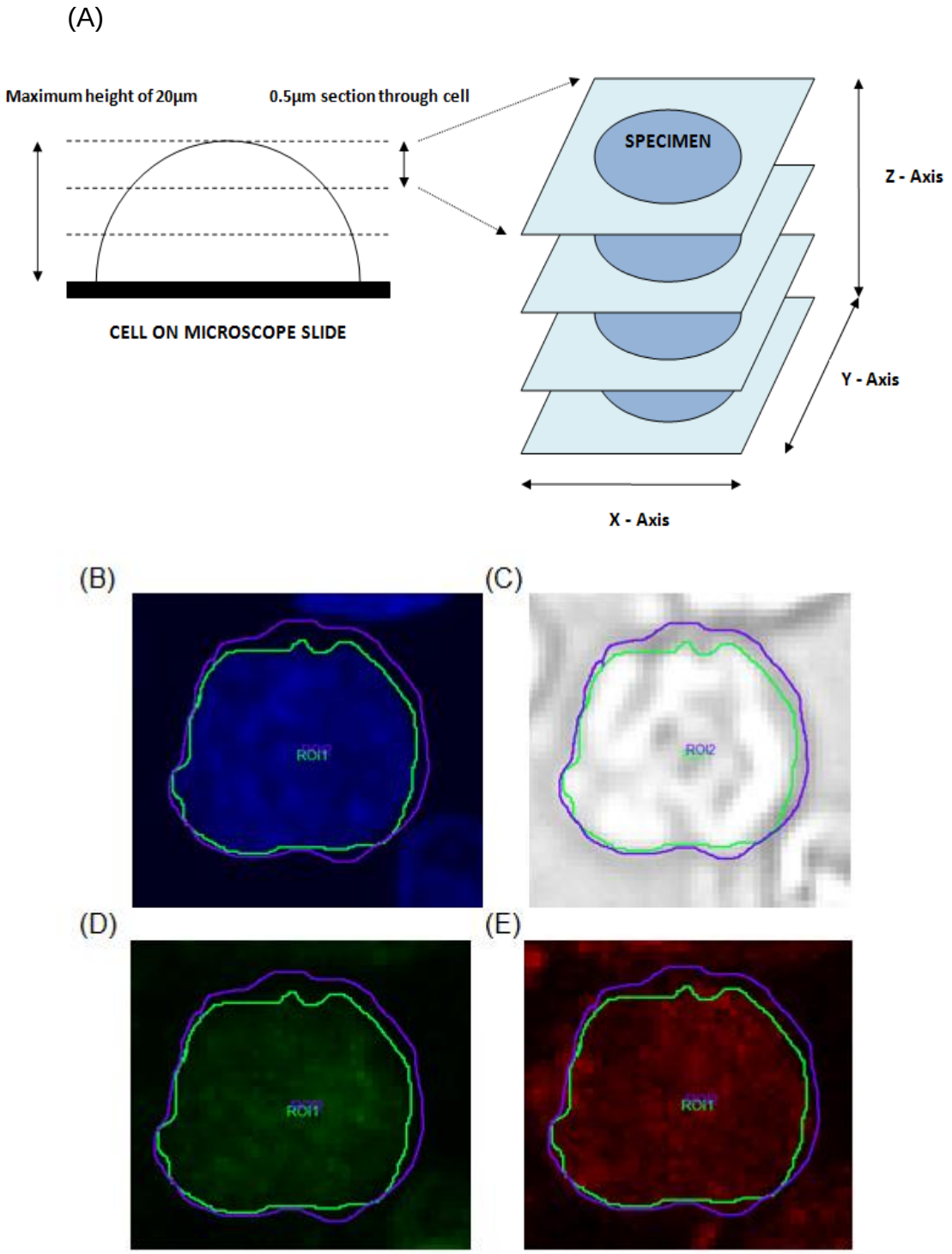


Figure 4.2 Schematic of Z-stack imaging carried out by confocal laser scanning microscope. Cells were imaged in 0.5µm sections throughout the depth of the cell (as shown in A) and merged to form an overlaid image of all sections, called a maximum projection image. Maximum projection files were acquired using (B) the 405nm laser for DAPI nuclear staining, (C) phase contrast to show entire cell outline, (D) the 488nm laser for Alexa 488 staining to detect ERCC1 and (E) the 543nm laser for Alexa 546 staining to detect XPF. The regions of interest (ROI) 1 and 2 represent the nucleus and perimeter of the cell respectively.

treatment for each cell group was performed using a one-way ANOVA, using the mean RFI values for each cell line carrying the same *ERCC4* variant. Where a statistically significant interaction was identified, a Tukey post-test was performed to identify statistical significance between specific cell groups.

4.2.5 Construction of *S. pombe* strains carrying *rad16* variants

All *S. pombe* strains were constructed using site directed mutagenesis to induce specific mutations in *rad16* into each strain. Firstly, a *rad16* deletion base strain was constructed and transformed into a wild type strain of *S. pombe* to produce a knockout strain. This included incorporation of lox sites to allow future homologous incorporation of nonsynonymous *rad16* variants by colony PCR and cloning. The deletion base strain was distinguished from wild type cells via enrichment of UV sensitivity. The truncating Ser585X variant and remaining nonsynonymous variants were then incorporated by transfection of vectors. All vectors used for cloning contained an N-terminal (5') His tag to allow probing via immunofluorescence. All *rad16* variants are listed with their respective human homologs in Table 4.3. All strains were constructed and kindly donated by Dr Oliver Fleck (Bangor University). Full details of the *S. pombe* strain construction process can be found in Dr Hannah West's PhD thesis (West, 2013, Chapter 6).

4.2.6 Immunofluorescence of *S. pombe* without treatment

During optimisation, wild type and variant *S. pombe* cells were treated with different concentrations of zymolyase (0.1-1mg/ml) to achieve cell wall digestion, and were permeabilised using varying concentrations of Triton X-100 (2-5%). Different concentrations of primary His tag antibody (5-30µg/ml) were also tested to achieve optimum staining conditions.

Optimised conditions are outlined in section 2.5.4.6; antibody species and concentrations are shown in Table 4.4a. Cell lines were imaged using a Leica SP5 confocal laser scanning microscope using optimised imaging parameters (Table 4.4b).

4.2.7 Immunofluorescence of *S. pombe* with oxaliplatin treatment

During optimisation, different concentrations of primary His tag antibody (5-30µg/ml) and secondary Alexa 488 antibody (2-20µg/ml) were tested. During

Strain name	***Rad16 variant***	***Human homolog (ERCC4 variant)***
rad16^{WT}	Wild type	Wild type
rad16^{Pro361Ser}	Pro361Ser	Pro379Ser
rad16^{Arg399Gln}	Arg399Gln	Arg415Gln
rad16^{Arg548Thr}	Arg548Thr	Arg576Thr
rad16^{Ser585X}	Ser585X	Ser613X
rad16^{Glu844Gly}	Glu844Gly	Glu875Gly

Table 4.3 Summary of *S. pombe* strains with their *rad16* variants and respective human homologs.

A.

Antibody	***Primary/ Secondary***	***Species***	***Concentration***	***Dilution***	***Staining conditions***
His tag	Primary	Rabbit polyclonal	10µg/mL	1:100	Overnight at RT Diluted in PEMBALG buffer
Alexa 488	Secondary	Goat anti-rabbit IgG	5µg/mL	1:100	4 hours at RT Diluted in PEMBALG buffer

B.

Fluorescent stain	***DAPI***	***Rad16 - Alexa 488***
Laser	405 Diode	Argon
Overall laser power	30%	30%
Individual laser powers	10% 30%*	20%
Format	512x512	512x512
Speed	8000 Hz	8000 Hz
Zoom factor	3.0	3.0
Image size	82.01 x 82.01µm	82.01 x 82.01µm
Pixel size	160.49 x 160.49 x 49.4nm	160.49 x 160.49 x 49.4nm
Section thickness	0.772µm	0.772µm
Line/Frame average	4/4	4/4
Rotation	0.00	0.00
Voltage of channel	611.0; 660.0*	800.0; 1100.0*
Section thickness of Z-stacks	0.49µm	0.49µm

Table 4.4 Summary of antibody concentrations and imaging parameters for immunofluorescence using *S. pombe* strains. A) Summary of staining conditions after optimisation of the immunofluorescence assay showing conditions with* and without oxaliplatin (1mM) treatment. B) Summary of imaging parameters used to image cells on the Leica SP5 confocal microscope using the 63x oil objective, showing conditions with* and without oxaliplatin (1mM) treatment.

confocal imaging, a higher voltage was tested to achieve higher intensity images whilst maintaining low background fluorescence.

Optimised conditions were as follows: cells were treated with oxaliplatin (1mM) for 18 hours in YEL as outlined in section 2.5.4.5. Fixation, cell wall digestion, membrane permeabilisation and blocking were carried out as described in section 2.5.4.6. Cells were incubated with antibodies and DAPI nuclear stain in the same way as untreated cells, and mounted and imaged as in 4.2.5 (imaging parameters in Table 4.4b).

4.2.8 Quantifying localisation of rad16 in *S. pombe*

Nuclear and cytoplasmic rad16 was quantified using the same method as that used for ERCC1/XPF localisation in cell lines (section 4.2.3), using 20 cells per *S. pombe* strain. RFI values were calculated as outlined in section 4.2.3.

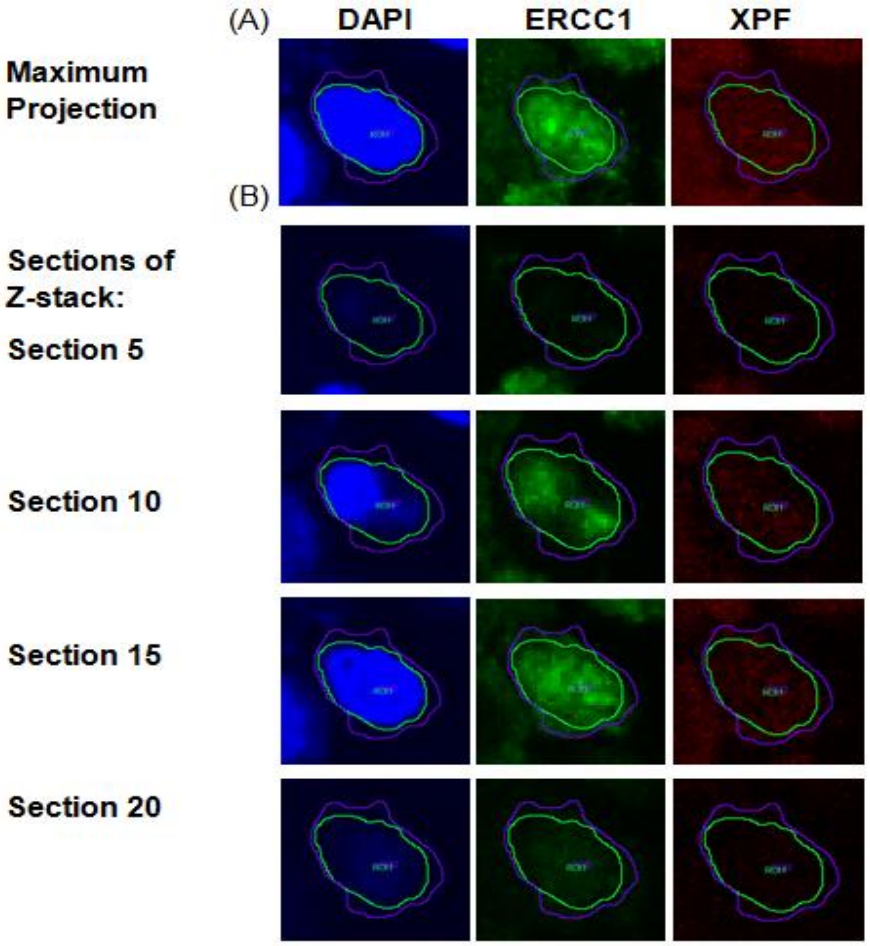
4.2.9 Statistical analysis of rad16 localisation

Statistical analysis of percentage localisation values was performed in SPSS v20.0 using a two-way ANOVA, with mutation status ($rad16^{WT}$, $rad16^{Pro361Ser}$, $rad16^{Arg399Gln}$, $rad16^{Arg548Thr}$, $rad16^{Ser585X}$ and $rad16^{Glu844Gly}$) and treatment type (no treatment and oxaliplatin treatment) as independent variables. Percentage nuclear localisation of rad16 was used as the dependent variable. Percentage localisation values (n=20) per strain with and without treatment were given as input for comparison of each strain by ANOVA; however, despite this, individual localisation values for the same *S. pombe* strain cannot be considered as different experiments since all localisation values were acquired during the same experiment. RFI values (n=20) per strain were given as input for a one-way ANOVA to compare RFI between variant strains; similarly these values were acquired during the same experiment and therefore cannot be considered as different experiments. Where a statistically significant interaction was identified, a Tukey post-test was performed to identify statistical significance between specific cell groups.

4.3 Results

4.3.1 Percentage localisation of ERCC1 and XPF without treatment

During optimisation, wild type cells were seeded onto coverslips coated with poly-L-lysine, fibronectin and laminin prior to fixation, permeabilisation and



(C)

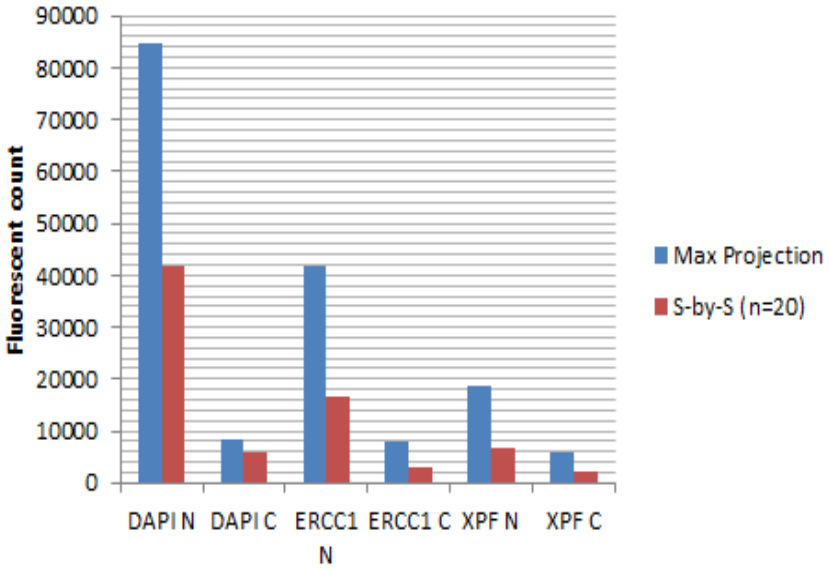


Figure 4.3 Comparison of total fluorescence in cells quantified as maximum projection vs. section-by-section files. Immunofluorescent staining of wild type cells were imaged as a maximum projection file (A) and as a Z stack comprised of 0.5µm sections (B). Fluorescent counts in DAPI, ERCC1 and XPF channels were quantified as maximum projection and section-by-section (S-by-S) files for comparison and are plotted in (C). Total fluorescence of twenty cells stained and imaged during the same experiment was quantified.

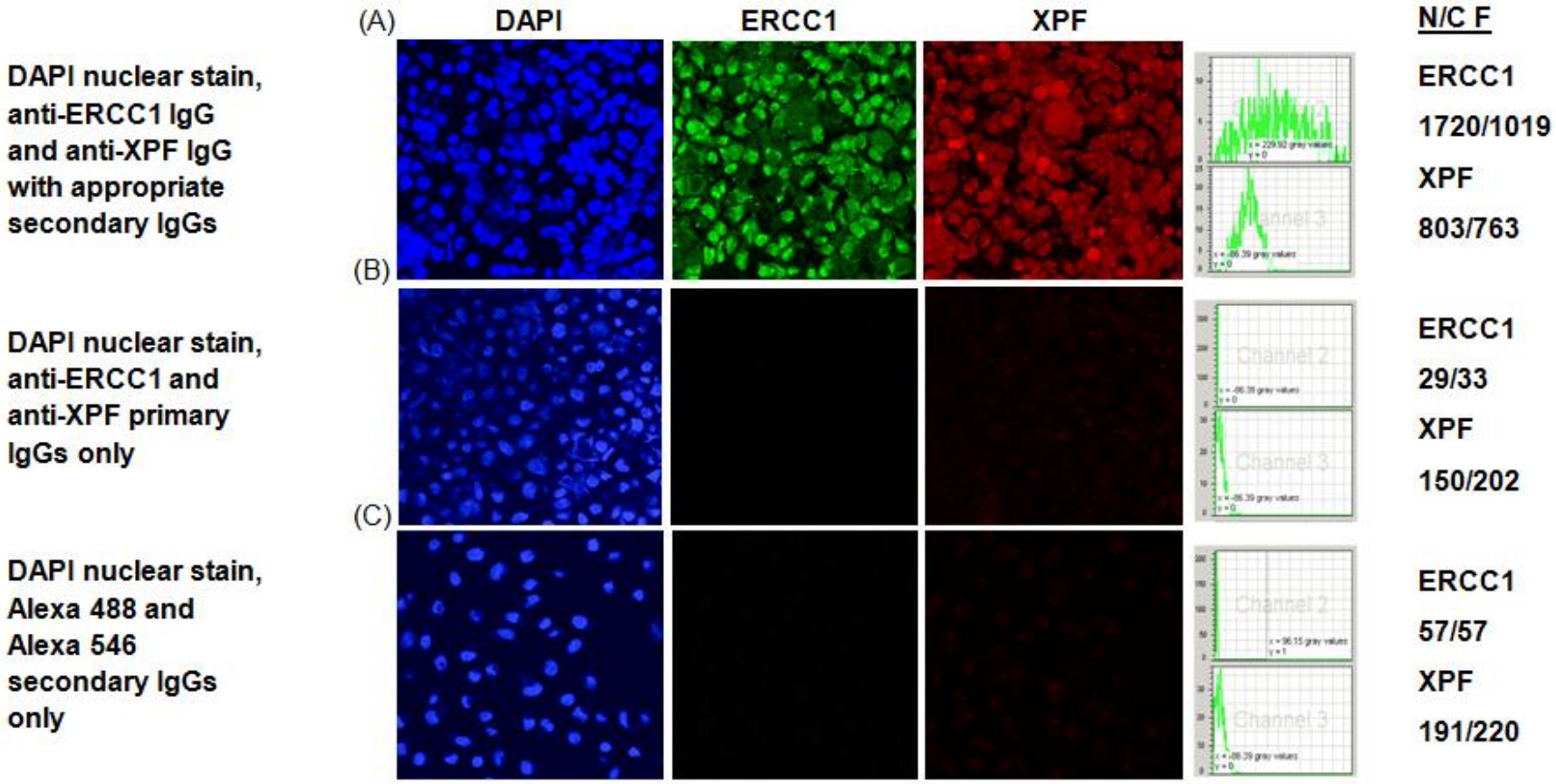


Figure 4.4 Negative controls in wild type cell lines without treatment. Wild type cells (no starvation) without drug treatment stained with (A) DAPI nuclear stain with anti-ERCC1 and anti-XPF IgGs with corresponding Alexa 488 and Alexa 546 IgGs, (B) DAPI nuclear stain with anti-ERCC1 and anti-XPF primary IgGs only and (C) DAPI nuclear stain with Alexa 488 and Alexa 564 secondary IgGs only. Intensity peaks (right) indicate intensity of ERCC1 (above) and XPF (below) staining respectively and nuclear/cytoplasmic fluorescent count (N/C F) is shown on the far right. Fluorescent counts are the average of fifty cells per cell line (150 cells per cell group) stained and imaged as part of the same biological repeat experiment.

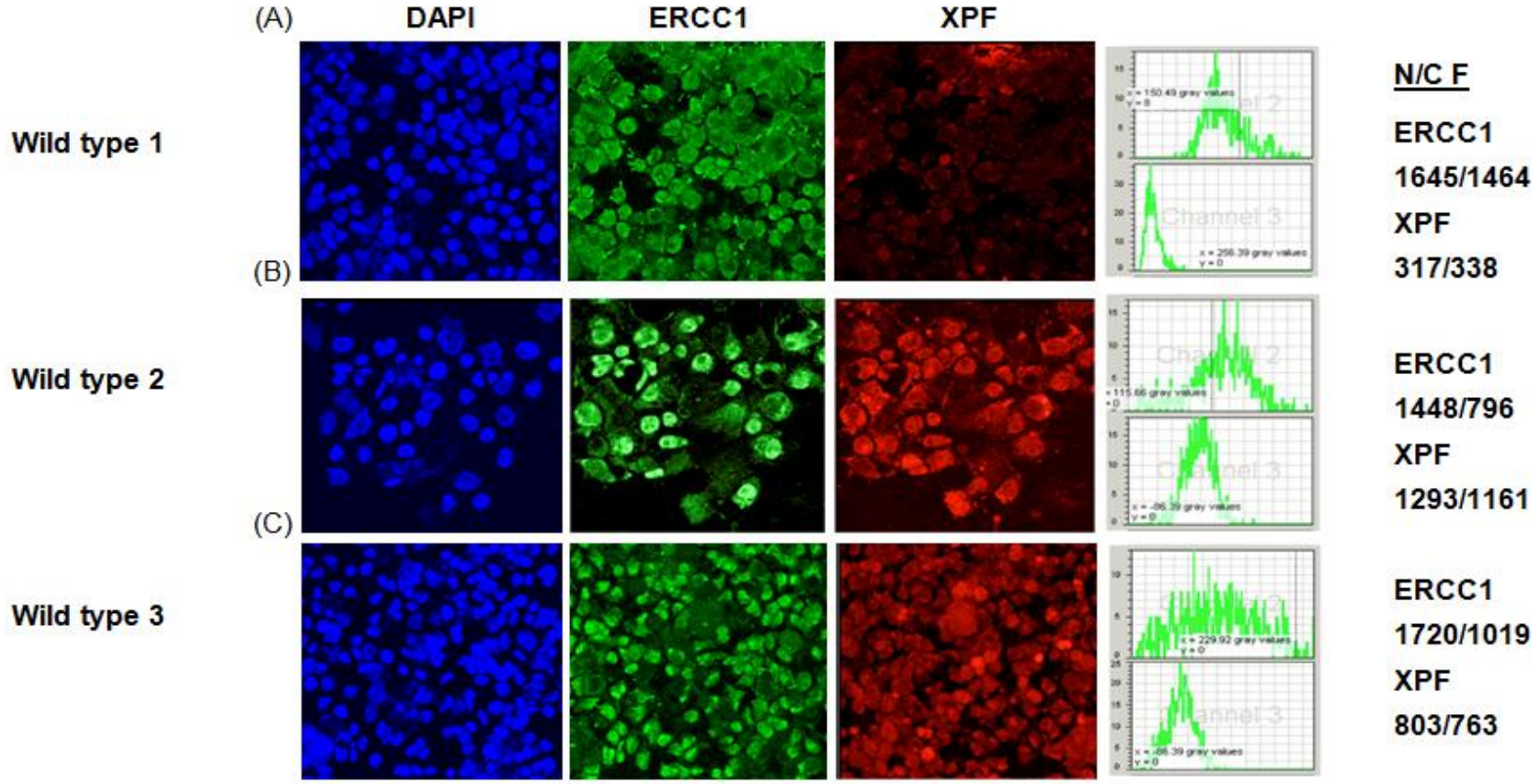


Figure 4.5 Immunofluorescence of wild type cell lines without treatment. Wild type cell lines for *ERCC4* without drug treatment stained with DAPI nuclear stain, anti-ERCC1 IgG with corresponding Alexa 488 IgG and anti-XPF IgG with corresponding Alexa 546 IgG, showing (A) wild type 1, (B) wild type 2 and (C) wild type 3 cell lines (HRC Collection). Intensity peaks (right) indicate intensity of ERCC1 (above) and XPF (below) staining respectively and nuclear/cytoplasmic fluorescent count (N/C F) is shown on the far right. Fluorescent counts are the average of fifty cells per cell line (150 cells per cell group) stained and imaged as part of the same biological repeat experiment.

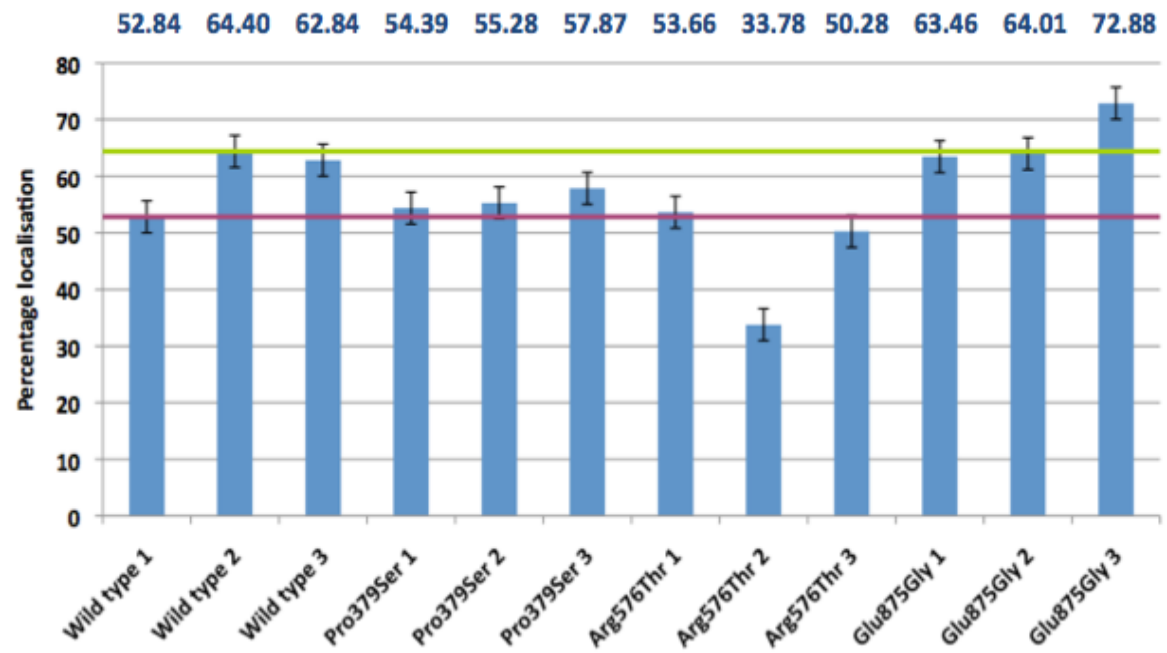
blocking. Microscopic examination between wash steps revealed that our cell lines did not successfully adhere to coverslips (data not shown).

Cells were cytospun onto glass slides and fixed, permeabilised and blocked according to 4.2.1. Different concentrations of Triton-X100 for permeabilisation were tested (data not shown). Cells were imaged as MP files and as 0.5µm section-by-section files (Figure 4.3a and b). Nuclear and cytoplasmic fluorescent counts in DAPI, ERCC1 and XPF channels were compared for MP and section-by-section files, which indicated higher fluorescence values for MP files. Although fluorescent counts for MP and individual section files differed considerably, there was an overall trend showing that fluorescence counted across individual sections amounted to approximately half of the total fluorescence in MP files of the same region and antibody (Figure 4.3c).

Wild type cells were stained with primary and secondary antibodies for ERCC1 and XPF, and with primary and secondary antibodies only as negative controls (Figure 4.4). This identified negligible levels of fluorescence in primary only and secondary only stained cells.

Wild type cell lines fully stained with primary and secondary antibodies without treatment are shown in Figure 4.5. Quantification of MP fluorescence in wild type cell lines without treatment showed reasonable variation in average nuclear and cytoplasmic ERCC1 and XPF percentage localisation (Figure 4.6). This apparent variation between individual wild type cell lines meant that it was important to consider percentage localisation of cell lines carrying the same *ERCC4* variant collectively. Cell lines carrying Pro379Ser, Arg576Thr and Glu875 fully stained with primary and secondary antibodies without treatment are shown in Figures 4.7, 4.8 and 4.9 respectively. Average percentage localisation of nuclear ERCC1 in cell lines carrying Pro379Ser and Arg576Thr showed a trend towards the lower end of the wild type localisation range, with Arg576Thr showing the greatest localisation to the cytoplasm, whereas cell lines carrying Glu875Gly localised towards the upper end of the wild type range (Figure 4.6a). Average localisation of nuclear XPF in all cell lines showed a very similar pattern, with slightly lower percentage values (Figure 4.6b).

(A)



(B)

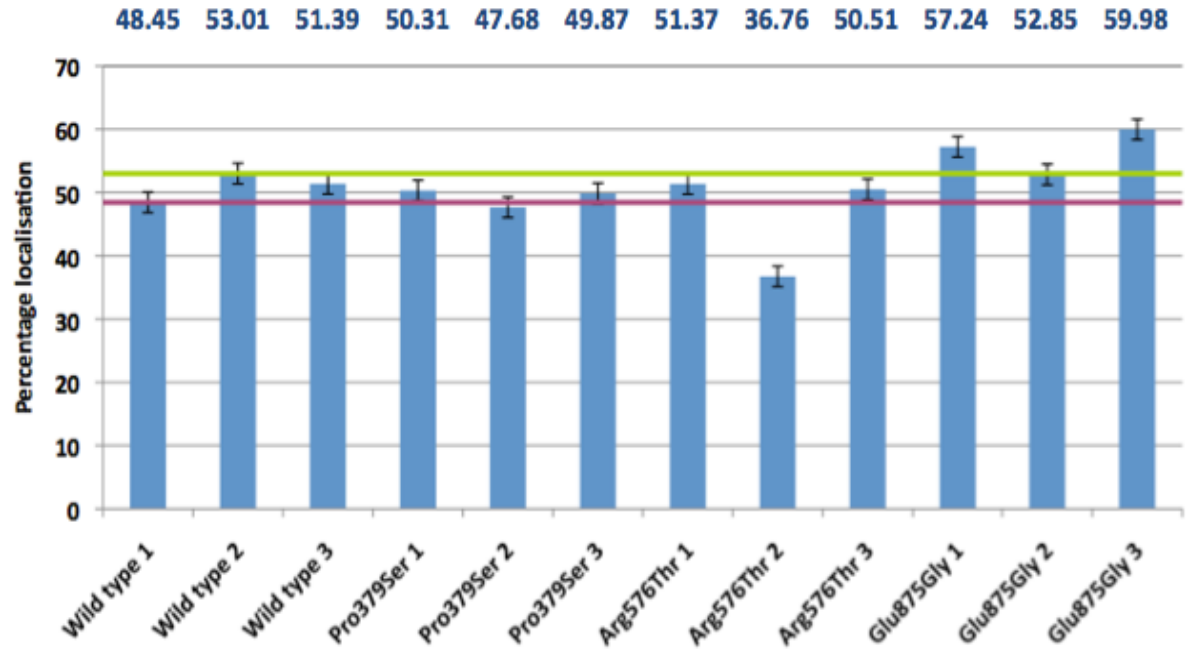


Figure 4.6 Percentage localisation of nuclear ERCC1 and XPF without treatment. Average percentage localisation of nuclear (A) ERCC1 and (B) XPF across all cell lines without treatment (n=50), with standard error bars. Horizontal lines indicate the maximum (green) and minimum (red) percentage localisation in wild type cell lines. Nuclear percentage localisation values for each cell line are indicated in blue on each plot. Percentage localisation values are the average of fifty cells per cell line stained and imaged as part of the same biological repeat experiment.

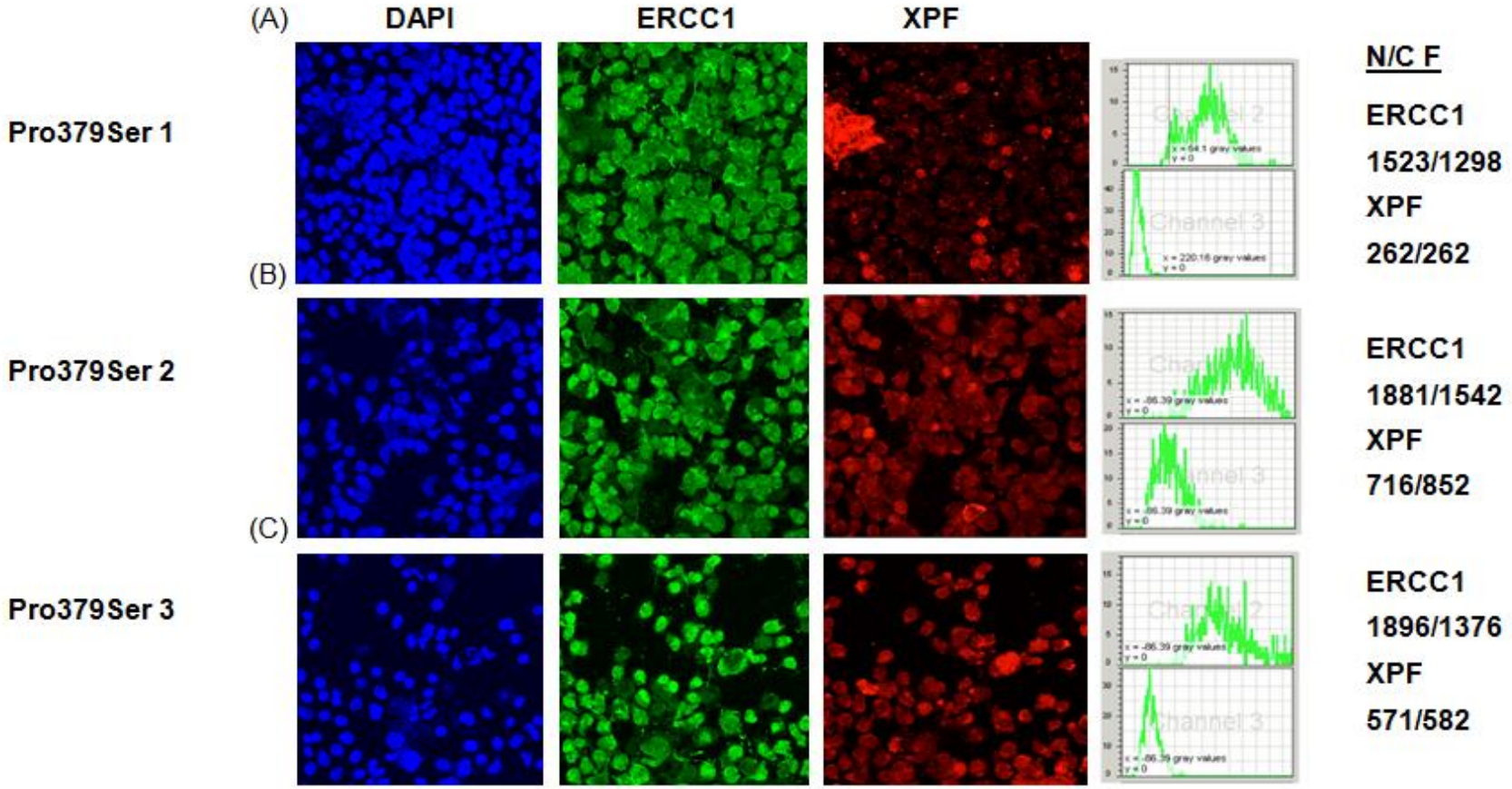


Figure 4.7 Immunofluorescence of cell lines carrying Pro379Ser without treatment. Cell lines carrying Pro379Ser without drug treatment stained with DAPI nuclear stain, anti-ERCC1 IgG with corresponding Alexa 488 IgG and anti-XPF IgG with corresponding Alexa 546 IgG, showing (A) Pro379Ser 1, (B) Pro379Ser 2 and (C) Pro379Ser 3 cell lines (HRC Collection). Intensity peaks (right) indicate intensity of ERCC1 (above) and XPF (below) staining respectively and nuclear/cytoplasmic fluorescent count (N/C F) is shown on the far right. Fluorescent counts are the average of fifty cells per cell line (150 cells per cell group) stained and imaged as part of the same biological repeat experiment.

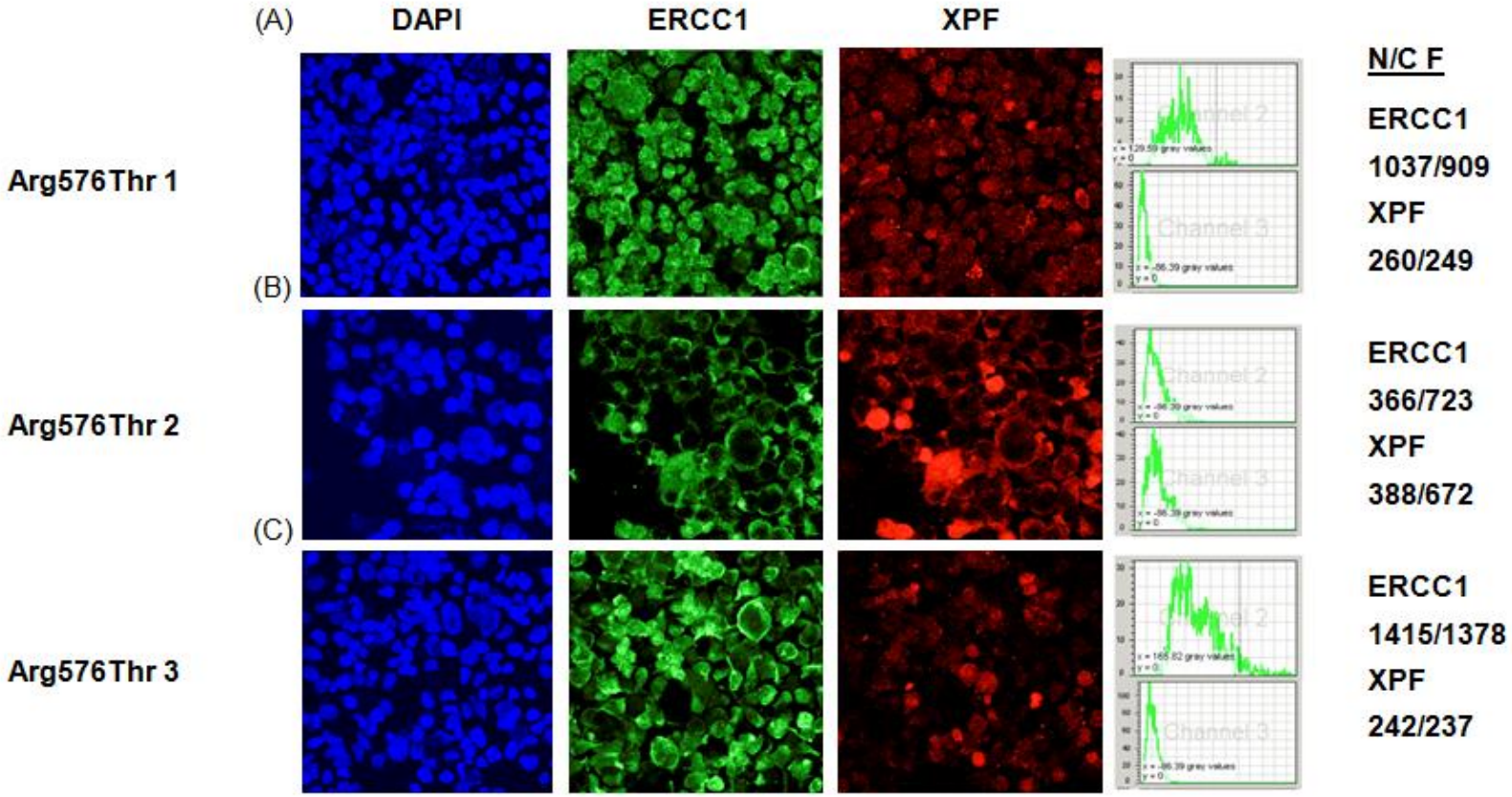


Figure 4.8 Immunofluorescence of cell lines carrying Arg576Thr without treatment. Cell lines carrying Arg576Thr without drug treatment stained with DAPI nuclear stain, anti-ERCC1 IgG with corresponding Alexa 488 IgG and anti-XPF IgG with corresponding Alexa 546 IgG, showing (A) Arg576Thr 1, (B) Arg576Thr 2 and (C) Arg576Thr 3 cell lines (HRC Collection). Intensity peaks (right) indicate intensity of ERCC1 (above) and XPF (below) staining respectively and nuclear/cytoplasmic fluorescent count (N/C F) is shown on the far right. Fluorescent counts are the average of fifty cells per cell line (150 cells per cell group) stained and imaged as part of the same biological repeat experiment.

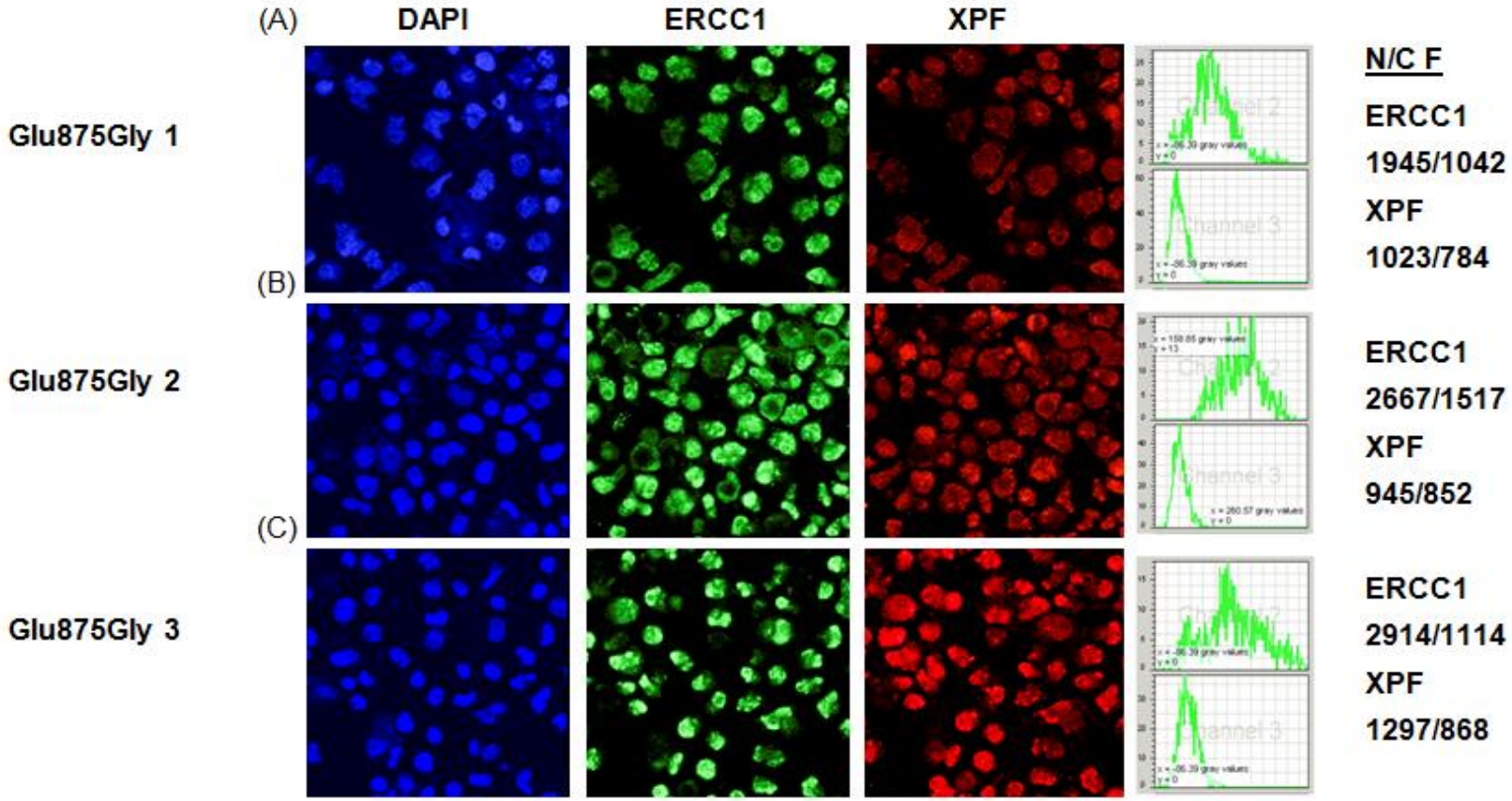


Figure 4.9 Immunofluorescence of cell lines carrying Glu875Gly without treatment. Cell lines carrying Glu875Gly without drug treatment stained with DAPI nuclear stain, anti-ERCC1 IgG with corresponding Alexa 488 IgG and anti-XPF IgG with corresponding Alexa 546 IgG, showing (A) Glu875Gly 1, (B) Glu875Gly 2 and (C) Glu875Gly 3 cell lines (HRC Collection). Intensity peaks (right) indicate intensity of ERCC1 (above) and XPF (below) staining respectively and nuclear/cytoplasmic fluorescent count (N/C F) is shown on the far right. Fluorescent counts are the average of fifty cells per cell line (150 cells per cell group) stained and imaged as part of the same biological repeat experiment.

Comparison of nuclear ERCC1 and XPF localisation between cell groups was performed using a two-way ANOVA; therefore the results of statistical analysis are outlined in section 4.3.2.

4.3.2 Percentage localisation of ERCC1 and XPF with oxaliplatin treatment

During optimisation, we tested the effect of serum starving duration (0-16 hrs) prior to oxaliplatin treatment on staining intensity in wild type cell lines. Fluorescent count was significantly higher in the presence of serum. Removal of serum resulted in a reduction in fluorescent count of approximately 80%, with similar low levels of fluorescence regardless of duration of serum starvation (data not shown).

All wild type cell lines were subjected to different serum starving conditions with and without oxaliplatin (125µM) treatment. Cells treated with oxaliplatin in normal tissue culture media showed high fluorescent counts. Removal of serum resulted in reduced fluorescence levels, irrespective of whether oxaliplatin was added or not (data not shown).

Different concentrations of secondary antibodies were tested, which resulted in brighter and specific staining for ERCC1 only (data not shown). Primary and secondary antibodies for XPF detection were substituted, resulting in brighter and specific XPF staining (data not shown). Staining with primary and secondary antibodies only as a negative control produced negligible fluorescent counts in comparison with fully stained cells (Figure 4.10).

Wild type and variant cell lines fully stained with primary and secondary antibodies after oxaliplatin treatment are shown in Figure 4.11. Nuclear and cytoplasmic percentage localisation of ERCC1 and XPF in individual wild type cell lines after oxaliplatin treatment showed considerable variation (approximately 10% and 9% range respectively) (Figure 4.12).

Average nuclear ERCC1 in cell lines carrying Pro379Ser and Arg576Thr localised towards the lower end of the wild type localisation range, whereas cell lines carrying Glu875Gly localised towards the upper end of the wild type range (Figure 4.12a), with marked variation between localisation in cell lines carrying the same variant. Average localisation of nuclear XPF showed a very similar pattern in all cell lines (Figure 4.12b).

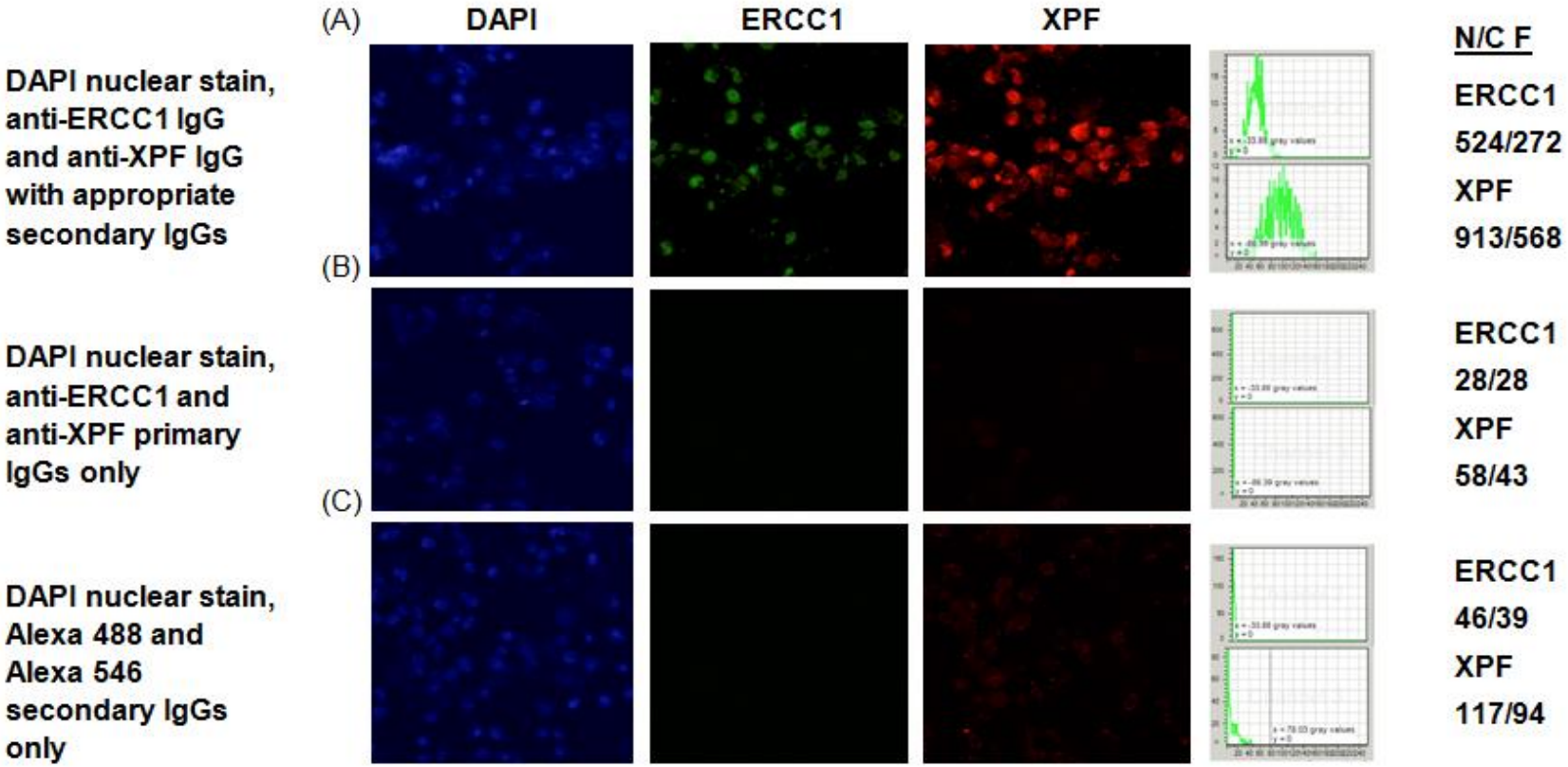


Figure 4.10 Negative controls in wild type cell lines with oxaliplatin treatment. Wild type cells starved for 1 hour in SF media followed by oxaliplatin (125µM) treatment for 1 hour. Cells were incubated for 2 hours after drug treatment and stained with (A) DAPI nuclear stain with anti-ERCC1 and anti-XPF IgGs with corresponding Alexa 488 and Alexa 546 IgGs, (B) DAPI nuclear stain with anti-ERCC1 and anti-XPF primary IgGs only and (C) DAPI nuclear stain with Alexa 488 and Alexa 564 secondary IgGs only. Intensity peaks (right) indicate intensity of ERCC1 (above) and XPF (below) staining respectively and nuclear/cytoplasmic fluorescent count (N/C F) is shown on the far right. Fluorescent counts are the average of twenty cells per cell line (sixty cells per cell group) stained and imaged as part of the same biological repeat experiment.

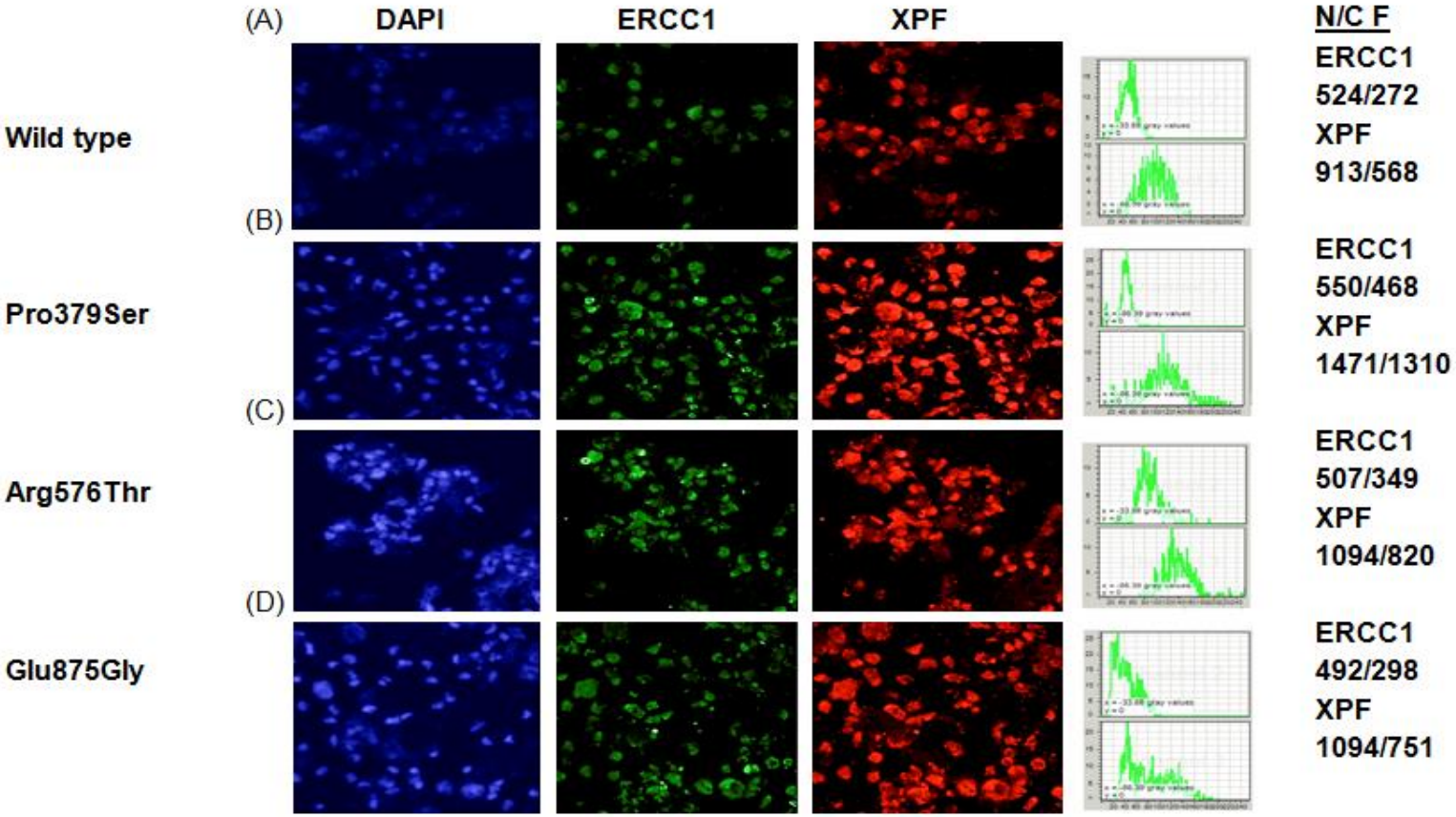
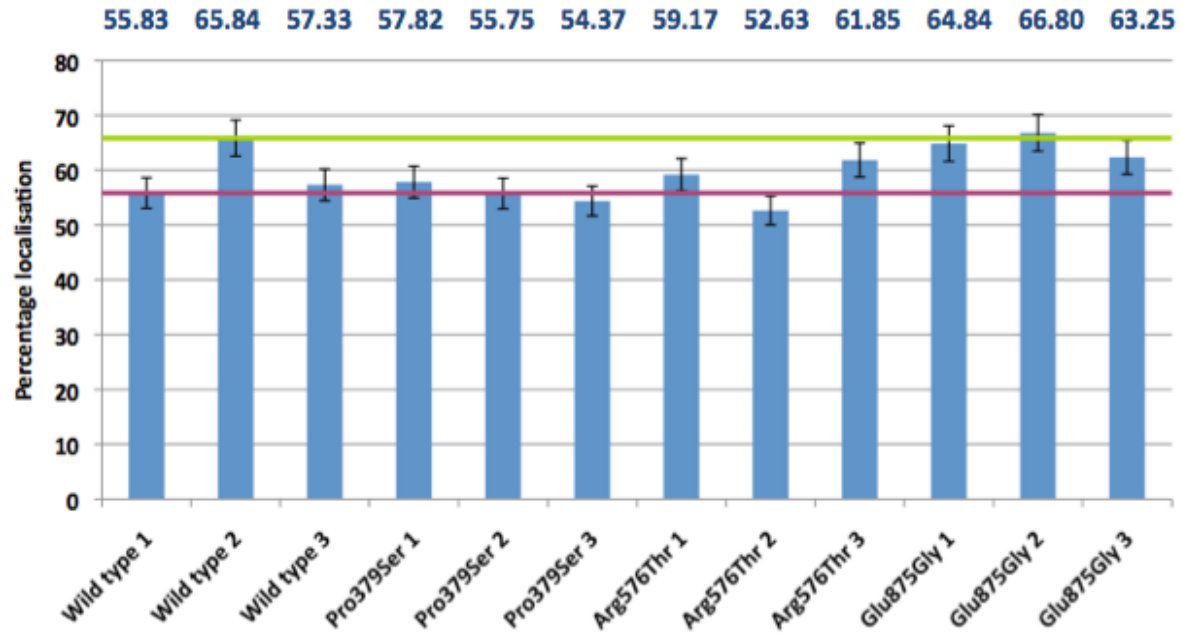


Figure 4.11 Immunofluorescence of all cell lines with oxaliplatin treatment. Cell lines treated with oxaliplatin (125µM) for 1 hour in SF media, followed by 2 hour incubation and staining with DAPI nuclear stain and anti-ERCC1 and anti-XPF IgGs with corresponding Alexa 488 and Alexa 546 IgGs, showing (A) wild type cells and cells carrying (B) Pro379Ser, (C) Arg576Thr and (D) Glu875Gly. Intensity peaks (right) indicate intensity of ERCC1 (above) and XPF (below) staining respectively and nuclear/cytoplasmic fluorescent count (N/C F) is shown on the far right. Fluorescent counts are the average of twenty cells per cell line (sixty cells per cell group) stained and imaged as part of the same biological repeat experiment.

(A)



(B)

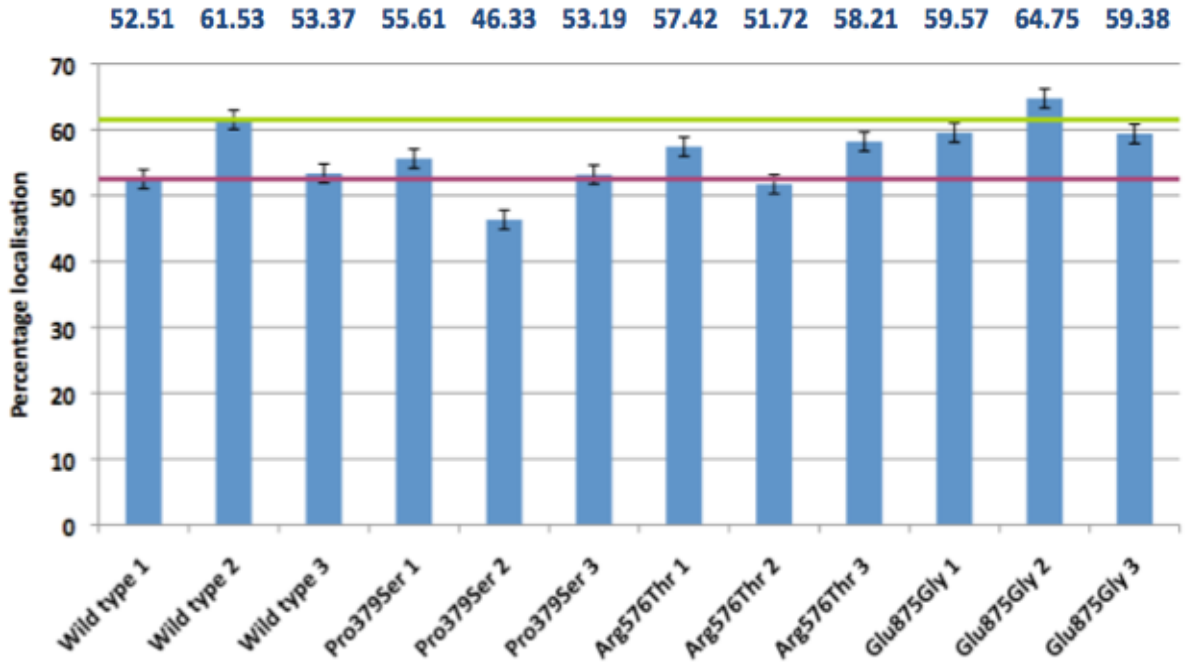


Figure 4.12 Percentage localisation of nuclear ERCC1 and XPF with oxaliplatin treatment. Average percentage localisation of nuclear (A) ERCC1 and (B) XPF across all cell lines with oxaliplatin (125µM; 1 hour) treatment (n=20), with standard error bars. Horizontal lines indicate the maximum (green) and minimum (red) percentage localisation in wild type cell lines. Nuclear percentage localisation values for each cell line are indicated in blue on each plot. Percentage localisation values are the average of twenty cells per cell line stained and imaged as part of the same biological repeat experiment.

Statistical analysis of nuclear ERCC1 and XPF localisation was performed using a two-way ANOVA, comparing mean localisation values for each cell line carrying the same *ERCC4* variant. Comparison of localisation per cell group in this way revealed no statistically significant interaction between the effects of mutation and treatment on the percentage localisation of ERCC1 and XPF ($P = 0.147$ and $P = 0.569$, respectively; Table 4.5b). However, there was a statistically significant difference between cell groups with regard to mutation for ERCC1 localisation ($P = 0.004$) and with regard to mutation and treatment type for XPF localisation ($P = 0.017$ and $P = 0.009$, respectively; Table 4.5b).

A Tukey post-test was performed to identify which cell groups differed from one another with statistical significance with regard to percentage localisation. This revealed statistically significant differences in localisation of ERCC1 and XPF between Pro379Ser and Glu875Gly cell groups ($P = 0.033$ and $P = 0.021$, respectively), and between Arg576Thr and Glu875Gly cell groups ($P = 0.003$ and $P = 0.031$, respectively; Table 4.5c). All other associations showed no statistical significance, which was consistent with the relatively small percentage localisation differences across all cell lines (Figure 4.12).

4.3.3 Relative-fold increase of ERCC1-XPF after damage

The RFI of nuclear ERCC1 and XPF localisation after oxaliplatin damage was calculated for all cell lines. In wild type cell lines, we observed influx of ERCC1 and XPF in two and three wild type cell lines respectively. Cell lines carrying Pro379Ser and Glu875Gly showed similar recruitment levels of both ERCC1 and XPF as wild type cell lines, whereas cell lines carrying Arg576Thr demonstrated greater recruitment overall than wild type cell lines. Recruitment profiles for each cell line for both ERCC1 and XPF were very similar (Figure 4.13).

Statistical analysis of nuclear RFI for ERCC1 and XPF was performed using a one-way ANOVA, comparing mean RFI values for each cell line carrying the same *ERCC4* variant. Comparison of RFI per cell group in this way

A.

		ERCC1		XPF	
Mutation	**Treatment**	**Mean**	**SD**	**Mean**	**SD**
Wild Type	NT	60.02667	6.272522	50.95000	2.311623
	OXA	59.66667	5.398614	55.80333	4.978045
Pro379Ser	NT	55.84667	1.807881	49.28667	1.408699
	OXA	55.98000	1.736462	51.71000	4.813772
Arg576Thr	NT	45.90667	10.637111	46.21333	8.198112
	OXA	57.88333	4.742756	55.78333	3.541050
Glu875Gly	NT	66.78333	5.287025	56.69000	3.596679
	OXA	64.66333	2.230254	61.23333	3.047004

B.

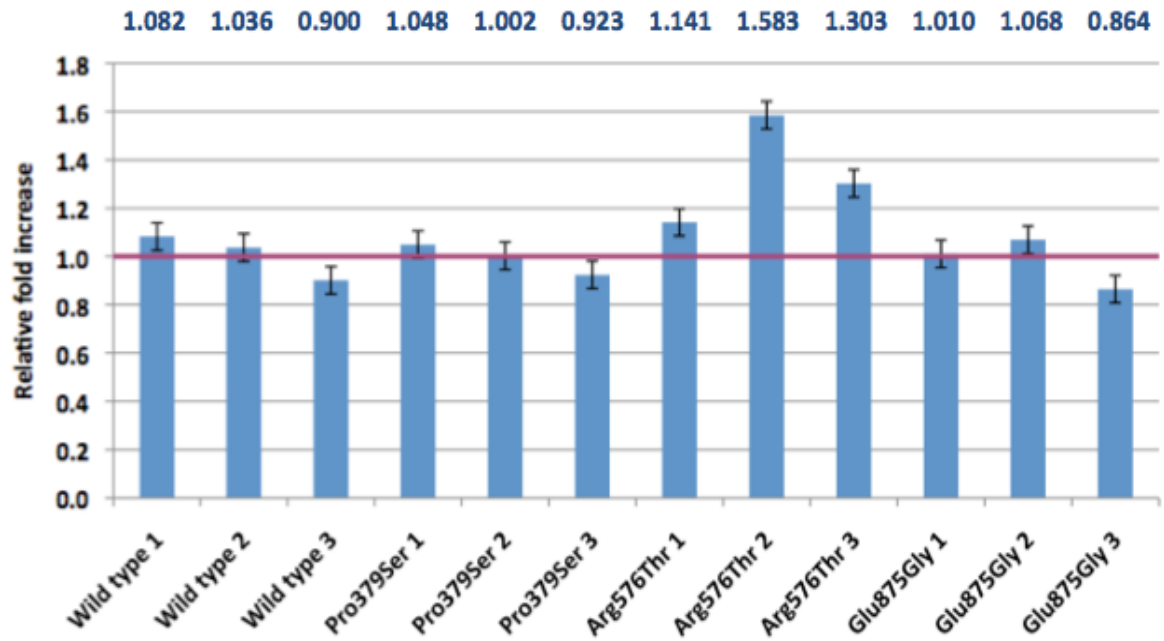
	ERCC1	XPF
Mutation	0.004	0.017
Treatment	0.301	0.009
Mutation*Treatment	0.147	0.569

C.

Mutation		P value (ERCC1)	P value (XPF)
Wild Type	**Pro379Ser**	0.615	0.680
	Arg576Thr	0.099	0.790
	Glu875Gly	0.289	0.171
Pro379Ser	**Wild Type**	0.615	0.680
	Arg576Thr	0.599	0.997
	Glu875Gly	0.033	0.021
Arg576Thr	**Wild Type**	0.099	0.790
	Pro379Ser	0.599	0.997
	Glu875Gly	0.003	0.031
Glu875Gly	**Wild Type**	0.289	0.171
	Pro379Ser	0.033	0.021
	Arg576Thr	0.003	0.031

Table 4.5 Statistical analysis of percentage localisation values of nuclear ERCC1 and XPF in cell groups with and without treatment. Percentage localisation values for ERCC1 and XPF per cell line were averaged and mean localisation values were compared for cell lines carrying the same *ERCC4* variant using a two-way ANOVA. (A) Mean localisation values and their standard deviations (SD) are given for each cell group. (B) P values are given for mutation, treatment type and the combination of both variables (mutation*treatment) for ERCC1 and XPF between different cell groups. (C) Results of the Tukey post test comparing mean localisation of ERCC1 and XPF across cell groups are listed with P values. Statistically significant values (<0.05) are highlighted. NT = no treatment; OXA = oxaliplatin (125µM; 1 hour) treatment.

(A)



(B)

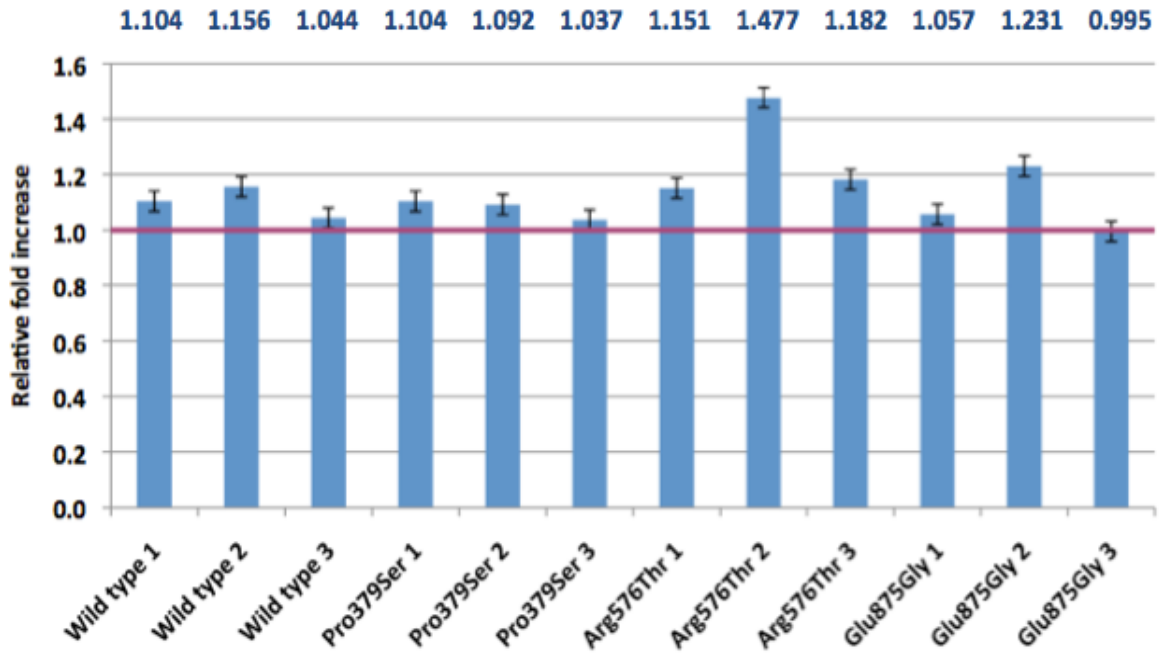


Figure 4.13 Relative-fold increase of nuclear ERCC1 and XPF after oxaliplatin treatment. Average relative fold increase of (A) ERCC1 and (B) XPF localisation to the nucleus after DNA damage by oxaliplatin (n=20), with standard error bars. A relative fold increase of 1.0 indicates no recruitment (shown in red). RFI values are shown in blue on each plot. A RFI value greater than 1 indicates recruitment into the nucleus, whereas a value less than 1 indicates recruitment from the nucleus into the cytoplasm upon DNA damage. RFI values were calculated using percentage localisation values from twenty cells without treatment and twenty cells following oxaliplatin treatment per cell line, taken from two separate biological experiments.

revealed a statistically significant difference in RFI between cell groups for ERCC1 only (P = 0.031; Table 4.6b). A Tukey post-test was performed to identify which cell groups differed from one another with statistical significance. This revealed statistically significant differences in RFI of ERCC1 between Arg576Thr and Glu875Gly cell groups (P = 0.047). No RFI values for XPF were statistically significant (Table 4.6c).

4.3.4 Percentage localisation of rad16 without treatment

During optimisation, cells from the wild type and rad16^{Pro361Ser} strains were treated with varying concentrations of zymolyase to catalyse cell wall degradation. We observed the greatest degree of cell wall destruction at the lowest concentration tested (0.1mg/ml) when incubated for up to 90 minutes. We tested different concentrations of Triton X-100 for cell membrane permeabilisation, which revealed optimum antibody penetration into cells when using a Triton X-100 concentration of 4% in combination with 0.1mg/ml zymolyase (data not shown). Different primary His tag antibody concentrations (5-30µg/ml) were also tested to identify the maximum concentration that could be used whilst maintaining low background noise (data not shown).

The wild type and rad16^{Pro361Ser} strains were stained with primary and secondary antibodies for the His tag conjugated to rad16, and with primary and secondary antibody only as negative controls (Figure 4.14). This revealed negligible fluorescence levels in primary only and secondary only stained cells.

Wild type and variant strains fully stained with primary and secondary antibodies without treatment are shown in Figure 4.15. Average percentage localisation of rad16 in wild type and variant strains without treatment was very similar (Figure 4.16a). Comparison of nuclear rad16 localisation between *S. pombe* strains was performed using a two-way ANOVA; therefore the results of statistical analysis are outlined in section 4.3.5.

4.3.5 Percentage localisation of rad16 with oxaliplatin treatment

During optimisation, *S. pombe* strains were prepared for oxaliplatin treatment and subsequent immunofluorescence using identical conditions for fixation, cell wall digestion, permeabilisation and blocking as those used in strains without treatment. Different primary His tag antibody and secondary

A.

Mutation	ERCC1		XPF	
	Mean	SD	Mean	SD
Wild Type	1.00587	0.094101	1.10154	0.056347
Pro379Ser	0.99103	0.063116	1.07762	0.035810
Arg576Thr	1.34284	0.224836	1.27033	0.180077
Glu875Gly	0.98066	0.105328	1.09437	0.122005

B.

	P value (ERCC1)	P value (XPF)
Mutation (Between cell groups)	0.031	0.211

C.

Mutation		P value (ERCC1)	P value (XPF)
Wild Type	**Pro379Ser**	0.999	0.994
	Arg576Thr	0.064	0.333
	Glu875Gly	0.996	1.000
Pro379Ser	**Wild Type**	0.999	0.994
	Arg576Thr	0.053	0.240
	Glu875Gly	1.000	0.998
Arg576Thr	**Wild Type**	0.064	0.333
	Pro379Ser	0.053	0.240
	Glu875Gly	0.047	0.302
Glu875Gly	**Wild Type**	0.996	1.000
	Pro379Ser	1.000	0.998
	Arg576Thr	0.047	0.302

Table 4.6 Statistical analysis of relative-fold increase values of nuclear ERCC1 and XPF after oxaliplatin treatment of cell groups. RFI values for ERCC1 and XPF per cell line were averaged and mean localisation values were compared for cell lines carrying the same *ERCC4* variant using a one-way ANOVA. (A) Mean RFI values and their standard deviations (SD) are given for each cell group. (B) P values are given for comparison of RFI between different mutation groups for ERCC1 and XPF. (C) Results of the Tukey post-test comparing mean RFI of ERCC1 and XPF across cell groups are listed with P values. Statistically significant values (<0.05) are highlighted.

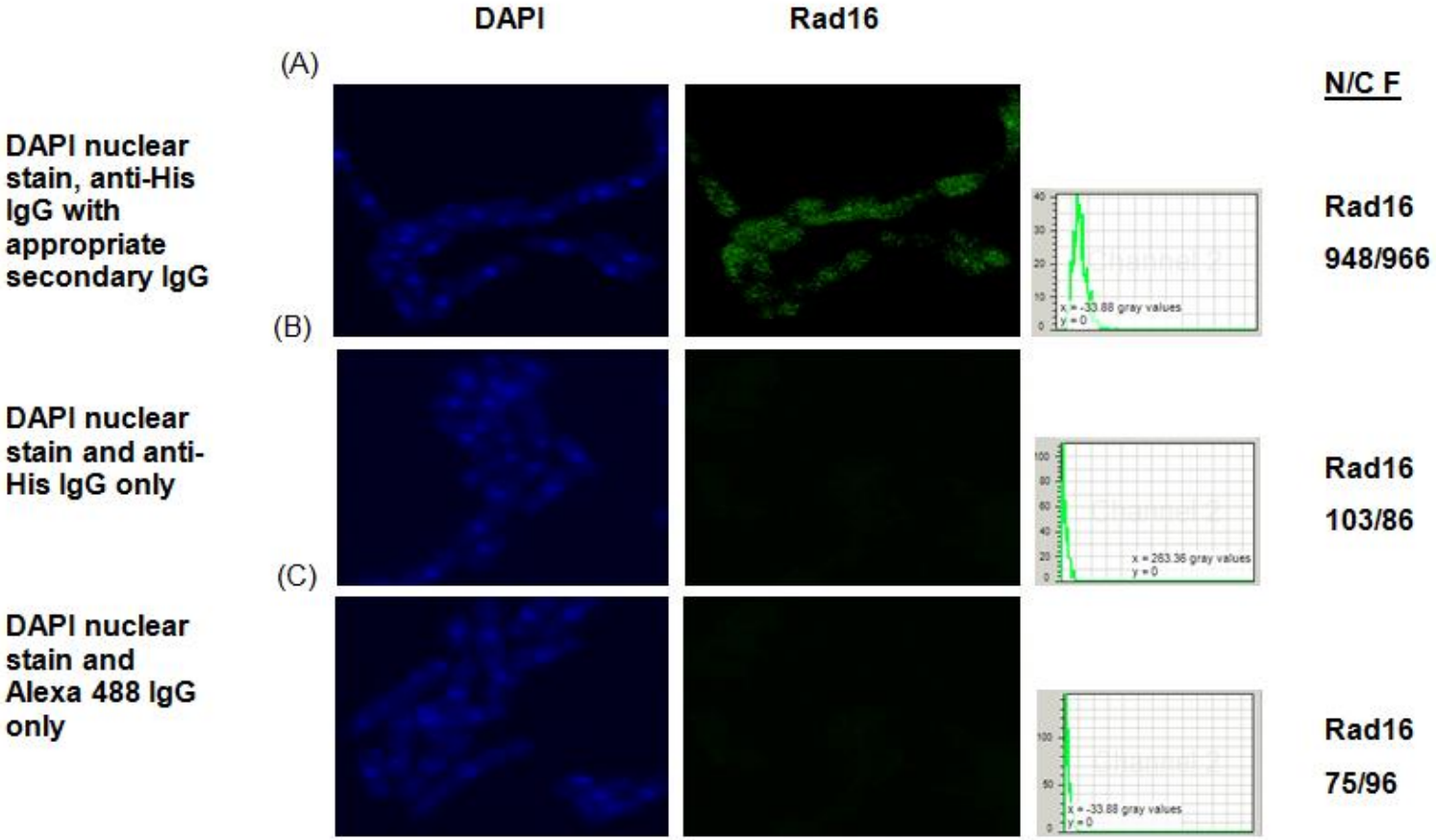


Figure 4.14 Negative controls in cells from the *S. pombe* wild type strain without treatment. Wild type *S. pombe* cells stained with (A) DAPI nuclear stain with primary anti-His tag IgG with corresponding secondary Alexa 488 IgG, (B) DAPI nuclear stain with anti-His tag IgG only, and (C) DAPI nuclear stain with Alexa 488 IgG only. Intensity peaks (right) indicate intensity of His tag staining and nuclear/cytoplasmic fluorescent count (N/C F) is shown on the far right. Fluorescent counts are the average of twenty cells per strain stained and imaged as part of the same biological repeat experiment.

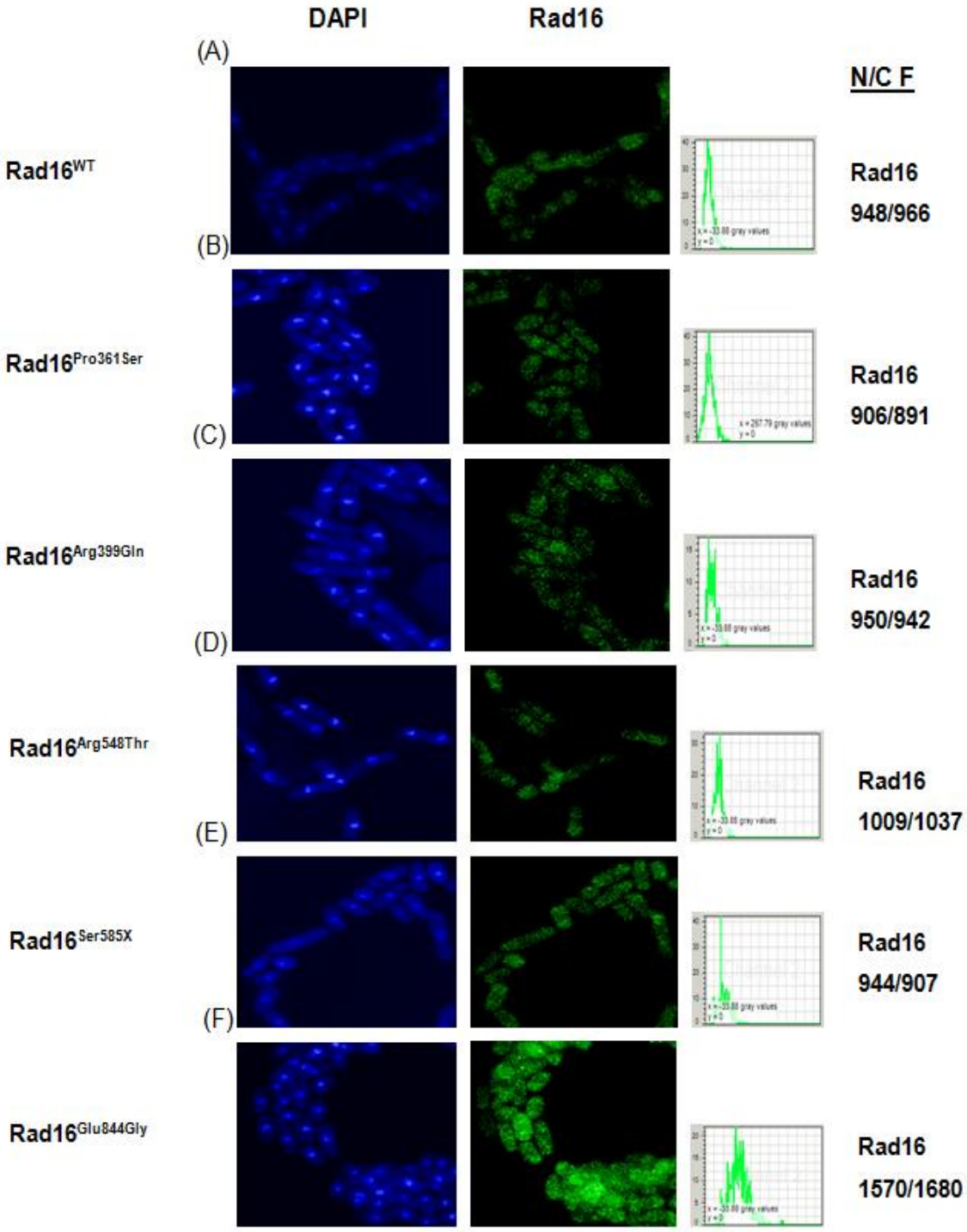
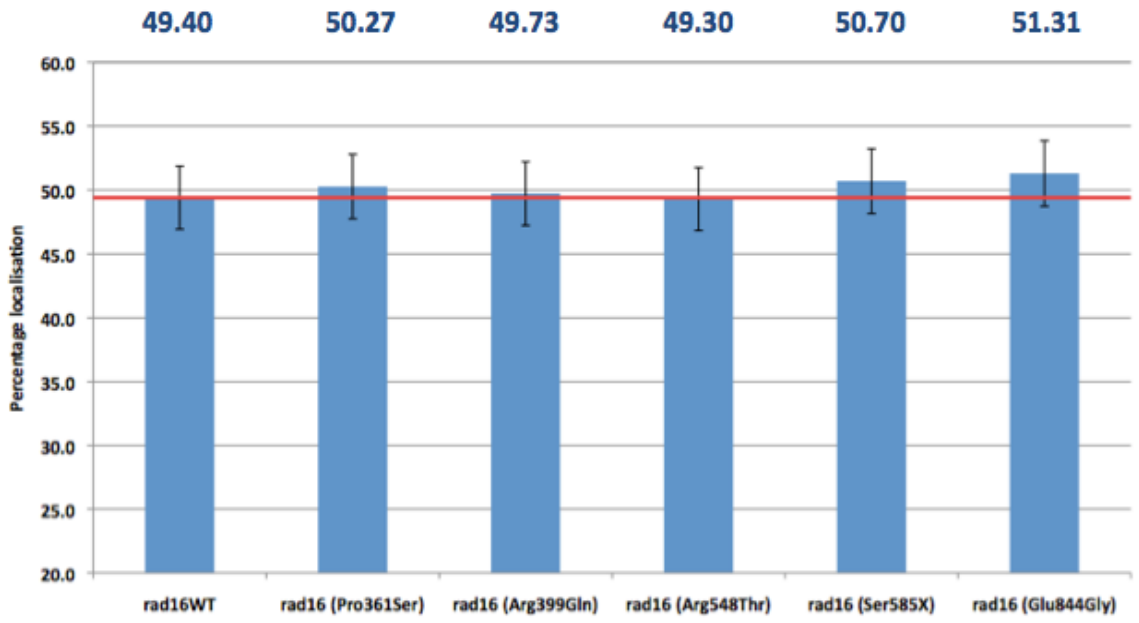


Figure 4.15 Immunofluorescence of wild type and variant *S. pombe* strains without treatment. Wild type and variant *rad16* strains without drug treatment stained with DAPI nuclear stain, anti-His tag IgG and corresponding Alexa 488 IgG, showing (A) wild type, (B) rad16^{Pro361Ser}, (C) rad16^{Arg399Gln}, (D) rad16^{Arg548Thr}, (E) rad16^{Ser585X} and (F) rad16^{Glu844Gly}. Intensity peaks (right) indicate intensity of His tag staining and nuclear/cytoplasmic fluorescent count (N/C F) is shown on the far right. Fluorescent counts are the average of twenty cells per strain stained and imaged as part of the same biological repeat experiment.

(A)



(B)

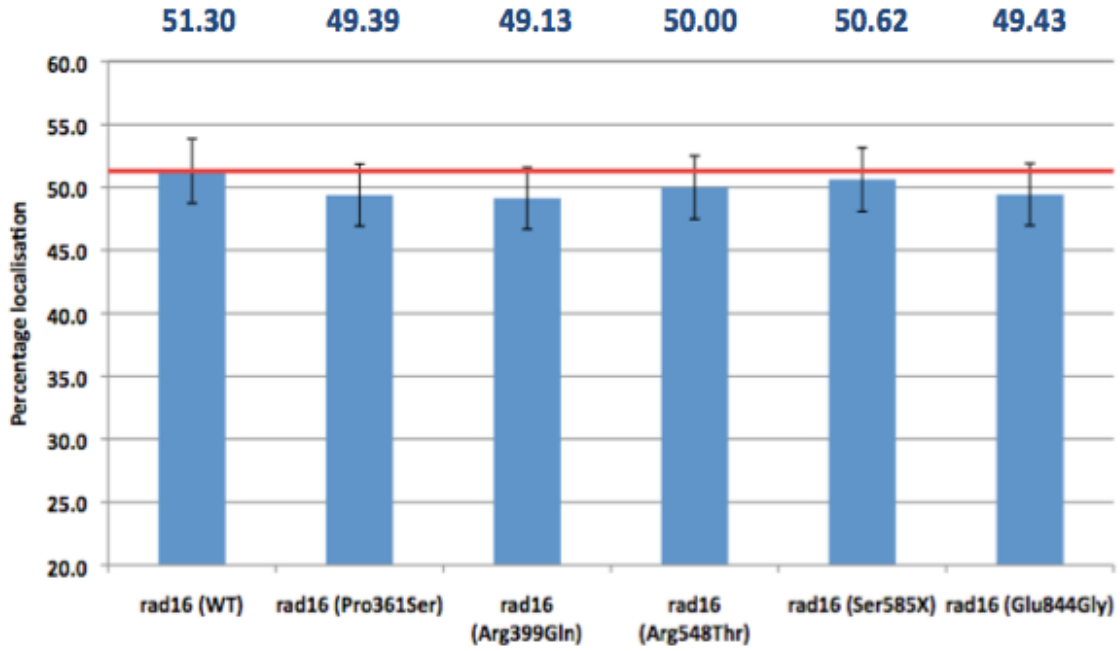


Figure 4.16 Percentage localisation of nuclear rad16 with and without oxaliplatin treatment. Average percentage localisation of nuclear rad16 across all *S. pombe* strains (A) without treatment and (B) with oxaliplatin (1mM; 1 hour) treatment (n=20), with standard error bars. Horizontal red lines indicate the average percentage localisation of rad16 in the wild type strain. Percentage localisation values for each strain are indicated in blue on each plot. Percentage localisation values are the average of twenty cells per strain stained and imaged as part of the same biological repeat experiment.

Alexa 488 antibody concentrations were tested, resulting in optimum staining conditions when the maximum concentrations were used (data not shown). Additionally, during confocal imaging, higher voltages were tested for imaging strains treated with oxaliplatin compared to those without treatment. This revealed that strains treated with oxaliplatin and stained immunofluorescently could be imaged at higher voltages whilst maintaining low background noise.

Wild type and rad16^{Pro361Ser} strains were stained with primary and secondary antibodies for the His tag conjugated to rad16, and with primary and secondary antibody only as negative controls, which revealed negligible fluorescence levels in primary only and secondary only stained cells (data not shown).

Wild type and variant strains fully stained with primary and secondary antibodies after oxaliplatin treatment are shown in Figure 4.17. Similar to untreated strains, average percentage localisation of nuclear rad16 in wild type and variant strains was very similar (approximately 2% range; Figure 4.16b).

Statistical analysis of nuclear rad16 localisation was performed using a two-way ANOVA, comparing all percentage localisation values (n=20) for each *S. pombe* strain. Since these values were acquired during the same experiment, these cannot be considered different experiments. Comparison of localisation per strain in this way revealed a statistically significant interaction between the effects of mutation and treatment on the percentage localisation of rad16 ($P = 0.029$; Table 4.7b). However, there was no statistically significant difference between variant strains with regard to mutation or treatment only for rad16 localisation ($P = 0.242$ and $P = 0.670$, respectively; Table 4.7b). A Tukey post-test confirmed that there were no statistically significant differences between variant strains with regard to rad16 localisation (Table 4.7c). This was consistent with the small range of percentage localisation between all strains (range of approximately 2%; Figure 4.16b).

4.3.6 Relative-fold increase of rad16 after damage

The RFI of nuclear rad16 localisation after DNA damage by oxaliplatin was calculated for all *S. pombe* strains. In wild type cells, we observed marginal recruitment into the nucleus upon DNA damage, whilst the rad16^{Ser585X} strain displayed no recruitment. All remaining strains showed virtually no recruitment

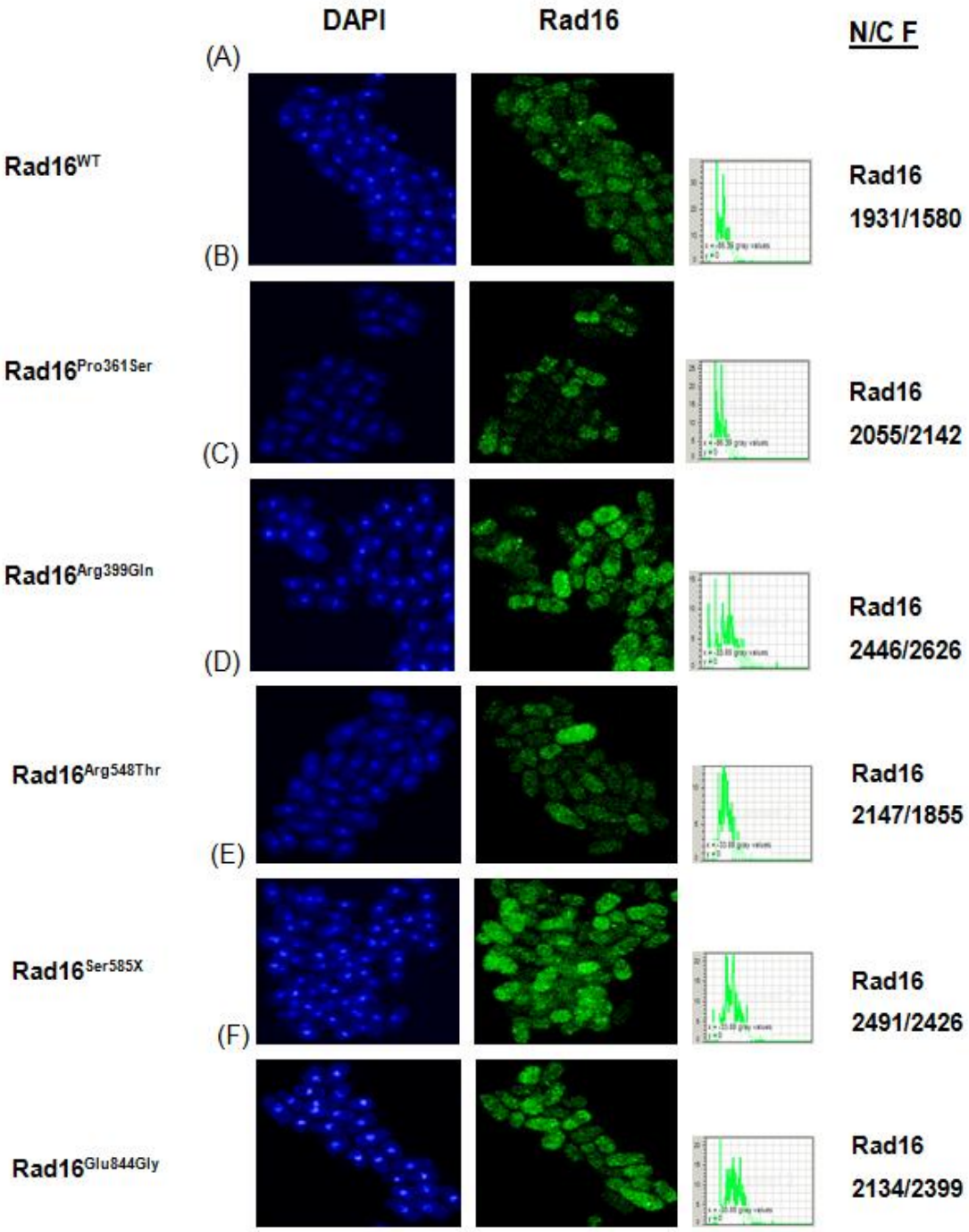


Figure 4.17 Immunofluorescence of wild type and variant *S. pombe* strains with oxaliplatin treatment. Wild type and variant *rad16* strains with oxaliplatin (1mM) drug treatment stained with DAPI nuclear stain, anti-His tag IgG and corresponding Alexa 488 IgG, showing (A) wild type, (B) rad16^{Pro361Ser}, (C) rad16^{Arg399Gln}, (D) rad16^{Arg548Thr}, (E) rad16^{Ser585X} and (F) rad16^{Glu844Gly}. Intensity peaks (right) indicate intensity of His tag staining and nuclear/cytoplasmic fluorescent count (N/C F) is shown on the far right. Fluorescent counts are the average of twenty cells per strain stained and imaged as part of the same biological repeat experiment.

A.

		Rad16	
Mutation	**Treatment**	**Mean**	**SD**
$Rad16^{WT}$	NT	49.40259	1.899078
	OXA	51.29602	2.883166
$Rad16^{Pro361Ser}$	NT	50.27336	1.989772
	OXA	49.38623	3.069647
$Rad16^{Arg399Gln}$	NT	49.73494	2.834460
	OXA	49.12789	3.716874
$Rad16^{Arg548Thr}$	NT	49.30107	1.768717
	OXA	49.99595	3.255106
$Rad16^{Ser585X}$	NT	50.69731	2.219011
	OXA	50.61944	2.344482
$Rad16^{Glu844Gly}$	NT	51.30586	2.122648
	OXA	49.42664	2.401261

B.

	Rad16
Mutation	0.242
Treatment	0.670
Mutation*Treatment	0.029

C.

Mutation		P value (Rad16)
Rad16^{WT}	**Rad16^{Pro361Ser}**	0.948
	Rad16^{Arg399Gln}	0.616
	Rad16^{Arg548Thr}	0.836
	Rad16^{Ser585X}	0.995
	Rad16^{Glu844Gly}	1.000
Rad16^{Pro361Ser}	**Rad16^{WT}**	0.948
	Rad16^{Arg399Gln}	0.984
	Rad16^{Arg548Thr}	1.000
	Rad16^{Ser585X}	0.714
	Rad16^{Glu844Gly}	0.941
Rad16^{Arg399Gln}	**Rad16^{WT}**	0.616
	Rad16^{Pro361Ser}	0.984
	Rad16^{Arg548Thr}	0.999
	Rad16^{Ser585X}	0.289
	Rad16^{Glu844Gly}	0.597
Rad16^{Arg548Thr}	**Rad16^{WT}**	0.836
	Rad16^{Pro361Ser}	1.000
	Rad16^{Arg399Gln}	0.999
	Rad16^{Ser585X}	0.512
	Rad16^{Glu844Gly}	0.821
Rad16^{Ser585X}	**Rad16^{WT}**	0.995
	Rad16^{Pro361Ser}	0.714
	Rad16^{Arg399Gln}	0.289
	Rad16^{Arg548Thr}	0.512
	Rad16^{Glu844Gly}	0.996
Rad16^{Glu844Gly}	**Rad16^{WT}**	1.000
	Rad16^{Pro361Ser}	0.941
	Rad16^{Arg399Gln}	0.597
	Rad16^{Arg548Thr}	0.821
	Rad16^{Ser585X}	0.996

Table 4.7 Statistical analysis of percentage localisation values of nuclear rad16 in *S. pombe* strains with and without treatment. Percentage localisation values for rad16 per *S. pombe* strain (n=20) were compared using a two-way ANOVA. All twenty localisation values per strain were acquired using the same slides that were immunofluorescently stained at the same time, therefore these values cannot be considered as separate experiments. (A) Mean localisation values and their standard deviations (SD) are given for strain. (B) P values are given for mutation, treatment type and the combination of both variables (mutation*treatment) for rad16 between different *S. pombe* strains. (C) Results of the Tukey post-test comparing mean localisation of rad16 across between strains are listed with P values. Statistically significant values (<0.05) are highlighted. NT = no treatment; OXA = oxaliplatin (125µM; 1 hour) treatment.

in comparison with wild type cells (Figure 4.18).

Statistical analysis of nuclear RFI for rad16 was performed using a one-way ANOVA, comparing all RFI values (n=20) for each *S. pombe* strain. Since these values were acquired during the same experiment, these cannot be considered different experiments. Comparison of RFI per strain in this way revealed a statistically significant difference in RFI between strains for rad16 RFI (P = 0.024; Table 4.8b). A Tukey post-test was performed to identify which cell groups differed from one another with statistical significance. This revealed statistically significant differences in RFI of rad16 between wild type and Glu844Gly strains only (P = 0.014; Table 4.8c).

4.4 Discussion

4.4.1 Optimisation of immunofluorescence assay in cell lines

Staining for both ERCC1 and XPF was pan-cellular throughout the vast majority of cells seen in all cell lines. It was not possible to ascertain by eye whether fluorescence was primarily nuclear or cytoplasmic in nature; therefore the method used to quantify fluorescence was important to optimise correctly. Comparison of the same wild type cells imaged as MP and section-by-section files revealed large discrepancies in total fluorescent count, despite an overall trend in a similar direction for each stain. This is potentially due to the incorporation of numerous 0.5µm sections without any fluorescence at the top and bottom of the cell, which could skew the overall fluorescent count towards the lower range when using the section-by-section approach. Additionally, this approach required manual encircling of ROI1 and ROI2 areas for every section throughout the Z-stack, resulting in a much larger degree of human error than is involved for MP files. This large degree of human error, coupled with the fact that MP files showed a much higher degree of fluorescence, made data acquisition by MP files the preferable imaging method.

Wild type cells stained with primary antibody and secondary antibody only were used as negative controls to ensure fluorescence was specific. It was evident from the low background fluorescence that immunofluorescence was detected only within the cells. Negative control samples were quantified,

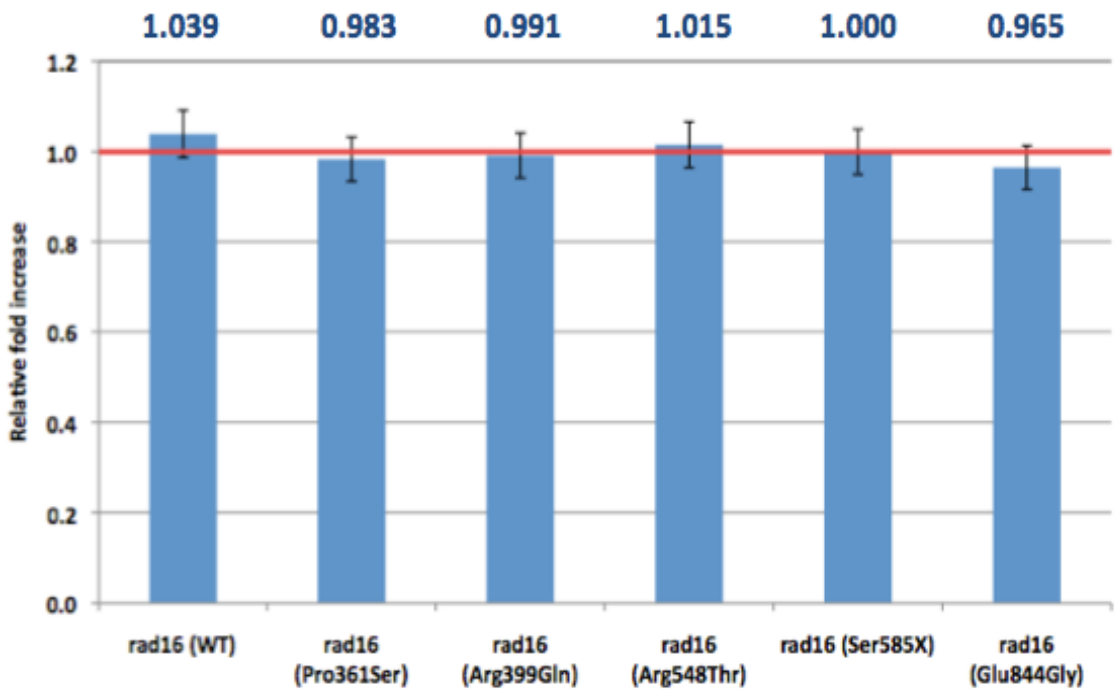


Figure 4.18 Relative fold increase of nuclear rad16 after oxaliplatin treatment. Average relative fold increase of rad16 localisation to the nucleus after DNA damage by oxaliplatin (n=20), with standard error bars. A relative fold increase of 1.0 indicates no recruitment (shown in red). A RFI value greater than 1 indicates recruitment into the nucleus, whereas a value less than 1 indicates recruitment from the nucleus into the cytoplasm upon DNA damage. RFI values for each strain are indicated in blue above each column on the graph. RFI values were calculated using percentage localisation values from twenty cells without treatment and twenty cells following oxaliplatin treatment per strain, taken from two separate biological experiments.

A.

Mutation	Rad16	
	Mean	**SD**
$Rad16^{WT}$	1.03894	0.056186
$Rad16^{Pro361Ser}$	0.98311	0.061508
$Rad16^{Arg399Gln}$	0.99143	0.099953
$Rad16^{Arg548Thr}$	1.01529	0.073366
$Rad16^{Ser585X}$	0.99987	0.056152
$Rad16^{Glu844Gly}$	0.96502	0.063826

B.

	P value (Rad16)
Mutation (Between cell groups)	0.024

C.

Mutation		P value (Rad16)
Rad16^{WT}	**Rad16^{Pro361Ser}**	0.128
	Rad16^{Arg399Gln}	0.274
	Rad16^{Arg548Thr}	0.894
	Rad16^{Ser585X}	0.495
	Rad16^{Glu844Gly}	0.014
Rad16^{Pro361Ser}	**Rad16^{WT}**	0.128
	Rad16^{Arg399Gln}	0.999
	Rad16^{Arg548Thr}	0.696
	Rad16^{Ser585X}	0.974
	Rad16^{Glu844Gly}	0.964
Rad16^{Arg399Gln}	**Rad16^{WT}**	0.274
	Rad16^{Pro361Ser}	0.999
	Rad16^{Arg548Thr}	0.890
	Rad16^{Ser585X}	0.999
	Rad16^{Glu844Gly}	0.841
Rad16^{Arg548Thr}	**Rad16^{WT}**	0.894
	Rad16^{Pro361Ser}	0.696
	Rad16^{Arg399Gln}	0.890
	Rad16^{Ser585X}	0.982
	Rad16^{Glu844Gly}	0.217
Rad16^{Ser585X}	**Rad16^{WT}**	0.495
	Rad16^{Pro361Ser}	0.974
	Rad16^{Arg399Gln}	0.999
	Rad16^{Arg548Thr}	0.982
	Rad16^{Glu844Gly}	0.619
Rad16^{Glu844Gly}	**Rad16^{WT}**	0.014
	Rad16^{Pro361Ser}	0.964
	Rad16^{Arg399Gln}	0.841
	Rad16^{Arg548Thr}	0.217
	Rad16^{Ser585X}	0.619

Table 4.8 Statistical analysis of relative-fold increase values of nuclear rad16 after oxaliplatin treatment of *S. pombe* strains. RFI values for rad16 per S. pombe strain (n=20) were compared using a one-way ANOVA. All twenty RFI values per strain were acquired using the same slides that were immunofluorescently stained at the same time, therefore these values cannot be considered as separate experiments. (A) Mean RFI values and their standard deviations (SD) are given for each cell group. (B) P values are given for comparison of RFI between different mutation groups for rad16. (C) Results of the Tukey post-test comparing mean RFI of rad16 between strains are listed with P values. Statistically significant values (<0.05) are highlighted.

demonstrating negligible fluorescence in the absence of either primary or secondary antibody. All negative controls showed slightly higher background fluorescence in the XPF (red) channel than the ERCC1 (green) channel, most likely resulting from autofluorescence of cellular structures in the red spectrum. Several cellular flavins have been associated with autofluorescence in the red spectrum (approx. 500-700nm), including flavin adenine dinucleotide (FAD), a redox cofactor naturally present in all cells, which has been shown to auto-fluoresce in the region of 500-600nm (Sivabalan *et al.* 2010).

Prior to oxaliplatin treatment, serum was removed from tissue culture media to prevent the drug from binding to serum molecules (Yue *et al.* 2009), which could potentially prevent DNA damage from occurring. Interestingly, removal of serum resulted in significantly reduced fluorescence, regardless of duration of serum starving. We hypothesised that this could be due to (1) a reduction in protein turnover rate, (2) autofluorescence caused by serum when present or (3) protein-protein interactions resulting from starving that could mask one or both antibody epitopes. We observed reduction of fluorescence immediately after removal of serum, suggesting that a reduction in protein turnover rate is unlikely. Similarly, we would expect medium-high background fluorescence if serum presence produced autofluorescence, which was not apparent in cell lines stained in normal media without treatment. Therefore, the most likely assumption is that starving alters the epitope structure in some way, although there is currently nothing in the literature to support this hypothesis.

In order to ensure that this reduction in fluorescence was due to removal of serum and not a consequence of oxaliplatin treatment, wild type cells were stained after removal of serum in the presence and absence of oxaliplatin, which confirmed that oxaliplatin treatment had no effect on fluorescence reduction.

The reduction of fluorescence after serum removal proved problematic for localisation detection. Higher concentrations of secondary antibody were tested and staining with primary antibody only and secondary antibody only revealed specific staining for ERCC1 detection. Specific staining for XPF detection was achieved by substitution of antibodies. These changes in serum

and antibody conditions mean that it was necessary to repeat the negative controls after oxaliplatin treatment, which again confirmed that fluorescence was specific.

4.4.2 Percentage localisation of ERCC1 and XPF

Percentage localisation of ERCC1 and XPF for each cell line was very similar, both before and after oxaliplatin treatment, suggesting that our immunofluorescence assay was consistent and was probing for conjugated repair protein. This also indicated that the method used to quantify fluorescence was consistent across all cell lines, even where different antibodies were used for XPF detection after oxaliplatin treatment.

There was considerable variation in percentage localisation of ERCC1 and XPF between individual wild type cell lines, both before and after oxaliplatin treatment. This could be the result of the subtle nature of the *ERCC4* variants, or due to the presence of additional polymorphisms in other NER genes in these cell lines. We genotyped all cell lines for common polymorphisms in *ERCC1* and *ERCC4* (section 3.3.1); however, it is possible that there were polymorphisms present in other NER-related genes that we were not aware of, which could have additional functional effects on localisation in wild type cell lines. Alternatively, this may be the result of an assay limitation in either the detection of ERCC1 and XPF, or the method used for quantification of fluorescence.

The percentage localisation differences between cell groups for ERCC1 and XPF were relatively small. All cell lines carrying Arg576Thr showed reduced nuclear localisation of ERCC1 and XPF without treatment, compared with wild type cell lines. In contrast, cell lines carrying Glu875Gly showed markedly increased nuclear localisation of ERCC1 and XPF without treatment compared with wild type cell lines.

Despite the variation in percentage localisation of ERCC1 and XPF between individual wild type cell lines, we decided to analyse the percentage viability values for cell lines carrying the same *ERCC4* variant collectively to allow comparison of localisation between cell groups. The percentage

localisation values obtained per cell line were acquired using the same microscope slides that were immunofluorescently stained during the same experiment; therefore these localisation values cannot be considered independent experiments. As a result, localisation values acquired for each cell line were averaged, and the mean values for cell lines carrying the same *ERCC4* variant were analysed by ANOVA.

The differences in percentage localisation between wild type and variant cell lines were relatively small, before and after treatment. The differences in localisation were not large enough to suggest that our *ERCC4* variants affect localisation, either before or after damage. Statistical analysis of ERCC1 and XPF localisation between cell groups identified statistically significant differences in localisation between Pro379Ser and Glu875Gly cell groups, and between Arg576Thr and Glu875Gly cell groups only. All other cell groups showed no statistical significance during comparison by ANOVA, which was consistent with the relatively small percentage localisation differences between wild type and variant cell groups. Therefore, the results do not appear to represent a functional difference in the localisation of the repair complex.

Recruitment of ERCC1 and XPF into the nucleus after oxaliplatin damage, indicated by the RFI value, was similar across all cell lines for both ERCC1 and XPF. Following damage by oxaliplatin, we observed ERCC1 recruitment into the nucleus in nine of twelve cell lines and XPF recruitment in eleven cell lines. This indicated that our recruitment assay was consistent, including where substitution of antibodies was necessary for XPF detection. Statistical analysis of RFI values by ANOVA revealed a statistically significant difference between Arg576Thr and Glu875Gly cell groups for ERCC1 recruitment only, despite very small differences in percentage localisation before and after oxaliplatin treatment. However, all other cell groups showed no statistical significance with regard to RFI, suggesting that our results do not represent a functional difference in recruitment of the repair complex in the presence of *ERCC4* variants.

One variant in particular (Arg576Thr) showed the highest degree of recruitment into the nucleus after damage across all cell lines carrying this

variant, with a markedly higher recruitment in one cell line in particular (Arg576Thr 2). Additionally, cell lines carrying Arg576Thr showed the highest degree of mislocalisation to the cytoplasm without treatment, followed by the highest degree of nuclear recruitment after oxaliplatin treatment. Despite this, comparison of the Arg576Thr cell group with all other cell groups identified statistical significance with the Glu875Gly group only.

Mislocalisation of ERCC1 and XPF to the cytoplasm has been shown previously in monoallelic XPF mutant cell lines by immunofluorescence, including cells carrying Pro379Ser, without treatment. Similar to our data, these XPF mutant cell lines also showed pancellular XPF localisation with some cells demonstrating exclusively cytoplasmic localisation, which was consistent with ERCC1 staining (Ahmad *et al.* 2010). Our variant cell lines were haploinsufficient due to the presence of only one variant allele, and therefore consistently demonstrated pancellular localisation of ERCC1 and XPF with and without treatment.

Despite the consistency with regard to percentage localisation values acquired using this assay, there is a lack of similarity in the staining patterns between ERCC1 and XPF, which is demonstrated in Figures 4.2 and 4.3. During optimisation, the specificity of the primary antibodies was not validated; therefore, although we have confirmed that the secondary antibodies bind correctly to the primary antibodies, it is possible that the primary antibodies are not specific enough to detect only ERCC1 and XPF, respectively. This could result in observation and quantification of additional proteins, which could potentially produce unreliable percentage localisation values. The specificity of the primary antibodies could have been confirmed by transfection of bacterial plasmids containing the protein of interest and a fluorescent tag, which could be immunofluorescently stained and viewed by confocal microscopy. A specific antibody would co-localise with the fluorescent construct containing the protein of interest. An alternative positive control assay for specificity would involve transfection of cells with small interfering RNA (siRNA) to knock down the expression of ERCC1 and XPF, resulting in reduced or negligible fluorescence following immunofluorescent staining, depending on the degree of transfection that is achieved.

4.4.3 Optimisation of immunofluorescence assay in *S. pombe*

During optimisation, we tested different conditions for zymolyase treatment and cell membrane permeabilisation, as these are the two most critical factors of the protocol to enable antibody molecules to diffuse into the cell's interior. Once this was successfully optimised, different primary antibody concentrations were tested to maximise the fluorescent signal.

We stained both wild type and rad16^{Pro361Ser} strains without treatment as negative controls to account for the different cell wall thicknesses exhibited by different strains of *S. pombe* (Osumi *et al.* 1998; Calonge *et al.* 2000). This could adversely affect the amount of antibody that is able to penetrate the cell membrane; by testing optimised conditions in multiple strains we ensured that the staining conditions would be applicable to all strains of interest.

Similar to the fluorescence pattern seen in human cell lines, rad16 fluorescence was pancellular across all *S. pombe* strains. As a result, we decided to image the *S. pombe* strains as MP files using similar parameters to those used for imaging cell lines, although a higher zoom factor was necessary to account for the smaller size of cells (10μm compared with 30μm in human lymphoblastoid cells; Mitchison and Nurse, 1985). This avoided the need for repeating the optimisation process for the confocal microscope.

4.4.4 Percentage localisation of rad16

Percentage localisation of rad16 was very similar across wild type and variant strains of *S. pombe*, with and without oxaliplatin treatment. The percentage localisation range was approximately 2%, compared with >10% in human cell lines for localisation of ERCC1 and XPF. Statistical analysis of percentage localisation of rad16 between strains was performed by ANOVA, using all localisation values per strain (n=20) as input. Despite this, all values acquired for each strain cannot be considered independent experiments, since all values were acquired using the same microscope slides that were immunofluorescently stained and imaged during the same experiment. Statistical analysis of rad16 localisation between strains produced no statistically significant findings, which was consistent with the small differences in percentage localisation. This suggested that our assay was unable to detect

differences in rad16 localisation between wild type and variant *S. pombe* strains.

Following oxaliplatin treatment, the wild type strain demonstrated very little recruitment of rad16 into the nucleus to facilitate DNA repair as we would expect. We also observed no recruitment of rad16 in the rad16^{Ser585X} strain; since this strain carried the novel stop gain truncating variant present in Patient 8, we would have expected this strain to show considerably higher recruitment than wild type and remaining variant cells. However, this strain demonstrated no recruitment, whilst all other variant strains demonstrated very little recruitment. This suggested that we cannot conclude that there was any trend in rad16 recruitment after damage by oxaliplatin.

Statistical analysis of RFI following DNA damage by ANOVA showed a significant difference in recruitment patterns between the wild type and rad16^{Glu844Gly} variant strain, whereas comparison of RFI between all remaining strains demonstrated no statistical significance. Statistical analysis of RFI between strains was performed by ANOVA, using all RFI values per strain (n=20) as input. Therefore, this statistically significant difference was most likely the result of comparison of large cell numbers per strain. All RFI values per strain were acquired using the same microscope slides that were immunofluorescently stained and imaged during the same experiment, therefore they cannot be considered independent experiments.

Since *S. pombe* is a haploid model organism, there is no issue with haploinsufficiency, unlike in the variant cell lines. Therefore we hoped that, by probing for the His tag conjugated to rad16 before and after damage, we would observe a more distinct difference in the localisation and recruitment of the repair protein than that seen in the heterozygous mammalian cell lines. Unfortunately, this was not the case. It is possible that there was a limitation within the assay. The low background noise seen during imaging of *S. pombe* suggested that the antibody was specific. However, it is possible that the antibody was not sensitive enough to detect the levels of His tag present, or alternatively the protocol may have required further optimisation to increase cell wall degradation or cell membrane permeabilisation. If either of these steps

were not entirely successful, the antibody's ability to reach its target antigen could have been compromised, resulting in a reduction in fluorescent count. Alternatively, there may have been other novel repair pathways involved; it has been reported that *S. pombe* have an additional repair pathway to remove UV-induced lesions, the UV-damaged DNA endonuclease (UVDE)-dependent excision repair (UVER) pathway (Yasuhira and Yasui, 2000). Therefore it is possible that *S. pombe* could harbour other novel DNA repair pathways that could be repairing damage before rad16 localisation becomes necessary to carry out repair via the typical NER pathway.

4.4.5 Alternative assays for localisation studies

Alternative assays may be useful for protein localisation studies; these include co-immunoprecipitation (co-IP) and western blotting (WB), both of which involve cell lysis followed by probing for proteins using antibodies. For these assays, cells are lysed into nuclear and cytoplasmic fractions and protein localisation is identified by development of bands in a dark room using a chemiluminescent substrate. These can be viewed subjectively or quantified by densitometry (Bronstein *et al.* 1992).

During WB, cells are lysed into cellular fractions and polypeptides are separated by electrophoresis in polyacrylamide gels. Proteins are isolated from electrophoretic gels and blotted directly onto membranes, typically by electro-blotting which has a high degree of speed and efficiency (Kurien and Scofield, 2006). This potentially increases accuracy of protein yield, prior to probing with antibodies, which may be advantageous when compared with immunofluorescence, which requires optimisation of several steps for effective antigen detection. If any one of the fixation, permeabilisation and blocking steps are not efficient enough, antigen detection is compromised (Mosedale *et al.* 1996).

However, similar to immunofluorescence, the specificity and sensitivity of these assays is largely dependent on the antibodies used. For this reason, the quantitation of protein localisation by WB is limited to the activity of the antibody, meaning that co-IP and WB are most likely alternative assays with equally limited sensitivity and specificity (Chang and Lovett, 2011).

Chapter Five – Rate of repair in wild type and variant *ERCC4* cell lines after DNA damage

5.1 Introduction

Mutations in XP genes have been linked to defects in numerous repair pathways, including NER, DSB repair and ICL repair by homologous recombination (HR). DNA repair studies in a variety of model organisms carrying mutations in XP genes are outlined in Table 5.1.

A study designed to assess XPF function *in vivo* using homozygous XPF-deficient mice and a separate group of homozygous ERCC1-deficient mice revealed severe postnatal growth defects followed by premature death at three weeks of age. Embryonic fibroblasts isolated from these mice were hypersensitive to UV irradiation and mitomycin C treatment, indicative of an NER defect (Tian *et al.* 2004). Other studies using fibroblasts from homozygous *ERCC4*- and *ERCC1*-deficient mice have demonstrated a role for the ERCC1-XPF complex in processing ICL-induced DSBs in preparation for repair by HR. This was indicated by chromatid fusions and inhibition of downstream DSB repair in mutant cell lines after mitomycin C treatment. Identical effects were shown in *ERCC1*- and *ERCC4*-deficient mice, consistent with the functional link between ERCC1 and XPF (Niedernhofer *et al*. 2004). This was supported by further work by Zhang *et al.* (2007) using a combination of heterozygous human lymphoid cell lines and homozygous mouse fibroblast cells, who reported that ERCC1-XPF acts in conjunction with the mismatch repair protein MSH2 to uncouple the ICL lesion prior to HR. This suggested a more specific role for ERCC1-XPF in HR, in addition to recruitment to the DSB site.

The majority of repair studies investigating ERCC1 and XPF were carried out using biallelic mutant models. We hoped to further previous work by investigating the rate of repair after damage in cell lines carrying heterozygous variants in *ERCC4,* which may affect function of the ERCC1-XPF complex responsible for 5’ incision during NER.

Propidium iodide (PI) is a fluorescent dye that permeates the damaged cell membranes of non-viable cells, making it ideal for exclusion of non-viable

Gene (Protein)	Model organism	Effect seen	References
***XPA* (XPA)**	Human *XPA*-deficient (heterozygous) lymphoblast and fibroblast cell lines	Reduced affinity between XPA and ERCC1, required for correct positioning of ERCC1-XPF during NER, leading to defective NER	Li *et al*. 1995
***XPA/ERCC2* (XPA/XPD)**	Mice carrying double homozygous mutations in *XPA* and *ERCC2*	Accelerated ageing phenotype when *XPA* and *ERCC2* mutations were combined	De Boer *et al*. 2002
***XPE* (DDB2)**	Heterozygous *XPE* human lymphoblastoid and fibroblast cell lines	No degradation of p48, suggesting a functional connection between UV-damaged DNA binding protein complex (UV-DDB binding) and proteasomal degradation of p48 in early NER, which is impaired with *XPE* mutation	Rapic-Otrin *et al*. 2002
***ERCC4* and *ERCC1* (XPF and ERCC1)**	*ERCC4*- and ERCC1-deficient mice (homozygous mutant)	Severe postnatal growth defect resulting in death after 3 weeks; embryonic fibroblasts showed hypersensitivity to UV and mitomycin C; effects seen in mice deficient for *ERCC1* and *ERCC4* independently	Tian *et al*. 2004
***ERCC4* and *ERCC1* (XPF and ERCC1)**	Homozygous *ERCC4* and *ERCC1* embryonic stem cell fibroblasts isolated from mice	Mitomycin C treatment caused chromatid fusions and inhibition of downstream DSB repair, highlighting a role in DSB repair.	Niedernhofer *et al.* 2004
***ERCC4* and *ERCC1* (XPF and ERCC1)**	Mutant heterozygous *ERCC4* and *ERCC1* human lymphoid cell lines, homozygous *ERCC4*- and *ERCC1*-deficient mouse fibroblasts	Cell lines were deficient in ICL-induced HR, suggesting that ERCC1-XPF has a role more specific to HR of ICL-induced DSBs, rather than being recruited to the site of DSBs only.	Zhang *et al.* 2007
***TTD-A* (Smallest subunit of TFIIH)**	Homozygous fibroblasts isolated from human patients with photosensitive TTD-A	Severely delayed repair of UV damage in TTD-A mutant cells compared with wild type cells, which could explain relatively mild photosensitivity suffered by patients with TTD-A.	Theil *et al.* 2011
***ERCC2* (XPD)**	Fibroblasts from a patient with TTD caused by homozygous *XPD* mutations	Deficient in repair of both CPDs and (6-4)PPs after UV irradiation, suggesting that cancer-free phenotype in TTD may be linked to a transcription defect.	Nishiwaki *et al.* 2004; Berneburg *et al.* 2000
***ERCC2* (XPD)**	Baculovirus expression system transfected with heterozygous *ERCC2* variants	Normal DNA repair and transcription, suggesting that single polymorphisms in *ERCC2* may be benign; numerous polymorphisms required for cancer-prone phenotype in XP.	Laine *et al.* 2007

Gene (Protein)	Model organism	Effect seen	References
ERCC2 **(XPD)**	DNA isolated from female Caucasian participants (heterozygous and homozygous variants identified at codons 312 and 751)	Homozygous *ERCC2* mutations at the Lys/Lys codon 751 locus associated with significantly reduced repair of X-ray-induced DNA damage	Lunn *et al.* 2000
ERCC8 **(CSA)**	DNA isolated from dermal fibroblasts of a patient with CSA (homozygous or dominant heterozygous for *ERCC8* variant)	Impaired removal of UV-induced lesions by TC-NER, without affecting the protein's role in oxidative damage response. Variant could be homozygous or dominant heterozygous for mutation (paternal DNA not available for sequencing).	Nardo *et al.* 2009
ERCC8/ ERCC6 **(CSA/CSB)**	Human wild type, CSA, CSB and XPA fibroblast cell lines (heterozygous)	Reduced transcription (indicated by reduced RNA synthesis) after UV damage in CS cell extracts correlated with a marked reduction in RNA polymerase II (RNAPII); suggesting that repression of transcription initiation is partially responsible for inhibition of transcription after damage in CS patients.	Rockx *et al.* 2000
ERCC8/ ERCC6 **(CSA/CSB)**	*ERCC6*- and *ERCC8*-deficient embryonic fibroblasts and stem cells isolated from mice (homozygous mutant)	Distinct differences between CSA and CSB in terms of sensitivity to oxidative damage by gamma ray treatment, despite similar clinical manifestations of CSA and CSB patients; this suggests a role of CSB in repair of oxidative DNA damage.	De Waard *et al.* 2004
ERCC5 **(XPG)**	Fibroblasts from non-sun exposed skin isolated from siblings with XP (homozygous variants in *ERCC5)*	Corrected repair of oxidative stress but not UV damage; this suggests that the XP phenotype in these patients is the result of novel *ERCC5* mutations and their inability to repair UV damage.	Soltys *et al.* 2013

Table 5.1 Summary of DNA repair studies in model organisms carrying mutations in NER genes.

cells during FACS (Banerjee *et al*. 2014). There are a number of examples in the literature where removal of non-viable cells *in vitro* has been used to improve experimental conditions. Rogers *et al.* (2010) showed that bacterial DNA-based techniques are unable to differentiate between viable, non-viable bacteria and extracellular DNA, resulting in false positive results and an inability to predict therapy response in cystic fibrosis patients, requiring removal of non-viable DNA prior to analysis. Additionally, Gregory *et al* (2009) demonstrated that removal of non-viable cells greatly improves antibody production *in vitro*.

UV light produces intrastrand crosslinks of two adjacent pyrimidines, resulting in cyclobutane pyrimidine dimers (CPDs) and 6-4 photoproducts [(6-4)PPs] (Dreze *et al.* 2014). (6-4)PPs are highly distorting lesions that can be rapidly repaired by TC- NER, whereas CPDs are able to form base pairs with the opposite DNA strand, resulting in slower repair by global genome (GG)-NER (Hyka-Nouspikel *et al.* 2012). Several studies have shown that the ratio of overall formation of CPDs to (6-4)PPs by UV-C light is 3:1 (Mitchell, 1988), making CPDs the preferable lesion to probe for to investigate repair rates.

Here, we sought to identify the functional consequences of *ERCC4* variants on repair capacity after DNA damage by enzyme-linked immunosorbent assay (ELISA). We administered cell lines with UV-C irradiation and oxaliplatin to examine the repair rate of lesions in wild type and variant cell lines.

5.2 Materials and Methods

5.2.1 UV irradiation

During optimisation, wild type cells were seeded onto poly-L-lysine coated dishes and UV (20J) treated in SF media, prior to ELISA using a commercial kit to detect CPD formation within cells. Wild type cells and cells carrying Pro379Ser were irradiated in suspension with different UV doses (20-400J) over 48 hours under different serum starving conditions. UV irradiation was carried out with cells suspended in SF media and PBS, and cells were UV treated using a germicidal lamp and a crosslinker for comparison.

Optimised conditions were as follows: cells were centrifuged at 1,200 RPM for five minutes to produce a cell pellet. The supernatant was removed and cells were suspended in SF media. Cells were irradiated with UV-C (70J) in large single well dishes using a UV crosslinker, followed by immediate re-suspension in normal tissue culture media containing FBS and antibiotics. Aliquots of cell suspension were removed at 0, 4, 24 and 48 hours after UV treatment.

5.2.2 Oxaliplatin treatment

During optimisation, wild type cells were treated with different concentrations (1.25-10µg/ml) of primary cisplatin modified-DNA antibody and for different incubation times (1-3 hours), prior to DNA extraction and assay by ELISA.

Optimised conditions were as follows: cells were suspended in SF media by centrifugation as outlined in 5.2.1, followed by immediate addition of oxaliplatin (125µM) to tissue culture media for 1 hour at 37°C. After treatment, cells were re-suspended in normal media containing FBS and antibiotics and aliquots of cell suspension were removed at 0, 4, 24 and 48 hours.

5.2.3 FACS

Aliquots of cell suspension were sorted by FACS as described in section 2.5.3.4. Only viable cells were retrieved in normal tissue culture media containing FBS and antibiotics (purity > 98%) prior to DNA extraction.

5.2.4 DNA extraction

Genomic DNA was extracted from viable cell suspensions using a commercial extraction kit according to the manufacturer's instructions (Qiagen; Section 2.5.3.5). DNA was eluted in 20µl elution buffer and stored at -20°C for long-term storage.

5.2.5 ELISA for CPD detection

DNA samples were probed for CPDs using a commercial ELISA kit according to the manufacturer's instructions (Cell BioLabs). Briefly, DNA

samples and a DNA standard conjugated to a known amount of CPDs were converted to single stranded (ss) DNA by incubation at 96°C for ten minutes followed by rapid chilling on ice for ten minutes. DNA was diluted to 200ng/ml in cold PBS and incubated overnight at 4°C prior to ELISA.

Wells were washed twice with cold PBS, followed by hour-long incubations at room temperature with assay diluent, primary CPD antibody, blocking reagent and secondary horseradish peroxidase (HRP)-conjugated antibody, with wash steps between each incubation. Substrate solution was added to each well and incubated for 20 minutes at room temperature, before addition of stop solution to terminate the enzyme reaction. Absorbance was read at 450nm using a plate reader, with a reference range of 620nm.

DNA samples from treated and untreated cell lines were assayed for CPDs by ELISA and CPD formation was plotted against time. Duplicate DNA samples from wild type and variant cell lines with and without UV treatment were assayed on separate ELISA plates and CPD levels were averaged. This initial data was then validated by repeating UV treatments across all cell lines and again assaying duplicate DNA samples on separate plates.

5.2.6 ELISA for oxaliplatin adduct detection

DNA samples were probed for oxaliplatin adducts using the same commercial ELISA kit in 5.2.5, with substituted primary and secondary antibody. Antibodies used were a rat cisplatin modified-DNA primary antibody and goat anti-rat HRP-conjugated secondary antibody (both purchased from Abcam). Absorbance was read in the same way as in 5.2.5.

There is no commercially available antibody for detection of oxaliplatin-DNA adducts or a cisplatin-DNA standard to allow quantification of lesions by ELISA. As a result, oxaliplatin lesions were quantified as arbitrary units (AU) based on the absorbance reading of each sample. The cisplatin-modified DNA primary antibody was selected because it probes for the same epitope as that produced by oxaliplatin adduct formation.

5.2.7 Statistical analysis

Statistical analysis was performed in SPSS v.20.0 using a two-way ANOVA, with mutation status (wild type, Pro379Ser, Arg576Thr and Glu875Gly) and treatment (no treatment and UV treatment) as the independent variables. The dependent variable was CPD formation (ng/ml) as a measure of DNA damage. CPD formations for all cell lines carrying the same *ERCC4* variant were averaged to allow comparison of CPD formation with and without UV treatment for individual cell groups. Individual ANOVAs were run on CPD formations acquired at 24 and 48 hours. The duplicate values for cell groups acquired during the initial test experiment were analysed independently of duplicate values acquired during validation. Only CPD values from samples run on the same ELISA plate were compared with one another, which meant that one set of wild type samples were analysed with Pro379Ser and Arg576Thr samples, and another set of wild type samples were analysed with Glu875Gly samples. Oxaliplatin treated samples showed no difference in the primary results, therefore no statistics were applied to this data set.

5.3 Results

5.3.1 Optimisation of UV assay

A summary of the optimisation stages is shown in Figure 5.1. Wild type cells were seeded onto tissue culture dishes coated with poly-L-lysine, fibronectin and laminin and allowed to form a monolayer overnight, prior to UV (20J) treatment in SF media. Cells were probed for CPD formation over 24 hours using a cellular ELISA kit (Cell BioLabs), which resulted in low CPD formation. Microscopic examination of the dishes indicated that the vast majority of cells were washed off during the wash steps of the protocol, irrespective of adhesion matrix used (data not shown).

We therefore extracted DNA from cells for analysis. Wild type cells and cells carrying Pro379Ser were irradiated with UV (20J) using a germicidal lamp in SF media after overnight starvation. DNA was extracted at 0, 4 and 24 hours from UV treated and untreated cell lines and assayed for CPD formation using

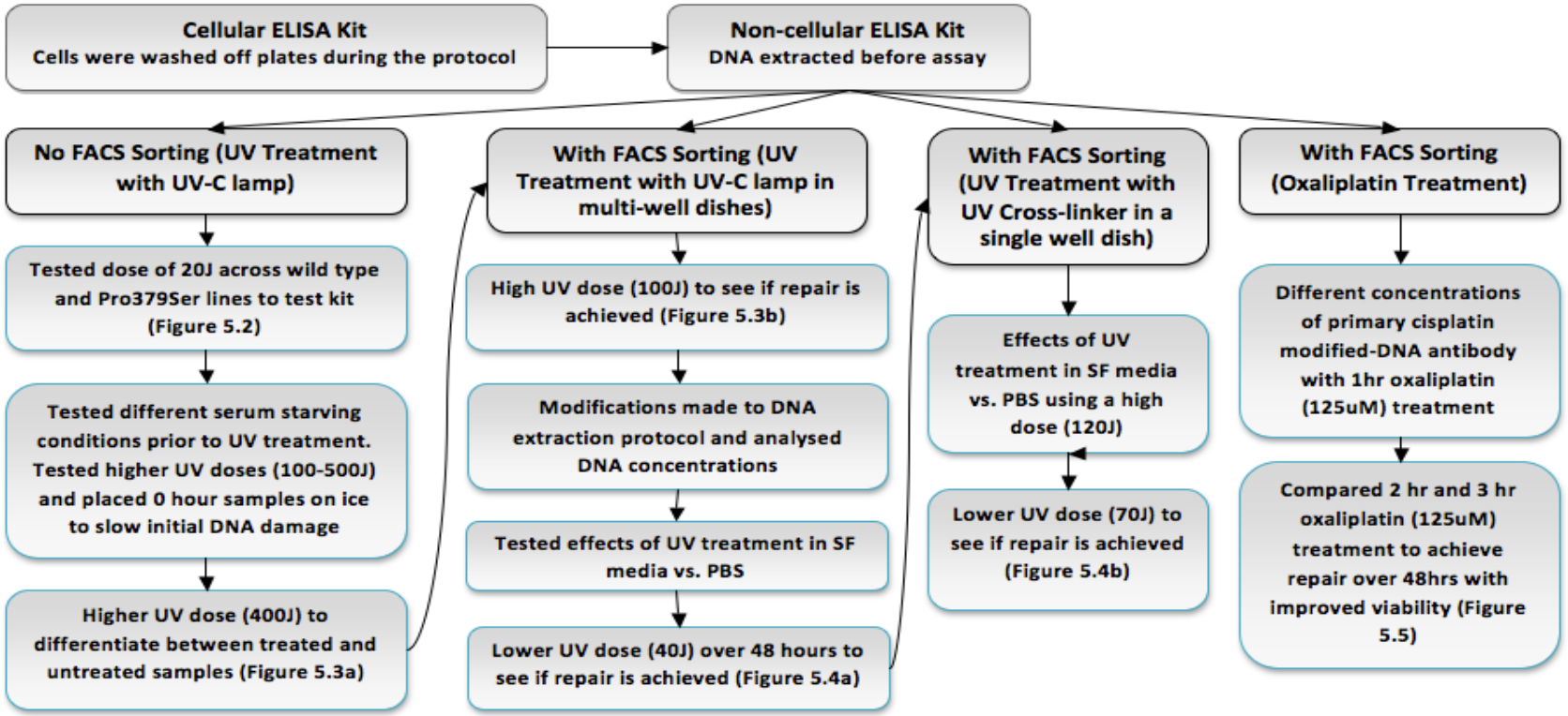


Figure 5.1 Schematic diagram of optimisation of both UV and oxaliplatin ELISAs. A cellular ELISA kit for CPD detection was tested in cells seeded onto tissue culture plates, which showed cells being removed from plates during wash steps. A non-cellular ELISA kit was therefore used, requiring DNA extraction prior to ELISA. Initial experiments using this kit showed that the ELISA could differentiate between untreated and UV irradiated samples, but no repair was seen. We incorporated fluorescence activated cell sorting (FACS) to remove non-viable cells prior to DNA extraction, which revealed repair after UV treatment. However other technical challenges arose, resulting in substitution of the original germicidal lamp for a UV cross-linker to allow a single UV dose per experiment, which allowed collection of experimental data. The non-cellular ELISA kit was modified for oxaliplatin adduct detection by substituting primary and secondary antibodies. Various concentrations of primary antibody and duration of oxaliplatin treatment (1-3hr) were tested, resulting in optimum conditions for the oxaliplatin adduct ELISA.

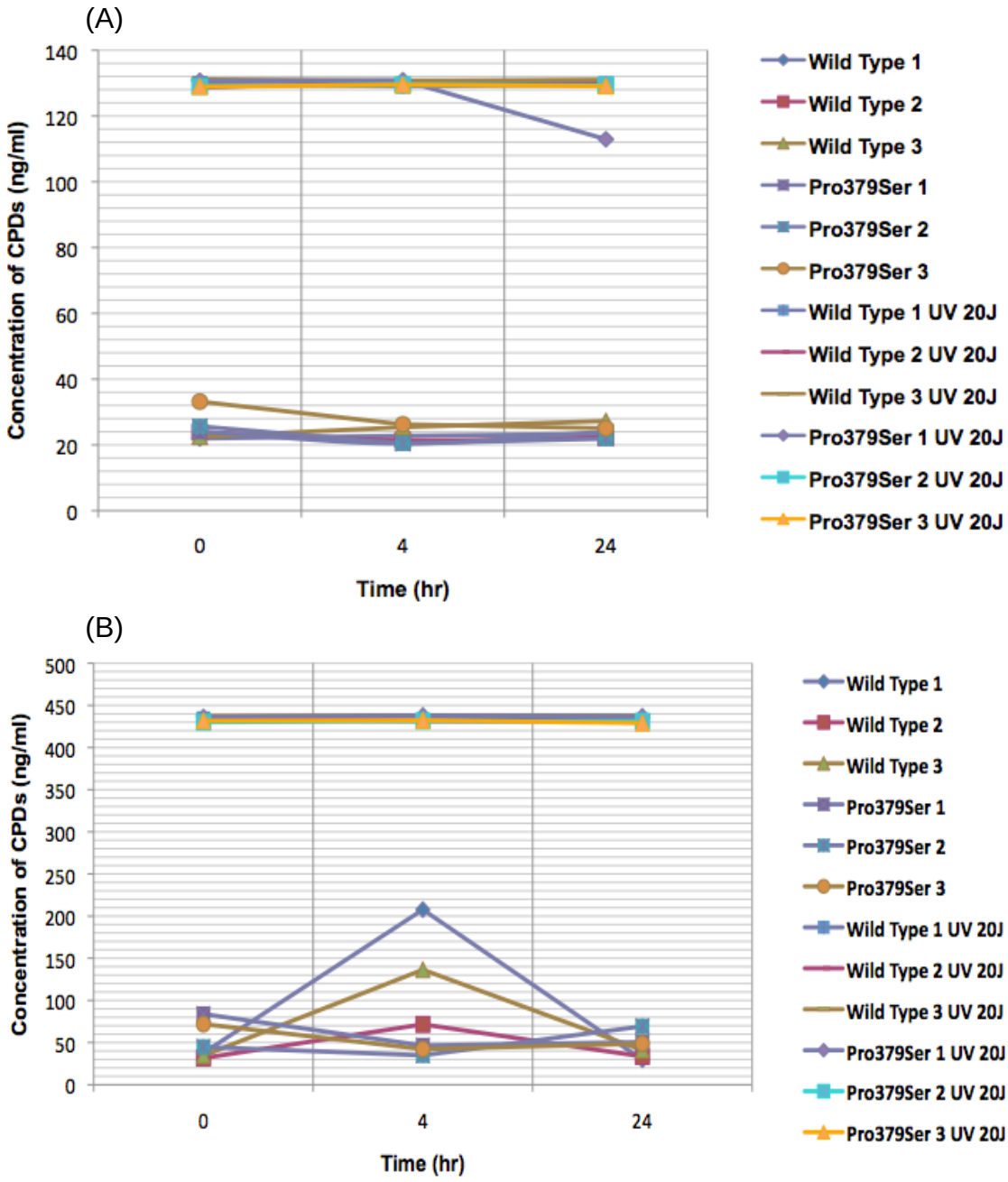


Figure 5.2 Initial testing of UV treated and untreated DNA samples in the commercial ELISA kit. Wild type and cell lines carrying Pro379Ser were incubated in SF media overnight prior to UV irradiation (20J). DNA was extracted at 0, 4 and 24hrs post-treatment and assayed for CPDs using a commercial ELISA kit. CPD formation (ng/ml) was plotted against time (hr). The data shown in A) and B) were acquired in separate experiments under identical conditions, highlighting a discrepancy in the amount of damage cells receive when using the UV-C germicidal lamp (maximum CPD formation of approximately 130 and 440ng/ml in A) and B) respectively). This data indicated that the ELISA kit can differentiate between UV irradiated and untreated DNA samples, but further improvements were needed to address the inconsistencies in CPD formation and lack of repair.

the non-cellular CPD ELISA kit. Separate experiments under identical conditions revealed that the non-cellular ELISA kit was capable of differentiating between untreated and treated cells (Figure 5.2), although no repair was seen. Additionally, there were discrepancies in the maximum CPD formation seen under identical experiments, suggesting a potential problem with reproducibility of UV irradiation when using the germicidal lamp (Figure 5.2). We tested higher UV doses (100-500J) in SF media without prior starvation in wild type cell lines and carried out DNA extraction at 0, 0.5, 4, 24 and 48 hours. Cells used for DNA extraction at 0 hours were placed on ice immediately after UV irradiation, in an attempt to slow the damage process. This had no effect on reducing the initial CPD formation seen at 0 hours with UV treatment, and again no repair was seen (data not shown).

Wild type cell lines and cells carrying Pro379Ser were irradiated with high dose UV (100-400J) in SF media with and without fluorescence activated cell sorting (FACS) to remove non-viable cells prior to DNA extraction and ELISA. We observed no apparent repair over 24 hours without FACS. However, once FACS was incorporated, repair was seen between 4 and 48 hours after UV treatment (Figure 5.3). However, when conditions were tested in remaining cell lines, results were still inconsistent (data not shown).

In an attempt to reduce these inconsistencies, we (i) modified the DNA extraction protocol to prevent carry-over of wash buffers into eluted DNA, (ii) analysed the accuracy of DNA extractions and (iii) tested the effect of SF media and PBS on CPD formation after UV treatment. These modifications still produced inconsistent results, despite revealing that the DNA concentrations by nanodrop were accurate. During FACS we observed a higher viability of cells treated in SF media compared with PBS, which made the FACS process quicker and the DNA yield higher. Following FACS and dilution of DNA prior to ELISA, we observed comparable CPD formation in UV irradiated samples, irrespective of suspension medium used (data not shown).

We considered whether the method of UV irradiation could be the reason why no consistent repair had been demonstrated, even after removal of non-viable cells by FACS. When using a germicidal UV-lamp for UV irradiation,

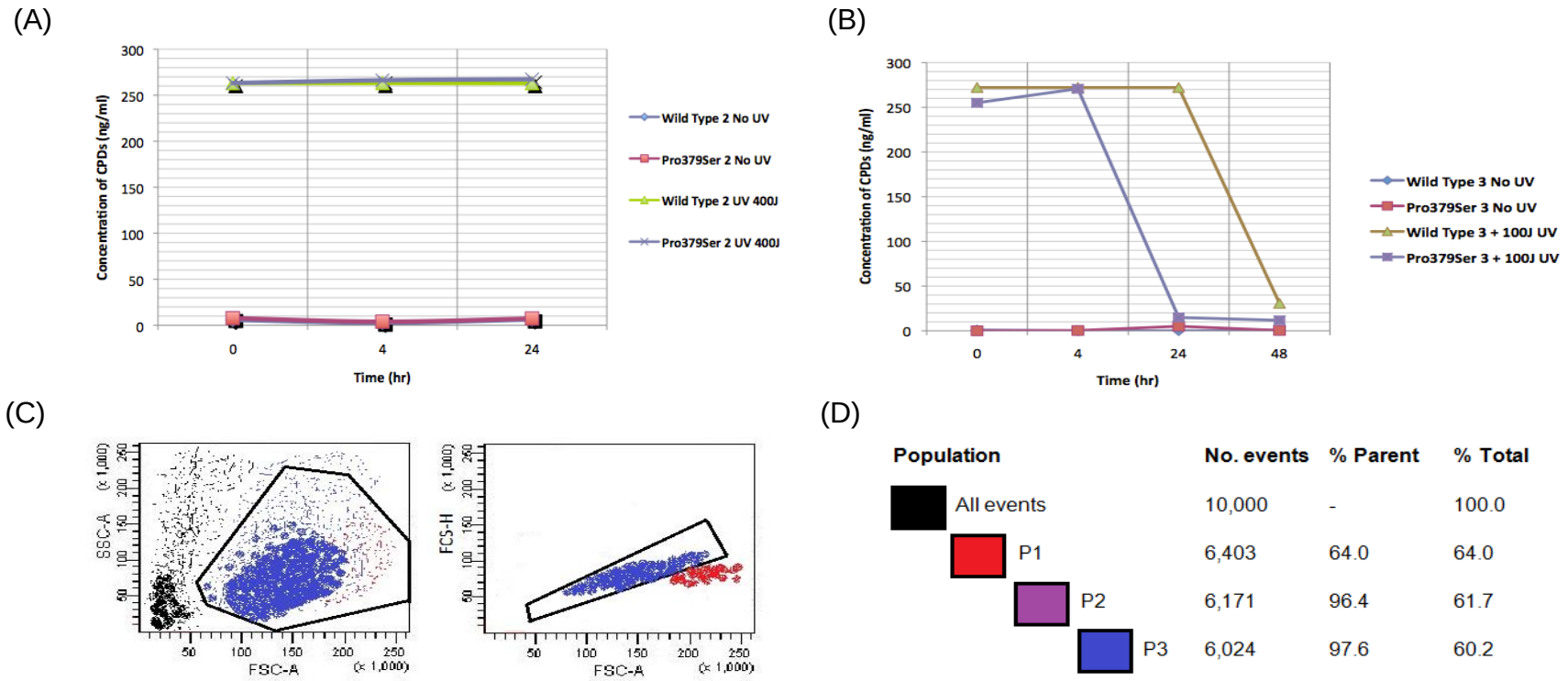
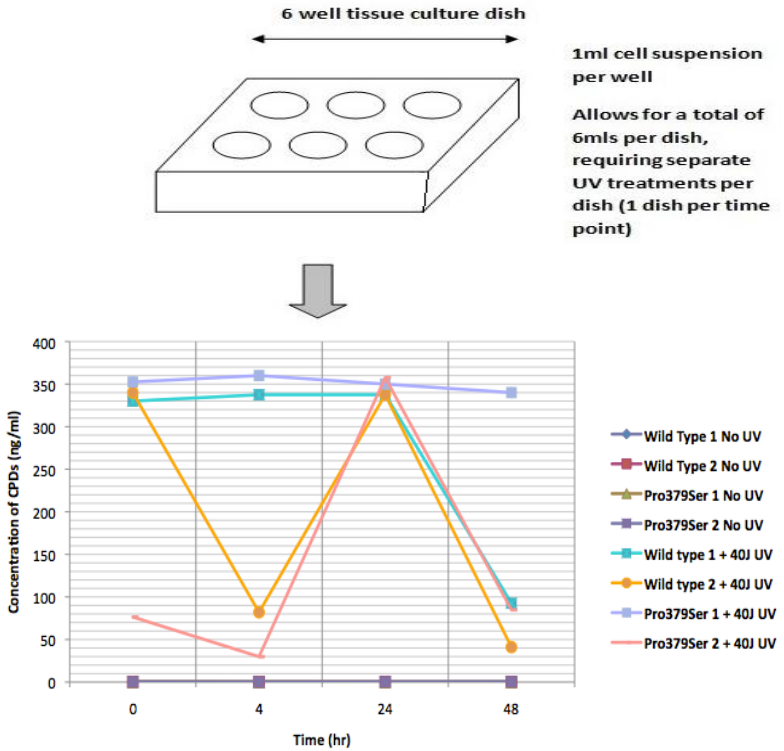


Figure 5.3 Incorporation of FACS to remove non-viable cells before DNA extraction. Wild type cells and cells carrying Pro379Ser were treated with high dose UV in SF media and incubated for 24 and 48 hours. DNA was extracted at the time points shown A) without FACS and B) after FACS prior to ELISA (sample purity 98-99% at time of sorting). In both experiments the kit was able to differentiate between treated and untreated samples. However, repair between 4 and 48 hours was only seen after removal of non-viable cells by FACS prior to DNA extraction. During FACS, cells are analysed by forward scatter according to area and height (FSC-A and FSC-H, respectively) and side scatter (SSC-A) and gated to allow passage of only viable cells through the sorting apparatus, as shown in (C). The number of events and percentage of parent and total populations at each gating stage (P1-3) are shown in (D). P1 = cells selected based on morphology (FSC-A/FSC-H versus SSC-A); P2 = single unstained cells; P3 = negative (viable) cells.

(A)



(B)

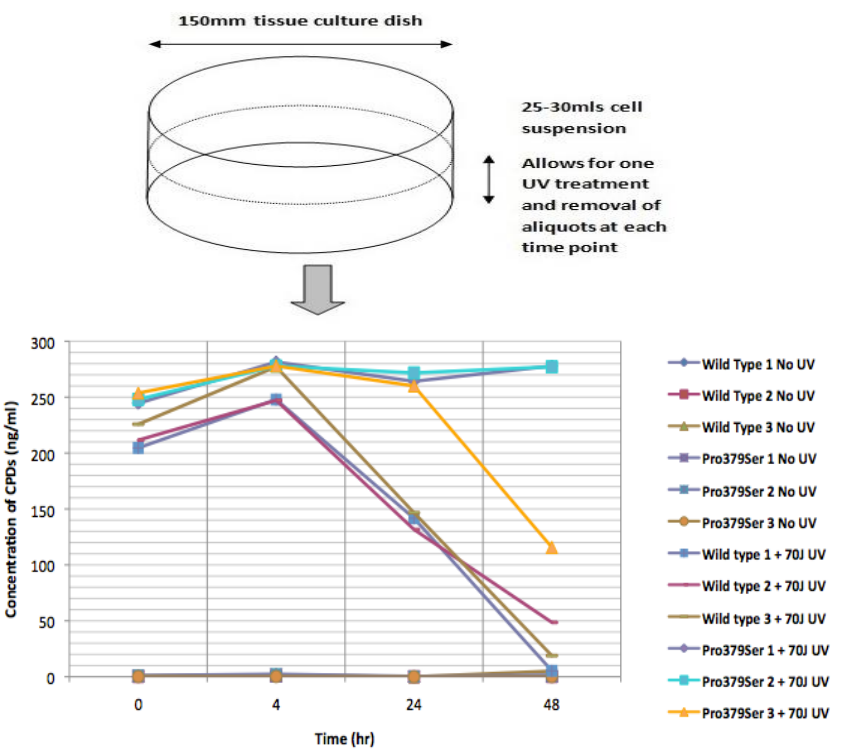


Figure 5.4 Substitution of a UV-C germicidal lamp for a UV-crosslinker to produce more consistent results. Wild type cells and cells carrying Pro379Ser were treated with (A) UV (40J) using the UV-C germicidal lamp and (B) UV (70J) using a UV-crosslinker in SF media. The substitution of the UV-crosslinker enabled a switch to large tissue culture dishes, allowing for one uniform UV treatment of the entire cell suspension. At each time point, cells were sorted by FACS and DNA was extracted using the modified method, prior to assay for CPDs by ELISA. Individual UV treatments per time point using the germicidal lamp produced inconsistent results (A), suggesting that a uniform UV treatment using the UV-crosslinker followed by removal of aliquots at each time point might be a better approach. This modification to the method produced more consistent results (B).

there was a maximum area (a circle of approx. 100mm diameter) that can be irradiated to ensure uniform treatment of all cells within the suspension. This required a separate tissue culture dish per time point and several different UV treatments per experiment. Alternatively, a UV crosslinker had a much larger area for uniform UV irradiation (approx. 200 x 300mm). We compared UV (40J) treatment using a germicidal lamp with UV (70J) treatment using a UV crosslinker in wild type cells and cells carrying Pro379Ser. Cells irradiated with the germicidal lamp showed inconsistent results (Figure 5.4a). UV treatment using a crosslinker in a single dish followed by removal of aliquots at each time point produced more consistent results, demonstrating repair in all wild type cell lines between 4 and 48 hours after treatment, with delayed repair in all cell lines carrying Pro379Ser (Figure 5.4b).

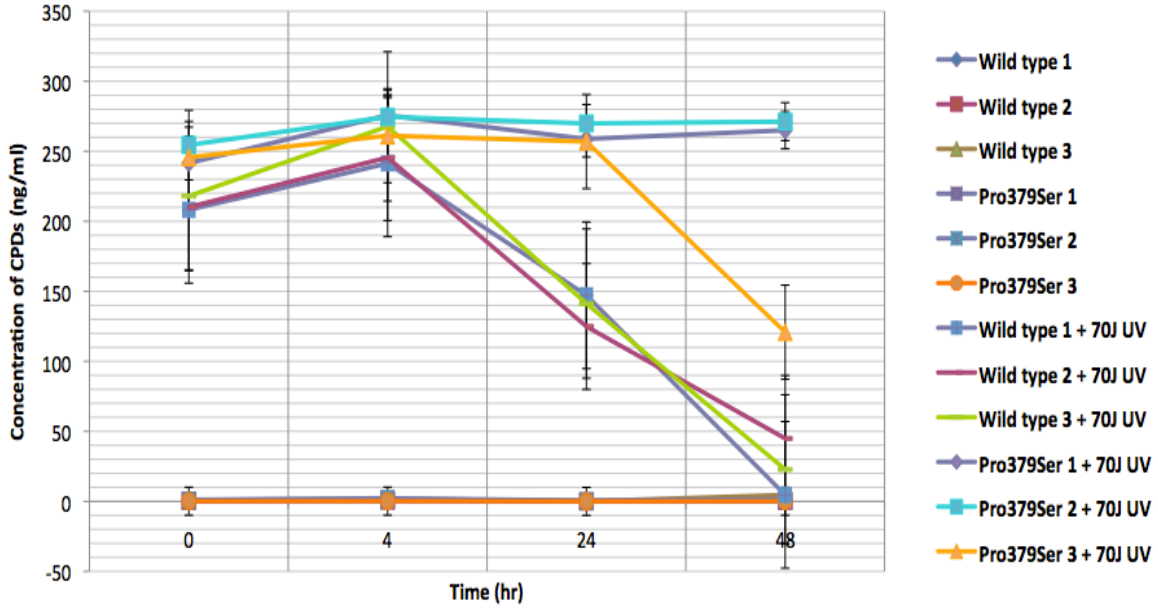
5.3.2 Effect of UV light on repair in variant cell lines

Wild type and *ERCC4* variant cell lines were irradiated with UV (70J) light in SF media and CPD formation was measured by ELISA over 48 hours. All cell lines without UV treatment showed negligible CPD formation (<1ng/ml) at all time points.

Wild type cell lines with UV treatment achieved CPD levels of approximately 210-260ng/ml at 0 and 4 hours, followed by noticeable repair between 4 and 48 hours, reaching comparable levels with untreated samples after 48 hours, with minimal variation between individual wild type cell lines. Cell lines carrying Pro379Ser with UV treatment achieved similar CPD levels to wild type cell lines at 0 and 4 hours, but failed to repair lesions after 48 hours in two cell lines (Pro379Ser 1 and 2) and demonstrated delayed repair in one cell line (Pro379Ser 3) during the test experiment (Figure 5.5a). The Pro379Ser 3 cell line began repair after 24 hours, indicated by CPD levels of 257ng/ml and 121ng/ml at 24 and 48 hours respectively. Validation of this data showed very similar results, with the exception of a lack of repair after 48 hours in the Pro379Ser 3 cell line (Figure 5.5b).

Cell lines carrying Arg576Thr treated with UV similarly showed relatively stable levels of CPDs between 0 and 24 hours before commencement of repair, indicated by CPD formation ranges of approximately 259-260ng/ml and 104-

(A)



(B)

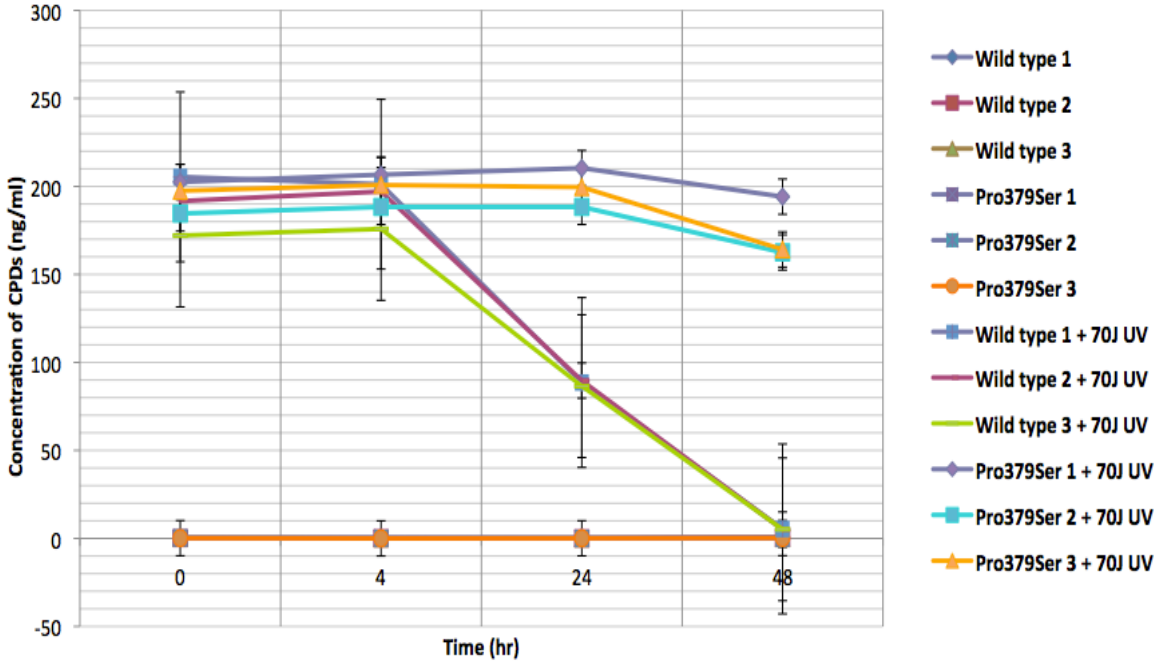
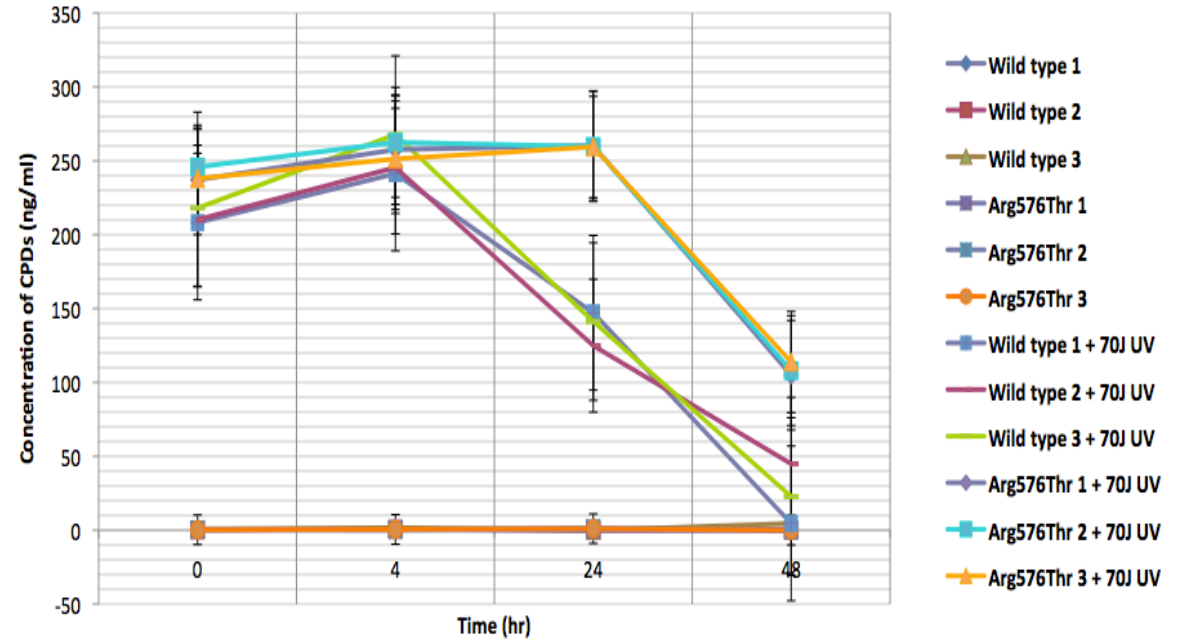


Figure 5.5 Effect of UV treatment on repair in cell lines carrying Pro379Ser. Average CPD formation (ng/ml) in wild type cells and cells carrying Pro379Ser in *ERCC4* with and without UV-C (70J) irradiation, showing the initial (A) and validation (B) experiment. CPD concentrations for each experiment are plotted as an average of two duplicate samples from the same experiment run on separate ELISA plates, which are plotted against time with standard error bars. The test and validation experiments were biological repeats of the same experiment performed on separate days under identical conditions.

(A)



(B)

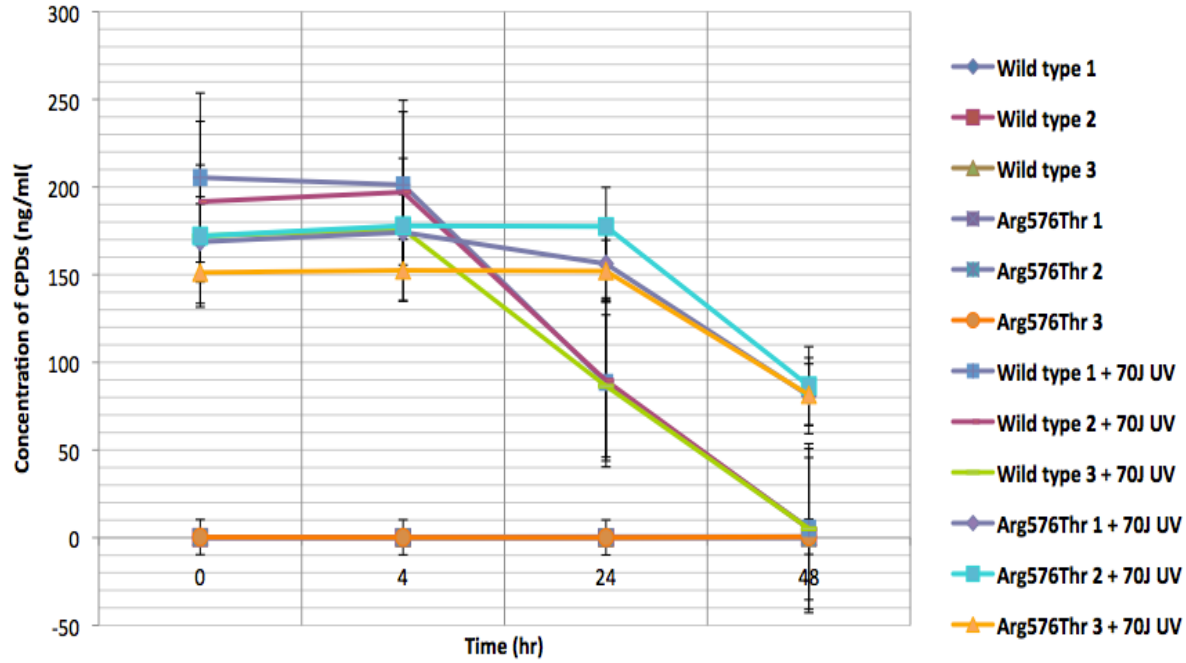


Figure 5.6 Effect of UV treatment on repair in cell lines carrying Arg576Thr. CPD formation (ng/ml) in wild type cells and cells carrying Arg576Thr in *ERCC4* with and without UV-C (70J) irradiation, showing the initial (A) and validation (B) experiment. CPD concentrations for each experiment are plotted as an average of two duplicate samples from the same experiment run on separate ELISA plates, which are plotted against time with standard error bars. The test and validation experiments were biological repeats of the same experiment performed on separate days under identical conditions.

113ng/ml at 24 and 48 hours respectively (Figure 5.6a). Validation of this data showed similar results (Figure 5.6b).

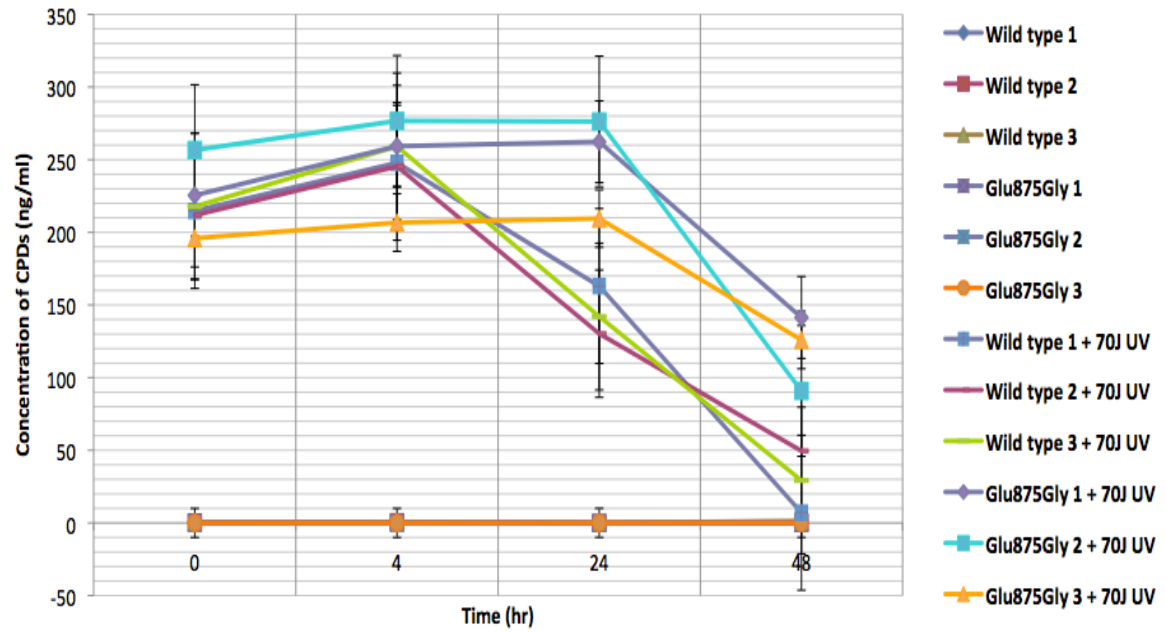
Cell lines carrying Glu875Gly treated with UV showed a very similar repair trend to cells carrying Arg576Thr, with high levels of CPDs between 0 and 24 hours and repair between 24 and 48 hours, indicated by CPD level ranges of 209-276ng/ml and 90-141ng/ml at 24 and 48 hours respectively (Figure 5.7a). Validation of this data showed similar results (Figure 5.7b).

Statistical analysis of CPD formation in cell groups was performed by two-way ANOVA, as separate analyses for the initial test experiment and the validation experiment. Comparison of each cell group based on association between *ERCC4* mutation and treatment type during the initial test experiment revealed a statistically significant association at both 24 and 48 hours (P = 0.000; Table 5.2). Further investigation using a post hoc Tukey test identified a statistically significant difference in CPD formation between wild type cells and all *ERCC4* variant cell groups (P = 0.000) at 24 hours. Additionally, after 48 hours, statistical significance was identified between Pro379Ser and Arg576Thr cell groups (P = 0.000), and between Pro379Ser and Glu875Gly cell groups undergoing UV treatment (P = 0.001; Table 3.2). Statistical analysis of CPD formations acquired during validation showed very similar results, with additional statistically significant differences between Pro379Ser and Arg576Thr cell groups (P = 0.000), and between Pro379Ser and Glu875Gly cell groups (P = 0.000) after 24 hours (Table 3.3).

5.3.3 Optimisation of oxaliplatin assay

Wild type cells were treated with oxaliplatin (125µM) for 1 hour in SF media, followed by incubation in normal media for up to 48 hours. At 2 and 4 hours after treatment, cells were sorted by FACS and DNA was extracted, followed by ELISA using substituted antibodies. This revealed low adduct formation in untreated cells and a peak in adduct formation after 2 hours in treated cells, indicating potential for the antibodies used to assay for oxaliplatin adducts (data not shown).

(A)



(B)

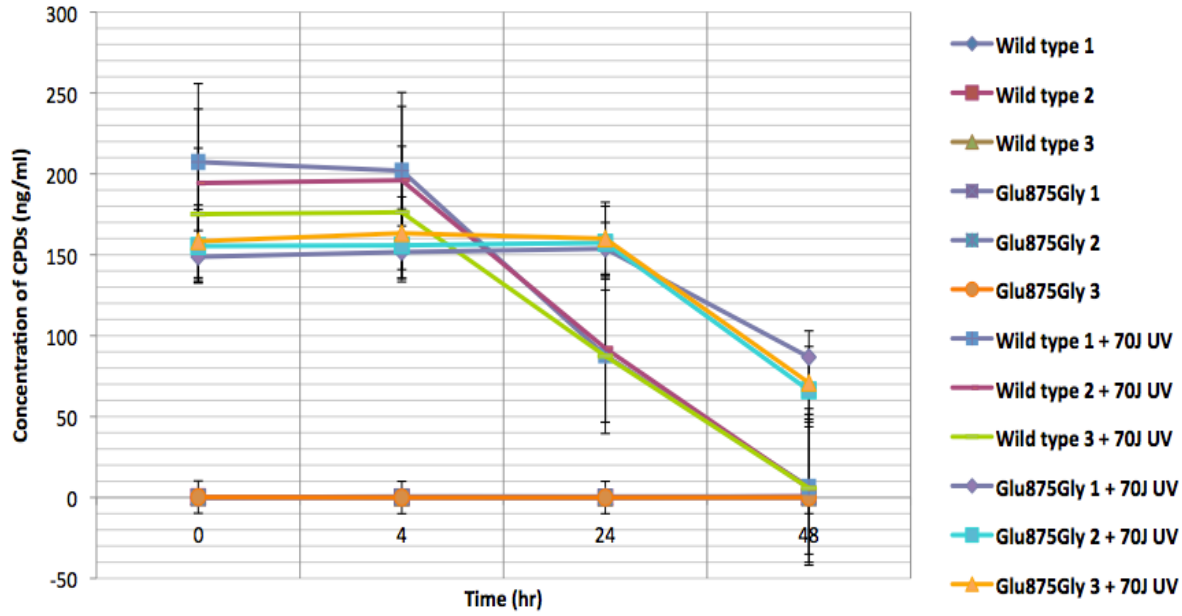


Figure 5.7 Effect of UV treatment on repair in cell lines carrying Glu875Gly. CPD formation (ng/ml) in wild type cells and cells carrying Glu875Gly in *ERCC4* with and without UV-C (70J) irradiation, showing the initial (A) and validation (B) experiment. CPD concentrations for each experiment are plotted as an average of two duplicate samples from the same experiment run on separate ELISA plates, which are plotted against time with standard error bars. The test and validation experiments were biological repeats of the same experiment performed on separate days under identical conditions.

A.

		24hr		48hr	
Mutation	**Treatment**	**Mean**	**SD**	**Mean**	**SD**
Wild Type	NT	0.0000	0.00000	1.8117	2.18417
	UV	137.7533	12.19263	24.1100	18.38379
Pro379Ser	NT	0.2267	0.23577	0.5550	0.64246
	UV	261.8800	7.46706	218.9933	76.70514
Arg576Thr	NT	1.0567	0.46620	0.0000	0.00000
	UV	259.7817	7.65487	108.9217	6.76938
Glu875Gly	NT	0.0183	0.03251	0.0100	0.01673
	UV	249.3883	32.44602	119.3883	26.15201

B.

	24hr	48hr
Mutation	0.000	0.000
Treatment	0.000	0.000
Mutation*Treatment	0.000	0.000

C.

Mutation		**P value (24hr)**	**P value (48hr)**
Wild Type	**Pro379Ser**	0.000	0.000
	Arg576Thr	0.000	0.007
	Glu875Gly	0.000	0.002
Pro379Ser	**Wild Type**	0.000	0.000
	Arg576Thr	0.999	0.000
	Glu875Gly	0.623	0.001
Arg576Thr	**Wild Type**	0.000	0.007
	Pro379Ser	0.999	0.000
	Glu875Gly	0.697	0.972
Glu875Gly	**Wild Type**	0.000	0.002
	Pro379Ser	0.623	0.001
	Arg576Thr	0.697	0.972

Table 5.2 Statistical analysis of cell groups during initial test experiment following UV treatment. CPD formations for cell lines carrying the same *ERCC4* variant with and without UV (70J) treatment were averaged and compared using a two-way ANOVA. (A) Mean and standard deviations for each cell group before and after UV treatment at 24 and 48 hours are shown. (B) P values obtained by ANOVA are given for CPD formation after 24 and 48 hours for the variables mutation, treatment and the association between mutation and treatment (mutation*treatment). Highlighted values were statistically significant (P<0.05). A post hoc Tukey test was performed to identify which groups significantly differed from one another at 24 and 48 hours, with significant p values (<0.05) highlighted. NT = no treatment; SD = Standard deviation.

A.

		24hr		48hr	
Mutation	**Treatment**	**Mean**	**SD**	**Mean**	**SD**
Wild Type	NT	0.2183	0.26716	0.3767	0.19075
	UV	88.2583	1.68842	5.1950	.93104
Pro379Ser	NT	0.0233	0.05715	0.1933	0.13895
	UV	199.4450	11.88586	173.6117	16.97074
Arg576Thr	NT	0.2517	0.29963	0.2533	0.34180
	UV	161.9433	13.86127	83.0417	5.65300
Glu875Gly	NT	0.1067	0.13779	0.0617	0.08256
	UV	157.0833	8.59053	74.6017	9.81068

B.

	24hr	**48hr**
Mutation	0.000	0.000
Treatment	0.000	0.000
Mutation*Treatment	0.000	0.000

C.

Mutation		**P value (24hr)**	**P value (48hr)**
Wild Type	**Pro379Ser**	0.000	0.000
	Arg576Thr	0.000	0.000
	Glu875Gly	0.000	0.000
Pro379Ser	**Wild Type**	0.000	0.000
	Arg576Thr	0.000	0.000
	Glu875Gly	0.000	0.000
Arg576Thr	**Wild Type**	0.000	0.000
	Pro379Ser	0.000	0.000
	Glu875Gly	0.827	0.468
Glu875Gly	**Wild Type**	0.000	0.000
	Pro379Ser	0.000	0.000
	Arg576Thr	0.827	0.468

Table 5.3 Statistical analysis of cell groups during the validation experiment following UV treatment. CPD formations for cell lines carrying the same *ERCC4* variant with and without UV (70J) treatment were averaged and compared using a two-way ANOVA. (A) Mean and standard deviations for each cell group before and after UV treatment at 24 and 48 hours are shown. (B) P values obtained by ANOVA are given for CPD formation after 24 and 48 hours for the variables mutation, treatment and the association between mutation and treatment (mutation*treatment). Highlighted values were statistically significant (P<0.05). A post hoc Tukey test was performed to identify which groups significantly differed from one another at 24 and 48 hours, with significant p values (<0.05) highlighted. NT = no treatment; SD = Standard deviation.

We tested longer drug incubation times and compared viability of cells during FACS. Wild type cells were treated with oxaliplatin (125µM) for 2 and 3 hours, followed by incubation in normal media for up to 48 hours. Removal of aliquots for FACS and DNA extraction followed by ELISA revealed a peak in oxaliplatin adduct formation at 4 hours, followed by repair between 4 and 48 hours in treated cells (data not shown). Additionally, during FACS we observed a higher viability in cells treated for 2 hours, resulting in a more rapid FACS process and a higher yield of DNA for ELISA.

5.3.4 Effect of oxaliplatin on repair in variant cell lines

Wild type and *ERCC4* variant cell lines were treated with oxaliplatin (125µM) for 2 hours, followed by incubation for 48 hours. DNA samples extracted from untreated and treated cell lines were assayed for oxaliplatin adducts by ELISA and adduct levels were plotted against time (Figure 5.8). We observed low adduct levels in all cell lines without treatment (<0.1 AU). Oxaliplatin treated cell lines demonstrated a peak in adduct formation at either 0 or 4 hours, followed by repair between 4 and 48 hours, with no noticeable difference in repair rate between wild type and variant cell lines. At 48 hours, all cell lines displayed oxaliplatin adduct levels comparable with untreated cell lines, regardless of *ERCC4* status. For all three variants in question, between 4 and 48 hours, variant cell lines consistently showed similar repair rates to wild type cell lines.

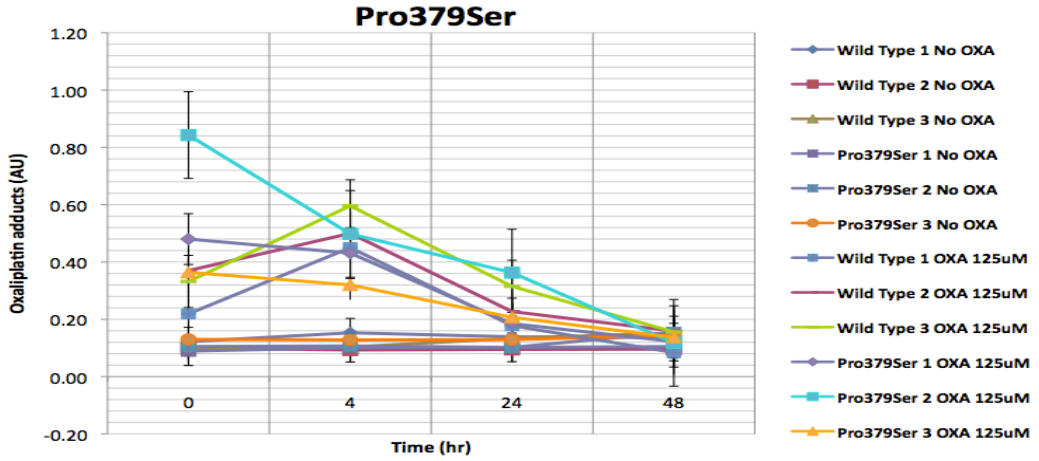
5.4 Discussion

5.4.1 Optimisation of ELISA

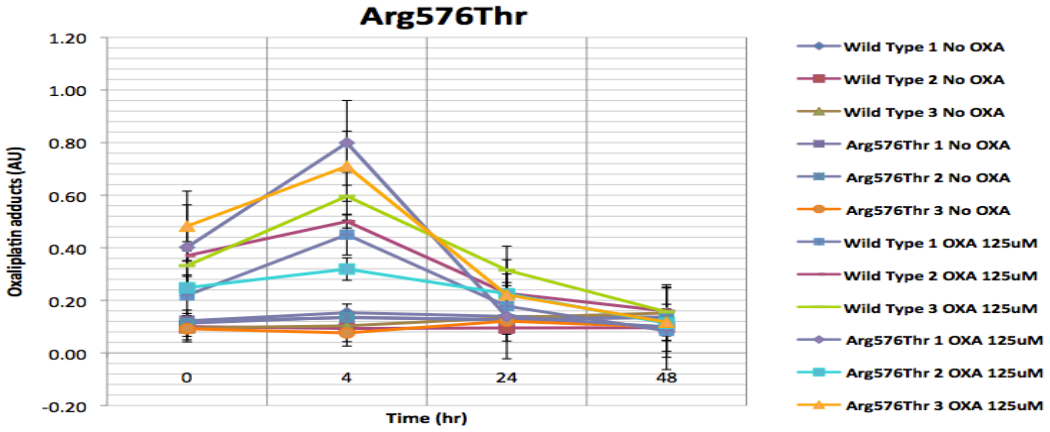
Previous experiments outlined in chapter 4 have shown that our cell lines do not adhere successfully to adhesion matrices, since they naturally exist in suspension. Initial experiments using a cellular ELISA kit showed that cells were removed from the surface of poly-L-lysine coated wells during wash steps, confirming that a cellular ELISA kit was not appropriate for this assay.

A substitution of a non-cellular ELISA kit to detect CPD formation required extraction of genomic DNA prior to assay. Initial testing of the non-

(A)



(B)



(C)

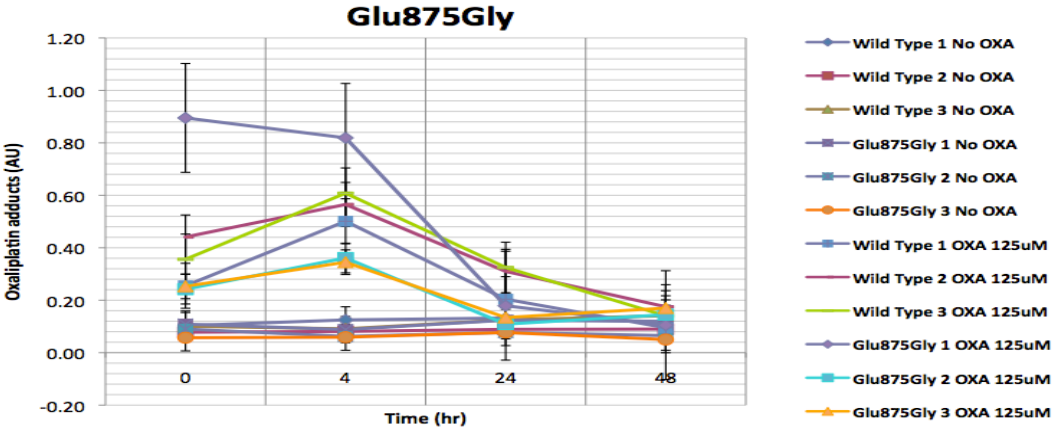


Figure 5.8 DNA repair in *ERCC4* variant and wild type cell lines, with and without oxaliplatin treatment. Oxaliplatin adduct formation (AU) in wild type and cells carrying (A) Pro379Ser, (B) Arg576Thr and (C) Glu875Gly with and without oxaliplatin (OXA; 125µM). Adduct formation was measured as optical density. Optical density values for each experiment are plotted as an average of two duplicate samples from the same experiment run on separate ELISA plates, which are plotted against time with standard error bars.

cellular ELISA kit with DNA samples from untreated and UV treated cell lines showed high CPD formations in UV treated cell lines compared with low levels in untreated cell lines. This suggested that extracted genomic DNA adhered correctly to the wells, therefore resolving the previous issue seen with cell removal using the cellular ELISA kit.

DNA extracted from UV treated cell lines initially appeared to show no repair over 24 hours. We incorporated FACS to remove non-viable cells prior to DNA extraction, which led to apparent repair between 4 and 48 hours after UV treatment. This suggested that the presence of DNA from non-viable cells masked repair previously. Despite this modification to the protocol, when these conditions were tested in remaining cell lines, inconsistent results were still seen. We previously noticed a potential discrepancy in CPD formation after UV irradiation by germicidal lamp, despite treating cell lines with the same dose at the same intensity. We therefore tested UV treatment in wild type cell lines using a UV crosslinker, which allowed for one single UV treatment of the entire cell suspension, followed by removal of aliquots at each time point for FACS and DNA extraction. This amendment led to more consistent results, most likely as a result of more consistent UV irradiation. UV treatment using the germicidal lamp required different dishes per time point, resulting in numerous UV treatments per cell line per experiment, therefore increasing the potential for inconsistent UV doses to reach cells. This made UV irradiation by UV crosslinker the preferable irradiation method.

The non-cellular ELISA kit for CPD detection was modified to probe for oxaliplatin-DNA adducts by substituting cisplatin-modified DNA primary antibody and an appropriate species specific HRP-conjugated secondary antibody, while using all other reagents according to the manufacturer's protocol. We tested 2 hour and 3 hour incubation times with oxaliplatin, prior to FACS, DNA extraction and ELISA, which demonstrated a peak in damage at 4 hours followed by repair between 4 and 48 hours. This was irrespective of treatment time; however, the shorter treatment time resulted in a higher viability of cells, meaning that sorting by FACS was more efficient and the overall DNA yield was higher. This was important because, more rapid FACS resulted in a shorter time between FACS

and DNA extraction, which reduced the opportunity for cell death and allowed for higher purity samples for analysis by ELISA.

5.4.2 Repair of ERCC4 variant cell lines after UV treatment

Untreated samples across all cell lines showed negligible CPD levels (<2ng/ml), which showed there were no problems with the specificity of the commercial ELISA kit for CPD detection, or its ability to differentiate between treated and untreated samples. This was true of both the commercial ELISA kit for CPD detection and the use of substituted antibodies to detect oxaliplatin adducts.

UV treatment of wild type cell lines showed high CPD formation at 0 and 4 hours, followed by marked repair between 4 and 48 hours. The high CPD levels seen at 0 hours suggests that damage occurs within the first 30 minutes after treatment, since this is the time that elapses between UV treatment and cell lysis during DNA extraction. Our data suggests that wild type cell lines undergo repair within the first 24 hours, as would be expected (Mailand *et al.* 2000). Additionally, CPD levels at each time point were remarkably similar across all wild type cell lines, during both test and validation experiments.

UV treatment of cell lines carrying Pro379Ser showed a distinct lack of repair in two cell lines and delayed repair after 24 hours in one cell line. Cell lines carrying Arg576Thr and Glu875Gly showed delayed repair that began 24 hours after UV treatment, compared with repair after 4 hours in wild type cell lines. Validation of this data showed very similar results, with consistent CPD formations between cell lines carrying the same *ERCC4* variant.

Statistical analysis showed significant differences in CPD formation between wild type and all *ERCC4* variant groups at 24 hours, followed by significant differences in CPD formation between the majority of cell groups at 48 hours. Statistical comparison of the validation experiment yielded very similar results. The only exception was a lack of a significant difference between Arg576Thr and Glu875Gly cell groups following UV treatment. These data suggest that the presence of our *ERCC4* variants adversely affected DNA repair.

Although the data suggests that the presence of an *ERCC4* variant disrupts NER following damage by UV light, it is important to consider that the absolute concentrations of CPDs measured may have been confounded by proliferation of the cells between UV exposure and CPD measurement. The assay involved UV irradiation of cell lines followed by incubation in normal tissue culture media for up to 48 hours prior to CPD measurement. Since the assay used did not account for cell proliferation during this incubation time, it is possible that our measurements are lower than would have been measured in a population of cells that have not undergone proliferation.

5.4.3 Repair of ERCC4 variant cell lines after oxaliplatin treatment

Oxaliplatin adduct formation in untreated cell lines was negligible, suggesting that the use of substituted primary and secondary antibodies in the commercial ELISA kit was specific for measurement of oxaliplatin-induced lesions. Wild type cell lines treated with oxaliplatin demonstrated a peak in oxaliplatin adducts at either 0 or 4 hours, followed by subsequent repair after 4 hours, reaching lesion levels comparable with untreated samples after 48 hours. There appeared to be some degree of variation in peak damage levels and repair rates between wild type cell lines.

Oxaliplatin treatment of cell lines carrying Pro379Ser, Arg576Thr and Glu875Gly showed repair rates that were similar to those of wild type cell lines, with no distinct differences in repair rate. These findings differed from the effect seen in variant cell lines after UV treatment, which could indicate that the variants in *ERCC4* have little or no effect on the cell's ability to perform NER after oxaliplatin treatment. This may suggest that other pathways are involved in repair of damage by drugs such as oxaliplatin, but not UV, which may be upregulated to compensate for reduced NER.

Alternatively, the repair trend seen after oxaliplatin treatment could be due to an issue with the specificity of the antibody used, since the primary antibody has only been tested in cisplatin treated model organisms. We observed consistent repair across all cell lines after oxaliplatin treatment, which suggests that the repair phenotype is present. However, we cannot be certain that the similar repair pattern between wild type and variant cell lines is a

consequence of a reduction in the sensitivity of the antibody as a direct consequence of its non-specific nature, resulting in an inability to detect small differences in repair rates.

The ELISA for detection of oxaliplatin adducts was performed in a similar way to the ELISA for CPD detection, such that the cells were treated with oxaliplatin and maintained in normal tissue culture media for up to 48 hours prior to ELISA. Therefore, it is similarly possible that the oxaliplatin adduct levels measured by ELISA were confounded by cell proliferation, therefore potentially resulting in lower or skewed optical density values for this assay.

5.4.4 Alternative assays for investigation of repair rates

Alternative assays may be useful for the investigation of repair after DNA damage. A potential alternative is the slot blot (SB) assay, a semi-quantitative assay in which a mixture containing biomolecules of interest is applied directly to a membrane, followed by detection by either nucleotide probes for northern and southern blotting, or by antibodies for western blotting (WB). The SB assay has been used previously to show a role of the 19S subunit of the proteasome in the regulation of NER (Gillette *et al.* 2001), in addition to the identification of two distinct mechanisms of the ubiquitin-proteasome pathway that control NER in *Saccharomyces cerevisiae* (Gillette *et al.* 2006).

The assay differs from traditional WB because samples are not separated by electrophoresis prior to assay, and there are no complex electro-blotting procedures required, which offers the significant advantage of saving time (Kumar *et al.* 2014; Walsh *et al.* 1992). It is also more sensitive than ELISA in terms of detection, since lower nucleic acid content samples can be run successfully in a SB assay, compared with ELISA (allowing detection of picograms and nanograms, respectively). Additionally, there is potential for increased variability during the numerous wash steps during ELISA (Alamdari *et al.* 2005).

However, unlike WB, the assay provides no information on the size of the target nucleic acid, and where more than one DNA lesion is present, the assay is unable to detect the proportions of each nucleic acid; therefore it can only

confirm the presence or absence of DNA lesion detected by the antibodies being used (Dalle-Donne *et al.* 2003). Although the immunoblots produced may be quantified by densitometry, this can only provide data on the relative abundance of different nucleic acids in the same mixture (Brown *et al.* 2004), whereas quantitative ELISA can provide an absolute concentration.

Other DNA repair capacity assays have been designed, including a cellular phenotype comet assay for detection of NER, by probing for benzo(a)pyrene diolepoxide (BPDE)-DNA adducts. Measurement of inter- and intra-assay variation using this assay identified low variation in peripheral blood mononuclear cells (PBMCs), but noticeably higher variation in lymphoblastoid cells (Allione *et al.* 2013), suggesting that this assay was not a suitable alternative for our investigation.

More recently, the SB and ELISA assays were combined for detection of carcinoembryonic antigen (CEA) in urine samples from patients with gastrointestinal (GI) cancers. The aim was to provide a more rapid, sensitive, specific and non-invasive immunodiagnostic test for GI malignancies, in an attempt to produce an assay more appropriate for mass screening, in comparison to more traditional assays that required expensive specialist equipment and longer time periods for incubation (El-Masry *et al.* 2006).

More recent advances in the DNA repair field have led to the combination of chromatin immunoprecipitation (ChIP) with either microarray or sequencing technology (ChIP-chip and ChIP-Seq, respectively), to allow high-throughput analysis of repair events throughout the entire genome. These assays may be used to investigate epigenetic modifications and DNA-protein binding in cells, and can be applied to all DNA damage pathways as long as antibodies or tagged DNA damage recognition enzymes are available for immunoprecipitation of the respective lesions (Wu *et al.* 2006; Teng *et al.* 2010).

ChIP-chip technology involves crosslinking DNA and protein or lesion with formaldehyde *in vivo*, immunoprecipitation using antibodies raised against the protein or lesion of interest, followed by sonication of DNA bound to protein to produce small fragments (0.2-2kb in length). The DNA fragments and relevant controls are fluorescently labelled and applied to microscope slides for

microarray analysis, in which control DNA acts as a background control and comparison of this DNA with immunoprecipitated DNA ensures accurate profiling of the position of specific proteins within the genome (Wu *et al.* 2006). This technology is advantageous due to its high specificity for detection of genes directly bound by the protein of interest, unlike other classic expression arrays that cannot distinguish between genes that are directly regulated or modified indirectly. Additionally, its 'reverse-genetic' approach allows genome-wide detection of specific proteins (Buck and Lieb, 2004).

However, ChIP-chip technology is limited by the amount of starting material required for successful array hybridization, which varies considerably between applications; previous studies have reported that <10ng DNA is hybridised from >50µg sample DNA, requiring hundreds of individual ChIP DNA samples to acquire enough DNA for hybridization (Oberley *et al.* 2004; Orlando *et al.* 1997). The starting material may be amplified by PCR, although in many cases PCR bias occurs, particularly in mammalian models where large areas of nucleotide repeats can skew data. Additionally, since microarrays covering all human chromosomes are not commercially available, the assay is only applicable to certain organisms (Wu *et al.* 2006).

Where appropriate microarray technology has not been developed for the organism of interest, ChIP-Seq technology can be used; this involves DNA-protein crosslinking and immunoprecipitation using an identical protocol to that of ChIP-chip, followed by direct NGS of small (~35bp) DNA fragments of interest (Park, 2009). As a result of rapid progress in NGS technology, ChIP-Seq provides an alternative assay with higher resolution, reduced noise and fewer artefacts than ChIP-chip (Bentley, 2006; Shendure and Ji, 2008). Additionally there is markedly greater genome coverage associated with ChIP-Seq, which is particularly useful for analysis of repetitive regions that are typically masked out on arrays (Park, 2009). Similar to ChIP-chip, ChIP-seq technology is highly specific for genes of interest, with approximately 80% of the human genome which can be readily aligned to the correct DNA fragment using 30bp short reads (Rozowsky *et al.* 2009).

However, despite the numerous advantages of ChIP-Seq over ChIP-chip, the major limitation is the cost and availability of the technology. Additional challenges include presence of sequencing errors, bias towards GC-rich content in fragment selection and loss of sensitivity and specificity in enriched regions, despite significant improvements in optimisation of the assay (Hillier *et al.* 2008; Quail *et al.* 2008). Technical issues surrounding loading samples can also contribute to low quality sequencing data (Park, 2009).

Chapter Six – Exome wide approach to identify genes that may cause PNAO

6.1 Introduction

Prior to the advent of genome-wide analyses, the traditional approach for pharmacogenomic analysis was the candidate gene approach, in which a group of related genes or a pathway that is relevant to a specific phenotype is selected based on literature searches. Otherwise known as a 'prior hypothesis' approach, this typically involves selection of a gene or pathway for further analysis (Hartford and Dolan, 2007). With regard to toxicities in the field of oncology, this approach has been useful in the study of genes involved in drug metabolism, transport and DNA repair, and has provided clinically relevant information that has aided treatment decisions for several drugs (Bosch *et al.* 2006; Efferth and Volm, 2005). Well known examples of polymorphisms identified via a candidate gene approach include the effects of variants in the cytochrome P450 enzyme on side effects to tamoxifen and the effect of the *UGT1A1*6* polymorphism on irinotecan-induced toxicity (discussed in section 1.1.1.2). Genetic variants in another drug metabolising enzyme, thiopurine methyltransferase, encoded by the *TPMT* gene, have been associated with severe thiopurine-induced toxicity. The enzyme catalyses the S-methylation of the thiopurine class of drugs, which are used in the treatment of leukaemia. Up to 20 variant alleles in *TPMT* have been identified, of which three (*TPMT*2*, **3A* and **3C*) are defective alleles that result in poor enzymatic activity, leading to elevated levels of cytotoxic thiopurine nucleotides and severe haematological toxicity (Ma and Lu, 2011).

The candidate gene approach is advantageous for several reasons; the techniques and genetic analyses used are well established and the costs incurred are reasonable. Additionally, this approach is considered useful for furthering our understanding of the functional implications of genetic variation within a gene or pathway, and provides a rationale for bringing this information into the clinic (Hartford and Dolan, 2007). However, several limitations exist, including the potential to overlook novel genes or pathways associated with disease or toxicity. Additionally, the simplicity of this approach means that the

influences of other genetic variants on cell behaviour are not considered (Schully *et al.* 2011).

The candidate gene approach is most successful for detection of single genes carrying variants that disrupt a drug's metabolism and are present in a reasonable proportion of the population (Hartford and Dolan, 2007). However, since chemotherapy-induced toxicity is most likely a multigenic trait (Moen *et al.* 2012) and may be influenced by genes present in a smaller proportion of the population, a broader approach is typically required. In contrast, a whole-genome or whole-exome approach may be used to consider a patient's entire genome or exome and evaluate multiple genes to identify genetic variants that contribute to a phenotype without bias. Previously, most candidate-based studies have revealed genetic variants that influence drug pharmacokinetics, whilst GWAS and exome sequencing methods have offered the potential for identifying rare and novel variants that influence drug pharmacodynamics and gene-phenotype interactions (Low *et al.* 2014). For multigenic traits, subtle variants in numerous genes may contribute in a cumulative manner to produce a more severe phenotype than a dramatic change in a single gene. For example, Easton *et al*. (2007) used a GWAS to identify five novel susceptibility loci for breast cancer in four genes (*FGFR2*, *TNRC9*, *MAP3K1* and *LSP1*) that had not previously been identified using candidate gene approaches, which contributed to overall risk in a cumulative manner. Additionally, a rare susceptibility locus on chromosome 8q24 for colorectal cancer was identified through GWAS, which similarly had not been identified using other approaches (Zanke *et al.* 2007). A rare variant at this susceptibility locus has since been associated with prostate cancer using a genome-wide sequencing approach (Gudmundsson *et al.* 2012).

However, limitations of genome-wide technologies still exist, including primarily high rates of false-positive variant predictions, as well as a high chance of missing rare genetic variants that could be important. Additionally, GWAS is limited in terms of accessibility to validation cohorts, since large control populations are needed to validate rare variants (Moen *et al.* 2012). Additional limitations include the cost of genome-wide and exome-wide

technologies; therefore experiments must be well designed prior to committing to the use of expensive technologies (Low *et al.* 2014).

Previous work by Dr Hannah West identified an association between variants in *ERCC4* and an increased incidence of PNAO in patients with CRC via a candidate based approach (section 1.9.5). Here we sought to identify other genes that may cause peripheral neuropathy by identifying rare or novel truncating variants in patients with PNAO using an exome wide approach.

6.2 Materials and Methods

6.2.1 Selection of patients

Patients were selected from 2,445 individuals with colorectal cancer undergoing therapy with 5-fluorouracil or capecitabine, oxaliplatin and, in a proportion of cases, cetuximab during the MRC COIN trial. A panel of patients (n=10) with severe PNAO (Grade ≥3) after oxaliplatin treatment were selected after review of their toxicity data (see section 1.9.2). An additional 45 patients were selected based on their clinical data, which indicated that they experienced a maximum of PNAO grade 0 or 1 throughout the duration of the trial. For the purpose of this chapter, we will refer to these no/low grade PNAO patients as controls. DNA from these patients was isolated for exome resequencing.

6.2.2 Exome resequencing

Exome resequencing, read alignment and variant calling of DNA from the ten patients with severe PNAO was performed by James Colley as outlined in section 1.9.2. Exome resequencing of 45 control exomes was performed independently by BGI (Hong Kong), using the Agilent SureSelectAll system for exome capture and sequencing at a depth of up to 50x coverage (section 2.3.6). Library fragments of exomic DNA from the 45 control patients were randomly fragmented by Covaris and adapters were ligated to both ends of each fragment. The adapter-ligated fragments were purified using Agencourt AMPure SPRI breads, which allowed extraction of fragments of approximately 250bp in size. Extracted DNA was amplified by ligation-mediated (LM) PCR, purified and hybridised to the SureSelect Biotinylated RNA Library for

enrichment. The magnitude of enrichment was assessed using the Agilent 2100 Bioanalyzer, followed by high throughput sequencing using the Hiseq2000 platform. Fastq files were processed by Illumina Basecalling software v1.7 and mapping of reads was carried out using SOAPaligner v2.21.

6.2.3 Selection of variants of interest

The bioinformatic output from the pipeline was altered to include only variants present in the 10 exomes from patients with grade ≥3 PNAO and absent from 45 control exomes. The initial filtering strategy (Strategy 1) involved exclusion of all non-synonymous SNPs, leaving only likely truncating mutations (stop gains, frame shifts and splice donor and acceptor sites) present in the same gene in ≥2 patients with PNAO. Common variants (those with MAF >1%) were removed by searching for MAF data in dbSNP v.132 and 1000 Genomes database v.3 (released 23/11/2010). Intronic variants were excluded by searching for all alternatively spliced transcripts in Ensembl.

6.2.4 Assessment of quality status of variants

Quality status of predicted variants was assessed during bioinformatic analysis. During Fastq file processing, variants were assigned a quality filter (P = high quality; VQSRT = medium quality; LQ = low quality), which combined the quality of the variant read during exome resequencing with the quality of alignment during bioinformatic analysis. Recalibration of quality scores was performed during bioinformatic analysis as described in section 2.5.6.

6.2.5 Validation by PCR and Sanger sequencing

All predicted variants of interest were validated by Sanger sequencing using an independent PCR product. All primers for PCR were designed by Marc Naven using a custom bioinformatic pipeline incorporating information from Primer3 v4.0.0 (section 2.5.6). PCR of individual variants, verification by agarose gel electrophoresis, product purification, Sanger sequencing and clean up were carried out as outlined in sections 2.5.7 to 2.5.11. Sequences were analysed using Sequencher v5.3.

Approaches for filtering for variants of interest	Description
Strategy 1	≥2 out of ten patients with PNAO carrying truncating variants in the same gene
Strategy 2	1 patient with PNAO carrying a truncating variant
Strategy 3	≥2 patients with PNAO, both carrying one truncating and one nonsynonymous variant in the same gene
Strategy 4	1 patient with PNAO carrying both one truncating and one nonsynonymous variant in the same gene
Strategy 5	≥2 patients with PNAO carrying either one truncating or one nonsynonymous variant in the same gene

Table 6.1 Approaches used to filter the bioinformatic pipeline's output of exome resequencing data, to identify variants that cause PNAO.

6.2.6 Identification of gene function

A literature review of one identified candidate gene was carried out using NCBI and other internet search engines, to identify any previously known association between the gene of interest and neuron function or peripheral neuropathy.

6.2.7 Alternative strategies for identifying genes of interest

The bioinformatic pipeline's output of exome resequencing data was filtered according to several different strategies; 1) ≥2 patients with PNAO carrying truncating variants in the same gene, 2) 1 patient with PNAO carrying a truncating variant, 3) ≥2 patients with PNAO both carrying one truncating and one nonsynonymous variant in the same gene, 4) 1 patient with PNAO carrying one truncating and one nonsynonymous variant in the same gene, and 5) ≥2 patients with PNAO carrying either one truncating or one nonsynonymous variant in the same gene (Table 6.1).

6.2.8 Removal of indels containing homopolymers and STRs

The exome resequencing data acquired by each filtering strategy was reduced further by removing indels containing repetitive base sequences in their flanking regions. A bioinformatic pipeline was established to find flanking regions of 100bp upstream and downstream of each variant. This information was used to manually filter out truncating indels that contained homopolymer runs of ≥5 single bases or ≥4 double base pairs (e.g. runs of GTs) and ≥2 STRs of ≥3bp.

6.3 Results

6.3.1 Selection criteria 1

All stages of filtering for variants of interest using Strategy 1 are outlined in Figure 6.1. All variants that were called during exome resequencing in 10 patients with severe PNAO and 45 control patients amounted to 643,354 variants. The list was reduced by filtering for variants present in our ten patients with PNAO and absent from all control exomes (3,578 variants). We then selected only truncating mutations present in the same gene in ≥2 PNAO

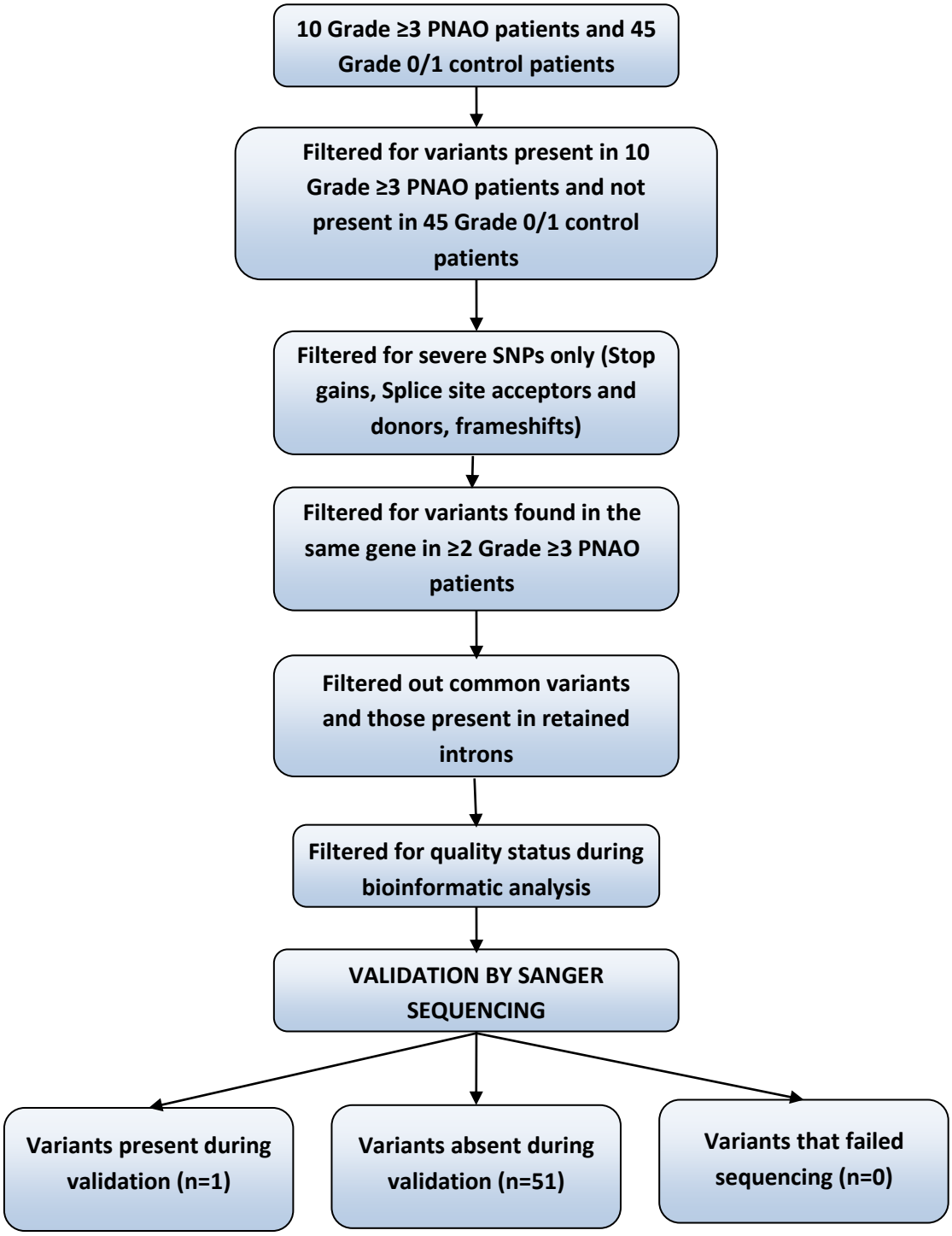


Figure 6.1 Strategy 1 – Schematic diagram of stages involved in filtering variants. All variants called during exome resequencing of DNA from ten patients with severe PNAO and 45 control patients were filtered and validated by Sanger sequencing to identify a list of candidate genes that may cause PNAO.

Stage of filtering		No. variants (genes) present	No. of variant types	
			Stop gains	Indels
Variants in 10 PNAO exomes and not in control exomes		3,578 (2,997)	N/A	N/A
Truncating variants only		82 (23)	12	70
Truncating variants in the same gene present in two or more PNAO patients		74 (20)	12	62
Removal of common (MAF>1%) variants		64 (18)	10	54
Removal of intronic variants		61 (17)	10	51
Exclusion of low quality variants		52 (14)	2	50
VALIDATION BY SANGER SEQUENCING	Genuine variants	1 (1)	0	1
	False positives	51 (14)	2	49
	Failed sequencing	0 (0)	0	0
Final list of variants		1 (1)	0	1

Table 6.2 Strategy 1 – Total numbers of variants and genes present at each stage of filtering. Validation of the two predicted variants in *PAPLN* revealed that only one variant was genuine, and therefore this gene is relevant only to filtering Strategy 2.

patients, resulting in 74 variants in 20 genes (*MACF1*, *OR2T35*, *FGFR3*, *FGFR4*, *SLC17A5*, *SMOC2*, *LRCC4*, *PRDM12*, *ANXA7*, *TAF1D*, *SPERT*, *TMCO3*, *PAPLN*, *ASGR2*, *GOSR1*, *STXBP2*, *KRI1*, *ANKLE1*, *ACSS2* and *ZFP64*). The numbers of variants (stop gains and indels) and genes at each stage of filtering are shown in Table 6.2.

6.3.2 Selection of rare and novel protein-encoding variants

Investigation of common variants (MAF >1%) resulted in exclusion of three variants present in ten patients: one novel deletion in *SMOC2* (MAF 14.4%), one novel stop gain in *SPERT* (MAF 7.4%) and one common insertion in *STXBP2* (rs35042054; MAF 43.9%). This resulted in 64 variants in 18 genes.

We then cross-referenced all predicted frame shift and splice site variants with their transcript IDs in Ensembl, which identified two novel frameshift variants located deep within the intron of *ASGR2* (ATG→A deletion at position 7011116 in two patients; A→AC insertion at position 7011120 in one patient; Ensembl transcript ID ENST00000450034). Exclusion of these variants resulted in 61 variants in 17 genes.

6.3.3 Exclusion of low quality variants

We excluded all variants assigned a VQSRT or LQ quality score, which resulted in removal of nine variants in five genes (*OR2T35*, *SLC17A5*, *PRDM12*, *ANXA7* and *TAF1D*). Other predicted variants in these genes showed good quality scores, and therefore this exclusion resulted in 52 variants in 14 genes.

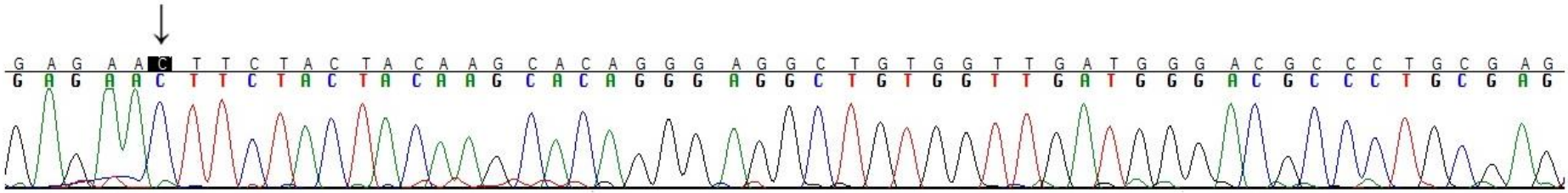
6.3.4 Validation of variants by Sanger sequencing

We validated one novel frameshift in the *PAPLN* gene in one patient with severe PNAO by Sanger sequencing. We failed to validate the second predicted variant in *PAPLN*, a rare splice acceptor site variant, and therefore *PAPLN* can only be considered as an output of filtering Strategy 2 (Figure 6.2).

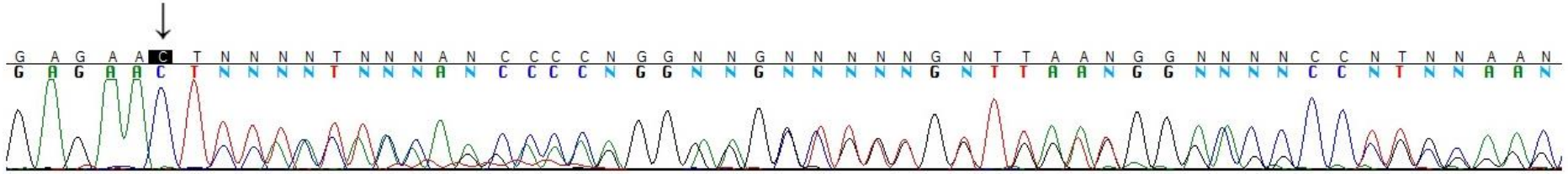
We failed to validate a total of fifty-one variants in fourteen genes (Table 6.3), including the rare splice acceptor site variant in *PAPLN*, which amounted to a 98.1% rate of false-positive variant calls. Chromatograms for a selection of

(A) PAPLN (Chromosome 14:73716735) – Control and Patient 7

(Control) No deletion

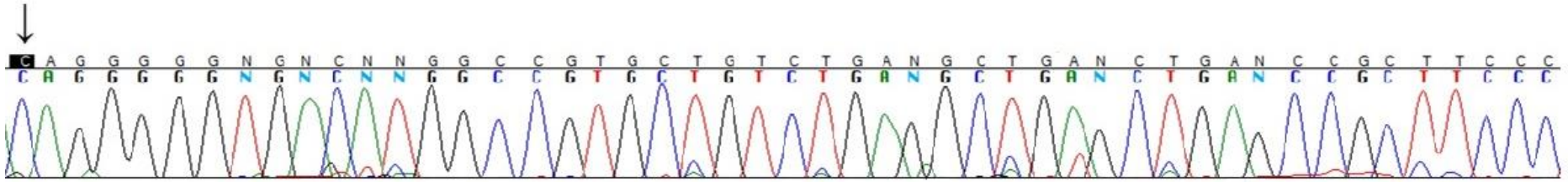


(Patient 7) CT→C deletion



(B) PAPLN (Chromosome 14:73731293) – Control and Patient 6

(Control) No deletion



(Patient 6) No deletion

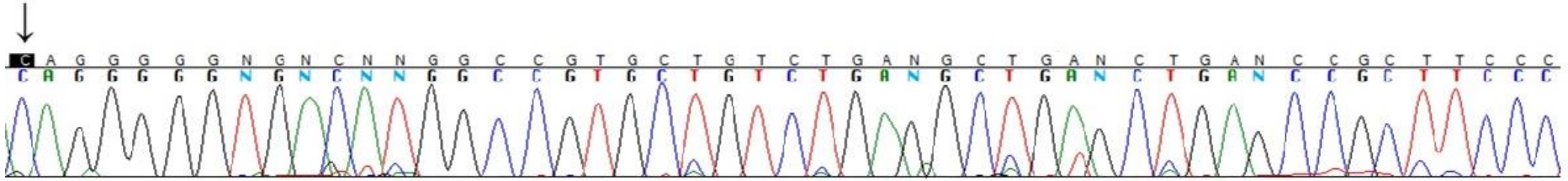


Figure 6.2 Chromatograms showing (A) one genuine frameshift variant and (B) one false-positive splice accepter site variant in *PAPLN* in 2 patients with grade ≥3 PNAO. The arrow indicates the variant allele in the control and PNAO exome.

Gene	Variants not present upon validation (Wild type → variant allele)	Zygosity	Flanking region of variant (60bp up/downstream of variant)
ACSS2	Splice donor site (AGT→A) (rs72449308)	Het	GATGCCGAGGACCCACTCTTCATCCTGTACACCAGTGGCTCCACAGGCAA ACCCAAGGCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT GTGTATTATGTAGGGGTAAG
ANKLE1	Splice acceptor site (GTGTT→ G) (rs10606406) Splice acceptor site (GTT→G) (novel) Frame shift (GTGTGTT→G) (novel)	Hom Hom Het	CATCTCTGGTTTCAGAAGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG TGTGTGTGTGTGTTTGTGTGTGTGTGTGTGTGTAGGGAGCACCCAGGC AGATCTCCCCCAGGCTGAGA TCTCTGGTTTCAGAAGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG TGTGTGTGTGTTTGTGTGTGTGTGTGTGTGTAGGGAGCACCCAGGCAG ATCTCCCCCAGGCTGAGAGA CCCATCTCTGGTTTCAGAAGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTG TGTGTGTGTGTGTTTGTGTGTGTGTGTGTGTGTAGGGAGCACCCAG GCAGATCTCCCCCAGGCTGA
FGFR3	Frame shift (CGT→C) (rs34003391)	Hom	GAGACTCAGTGCAGATGGAGAGACAGCTACACAGAGCTTTGGTCTGTGTG TGTGTGTGTGCGTGTGTGTGTGTGTGTGCACATCCGCGTGTGCCTGTG TGCGTGCGCATCTTGCCTCCA
FGFR4	Frame shift (C→CGTGT) (rs59390048)	Het	ATGTGGCACCAACAGCATGTGCCTTGTGTGTGTGTGTGTCCGTATGTGTGT GTGTGTATGCGTGTGTGTGTGTGTGTGTGTCTTGGCCACTGTCGTGTGCAC TAAATGCTGTGTGTGTGAC

Gene	Variants not present upon validation (Wild type → variant allele)	Zygosity	Flanking region of variant (60bp up/downstream of variant)
GOSR1	Frame shift (G→GT) (novel)	Het	AGAGACATATTGCAGGTAATATATTGGGCTCGAGATGTTTTCATTATCACAG GAGTTTGGGTTTTTTTTTTTTTAATCCCTGCATTGGATATGTGCACATATATT TAAAAGGACAAAGAGA
KRI1	Frame shift (G→GC) (rs34864064)	Hom	GAGTCGCTGCTGCTGTCTCGGTCCCCGTAGCGATCCTTCACTGCGGGACA CAGACGGGATGCCCCCCCCCCAGGTCAGCCGGCGGGGCCACTCGACTCCC CACTTCCCAACCGGCCGGCGCG
LRCC4	Frame shift (T→TG) (novel)	Het	CAAAGGCCCCGCTAGGGATGACTGTCAGCCAGTTGTCGAACAGCTCCAGG GTGTTGAGGCTGGCCAGGCCGTTGAAGGCCCCCACCTCAATCTGCCGGAT GGAGTTCCTGCCCAACTGCAG
MACF1	Frame shift (A→AG) (novel)	Het	CAGAACATCGCAGGTGGTAGTATGATGATGTCAGAAAAGACCGATGAGGA AGATAGTGGCAGGGAAATTTTTTCTGTCATGCAGTCATCCATTAGAATTGCTT GAAGAAGCTACCTTAAATG
PRDM12	Splice donor site (GGT→G) (novel)	Het	ACAGGAGCAGAACCTGGAGGTGGTCCAGATCGGCACCAGCATCTTCTACA AGGCCATTGAGGTGTGTGTGTGTGTGTGTGCACTGTTGTGTAGGGACCAGCC GGTAAACCCGGCGGGGGGGAGG
STXBP2	Splice acceptor site (TATGTGTGTGC→T) (rs11278818)	Het	TATGTCTGCATGCATGTGATTGCATGTGTGCATGTGTATACGTGTGCATGT GTCCATGTGTATGTGTGTGCATGTGTGTGCATGTGTGTGCATCTGTGTA TGCATGTGTGTGCGTGTTT

Gene	Variants not present upon validation (Wild type → variant allele)	Zygosity	Flanking region of variant (60bp up/downstream of variant)
STXBP2	Frame shift (A→ATG) (rs112703251)	Het	GCATGTGTGTATGCGTGTGTGTGTGCATCAGTGTCTGCATGTGTGTATATG TGTGTATGTATGTGTGCGCGCGCATCTGTGTGTGTGCATGTGTGTATGTATT GTGTGTGCATCTCTGTGTG
	Splice acceptor site (ATG→A) (novel)	Het	GTGTGTCTGTGGGTCTGTGTGTGCATTTGTGTCTGTGCATGTGTGTATGCG TGTGTATGTATGTGTGTGTGCATCTGTGTGCATCTGTATGTGTGTGTGTGC GTCTGTCTGTGTGCATGTG
	Frame shift (CATCTGTGTGTGTGT→C) (rs76937271)	Het	TGTGCATGTGTGTATGTATGTGTGTGCATCTCTGTGTGTGCATGTGTGTAT GTGTGTGTGCATCTGTGTGTGTGCATGTGTGCATGCGTGTGTATGTGTG TGTGCGTGCGTGCATCTGT
TMCO3	Frame shift (G→GGTTTT) (novel)	Het	GTCGTGGAAGTTCTCCGAATCCTGGTTTTGATTGGTCAGATTCTTTTTTCAC TAGCGGCGGGTTTTTCTTTTATGTCTTGTTTTTCTTTTATGTCTTGTTATAAA GAAGTATCTCATTGGACCCTATTATCGGAAGCTGC
ZFP64	Frame shift (TA→T) (rs72411807)	Het	GTGAATTTTTAAATGCCGCTCCATGTCCTTCATGCCATAAGCAGTCTTGAAT TGGCAACCTAAAAAAAAAAAGGAAAGATTAATTACAAAGGAAAACTTAGTAA CTTTTAGCATGGAGAAA

Table 6.3 Variants excluded during validation by Sanger sequencing, with wild type and variant alleles. Flanking regions of 60 base pairs (bp) upstream and downstream of the variant are shown, with the position of the variant underlined. Repetitive sequences are highlighted.

GOSR1 – Control and Patients 6 and 10

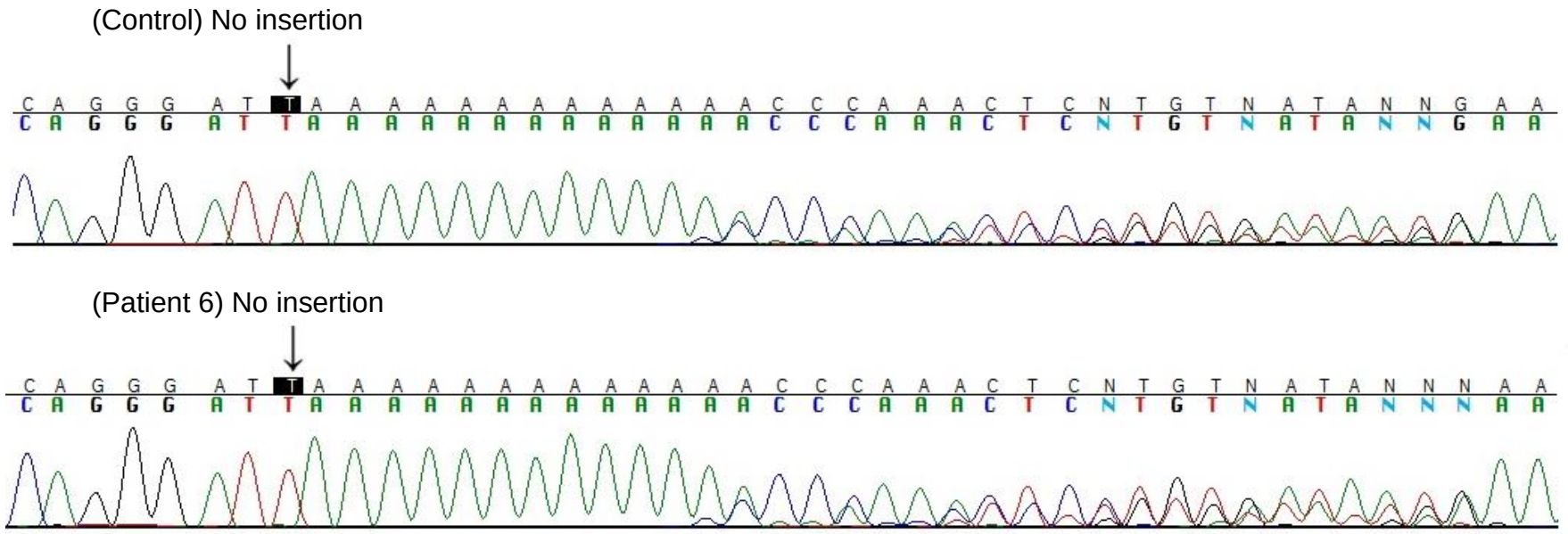


Figure 6.3 A false positive insertion (G→GT) in *GOSR1* predicted during bioinformatic analysis. The arrow indicates the locus of the predicted variant allele.

these variants are shown in Figure 6.3. A high proportion (49/51; 96.1%) of these non-validated variants contained homopolymer runs and short tandem repeats (STRs) flanking within 100 bp upstream or downstream of the variant, potentially increasing the likelihood of false positive variant calling by the bioinformatics pipeline (Table 6.3). We observed ten predicted variants containing homopolymer runs of ≥8 guanine-thymine pairs (GTs) before or after the variant. Two variants in *GOSR1* and *TMCO3* contained a run of ≥6 thymine bases after the variant; one variant in *KRI1* carried a run of 9 cytosine bases and one variant in *ZFP64* carried a run of 11 adenine bases after the variant. Additionally, one variant in *STXBP2* carried two STRs of ten bases in length (TGTGTGTGCA) after the variant (Table 6.3).

6.3.5 Literature search of candidate genes associated with PNAO

One variant in a single gene, *PAPLN*, was validated by Sanger sequencing, which identified *PAPLN* as a potential candidate gene for PNAO using our exome wide filtering Strategy 2. *PAPLN* has been cited in eight publications listed in PubMed since 1999, primarily in publications using GWAS of patients with depression to study suicide ideation. The majority of these publications concur that polymorphisms in *PAPLN* are linked to suicidal ideation in patients undergoing therapy for depression (Perroud, 2011; Brent *et al.* 2010; Laje *et al.* 2009; Laje *et al.* 2009*; Souslova *et al.* 2013). More recently, *PAPLN* was cited as one of four genes that encode extracellular matrix proteins that demonstrated substantial changes in expression after spinal cord injury in the neonatal South American grey short-tailed opossum (*Monodelphis domestica*; Saunders *et al.* 2014). *PAPLN* has also been associated with tooth development in mice (Pemberton *et al.* 2007).

6.3.6 Alternative filtering strategies for identifying variants of interest

The numbers of variants of interest for each strategy are depicted in Figure 6.4. There was considerable overlap in the genes listed for filtering Strategies 1-3, whilst the output for Strategy 4 largely overlapped with that of Strategy 2. The total numbers of variants for each filtering strategy ranged from 59 to 220 variants. In order to reduce the total numbers of variants for molecular

analysis, we developed a bioinformatic approach to remove indels that were likely to be false-positive calls based on the results outlined in section 6.3.4.

6.3.7 Removal of indels containing repetitive sequences in flanking regions

Removal of predicted indels with homopolymers and STRs present in their flanking regions resulted in an average of a 54.9% reduction in the total number of variants for molecular analysis (50.0% reduction in Strategy 1 output; 31.9% reduction in Strategy 2 output; 100% reduction in Strategy 3 output; 35.6% reduction in Strategy 4 output; 56.8% reduction in Strategy 5 output). A comparison of the numbers of variants before and after exclusion of indels with repetitive base sequences for each filtering strategy is given in Table 6.4. After removal of indels that met the selection criteria, the total numbers of variants listed for each filtering strategy ranged from 0 to 139 variants.

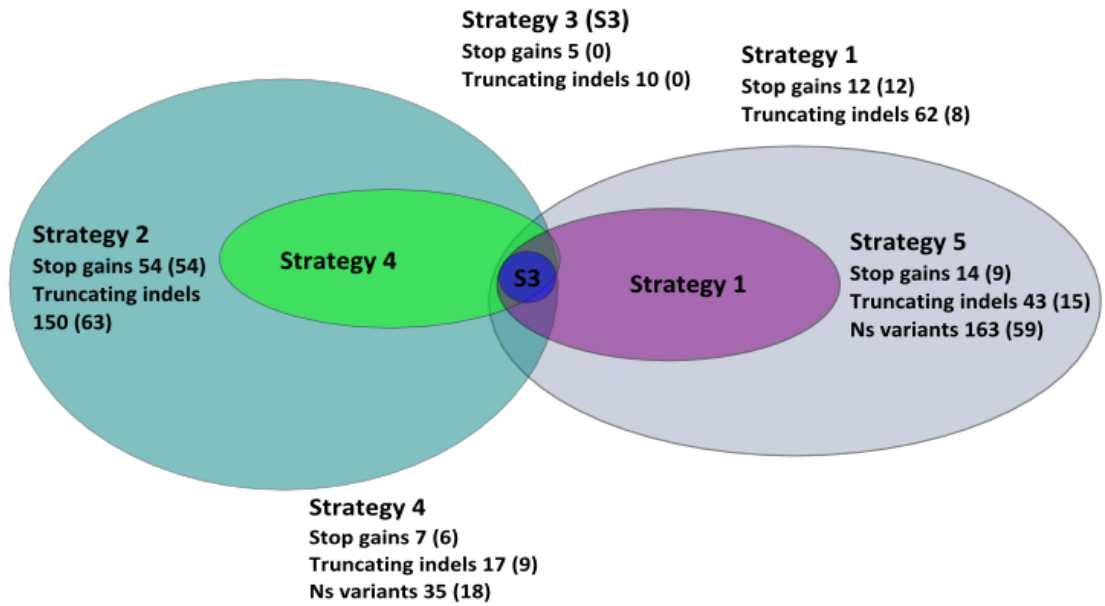
We identified five genes (*ANXA7*, *OR2T35*, *SMOC2*, *SPERT* and *TAF1D*) that were common across strategies 1 and 2, whilst a further three genes (*ASGR2*, *PAPLN* and *SLC17A5*) were common across Strategies 1, 2 and 5. We also observed overlap between Strategies 2, 4 and 5; seven genes were common across Strategies 2 and 4, whilst a further five genes were also present in Strategy 5. A schematic Venn diagram illustrating the overlap of genes identified via different filtering strategies is shown in Figure 6.4. All genes present in the exome resequencing output for each filtering strategy are listed in Table 6.5.

6.4 Discussion

6.4.1 False-positive variant calling of indels

Validation of the variants predicted using filtering Strategy 1 by Sanger sequencing showed a high rate of false-positive variant calls (98.1%), all of which were indels. Of these false-positives, 96.1% contained repetitive base sequences in their flanking regions. These were primarily homopolymers and runs of consecutive guanine-thymine pairs (GTs), although one variant carried a short tandem repeat (STR) of eight bases in length. This suggested that the

(A)



(B)

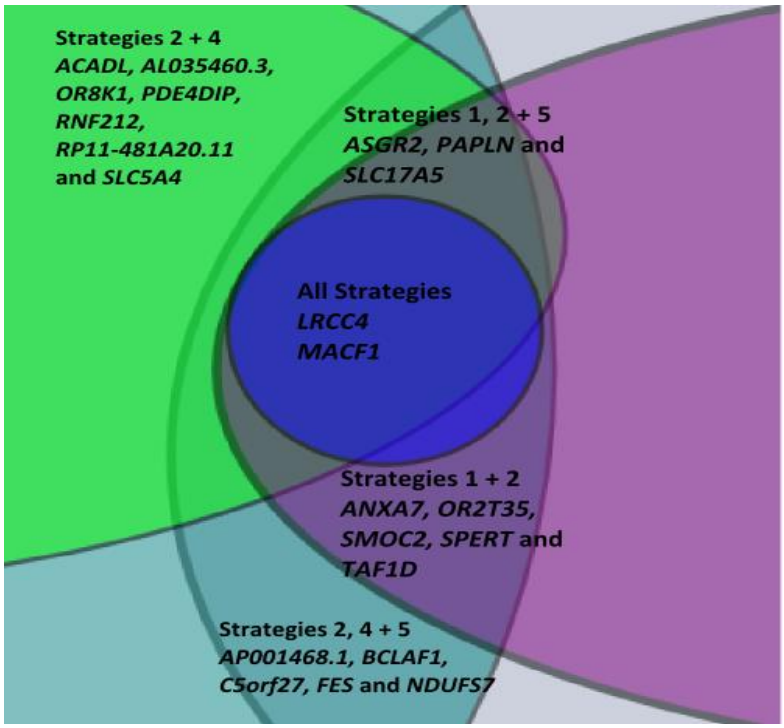


Figure 6.4 Venn diagram of NGS output from each filtering strategy. (A) Venn diagram of numbers of variants (stop gains, truncating indels and nonsynonymous [Ns] variants) present for each filtering strategy. Variant numbers after exclusion of predicted indels containing homopolymers and STRs in their flanking regions and following exclusion of variants not present during validation are shown in brackets. (B) Enlarged view of central section of Venn diagram, showing genes present in overlapping regions of different filtering strategies. Of the genes listed here, only *LRCC4* and *MACF1* were not validated, since the variants called in these genes were false positives.

	Strategy 1		Strategy 2		Strategy 3		Strategy 4		Strategy 5	
	Initial data output	**After indel exclusion**	**Initial data output**	**After indel exclusion**	**Initial data output**	**After indel exclusion**	**Initial data output**	**After indel exclusion**	**Initial data output**	**After indel exclusion**
Total genes	20	8	145	99	2	0	18	12	39	18
Total variants	74	20	204	117	15	0	59	33	220	83
Stop gain variants	12	12	54	54	0	0	7	6	14	9
Truncating indels	62	8	150	63	5	0	17	9	43	15
Nonsynonymous variants	0	0	0	0	10	0	35	18	163	59

Table 6.4 Summary of gene/variant numbers for each filtering strategy. Numbers of genes and variants are shown for each filtering strategy before and after bioinformatic analysis to exclude indels containing homopolymers or STRs in their flanking regions, and after exclusion of variants not present during validation.

Strategy 1 (n=8)	Strategy 2 (n=99)	Strategy 4 (n=12)	Strategy 5 (n=18)
ANXA7 *ASGR2* *OR2T35* *PAPLN* *SLC17A5* *SMOC2* *SPERT* *TAF1D*	*ABCA7; ACADL; ADC; AFAP1L1; AF131215.4; AKNAD1; ALDH3A1; AL035460.3; AL355149.1; AL391421.1; ANGEL1; ANKRD46; ANXA7; AP001468.1; ASCL3; ASGR2; ATP6V1H; AQP7; BCLAF1; BGLAP; BTNL9; CCDC70; CCDC83; CCDC87 CCDC108; CSTF2; CUZD1; CWH43; C5orf27; C9orf85; C11orf66; C12orf32; C15orf42; DCAKD; DNAH8; DSG4; ERCC4; ETNK2; EVPL; FAM109A; FAM5B; FAM53A; FES; FIBCD1; FRMD4A; FSTL5; GJA9; GPNMB; GRM3; KHNYN; KRT33B; KRT82; KRTAP19-4; LCE3C; LTK; MAN1B1; MRO; NBPF10; NDUFS7; NKIRAS1; NPIPL2; OR2T35; OR6F1; OR8K1; PABPC3; PAN2; PAPLN; PAX6; PCDHA11; PDE4DIP; PIK3R3; PP1R3A; QRICH2; RANBP10; RBMX; REG1B; RGS16; RNF212; RP11-10G12.2-001; RP11-481A20.11; RP11-867G23.8; SERPINA12; SLAMF8; SLC17A4; SLC17A5; SLC5A4; SMOC2; SMR3B; SNAPC1; SPERT; STOML3; TAF1D; TMEM159; TMEM54; TRIM16L; VRK2; ZNF135; ZNF354C; ZNF418; ZNF562.*	*ACADL* *AL035460.3* *AP001468.1* *BCLAF1* *C5orf27* *FES* *NDUFS7* *OR8K1* *PDE4DIP* *RNF212* *RP11-481A20.11* *SLC5A4*	*ABCA7* *AF131215.4* *AP001468.1* *ASGR2* *BCLAF1* *C5orf27* *DNAH8* *FES* *MAN1B1* *NDUFS7* *NPIPL2* *PABPC3* *PAPLN* *PDE4DIP* *PIK3R3* *RBMX* *SLC17A5* *ZNF135*

Table 6.5 Summary of genes listed for each filtering strategy. Genes present in the output for each filtering strategy after bioinformatic analysis to exclude indels containing homopolymer runs or STRs in their flanking regions and after exclusion of genes not present during validation. Genes highlighted in green are common across Strategies 1-2; genes highlighted in yellow are common across Strategies 1, 2 and 5.

presence of homopolymers and STRs reduced the accuracy of bioinformatic analysis when mapping the short read lengths to the correct sequence within these variants' flanking regions, due to slippage of DNA polymerases at the repetitive site, resulting in false-positive calling of small indels.

Identification of indels by NGS has been widely considered more challenging than SNP calling (Mullaney *et al.* 2010; Albers *et al.* 2011), in particular where sequencing errors and PCR enrichment bias is present (Fang *et al.* 2014). The majority of known indels range from 1-10bp in length (Zhang and Gerstein, 2003); however, within the human genome there are known indels that range up to 10,000bp in length (Mills *et al.* 2011). Detection of indels longer than 10bp provides a significant challenge for current NGS technologies, primarily due to the short read length used, making it difficult to accurately call indels that are longer than the short read length and increasing the likelihood of false-positive calls.

However, here we have identified an issue with accurate variant calling of relatively small indels (2-15bp in length), which suggests that further optimisation of read mapping may be necessary when employing the use of NGS. Published data has reported previously that assembly based callers show a significantly higher sensitivity for detection of indels >5bp in length compared with alignment based callers (Fang *et al.* 2014). Therefore, the use of an alignment mapping tool for variant calling of our data could provide one explanation for the high incidence of false positive calls.

The presence of homopolymers has been identified as an issue surrounding accurate indel calling by NGS previously (Koboldt *et al.* 2010; Krawitz *et al.* 2010; McElroy *et al.* 2012; De Beuf *et al.* 2012). A large scale validation experiment using 600 loci characterised by whole genome sequencing (WGS) and whole exome sequencing (WES) identified a low concordance (53%) in indel detection between WGS and WES of the same data, and a notably higher incidence of validated indels in WGS compared with WES data (84% vs. 57%, respectively; Fang *et al.* 2014), suggesting that WES may not be the most sensitive approach for large indel detection. Other issues have been reported including introduction of higher indel error rates following

poor PCR amplification and capture deficiency, in particular where homopolymers were present. Additionally, where homopolymers or STRs were located, the bioinformatic pipeline often called multiple variants at the same locus (Fang *et al.* 2014; Macalalad *et al.* 2012).

There is much debate about whether to use WGS for personal exomes since it is free from exome capture deficiency issues and could potentially reveal structural variants (SVs) and maintain greater uniformity of coverage than that seen in WES (Metzker, 2010; Zhu *et al.* 2012). Of course, the cost of sequencing is considerably cheaper using WES, although prices continue to be reduced as new technology is developed (Cirulli and Goldstein, 2010).

Another potential issue could be the level of coverage required to accurately call indels, which can vary depending on whether the variant is homozygous or heterozygous. The latter typically requires up to 60x coverage, which is significantly higher than the typical 30x coverage required for detection of homozygous variants (Fang *et al.* 2014; Li, 2011). Additionally, heterozygous indels are considerably more common than homozygous indels, indicating a need for higher coverage sequencing (Fang *et al.* 2014). Since 85.5% of the Strategy 1 output (53 of 62 indels) were heterozygous, this could provide an alternative explanation for such a high false-positive rate of indel calls.

6.4.2 Alternative strategies for identifying genes that may cause PNAO

Strategies 2 and 5 yielded large lists of variants, whilst Strategies 1-3 provided more focussed lists of predicted variants for molecular analysis. For the latter strategies, we chose to select variants present in ≥2 patients with PNAO to account for *de novo* mutations that may arise in a single patient as a result of individual variation. A large proportion of variants listed for each strategy were truncating indels and therefore there was a high likelihood of false-positive variant predictions. In order to reduce the lists of genes, we removed indels that could be false-positive calls.

6.4.3 Reduction of candidate gene lists by indel exclusion

We were unable to develop a bioinformatic pipeline to filter out predicted indels containing homopolymers and STRs in their flanking regions, due to the complexity required to account for differences in sequence length and base order of repetitive sequences. Therefore, any indels that met these criteria were removed manually.

Following removal of appropriate truncating indels and non-validated variants, a total of eight genes were identified that were common across multiple strategies (a combination of Strategies 1-2, 1-3 and 1, 2 and 5), suggesting that these might be implicated in PNAO. Of these eight genes, three were identified in Strategies 1, 2 and 5 (*ASGR2*, *PAPLN* and *SLC17A5*). *PAPLN* is discussed in section 6.4.4. *SLC17A5* encodes the sialic acid cotransporter protein, or sialin, which is involved in trafficking of free sialic acid molecules around the cell following protein and fat degradation (Qin *et al.* 2012). Mutations in *SLC17A5* are associated with sialic acid storage diseases in which a defective proton-driven transport system results in abnormally high urinary sialic acid levels and accumulation of free sialic acid in fibroblasts. A common example is Salla disease (OMIM #604369; Verheijen *et al.* 1999), an autosomal recessive neurodegenerative disorder characterised by delayed psychomotor development, dysmorphic facial features and ataxia (Renlund *et al.* 1986). In contrast, *ASGR2*, which encodes the asiaglycoprotein receptor 2 that functions to regulate serum glycoprotein homeostasis, has not been associated with neuron development or neuropathy (Veselkin *et al.* 2011).

The remaining five genes were present in Strategies 1 and 2 only (*ANXA7*, *OR2T35*, *SMOC2*, *SPERT* and *TAF1D*); of these, two were associated with neuron development or neural stem cell differentiation in some way. *SMOC2* encodes the SPARC related modular calcium binding 2 protein, which is primarily involved in extracellular matrix assembly and endothelial cell proliferation and migration (Bradshaw, 2012). LeDoux *et al*. (2006) reported a 16-fold increase in *SMOC2* expression in murine dorsal root ganglia (DRG) compared with spinal cord cells, whilst in drosophila, *SMOC2* was identified as one of approx. 400 genes expressed in sympathetic neuronal cells during

development (Apostolova *et al.* 2007). However, despite these findings, no direct association with neuropathy in humans has yet been identified (Basciano *et al.* 2011). *TAF1D* encodes TATA box binding protein (TBP)-associated factor, RNA polymerase I, D, a protein essential for RNA polymerase I transcription (Gorski *et al.* 2007). *TAF1D* has been shown to be expressed in pluripotent stem cells and in the early stages of neural stem cell differentiation in mice (Guzman-Ayala *et al.* 2015; Wang *et al.* 2012). The apparent function of these genes in early neuron development and differentiation may increase the likelihood that they are implicated in PNAO.

In contrast, the remaining three genes appear to play no role in neuron development or neuropathy; *ANXA7* encodes the Annexin-A7 tumour suppressor, which functions as a voltage-dependent calcium channel and has been implicated in prostate cancer progression (Srivastava *et al.* 2001). Although not directly associated with neuropathy, splicing of an alternative exon within *ANXA7* has been shown to promote tumour progression in glioblastoma cells (Ferrarese *et al.* 2014). *OR2T35* encodes olfactory receptor, family 2, subfamily T, member 35, (Prasad *et al.* 2014), which has not been associated with neuron development. Finally, *SPERT* encodes spermatid-associated protein, which appears to be involved in spermatozoa formation, although its exact role remains elusive (Feige *et al.* 2002). There are currently no citations in PubMed that associate *SPERT* with neuron function or neuropathy.

6.4.4 Identification of a candidate gene that may cause PNAO

Validation of all variants identified using filtering Strategy 1 revealed that one variant in the *PAPLN* gene was genuine, whilst the second predicted indel was a false-positive call. Therefore, although we used the filtering criteria for Strategy 1, *PAPLN* is only applicable to filtering Strategy 2.

PAPLN encodes the papilin proteoglycan-like sulfated glycoprotein (papilin), an extracellular matrix glycoprotein believed to influence cell rearrangement and migration, formation of basement membranes, and formation of the ectoderm, endoderm and mesoderm during embryogenesis. Papilin has been shown to be present in multiple basement membranes of the human eye (Kramerova *et al.* 2000; Uechi *et al.* 2014). The precise role of

papilin remains elusive, although *PAPLN* has been reported in a few high-throughput studies. *PAPLN* was identified as one of 50 genes with upregulated expression and a hypomethylated promoter in neurally differentiated mesenchymal stem cells (MSC), suggesting that *PAPLN* may influence nervous system development (TacGhee *et al.* 2014). Additionally, variants in *PAPLN* have been reported to place patients with severe depression at increased risk of suicidal behaviour following treatment with citalopram, a commonly used anti-depressant drug (Mann *et al.* 2009).

Papilin shares a set of protein domains known as the 'papilin cassette' with a disintegrin-like and metalloprotease with thrombospondin type 1 motif (ADAMTS) superfamily of proteins. The ADAMTS family consists of 19 ADAMTS zinc metalloproteases with ancillary domains and 7 ADAMTS-like (ADAMTSL) proteins, which resemble ADAMTS ancillary domains but lack proteolytic activity (Apte, 2009). An important paralog of *PAPLN* is the *ADAMTSL4* gene, which is associated with nervous system development (Neuhann *et al.* 2011). Papilin has been shown to non-competitively bind to ADAMTS2 *in vitro*, demonstrating their markedly similar structures (Kramerova *et al.* 2000).

Other paralogs of *PAPLN* include *ADAMTS7*, *ADAMTS12*, *ADAMTS13*, *THSD4*, *ADAMTSL1*, *ADAMTSL2*, *ADAMTSL3* and *ADAMTSL5* (GeneCards, 2015). Although there is nothing in the literature to suggest that variants in *PAPLN* directly cause disease, several of *PAPLN*'s paralogs have been associated with human genetic diseases caused by impaired extracellular matrix function; variation at the *ADAMTS7* locus has recently been associated with increased risk of inherited coronary artery disease (Schunkert *et al.* 2011). Additionally, missense variants in *ADAMTS13* are responsible for the rare, life threatening disease, thrombotic thrombocytopenic purpura (TTP), characterised by thrombocytopenia and microangiopathic haemolytic anaemia (Lotta *et al.* 2009).

Additionally, variants in *ADAMTS10*, which shares homology with *PAPLN*, have been associated with Weill-Marchesani syndrome, an autosomal recessive disorder of connective tissue characterised by short stature, eye

abnormalities, bradydactyly and stiffness of the joints (Le Goff and Cormier-Daire, 2011). Similarly, ADAMTS17 variants have been linked to Weill-Marchesani-like syndrome, in which sufferers present with all typical symptoms with the exception of bradydactyly and decreased joint mobility (Spranger *et al.* 1971).

Finally, variants in the *ADAMTSL2* and *ADAMTSL4* genes have been reported to cause geleophysic dysplasia (GD) and autosomal recessive ectopia lentis (EL), respectively. GD is similar to Weill-Marchesani syndrome and is characterised by short stature, joint limitations, thick skin and progressive cardiac valvular thickening resulting in premature death (Le Goff *et al.* 2008). EL is a rare disorder characterised by partial or entire displacement of the ocular lens from its space, leading to marked loss of vision, which can progress to cataracts, glaucoma and entire retinal detachment (Ahram *et al.* 2009).

6.4.5 Future directions

All eight genes of interest (*ANXA7*, *ASGR2*, *OR2T35*, *PAPLN*, *SLC17A5*, *SMOC2*, *SPERT* and *TAF1D*) should be sequenced in a validation panel of 54 further patients with PNAO to identify further stop gain mutations and nonsynonymous variants. These can then be assayed in a control cohort to look for variants that are over represented in patients with the PNAO phenotype.

Chapter Seven – General Discussion

7.1 Optimisation of cell-based assays for analysis of XPF function

A large component of the experimental work in this thesis involved the optimisation of assays required to investigate cell function in mammalian cell lines and *S. pombe* strains. As mentioned previously, it has been reported that serum proteins can bind irreversibly to oxaliplatin and absorb UV light (Graham *et al.* 2000; Yue *et al.* 2009; Polet and Steinhardt, 1968), whilst complete removal of serum from tissue culture media containing mammalian cells can result in quiescence and adversely affect DNA repair and apoptotic pathways (Jia *et al.* 1997; Collier *et al.* 2006). This meant that the conditions for DNA damaging treatments required a delicate balance between ensuring damage took place and allowing cells to carry out their damage recognition and repair processes, without influencing the results.

Additionally, the immunofluorescence and ELISA protocols involved multiple stages that required optimisation. It was particularly important to select an appropriate UV dose that would allow repair and ensure a high purity and yield of DNA from viable cells for ELISA. Alternatively, for immunofluorescence, it was essential to carefully optimise the fixation, permeabilisation and wash steps of the protocol, to allow sufficient entry of antibodies into cells without creating oversaturation of fluorescence or high background noise due to inadequate washing of cells.

7.2 Assaying the effects of *ERCC4* variants on mammalian cell function

In this thesis, we hoped to identify whether variants in *ERCC4* directly affect viability, repair rate and localisation of ERCC1-XPF in mammalian cells *in vitro*. A summary of the key findings is outlined in Table 7.1. Previous research has indicated that altered expression of the ERCC1 and XPF subunits could affect side effects of and response to platinum therapy, most likely through impaired ability to repair damaged DNA (Section 1.5.1). Increased expression of one or both subunits is typically correlated with a worse prognosis due to increased resistance to therapy. Alternatively, decreased expression results in a

Result	Assay and model used	Chapter
Negligible difference in viability after damage by either UV light or oxaliplatin between wild type and variant *ERCC4* cell lines	Viability measurement; mammalian cell lines	Three
Negligible difference in localisation of ERCC1/XPF complex after damage by oxaliplatin between wild type and variant *ERCC4* cell lines	Immunofluorescence; mammalian cell lines	Four
Negligible difference in localisation of rad16 after damage by oxaliplatin between wild type and variant *rad16* strains	Immunofluorescence; *S. pombe* strains	Four
Delayed repair in variant *ERCC4* cell lines compared with wild type cell lines after damage by UV light	ELISA; mammalian cell lines	Five
Very similar repair rates in wild type and variant *ERCC4* cell lines after oxaliplatin treatment	ELISA; mammalian cell lines	Five
High proportion (>95%) of false positive indel calls by whole exome NGS technology	NGS; DNA isolated from MRC COIN trial patients with PNAO	Six
Selection criteria established to remove predicted indels flanked by homopolymer runs or STRs	Bioinformatic analysis of NGS data	Six
PAPLN identified as a gene that may cause PNAO	NGS calling; validated by Sanger sequencing using DNA isolated from MRC COIN trial patients with PNAO	Six

Table 7.1 Summary of key findings in this thesis.

heightened sensitivity to treatment, which could improve response. However, this is likely to increase the severity of side effects due to reduced DNA repair, potentially resulting in an accumulation of adducts and increased apoptosis of healthy cells.

The mammalian cell lines were heterozygous for variants in *ERCC4*, suggesting that XPF expression was likely to be reduced, not absent. Therefore, we hypothesised that the presence of these variants would have a potentially subtle effect on viability, rate of DNA repair and cellular localisation of the repair complex.

7.2.1 Viability

Previously, embryonic fibroblasts isolated from homozygous XPF-deficient mice have displayed hypersensitivity to UV and mitomycin C (MMC), indicated by markedly reduced viability in *ERCC4*$^{-/-}$ cells compared with wild type and heterozygous cell lines (Tian *et al.* 2004). This is consistent with the role of XPF in the incision of damaged DNA during NER, suggesting that defective NER as a consequence of *ERCC4* variants results in eventual apoptosis. However, despite this, our data showed very similar trends in viability post treatment between wild type and variant cell groups, potentially as a result of haploinsufficiency of the variant cell lines due to their heterozygous nature, which may result in a subtle phenotype (Section 3.4.2).

Previous work in our laboratory carried out by Dr Hannah West used haploid *S. pombe* strains with an incorporated novel stop gain truncating mutation and nonsynonymous variants (that mimicked the variants studied herein) in *rad16* to investigate the effects of these variants on viability after damage. In contrast to the cell line data, this revealed a statistically significant decrease in viability (<20% compared with 100% in rad16^{WT} cells) after both oxaliplatin and UV treatment in the fission yeast model carrying the novel stop gain mutation. In contrast, all strains carrying nonsynonymous *rad16* variants displayed a smaller reduction in viability after damage (<60% viability compared with 100% in rad16^{WT} cells) by both oxaliplatin and UV (West, 2013, Chapter 7).

The significantly reduced viability of the rad16^{Ser585X} strain following UV and oxaliplatin treatment suggested that the incorporation of Ser585X severely

disrupted the ability of *rad16* to act normally in the repair of DNA damage, and was potentially as deleterious as removal of the entire gene. Additionally, the pattern of reduced viability in strains carrying nonsynonymous variants in *rad16* supported the hypothesis that *rad16* variants contribute to an NER deficiency. The differences in viability data between the diploid mammalian cell lines and haploid *S. pombe* strains were most likely the result of a single mutated *rad16* allele in all *S. pombe* strains, meaning that potential haploinsufficiency of the variant strains was not an issue.

Other studies have shown reduced viability in cell lines carrying variants in NER genes; Schafer *et al*. (2013) reported reduced viability after UV-C treatment and decreased XPG mRNA expression in XPG-deficient fibroblast cell lines isolated from one homozygous and two compound heterozygous patients. This is consistent with the suggestion that variants that impair incision of the damaged DNA strand during NER also impair cell viability.

It is commonly known that platinum compounds exert their cytotoxic effects by formation of lesions that inhibit transcription, resulting in impaired DNA repair and eventual apoptosis. This has been demonstrated in a number of *in vitro* cell-based assays and reconstituted systems (Mello *et al.* 1995; Jung and Lippard, 2007; Tremeau-Bravard *et al.* 2004; Tornaletti *et al.* 2003). More recently, Ang *et al*. (2010) demonstrated significant transcription inhibition in live mammalian cells using a site-specific platinated probe containing a single platinum-DNA adduct, followed by restoration of transcription following repair of adducts by NER. These findings were consistent with the hypothesis that platinum-DNA adducts form a physical barrier to block transcription by stalling RNA polymerase II. This results in cell cycle arrest in G2 phase because cells are unable to undergo mRNA synthesis in order to pass into mitosis, therefore implicating transcription inhibition as a critical determinant in the apoptotic pathway triggered by platinum drugs (Todd and Lippard, 2009). It is possible that DNA damage of our cell lines carrying heterozygous variants in *ERCC4* resulted in apoptosis of only a proportion of the cell population due to transcription stalling and cell cycle arrest. As such, the sensitivity of our assay may have not been sufficient to detect subtle differences in viability between wild type and variant cell lines.

7.2.2 ERCC1-XPF/Rad16 localisation

Previously, analysis of ERCC1 and XPF in cells isolated from XPE and XPF patients displayed cytoplasmic mislocalisation of both proteins, which could contribute to a reduced DNA repair capacity in these patients. Wild type cell lines showed ubiquitous expression of ERCC1 and XPF throughout the cell. Two cell lines from patients carrying one Pro379Ser allele in *ERCC4* both demonstrated cytoplasmic mislocalisation of XPF without treatment (Ahmad *et al.* 2010).

Despite this, our data showed similar localisation of ERCC1 and XPF in cell lines carrying Pro379Ser compared with wild type cell lines without treatment. In contrast, we observed a small trend of possible cytoplasmic mislocalisation in cell lines carrying Arg576Thr and marginally higher nuclear localisation of ERCC1 and XPF in cell lines carrying Glu875Gly. The localisation trends were very similar after treatment with oxaliplatin, with an overall trend showing small recruitment of ERCC1 and XPF into the nucleus after DNA damage in cell lines carrying Arg576Thr. Although the published results by Ahmad *et al*. (2010) differ from our data, it is possible that this is due to differences in method such as antibodies used or method of quantifying fluorescence. Our method involved analysis of fluorescent counts based on pixel number and intensity, whereas theirs involved manually observing cells and tallying the numbers of cells with exclusively nuclear, cytoplasmic and pancellular localisation of ERCC1 and XPF. It is therefore likely that our results are more accurate.

Other diseases have revealed mislocalisation of cellular proteins as part of their aetiology, including mislocalisation of mutant forms of the cystic fibrosis transmembrane conductance regulator (CFTR) chloride channel to the endoplasmic reticulum, resulting in premature protein degradation (Koch and Hoiby, 1993).

In order to address the potential issue of haploinsufficiency seen with heterozygous cell line models, we decided to mimic this experiment using a haploid model organism. As mentioned in section 7.2.1, viability studies using *S. pombe* strains carrying *rad16* mutations revealed a clear pattern of reduced

viability following DNA damage in variant strains, which would imply a repair deficiency that resulted in apoptosis. However, despite these findings, we were unable to demonstrate cytoplasmic to nuclear recruitment of rad16 after damage by immunofluorescence, despite the haploid nature of the organism. We observed no recruitment in rad16^{Ser585X} cells despite the presence of the novel stop gain mutation identified in Patient 8. This suggested that the stop gain mutation did not influence rad16 localisation after DNA damage by oxaliplatin.

A number of studies have confirmed the presence of a novel UV repair pathway present in *S. pombe*, known as the UVER pathway (McCready *et al.* 1993; McCready *et al.* 2000; Herrero *et al.* 2006). The presence of an additional damage repair pathway in *S. pombe* but not in other species of yeast suggests that there could be novel pathways that exist to repair chemical-induced damage by drugs including oxaliplatin. Additionally we had no way to introduce *rad16* variants without the presence of Cre-loxP sites, meaning that it was not possible to determine whether the introduction of these constructs would affect the strain's phenotype in terms of repair or localisation.

Alternatively, there may be issues with damage recognition in the presence of NER variants. Research by Giannattasio *et al*. (2004) using *Saccharomyces cerevisiae* (*S. cerevisiae*) revealed that NER plays a role in DNA damage checkpoint function in the presence of UV-induced lesions via a physical interaction between Rad14 and Ddc1 (*S. cerevisiae* homologs of human XPA and PCNA, respectively). This physical interaction enables checkpoint complexes to identify DNA lesions, therefore suggesting that damage recognition relies on functioning NER factors. Similar checkpoint proteins become activated following UV damage in *S. pombe*, leading to cell cycle arrest via two distinct sub-pathways; a Chk1-dependent mitotic arrest pathway and a Cds1-dependent pathway resulting in maintenance of DNA damage during S phase of the cell cycle (Murray *et al.* 1997). Therefore, it is possible that, in the presence of mutated NER genes yielding ineffective proteins, the damage checkpoint may not be activated, resulting in ineffective damage recognition. If this were true of *S. pombe* in response to chemical damage by oxaliplatin, NER would not be initiated, which could therefore

account for the lack of Rad16 recruitment seen in the *S. pombe* strains. However, further work would be required to verify this.

7.2.3 DNA repair rate

Our data revealed a distinct lack of repair in cell lines carrying Pro379Ser and delayed repair in cell lines carrying Arg576Thr and Glu875Gly after UV treatment. This suggested that our variants in *ERCC4* adversely affected function of the XPF protein in some way, perhaps by retarding the protein's capacity to facilitate repair by NER. The major protein-protein interaction required for formation of the ERCC1-XPF complex is dimerisation of each protein's hydrophobic C-terminal residues, which results in a stable heterodimer via the double helix-hairpin-helix regions of their HhH_2 domains. It is believed that XPF acts as a scaffold protein for ERCC1 during folding, suggesting that in the absence of XPF, ERCC1 may be unable to fold correctly *in vitro*, resulting in impaired endonuclease activity (Tripsianes *et al.* 2005). Recently, the Glu875Gly variant was reported to reduce binding affinity between XPF and the damaged DNA strand, despite having no effect on the protein-protein interaction of the repair complex (Allione *et al.* 2013b). Additionally, the Pro379Ser variant has previously been identified as a pathogenic variant when present in combination with other *ERCC4* mutations in a compound heterozygous state (Gregg *et al.* 2011), which may account for the most severe repair-deficient phenotype after UV damage in cells carrying this variant.

There are other instances in the literature of delayed repair associated with NER disorders. Recently, primary fibroblasts isolated from patients with TTD-A, a mildly photosensitive form of TTD caused by variants in *TFIIH*, displayed retarded downstream binding of NER factors after damage by UV light, resulting in delayed NER by approximately 24 hours (Theil *et al.* 2011). Additionally, delayed repair of 6-4 PPs following UV damage was reported in human XPE cell lines carrying a compound heterozygous mutation in the *p48* gene, which indicated that binding of damage recognition proteins is retarded in these cells (Tang *et al.* 2000).

Interestingly, despite the clear pattern of deficient DNA repair observed in variant cell groups compared with the wild type group following UV treatment,

in contrast this repair deficiency did not result in a reduction in viability of variant cell groups. This is surprising because we would expect a reduction in viability following an accumulation of UV-induced lesions, due to activation of apoptotic pathways in the absence of adequate DNA repair. As discussed previously, this may have been due to an issue regarding specificity of the assay used for viability measurements. Alternatively, there may be confounding due to cell proliferation during the 48 hour incubation period, which could have contributed to higher viability measurements at 24 and 48 hours, in addition to potentially lower levels of damage measured by ELISA. There is the additional consideration that the cell lines in question were EBV-transformed lymphoblastoid cell lines, and as such their ability to undergo apoptosis may have been compromised, which could account for the negligible differences in viability between wild type and variant cell groups.

7.3 Involvement of NER in neuron function and PNAO

Since neurons are terminally differentiated, replication of the genome does not occur; however, the stability of neuronal DNA is essential for effective transcription. Maintenance of neuronal DNA is reportedly regulated by the various NER pathways, in particular TC-NER (Jaarsma *et al.* 2011). This is consistent with an association between reduced NER capacity and amplified adduct levels in the dorsal root ganglia (DRG) of *XPA*$^{-/-}$ and *XPC*$^{-/-}$ mice during cisplatin treatment (Dzagnidze *et al.* 2007). Oxaliplatin has been shown to produce fewer platinum adducts in the DRG than equimolar concentrations of cisplatin; however, both drugs still display their cytotoxic action via adduct formation and apoptosis of DRG cells (Ta *et al.* 2006). In addition, 20-30% of patients carrying mutations in multiple complementation groups of XP (XPA, XPD and XPG) exhibit neurological symptoms, including progressive mental deterioration, ataxia and peripheral neuropathy (Kraemar *et al.* 1987; Anttinen *et al.* 2008; Nakano *et al.* 2014). This suggests a role of NER in the development of peripheral neuropathy.

Overall, XPF patients have exhibited symptoms of a milder neurological phenotype (Gregg *et al.* 2011), with one patient presenting with late onset axonal polyneuropathy (Sijbers *et al.* 1998). Additionally, reduced expression of

the ERCC1-XPF complex has been shown in mice in an attempt to mimic the XFE syndrome that occurs as a result of functional disruption of ERCC1. This reduced expression resulted in a distinct phenotype associated with peripheral neuropathy, consisting of morphological loss of neurons and accelerated ageing related neuronal dysregulation (Goss *et al.* 2011).

PNAO has been reported previously to be a consequence of direct oxalate toxicity on neuronal cells via their influence on voltage-gated sodium channels (Grolleau *et al.* 2001; Jaggi and Singh, 2012). In contrast, previous work by Dr. Hannah West suggested that defects in DNA repair mechanisms could contribute to PNAO in the acute setting, which was demonstrated by an inability to repair oxaliplatin-induced damage in *S. pombe*. We hoped to demonstrate this using mammalian cell lines; however, the functional analyses we present here suggest that deficiencies in DNA repair mechanisms could contribute to accumulation of UV lesions, but not oxaliplatin adducts. A deficiency in the repair of DNA adducts following either UV or oxaliplatin treatment could theoretically increase the rate of neuronal apoptosis and therefore contribute to an elevated rate of onset or severity of PNAO. However, we would need to demonstrate deficient repair following oxaliplatin treatment in a mammalian model to suggest with any certainty that this is the case.

7.4 NGS of patients with adverse reactions to chemotherapy

It is commonly known that genetic variation between individuals undergoing identical chemotherapy regimens can give rise to differences in the severity of adverse effects (Hertz and McLeod, 2013). Severe side effects associated with chemotherapeutic drugs often lead to treatment discontinuation or dose reduction, which could be detrimental to the treatment of the tumour. Improved understanding of the underlying genetics that influence cellular processing and efflux of drugs could significantly improve clinical decisions surrounding cancer treatment. This would potentially allow tailoring of treatment for individuals to maximise drug response with minimum toxicity.

The limitations of candidate gene approaches and their potential to overlook novel genes have been reported in several studies (Easton *et al.* 2007; Zanke *et al.* 2007; Gudmundsson *et al.* 2012). In contrast, GWAS offers the

ability to detect novel genes but may lack a large enough sample size to validate any association identified, particularly when rare variants are identified (Moen *et al.* 2012; Section 6.1).

The genomic work used to identify the nonsynonymous variants in *ERCC4* utilised a pathway approach. We applied another comprehensive approach to identify causal alleles for PNAO via a 'no prior hypothesis' approach, using NGS data from ten patients with severe PNAO and a suitably sized cohort of 'control' patients without PNAO. The benefit of such an approach was the elimination of selection bias that is present during candidate gene or pathway analyses.

7.4.1 False positive calling of indels

The avoidance of false positive and false negative results owing to coverage issues is essential for analysis of whole exome NGS data. In this study, we failed to validate >95% of variants, and of these, a large proportion (96.1%) were indels with repetitive flanking sequences, including homopolymers or STRs. This revealed an issue with the bioinformatic pipeline with regard to accurate prediction of small indels, ranging 2-15bp in length. It is possible that the exome resequencing protocol may have required higher coverage, in particular for detection of heterozygous indels, as this has been reported as an issue previously (Fang *et al.* 2014; Li, 2011). Greater than 90% of all false-positive indels in our data set were heterozygous in nature, which would support this hypothesis. Alternative explanations may lie with the bioinformatic pipeline, as has been discussed in section 6.4.1.

Issues with indel calling by NGS have been highlighted in a number of studies previously (O'Rawe *et al.* 2013; Zook *et al.* 2014; Quail *et al.* 2012; Ley *et al.* 2008). Comparison of variant calls from validated data sets run on various SOLiD platforms revealed that >90% of variant calls were unique to only one platform. This was true for other NGS platforms also, with >20% of variant calls unique to only one Illumina variant caller, suggesting that consistent variant calling of indels between platforms from the same manufacturer remains an issue (Challis *et al.* 2012). Since indel calling is particularly challenging, it is necessary to perform multi-alignment for reads mapped to the same locus (Li

and Homer, 2010). However, even in cases where a sophisticated indel caller such as SAMtools was employed for read alignment, there remained a high rate of false-positive indel predictions (O'Rawe *et al.* 2013), which has been reported in other NGS studies (Albers *et al.* 2011; Challis *et al.* 2012; Yeo *et al.* 2012; Ding *et al.* 2010).

7.4.2 Different approaches to identify causal alleles for PNAO

The initial filtering strategy (Strategy 1) revealed an issue with high rates of false-positive indel calls; the only variant that was present upon validation was in the *PAPLN* gene (discussed in section 7.4.4). In order to broaden our search for additional novel genes that may be implicated in PNAO, we employed five filtering strategies to identify different combinations of novel and low frequency (<1%) truncating and nonsynonymous variants present in genes that were mutated in ≥1 of the 10 patients with PNAO. Verification of the presence of these variants was carried out using Sanger sequencing (Section 6.3.6).

Similar exome wide, 'no prior hypothesis' approaches have been used for identification of novel genes associated with a certain phenotype. This approach has proven useful for the detection of causal Mendelian disease genes and is successful in approximately 60% of projects, depending on the availability of well-phenotyped patients and MAF of variants identified (Gilissen *et al.* 2012). Ng *et al*. (2010) employed a similar exome wide approach for manipulation of exome data from ten patients with a diagnosis of Kabuki syndrome (OMIM #147920), a monogenic disease characterised by short stature, skeletal abnormalities and delayed mental development. This revealed variants in *MLL2* that were the primary cause of the disease, which had been previously overlooked during candidate approaches.

Exome wide approaches can also be applied to multigenic diseases associated with *de novo* mutations, in particular where family-based exome sequencing is possible. This would involve filtering out all inherited variants to identify only sporadic variants that could cause the disease phenotype (Gilissen *et al.* 2012). However, additional issues are associated with exome wide approaches irrespective of filtering strategy used, including instances of copy

number variations, translocations or inversions that overlap with SNPs. These variants often cause a disease phenotype but may be easily missed by NGS (Norton *et al.* 2011).

7.4.3 Removal of indels containing homopolymer runs

Work has begun in the laboratory to develop a bioinformatic pipeline capable of removing indels with repetitive flanking regions that are likely to be false-positive calls. The aim was to reduce the numbers of false variants identified for molecular analysis.

Despite the intention to remove indels containing repetitive flanking sequences using a bioinformatic approach, the reality was that this was not possible to develop quickly. Therefore, indels that met these criteria were removed manually. This resulted in an average reduction of 54.9%, which resulted in a more focussed list of genes for molecular analysis. Following indel exclusion, in addition to *PAPLN*, a further seven genes were identified (*ASGR2*, *ANXA7*, *OR2T35*, *SLC17A5*, *SMOC2*, *SPERT* and *TAF1D*), which could provide interesting candidate genes for future sequencing analyses to investigate PNAO pathogenesis.

7.4.4 Potential genes that may cause PNAO

An initial filtering strategy (Strategy 1), followed by validation of predicted variants by Sanger sequencing, revealed one variant in the *PAPLN* gene, suggesting that this may be implicated in PNAO. Despite the fact that the second predicted variant was a false-positive call, therefore rendering the *PAPLN* gene a product of the filtering Strategy 2, it is possible that other polymorphisms in this gene could contribute to PNAO. This gene was not detected using previous candidate pathway approaches. Additionally, the presence of *PAPLN* across multiple filtering strategies after exclusion of variants with repetitive flanking regions would suggest that it may influence PNAO pathogenesis.

7.5 Future directions

7.5.1 Analysis of transcription recovery in human cell lines

In chapter three, we revealed no apparent reduction in survival of cells carrying an *ERCC4* variant. Despite this, in chapter five we observed deficient repair in variant cell lines following UV treatment only, with normal repair taking place following oxaliplatin treatment. Recent research in the field of defective DNA repair has identified a link between disorders of repair and impaired transcription. In response to DNA damage by UV light, local inhibition of transcription occurs via RNA polymerase II stalling; this DNA-damage induced transcription arrest is necessary to prevent production of aberrant RNA transcripts (Section 7.2.1; Svejstrup, 2010; Adam *et al.* 2013). Displacement of stalled RNA polymerase II results in completion of DNA repair and subsequent recovery of transcriptional activity (Gaillard and Aguilera, 2013).

Therefore, it is desirable to identify whether the *ERCC4* variants are exhibiting a functional effect following oxaliplatin treatment at the level of transcription as opposed to repair. It is possible that damage by oxaliplatin could adversely affect recovery of transcription, which could explain why we could not detect changes in repair rate or percentage survival in cell lines. This effect could be assayed for *in vitro* by examining mRNA synthesis following damage by real time quantitative PCR (qPCR) using both the mammalian cell lines and *S. pombe* strains as model organisms.

7.5.2 Sanger sequencing of *PAPLN*

In chapter six, we identified one candidate gene, *PAPLN*, which may be implicated in PNAO. *PAPLN* currently has no direct association with neuron development or function according to the literature, and has therefore not been identified previously using a pathway approach. It would be desirable to sequence all protein coding regions of this gene in a cohort of patients with PNAO to investigate the incidence of truncating variants in patients displaying the phenotype. This could then be compared with a suitably sized cohort of control patients without PNAO (Grade <1) to determine whether variants in

these genes are over represented in a population exhibiting the toxicity phenotype.

7.5.3 Sanger sequencing of additional candidate genes

In chapter six, a further seven candidate genes were identified by taking several exome wide approaches to filtering the NGS data and cross-referencing the output for each filtering strategy (Section 6.4.3). In addition to Sanger sequencing of *PAPLN*, it would also be desirable to sequence all protein coding regions of these seven genes using the same cohorts of patients with PNAO and controls for comparison. A proportion of these genes were associated with neuron development or neuropathy according to the literature. This could identify some interesting associations between novel genes not previously detected via pathway approaches and PNAO.

7.5.4 Improved optimisation of indel calling

Our findings suggested that accurate indel calling is an issue for both larger indels, as evident from scientific literature, and smaller indels. The small reads used and high error rates associated with NGS methodologies are partially responsible. In order to address this issue, the systematic errors that contribute to variant calling inaccuracies must be reduced to ensure that false-positive and -negative rates are acceptable for sequencing-based association studies. Generally speaking, this would require simultaneous optimisation of library preparation, sequence alignment tools and base calling algorithms (Harismendy *et al.* 2009). However, with regard to this particular project, we could begin by optimising the base calling algorithms used to predict the presence of indels and revisit the exome sequencing data files in an attempt to reduce the false-positive call rate.

References

Aarnio M *et al*. (2006). Life-time risk of different cancers in hereditary non-polyposis colorectal cancer (HNPCC) syndrome. *Human Cancer*; **64**(6): pp 430-3.

Abbasi R *et al*. (2009). Laryngeal cancer risk associated with smoking and alcohol consumption is modified by genetic polymorphisms in *ERCC5*, *ERCC6* and *RAD23B* but not by all polymorphisms in five other nucleotide excision repair genes. *Intl J Cancer*; **125**: pp 1431-9.

Adam S *et al*. (2013). Transcription recovery after DNA damage requires chromatin priming by the H3.3 histone chaperone HIRA. *Cell*; **155**: pp 94-106.

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**: pp 251-8.

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.

Agero AL *et al*. (2006). Dermatologic side effects associated with the epidermal growth factor receptor inhibitors. *J Am Acad Dermatol*; **55**: pp 657-70.

Ahel I *et al.* (2006). The neurodegenerative disease protein aprataxin resolves abortive DNA ligation intermediates. *Nature*; **443**: pp 713-6.

Ahmad A *et al*. (2010). Mislocalisation of XPF-ERCC1 nuclease contributes to reduced DNA repair in XP-F patients. *PLOS Genetics*; **6** (3): pp e1000871.

Ahram D *et al*. (2009). A homozygous mutation in ADAMTSL4 causes autosomal-recessive isolated ectopia lentis. *Am J Hum Genet*; **84**: pp 274-8.

Aitman TJ *et al*. (2011). The future of model organisms in human disease research. *Nature Rev Genet*; **12**: pp 575-82.

Alamdari DH *et al*. (2005). High sensitivity enzyme-linked immunosorbent assay (ELISA) method for measuring protein carbonyl in samples with low amounts of protein. *Free Radical Biol Med*; **39** (10): pp 1362-7.

Albers CA *et al*. (2011). Dindel: accurate indel calls from short-read data. *Genome Res*; **21**: pp 961-73.

Albert TJ *et al*. (2007). Direct selection of human genomic loci by microarray hybridization. *Nat Methods*; **4**(11): pp 903-905.

Albertson DG *et al*. (2003). Chromosome aberrations in solid tumours. *Nature Genetics*; **34**(4): pp 369-76.

Alcindor T and Beauger N. (2011). Oxaliplatin: a review in the era of molecularly targeted therapy. *Curr Oncol*; **18**: pp 18-25.

Allione A *et al*. (2013). Validation of the nucleotide excision repair comet assay on cryopreserved PBMCs to measure inter-individual variation in DNA repair capacity. *Mutagenesis*; **28**: pp 65-70.

Allione A *et al.* (2013b). Inter-individual variation in nucleotide excision repair pathway is modulated by non-synonymous polymorphisms in ERCC4 and MBD4 genes. *Mutat Res*; **751-2**: pp 49-54.

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 1626-34.

Amersi F *et al*. (2005). Colorectal cancer: epidemiology, risk factors, and health services. *Clin Colon Rectal Surg*; **18**(3): pp 133-40.

Amstutz U *et al*. (2011). Dihydropyrimidine dehydrogenase gene as a major predictor of severe 5-fluorouracil toxicity. *Pharmacogenom*; **12**(9): pp 1321-36.

Ando Y *et al*. (1998). UGT1A1 genotypes and glucuronidation of SN-38, the active metabolite of irinotecan. *Ann Oncol*; **9**: pp 845-7.

Andre T *et al*. (2009). Improved overall survival with oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment in stage II or III colon cancer in the MOSAIC trial. *J Clin Oncol*; **27**(19): pp 3109-16.

Ang WH *et al*. (2010). Transcription inhibition by platinum-DNA cross-links in live mammalian cells. *J Am Chem Soc*; **132**: pp 7429-35.

Anttinen A *et al*. (2008). Neurological symptoms and natural course of xeroderma pigmentosum. *Brain*; **131**(8): pp 1979-89.

Apostolova G *et al*. (2007). Neurotransmitter phenotype-specific expression changes in developing sympathetic neurons. *Mol Cell Neurosci*; **35**(3): pp 397-408.

Apte SS. (2009). A disintegrin-like and metalloprotease (Reprolysin-type) with thrombospondin type 1 motif (ADAMTS) superfamily: functions and mechanisms. *J Biol Chem*; **284**(46): pp 31493-7.

Argyriou AA *et al*. (2008). A review on oxaliplatin-induced peripheral nerve damage. *Cancer Treat Rev*; **34**(4): pp 368-77.

Argyriou AA *et al*. (2008). Bortezomib-induced peripheral neuropathy in multiple myeloma: a comprehensive review of the literature. *Blood*; **112**(5): pp 1593-9.

Arpino G *et al*. (2005). Estrogen receptor-positive, progesterone receptor-negative breast cancer: association with growth factor receptor expression and tamoxifen resistance. *J Natl Cancer Inst*; **97**(17): pp 1254-61.

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.

Baldwin RM *et al*. (2012). A genome-wide association study identifies novel loci for paclitaxel-induced sensory peripheral neuropathy in CALGB 40101. *Clin Cancer Res*; **18**(18): pp 5099-109.

Banerjee A et al. (2014). The DNA intercalators ethidium bromide and propidium iodide also bind to core histones. *FEBS Open Bio*; **4**: pp 251-9.

Bao S *et al.* (2011). Evaluation of next-generation sequencing software in mapping and assembly. *J Hum Genet*; **56**(6): pp 406-414.

Barcena C *et al*. (2014). Exome sequencing identifies a novel mutation in PIK3R1 as the cause of SHORT syndrome. *BMC Med Genet*; **15**: pp 51-6.

Barzilai A. (2007). The contribution of the DNA damage response to neuronal viability. *Antioxid Redox Signal*; **9**: pp 211-8.

Basciano L *et al*. (2011). Long term culture of mesenchymal stem cells in hypoxia promotes a genetic program maintaining their undifferentiated and multipotent status. *BMC Cell Biology*; **12**: pp 12-23.

Beevers G *et al*. (2001). The pathophysiology of hypertension. *BMJ*; **322**(7291): pp 912-6.

Bentley DR. (2006). Whole-genome re-sequencing. *Curr Opin Genet Dev*; **16**: pp 545-52.

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**: pp 2643-54.

Berends MJ *et al*. (2002). Molecular and clinical characteristics of MSH6 variants: an analysis of 25 index carriers of a germline variant. *Am J Hum Genet*; **70**: pp 26-37.

Berhane N *et al*. (2012). DNA repair genes polymorphism (XPG and XRCC1) and association of prostate cancer in a north Indian population. *Mol Biol Reports*; **39**(3): pp 2471-9.

Berneburg M *et al*. (2000). The cancer-free phenotype in trichothiodystrophy is unrelated to its repair defect. *Cancer Res*; **60**: pp 431-8.

Bernstein C *et al*. (2002). DNA repair/pro-apoptotic dual-role proteins in five major DNA repair pathways: fail-safe protection against carcinogenesis. *Mutation Res*; **511**(2): pp 145-78.

Biggerstaff M *et al*. (1993). Co-correction of ERCC1, ERCC4 and xeroderma pigmentosum group F DNA repair defects in vitro. *EMBO J.;* **12**: pp 3685-92.

Bindu PS *et al*. (2005). Peripheral neuropathy in metachromatic leucodystrophy: a study of 40 cases from South India. *J Neurol Neurosurg Psychiatry*; **76**: pp 1698-701.

Blandizzi C *et al*. (2001). Characterisation of a novel mechanism accounting for the adverse cholinergic effects of the anticancer drug irinotecan. *Br J Pharmacol*; **132**: pp 73-84.

Bodmer W and Bonilla C. (2008). Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet*; **40**(6): pp 695-701.

Bogliolo M *et al*. (2013). Mutations in *ERCC4*, encoding the DNA repair endonuclease XPF, cause Fanconi anemia. *Am J Hum Genet*; **92**: pp 800-6.

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 663-71.

Bolton KL *et al*. (2012). Association between BRCA1 and BRCA2 mutations and survival in women with invasive epithelial ovarian cancer. *JAMA*; **307**: pp 382-90.

Bose R *et al*. (2013). Activating HER2 mutations in HER2 gene amplification negative breast cancer. *Cancer Discovery*; **3**(2): pp 224-37.

Bosch TM *et al*. (2006). Genetic polymorphisms of drug-metabolising enzymes and drug transporters in the chemotherapeutic treatment of cancer. *Clin Pharmacokinet*; **45**(3): pp 253-85.

Botstein D and Risch N. (2003). Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. *Nat Genet*; **33**: pp 228-37.

Botta E *et al*. (2002). Reduced level of the repair/transcription factor TFIIH in trichothiodystrophy. *Hum Mol Genet*; **11**(23): pp 2919-28.

Bou-Assaly W and Mukherji S. (2010). Cetuximab (Erbitux). *American Journal of Neuroradiology*; **31**: pp 626-7.

BoulikasT. (2007). Molecular mechanisms of cisplatin and its liposomally encapsulated form, Lipoplatin™. Lipoplatin™ as a chemotherapy and antiangiogenesis drug. *Cancer Ther*; **5**: pp 351-76.

Bradshaw AD. (2012). Diverse biological functions of the SPARC family of proteins. *Intl J Biochem and Cell Biol*; **44**(3): pp 480-8.

Brent D *et al*. (2010). Pharmacogenomics of suicidal events. *Pharmacogenomics*; **11**(6): pp 793-807.

Britland ST *et al*. (1990). Association of painful and painless diabetic polyneuropathy with different patterns of nerve fiber degeneration and regeneration. *Diabetes*; **39**(8): pp 898-908.

Bronstein I *et al*. (1992). Improved chemiluminescent western blotting procedure. *BioTechniques*; **12**(5): pp 748-53.

Brooks PJ *et al.* (2008). Do all of the neurological diseases in patients with DNA repair gene mutations result from the accumulation of DNA damage? *DNA Repair*; **7**: pp 834-48.

Brown T *et al*. (2004). Unit 4.9 Analysis of RNA by northern and slot blot hybridization. *Current Protocols in Molecular Biology*; **4.9**: pp 1-19.

Bruins W *et al*. (2004). Increased sensitivity to UV radiation in mice with a p53 point mutation at Ser389. *Mol Cell Biol*; **24**(20): pp 8884-94.

Bruno R *et al*. (2003). Preliminary population pharmacokinetics (PPK) and exposure-safety (E-S) relationships of erlotinib HCI in patients with metastatic breast cancer (MBC). *Proc Am Soc Clin Oncol*; **22**: pp 205-13.

Buck D *et al*. (2006). Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell*; **124**: pp 301-13.

Buck MJ and Lieb JD. (2004). ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics*; **83**(3): pp 349-60.

Burglen L *et al*. (1997). The gene encoding p44, a subunit of the transcription factor TFIIH, is involved in large-scale deletions associated with Werdnig-Hoffmann disease. *Am J Hum Genet*; **60**: pp 72-9.

Burt R and Neklason DW. (2005). Genetic testing for inherited colon cancer. *Gastroenterol*; **128**(6): pp 1696-716.

Busam KJ *et al*. (2001). Cutaneous side-effects in cancer patients treated with the antiepidermal growth factor receptor antibody C225. *Br J Dermatol*; **144**: pp 1169-76.

Calonge TM *et al*. (2000). *Schizosaccharomyces pombe* Rho2p GTPase regulates cell wall α-glucan biosynthesis through the protein kinase Pck2p. *Mol Biol Cell*; **11**(12): pp 4393-401.

Cancer Genome Atlas Research Network. (2011). Integrated genomic analyses of ovarian carcinoma. *Nature*; **474**: pp 609-15.

Cancer Genome Atlas Network. (2012). Comprehensive molecular characterization of human colon and rectal cancers. *Nature*; **487**: pp 330-7.

Cancer Research UK. (2010). Bowel cancer statistics. Retrieved May 11, 2015, from Cancer Research UK: http://www.cancerresearchuk.org/cancer-info/cancerstats/types/bowel/incidence/uk-bowel-cancer-incidence-statistics

Carethers JM. (2008). Systemic treatment of advanced colorectal cancer: tailoring therapy to the tumor. *Therap Adv Gastroenterol*; **1**: pp 33-42.

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 566-75.

Cassidy J *et al*. (2004). XELOX (capecitabine plus oxaliplatin): active first-line treatment for patients with metastatic colorectal cancer. *J Clin Oncol*; **22**(11): pp 2084-91.

Cassidy J and Misset JL. (2002). Oxaliplatin-related side effects: characteristics and management. *Semin Oncol*; **29**(5): pp 11-20.

Cavaletti G *et al*. (1995). Peripheral neurotoxicity of taxol in patients previously treated with cisplatin. *Cancer*; **75**(5): pp 1141-50.

Challis D *et al*. (2012). An integrative variant analysis suite for whole exome next-generation sequencing data. *BMC Bioinformatics*; **13**: pp 8-19.

Chaney SG *et al*. (2005). Recognition and processing of cisplatin- and oxaliplatin-DNA adducts. *Critical Reviews in Oncology/Hematology*; **53**: pp 3-11.

Chang MM and Lovett J. (2011). A laboratory exercise illustrating the sensitivity and specificity of Western blot analysis. *Biochem Mol Biol Educ*; **39**(4): pp 291-7.

Cheadle JP and Sampson JR. (2007). MUTYH-associated polyposis - from defect in base excision repair to clinical genetic testing. *DNA Repair (Amst)*; **6**(3): pp 274-279.

Chen M and Wang J. (2002). Initiator caspases in apoptosis signalling pathways. *Apoptosis*; **7**: pp 313-9.

Chen H and Kuo MT. (2010). Role of glutathione in the regulation of cisplatin resistance in cancer chemotherapy. *Met Based Drugs*; **2010**: pp 1-7.

Cheng C *et al*. (2015). Use of atropine-diphenoxylate compared with hyoscyamine to decrease rates of irinotecan-related cholinergic syndrome. *J Community Support Oncol*; **13**: pp 3-7.

Ciccia A and Elledge SJ. (2010). The DNA damage response: making it safe to play with knives. *Mol Cell*; **40** (2): pp 179-204.

Ciccolini J *et al*. (2010). Routine dihydropyrimidine dehydrogenase testing for anticipating 5-fluorouracil-related severe toxicities: hype or hope? *Clin Colorectal Cancer*; **9**(4): pp 224-8.

Cingolani P *et al*. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*; **6**(2): pp 80-92.

Cirulli ET and Goldstein DB. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Rev Genet*; **11**: pp 415-25.

Cleaver JE. (2005). Cancer in xeroderma pigmentosum and related disorders of DNA repair. *Nat Rev Cancer*; **5**: pp 564-73.

Cliffer KD *et al*. (1998). Physiological characterisation of Taxol-induced large-fiber sensory neuropathy in the rat. *Ann Neurol*; **43**: pp 46-55.

Collier HA *et al*. (2006). A new description of cellular quiescence. *PLOS Biology*; **4**(3): pp 329-49.

Collins AR *et al*. (1997). The comet assay: what can it really tell us? *Mutation Res*; **375**(2): pp 183-93.

Collins AR. (2004). The comet assay for DNA damage and repair. *Mol Biotech*; **26**(3): pp 249-61.

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 628-40.

Crawford NP *et al*. (2006). Germline polymorphisms in SIPA1 are associated with metastasis and other indicators of poor prognosis in breast cancer. *Breast Cancer Res*; **8**: pp R16.

Crawford NP *et al*. (2008). Bromodomain 4 activation predicts breast cancer survival. *Proc Natl Acad Sci USA*; **105**: pp 6380-5.

Culy CR *et al*. (2000). Oxaliplatin. *Drugs*; **60** (4): pp 895-924.

Cunningham D et al. (2001). Optimizing the use of irinotecan in colorectal cancer. *Oncologist*; **6**: pp 17-23.

Cunningham D *et al*. (2004). Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *New England J Med*; **351**: pp 337-45.

D'Andrea AD and Grompe M. (2003). The Fanconi anemia/BRCA pathway. *Nat Rev Cancer*; **3**: pp 23-34.

Dalle-Donne I *et al*. (2003). Protein carbonyl groups as biomarkers of oxidative stress. *Clinica Chimica Acta*; **329** (1-2): pp 23-38.

De Beuf K *et al*. (2012). Improved base-calling and quality scores for 454 sequencing based on a Hurdle Poisson model. *BMC Bioinformatics*; **13**: pp 303-13.

De Boer J *et al*. (2002). Premature aging in mice deficient in DNA repair and transcription. *Science*; **296**(5571): pp 1276-9.

De Boer J and Hoeijmakers JHJ. (2000). Nucleotide excision repair and human syndromes. *Carcinogenesis*; **21**(3): pp 453-60.

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 2938-47.

De Laat WL *et al*. (1999). Molecular mechanism of nucleotide excision repair. *Genes and Dev*; **13**: pp 768-85.

De La Chapelle A. (2004). Genetic predisposition to colorectal cancer. *Nature Rev Cancer*; **4**: pp 769-80.

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. *Annals of Oncology*; **19**(3): pp 508-15.

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 753-62.

De Waard H *et al*. (2004). Different effects of CSA and CSB deficiency on sensitivity to oxidative DNA damage. *Mol Cell Biol*; **24**(18): pp 7941-8.

Dean M *et al*. (2005). Tumour stem cells and drug resistance. *Nature Reviews Cancer*; **5**: pp 275-84.

Delague V *et al*. (2007). Mutations in FGD4 encoding the Rho GDP/GTP exchange factor FRABIN cause autosomal recessive Charcot-Marie-Tooth type 4H. *Am J Hum Genet*; **81**: pp 1-16.

DePristo MA *et al*. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*; **43**: pp 491-8.

Desoize B and Madoulet C. (2002) Particular aspects of platinum compounds used at present in cancer treatment. *Crit Rev Oncol/Hematol*; **42**(3): pp 317-25.

Di Fiore F *et al*. (2007). Clinical relevance of KRAS mutation detection in metastatic colorectal cancer treated by cetuximab plus chemotherapy. *Br J Cancer*; **96**: pp 1166-9.

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 5705-12.

Ding L *et al*. (2010). Analysis of next generation genomic data in cancer: accomplishments and challenges. *Hum Mol Genet*; **19**: pp R188-96.

Douillard JY *et al*. (2000). Irinotecan combined with fluorouracil compared with fluorouracil alone as first-line treatment for metastatic colorectal cancer: a multicentre randomised trial. *Lancet*; **355**(9209): pp 1041-7.

Douillard JY *et al*. (2010). Randomized, phase III trial of panitumumab with infusional fluorouracil, leucovorin, and oxaliplatin (FOLFOX4) versus FOLFOX4 alone as first-line treatment in patients with previously untreated metastatic colorectal cancer: the PRIME study. *J Clin Oncol*; **27**: pp 4860-8.

Dreze M *et al*. (2014). Monitoring repair of UV-induced 6-4-photoproducts with a purified DDB2 protein complex. *PLoS One*; **9**: pp e85896-904.

Dubois S and Demetri G. (2006). Markers of angiogenesis and clinical features in patients with sarcoma. *Cancer*; **109**(5): pp 813-9.

Dulak AM *et al*. (2013). Exome and whole-genome sequencing of oesophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nature Genetics*; **45**: pp 478-86.

Dunlop MG *et al.* (2012). Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42 103 individuals. *Gut*; **62**(6): pp 871-881.

Durbin RM *et al*. (2010). A map of human genome variation from population-scale sequencing. *Nature*; **467**: pp 1061-73.

Dworkin AM *et al*. (2010). Germline variation controls the architecture of somatic alteration in tumors. *PLoS Genet*; **6**: pp e1001136.

Dyck PJ and Giannini C. (1996). Pathologic alterations in the diabetic neuropathies of humans: a review. *J Neuropathol Exp Neurol*; **55**(12): pp 1181-93.

Dzagnidze A *et al*. (2007). Repair capacity for platinum-DNA adducts determines severity of cisplatin-induced peripheral neuropathy. *J Neuroscience*; **27**(35): pp 9451-7.

Easton DF *et al*. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*; **447**: pp 1087-93.

Efferth T and Volm M. (2005). Pharmacogenetics for individualised cancer chemotherapy. *Pharmacol Ther*; **107**(2): pp 155-76.

El-Khamisy SF and Caldecott KW. (2007). DNA single-strand break repair and spinocerebellar ataxia with axonal neuropathy-1. *Neurosci*; **145**(4): pp 1260-6.

El-Masry S *et al*. (2006). Utility of slot-blot-ELISA as a new, fast, and sensitive immunoassay for detection of carcinoembryonic antigen in the urine samples of patients with various gastrointestinal malignancies. *J Immunoassay Immunochem*; **28**(2): pp 91-105.

Ellis LM. (2006). Mechanisms of action of bevacizumab as a component of therapy for metastatic colorectal cancer. *Semin Oncol*; **33**(5): pp 1-7.

Elmore S. (2007). Apoptosis: a review of programmed cell death. *Toxicol Pathol*; **35**(4): pp 495-516.

Etienne-Grimaldi MC *et al*. (2010). Methylenetetrahydrofolate reductase (MTHFR) gene polymorphisms and FOLFOX response in colorectal cancer patients. *Br J Clin Pharmacol*; **69**: pp 58-66.

Fairbairn DW *et al*. (1995). The comet assay: a comprehensive review. *Mutation Research/Reviews in Genetic Toxicology*; **339**: pp 37-59.

Faivre S *et al*. (2003). DNA strand breaks and apoptosis induced by oxaliplatin in cancer cells. *Biochem Pharmacol*; **66**(2): pp 225-37.

Fang H *et al*. (2014). Reducing INDEL calling errors in whole genome and exome sequencing data. *Genome Med*; **6**: pp 89-106.

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 15992-7.

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.

Feige E *et al*. (2002). Nurit, a novel leucine-zipper protein, expressed uniquely in the spermatid flower-like structure. *Mech Dev*; **117**(1-2): pp 369-77.

Felici A *et al*. (2002). Dosing strategies for anticancer drugs: the good, the bad and body-surface area. *Eur J Cancer*; **38**: pp 1677-84.

Ferrarese R *et al*. (2014). Lineage-specific splicing of an alternative exon of ANXA7 promotes EGFR signalling activation and tumour progression in glioblastoma. *Neuro Oncol*; **16**(2): pp 28.

Ferrari P *et al*. (2007). Lifetime and baseline alcohol intake and risk of colon and rectal cancers in the European prospective investigation into cancer and nutrition (EPIC). *Intl J Cancer*; **121**(9): pp 2065-72.

Flicek P *et al.* (2013). Ensembl 2013. *Nucleic Acids Res*; **41**(Database issue): pp D48-55.

Flis S and Splawinski J. (2009). Inhibitory effects of 5-fluorouracil and oxaliplatin on human colorectal cancer cell survival are synergistically enhanced by sulindac sulfide. *Anticancer Res*; **29**: pp 435-41.

Fousteri M and Mullenders LHF. (2008). Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. *Cell Res*; **18**: pp 73-84.

Fox LP. (2006). Pathology and management of dermatologic toxicities associated with anti-EGFR therapy. *Oncology (Williston Park)*; **20**(2): pp 26-34.

Franz DN. (2011). Everolimus: an mTOR inhibitor for the treatment of tuberous sclerosis. *Expert Rev Anticancer Ther*; **11**: pp 1181-92.

Frappart PO and McKinnon PJ. (2006). Ataxia-telangiectasia and related diseases. *Neuromolecular Med*; **8**: pp 495-511.

Frattini M *et al*. (2007). PTEN loss of expression predicts cetuximab efficacy in metastatic colorectal cancer patients. *Br J Cancer*; **97**: pp 1139-45.

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 10722-7.

Friday BB and Adjei AA. (2008). Advances in targeting the Ras/Raf/MEK/Erk mitogen-activated protein kinase cascade with MEK inhibitors for cancer therapy. *Clin Cancer Res*; **14**(2): pp 342-6.

Gaillard H and Aguilera A. (2013). Transcription coupled repair at the interface between between transcription elongation and mRNP biogenesis. *Biophys Acta*; **1829**: pp141-50.

Gaillard PHL and Wood RD. (2001). Activity of individual ERCC1 and XPF subunits in DNA nucleotide excision repair. *Nucleic Acids Res*; **29**(4): pp 872-879.

Gamelin L *et al*. (2007). Predictive factors of oxaliplatin neurotoxicity: the involvement of the oxalate outcome pathway. *Clin Cancer Res*; **13**(21): pp 6359-68.

Ganem NJ and Pellman D. (2012). Linking abnormal mitosis to the acquisition of DNA damage. *J Cell Biol*; **199**: pp 871-81.

Gasser HS and Erlanger J. (1927). The role played by the sizes of the constituent fibers of a nerve-trunk in determining the form of its action potential wave. *Am J Physiol*; **80**: pp 522-47.

Genin E *et al*. (2008). Identifying modifier genes of monogenic disease: strategies and difficulties. *Hum Genet*; **124**: pp 357-68.

Gerlinger M *et al*. (2013). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*; **366**: pp 883-92.

Giannattasio M *et al*. (2004). Physical and functional interactions between nucleotide excision repair and DNA damage checkpoint. *EMBO J*; **23**(2): pp 429-38.

Gianni L *et al*. (2010). Neoadjuvant chemotherapy with trastuzumab followed by adjuvant trastuzumab versus neoadjuvant chemotherapy alone, in patients with HER2-positive locally advanced breast cancer (the NOAH trial): a randomised controlled superiority trial with a parallel HER2-negative cohort. *Lancet*; **375**(9712): pp 377-84.

Giantonio 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. *J Clin Oncol*; **25**(12): pp 1539-44.

Giglia-Mari G *et al*. (2004). A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. *Nature Genetics*; **36**: pp 714-9.

Gilissen C *et al.* (2012). Disease gene identification strategies for exome sequencing. *Eur J Hum Genet*; **20**(5): pp 490-497.

Gillet LC and Scharer OD. (2006). Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem Rev*; **106**(2): pp 253-76.

Gillette TG *et al*. (2001). The 19S complex of the proteasome regulates nucleotide excision repair in yeast. *Genes and Development*; **15**: pp 1528-39.

Gillette TG *et al*. (2006). Distinct functions of the ubiquitin-proteasome pathway influence nucleotide excision repair. *EMBO J*; **25**(11): pp 2529-38.

Goel HL and Mercurio AM. (2013). VEGF targets the tumour cell. *Nat Rev Cancer*; **13**(12): pp 871-82.

Goetz MP *et al*. (2011). Evaluation of CYP2D6 and efficacy of tamoxifen and raloxifene in women treated for breast cancer chemoprevention: results from the NSABP P1 and P2 clinical trials. *Clin Cancer Res*; **17**: pp 6944-51.

Goldberg RM *et al*. (2004). A randomized controlled trial of fluorouracil plus leucovorin, irinotecan, and oxaliplatin combinations in patients with previously untreated metastatic colorectal cancer. *J Clin Oncol*; **22**: pp 23-30.

Goldhirsch A *et al*. (2009). Thresholds for therapies: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer. *Ann Oncol*; **20**: pp 1319-29.

Gorski JJ *et al*. (2007). A novel TBP-associated factor of SL1 functions in RNA polymerase I transcription. *EMBO J*; **26**(6): pp 1560-8.

Goss JR *et al*. (2011). Premature ageing-related peripheral neuropathy in a mouse model of progeria. *Mechanisms of Ageing and Development*; **132**(8-9): pp 437-42.

Graham MA *et al*. (2000). Clinical pharmacokinetics of oxaliplatin: a critical review. *Clin Cancer Res*; **6**(4): pp 1205-18.

Green ED *et al*. (2011). Charting a course for genomic medicine from base pairs to bedside. *Nature*; **470**: pp 204-13.

Greene FL and Sobin LH. (2008). The staging of cancer: a retrospective and prospective appraisal. *CA Cancer J Clin*; **58**: pp 180-90.

Gregg SQ *et al*. (2011). Physiological consequences of defects in ERCC1-XPF DNA repair endonuclease. *DNA Repair (Amst)*; **10**(7): pp 781-791.

Gregory CD *et al*. (2009). Inhibitory effects of persistent apoptotic cells on monoclonal antibody production in vitro. *mAbs*; **1**(4): pp 370-6.

Gressett SM *et al*. (2006). Management of hand-foot syndrome induced by capecitabine. *J Oncol Pharm Pract*; **12**(3): pp 131-41.

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 2293-7.

Grothey A. (2003). Oxaliplatin-safety profile: neurotoxicity. *Semin Oncol;* **4** (Suppl 15): pp 5-13.

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.

Grothey A and Goldberg AM. (2004). A review of oxaliplatin and its clinical use in colorectal cancer. *Expert Opin Pharmacother;* **5**(10): pp 2159-70.

Grothey A *et al*. (2013). Regorafenib monotherapy for previously treated metastatic colorectal cancer (CORRECT): an international, multicentre, randomised, placebo-controlled, phase 3 trial. *Lancet*; **381**: pp 303-12.

Gudmundsson J *et al*. (2012). A study based on whole-genome sequencing yields a rare variant at 8q24 associated with prostate cancer. *Nature Genetics*; **44**: pp 1326-9.

Guglielmi AP and Sobrero AF. (2007). Second-line therapy for advanced colorectal cancer. *Gastrointest Cancer Res*; **1**(2): pp 57-63.

Gupta A and Lutsenko S. (2010). Human copper transporters: mechanism, role in human diseases and therapeutic potential. *Future Med Chem*; **1**(16): pp 1125-42.

Guzman-Ayala M *et al*. (2015). Chd1 is essential for the high transcriptional output and rapid growth of the mouse epiblast. *Development*; **142**: pp 118-27.

Haggar FA and Boushey RP. (2009). Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clin Colon Rectal Surg*; **22** (4): pp 191-7.

Han B *et al*. (2011). Association of ABCC2 polymorphisms with platinum-based chemotherapy response and severe toxicity in non-small cell lung cancer patients. *Lung Cancer*; **72**(2): pp 238-43.

Harding AE and Thomas PK. (1980). The clinical features of hereditary motor and sensory neuropathy types I and II. *Brain*; **103**: pp 259-80.

Harismendy O *et al*. (2009). Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biology*; **10**(3): pp R32-45.

Harper JW and Elledge SJ. (2007). The DNA damage response: ten years after. *Mol Cell;* **28**: pp 739-45.

Harrigan JA and Bohr VA. (2003). Human diseases deficient in RecQ helicases. *Biochimie*; **85**: pp 1185-93.

Hartford CM and Dolan ME. (2007). Identifying genetic variants that contribute to chemotherapy-induced cytotoxicity. *Pharmacogenom*; **8**(9): pp 1159-68.

Hedges SB. (2002). The origin and evolution of model organisms. *Nature Rev Genet*; **3**: pp 838-49.

Helleday T *et al*. (2007). DNA double-strand break repair: from mechanistic understanding to cancer treatment. *DNA Repair*; **6**: pp 923-35.

Henderson S *et al.* (1991). Induction of bcl-2 expression by Epstein-Barr virus latent membrane protein 1 protects infected B cells from programmed cell death. *Cell*; **65**(7): pp 1107-15.

Herrero AB *et al*. (2006). Cross-talk between nucleotide excision and homologous recombination DNA repair pathways in the mechanism of action of antitumor trabectedin. *Cancer Res*; **66**: pp 8155-62.

Hertz DL and McLeod HL. (2013). Use of pharmacogenetics for predicting cancer prognosis and treatment exposure, response and toxicity. *J Hum Genet*; **58**: pp 346-52.

Hickson ID. (2003). RecQ helicases: caretakers of the genome. *Nat Rev Cancer*; **3**: pp 169-78.

Higa G. (2011). Retrieved May 26, 2015, from Clinical Trials, US National Institute of Health: https://clinicaltrials.gov/ct2/show/NCT01499940.

Hillier LW *et al*. (2008). Whole-genome sequencing and variant discovery in *C. elegans. Nature Methods*; **5**: pp 183-8.

Hirota S *et al*. (2003). Gain-of-function mutations of platelet-derived growth factor receptor alpha gene in gastrointestinal stromal tumours. *Gastroenterol*; **125**: pp 660-7.

Hockel M. (2001). Tumour hypoxia: definitions and current clinical, biologic, and molecular aspects. *J Natl Cancer Inst*; **93**(4): pp 266-76.

Hoeijmakers JHJ. (2001). Genome maintenance mechanisms for prevention of cancer. *Nature*; **411**(6835): pp 366-74.

Holmes RP and Assimos DG. (1998). Glycoxylate synthesis, and its modulation and influence on oxalate synthesis. *J Urol*; **160**(5): pp 1617-24.

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.

Hoskins JM *et al*. (2007). UGT1A1*28 genotype and irinotecan-induced neutropaenia: dose matters. *J Natl Cancer Inst*; **99**: pp 1290-5.

Hunter N. (2008). The RecQ DNA helicases: Jacks-of-all-trades or master-tradesmen? *Cell Res*; **18**: pp 328-30.

Hsieh SM *et al*. (2009). Distinct inherited metastasis susceptibility exists for different breast cancer subtypes: a prognosis study. *Breast Cancer Res*; **11**: pp R75.

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.

Hurwitz H *et al*. (2004). Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *New Engl J Med*; **350**: pp 2335-42.

Hyka-Nouspikel N *et al*. (2012). Deficient DNA damage response and cell cycle checkpoints lead to accumulation of point mutations in human embryonic stem cells. *Stem Cells*; **30**(9): pp 1901-10.

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.

Innocenti F *et al*. (2004). Genetic variants in the UDP-glucuronosyltransferase 1A1 gene predict the risk of severe neutropaenia of irinotecan. *J Clin Oncol*; **22**(8): pp 1382-8.

International HapMap Consortium. (2003). The International HapMap Project. *Nature*; **426**(6968): pp 789-96.

Isenberg JS *et al*. (2009). Regulation of nitric oxide signalling by thrombospondin 1: implications for anti-angiogenic therapies. *Nature Reviews Cancer*; **9**: pp 182-94.

Itoh M *et al*. (1999). Neurodegeneration in hereditary nucleotide repair disorders. *Brain Dev*; **21**: pp 326-33.

Itoh T *et al*. (2001). Abnormal regulation of DDB2 gene expression in xeroderma pigmentosum group E strains. *Oncogene*; **20**(48): pp 7041-50.

Itzkowitz SH and Yio X. (2004). Inflammation and cancer IX. Colorectal cancer in inflammatory bowel disease: the role of inflammation. *Am J Physiol Gastrointest Liver Physiol*; **287**: pp 7-17.

Iyer G *et al*. (2012). Genome sequencing identifies a basis for everolimus sensitivity. *Science*; **338**: pp 221.

Iyer L *et al*. (1998). Genetic predisposition to the metabolism of irinotecan (CPT-11). *J Clin Invest*; **101**(4): pp 847-54.

Iyer L *et al*. (1999). Phenotype-genotype correlation of in vitro SN-38 (active metabolite of irinotecan) and bilirubin glucuronidation in human liver tissue with UGT1A1 promoter polymorphism. *Clin Pharmacol Ther*; **65**: pp 576-82.

Iyer L *et al*. (2002). UGT1A1*28 polymorphism as a determinant of irinotecan disposition and toxicity. *Pharmacogenom*; **2**: pp 43-7.

Jaarsma D *et al*. (2011). Age-related neuronal degeneration: complementary roles of nucleotide excision repair and transcription coupled repair in preventing neuropathology. *PLoS genetics*; **7**(12): pp 1-17.

Jabir RS *et al*. (2012). Pharmacogenetics of taxanes: impact of gene polymorphisms of drug transporters on pharmacokinetics and toxicity. *Pharmacogenom*; **13**(16): pp 1979-88.

Jackson SP and Bartek J. (2009). The DNA-damage response in human biology and disease. *Nature*; **461**: pp 1071-8.

Jaggi AS and Singh N. (2012). Mechanisms in cancer-chemotherapeutic drugs-induced peripheral neuropathy. *Toxicol*; **291**(1-3): pp 1-9.

Jasperson KW *et al*. (2010). Hereditary and familial colon cancer. *Gastroenterol*; **138**(6): pp 2044-58.

Jemal A *et al*. (2010). Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiol Biomarkers Prev*; **19**(8): pp 1893-907.

Jia L *et al*. (1997). Mitochondrial electron transport chain activity, but not ATP synthesis, is required for drug-induced apoptosis in human leukaemic cells: a possible novel mechanism of regulating drug resistance. *Br J Haematol*; **98**(3): pp 686-98.

Jin Y et al. (2005). CYP2D6 genotype, antidepressant use, and tamoxifen metabolism during adjuvant breast cancer treatment. *J Natl Cancer Inst*; **97**: pp 30-9.

Jones S *et al*. (2002). Biallelic germline mutations MYH predispose to multiple colorectal adenoma and somatic G:C→T:A mutations. *Hum Mol Genet*; **11**(23): pp 2961-7.

Jonsson G *et al*. (2005). Distinct genomic profiles in hereditary breast tumours identified by array-based comparative genomic hybridization. *Cancer Res*; **65**: pp 7612-21.

Jost M *et al*. (2000). The EGF receptor – an essential regulator of multiple epidermal functions. *Eur J Dermatol*; **10**: pp 505-10.

Jung Y and Lippard SJ. (2007). Direct cellular responses to platinum-induced DNA damage. *Chemical Reviews*; **107**(5): pp 1387-407.

Kabbinavar F *et al*. (2003). Phase II, randomized trial comparing bevacizumab plus fluorouracil (FU)/leucovorin (LV) with FU/LV alone in patients with metastatic colorectal cancer. *J Clin Oncol*; **21**: pp 60-5.

Karapetis CS *et al*. (2008). *KRAS* mutations and benefit from cetuximab in advanced colorectal cancer. *New Engl J Med*; **359**: pp 1757-65.

Kerem B *et al*. (1989). Identification of the cystic fibrosis gene: genetic analysis. *Science*; **245**: pp 1073-80.

Kern W *et al.* (1999). Oxaliplatin pharmacokinetics during a four-hour infusion. *Clin Cancer Res*; **5**(4): pp 761-765.

Kersemaekers AF *et al*. (1999) Oncogene alterations in carcinomas of the uterine cervix: overexpression of the epidermal growth factor receptor is associated with poor prognosis. *Clin Cancer Res*; **5**: pp 577-86.

Kidani Y *et al*. (1978). Antitumor activity of 1,2-diaminocyclohexaneplatinum complexes against Sarcoma-180 ascites form. *J Med Chem*; **21**(12): pp 1315-8.

Kiemeney LA *et al*. (2010). A sequence variant at 4p16.3 confers susceptibility to urinary bladder cancer. *Nature Genet*; **42**: pp 415-9.

Kilpivaara O *et al*. (2009). A germline JAK2 SNP is associated with predisposition to the development of JAK2(V617F)-positive myeloproliferative neoplasms. *Nature Genet*; **41**: pp 455-9.

Kim HS *et al*. (2009). Genetic polymorphisms affecting clinical outcomes in epithelial ovarian cancer patients treated with taxanes and platinum compounds: A Korean population-based study. *Gynecol Oncol*; **113**(2): pp 264-9.

Kimura H *et al*. (2007). Antibody-dependent cellular cytotoxicity of cetuximab against tumor cells with wild-type or mutant epidermal growth factor receptor. *Cancer Sci*; **98** (8): pp 1275-80.

Kinzler KW and Vogelstein B. (1996). Lessons from hereditary colorectal cancer. *Cell*; **87**(2): pp 159-70.

Kiyotani K *et al*. (2010). Lessons for pharmacogenomics studies: association study between *CYP2D6* genotype and tamoxifen response. *Pharmacogenet Genom*; **20**: pp 565-8.

Klein G and Ernberg I. 2007. Chapter 29: Effects on apoptosis, cell cycle and transformation, and comparative aspects of EBV with other known DNA tumor viruses. In: Arvin A *et al*. ed. *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis.* Cambridge: Cambridge University Press, pp 514-39.

Knudson AG. (1985). Hereditary cancer, oncogenes, and antioncogenes. *Cancer Res*; **45**(4): pp 1437-43.

Knudson AG. (1996). Hereditary cancer: two hits revisited. *J Cancer Res Clin Oncol*; **122**: pp 135-40.

Kobayashi N and Mundel P. (1998). A role of microtubules during the formation of cell processes in neuronal and non-neuronal cells. *Cell Tissue Res*; **291**(2): pp 163-74.

Koboldt DC *et al*. (2010). Challenges of sequencing human genomes. *Brief Bioinform*; **11**(5): pp 484-98.

Koch C and Hoiby N. (1993). Pathogenesis of cystic fibrosis. *Lancet*; **341**(8852): pp 1065-9.

Kosuri KV *et al*. (2010). An epigenetic mechanism for capecitabine resistance in mesothelioma. *Biochem Biophys Res Comm*; **391**(3): pp 1465-70.

Kraemar KH *et al*. (1987). Xeroderma pigmentosum: cutaneous, ocular, and neurologic abnormalities in 830 published cases. *JAMA Dermatol*; **123**(2): pp 241-50.

Kraemar KH *et al*. (1994). The role of sunlight and DNA repair in melanoma and nonmelanoma skin cancer: the xeroderma pigmentosum paradigm. *JAMA Dermatol*; **130**(8): pp 1018-21.

Kraemer KH *et al*. (2007). Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: a complex genotype-phenotype relationship. *Neurosci*; **145**(4): pp 1388-96.

Kramerova IA *et al*. (2000). Papilin in development; a pericellular protein with a homology to the ADAMTS metalloproteinases. *Development*; **127**: pp 5475-85.

Krause AW *et al*. (1984). Fluorescent erythrosin B is preferable to trypan blue as a vital exclusion dye for mammalian cells in monolayer culture. *J Histochem Cytochem*; **32**(10): pp 1084-90.

Krawitz P *et al*. (2010). Microindel detection in short-read sequence data. *Bioinformatics*; **26**(6): pp 722-9.

Ku CS *et al*. (2012). Exome sequencing: dual role as a discovery and diagnostic tool. *Ann Neurol*; **71**: pp 5-14.

Kumar S *et al*. (2014). A slot blot immunoassay for quantitative detection of *Plasmodium falciparum* circumsporozoite protein in mosquito midgut oocyst. *PLOS One*; **9**(12): pp 1-20.

Kurien BT and Scofield RH. (2006). Western blotting. *Methods*; **38**(4): pp 283-93.

Kusumoto R *et al*. (2001). Diversity of the damage recognition step in the global genomic nucleotide excision repair in vitro. *Mutation Res/DNA Repair*; **485**(3): pp 219-27.

Kweekel DM *et al*. (2005). Pharmacology of oxaliplatin and the use of pharmacogenomics to individualise therapy. *Cancer Treatment Rev*; **31**: pp 90-105.

Kwon HC *et al*. (2007). Prognostic value of expression of *ERCC1*, thymidylate synthase, and glutathione S-transferase P1 for 5-fluorouracil/oxaliplatin chemotherapy in advanced gastric cancer. *Ann Oncol*; **18**(3): pp 504-9.

Laine JP *et al*. (2006). When transcription and repair meet: a complex system. *Trends Genet*; **22**(8): pp 430-6.

Laine JP *et al*. (2007). Common XPD (ERCC2) polymorphisms have no measurable effect on nucleotide excision repair and basal transcription. *DNA Repair*; **6**: pp 1264-70.

Laje G *et al.* (2009). Pharmacogenetics studies in STAR*D: strengths, limitations, and results. *Psychiatr Serv*; **60**(11): pp 1446-57.

Laje G *et al.* (2009*). Genome-wide association study of suicidal ideation emerging during citalopram treatment of depressed outpatients. *Pharmacogenet Genomics*; **19**(9): pp 666-74.

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 2215-21.

Lamont EB and Schilsky RL. (1999). The oral fluoropyrimidines in cancer chemotherapy. *Clin Cancer Res*; **5**: pp 2289-96.

Lander ES *et al*. (1987). Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science*; **236**: pp 1567-70.

Landi MT *et al*. (2006). MC1R germline variants confer risk for BRAF-mutant melanoma. *Science*; **313**: pp 521-2.

Lankisch TO *et al*. (2008). Gilbert's syndrome and irinotecan toxicity: combination with UDP-glucuronosyltransferase 1A7 variants increases risk. *Cancer Epidemiol Biomarkers Prev*; **17**: pp 695-702.

Laugel V. (2013). Cockayne syndrome: the expanding clinical and mutational spectrum. *Mech Age Dev*; **134**(5-6): pp 161-70.

Laurent-Puig P *et al*. (2009). Analysis of *PTEN*, *BRAF*, and *EGFR* status in determining benefit from cetuximab therapy in wild-type *KRAS* metastatic colorectal cancer. *J Clin Oncol*; **27**(35): pp 5924-30.

Le Goff C *et al*. (2008). ADAMTSL2 mutations in geleophysic dysplasia demonstrate a role for ADAMTS-like proteins in TGF-beta bioavailability regulation. *Nat Genet*; **40**: pp 1119-23.

Le Goff C. and Cornier-Daire V. (2011). The ADAMTS(L) family and human genetic disorders. *Human Molecular Genetics*; **20**(R2): pp 163-7.

Leandro-Garcia LJ *et al*. (2012). Regulatory polymorphisms in beta-tubulin IIa are associated with paclitaxel-induced peripheral neuropathy. *Clin Cancer Res*; **18**: pp 4441-8.

LeDoux MS *et al*. (2006). Murine central and peripheral nervous system transcriptomes: comparative gene expression. *Brain Research*; **1107**: pp 24-41.

Lecomte T *et al*. (2004). Thymidylate synthase gene polymorphism predicts toxicity in colorectal cancer patients receiving 5-fluorouracil-based chemotherapy. *Clin Cancer Res*; **10**: pp 5880-9.

Lecomte T *et al*. (2006). Glutathione S-transferase P1 polymorphism ($Ile^{105}Val$) predicts cumulative neuropathy in patients receiving oxaliplatin-based chemotherapy. *Clin Cancer Res*; **12**: pp 3050-7.

Lenardo MJ. (2009). Autophagic cell death. *Methods Enzymol*; **453**: pp 17-31.

Lenz HJ. (2006). Anti-EGFR mechanism of action: antitumor effect and underlying cause of adverse effects. *Oncology*; **20**(5): pp 5-13.

Levi F *et al*. (2000). Oxaliplatin: Pharmacokinetics and chronopharmacological aspects. *Clin Pharmacokinet*; **38**: pp 1-21.

Ley TJ *et al*. (2008). DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature*; **456**: pp 66-72.

Li H and Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*; **25**(14): pp 1754-60.

Li H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics*; **27**(21): pp 2987-93.

Li L *et al*. (1995). Mutations in XPA that prevent association with ERCC1 are defective in nucleotide excision repair. *Mol Cell Biol*; **15**: pp 1993-8.

Li H and Homer N. (2010). A survey of sequence alignment algorithms for next-generation sequencing. *Briefings in Bioinformatics*; **11**(5): pp 473-83.

Lieber MR *et al*. (2003). Mechanism and regulation of human non-homologous DNA end-joining. *Nat Rev Mol Cell Biol*; **4**: pp 712-20.

Lievre A *et al*. (2006). KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer Res*; **66**: pp 3992-5.

Lievre 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 374-9.

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 962-72.

Lipton L *et al*. (2003). Carcinogenesis in MYH-associated polyposis follows a distinct genetic pathway. *Cancer Res*; **63**(22): pp 7595-9.

Liu W *et al*. (2011). Functional EGFR germline polymorphisms may confer risk for EGFR somatic mutations in non-small cell lung cancer, with a predominant effect on exon 19 microdeletions. *Cancer Res*; **71**: pp 2423-7.

Ljungman M and Lane DP. (2004). Transcription – guarding the genome by sensing DNA damage. *Nat Rev Cancer*; **4**: pp 727-37.

Longley DB *et al*. (2003). 5-Fluorouracil: mechanisms of action and clinical strategies. *Nature Reviews Cancer*; **3**: pp 330-8.

Lopez-Lopez E *et al*. (2011). Polymorphisms of the SLCO1B1 gene predict methotrexate-related toxicity in childhood acute lymphoblastic leukemia. *Pediatr Blood Cancer*; **57**: pp 612-9.

Lotta LA *et al*. (2009). *ADAMTS13* mutations and polymorphisms in congenital thrombotic thrombocytopenic purpura. *Hum Mutation*; **31**: pp 11-19.

Loupakis F *et al*. (2009). KRAS codon 61, 146 and BRAF mutations predict resistance to cetuximab plus irinotecan in KRAS codon 12 and 13 wild-type metastatic colorectal cancer. *Genet Genom*; **101**: pp 715-21.

Low SK *et al*. (2014). Genome-wide association study: a useful tool to identify common genetic variants associated with drug toxicity and efficacy in cancer pharmacogenomics. *Clin Cancer Res*; **20**(10): pp 2541-52.

Lunn RM *et al*. (2000). XPD polymorphisms: effects on DNA repair proficiency. *Carcinogenesis*; **21**(4): pp 551-5.

Luo FR *et al*. (1999). Pharmacokinetics and biotransformations of oxaliplatin in comparison with ormaplatin following a single bolus intravenous injection in rats. *Cancer Chemother Pharmacol*; **44**: pp 19-28.

Lv H *et al.* (2013). Genetic polymorphism of *XRCC1* correlated with response to oxaliplatin-based chemotherapy in advanced colorectal cancer. *Cancer Invest*; **31**: pp 24-8.

Lynch HT and De la Chapelle A. (2003). Hereditary colorectal cancer. *NEJM*; **348**: pp 919-32.

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 TJ *et al*. (2007). Epidermal growth factor receptor inhibitor-associated cutaneous toxicities: an evolving paradigm in clinical management. *The Oncologist*; **12**: pp 610-21.

Ma H *et al*. (2009). *ERCC6/CSB* gene polymorphisms and lung cancer risk. *Cancer Letters*; **273**: pp 172-6.

Ma Q and Lu AYH. (2011). Pharmacogenetics, pharmacogenomics, and individualised medicine. *Pharmacol Rev*; **63**: pp 437-59.

Macalalad AR *et al*. (2012). Highly sensitive and specific detection of rare variants in mixed viral populations from massively parallel sequence data. *PLoS Computat Biol*; **8**(3): pp e1002417.

Mackay FS *et al*. (2007). A potent cytotoxic photoactivated platinum complex. *Proc Natl Acad Sci USA*; **104**(52): pp 20743-8.

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. *Pharmacogenet*; **14**(5): pp 319-27.

Mandrekar SJ and Sargent DJ. (2009). Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. *J Clin Oncol*; **27**: pp 4027-34.

Maniolo TA *et al*. (2009). Finding the missing heritability of complex diseases. *Nature*; **461**: pp 747-53.

Mann JJ *et al*. (2009). Candidate endophenotypes for genetic studies of suicidal behaviour. *Biol Psych*; **65**(7): pp 556-63.

Manzano GM *et al*. (2008). A brief historical note on the classification of nerve fibers. *Arq Neuropsiquiatr*; **66**: pp 117-9.

Marsh S *et al*. (2007). Pharmacogenetic assessment of toxicity and outcome after platinum plus taxane chemotherapy in ovarian cancer: The Scottish randomised trial in ovarian cancer. *J Clin Oncol*; **25**(29): pp 4528-35.

Marsh S *et al*. (2009). Platinum pathway. *Pharmacogenet Genom*; **19**(7): pp 563-4.

Masutani C *et al*. (1999). The *XPV* (xeroderma pigmentosum variant) gene encodes human DNA polymerase η. *Nature*; **399**: pp 700-4.

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 969-74.

Maughan 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 2103-14.

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 885-90.

McCready SJ *et al*. (2000). Repair of UV damage in the fission yeast *Schizosaccharomyces pombe*. *Mutation Res*; **451**(1-2): pp 197-210.

McElroy KE *et al*. (2012). GemSIM: general, error-model based simulator of next-generation sequencing data. *BMC Genomics*; **13**: pp 74-82.

McKenna A *et al*. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*; **20**(9): pp 1297-303.

McKinnon PJ. (2004). ATM and ataxia telangiectasia. *EMBO Rep*; **5**: pp 772-6.

McKinnon PJ. (2009). DNA repair deficiency and neurological disease. *Nat Rev Neurosci*; **10**(2): pp 100-12.

McLeay SC *et al*. (2012). The relationship between drug clearance and body size: systematic review and meta-analysis of the literature published from 2000 to 2007. *Clin Pharmacokinet*; **51**: pp 319-30.

McLeod HL and Evans WE. (2001). Pharmacogenomics: unlocking the human genome for better drug therapy. *Annu Rev Pharmacol Toxicol*; **41**: pp 101-21.

McLeod HL and Yu J. (2003). Cancer pharmacogenomics: SNPs, chips, and the individual patient. *Cancer Invest*; **21**: pp 630-40.

McLeod HL. (2013). Cancer pharmacogenomics: early promise, but concerted effort needed. *Science*; **339**: pp 1563-6.

McNally R *et al*. (2012). Understanding the ‘intensive’ in ‘data intensive research’: data flows in next generation sequencing and environmental networked sensors. *Intl J Digital Curation*; **7**: pp 81-94.

Meikle L *et al*. (2008). Response of a neuronal model of tuberous sclerosis to mammalian target of rapamycin (mTOR) inhibitors: effects on mTORC1 and Akt signalling lead to improved survival and function. *J Neurosci*; **28**: pp 5422-32.

Mejare M and Bulow L. (2001). Metal-binding proteins and peptides in bioremediation and phytoremediation of heavy metals. *Trends in Biotech*; **19**(2): pp 67-73.

Mello JA *et al*. (1995). DNA adducts of cis-Diamminedichloroplatinum(II) and its trans isomer inhibit RNA polymerase II differentially in vivo. *Biochemistry*; **34**(45): pp 14783-91.

Messerli FH. (1999). Combinations in the treatment of hypertension: ACE inhibitors and calcium antagonists. *Am J Hypertension*; **12**(8): pp 86S-90S.

Metzker ML. (2010). Sequencing technologies – the next generation. *Nature Rev Genet*; **11**: pp 31-46.

Milano G *et al*. (2008). Candidate mechanisms for capecitabine-related hand-foot syndrome. *Br J Clin Pharmacol*; **66**: pp 88-95.

Miller CL *et al.* (1994). An integral membrane protein (LMP2) blocks reactivation of Epstein-Barr virus from latency following surface immunoglobulin crosslinking. *Proc Natl Acad Sci USA*; **91**(2): pp 772-6.

Mills RE *et al*. (2006). An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res*; **16**: pp 1182-90.

Mills RE *et al*. (2011). Natural genetic variation caused by small insertions and deletions in the human genome. *Genome Res*; **21**: pp 830-9.

Mimaki T *et al*. (1986). Neurological manifestations in xeroderma pigmentosum. *Ann Neurol*; **20**: pp 70-5.

Mirchandani KD and D'Andrea AD. (2006). The Fanconi anemia/BRCA pathway: a coordinator of cross-link repair. *Exp Cell Res*; **312**: pp 2647-53.

Mitchell DL. (1988). The relative cytotoxicity of (6-4) photoproducts and cyclobutane dimers in mammalian cells. *Photochem Photobiol*; **48**: pp 51-7.

Mitchison JM and Nurse P. (1985). Growth in cell length in the fission yeast Schizosaccharomyces pombe. *J Cell Science*; **75**: pp 357-76.

Miyaki M *et al*. (1997). Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nat Genet*; **17**: pp 271-2.

Moen EL *et al*. (2012). Pharmacogenomics of chemotherapeutic susceptibility and toxicity. *Genome Med*; **4**(11): pp 90-9.

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. *Nature Med*; **18**: pp 221-3.

Morel A *et al*. (2006). Clinical relevance of different dihydropyrimidine dehydrogenase gene single nucleotide polymorphisms on 5-fluorouracil tolerance. *Mol Cancer Ther*; **5**: pp 2895-904.

Morley N *et al*. (2006). UVA-induced apoptosis studied by the new apo/necro-Comet-assay which distinguishes viable, apoptotic and necrotic cells. *Mutagenesis*; **21**(2): pp 105-14.

Mosedale DE *et al*. (1996). Optimisation of immunofluorescence methods by quantitative image analysis. *J Histochem Cytochem*; **44**(9): pp 1043-50.

Moser J *et al.* (2007). Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase IIIα in a cell-cycle-specific manner. *Mol Cell*; **27**(2): pp 311-23.

Mullaney JM *et al*. (2010). Small insertions and deletions (INDELs) in human genomes. *Hum Mol Genet*; **19**: pp R131-6.

Muller A and Fischel R. (2002). Mismatch repair and the hereditary non-polyposis colorectal cancer syndrome (HNPCC). *Cancer Invest*; **20**: pp 102-9.

Murray JM *et al*. (1997). Role of *Schizosaccharomyces pombe* RecQ homolog, recombination, and checkpoint genes in UV damage tolerance. *Mol Cell Biol*; **17**(12): pp 6868-75.

Mwenifumbo JC and Marra MA. (2013). Cancer genome-sequencing study design. *Nature Rev Genet*; **14**: pp 321-32.

Nakano E *et al*. (2014). Differences in clinical phenotype among patients with XP complementation group D: 3D structure and ATP-docking of XPD in silico. *J Invest Dermatol*; **134**: pp 1775-8.

Nance MA and Berry SA. (1992). Cockayne syndrome: review of 140 cases. *Am J Med Genet*; **42**: pp 68-84.

Nanney LB *et al*. (1990). Immunolocalisation of epidermal growth factor receptors in normal developing human skin. *J Invest Dermatol*; **94**: pp 742-8.

Nardo T *et al*. (2009). A UV-sensitive syndrome patient with a specific *CSA* mutation reveals separable roles for CSA in response to UV and oxidative DNA damage. *PNAS*; **106**(15): pp 6209-14.

National Cancer Institute SEER database, 2014. *What are the survival rates for colorectal cancer by stage?* [Online]. Available at: http://www.cancer.org/cancer/colonandrectumcancer/detailedguide/colorectal-cancer-survival-rates. [Accessed: 27th August 2015].

Naven M., 2015. *Development of a pipeline and protocols for next generation sequencing of blood and formalin-fixed, paraffin-embedded tumour DNA samples*. Unpublished PhD Thesis, Cardiff University.

NCBI Resource Coordinators. (2013). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*; **41**(D1): D8-D20.

Nejdl L *et al*. (2015). Mechanisms of uptake and interaction of platinum based drugs in eukaryotic cells. *Platinum Metals in the Environment*; **1**: pp 401-15.

Neuhann TM *et al*. (2011). A homozygous microdeletion within *ADAMTSL4* in patients with isolated ectopia lentis: evidence of a founder mutation. *Invest Opthalmol Visual Sci*; **52**(2): pp 695-700.

Ng SB *et al*. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*; **461**: pp 272-6.

Ng SB *et al*. (2010). Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet*; **42**: pp 790-3.

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 5776-87.

Nielsen M *et al*. (2009). Colorectal carcinomas in MUTYH-associated polyposis display histopathological similarities to microsatellite unstable carcinomas. *BMC Cancer*; **9**: pp 184.

Nielsen TO et al. (2010). A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin Cancer Res*; **16**: pp 5222-32.

Nikiforova MN *et al*. (2013). Targeted next-generation sequencing panel (ThyroSeq) for detection of mutations in thyroid cancer. *J Clin Endocrinol Metab*; **98**(11): pp e1852-60.

Nishiwaki Y *et al*. (2004). Trichothiodystrophy fibroblasts are deficient in the repair of ultraviolet-induced cyclobutane pyrimidine dimers and (6-4) photoproducts. *J Invest Dermatol*; **122**: pp 526-32.

Norton N *et al*. (2011). Genome-wide studies of copy number variation and exome sequencing identify rare variants in BAG3 as a cause of dilated cardiomyopathy. *Am J Hum Genet*; **88**: pp 273-82.

Nowell SA *et al*. (2005). Association of genetic variation in tamoxifen-metabolising enzymes with overall survival and recurrence of disease in breast cancer patients. *Breast Cancer Res*; **91**: pp 249-58.

O'Donovan A *et al*. (1994). Isolation of active recombinant XPG protein, a human repair endonuclease. *J Biol Chem*; **269**: pp 15965-8.

O'Rawe J *et al*. (2013). Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. *Genome Med*; **5**: pp 28-45.

Oberley MJ *et al*. (2004). High-throughput screening of chromatin immunoprecipitates using CpG-island microarrays. *Methods Enzymol*; **376**: pp 315-34.

Oetting WS. (2012). Exome and genome analysis as a tool for disease identification and treatment: The 2011 human genome variation society scientific meeting. *Hum Mutation*; **33**(3): pp 586-90.

Ogi T *et al*. (2010). Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. *Mol Cell*; **37**(5): pp 714-27.

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.

Oh KS *et al*. (2006). Phenotypic heterogeneity in the XPB DNA helicase gene (ERCC3): xeroderma pigmentosum with and without Cockayne syndrome. *Hum Mutation*; **27**(11): pp 1092-103.

Okishiro M *et al*. (2009). Genetic polymorphisms of CYP2D6,10 and CYP2C1,9,2,3 are not associated with prognosis, endometrial thickness or bone mineral density in Japanese breast cancer patients treated with adjuvant tamoxifen. *Cancer*; **115**: pp 952-61.

Olivier M et al. (2006). The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer. *Clin Cancer Res*; **12**: pp 1157-68.

Orditura M *et al*. (2009). Correlation between efficacy and skin rash occurrence following treatment with the epidermal growth factor receptor inhibitor cetuximab: a single institution retrospective analysis. *Oncol Rep*; **21**(4): pp 1023-8.

Orlandi A *et al*. (2015). ERCC1 induction after oxaliplatin exposure may depend on KRAS mutational status in colorectal cancer cell line: In vitro veritas. *J Cancer*; **6**: pp 70-81.

Orlando V *et al*. (1997). Analysis of chromatin structure by *in vivo* formaldehyde crosslinking. *Methods*; **11**: pp 205-14.

Osmani K *et al*. (2012). Taxane-induced peripheral neuropathy has good long-term prognosis: a 1- to 13-year evaluation. *J Neurol*; **259**(9): pp 1936-43.

Osumi M *et al*. (1998). Dynamics of cell wall formation in fission yeast, Schizosaccharomyces pombe. *Fungal Genet Biol*; **24**: pp 178-206.

Pao W *et al*. (2005). Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation on the EGFR kinase domain. *PLoS Medicine*; **2**(3): pp e73.

Paoluzzi L *et al*. (2004). Influence of genetic variants in UGT1A1 and UGT1A9 on the *in vivo* glucuronidation of SN-38. *J Clin Pharmacol*; **44**: pp 854-60.

Palles C *et al*. (2013). Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nature Genetics*; **45**: pp 136-44.

Park DJ *et al*. (2003). Tailoring chemotherapy in advanced colorectal cancer. *Curr Op Pharmacol*; **3**(4): pp 378-85.

Park YG *et al*. (2005). SIPA1 is a candidate for underlying the metastasis efficiency modifier locus Mtes1. *Nat Genet*; **37**: pp 1055-62.

Park PJ. (2009). ChIP-seq: advantages and challenges of a maturing technology. *Nature Rev Genet*; **10**: pp 669-80.

Parsons DW *et al*. (2008). An integrated genomic analysis of human glioblastoma multiforme. *Science*; **321**(5897): pp 1807-12.

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.

Patt YZ *et al*. (2007). Capecitabine plus 3-weekly irinotecan (XELIRI regimen) as first-line chemotherapy for metastatic colorectal cancer: phase II trial results. *Am J Clin Oncol*; **30**(4): pp 350-7.

Peeters M *et al*. (2013). Massively parallel tumor multigene sequencing to evaluate response to panitumumab in a randomised phase III study of metastatic colorectal cancer. *Clin Cancer Res*; **19**: pp 1902-13.

Pei R *et al*. (2013). Association of SIPA1 545 C>T polymorphism with survival of Chinese women with metastatic breast cancer. *Front Med*; **7**: pp 138-42.

Peklak-Scott C *et al*. (2008). Role of glutathione S-transferase P1-1 in the cellular detoxification of cisplatin. *Mol Cancer Ther*; **7**: pp 3247-56.

Peltomaki P. (2003). Role of DNA mismatch repair defects in the pathogenesis of human cancer. *J Clin Oncol*; **21**: pp 1174-9.

Pemberton TJ *et al.* (2007). Identification of novel genes expressed during mouse tooth development by microarray gene expression analysis. *Dev Dyn*; **236**(8): pp 2245-57.

Perlman S *et al*. (2003). Ataxia-telangiectasia: diagnosis and treatment. *Pediatr Neurol*; **10**: pp 173-82.

Perrone F *et al*. (2009). PI3KCA/PTEN downregulation contributes to impaired responses to cetuximab in metastatic colorectal cancer patients. *Ann Oncol*; **20**: pp 84-90.

Perroud N. (2011). Suicidal ideation during antidepressant treatment: do genetic predictors exist? *CNS Drugs*; **25**(6): pp 459-71.

Peus D *et al*. (1997). EGF-receptor tyrosine kinase inhibition induces keratinocyte growth arrest and terminal differentiation. *J Invest Dermatol*; **109**: pp 751-6.

Pinto C *et al*. (2011). Management of skin toxicity associated with cetuximab treatment in combination with chemotherapy or radiotherapy. *Oncologist;* **16**(2): pp 228-38.

Polet H and Steinhardt J. (1968). Binding-induced alterations in ultraviolet absorption of native serum albumin. *Biochemistry*; **7** (4): pp 1348-56.

Pommier Y. (2013). Drugging topoisomerases: lessons and challenges. *ACS Chem Biol*; **8**: pp 82-95.

Porreca GJ *et al.* (2007). Multiplex amplification of large sets of human exons. *Nat Methods*; pp 931-6.

Prasad R *et al*. (2014). Thioredoxin reductase 2 (TXNRD2) mutation associated with familial glucocorticoid deficiency (FGD). *J Clin Endocrinol Metab*; **99**(8): pp E1556-63.

Pritchard JK. (2001). Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet*; **69**: pp 124-37.

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.

Qin L *et al*. (2012). Sialin (SLC17A5) functions as a nitrate transporter in the plasma membrane. *Proc Natl Acad Sci USA*; **109**(33): pp 13434-9.

Quail MA *et al*. (2008). A large genome center's improvements to the Illumina sequencing system. *Nature Methods*; **5**: pp 1005-10.

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 (DPYD) 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**: pp 2832-40.

Ramsey LB *et al*. (2012). Rare versus common variants in pharmacogenetics: SLCO1B1 variation and methotrexate disposition. *Genome Res*; **22**: pp 1-8.

Rapic-Otrin V *et al*. (2002). Sequential binding of UV DNA damage binding factor and degradation of the p48 subunit as early events after UV irradiation. *Nucleic Acids Research*; **30**(11): pp 2588-98.

Rass U *et al*. (2007). Defective DNA repair and neurodegenerative disease. *Cell*; **130**: pp 991-1004.

Raymond E *et al*. (1998). Oxaliplatin: mechanism of action and antineoplastic activity. *Semin Oncol*; **25**(2 Suppl 5): pp 4-12.

Raymond E *et al*. (2002). Cellular and molecular pharmacology of oxaliplatin. *Mol Cancer Ther*; **1**: pp 227-35.

Regan MM *et al*. (2012). CYP2D6 genotype and tamoxifen response in postmenopausal women with endocrine-responsive breast cancer. *J Natl Cancer Inst*; **104**: pp 441-51.

Reid S *et al*. (2007). Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. *Nat Genet*; **39**: pp 162-4.

Relling MV *et al*. (1999). Mercaptopurine therapy intolerance and heterozygosity of the thiopurine S-methyltransferase gene locus. *J Natl Cancer Inst*; **91**: pp 2001-8.

Relling MV *et al*. (2011). Clinical pharmacogenetics implementation consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing. *Clin Pharmacol Ther*; **89**: pp 387-91.

Renaud E *et al*. (2011). Differential contribution of XPC, RAD23A, RAD23B and CENTRIN 2 to the UV-response in human cells. *DNA Repair*; **10**(8): pp 835-47.

Renlund M *et al*. (1986). Defective sialic acid egress from isolated fibroblast lysosomes of patients with Salla disease. *Science*; **232**(4751): pp 759-62.

Rivory LP *et al*. (2002). New drugs for colorectal cancer – mechanisms of action. *Australian Prescriber*; **25**(5): pp 108-10.

Rixe O *et al*. (1996). Oxaliplatin, tetraplatin, cisplatin, and carboplatin: spectrum of activity in drug-resistant cell lines and in the cell lines of the national cancer institute's anticancer drug screen panel. *Biochem Pharmacol*; **52**(12): pp 1855-65.

Roach JC *et al*. (2010). Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science*; **328**: pp 636-9.

Robbins JH *et al.* (2002). Adult-onset xeroderma pigmentosum neurological disease – observations in an autopsy case. *Clin Neuropathol*; **21**: pp 18-23.

Robinson PN *et al.* (2011). Strategies for exome and genome sequence data analysis in disease-gene discovery projects. *Clin Genet*; **80**(2): pp 127-32.

Roche-Lestienne C *et al*. (2002). Several types of mutations of the Abl gene can be found in chronic myeloid leukaemia patiants resistant to ST1571, and they can pre-exist to the onset of treatment. *Blood*; **100**(3): pp 1014-8.

Rockx DAP *et al*. (2000). UV-induced inhibition of transcription involves repression of transcription initiation and phosphorylation of RNA polymerase II. *PNAS*; **97** (19): pp 10503-08.

Rogers GB *et al*. (2010). The exclusion of dead bacterial cells is essential for accurate molecular analysis of clinical samples. *Clin Microbiol Infec;* **16**(11): pp 1656-8.

Romond EH *et al*. (2005). Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *New Engl J Med*; **353**: pp 1673-84.

Ross DD *et al*. (1989). Estimation of cell survival by flow cytometric quantification of fluorescein diacetate/propidium iodide viable cell number. *Cancer Res*; **49**: pp 3776-82.

Rothenberg ML *et al*. (2008). Capecitabine plus oxaliplatin (XELOX) versus 5-fluorouracil/folinic acid plus oxaliplatin (FOLFOX-4) as second-line therapy in metastatic colorectal cancer. *Ann Oncol*; **19**(10): pp 1720-6.

Roychowdhury S *et al*. (2011). Personalised oncology through integrative high-throughput sequencing: a pilot study. *Sci Transl Med*; **3**: pp 111-21.

Rozowsky J *et al*. (2009). PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls. *Nature Biotech*; **27**: pp 66-75.

Rual JF *et al*. (2005). Towards a proteome-scale map of the human protein-protein interaction network. *Nature*; **437**: pp 1173-8.

Rudolf E *et al*. (2013). The role of p38 in irinotecan-induced DNA damage and apoptosis of colon cancer cells. *Mutation Res*; **741-2**: pp 27-34.

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 1247-54.

Ryan JL *et al*. (2006). Clonal evolution of lymphoblastoid cell lines. *Lab Invest*; **86**: pp 1193-1200.

Sai K *et al*. (2004). UGT1A1 haplotypes associated with reduced glucoronidation and increased serum bilirubin in irinotecan-administered Japanese patients with cancer. *Clin Pharmacol Ther*; **75**(6): pp 501-15.

Saito Y *et al*. (2009). Close association of UGT1A9 IVS1 + 399C>T with UGT1A1*28, *6, or *60 haplotype and its apparent influence on 7-ethyl-10-hydroxycamptothecin (SN-38) glucuronidation in Japanese. *Drug Metab Dispos*; **37**: pp 272-6.

Saltz LB *et al*. (2000). Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. *New Engl J Med*; **343**: pp 905-14.

Saltz LB *et al*. (2003). Presence and intensity of the cetuximab-induced acne-rash predicts increased survival in studies across multiple malignancies. *Proc Am Soc Clin Oncol*; **22**: pp 204-11.

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 1201-8.

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 2013-9.

Sands AT *et al*. (1995). High susceptibility to ultraviolet-induced carcinogenesis in mice lacking XPC. *Nature*; **377**: pp 162-5.

Sangrithi MN *et al.* (2005). Initiation of DNA replication requires the RECQL4 protein mutated in Rothmund-Thomson syndrome. *Cell*; **121**(6): pp 887-98.

Sartore-Bianchi A *et al*. (2009). PI3KCA mutations in colorectal cancer are associated with clinical resistance to EGFR-targeted monoclonal antibodies. *Cancer Res*; **69**: pp 1851-9.

Saunders NR *et al.* (2014). Age-dependent transcriptome and proteome following transection of neonatal spinal cord of Monodelphis domestica (South American grey short-tailed opossum). *PLoS One*; **9**(6): pp e99080.

Saxowsky TT and Doetsch PW. (2006). RNA polymerase encounters with DNA damage: transcription-coupled repair or transcriptional mutagenesis? *Chem Rev*; **106**: pp 474-88.

Sboner A *et al*. (2011). The real cost of sequencing: higher than you think. *Genome Biol*; **12**: pp 125-34.

Scartozzi M *et al*. (2009). Arterial hypertension correlates with clinical outcome in colorectal cancer patients treated with first-line bevacizumab. *Ann Oncol*; **20**(2): pp 227-30.

Schafer A *et al*. (2013). Characterization of three XPG-defective patients identifies three missense mutations that impair repair and transcription. *J Invest Dermatol*; **133**: pp 1841-9.

Schroth W *et al*. (2007). Breast cancer treatment outcome with adjuvant tamoxifen relative to patient *CYP2D6* and *CYPC19* genotypes. *J Clin Oncol*; **25**: pp 5187-93.

Schroth W *et al*. (2009). Association between CYP2D6 polymorphisms and outcome among women with early stage breast cancer treated with tamoxifen. *JAMA*; **302**: pp 1429-36.

Schully SD *et al*. (2011). Cancer GAMAdb: database of cancer genetic associations from meta-analyses and genome-wide association studies. *Eur J Hum Genet*; **19**(8): pp 928-30.

Schunkert H *et al*. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genetics*; **43**: pp 333-8.

Schutte DL *et al*. (2003). A LRPAP1 intronic insertion/deletion polymorphism and phenotypic variability in Alzheimer disease. *Res Theory Nurs Pract*; **17**: pp 301-19.

Schwab M *et al*. (2008). Role of genetic and nongenetic factors for fluorouracil treatment-related severe toxicity: a prospective clinical trial by the German 5-FU toxicity study group. *J Clin Oncol*; **26**(13): pp 2131-8.

Schweiger MR *et al*. (2011). The power of NGS technologies to delineate the genome organisation in cancer: from mutations to structural variations and epigenetic alterations. *Cancer Metastasis Rev*; **30**: pp 199-210.

Scripture CD and Figg WD. (2006). Drug interactions in cancer therapy. *Nature*; **6**: pp 546-59.

Scripture CD *et al*. (2006). Peripheral neuropathy induced by paclitaxel: recent insights and future perspectives. *Curr Neuropharmacol*; **4**(2): pp 165-72.

Seck K *et al*. (2005). Analysis of the *DPYD* gene implicated in 5-fluorouracil catabolism in a cohort of Caucasian individuals. *Clin Cancer Res*; **11**(16): pp 5886-92.

Seidle HF *et al*. (2005). Disease-associated mutations inactivate AMP-lysine hydrolase activity of aprataxin. *J Biol Chem*; **280**: pp 20927-31.

Seker H *et al*. (2001). Functional significance of XPD polymorphic variants: Attenuated apoptosis in human lymphoblastoid cells with the XPD 312 Asp/Asp genotype. *Cancer Res*; **61**: pp 7430-8.

Selvaggi G *et al*. (2004). Epidermal growth factor receptor overexpression correlates with a poor prognosis in completely resected non-small-cell lung cancer. *Ann Oncol*; **15**: pp 28-32.

Senter L *et al*. (2008). The clinical phenotype of Lynch syndrome due to germline PMS2 mutations. *Gastroenterol*; **135**: pp 419-28.

Shawan A *et al*. (2006). Atypical presentation of ataxia-oculomotor apraxia type 1. *Dev Med Child Neurol*; **48**(6): pp 529-32.

Shendure J and Ji H. (2008). Next-generation DNA sequencing. *Nature Biotech*; **26**: pp 1135-45.

Shibutani S *et al*. (1991). Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature*; **349**(6308): pp 431-4.

Shigematsu H *et al*. (2005). Somatic mutations of the HER2 kinase domain in lung adenocarcinomas. *Cancer Res*; **65**(5): pp 1642-6.

Shiloh Y. (2003). The ATM-mediated DNA-damage response: taking shape. *Trends Biochem Sci*; **31**: pp 402-10.

Shimizu Y *et al*. (2011). SPA-1 controls the invasion and metastasis of human prostate cancer. *Cancer Sci*; **102**: pp 828-36.

Shinmura K *et al*. (2012). Role of base excision repair enzyme MUTYH in the repair of 8-hydroxyguanine and MUTYH-associated polyposis (MAP). *Hereditary Genet*; **1**(3): pp 1-10.

Siddiq A *et al*. (2012). A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Hum Mol Genet*; **21**: pp 5373-84.

Sijbers AM et al. (1996). Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell*; **86**: pp 811-22.

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**: pp 832-6.

Singh RR *et al*. (2013). Clinical validation of a next-generation sequencing screen for mutational hotspots in 46 cancer-related genes. *J Mol Diagnostics*; **15**(5): pp 607-22.

Sivabalan S *et al*. (2010). In vivo native fluorescence spectroscopy and nicotinamide adenine dinucleotide/flavin adenine dinucleotide reduction and oxidation states of oral submucous fibrosis for chemopreventive drug monitoring. *J Biomed Opt*; **15**: pp 17010-8.

Skof E *et al*. (2009). Capecitabine plus irinotecan (XELIRI regimen) compared to 5-FU/LV plus irinotecan (FOLFIRI regimen) as neoadjuvant treatment for

patients with unresectable liver-only metastases of metastatic colorectal cancer: a randomized prospective phase II trial. *BMC Cancer*; **9**: pp 120-9.

Sok JC *et al*. (2006). Mutant epidermal growth factor receptor (EGFRvIII) contributes to head and neck cancer growth and resistance to EGFR targeting. *Clin Cancer Res*; **12**: pp 5064-73.

Soltys DT *et al*. (2013). Novel XPG (ERCC5) mutations affect DNA repair and cell survival after ultraviolet but not oxidative stress. *Hum Mutat*; **34**(3): pp 481-9.

Sorlie T et al. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *PNAS*; **98**(19): pp 10869-74.

Souslova T *et al.* (2013). Personalised medicine in Alzheimer's disease and depression. *Contemp Clin Trials*; **36**(2): pp 616-23.

Spano JP *et al*. (2005). Impact of EGFR expression on colorectal cancer patient prognosis and survival. *Ann Oncol*; **16**: pp 102-8.

Spicer DV *et al*. (1988). Re-evaluation of the maximum tolerated dose of continuous venous infusion of 5-fluorouracil with pharmacokinetics. *Cancer Res*; **48**: pp 459-66.

Spranger JW *et al*. (1971). Geleophysic dwarfism – a 'focal' mucopolysaccharidosis? *Lancet*; **10**: pp 97-8.

Springler B *et al*. (2003). A crystal structure of an oxaliplatin 1,2-d(GpG) intrastrand crosslink in a DNA dodecamer duplex. *Inorg Chem*; **40**(22): pp 5596-602.

Sprowl JA *et al*. (2013). Polymorphic transporters and platinum pharmacodynamics. *Drug Metab Pharmacokinet*; **28**: pp 19-27.

Srivastava M *et al*. (2001). ANX7, a candidate tumour suppressor gene for prostate cancer. *Proc Nat Acad Sci*; **98**: pp 4575-80.

Stangl R *et al*. (1994). Factors influencing the natural history of colorectal liver metastases. *Lancet*; **343**(8910): pp 1405-10.

Stearns V *et al*. (2002). Hot flushes. *Lancet*; **360**: pp 1851-61.

Stefanini M. (2000). Trichothiodystrophy: a disorder highlighting the crosstalk between DNA repair and transcription. *Madame Curie Bioscience*; **9**: pp 2-10.

Stiff T *et al*. (2006). ATR-dependent phosphorylation and activation of ATM in response to UV treatment or replication fork stalling. *EMBO J*; **25**: pp 5775-82.

Stohr H *et al*. (1998). Refined mapping of the gene encoding the p127 kDa UV-damaged DNA-binding protein (DDB1) within 11q12-q13.1 and its exclusion in Best's vitelliform macular dystrophy. *Eur J Hum Genet*; **6**: pp 400-5.

Sturgis EM *et al*. (2000). *XPD/ERCC2* polymorphisms and risk of head and neck cancer: a case-control analysis. *Carcinogenesis*; **21**(12): pp 2219-23.

Sucheston L *et al*. (2011). Genetic predictors of taxane-induced neurotoxicity in a SWOG phase III intergroup adjuvant breast cancer treatment trial (S0221). *Breast Cancer Res Treat*; **130**: pp 993-1002.

Sugasawa K *et al*. (2001). A multistep damage recognition mechanism for global genomic nucleotide excision repair. *Genes and Dev*; **15**: pp 507-21.

Surjushe A *et al*. (2008). Hand-foot syndrome due to capecitabine. *Indian J Dermatol*; **53**: pp 43-44.

Suzuki Y *et al*. (2002). Targeted disruption of LIG-1 results in psoriasiform epidermal hyperplasia. *FEBS Letters*; **521**(1-3): pp 67-71.

Svejstrup JQ. (2002). Mechanisms of transcription-coupled DNA repair. *Nature Reviews Mol Cell Biol*; **3**: pp 21-9.

Svestrup JQ. (2010). The interface between transcription and mechanisms maintaining genome integrity. *Trends Biochem Sci*; **35**: pp 333-8.

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.

Tabor HK *et al*. (2002). Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Rev Genet*; **3**: pp 391-7.

TacGhee Y *et al*. (2014). Molecular characterization of neutrally differentiated human bone marrow-derived clonal mesenchymal stem cells. *Immune Network*; **14**: pp 54-65.

Takasuna K *et al*. (2006). Optimal antidiarrhea treatment for antitumor agent irinotecan hydrochloride (CPT-11)-induced delayed diarrhea. *Cancer Chemother Pharmacol*; **58**: pp 494-503.

Tang JY *et al*. (2000). Xeroderma pigmentosum *p48* gene enhances global genomic repair and suppresses UV-induced mutagenesis. *Mol Cell*; **5**: pp 737-44.

Teng Y *et al*. (2010). A novel method for the genome-wide high resolution analysis of DNA damage. *Nucleic Acids Res*; **39**(2): pp e10-6.

Theil AF *et al*. (2011). Slowly progressing nucleotide excision repair in Trichothiodystrophy group A patient fibroblasts. *Mol Cell Biol*; **31**(17): pp 3630-8.

Theodoratou E *et al*. (2012). Systematic meta-analyses and field synopsis of genetic association studies in colorectal cancer. *J Natl Cancer Inst*; **104**: pp 1433-57.

Thomas PK. (1999). Diabetic neuropathy: mechanisms and future treatment options. *J Neurol Neurosurg Psychiatry*; **67**(3): pp 277-9.

Thompson CB *et al*. (1984). Size-dependent B lymphocyte subpopulations: relationship of cell volume to surface phenotype, cell cycle, proliferative response, and requirements for antibody production to TNP-Ficoll and TNP-BA. *J Immunol*; **133**(5): pp 2333-42.

Tian C *et al*. (2012). Common variants in ABCB1, ABCC2, and ABCG2 genes and clinical outcomes among women with advanced stage ovarian cancer treated with platinum and taxane-based chemotherapy: A gynaecologic oncology group study. *Gynecologic Oncol*; **124**(3): pp 575-81.

Tian M *et al*. (2004). Growth retardation, early death, and DNA repair defects in mice deficient for the nucleotide excision repair enzyme XPF. *Mol Cell Biol*; **24**(3): pp 1200-5.

Tierney R *et al*. (2000). The Epstein-Barr virus promoter initiating B-cell transformation is activated by RFX proteins and the B-cell-specific activator protein BSAP/Pax5. *J Virol*; **74**(22): pp 10458-67.

Tierney RJ *et al.* (2015). Unexpected patterns of Epstein-Barr virus transcription revealed by a high throughput PCR array for absolute quantification of viral mRNA. *Virology*; **474**: pp 117-30.

Timmermann B *et al*. (2010). Somatic mutation profiles of MSI and MSS colorectal cancer identified by whole exome next generation sequencing and bioinformatics analysis. *PLoS One*; **5**(12): pp e15661.

Todd RC and Lippard SJ. (2009). Inhibition of transcription by platinum antitumor compounds. *Metallomics*; **1**: pp 280-91.

Toffoli G *et al*. (2006). The role of UGT1A1*28 polymorphism in the pharmacodynamics and pharmacokinetics of irinotecan in patients with metastatic colorectal cancer. *J Clin Oncol*; **24**: pp 3061-8.

Toffoli G and De Mattia E. (2008). Pharmacogenetic relevance of *MTHFR* polymorphisms. *Pharmacogenom*; **9**(9): pp 1195-206.

Tornaletti S *et al*. (2003). Behaviour of T7 RNA polymerase and mammalian RNA polymerase II at site-specific cisplatin adducts in the template DNA. *J Biol Chem*; **278**: pp 35791-7.

Tremeau-Bravard A *et al*. (2004). Fate of RNA polymerase II stalled at a cisplatin lesion. *J Biol Chem*; **279**: pp 7751-9.

Trevino LR *et al*. (2009). Germline genetic variation in an organic anion transporter polypeptide associated with methotrexate pharmacokinetics and clinical effects. *J Clin Oncol*; **27**: pp 5972-8.

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**: pp 1849-58.

Tsimberidou AM *et al*. (2012). Personalised medicine in a phase I clinical trials program: the MD Anderson Cancer Centre initiative. *Clin Cancer Res*; **18**: pp 6373-83.

Tsodikov OV *et al*. (2007). Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA. *EMBO*; **26**: pp 4768-76.

Uechi G *et al*. (2014). Proteomic view of basement membranes from human retinal blood vessels, inner limiting membranes, and lens capsules. *J Proteome Res*; **13**: pp 3693-705.

Ueda S *et al*. (2004). The correlation between cytoplasmic overexpression of epidermal growth factor receptor and tumor aggressiveness: poor prognosis in patients with pancreatic ductal adenocarcinoma. *Pancreas*; **29**: pp 1-8.

Valle L. (2014). Genetic predisposition to colorectal cancer: where we stand and future perspectives. *World J Gastroenterol*; **20**(29): pp 9828-49.

Van Brabant AJ *et al*. (2000). DNA helicases, genomic instability, and human genetic disease. *Annu Rev Genomics Hum Genet*; **1**: pp 409-59.

Van Cutsem E *et al*. (2001). Oral capecitabine compared with intravenous fluorouracil plus leucovorin in patients with metastatic colorectal cancer: results of a large phase III study. *J Clin Oncol*; **19**(21): pp 4097-106.

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): pp 1658-64.

Van Cutsem E *et al*. (2011). Cetuximab plus irinotecan, fluorouracil, and leucovorin as first-line treatment for metastatic colorectal cancer: updated analysis of overall survival according to tumour KRAS and BRAF mutation status. *J Clin Oncol*; **28**: pp 1-10.

Van Cutsem E *et al*. (2012). Addition of aflibercept to fluorouracil, leucovorin, and irinotecan improves survival in a phase III randomized trial in patients with metastatic colorectal cancer previously treated with an oxaliplatin-based regimen. *J Clin Oncol*; **30**: pp 25-33.

Van Paassen BW *et al.* (2014). PMP22 related neuropathies: Charcot-Marie-Tooth disease type 1A and hereditary neuropathy with liability to pressure palsies. *Orphanet J Rare Dis*; **9**: pp 38-53.

Van Vuuren AJ *et al*. (1993). Evidence for a repair enzyme complex involving ERCC1, ERCC4, ERCC11 and the xeroderma pigmentosum group F proteins. *EMBO J.;* **12**: pp 3693-3701.

Verheijen FW *et al*. (1999). A new gene, encoding an anion transporter, is mutated in sialic acid storage diseases. *Nat Genet*; **23**: pp 462-5.

Veselkin E *et al*. (2011). A secreted form of the asiaglycoprotein receptor, sH2a, as a novel potential noninvasive marker for liver fibrosis. *PLoS One*; **6**(11): pp e27210.

Vissers LE *et al*. (2010). A de novo paradigm for mental retardation. *Nat Genet*; **42**: pp 1109-12.

Vyzula R *et al.* (2006). Raltitrexed plus oxaliplatin in the second-line treatment of metastatic colorectal cancer. *Neoplasma*; **53**(2): pp 119-27.

Waisif Saif M and Reardon J. (2005). Management of oxaliplatin-induced peripheral neuropathy. *Ther Clin Risk Management;* **1**(4): pp 249-258.

Walsh PS *et al*. (1992). A rapid chemiluminescent method for quantitation of human DNA. *Nucleic Acids Research*; **20**(19): pp 5061-5.

Walsh T *et al*. (2010). Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *PNAS*; **107**(28): pp 12629-33.

Wang K *et al.* (2010). ANNOVAR: functional annotation of genetic variants from high throughput sequencing data. *Nucleic Acids Res*; **38**(16): pp e164.

Wang M *et al*. (2010b). A novel XPF -357A>C polymorphism predicts risk and recurrence of bladder cancer. *Oncogene*; **29**: pp 1920-8.

Wang K *et al*. (2012). Systems approaches reveal the molecular networks involved in neural stem cell differentiation. *Protein and Cell*; **3**(3): pp 213-24.

Wang M *et al*. (2010b). A novel *XPF* – 357A>C polymorphism predicts risk and recurrence of bladder cancer. *Oncogene*; **29**: pp 1920-8.

Wasan H *et al*. (2014). Intermittent chemotherapy plus either intermittent or continuous cetuximab for first-line treatment of patients with KRAS wild-type advanced colorectal cancer (COIN-B): a randomised phase 2 trial. *Lancet Oncol*; **15**(6): pp 631-9.

Weeda G *et al*. (1997). A mutation in the XPB/ERCC3 DNA repair transcription gene, associated with trichothiodystrophy. *Am J Hum Genet*; **60**: pp 320-9.

West H., 2013. *Investigating the genetics and pharmacogenetics of bowel cancer*. PhD Thesis, Cardiff University. Found at: http://orca.cf.ac.uk/56820/ [Last accessed 14th September 2015].

Wheeler DA *et al*. (2008). The complete genome of an individual by massively parallel DNA sequencing. *Nature*; **452**: pp 872-6.

Wheeler HE *et al*. (2013). Cancer pharmacogenomics: strategies and challenges. *Nature Rev Genet*; **14**: pp 23-34.

Whinney SR *et al*. (2009). Platinum neurotoxicity pharmacogenetics. *Mol Cancer Ther*; **8**: pp 10-6.

Winawer S *et al*. (2003). Colorectal cancer screening and surveillance: Clinical guidelines and rationale – Update based on new evidence. *Gastroenterol*; **124**(2): pp 544-60.

Wolf S *et al*. (2008). Chemotherapy-induced peripheral neuropathy: prevention and treatment strategies. *Eur J Cancer*; **44**(11): pp 1507-15.

Wood RD. (1997). Nucleotide excision repair in mammalian cells. *J Biol Chem*; **272**: pp 23465-8.

Wood RD *et al*. (2005). Human DNA repair genes, 2005. *Mutation Res*; **577**: pp 275-83.

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 920-7.

Wray NR *et al*. (2011). Synthetic associations created by rare variants do not explain most GWAS results. *PLoS Biology*; **9**: pp e1000579.

Wu J *et al*. (2006). ChIP-chip comes of age for genome-wide functional analysis. *Cancer Res*; **66**(14): pp 6899-902.

Wolf S *et al*. (2008). Chemotherapy-induced peripheral neuropathy: Prevention and treatment strategies. *Eur J Cancer;* **44**(11): pp 1507-15.

Woodworth CD *et al*. (2005). Inhibition of the epidermal growth factor receptor increases expression of genes that stimulate inflammation, apoptosis, and cell attachment. *Mol Cancer Ther*; **4**: pp 650-8.

Wu X *et al*. (2006). Phosphorylation of nucleotide excision repair factor xeroderma pigmentosum group A by ataxia telangiectasia mutated and Rad3-related-dependent checkpoint pathway promotes cell survival in response to UV irradiation. *Cancer Res*; **66**: pp 2997-3004.

Xu C *et al*. (2011). MALAT-1: A long non-coding RNA and its important 3' end functional motif in colorectal cancer metastasis. *Intl J Oncol*; **39**: pp 169-75.

Yagi T *et al*. (1997). A low content of ERCC1 and a 120 kDa protein is a frequent feature of group F xeroderma pigmentosum fibroblast cells. *Mutagenesis*; **12**: pp 41-44.

Yamamoto M *et al*. (2008). Metabolism of irinotecan and its active metabolite SN-38 by intestinal microflora in rats. *Oncol Rep*; **20**(4): pp 727-30.

Yandell M *et al.* (2011). A probabilistic disease-gene finder for personal genomes. *Genome Res*; **21**(9): pp 1529-42.

Yang D *et al*. (2011). Association of BRCA1 and BRCA2 mutations with survival, chemotherapy sensitivity, and gene mutator phenotype in patients with ovarian cancer. *JAMA*; **306**: pp 1557-65.

Yang Y *et al*. (1993). The common variant of cystic fibrosis transmembrane conductance regulator is recognised by hsp70 and degraded in a pre-Golgi nonlysosomal compartment. *PNAS*; **90**(20): pp 9480-4.

Yasuhira S and Yasui. (2000). Alternative excision repair pathway of UV-damaged DNA in Schizosaccharomyces pombe operates both in nucleus and in mitochondria. *J Biol Chem*; **275**: pp 11824-8.

Yeo ZX *et al*. (2012). Improving indel detection specificity of the Ion Torrent PGM benchtop sequencer. *PLoS One*; **7**(9): pp e45798.

Yue Y *et al*. (2009). Spectroscopic investigation on the binding of antineoplastic drug oxaliplatin to human serum albumin and molecular modelling. *Colloids Surf B: Biointerfaces*; **69**: pp 51-7.

Zambrowicz BP and Sands AT. (2003). Knockouts model the 100 best-selling drugs – will they model the next 100? *Nature Rev Drug Discov*; **2**: pp 38-51.

Zanke BW *et al*. (2007). Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nature Genetics*; **39**: pp 989-94.

Zhang L *et al*. (2011). Nucleotide excision repair gene ERCC1 polymorphisms contribute to cancer susceptibility: a meta-analysis. *Mutagenesis*; **1**: pp 1-10.

Zhang N *et al*. (2000). Differential processing of UV mimetric and interstrand crosslink damage by XPF cell extracts. *Nucleic Acids Res*; **28**(23): pp 4800-4.

Zhang N *et al*. (2007). Double-strand breaks induce homologous recombinational repair of interstrand cross-links via cooperation of MSH2, ERCC1-XPF, REV3, and the Fanconi anemia pathway. *DNA Repair*; **6**: pp 1670-8.

Zhang S *et al*. (2006). Organic cation transporters are determinants of oxaliplatin cytotoxicity. *Cancer Res*; **66**(17): pp 8847-8857.

Zhang X *et al*. (2014). Impact of human pathogenic micro-insertions and micro-deletions on post-transcriptional regulation. *Hum Mol Genet*; **23**: pp 3024-34.

Zhang Z and Gerstein M. (2003). Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes. *Nucleic Acids Res*; **31**: pp 5338-48.

Zhu M *et al*. (2012). Using ERDS to infer copy-number variants in high-coverage genomes. *Am J Hum Genet*; **91**: pp 408-21.

Zhu X *et al*. (2000). Neuronal CDK7 in hippocampus is related to ageing and Alzheimer disease. *Neurobiol Ageing*; **21**(6): pp 807-13.

Zochodne DW. (1996). Neurotrophins and other growth factors in diabetic neuropathy. *Semin Neurol*; **16**(2): pp 153-61.

Zook JM *et al*. (2014). Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nature Biotech*; **32**(3): pp 246-53.

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.