Determining the effect of DNA repair capacity on chemotherapy toxicity during colorectal cancer treatment

Richard Webster

Cancer and Genetics Institute
Cardiff University School of Medicine
Cardiff University

A THESIS SUBMITTED TO CARDIFF UNIVERSITY FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

2015

DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed …….. (candidate) Date ……

STATEMENT 1

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

Signed …….. (candidate) 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. The views expressed are my own.

Signed …….. (candidate) Date ……

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository 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 online in the University's Open Access repository and for inter-library loans **after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.**

Signed …….. (candidate) Date ……

Summary

This study describes the translation of an assay developed for use in cell culture models to into a method of measuring patterns of DNA damage from platinum agents in human blood samples. These adduct patterns could potentially be used in future studies for the stratification of patients for response and toxicity to oxaliplatin chemotherapy.

Chapters 3 and 4 of this thesis describe the steps taken to translate our DIP-chip assay, a tool previously used in the study of DNA repair capacity in yeast and to measure induction of platinum-DNA adducts in cell culture models, into an assay capable of reproducibly analysing chemotherapy damage in human clinical samples. These results clearly demonstrate the protocol modifications required to use the assay on human blood samples, and show the reproducibly of the assay in detecting patterns of oxaliplatin induced DNA-adducts in clinical material.

Chapter 5 describes the development of novel bioinformatic tools and analysis methods for interpreting DIP-chip DNA-adduct microarray outputs. The translation of a genomic-scale laboratory technology into a tool for patient stratification is a technical and bioinformatic challenge. The tools developed are a significant advance on previously available bioinformatic functions, and are essential for the application of this technique as a clinically useful assay.

The final results section, chapter 6, documents the successful development of functional models to experimentally confirm links between single nucleotide polymorphisms in nucleotide excision repair genes with the development of oxaliplatin induced peripheral neuropathy (OIPN). This aspect of the study utilises new information, recently derived from experiments DNA-sequencing colorectal cancer patients, to develop a functional model of OIPN in *Saccharomyces cerevisiae*. This model is then used to demonstrate the impact of variations in DNA repair genes on the development of OIPN - a relationship that highlights the significance of DNA repair to the development of oxaliplatin toxicity.

Acknowledgements

My thanks go firstly to Professor Simon Reed for four years of supervision, guidance and support during all aspects of my research fellowship. My gratitude also to Professor Ray Waters and Professor J. Cheadle for the additional suggestions and guidance, all of which were invaluable.

The technique adapted for use in blood samples in this thesis has been developed over a decade under Simon's direction. The work described here is only one of many steps in the development of the assay to-date, so the contribution of Yumin, Shirong, James, and Simon, in the earlier development of the technique is to be fully acknowledged. My thanks particularly to James for the vital steps in modifying the assay for the detection of platinum damage and in translating the assay for use in human cell culture systems - critical steps allowing much of my project to proceed. In addition to Simon, a constant during all of these on-going developments is Mark Bennett, without whom the analysis of the assay would be impossible, and whose contribution to our projects has been huge and should be fully acknowledged.

All of the members of the laboratory team I have encountered over the course of the project have provided much needed advice, training and company, and through many discussions have helped develop the ideas expressed in this thesis. Thanks to Katie, Shirong, James, Mark, Yanbo, WenBin, Hammed and Trish for sharing my time in the lab, for the discussions and input that allowed me to complete this thesis. My thanks especially to Katie for the many trips to Starbucks which were often essential, and always appreciated, and for WenBin's lemonade poem – a verse that gave me the idea for a whole chapter of this thesis.

Finally, I thank my family. To my son, Zak, who has spent many hours on my knee bashing the keyboard during the months it took to write this thesis – (I think you were trying to help!), your entertainment, cheeky smile and giggles during the coffee breaks have kept me going. My greatest thanks go to my wife, Shabeena. You have been my greatest support, biggest source of strength, and the foundation of this success. This project has only been possible through your sacrifices, and I dedicate this work to both of you.

TABLE OF CONTENTS

List of Figures

List of Tables

List of Abbreviations

5-FU	5-Fluorouracil
5-FU/LV	5-Fluorouracil/leucovorin
AAS	Atomic Absorption Spectroscopy
AGC	Along Genome Correlation analysis
BSA	Bovine Serum Albumin
CPD	Cyclobutane pyrimidine dimer
CTC	Circulating Tumour Cell
CS	Cockayne syndrome
CV	Coefficient of Variance
DACH	Diaminocyclohexane
DIP	DNA Immunoprecipitation
DIP-chip	DNA immunoprecipitation coupled with microarray analysis
DRG	Dorsal Root Ganglia
DSB	Double Strand Break
DMSO	Di-methylsulphoxide
DNA	Deoxyribonucleic acid
ELSIA	Enzyme Linked Immunosorbent Assay
FDA	Federal Drug Administration
FE	Feature Extraction
FOLFOX	5-FU, leucovorin and oxaliplatin
GG-NER	Global Genome Nucleotide Excision Repair
GWAS	Genome Wide Association Study
ICL	Interstrand Cross Link
IP	Immunoprecipitated sample
IN	Input control sample
IV	Intravenous
LM-PCR	Ligation Mediated PCR
MIQE	Minimum Information for Publication of Quantitative Real-Time PCR Experiments
MMEJ	Micro homology-mediated End Joining
MMR	Miss-match repair
NCI-CTC	National Cancer Institute – Common Toxicity Criteria
NER	Nucleotide Excision Repair
NSCLC	Non-Small Cell Lung Cancer
NHEJ	Non-homologous End Joining
OIPN	Oxaliplatin Induced Peripheral Neuropathy
PARP	Poly-ADP-ribose
PBMC	Peripheral Blood Mononuclear Cell
PCR	Polymerase Chain Reaction
PBS	Phosphate Buffered Saline
PFGE	Pulsed-field gel electrophoresis
PLA	Post-labelling Assay
PN	Peripheral Neuropathy
QC	Quality Control
RFU	Relative Fluorescent Unit
RNA	Ribonucleic Acid

SCA	Single Channel Analysis
SEM	Standard Error of the Mean
SSA	Single strand annealing
TC-NER	Transcription Coupled Nucleotide Excision Repair
TLS	Translesion synthesis
UV	Ultra violet
WBS	Welsh Blood Service
WGA	Whole Genome Amplification
XP	Xeroderma Pigmentosum

Chapter 1 Introduction

1.1.1 Platinum chemotherapy

The first platinum compound to find clinical use as an anti-cancer agent, cisplatin (cis-platinum(II)-diamine-chloride), was first synthesised by Michele Peyrone in 1844 (Kauffman et al. 2010; Peyrone 1844). The anti-proliferative effect of cisplatin was not recognised until 1965 when, during an experiment to investigate the effect of an electric field on the growth of bacteria in liquid media, it was serendipitously discovered that the platinum salts generated from the electrodes in the experimental chamber inhibited bacterial cell growth (Rosenberg et al. 1965). The primary platinum compound responsible for this effect was identified as 'Peyrone's platinum' - cisplatin.

Over the following decades these properties have been thoroughly investigated and exploited; cisplatin has become one of the most frequently used antineoplastic agents in clinical use and is an essential component of many widely used combination chemotherapy regimes (Bowden 2014). Despite this, the response of an individual patient to cisplatin is unpredictable, and cisplatin is ineffective against several important malignancies, notably colorectal cancer. Even in patients who respond well to treatment, at therapeutic doses there is a significantly elevated risk of severe ototoxicity, nephrotoxicity and myelo-suppression that limits the use of this agent (Hartmann and Lipp 2003).

Several derivatives of cisplatin have been synthesised in an attempt to improve the therapeutic profile; increasing the clinical efficacy or reducing the risk of dose-limiting toxicities. Two agents have made a significant impact in the clinical setting; carboplatin (cis-diamine(cyclobutane-dicarboxylato-1,1(2-0)-0.0) platinum), which has a similar spectrum of activity as cisplatin, is less nephrotoxic but more myelo-suppressive, and oxaliplatin (oxalato(trans-$_{L}$-1,2-diaminocyclohexane)platinum) which is less ototoxic and nephrotoxic and has demonstrable activity against colorectal malignancies (Figure 1.1).

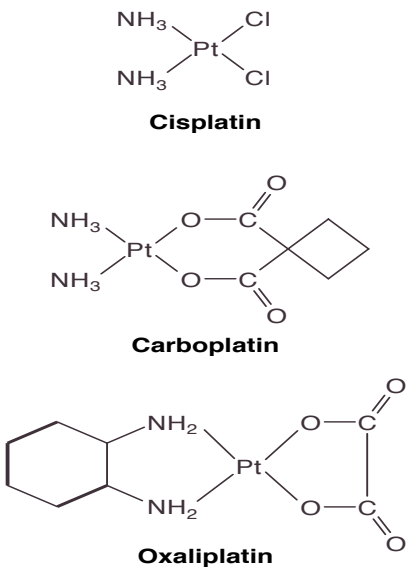


Figure 1.1: The structure of the platinum drugs.
The structure of cisplatin (cis-platinum(II)-diamine-chloride), carboplatin (cis-diamine(cyclobutane-dicarboxylato-1,1(2-0)-0.0) platinum), and oxaliplatin (oxalato(trans-L-1,2-diaminocyclohexane)platinum) (Adapted from (O'Dwyer et al. 2000))

Following FDA approval in 2004 oxaliplatin has become widely used in the treatment of colorectal cancers, a group of malignancies that remains one of the most common cancers worldwide and are notable for resistance to cisplatin and carboplatin (Di Francesco et al. 2002; Rabik and Dolan 2007).

One of the key dose-limiting toxicities of oxaliplatin is 'oxaliplatin induced peripheral neuropathy' (OIPN); damage to peripheral nerves resulting in numbness and painful paresthesia in the distal extremities, which may be permanent following treatment (Andre et al. 2009; Argyriou et al. 2014). Methods to predict which patients are at higher risk of this common and long-term toxicity are urgently needed, and would allow the use of alternative agents in patients at high risk, or could potentially improve efficacy by allowing higher doses to be used in patients at a lower risk of OIPN.

Central to developing tools to predict for the response and toxicity to platinum agents in general, and oxaliplatin on colorectal cancer in particular, is a detailed understanding of the mechanism of action of these agents from the initial drug administration through to drug metabolism and excretion, with a

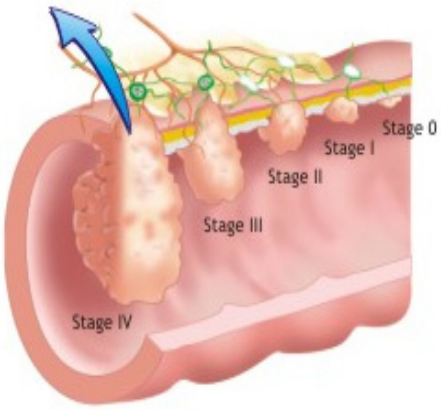
focus on the effect of platinum agents on the cellular machinery at the molecular level. The majority of the studies to elucidate the mechanisms of action of these agents at the molecular level have been conducted with cisplatin, but equally apply to oxaliplatin. Several key differences in action between cisplatin and oxaliplatin have been demonstrated and will be highlighted in the following discussion.

1.1.2 Colorectal cancer

Adenocarcinomas of the lower gastrointestinal tract (the colon and rectum – grouped as 'colorectal' cancer) are the third most common cancer in the U.K. with over 40,000 new cases registered annually (CRUK 2014a), and are the second most common cause of cancer death in the U.K., despite the fact that more than half of people with colorectal cancer survive for 10 years after diagnosis (CRUK 2014b). The number of people in the U.K. alive with colorectal cancer or having previously been treated for the disease was approximately 250,000 in 2008 and is increasing (NICE 2011). Globally, new colorectal malignancies are recorded in over 1,200,000 people annually and over 600,000 deaths from this disease are estimated to occur each year (Jemal et al. 2011).

In developed countries between 70% and 80% of new patients undergo attempts at curative surgery, usually by removing the primary tumour and local lymphatic nodes (NICE 2011). Without further treatment over the following five years approximately 40% of patients develop metastatic disease, almost invariably due to the subsequent growth of micro-metastatic disease present, but undetectable, at the time of initial diagnosis. As a consequence, and in the absence of further adjuvant therapy, the subsequent five-year survival rates are principally determined by the risk of micro-metastatic disease undetectable at presentation. The risk can be estimated by a combination of clinical and pathological factors; a histological assessment of the degree of de-differentiation of the tumour (the grade) and the pathological stage of the disease (figure 1.2) derived from an assessment of the depth of the primary tumour penetration into

the intestinal wall and the extent of local lymph node involvement (NICE 2011; Walker et al. 2014).



AJCC/TNM Staging		Tumor
Stage 0	Tis (carcinoma in-situ)	Superficially involves the mucosa. Has not grown beyond the mucosa
Stage I	T1N0 T2N0	Invades through mucosa Invades through submucosa
Stage II	T3N0 T4N0	Invades through muscle layers Invades nearby tissues or organs
Stage III	Any T, N1-N3	Lymph nodes involved
Stage IV	Any T, Any N, M1	Distant Spread

Figure 1.2: An overview of the staging of colorectal cancer.
http://alamocitycancercouncil.org/cancer/colorectal-cancer/diagnosis/staging/)

In stage II disease, approximately 25% of patients at the time of diagnosis, the primary tumour has locally invaded through the bowel wall and beyond the sub-mucosa, but there is no detectible involvement of regional lymph nodes or distant metastasis. After treatment with surgery alone over the following 5 years 25% of patients with stage II disease die from recurrent or metastatic disease (Cunningham and Starling 2007). However, this group is heterogeneous, and some patients have a significantly worse outcome, similar to more advanced stage III patients (O'Connell et al. 2004).

In stage III disease, again approximately 25% of patients at initial diagnosis, there is tumour invading through the bowel wall and, in addition, involvement of regional lymph nodes, but again without detectable distant metastasis. Without further treatment the chances of developing distant metastasis and death over the subsequent 5 years for a patient with stage III disease are in the order of 50% (Andre et al. 2009). In both high-risk stage II disease and stage III disease the addition of 6 months of oxaliplatin and fluoropyrimidine based adjuvant chemotherapy, given to treat any occult micro-

metastatic disease if present, can significantly reduce the risk of recurrence (Andre et al. 2009).

Patients with stage IV disease, with distant metastasis at the time of presentation, are typically offered palliative treatment with a combination of surgery, radiotherapy, biological agents and chemotherapy (usually fluoropyrimidine and oxaliplatin based), resulting in a median survival of 18-24 months (Seymour et al. 2007).

1.1.3 The development of oxaliplatin for use in colorectal cancer

1.1.3.1 Pre-clinical evidence of efficacy

Although the discovery of cisplatin and its introduction into clinical use in 1971 showed great promise (Hill et al. 1975) even at an early stage there was awareness of its limitations, particularly the significant toxicity profile and lack of efficacy against colorectal malignancies (Alcindor and Beauger 2011). Efforts to develop platinum compounds with improved efficacy or reduced adverse effects (a better 'therapeutic profile') resulted in the synthesis of several derivative compounds. Of particular interest were platinum compounds with 1,2-diaminocyclohexane (DACH) carrier ligands, first successfully synthesised in the early 1970s (Connors et al. 1972), and of which oxaliplatin is included. In initial studies involving oxaliplatin, and similar DACH-platinum based agents, were shown to be effective against cisplatin resistant mouse L1210 and p-388 leukaemia cell lines (Burchenal et al. 1979; Gale et al. 1974), and oxaliplatin in particular was demonstrated to be effective against a variety of cisplatin and carboplatin resistant human cancer cell lines and xenografts (Mathe et al. 1989). In view of the lack of activity of cisplatin against colorectal cancer considerable interest was generated by the demonstration that oxaliplatin was effective as a single agent against 5-FU resistant HT29 colorectal cancer cell lines, and after supra-additive effects were noted in combination with 5-FU in cisplatin resistant colon cancer cell lines (Raymond et al. 1998a).

1.1.3.2 Phase I clinical trials

In an initial phase I trial 44 patients with a variety of advanced malignancies received a combined total of 116 courses of oxaliplatin, with doses escalating from 45mg/m^2 to 200mg/m^2. Neither haematological nor renal toxicities were observed at doses up to 200mg/m^2, although at doses exceeding 135mg/m^2 the investigators noted a ‘‘peculiar toxicity” of a peripheral sensory neuropathy which developed in some individuals and was associated with cumulative dose. The development of a clinically severe peripheral neurotoxicity was noted in 6 of 44 patients treated at cumulative doses above 500mg/m^2. These patients were investigated with nerve conduction studies, revealing a sensory neuropathy characterised by axonal degeneration. The development of acute and transient peripheral paraesthesia was also noted, particularly triggered when patients touched cold surfaces. As a result of this trial a dose of 130mg/m^2 given every 3 weeks by IV infusion over 2 hours was recommended for future clinical studies (Extra et al. 1990).

1.1.3.3 Phase II clinical trials

The first clinical trials with oxaliplatin as a single agent in patients with 5-flourouracil (5-FU) resistant colorectal cancer demonstrated modest activity of 10% (Levi et al. 1993; Machover et al. 1996). An improved activity rate of up to 20% was noted using oxaliplatin as a single agent in previously untreated patients (Diaz-Rubio et al. 1998).

In combination with 5-FU response rates as high as 58% were achieved, as shown in a heterogeneous cohort of both previously treated and untreated patients with metastatic colorectal cancer conducted by Levi in 1993 (Levi et al. 1993). In the period following these initial studies the addition of folinic acid (leucovorin in the USA) to the then standard treatment of 5-FU was shown to be of benefit, and the activity of a combination of oxaliplatin, 5-FU and leucovorin was confirmed in subsequent phase II trials and given the commonly used acronym FOLFOX (de Gramont et al. 1997).

1.1.3.4 Phase III clinical trials

Initial phase 3 trials comparing FOLFOX to 5-FU and leucovorin, the previous standard of care, in untreated patients with stage IV colorectal cancer resulted in an improvement in response rate (50.7% vs. 22.3%) and progression free survival (9.0 months vs. 6.2 months) in favour of FOLFOX (de Gramont et al. 2000). The oral 5-FU pro-drug capecitabine (Xeloda, Roche) has been successfully substituted with 5-FU and leucovorin in the regime 'CAPOX' and is equally efficacious, indicating the choice between either intravenous or oral fluoropyrimidine is not clinically significant (Cassidy et al. 2004).

Given the beneficial response to treatment in the palliation of advanced disease the FOLFOX combination was subsequently tested as an adjuvant treatment following surgery in early stage patients. Previous studies with six months of adjuvant chemotherapy using 5-FU and leucovorin in patients who had undergone surgical resection for stage II and III colorectal cancer, when compared to observation only, reduced the relative risk of recurrence by 22% and death by 18% during a median follow-up of five and a half years (Cunningham and Starling 2007). The MOSAIC trial, a large international trial of 2,246 patients with stage II or III colorectal cancer compared the addition of oxaliplatin to six months adjuvant chemotherapy with 5-FU and leucovorin following primary surgical resection. The study demonstrated a relative risk reduction of death in six years following treatment of 20% in stage III disease in favour of the addition of oxaliplatin, although no benefit from the addition of oxaliplatin was demonstrated in stage II disease (Andre et al. 2009).

These trials, and confirmatory studies including NSABP C-07, a large US trial similar to MOSAIC (Kuebler et al. 2007), have resulted in the use of the FOLFOX combination chemotherapy regimen throughout the developed world as the standard adjuvant treatment for stage III disease and as a key palliative treatment in advanced disease (Attard et al. 2010; NCCN 2012; NICE 2011).

1.1.3.5 Clinical toxicity of oxaliplatin

The addition of oxaliplatin to fluoropyrimidine based chemotherapy in colorectal cancer has undoubtedly improved the response rate, time to disease progression, and increased the duration of survival in patients with advanced disease, and in those treated with curative intent when used as an adjuvant treatment. However, the addition of oxaliplatin undoubtedly results in a higher rate of side effects compared to fluoropyrimidine-based treatments (Andre et al. 2009; Ibrahim et al. 2004). Of particular importance is the trade-off between side effects and clinical efficacy in both patients with incurable disease and in patients who are treated with curative intent, the majority of whom are expected to have a normal life expectancy following treatment. Patients with advanced or terminal disease are prepared to accept higher rates of side effects whereas those treated for lower risk disease, who only gain small absolute benefit in terms of reduced risk of recurrence, and who have a normal life expectancy the adverse consequences of chronic side effects can be more problematic (Balmer et al. 2001; Matsuyama et al. 2006; Mende et al. 2013).

The 2004 FDA approval documentation for oxaliplatin, based on the early trials of patients with stage IV disease treated with palliative intent, compares the rates of toxicity from 5-FU/LV, single agent oxaliplatin, and from the addition of oxaliplatin to 5-FU/LV. The rate of any severe or life-threatening toxicity (NCI CTC toxicity grading 3 or 4(NCI 2009)) increases from 41% to 73% with the combination of agents, mainly from the increase in gastrointestinal, haematological and neuropathic side effects. The rate of transient and chronic OIPN is similar for single agent oxaliplatin and the combination treatment, and both are significantly higher than the minimal rate of neuropathy seen with 5-FU/LV, indicating that neuropathy is a consequence mainly of oxaliplatin, rather than the combination with 5-FU (figure 1.3) (Ibrahim et al. 2004).

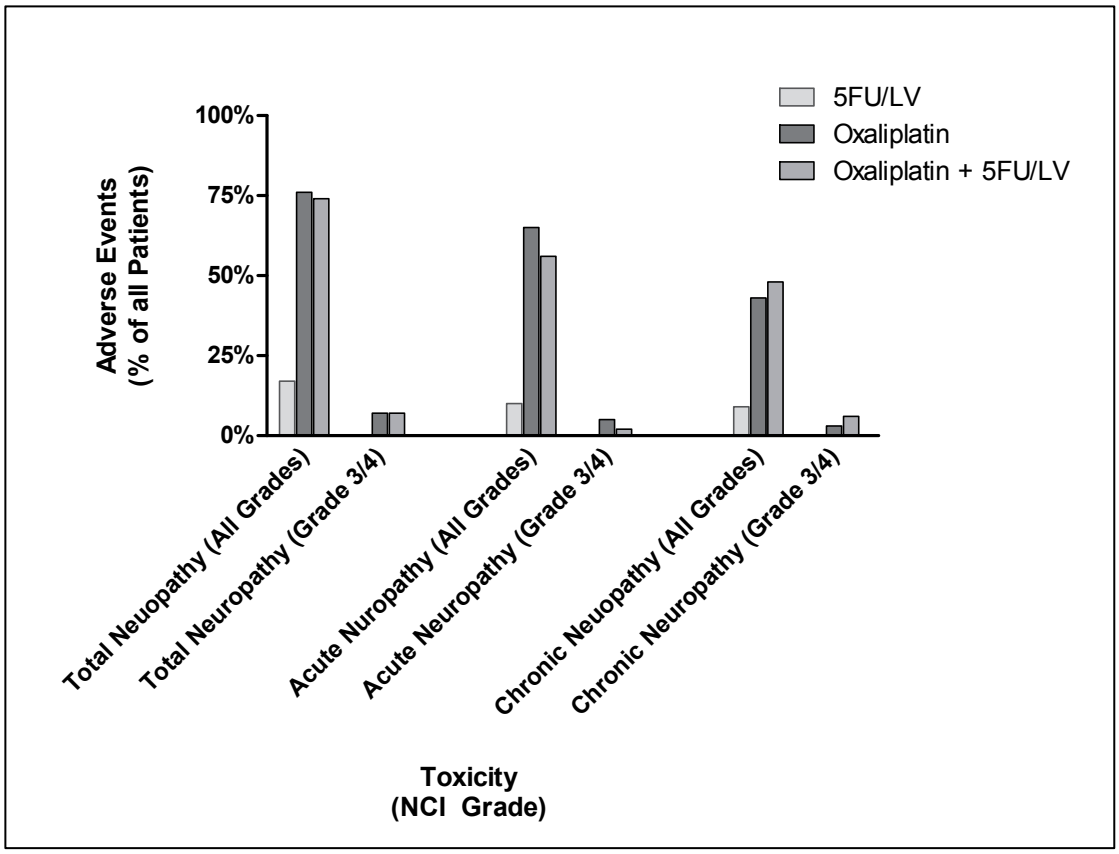


Figure 1.3: Rates of neuropathy after oxaliplatin-based chemotherapy.
The addition of oxaliplatin to 5FU\LV results in significant increase in acute and chronic neuropathy from chemotherapy. The increase is due to oxaliplatin, rather than the combination as shown by similar rates between oxaliplatin compared to oxaliplatin+5FU/LV. (Data taken from(Ibrahim et al. 2004))

Patients treated with curative intent with oxaliplatin are, by definition a different cohort of patients to those with advanced disease, and have different rates, and different expectancies of the risk of toxicities from chemotherapy (Balmer et al. 2001; Matsuyama et al. 2006). Data on the rate of toxicity in this cohort of patients from MOSAIC trial (of FOLFOX compared to 5-FU/LV) again demonstrate that diarrhoea, vomiting, neutropenia and neuropathy are the most common adverse events, with a 41.1% rate of neutropenia with the FOLFOX oxaliplatin compared to 4.1% with 5-FU/LV. In this group the rate of OIPN was also high when treated with FOLFOX with 92.1% of patients developing symptoms (of any severity grade), and 12.4% with OIPN graded as clinically severe and interfering with function, compared to 15.6% and 0.2% respectively in the non-oxaliplatin containing regimen. Even after follow-up of 36 months the

rates of residual OIPN of any grade in the FOLFOX group was still 17.1% (Figure 1.4) (Andre et al. 2009).

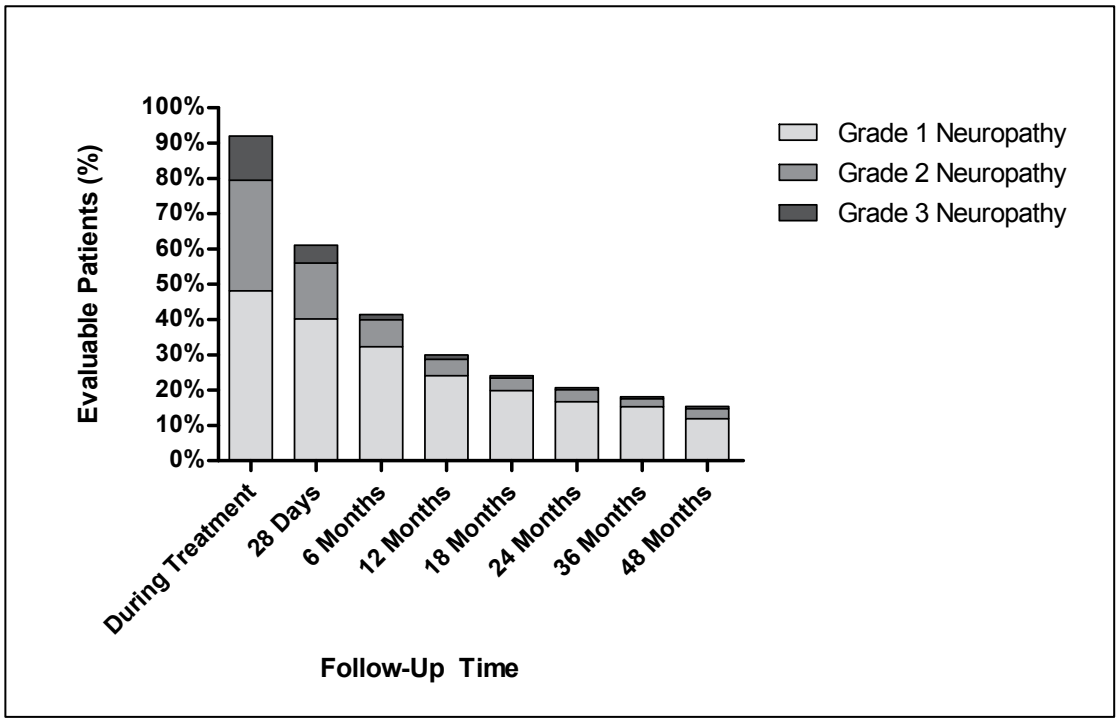


Figure 1.4: The time course of peripheral neuropathy.
Rates of peripheral neuropathy following treatment with oxaliplatin and 5FU/LV in the MOSAIC trial, using NCI CTC V1 criteria. Data taken from Andre 2009.

Several other studies confirm the high rates of long-term neurotoxicity from oxaliplatin. Using NCI CTC toxicity grading (NCI 2009) approximately 50% of patients receiving oxaliplatin have CTC grade 2 or more neuropathy with CTC grade 3 OIPN occurring in 10-20% (Allegra et al. 2009; de Gramont et al. 2000). A significant proportion of patients are left with symptoms more than two years after treatment, as demonstrated by a study with 44% of patients with CTC grade 2 or 3 symptoms at the end of treatment, and with 6% after 1 year and 4 % after 18 months (Weickhardt et al. 2011). In a European trial of oxaliplatin 26% of patients who had grade 3 OIPN during treatment still had persistent symptoms at a median follow up of 28 months (de Gramont et al. 2000). More than 10% of patients in a similar large US trial, NSABP C-07, had persistent neurological symptoms after 2 years (Land et al. 2007).

1.1.3.6 Conclusions from oxaliplatin clinical data

Whilst oxaliplatin is unquestionably of benefit to many patients with colorectal cancer, a significant proportion of patients are developing toxicity from treatment with significant long term economic, social and health consequences in a cohort of patients with a 10-year survival of over 55% (Berger et al. 2004; CRUK 2014b; Kautio et al. 2011; O'Connor 2009; Pike et al. 2012).

The development of tools to predict who will benefit from treatment and who is at greater risk of toxicity are urgently needed, and would allow rational investigation and modification of treatment schedules. These tools are needed to improve the cure rates, by increasing treatment intensity in patients unlikely to develop toxicity, and to reduce the burden of long-term toxicity in patients who are more sensitive to these agents. With this in mind, the next section will focus on the mechanism of action of oxaliplatin, with a view to understanding how, at the molecular level, methods to stratify patients for response and toxicity could be developed.

1.2 The molecular mechanism of action of oxaliplatin

Knowledge of the molecular mechanism of action of these drugs is crucial to predicting the spectrum of clinical effects of the platinum agents, and in understanding the similarities and differences between these agents. Particularly important in the context of this thesis is the effect of oxaliplatin on the neurones of the dorsal root ganglia, and the idiosyncratic response of these cells to DNA damage through the nucleotide excision repair (NER) pathway – a mechanism that may be responsible for the risk of the development of neuronal toxicity and could potentially be used as a predictive marker for OIPN.

The majority of studies investigating the mechanism of action of the platinum drugs at the molecular level are based on cisplatin, but apply also to oxaliplatin. Differences exist in molecular interactions between cisplatin and oxaliplatin and explain the differences in efficacy and toxicity of these agents. Several have been elucidated and will be highlighted in the following discussion.

1.2.1 Oxaliplatin pharmacokinetics

Pharmacokinetics studies analyse how the body affects a drug after administration, through the mechanisms of absorption, distribution, metabolism and the through the effects of excretion of the drug and of drug metabolites. The pharmacokinetics of oxaliplatin and cisplatin have been extensively studied. The central platinum ion of oxaliplatin is bound to an oxalate leaving group and a diaminocyclohexane (DACH) carrier ligand (figure 1.5). In comparison, cisplatin, a neutral inorganic compound, consists of a central platinum ion bound to two labile chloride ligand leaving groups in a square planar cis- configuration, and to two non-labile ammine carrier ligands (figure 1.5) (Arnesano et al. 2013).

Both compounds shed their leaving groups before entering the cell. As a result, it is primarily the differential cellular response to the DACH carrier ligand of oxaliplatin compared to the amine carrier ligands of cisplatin that results in the different spectrum of clinical activity and toxicity of these agents (Raymond et al. 1998b).

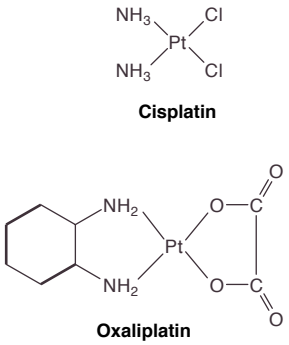


Figure 1.5: The chemical structure of cisplatin and oxaliplatin.
The structure of cisplatin (cis-platinum(II)-diamine-chloride) and oxaliplatin (oxalato(trans-L-1,2-diaminocyclohexane)platinum). The leaving groups are to the right of the central platinum ion and the carrier ligands to the left (Adapted from(O'Dwyer et al. 2000))

After intravenous administration the platinum agents undergo a process of rapid stepwise biotransformation – the process of modification of the chemical compound by the organism. Unlike most chemotherapy pharmacokinetic studies, investigation of the transformation of the platinum agents is limited by technological constraints: it is possible only to measure total platinum, rather than the specific biotransformation products of the drugs using the available techniques of atomic absorption spectroscopy (AAS) or inductively coupled mass spectrometry (Jerremalm et al. 2009). The lack of ability to differentiate between bio-transformants (the products of biotransformation) results in co-determination of the initial platinum agent and the short-lived platinum containing intermediate complexes, hence monitoring of platinum rather than specifically for cisplatin or oxaliplatin for pharmacokinetic assessment is generally conducted and accepted (Graham et al. 2000).

Oxaliplatin is administered usually as an intravenous infusion over two hours at a dose of 130mg/m^2 - the dose recommended in initial phase I trials (Extra et al. 1990). Subsequently, plasma levels rapidly decrease, with a half-life of 15 minutes. This occurs primarily a result of non-enzymatic transformation to reactive compounds through the displacement of the oxalate leaving group, forming a variety of reactive intermediates in both the blood and plasma, including monochloro-, dichloro-, and diaquo-platinum species (figure 1.6)

(Jerremalm et al. 2009). Oxaliplatin and the immediate derivatives undergo rapid distribution - by the end of the two-hour administration only 15% of detectible platinum remains in the blood; 85% has been distributed to tissues and a small fraction has been eliminated by urinary excretion.

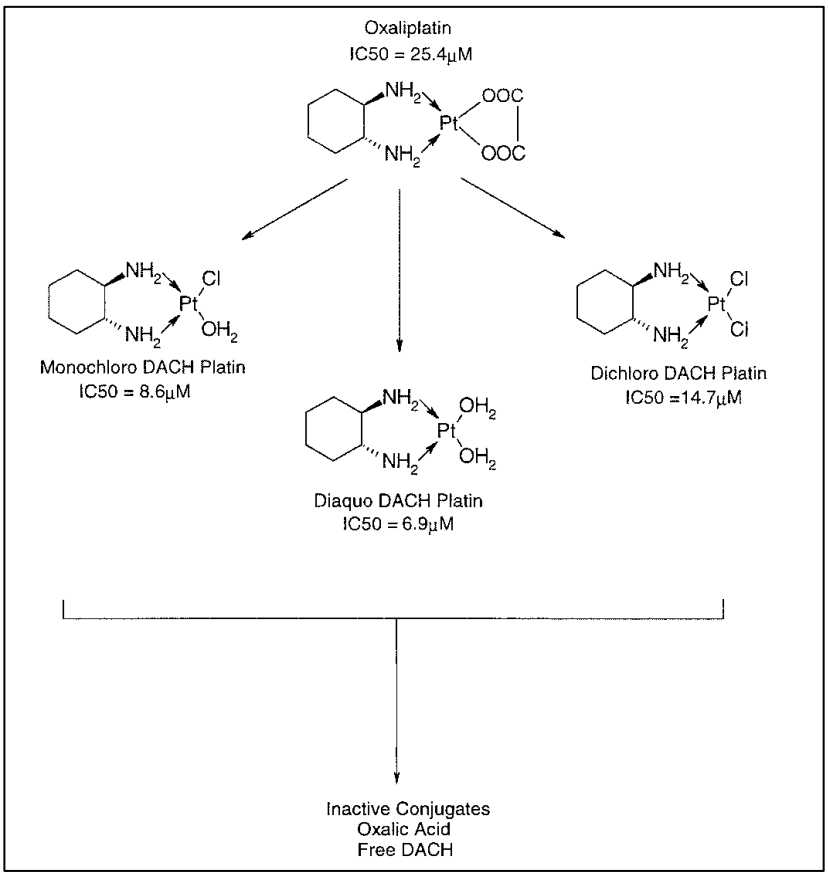


Figure 1.6: The biotransformation of oxaliplatin.
The biotransformation of oxaliplatin to monchloro, dichloro and diaquo compounds (Jerremalm et al. 2009).

A three-compartment model for the distribution of oxaliplatin in circulating blood has been proposed; total plasma platinum, ultra-filterable 'free' platinum and erythrocyte sequestered platinum (Extra et al. 1998; Graham et al. 2000). Of the 15% remaining in the systemic circulation after administration, 40% has been irreversibly sequestered in circulating erythrocytes and 30% has irreversibly bound to plasma proteins (Mani et al. 2002), including plasma gamma-globulins, albumin and haemoglobin (Allain et al. 2000) forming non-cytotoxic products (Luo et al. 1999). Only the unbound fraction, detectible after plasma ultra-filtration to remove the inert platinum bound to plasma proteins, is

the residual biologically active fraction, and is termed the ultra-filterable fraction.

The maximum plasma ultra-filtrate concentration of oxaliplatin (i.e. the maximum biologically active concentration) after a 2-hour infusion is approximately 1.21 +/- 0.1 μg/ml of plasma (a concentration of approximately 3μM). As a consequence of the rapid clearance of intact oxaliplatin and reactive dichloro- monochloro and diaquo-DACH platinum intermediates into tissues, or removal by renal filtration, the half-life of platinum in ultra-filtrate follows a tri-exponential pattern; a rapid alpha phase with a $T_{1/2}$ of 0.28 hours, a beta $T_{1/2}$ of 16.3 hours, and a long terminal gamma phase $T_{1/2}$ of 273 hours (figure 1.7 A). This pattern reflects the slow release of platinum amino-acid conjugates after the degradation of cellular macromolecules (Di Francesco et al. 2002; Graham et al. 2000). Notably, longer time-course experiments show that no significant accumulation occurs in plasma ultra-filtrate after multiple dosing at 130mg/m^2 every 3 weeks (figure 1.7C) (Graham et al. 2000).

The circulating ultra-filterable and biologically active fraction continues to be distributed to the tissues and enters the extra-cellular fluid for 24 hours after infusion, where it is available for uptake into both malignant and normal cells (Figure 1.7B) (Burz et al. 2009).

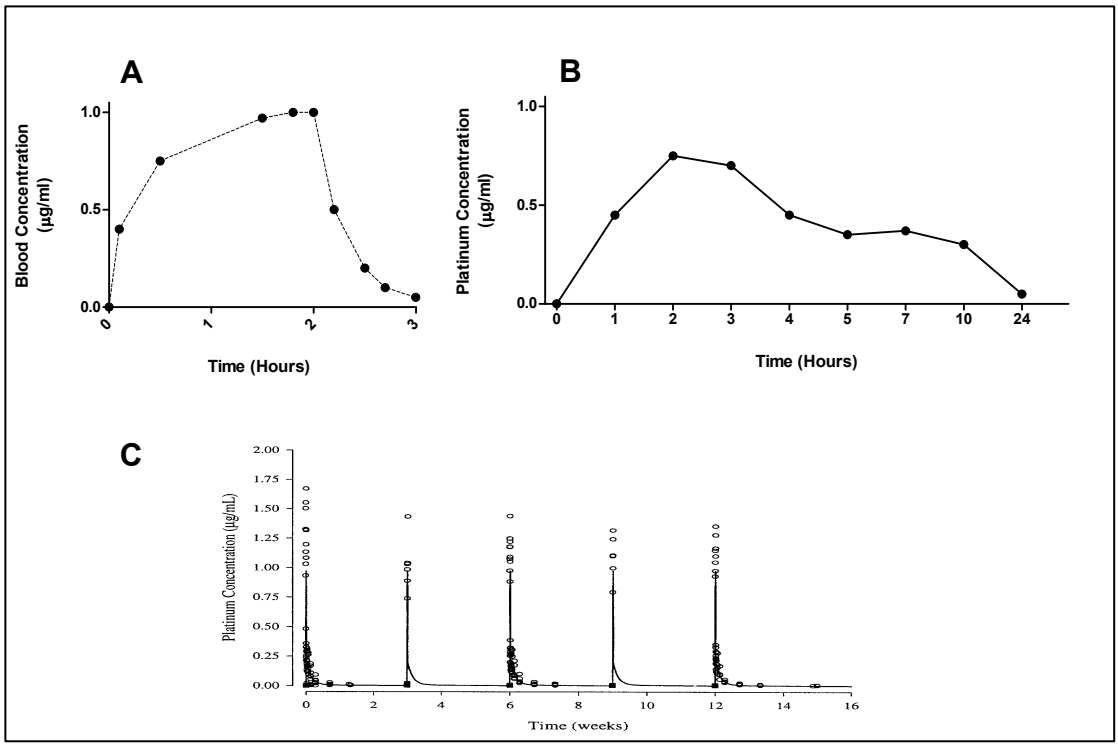


Figure 1.7: The pharmacokinetics of oxaliplatin.
Panel A shows the blood concentration of oxaliplatin during a 2-hour infusion and in the first hour afterwards (figure adapted from(Ehrsson et al. 2002)). Panel B shows the ultra-filterable compartment platinum concentration over 24 hours following infusion (adapted from (Burz et al. 2009)). Figure C shows no significant oxaliplatin accumulation when given 3 weekly for 5 cycles (Graham et al. 2000).

1.2.2 Cellular uptake of platinum - Influx vs. Efflux

Initially, experimental studies of the uptake of cisplatin suggested this class of drug is taken into the cell through passive diffusion. In these initial experiments demonstrated that uptake into the cell occurs in a concentration dependent linear manner over 60 minutes, the concentration dependent uptake of platinum could not be saturated, and was not inhibited by analogues, and that there was no optimum pH for entry; all suggesting that no significant active transport mechanism was operating (Binks and Dobrota 1990; Gately and Howell 1993; Hromas et al. 1987).

The evidence for an active component to drug influx came with the discovery that reactive aldehydes inhibit cisplatin entry into the cell. A putative mechanism involving a modification of membrane transporter proteins was proposed (Dornish and Pettersen 1990). The suggestion that the involvement of copper transporter mechanisms are involved in the influx and efflux of cisplatin

came with the observation that transfection of DNA coding for copper-transporting ATP7B efflux pumps into cell line models increased the cellular resistance to cisplatin as a direct result of enhanced cellular efflux (Komatsu et al. 2000). Several key copper transporting proteins have now been demonstrated to be involved in the active influx and efflux of cisplatin and its analogues. The principle influx transport channels are copper transporter 1 (CTR1) and organic cation transporter 3 (OCT3), with the efflux channels ATP7A and B primarily responsible for drug efflux, as will be discussed below.

1.2.2.1 Copper transporter 1 (CTR1)

CTR1, a 190 amino acid protein with three trans-membrane domains, is a high affinity copper transporter protein that mediates the influx of cellular copper (Safaei and Howell 2005). It is highly conserved during evolution, and orthologs from mammalian cells complements copper transport deficits in a yeast *S. cerevisiae* model (Howell et al. 2010). The yeast protein Ctrl1, a high affinity copper transporter, when absent results in decreased intra-cellular cisplatin accumulation and enhanced resistance to cisplatin (Ishida et al. 2002). A similar effect with other platinum agents, including oxaliplatin, has been described (Lin et al. 2002). In murine embryological fibroblasts (MEFs) the accumulation of platinum drugs into the cell is partially depended on CTR1, especially at physiologically achievable doses of platinum drugs. MEFs that are CTR1-/- accumulate one third of the amount of cisplatin and oxaliplatin that accumulates in CTR1+/+ MEFs after 1 hour of incubation at physiological doses of 2uM. At higher doses (of over 10uM) the difference in uptake of cisplatin between wildtype and CTR1-/- cells remains similar, whereas there is no differential uptake with oxaliplatin, suggesting a differential mechanism of uptake between these drugs at supra-physiological doses (Holzer et al. 2006).

The over-expression of hCTR1 in ovarian cancer cell lines results in a significant increase in the accumulation of intra-cellular cisplatin, although, of note, the amount of platinum-DNA adducts generated was not increased and there is only a marginal increase in resulting cytotoxicity (Holzer et al. 2004).

The association of platinum drug uptake and physiological copper transport may be a mutually exclusive effect. Copper and cisplatin reduce the uptake of one another, trigger degradation and delocalisation of CTR1 (Ishida et al. 2002) and show bi-directional cross-resistance (Katano et al. 2002; Safaei 2006).

1.2.2.2 Organic cation transporters (OCT1-3)

The organic cation transporters 1 to 3 (OCT1-3) are a subgroup of the SLC2 family of solute transporters. Facilitated delivery of platinum drugs by the organic cation transporters has been demonstrated, and this may correspond to a mechanism for cisplatin nephrotoxicity, as OCT2 has been shown to be important in mediating cisplatin uptake in renal proximal tubule cells. In these models an OCT2 substrate can be used, and inhibits cisplatin-induced renal cell apoptosis (Ciarimboli et al. 2005).

In human colorectal cancer cell lines OCT1 and OCT2 overexpression increases the accumulation of oxaliplatin, but not cisplatin or carboplatin, and may account for some of the spectrum of tumour specific cytotoxicity (Burger et al. 2010; Zhang et al. 2006). This is explained by the nature of the oxaliplatin leaving group, which has been found to give a degree of specificity for the OCT transporters with a differential uptake between the DACH leaving group compared to cisplatin and carboplatin (Burger et al. 2010; Zhang et al. 2006).

1.2.2.3 Efflux

After drug influx further biotransformation occurs in the cytoplasm as a consequence of the low intra-cellular chloride concentration, resulting in the formation of aquated species (figure 1.6) (Jerremalm et al. 2009). The resulting cationic molecules are unable to diffuse through the lipid bilayer, remaining trapped in the cell as a consequence of this biotransformation and the resulting gain of electrostatic charge (Jung and Lippard 2007).

Active drug efflux occurs via copper homeostasis mechanisms. The efflux pump copper transporting P type adenosine triphosphate (ATP7B) has an important role in regulating copper levels and is associated with cisplatin resistance *in vitro* (Komatsu et al. 2000) and in various cancers (Miyashita et al. 2003; Nakayama et al. 2002; Nakayama et al. 2004). Copper transporters ATP7A and ATP7B may have a roll specifically in oxaliplatin efflux, as over expression results in resistance, although the mechanism is unclear and may occur as a result of enhanced cytoplasmic sequestration rather than through increased efflux (Samimi et al. 2004). The levels of ATP7A have been shown to increase in response to oxaliplatin (Plasencia et al. 2006) and have been correlated with longer time to progression in some clinical samples (Martinez-Balibrea et al. 2009). Additionally, the expression of ATP7B in human carcinoma cells modulates sensitivity to cisplatin and copper by increasing efflux (Komatsu et al. 2000).

1.2.3 Intra-cellular biotransformation

In the chloride-rich environment outside of the cell (principally ultra-filterable plasma and extracellular fluid) the oxalate leaving-group of oxaliplatin is replaced by chloride ions to form dichloro- intermediates (figure 1.6) (Luo et al. 1998). Under these conditions, typically with a chloride concentration of 100mM, hydrolysis of the dichloro-platinum compounds is unfavourable (Kartalou and Essigmann 2001a). Following diffusion and active uptake to the cytoplasm the low intra-cellular chloride concentration (typically 4mM) results in aquation, replacing one of the chloride ions with water, to form a positively charged electrophilic aquo-chlorido species. The aquated species are much more biologically active, and have a high affinity for N7 of guanine or adenine bases (Kozelka 2009).

The interaction of platinum agents with DNA occurs only after aquation. Experimentally increasing chloride concentrations to reduce aquation proportionally slows the rate of adduct formation (Horacek and Drobnik 1971),

and the rate of reaction between salmon sperm DNA and cisplatin is almost identical to the rate of cisplatin hydrolysis (Bodenner et al. 1986).

1.2.4 Intracellular interactions

Following aquation, and the resulting accumulation of a positive charge, reactions with cellular components occurs rapidly and frequently. Over 90% of all platinum ions entering the cell are sequestered in the cytoplasm (Akaboshi et al. 1992; Akaboshi et al. 1994; Raymond et al. 2002) and there are a multitude of intra-cellular components that platinum agents can interact with, including proteins, membrane phospholipids, microfilaments, thio- containing molecules such as glutathione, DNA and RNA. The detailed nature of these reactions is uncertain because of the complexity of the intracellular environment and the lack of techniques available to investigate the subtleties of these interactions, although some information can be inferred from studies of platinum complexes with aqueous buffers and with growth medium components *in vitro* (Bancroft et al. 1990; Luo et al. 1999).

Although the weight of evidence suggests that DNA is the lethal target of platinum agents (as will be discussed in detail in the next section) it is possible that targets other than DNA may contribute to cytotoxic effect, however evidence to support this hypothesis is limited. Only 5-10% of covalently bound cisplatin is found in DNA whereas the rest is bound to protein (Akaboshi et al. 1992; Raymond et al. 2002). Cellular protein binding via sulphur atoms in cysteine and methionine residues may affect enzymatic functions, other proteins and the activity of cellular receptors, with the damage resulting in apoptosis induction (Woynarowski et al. 2000). Platinum-protein binding in A2780 cells is high (Pendyala and Creaven 1993) and the DACH moiety of oxaliplatin may direct the drug towards hydrophobic pockets, as opposed to polarised cisplatin (Woynarowski et al. 2000), and this effect may account for some of the differences seen between cisplatin and oxaliplatin adduct levels and equivalent cytotoxicity (Chaney and Sancar 1996).

1.2.5 Interaction of platinum agents with DNA

Having traversed the cytosolic compartments, platinum drugs enter the nucleus. Following entry to the nucleus the high concentration of nucleotides results in substantial interactions with nuclear DNA (Jung and Lippard 2007). There is extensive evidence that a defined family of platinum-DNA adducts form between the platinum atom and cis-N-donor ligands, resulting in the formation of cross-links with the nucleobases of DNA (Jamieson and Lippard 1999).

The evidence that DNA is the primary target of platinum drugs, among many potential cellular possibilities, has been extensively reviewed (Jamieson and Lippard 1999). Initial evidence came from the observation that cisplatin treated bacteria show phenotypes characteristic of those evoked by DNA damaging agents (Reslova 1971). More convincing evidence was demonstrated by experiments revealing that DNA repair-deficient cells are more sensitive to cisplatin (Brouwer et al. 1981; Popoff et al. 1987). Levels of platinum atoms bound to RNA and protein are too low to have significant targeted inhibitory effects, and hence are thought to be less likely to be significant (Blommaert et al. 1998). Cisplatin modification of cellular DNA can be measured using specific cisplatin-DNA adduct antibodies, and these can also be used to define the nature of crosslinks (Blommaert et al. 1998; Fichtinger-Schepman et al. 1990; Fichtinger-Schepman et al. 1987). A significant correlation typically occurs between platinum-DNA adduct levels and the sensitivity of treated cells to the drug (Boffetta et al. 1998; Welters et al. 1999a).

1.2.6 Adduct formation and base specificity

The formation of platinum adducts with DNA has been extensively reviewed (Kozelka 2009). Platinum agents initially bind a guanine nucleotide, resulting in the formation of a mono-adduct. Chemical manipulation of mono-adducts by trapping the platinum ion at the mono-adduct stage and allowing investigation of initial single intercalation is possible with the use of NH_4HC0_3 or with thiourea (Eastman 1986). These trapping experiments confirm that initial

mono-adducts form exclusively at guanine bases. Guanine bases are the preferred targets of cisplatin as the N7 position is the most reactive position in double stranded DNA - it combines a high electronic density and enhanced accessibility through its localisation in the major grove of the DNA helix.

Mono-intercalated platinum agents retain a single labile chloride or aqua ligand capable of di-adduct formation. The free second arm forms covalent bonds to the N7 positions of closely situated purine bases to form 1,2 or 1,3 intrastrand crosslinks. The negative electrostatic potential of the N7 position of guanine is increased when flanked by other guanine residues, resulting in an increased specificity for GG di-nucleotides (Gillet and Scharer 2006). A similar interaction with juxtaposed purines on the opposite strand results in intrastrand cross-links (ICLs), although at a much lower frequency (figure 1.8) (Wang and Lippard 2005).

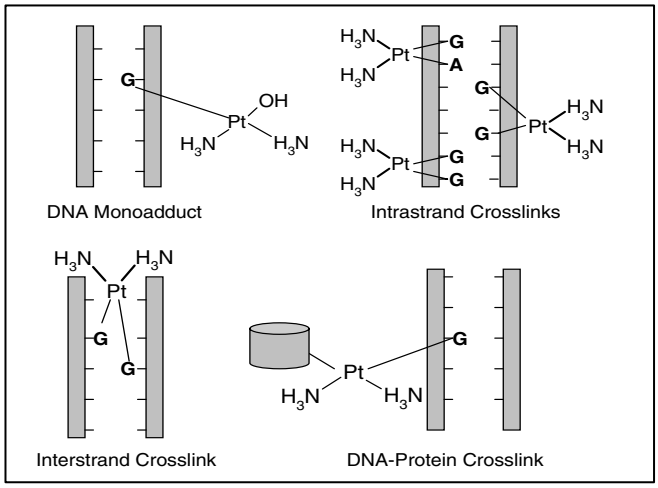


Figure 1.8: The platinum-DNA adducts.
Cisplatin-DNA monoadducts, intrastrand, interstrand and DNA-protein crosslinks. (Adapted from (Rabik and Dolan 2007))

The rate of formation of diadducts from monoadducts has been calculated using thio-urea trapping experiments (Eastman 1986; Kozelka 2009). After a short incubation of 15 minutes the majority of adducts are guanine monoadducts, although significant amounts of GG diadducts have already formed. Over the following 60 minutes incubation and after a 16-hour follow-up period all monoadducts are converted, principally to GG intrastrand diadducts (figure 1.8).

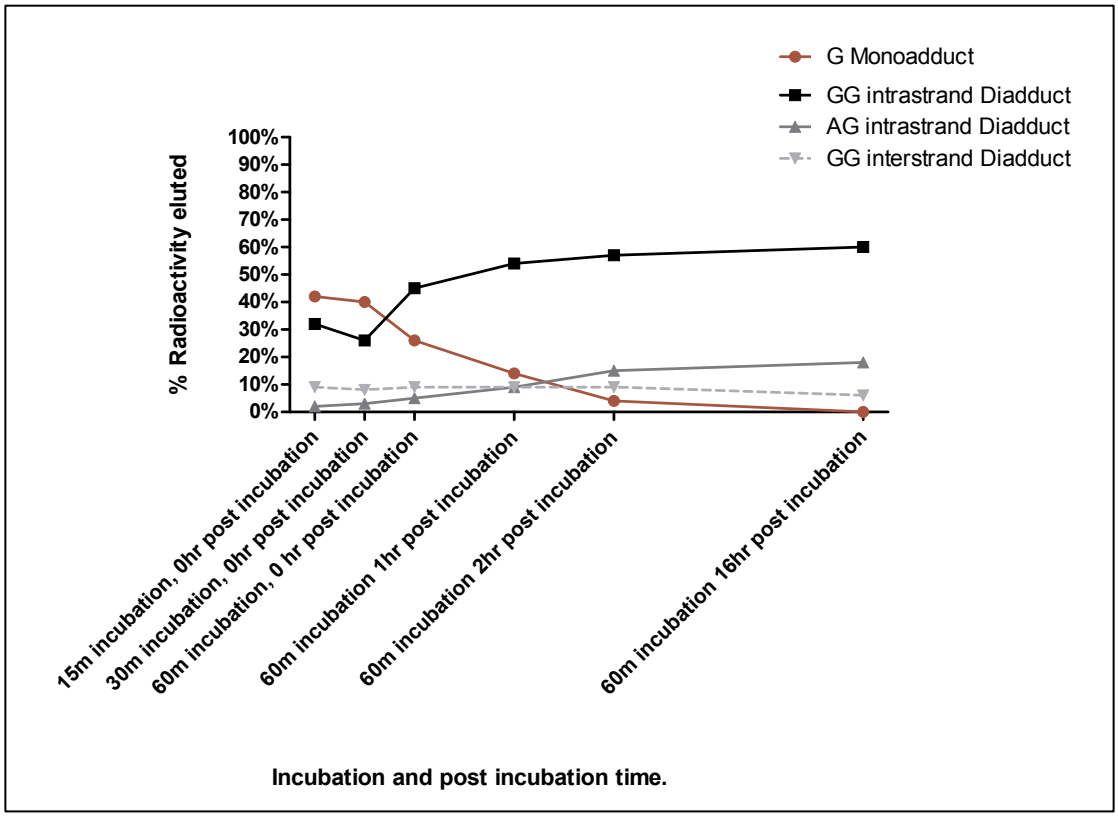


Figure 1.9: The rates of formation of platinum-DNA adducts.
The relative proportion of adducts formed between salmon sperm DNA and radiolabelled cisplatin at a variety of incubation and post-incubation times. (Adapted from (Eastman 1986; Kozelka 2009))

The relative proportions of dinucleotides involved in diadduct formation has also been elucidated by several other laboratories and is in agreement with the findings displayed in figure 1.9. To summarise, intrastrand adducts are the most prominent, with binding of two adjacent guanines, or less frequently AG dinucleotides. Analysis of purified DNA treated with cisplatin or DNA isolated from cisplatin treated patients confirms the presence of approximately 65% 1,2-d(GpG), 25% 1,2-d(ApG) and 5-10% 1,3 d(GpNpG) intrastrand cross links, and 1-5% interstrand cross links (Bruhn et al. 1990; Fichtinger-Schepman et al. 1985; Fichtinger-Schepman et al. 1987; Kartalou and Essigmann 2001b).

The formation of platinum-DNA adducts significantly affects the conformation of DNA, with studies showing unwinding and destabilisation of the duplex structure (Poklar et al. 1996). The presence of adducts and the effect on the structure of DNA has significant consequences for the ability of the duplex to function as a template for replication and transcription.

The effect of specific adducts on DNA have been elucidated (Chaney et al. 2005; Jamieson and Lippard 1999). The major 1,2 intrastrand adduct unwinds the DNA in the vicinity, bending towards the major groove and generating a widened and shallow minor groove (Coste et al. 1999; Gelasco and Lippard 1998). The cisplatin inter-strand DNA cross-link bends the helix towards the minor groove (Coste et al. 1999). There is a subtle variation in adduct-DNA distortion with different types of platinum drugs that may account for some spectrum of activity and toxicity differences between platinum agents, although no gross difference between oxaliplatin and cisplatin induced DNA distortion has been demonstrated (Wong and Giandomenico 1999).

1.2.7 The biological relevance of specific platinum-DNA adducts

The relevance of the different platinum-DNA adducts in respect of cytotoxicity is important. The key distinction is between the 1,2-d(GpG) intrastrand crosslink and the ICLs. Evidence for a difference in cytotoxicity comes from comparing the cytotoxicity of cisplatin to transplatin, *trans*-diamine-dichloroplatinum(II) (Figure 1.10). Transplatin is not cytotoxic, despite a very similar structure. It forms a different profile of adducts due to steric restriction resulting in inability to form 1,2-intrastrand crosslinks, but retains the ability to form 1,3-intrastrand crosslinks and interstrand crosslinks (Eastman 1987; Eastman et al. 1988). This is a strong indication that the primary cytotoxic lesion is the most prevalent 1,2 intrastrand crosslink.

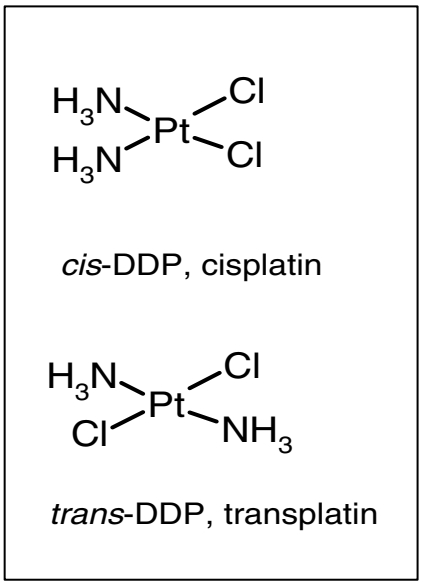


Figure 1.10: The structure of Cisplatin and transplatin.

Another indication of the relative cytotoxicity of the most common lesions is that oxaliplatin forms significantly fewer ICLs than cisplatin at equimolar doses, but is more potent. Additionally, the 1,3 intrastrand crosslink is more readily repaired than 1,2 intrastrand crosslink by NER, due to increased helical distortion and easier accessibility of the 1,3 lesion to damage recognition proteins (Wang and Lippard 2005).

1.2.8 Platinum DNA adducts in the context of chromatin structure

In the eukaryotic nucleus DNA is wound around positively charged histone proteins forming the nucleosome. These are further compacted to form the higher order chromatin structure (Figure 1.11).

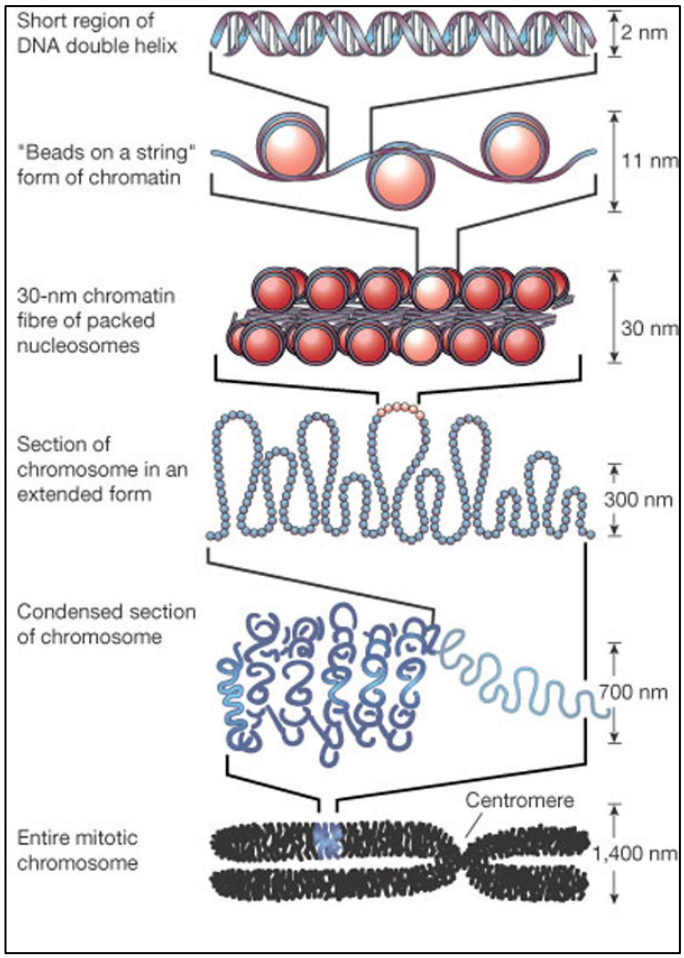


Figure 1.11: The organisation of DNA packaged in the chromatin structure.
DNA is wound around histone proteins to form the nucleosome. Stings of nucleosomes are folded to form the chromatin fibre, which is folded further into higher order structures (reproduced from(Felsenfeld and Groudine 2003))

Binding of platinum compounds to chromosomal DNA is different compared to platinum-DNA binding in naked DNA (Wang and Lippard 2005). Experiments reveal that linker DNA is a preferred target for platinum compounds (Foka and Paoletti 1986; Galea and Murray 2002) with a linker to nucleosome core ratio of 1:1.3 in reconstituted histone models treated with cisplatin and assessed using a stop-PCR based assay (Galea and Murray 2002).

Evidence also exists for increased adduct formation every 10bp over a nucleosome, a relationship that corresponds to the site of major groove exposure around the nucleosome core and coincides with the preferential nature of cisplatin to form adducts in the major groove (Galea and Murray 2002), although this is less pronounced at higher drug concentrations (Foka and Paoletti 1986; Galea and Murray 2002). Other studies have support the evidence for preferential targeting of linker regions (Foka and Paoletti 1986; Hayes and Scovell 1991).

The formation of platinum-DNA adducts on chromosomal DNA also depends on the level and nature of post-translational modifications of the associated histones (Bubley et al. 1996). Levels of platinum-DNA adducts are higher in human cancer cell lines following treatment with HDAC inhibitor agents, drugs that result in chromatin unfolding (Bubley et al. 1996). Structural changes in chromatin by transcriptional activation or protein binding have also been demonstrated to modulate cisplatin binding to DNA in human cells (Jung and Lippard 2007).

The structure of chromatin has significant impact on DNA replication, transcription and on rates of DNA repair. For example, NER in DNA containing site specific platinum adducts is less efficient in cell extracts than free DNA containing the same DNA (Reed 2005; Wang et al. 2003).

The influence of cisplatin on chromatin structure has been investigated both *in vivo* and *in vitro* (Hayes and Scovell 1991; Jung and Lippard 2007). Cisplatin binding has an effect on the higher order chromatin structure rather than that of the core nucleosome structure (Millard and Wilkes 2000). Cisplatin treatment of HeLa cells induces histone H3 phosphorylation and histone H4 hyperacetlyation (Wang and Lippard 2005), although whether this is a response to cisplatin binding on chromatin, or the effect of a downstream cellular pathway is uncertain (Jung and Lippard 2007).

1.2.9 Differences in DNA interaction between cisplatin and oxaliplatin

Oxaliplatin forms fewer DNA adducts than cisplatin, although oxaliplatin is at least equally as cytotoxic as cisplatin in cell models, indicating that, on a per-adduct basis, oxaliplatin-induced DNA damage is more cytotoxic than cisplatin damage (Saris et al. 1996). The lower adduct levels could result from oxaliplatin being less reactive than cisplatin with naked DNA, demonstrated *in vitro* and in A2780 cells with 3-fold less adducts formed in naked DNA than with cisplatin (Saris et al. 1996) detected using an antibody raised against cisplatin modified DNA.

In agreement, Woynarowski used a PCR-stop assay to demonstrate that oxaliplatin and cisplatin form lesions at the same sequence specificity and with a similar profile of adducts in naked SV40 DNA (Woynarowski et al. 1998). However the frequency of lesions in intracellular DNA in A2780 cells was 2 to 4 times lower with oxaliplatin, although oxaliplatin IC50 in A2780 cells was 0.56μM compared to 2.3μM for cisplatin, approximately a four fold difference.

The rate of formation of adducts may be different between these agents. Di-adduct formation occurs more slowly with oxaliplatin than with cisplatin, yet as overall oxaliplatin is more cytotoxic this suggests that the therapeutic effects do not depend solely on the rate of the intercalating effect of the platinum moiety (Di Francesco et al. 2002). Inter-strand cross links are also believed to contribute to toxicity, more so in cisplatin than in oxaliplatin treated cells (Zwelling et al. 1978).

The downstream response to oxaliplatin adducts, although similar in terms of NER - the predominant adduct repair pathway (Raymond et al. 2002), differs compared to cisplatin and carboplatin adducts, particularly as oxaliplatin fails to generate a mismatch repair (MMR) response. DNA synthesis is also inhibited to a greater extent with oxaliplatin adducts compared to cisplatin-DNA adducts because of the steric affect of the DACH moiety (Raymond et al. 1998b). The downstream processing of platinum adducts, and specific differences between cisplatin and oxaliplatin, will be discussed in the next section.

1.3 The cellular processing of platinum-DNA adducts

Once platinum-DNA adducts have formed, and have been recognised, several pathways in the cell are activated. Adducts are recognised by the binding to adduct-distorted DNA by wide variety of proteins, leading to the activation of cellular signalling cascades, and resulting in cell cycle arrest and the activation of mechanisms to attempt to repair the damaged DNA. During cell cycle arrest and the cessation of DNA replication, the inhibition of mRNA synthesis by platinum-DNA adducts additionally results in significant stresses to the cell. Unless platinum DNA adducts are repaired (by the mechanisms discussed later in section 1.4) the continued activation of DNA damage signal cascades, combined with the cellular response to failed replication and impaired transcription, results in cell death via apoptosis or necrosis (Figure 1.12), as will be discussed in this section (Jung and Lippard 2007; Wang and Lippard 2005).

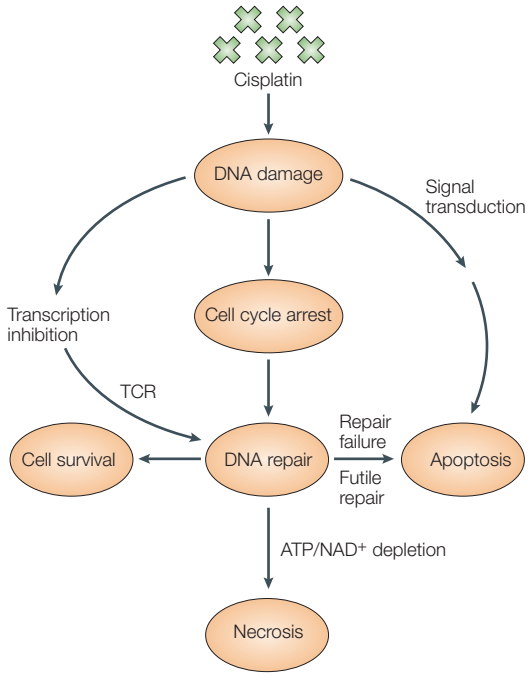


Figure 1.12: Overview of the cellular response to the presence of platinum-DNA adducts. Following the formation of platinum-DNA adducts, distortion in DNA structure results in protein-DNA binding, triggering multiple signalling cascades, resulting in cell cycle arrest and attempted DNA repair, and leading to either cell survival or cell death via apoptosis or necrosis (Adapted from (Wang and Lippard 2005))

1.3.1 Recognition of DNA platinum adducts

Many proteins that bind DNA in the presence of platinum-DNA adducts have been described and, in general, are proteins that bind primarily to structurally distorted DNA. They typically play important roles in chromatin structure modulation, transcription, DNA repair, DNA recombination, or DNA damage recognition (Chaney et al. 2005; Raymond et al. 2002). The number of interacting proteins is large, and includes HMG domain proteins (HMG1, HMG2, Ixr1, tsHMG, SRY, LEF-1 and hUBF) or non-HMG proteins (including TBP, histone H1, p53, XPE, Ku, XPA and RPA) (Chaney et al. 2005).

Platinum-DNA adduct-induced DNA distortion is central to protein binding in this context (Kartalou and Essigmann 2001b). The formation of platinum-DNA adducts results in distortion of the DNA duplex, and the degree of distortion is differs slightly between cisplatin and oxaliplatin. DNA is distorted towards the major groove by the most prevalent 1,2 diadducts, resulting in a wide and shallow minor groove surface to which several classes of proteins can bind, especially the high mobility group box proteins (HMG), DNA repair proteins, transcription factors and histone H1 (Jamieson and Lippard 1999; Kartalou and Essigmann 2001b; Vaisman et al. 1999).

The functional effects of the interaction of DNA-binding proteins with adduct distorted DNA are diverse, and include NER, replicative bypass, utilisation of transcription factors and/or action as damage sensors initiating signal cascades, and resulting in cell cycle arrest and apoptosis (Vaisman et al. 1999). Although many similarities occur, differences in the platinum-DNA adducts formed by oxaliplatin and cisplatin result in a different spectrum of protein binding, notably resulting in differences in the interaction of platinum-DNA adducts with the MMR pathway, as discussed in later section 1.4.2.2 (Vaisman et al. 1999).

Especially important is the interaction of platinated-DNA with HMG box protein 1 (HMGB1) which has been shown to interact with mismatch repair (MMR) protein MutSα (Yuan et al. 2004) and with p53 *in vitro* and enhances p53 DNA binding activity (Wang and Lippard 2005) providing a possible link with p53 mediated DNA repair (Imamura et al. 2001).

HMGB1 has been shown to interact with the nucleosome and facilitate the binding of ATP-utilising chromatin assembly and remodelling factor (ACF) to nucleosomal DNA, accelerating chromatin structural changes and demonstrated to bind irreversibly to chromatin in apoptotic cells putatively representing acetylation changes and HMGB1 may also act as a cytokine to signal tissue damage (Wang and Lippard 2005).

Some of the binding proteins, especially HMG-domain family, bind more tightly to cisplatin adducts compared to oxaliplatin adducts. This ability to discriminate between adducts has been shown with HMG1, TBP, and hUBF, although has not been looked at in several other damage recognition proteins (Chaney et al. 2005). The biological consequences are unclear, although HMG1 inhibits translesion synthesis (TLS) past cisplatin adducts *in vitro* more than oxaliplatin adducts (Vaisman et al. 1999).

1.3.2 Transduction of DNA damage signals

In addition to the interaction of HMGB1 proteins with the pathways discussed above, another important interaction with DNA distorted by the presence of platinum-DNA adducts is by c-ABL, a member of the SRC family of non-receptor tyrosine kinases. This occurs through the interaction c-ABL with adduct distorted DNA via a HMG-like domain, and results in initiation of several DNA damage signalling cascades, including activation of p53. Activation of p53 results in the subsequent initiation of the cell cycle arrest response or leads to a pro-apoptotic response, depending on the level of DNA damage detected and underlying levels of cellular stress (Wang and Lippard 2005). Several NER proteins, including XPC, TFIIH and RPA (see section 1.4 for details) interact with activated p53, enhancing the DNA repair response to platinum-DNA adducts (Wang and Lippard 2005).

Activated c-ABL additionally activates other protein kinase cascades, including the p38-MAPK pathway, activated via the activation of mitogen activated protein kinase kinases MKK3 and MKK6, resulting in changes to gene expression and chromatin state. Extracellular signal regulated kinase (ERK)

activation follows phosphorylation of MEK1 and MEK2, also mitogen activated protein kinases. The c-Jun N terminal kinase (JNK) signalling cascade is also activated, and has been proposed as a further mechanism of platinum induced cell death (Pandey et al. 1996)

1.3.3 Cell cycle arrest

Following DNA damage response activation, phosphorylation of ATM and ATR result in activation of CHK1 and CHK2 through phosphorylation, subsequently phosphorylating CDC25C, leading to CDC2 phosphorylation and ultimately to cell cycle arrest in G2 (Wang and Lippard 2005). As mitosis involving unrepaired DNA can pass DNA damage and mutation to daughter cells, or result in catastrophic mitosis through replication fork collapse, G2 arrest is a crucial component of the cellular response to cisplatin (Siddik 2003; Sorenson and Eastman 1988).

1.3.4 Inhibition of mRNA synthesis

Following platinum-DNA damage cell cycle arrest and DNA replication inhibition occurs, preventing mitosis whilst significant DNA damage is present. During cell cycle arrest, transcription is on-going. Inhibition of transcription by platinum-DNA adducts is an additional cellular stress that can have a bearing on whether the cell survives, or undergoes cell death through apoptosis or necrosis (Wang and Lippard 2005), although the magnitude of this contribution is debated (Roos and Kaina 2013).

Three main mechanisms of DNA-damage induced transcription inhibition that have been proposed (Todd and Lippard 2009). Firstly, binding of transcription factors occurs. Platinum-DNA adducts can serve as binding sites of transcription factors, especially those which have a strong affinity for platinum, preventing binding of transcription factors at promoter sites. Secondly, inhibition of RNA polymerases occurs. This has been demonstrated for cisplatin

and probably occurs with oxaliplatin, and occurs because distorted bases of platinum-DNA adducts are unable to enter the binding site of Pol II. Additionally, nucleosome-DNA adducts have the potential to block access by RNA polymerase to the DNA template (Todd and Lippard 2009).

1.3.5 Cell death: via apoptosis and necrosis

Exposure to cisplatin results in cell death through necrosis and apoptosis. The pathway activated is partly concentration dependent (Gonzalez et al. 2001). Higher concentrations of cisplatin results in necrosis in mouse renal cell models at 800uM, where as concentrations of 10uM over several days result in apoptosis (Lieberthal et al. 1996). One proposed mechanism for concentration dependent necrosis is through the hyper-activation of PARP at higher does with excessive DNA damage, resulting in NAD+/ATP depletion and ultimately in necrosis (Wang and Lippard 2005).

The most common cell death response to cytotoxic agents is through apoptosis, resulting from activation of the apoptosis-caspase-cascade. Cisplatin treated cells have been shown to increase cytochrome C in the cytoplasm, an indicator of activation of the extrinsic apoptotic cascade (Kojima et al. 1998). The importance of apoptotic pathways to platinum agents is demonstrated by the inability of caspase 3 deficient MCF7 breast cancer cell lines to undergo apoptosis in response to platinum treatment, which can be revered by stable transfection of CAP3 cDNA into MCF-7 cells (Blanc et al. 2000). Cisplatin can also up regulate Fas and Fas Ligand (FasL) proteins (Fulda et al. 1998).

1.4 Platinum-DNA adduct repair

As outlined above, the continuing presence of unrepaired platinum-DNA adducts triggers cell death through activation of the DNA damage response, leading to apoptosis or necrosis, especially if DNA replication and transcription is inhibited and if high levels of cellular stress persist. During cell cycle arrest, DNA repair is initiated. The repair of platinum-DNA damage has a significant impact on whether a cell enters apoptosis or survives (Wang and Lippard 2005), with corresponding pre-clinical and clinical correlations (see section 1.4.1.6).

Studies in yeast and mammalian cell lines have shown that platinum-DNA adducts are processed by a combination of nucleotide excision repair (NER), recombination repair (RR) and trans-lesion synthesis (TLS). The primary repair mechanism, responsible for repair of the predominant 1,2 intrastrand conformation of platinum-DNA adducts, is the NER pathway. The combination of the NER, RR and TLS pathways is involved in the repair of interstrand DNA damage (Chaney et al. 2005).

1.4.1 Nucleotide excision repair (NER)

The nucleotide excision repair (NER) pathway removes many different DNA lesions and specifically the predominant 1,2 and 1,3 intrastrand adducts formed by the interaction of platinum with DNA. For a recent review of NER see (Marteijn et al. 2014).

NER is highly conserved, comprising in eukaryotes a pathway of approximately 30 interacting proteins, and is a multi-step process that can be conceptualised in five distinct sub-processes: DNA damage detection, open DNA complex formation (separation of the double strand DNA via unwinding of the DNA helix), incision of the single stranded DNA 5' and 3' of the lesion, extrusion of the damage, and finally strand re-synthesis and ligation (figure 1.12) (Marteijn et al. 2014).

Two sub-pathways of NER operate with distinct mechanisms of DNA damage detection; global genome repair (GG-NER) and transcription-coupled

repair (TC-NER). These pathways differ in the initial DNA damage recognition phase, allowing NER to respond to a wide spectrum of damaging lesions. GG-NER removes lesions from all non-transcribed DNA and the non-transcribed stands of active genes, resulting in a genome wide DNA repair capability, whilst TC-NER occurs in actively transcribed genes, regions at which unresolved DNA damage is particularly problematic for the cell (Hanawalt and Spivak 2008; Tornaletti and Hanawalt 1999).

1.4.1.1 Stage I - DNA damage recognition

DNA lesions block the action of DNA and RNA polymerases, interfering with the critical cellular functions of replication and transcription. Appropriate recognition of these lesions is the first step to successful NER. Importantly, the initial damage recognition step in NER is the stage at which GG-NER and TC-NER differ, with different protein complexes involved in both sub-pathways.

1.4.1.1.1 Stage I – Global genome NER damage recognition

Global genome NER proteins continuously examine the genome for helical distortion resulting from DNA damage. Damage recognition occurs through the interaction of damaged and distorted DNA with a complex of XPC, RAD23B and centrin 2 (CETN2) proteins, with the assistance of the UV-DDB protein complex in specific circumstances (Friedberg 2005; Marteijn et al. 2014). The three-subunit complex of XPC, RAD23B and CETN2 recognises DNA structural distortion (Araki et al. 2001), with XPC binding to small single stranded DNA gaps that result from disruptions in base pairing (Maillard et al. 2007; Marteijn et al. 2014). NER is functional in the presence of XPC alone, however the addition of RAD23B and CETN2 enhances the reaction, with the 56 amino acids of RAD23B being particularly important (Gillet and Scharer 2006). The binding to damaged DNA of the XPC-RAD23B-CETN2 complex results in further distortion of the DNA helix allowing the recruitment and participation other necessary repair factors (Janicijevic et al. 2003).

After DNA damage the affinity of XPC for DNA is enhanced by poly-ubiquitinylation, without triggering an increase in protein degradation (Sugasawa et al. 2005).

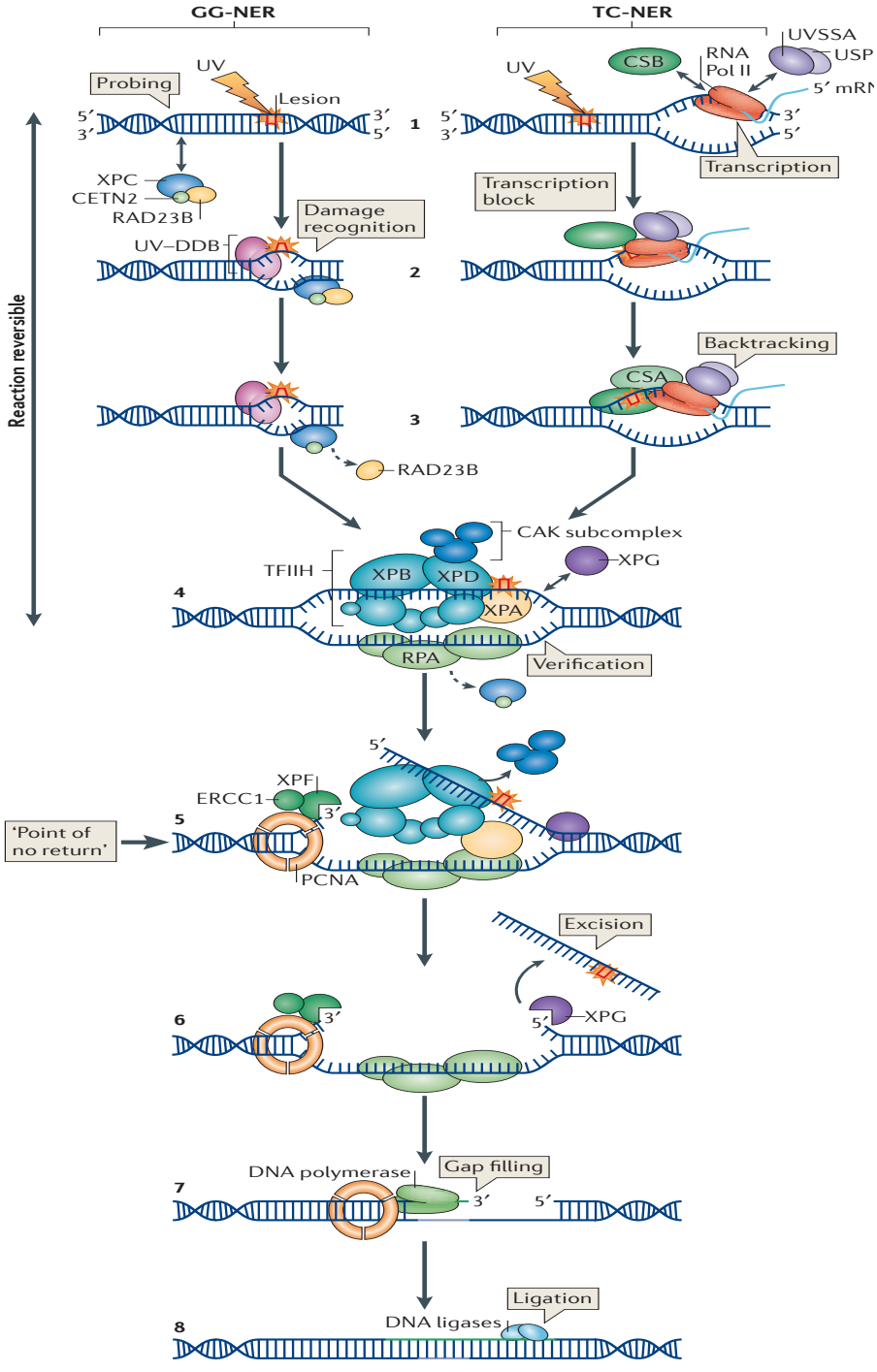


Figure 1.13: An overview of the NER pathway.
The NER pathway showing TC-NER and GG-NER damage recognition, followed by recruitment of the pre-incision complex, 5' and3' incision, oligonucleotide extrusion, and re-synthesis and ligation (Adapted from (Marteijn et al. 2014))

The concept of a bipartite system of damage recognition and confirmation of damage has been developed; a lesion has to both distort DNA and to have a chemical modification for NER to occur (Friedberg 2005; Hess et al. 1997). The distortion induced by XPC-RAD23B-CETN2 allows the entry of XPA, RPA and the TFIIH complex, necessary for the subsequent phases of NER. Entry of these proteins allows scanning in an 5' to 3' ATP-dependent fashion to identify chemically modified bases, arresting translocation at sites of damage and allowing NER to proceed. If chemically modified bases are not detected a stable open complex cannot form and NER does not proceed (Friedberg 2005; Sugasawa et al. 2001).

Damage recognition by the XPC-RAD23B-CETN2 complex is poor in UV-radiation induced damage (primarily UV induced cyclobutane-pyrimidine dimers (CPDs)), as these lesions result in only minor helical distortion. Instead, the UV DNA Damage Binding Protein complex (UV-DDB), a complex of DDB1 and DDB2, binds to UV induced lesions, increasing the degree of helical distortion, and in turn stimulating XPC-RAD23B-CENT2 binding (Chu and Chang 1988; Wakasugi et al. 2002). This mechanism increases the spectrum of lesions that can be identified by GG-NER.

1.4.1.1.2 Stage I – Transcription coupled NER - damage recognition

DNA damage in actively transcribed genes results in significant impairment of cellular function, particularly through impaired transcription; no RNA polymerases are known that can bypass lesions that distort the DNA helix. In addition, although the crucial process of DNA replication may continue through the function of alternative DNA polymerases (although this may induce mutations (Hendriks et al. 2010; Marietta and Brooks 2007)) or through homologous recombination-dependent template switching, impaired DNA replication activates cell cycle checkpoints, can result in the generation of lethal double strand breaks, and failure to resolve transcription arrest can trigger apoptosis (Ljungman and Zhang 1996). As cells are particularly sensitive to DNA damage in these regions, lesions in actively transcribed genes trigger NER through an additional pathway – transcription coupled NER (TC-NER).

TC-NER is initiated by the detection of stalled RNA polymerase II (RNAPII). Several lines of experimental evidence support this assertion. Lesions that induce the inhibition of transcription elongation by RNAPII such as CPDs or cisplatin adducts are repaired by TC-NER (Laine and Egly 2006; Sarker et al. 2005) whilst lesions that do not inhibit RNAPII elongation, such as an *N*-2-aminoflourine adduct, are not preferentially repaired within the transcribed strand (Donahue et al. 1996; Tang et al. 1989). The inhibition of RNAPII through chemical means or via temperature sensitive mutants prevents the preferential repair of lesions on the transcribed strand of a gene (Christians and Hanawalt 1992; Sweder and Hanawalt 1992). Additionally, the incision phase of TC-NER can be reconstituted *in vitro* at a lesion containing arrested RNAPII with recombinant NER proteins excluding XPC, indicating a the role of RNAPII stalling as a different mechanism for damage recognition to GG-NER (Laine and Egly 2006; Mu et al. 1996; Rademakers et al. 2003; Volker et al. 2001).

Defects in TC-NER result in Cockayne syndrome (CS); a condition characterised by UV-sensitivity, accelerated aging and neurological abnormalities (Nouspikel 2009). The majority of these patients have been shown to have a defect in one of two proteins, CSA and CSB (Licht et al. 2003).

CSA is a component of a ubiquitin ligase complex, along with the proteins cullin4A, Roc1 and the COP9 signalosome (CSN), which associates with RNAPII in a UV dependent fashion (Groisman et al. 2003). CSB is a DNA dependent ATPase of the SNF2 protein family (Eisen et al. 1995) and has a putative role in the modification of the interface between RNAPII and DNA. CSB has been identified as a target for the CSA-containing ubiquitin-ligase complex (Groisman et al. 2006). Ubiquitination of CSB leads to UV dependent degradation by the proteasome at late repair times and is important for the recovery of RNA synthesis after TC-NER. The ubiquitin-binding domain of CSB is essential also for damage incision during TC-NER (Anindya et al. 2010).

The presence of stalled RNAPII recruits CSA and CSB, resulting in the recruitment and assembly of several proteins specific to TC-NER, including UVSSA, USP7, XPAB2 and HMGN1, and recruitment of the core NER factors (excepting GG-NER specific XPC and UV-DDB GG-NER proteins) and progress to the second stage of NER (Marteijn et al. 2014; Vermeulen and Fousteri 2013).

1.4.1.2 Stage II - Open complex formation

Following the recognition of DNA damage recognition through GG-NER or TC-NER, a multi-protein 'pre-incision complex' (the PIC) is recruited and assembled at the site of damage, resulting in the separation of double stranded DNA around lesion (Mu et al. 1996; Wakasugi and Sancar 1998). The PIC structure allows dual incision 5' and 3' of the lesion, with extrusion of a short oligonucleotide containing the damage. Excepting GG-NER damage recognition specific XPC-RAD23B and TC-NER specific RNAPII, CSA and CSB, the proteins required for oligonucleotide extrusion to occur *in vitro* are identical, indicating that after damage recognition the same pathway operates for the remainder of NER (Laine and Egly 2006; Riedl et al. 2003).

The key components of the PIC are the human basal transcription initiation factor IIH (TFIIH), a ten-subunit protein complex, and the proteins XPA and RPA. Evidence suggests that TFIIH is the initial protein complex to join damage sites after detection has occurred (Riedl et al. 2003; Tapias et al. 2004; Volker et al. 2001). The TFIIH complex consists of XPB, XPD, GTF2H (aka p62), GTF2H4 (p52), GTF2H2 (p44), GTF2h3 (p34), GTF2H5 (TTDA).

XPD and XPB are the key functional subunits of TFIIH, and both have DNA helicase activity (Friedberg 2005). DNA unwinding is catalysed by these helicases, each with opposite polarity. The result of TFIIH activity, and XPD and XPB function, is the formation of an open DNA structure, allowing the recruitment of XPA, RPA and XPG (Tapias et al. 2004), the recruitment of which leads to stabilisation of the PIC. XPA has a high affinity for kinked DNA, rather than specifically for damaged DNA, and can sense the loss of normal hydrogen bonding as a consequence of the exposure of hydrophobic bases that are normally inaccessible (Friedberg 2005) and the DNA binding affinity of XPA is enhanced by the presence of RPA. The interaction of XPA, RPA and TFIIH ensures the correct siting of the PIC, and the correct placement on the damaged strand of the subsequently recruited XPF-ERCC1 and XPG endonucleases near each single strand-double strand junction (Fagbemi et al. 2011; Friedberg 2005).

RPA is a trimeric protein with a role in repair and recombination, and can bind to a single stranded region of unwound DNA (de Laat et al. 1998) with a

distinct polarity - with the RPA-1 subunit directed to the 5' end. RPA appears to bind the undamaged strand and can enhance the activity of the structure-specific endonuclease ERCC1-XPF on the opposite strand (de Laat et al. 1998). XPG interacts with TFIIH in manner dependent on RPA and XPA and this results in the formation of the pre-incision complex at the correct site, strand and orientation. The two endonucleases XPF and XPG can then perform dual incisions 5' and 3' of the lesion during the next stage of NER.

1.4.1.3 Stage III – Dual incision and oligonucleotide excision

XPF-ERCC1 forms a mutually supportive heterodimeric complex; if one of the two proteins is absent the other is unstable (Sijbers et al. 1996; van Vuuren et al. 1993). XPF is a nuclease, with the active nuclease domain at amino acids 670 to 740 (Friedberg 2005). As well as interaction with ERCC1, the function of the endonuclease also requires interaction with XPA (Friedberg 2005). The XPF-ERRCC1 heterodimer incises at the 5' side of the lesion.

XPG is also a structure specific endonuclease and has a preference for the single to double strand junction at the 3' end of the unwound DNA bubble (Mu et al. 1996). In the absence of XPG the TFIIH complex is unstable and results in dissociation, indicating a structural role in addition to the endonuclease function (Friedberg 2005). The presence, but not the function, of XPG is also required for the function of XPF-ERCC1, which indicates it may have role in the correct positioning of the opposite endonuclease (Mu et al. 1996; Wakasugi and Sancar 1998).

The exact order of incisions with XPF and XPG is uncertain and an area of current research, although evidence suggests these are ordered and not independent or simultaneous events, allowing better coordination of the repair and synthesis stages (Fagbemi et al. 2011) and preventing site single stranded gap occurring in the non-damaged strand which could potential trigger the activation of a further DNA damage response (Marteijn et al. 2014).

1.4.1.4 Stage IV – DNA synthesis and ligation

After oligonucleotide extrusion, gap filling occurs by DNA synthesis, followed by ligation of the synthesised DNA and the restoration of chromatin structure. After excision a 22 - 30 nucleotide gap needs to be filled, as the build-up of single stranded gaps may trigger a DNA damage signalling response (Marteijn et al. 2014). The 5' incision is enough to initiate DNA synthesis, even before XPG mediated 3' incision has completed. The DNA polymerases responsible are the replicate DNA polymerases Polδ or Polε that are associated with the activity of the protein PCNA. PCNA loads onto the 3' dsDNA-ssDNA junction and works to facilitate gap filling by recruiting polδ or polε to the site. Polδ also recruits polK - aiding efficient synthesis. Polδ and polK are typically active in non-replicating cells and Polε in replicating cells (Marteijn et al. 2014).

The ligase responsible for sealing the nick is ligase III with the XRCC1 partner in non-cycling cells, or both DNA ligase II/XRCC1 and DNA ligase I in dividing cells (Friedberg 2005; Nouspikel 2009; Wang and Lippard 2005).

1.4.1.5 Regulation of NER by post-translational modification and chromatin modification

The pathway discussed above describes how platinum-DNA adducts are removed throughout the genome, or specifically in the transcribed strand of active genes. The complex multistage process requires coordination and occurs in the context of the chromatin structure. There are several regulatory modifications necessary to optimise access to damaged DNA for the repair machinery. This thesis concerns the removal of platinum adducts, rather than the complex regulatory pathways and chromatin environment, and for this reason these pathways will not be discussed further. A recent review by Marteijn (Marteijn et al. 2014) includes a thorough review of this topic.

1.4.1.6 NER in the context of cisplatin and oxaliplatin response and toxicity

There are several lines of evidence to support the assertion that NER is the principle pathway for the repair of platinum-DNA adducts, with correspondingly important clinical correlations.

Evidence includes the observation that bacterial and mammalian cells deficient in NER are more sensitive to platinum compounds (Jamieson and Lippard 1999; Wozniak and Blasiak 2002), and in the human context, XP protein deficient cell lines are 5 to 10 times more sensitive to platinum drugs (Furuta et al. 2002). Extracts from XP cell lines fail to repair cisplatin modified DNA when compared to control cell extracts (Dijt et al. 1988; Hansson et al. 1991). NER is the primary process for the removal of platinum damaged DNA, specifically intrastrand 1,2 and 1,3 crosslinks, and cisplatin is removed by NER *in vivo* and *in vitro* (Furuta et al. 2002; Wang et al. 2003; Wang et al. 1993). XPC and ERCC1 mRNA is over expressed in ovarian cancer tissues resistant to cisplatin and carboplatin (Dabholkar et al. 1994), and enhanced sensitivity of human cancer cells to cisplatin has been demonstrated when ERCC1 was suppressed by siRNA (Selvakumaran et al. 2003).

Several clinical studies demonstrate a relationship between NER factors, through assessment of mRNA or protein expression, and the response to platinum agents in various cancers. The high rate of response of testicular cancer to cisplatin has been correlated with low tissue levels of several NER proteins, particularly XPA, ERCC1 and XPF (Koberle et al. 1999; Welsh et al. 2004). NER factors have been assessed especially when compared to cisplatin sensitivity and resistance, including XPC, CSB, XPA, XPB, XPD, XPF, XPG and ERCC1 levels, and many clinical studies are underway, especially to assess ERCC1 as a predictor of platinum response in lung cancer (Reviewed in (Bowden 2014)).

This discussion of NER, and the above clinical and preclinical observations indicate the importance of platinum DNA-adduct formation and the importance of NER repair in the response of patients to these agents. To-date, however, there are no proven methods to stratify patients for platinum based chemotherapy despite the considerable efforts and interest in the use of NER

factors as predictive biomarkers, primarily because of the contradictory results from measurements of a single NER factor. It has been hypothesised that this reflects NER action as a complex system of several interacting proteins, and measurements of a single factor will always have poor sensitivity and specificity with many potentially confounding variables (Bowden 2014). A functional measurement of total NER capacity would obviate these issues, and this approach will be discussed in the final section of this introduction.

1.4.2 Platinum-DNA adducts and interaction with alternative DNA repair pathways.

For completeness when discussing the molecular mechanism and effects of the platinum agents the effect of the interstrand crosslink (ICL) and ICL repair on platinum toxicity needs to be addressed, as does the effect of platinum adducts on the other key DNA repair pathways, notably mismatch repair.

1.4.2.1 Platinum-induced ICL damage and repair

After platinum adducts initially form as a monoadduct, the free second arm forms covalent bonds to the N7 positions of closely situated purine bases, with 5-10% forming with juxtaposed purines on the opposite DNA strand resulting in cross-links between DNA strands (an interstrand crosslink – (ICL)), preventing essential cellular processes of transcription and replication by preventing the separation of complementary DNA strands (Spanswick et al. 2012).

Unlike NER, with its specific and highly evolutionary conserved repair process, ICL repair is achieved by the interaction of several repair pathways, hypothesised to be a result of ICLs not representing a major evolutionary threat to cells (Vasquez and Legerski 2010). Two main sub-pathways result in ICL repair; the process is either recombination independent (repair occurs in G1/G0 and relies primarily on NER and translesion-synthesis (TLS)), or is

recombination dependent, referred to as a 'S phase pathway', and utilises the homologous recombination (HR) repair machinery (Huang and Li 2013). The precise mechanistic detail of the complex interplay of DNA repair pathways to repair ICLs is an area of considerable research, and the full details are not fully elucidated. The current knowledge of the ICL repair pathways is extensively reviewed by Deans (Deans and West 2011) and will be briefly discussed here.

The recombination-independent ICL mechanism, also described as 'mutagenic' ICL repair, requires XPC to recognise lesions followed by dual incision through the action of XPF-ERCC1 and XPG. Repair synthesis past the gap is performed by TLS polymerase ζ; a process recognised to result in potential mutation via this pathway (Huang and Li 2013).

Recombination-dependent ICL repair occurs during DNA replication. Repair is triggered when a replication fork is stalled by the lesion, resulting in initiation of the coordinated action of the NER, TLS and HR pathways. The NER factor XPF-ERCC1 makes dual incisions across the lesion on the same strand and TLS polymerases fill the resulting gap. The restored duplex is then repaired by NER. The remaining double strand break in the sister chromatid is repaired by classical homologous recombination using the repaired chromatid as the template (Huang and Li 2013).

The majority of evidence points to the intrastrand crosslink being the principle cytotoxic platinum-DNA adduct, as discussed above (section 1.4.1.6). A recent study has shown a correlation between ICL repair and the sensitivity of cell lines to cisplatin (Wynne et al. 2007) and the need to reassess the contribution of ICLs to platinum toxicity with further studies has been suggested (Enoiu et al. 2012).

1.4.2.2 Mismatch repair pathway (MMR)

Mismatch repair (MMR) corrects single base mismatches and repairs insertion/deletion loops (IDLs) which can arise as a consequence of base mismatches (Sameer et al. 2014). The MMR apparatus is highly conserved and operates specific to the damaged strand (Martin et al. 2008). A differential

response of the MMR pathway to platinum-DNA adducts is an important difference between the cellular responses to cisplatin adducts compared to oxaliplatin.

The details of the MMR pathway were initially elucidated in prokaryotes. The homologous system subsequently identified in eukaryotes requires the stepwise action of at least 6 Mut proteins. After detection of a nucleotide mismatch single strand incision occurs. Following this, recruitment of MMR proteins is triggered leading to strand incision, nucleotide excision, and the action of DNA polymerases to repair the defect (Martin et al. 2008).

In the prokaryotic MMR pathway the MutS protein acts as a sensor of mismatched DNA and triggers the MMR response. The eukaryotic system is more complex and five human MutS homologues (termed MSH proteins) have been identified. The functional unit of MSH proteins are heterodimeric, as combinations of MSH2, MSH3 and MSH6. The combination of MSH proteins depends on the specific mismatch to be repaired. MSH2 is required for the correct of the majority of mismatches in nuclear DNA, whilst MSH3 and MSH6 are required for distinct lesions (Marsischky et al. 1996). Specifically, the MSH2-MSH6 heterodimer, MutSα, recognises base-base mismatches and small insertion/deletion loops, whilst larger insertion/deletion loops are recognised by the MSH2-MSH3 heterodimer (MutSβ) (Jun et al. 2006).

Following mismatch or IDL recognition, human homologs of the *Escherichia coli* MutL protein (named MLH) are recruited to co-ordinate repair, again through formation of functional heterodimers. Four MLH proteins have been identified in humans: MLH1, MLH3, PMS1 and PMS2 (Jun et al. 2006). The assembly of the mismatch recognition complex with other MMR proteins required for repair are coordinated primarily by the MLH1-PMS2 heterodimer, MutLα. The other MLH heterodimers participate in a subset of IDL repair (Sameer et al. 2014). The interaction between MutSα and MutLα, PCNA and RPA leads to recruitment of EXO1 (Li 2014). Recruitment of EXO1 occurs to pre-existing or MutLα generated nicks 5' to the mismatch. EXO1 excises the mismatched base and creates a single strand gap that is filled by polymerase δ using the parental strand as the template. The nick is ligated by DNA ligase I (Li 2014).

Activation of the MMR pathway results in downstream activation of cell cycle regulators, including p53 and c-ABL, resulting in cell cycle arrest and/or apoptosis (Jun et al. 2006).

Cisplatin-DNA adducts can trigger the MMR pathway, unlike oxaliplatin, and in several studies deficiencies in the MMR pathway has been correlated with resistance to cisplatin (Jung and Lippard 2007). Cisplatin activation of downstream signalling requires functional MMR pathways, specifically the activity of the proteins MSH2 and MLH (Nehme et al. 1997). Two models have been proposed as to how MMR activation by cisplatin-DNA adducts induces cell death. One proposed mechanism is futile cycling. MMR may be triggered by a perceived mismatch between the cisplatin-guanine adduct and the opposing cytosine base. MMR proteins become engaged in cycles of attempted repair of the cytosine, without resolution of the initiating lesion, which can lead to futile cycles of repair, and ultimately to cell death (Kartalou and Essigmann 2001a). In the second proposed mechanism non-repairable lesions result in continuous DNA damage recognition through the MMR pathway acting as a sensor lesion, signalling through ATM/ATR to activate cell cycle arrest and apoptotic pathways (Bellacosa 2001).

A deficiency in MMR confers resistance to cisplatin but not to oxaliplatin (Fink et al. 1996). Defects in MMR diminish the cisplatin-induced expression of several damage response genes including c-jun and c-abl (Nehme et al. 1997). The same defects have little effect on the expression of the same genes following oxaliplatin treatment (Nehme et al. 1999). The DACH ligand of oxaliplatin decreases the affinity of MMR proteins for oxaliplatin-DNA adducts (Chaney et al. 2005; Di Francesco et al. 2002; Raymond et al. 2002) and MMR fails to recognise oxaliplatin adducts and trigger the repair pathway. This could explain some of the different therapeutic profile of these agents (Raymond et al. 1998b). The effectiveness of oxaliplatin in cisplatin resistant cell lines may be between due to repair or damage recognition established in the context of mismatch repair.

Defects in MMR also increase TLS past cisplatin adducts but not oxaliplatin (Vaisman et al. 1998), and cisplatin induced secondary mutations occur much more readily in MMR deficient cell lines (Chaney et al. 2005).

1.4.2.3 Trans-lesion synthesis (TLS)

Several DNA polymerases are able to synthesise DNA on the complementary strand to platinum-DNA adducts, a process termed translesion synthesis (TLS), albeit with reduced fidelity (Chijiwa et al. 2010). As discussed in section 1.4.2.1 TLS is a component of the repair of ICL, although the details of the interactions of the DNA repair pathway required for the repair of these lesions has not yet been fully elucidated (Huang and Li 2013). It has been demonstrated that XP-V cell lines proficient in NER but deficient in TLS are hypersensitive to cisplatin, indicating the importance of this DNA repair process in the repair of platinum–induced ICL (Chijiwa et al. 2010).

1.4.2.4 The Interaction of DNA repair pathways

In the discussion above the DNA repair pathways have been presented as independent pathways with separate roles in the identification, processing and repair of specific types of DNA damage. This reflects the development of the field, primarily through the use of experimental methods and techniques isolating and elucidating the specific details of individual pathways, initially in in model organism systems and using *ex vivo* models (Friedberg 2005). Recent evidence suggests, however, that complex interaction may occur between DNA repair pathways, particularly at the level of DNA damage detection and in the initiation of the DNA damage response (for an example see (Peng et al. 2014)). Whilst the mechanics of the majority of DNA repair pathways are reasonably well understood, this recent evidence, in combination with emerging evidence of the effect of chromatin structure on DNA repair (section 1.4.1.5), suggests that DNA repair *in vivo* will ultimately turn out to be more complex than initially envisaged (Friedberg 2005; Marteijn et al. 2014; Peng et al. 2014).

1.4.3 Summary: cisplatin-DNA adducts compared to oxaliplatin-DNA adducts

In general, the mechanism of action with respect to the profile of adducts formed by cisplatin and oxaliplatin is similar. At equimolar doses oxaliplatin forms fewer adducts than cisplatin but is equally potent (Chaney et al. 2005), indicating that oxaliplatin requires fewer inter and intrastrand cross links to achieve equivalent cytotoxicity to cisplatin (Luo et al. 1998; Raymond et al. 2002; Raymond et al. 1998b).

Studies indicate that once adducts have formed, the principle repair process, NER, does not discriminate between cisplatin and oxaliplatin adducts *in vitro* (Page et al. 1990; Reardon et al. 1999) and differences in nucleotide excision repair activity have an equal effect on cisplatin and oxaliplatin cytotoxicity *in vivo* (Petersen et al. 1996; Schmidt and Chaney 1993).

The bulky DACH carrier ligand of oxaliplatin results in differing interaction with damage recognition proteins, conferring less cross-resistance and enhanced activity compared to cisplatin (Raymond et al. 2002). The difference between cisplatin and oxaliplatin may be result of differential recognition damage recognition proteins such as high-mobility group box 1 (HMGB1), TATA box binding protein (TBP) and human upstream binding factor (UBF) and particularly of recognition and activation of the mismatch repair pathway, resulting in activity in cells which are resistant to cisplatin (Raymond et al. 1998a; Raymond et al. 1998b; Wang and Lippard 2005).

1.4.4 Oxaliplatin Induced Peripheral Neuropathy

One of the most devastating and long-term side effects from chemotherapy with platinum agents, and of oxaliplatin in particular, is damage to peripheral nerves resulting in permanent numbness and paresthesia in the distal extremities. For a detailed understanding of the potential role of predictive biomarkers in this condition it is important to understand the natural

history and molecular mechanisms of OIPN, as will be outlined in the following discussion.

1.4.4.1 Chemotherapy Induced Peripheral Nerve Damage

Neuronal tissue in the peripheral nervous system is especially vulnerable to the effects of cytotoxic agents; it is not protected by the blood-brain barrier, unlike the central nervous system. Drug-induced neurotoxicity can affect either the nerve fibres or neuronal bodies, although the typical effect is at the cell body in the dorsal root ganglia (DRG) of the primary sensory neurone. The neurotoxic effects of many of the commonly used chemotherapy drugs on the peripheral nervous system are a major clinical problem, given their widespread use, and peripheral neuropathy (PN) remains one of the main dose-limiting toxicities of several chemotherapy agents (for a review see (Argyriou et al. 2012a)). Even when PN is not dose limiting, the severity and duration of PN may significantly affect the quality of life of cancer patients and cause chronic discomfort (Kautio et al. 2011).

1.4.4.2 Natural history of OIPN

Peripheral neuropathy in oxaliplatin treated patients is common and occurs in two distinct forms – an acute, transient PN, and a chronic PN. Henceforth the chronic PN will be referred to as OIPN. The exact cause of each is not fully understood, although distinct but overlapping mechanisms have been proposed (Alcindor and Beauger 2011).

Acute neuropathy typically manifests as transient paraesthesia and dysaethesias of the distal extremities and perioral region. It typically occurs during and for a few hours following treatment, although occasionally symptoms can persist for one or two days. Symptoms are often precipitated by exposure to cold, and when occur are associated with symptoms of neuromuscular hyperactivity including jaw tightness, cramps and muscle fasciculations (Argyriou et al. 2013).

Chronic OIPN occurs frequently, with more lasting significance for many patients. It usually manifests during the several-month schedule of treatment, and, in some instances, can continue to worsen for many months after the full course of chemotherapy is complete. Approximately 50% of patients receiving oxaliplatin chemotherapy develop CTC grade 2 or more neuropathy, with CTC grade 3 (see figure 1.14) occurring in 10-20% (de Gramont et al. 2000).

NEUROLOGY						Page 4 of 5
		Grade				
Adverse Event	Short Name	1	2	3	4	5
Neuropathy: sensory	Neuropathy-sensory	Asymptomatic; loss of deep tendon reflexes or paresthesia (including tingling) but not interfering with function	Sensory alteration or paresthesia (including tingling), interfering with function, but not interfering with ADL	Sensory alteration or paresthesia interfering with ADL	Disabling	Death
REMARK: Cranial nerve sensory neuropathy is graded as Neuropathy: cranial – *Select*.						

Figure 1.14: NCI Common Toxicity Criteria for Adverse Advents Version 3
Sensory Neuropathy Criteria (http://ctep.cancer.gov)

The earliest clinical signs of OIPN are a decreased vibratory sensitivity in the toes and subtle reduction of ankle reflexes, associated with numbness, tingling or paraesthesia in the fingers and toes. Prolonged treatment may worsen symptoms and signs, resulting in the loss of deep tendon reflexes and gradual proximal spread of vibratory sensitivity impairment. Pin and temperature sensation, joint position and light touch perception are less severely affected (Alejandro et al. 2013; Argyriou et al. 2014).

Nerve conduction studies in the acute PN setting, 24-48 hours after oxaliplatin treatment, show neuro-myotonic discharges and repetitive compound muscle action potentials that fully resolve within 3 weeks of treatment (Lehky et al. 2004). In chronic OIPN nerve conduction studies consistently demonstrate sensory axonal damage with reduced sensory nerve action potentials (Argyriou et al. 2007; Krishnan et al. 2006). Motor nerve conduction velocities, compound muscle action potentials and F wave latencies remain unchanged, highlighting the specific sensory nature of this phenomenon (Cavaletti et al. 2001).

In general, there is no clinically validated tool for diagnosis or investigation of OIPN during treatment, and the chemotherapy schedule is modified according to self reported symptoms and clinician preferences, rather than an objective measure of neuropathy or risk of progression with on-going treatment. The management of OIPN is complicated by the fact that no treatment is available that has been shown to be effective in high quality large clinical trials (Ali 2010). High doses of analgesia may mask some of the symptoms of pain, but are of no benefit for numbness and have considerable risk of adverse effects (Ali 2010; O'Connor 2009).

Recovery from OIPN is frequently incomplete, even several years after treatment (Mangiameli et al. 2002). The condition is reversible to a degree in approximately 80% of patients, and it fully resolves in about 40% of sufferers six to eight months after treatment has completed (Argyriou et al. 2008). However, a significant proportion of patients are left with symptoms more than 2 years after treatment. In a large UK based trial of oxaliplatin and 5-FU/LV 44% of patients had CTC grade 2 or 3 OIPN at the end of treatment, 6% after 1 year and 4 % after 18 months (Andre et al. 2009) and in a large US trial of oxaliplatin as adjuvant therapy for colorectal cancer, NSABP C-07, more than 10% of patients still had persistent neurological symptoms 2 years after chemotherapy was complete (Land et al. 2007).

Long-term symptoms tend to occur in those with more severe symptoms at the time of chemotherapy. In a European trial of oxaliplatin treatment for colorectal cancer 26% of patients with CTC grade 3 OIPN still had persistent symptoms at a median follow up of 28 months (de Gramont et al. 2000).

These studies highlight that for patients with OIPN that fails to resolve after treatment it can remain long-term. In a cross-sectional study of patients treated with cisplatin for testicular cancer between 22 and 33 years previously, platinum-induced peripheral neuropathy was clinically detectable in up to 20% of patients, and 10% were still symptomatic (Glendenning et al. 2010). In agreement, in another study of cisplatin treated patients after a mean follow-up of 15 years: 38% and 28% had asymptomatic and symptomatic neuropathy respectively and PN was disabling in 6% (Strumberg et al. 2002). Although oxaliplatin has been used in the clinic over a much shorter timescale, so

comparable long duration studies of trends over decades are unavailable compared to cisplatin, similar rates appear to occur over the period oxaliplatin has been clinically available. For example, studies have documented long lasting OIPN in 35% of patients exposed to the drug 5-6 years after chemotherapy (Brouwers et al. 2009).

In general, patients with pre-existent or significant risk factors for neuropathy are excluded from clinical trials of oxaliplatin therapy (Adelsberger et al. 2000). However, common causes of peripheral neuropathy, such as diabetes mellitus, do not appear to be associated with an increased risk of OIPN (Ramanathan et al. 2010), nor does age (Goldberg et al. 2006). The principle risk factor appears to be the cumulative dose of chemotherapy drug (Glendenning et al. 2010). Chronic OIPN develops with increasing dose, and has been observed in the dose range 510-765 mg/m^2 in up to 10% of patients, but typically effects 50% in patients receiving >1000mg/m^2 (Argyriou et al. 2012b; Goekkurt et al. 2009a; Goekkurt et al. 2009b; Park et al. 2009).

It is the idiosyncratic response to oxaliplatin at moderate doses that may be most amenable to patient stratification approaches, allowing dose modifications in those at high risk and continuation of treatment in those unlikely to develop OIPN.

1.4.4.3 Molecular mechanism of OIPN

The specific molecular cause of OIPN appears to be through damage to peripheral nerves at the dorsal root ganglia (DRG) resulting in morphological and functional changes to DRG neural cells (Argyriou et al. 2014; Cavaletti et al. 2011; Meijer et al. 1999). In animal and in *in vitro* models, DRG cell apoptosis and neuronal atrophy are a common finding following oxaliplatin therapy (Ta et al. 2006), and the development of these changes is associated with impaired electrophysiological conduction, similar to that seen in patients with OIPN (Jamieson et al. 2005; Park et al. 2009; Renn et al. 2011).

There is a strong correlation between platinum-DNA adduct levels in DRG, induced neuronal changes and the severity of symptoms of OIPN in both animal models and human post-mortem studies (Dzagnidze et al. 2007; Gregg et al. 1992). Platinum-DNA adducts also appear disproportionately to accumulate in DRG cells compared to other tissues, indicating perhaps why OIPN is such a common toxicity experienced in patients treated with platinum agents (McDonald et al. 2005).

There is some debate about the exact mechanism through which increased levels of DRG platinum-DNA accumulation triggers neuronal changes of apoptosis, neuronal atrophy and the corresponding electrophysiological phenomena that give rise to the symptoms of OIPN.

1.4.4.4 NER in terminally differentiated cells

One theory, of particular interest to explain OIPN, is the idiosyncratic response of terminally differentiated cells to DNA damage, which may be a specific reason why peripheral nerves are particularly sensitive to oxaliplatin and platinum-DNA damage (Nouspikel 2007, 2011).

The functional capacity of NER is variable in different tissues, and aspects of the NER pathway, particularly GG-NER, are notably decreased in terminally differentiated cells, including neurones and peripheral blood mononuclear cells (PBMC) (Nouspikel 2007). As outlined in chapter 1 (section 1.4), the persistence

of unrepaired platinum-DNA adducts impairs two major molecular processes - transcription and DNA replication. TC-NER is maintained in post-differentiated cells as these continue to actively transcribe genes necessary for their specific cellular functions. Hence stalling of RNA polymerase II by DNA adducts and the failure of accurate transcription remains a significant problem. The fundamental importance of TC-NER for post-differentiated neuronal tissue is demonstrated by Cockayne syndrome, a failure of TC-NER, associated with developmental and neurodegenerative symptoms but without an increased risk of malignancy (Nance and Berry 1992).

Neurones, as an example of a post-differentiated cell, do not proliferate and are not replaced if lost. In terminally differentiated cells, including peripheral nerves, it has been demonstrated that NER is heavily down-regulated in the GG-NER sub-pathway (Nouspikel and Hanawalt 2000). As there is no requirement to repair non-transcribed regions of the genome in order to maintain the fidelity of DNA during replication, it has been hypothesised that the lack of evolutionary pressure to develop and maintain NER capacity in non-transcribed regions has resulted in the attenuation of the GG-NER pathway in comparison to TC-NER (Nouspikel and Hanawalt 2002).

Over time, an accumulation of DNA lesions in non-transcribed regions occurs in the context of attenuated GG-NER. This phenomenon has been demonstrated in post-mitotic human B-lymphocytes, which have also been shown to similarly down-regulate GG-NER. If B-lymphocytes are subsequently forced to re-enter the cell cycle and replicate high rates of mutation occur as a consequence of accumulated DNA damage in previously inactive genes (Hyka-Nouspikel et al. 2011).

An obvious hypothesis to explain the sensitivity of DRG cells for platinum adducts emerges from this body of work. If the neuronal cell is forced to re-enter the cell cycle the level of long-term accumulated damage (and additional burden of unrepaired platinum-DNA adducts accrued during months of chemotherapy) may be catastrophic, drug implications and the cellular resources available to repair the accumulated damage would be limited. This mechanism has been proposed to be a cause of Alzheimer's and other neurodegenerative diseases

(Nouspikel 2009), and could play a role in the exquisite sensitivity of DRG neurones for platinum-DNA damage.

Biopsies to obtain neuronal tissue for research and clinical investigation of peripheral neuropathies are rarely conducted and have a significant complication rate, including high rates of post-biopsy sensory loss, pain and paraesthesia in the distribution of the nerve from which material is taken (Hilton et al. 2007). This evidence suggests that other post-differentiated cells types may be a useful surrogate tissue for the investigation of OIPN, particularly PBMC, a cell type that is easily obtained by venesection.

1.4.4.5 Summary

The balance of evidence, and the widely accepted paradigm, is that chronic OIPN is a result of an idiosyncratic response to dose-dependent accumulation of platinum-DNA adducts in the dorsal root ganglia cell body. Adducts interfere with normal cellular function and trigger apoptosis, resulting in neuronal atrophy, although the exact mechanism by which apoptosis is induced is still a subject of considerable research (Sereno et al. 2014; Velasco et al. 2014).

The accumulation of platinum-DNA adducts also is a function of DNA repair, particularly NER. In NER deficient mouse models the accumulation of platinum-adducts at the DRG is increased, and the electrophysiological changes corresponding to clinical outcomes in humans are worse in these animals (Dzagnidze et al. 2007). Further evidence for this relationship is the correlation between polymorphisms in DNA repair genes and OIPN demonstrated in several genetic studies (Cavaletti et al. 2011). This area is discussed in more detail in chapter 6, section 6.1.3. Additional evidence supporting this link has been developed in a study by Professor J. Cheadle at Cardiff University, identifying several novel single nucleotide polymorphisms (SNPs) in NER genes associated with the development of OIPN in colorectal cancer patients treated with oxaliplatin-based chemotherapy (West 2013). These candidate OIPN biomarkers will be the focus of chapter 6 of this thesis, in which evidence will be

presented to demonstrate a functional link between DNA repair and oxaliplatin toxicity, confirming the importance of DNA repair as a critical factor in the development of OIPN.

The hypothesis surrounding the down-regulation of GG-NER and the accumulation of DNA damage is interesting, and, if correct, may be an area where our DIP-chip genome scale stratification technology could be applied. Our technique, capable of detecting accumulated DNA damage at high resolution on a genomic scale, could be used to identify patients with genomic DNA damage profiles (particularly in regions of attenuated GG-NER) that may be potential biomarkers of the risk of developing OIPN.

1.5 The measurement of platinum-DNA adducts and adduct DNA repair capacity as a biomarker for platinum response and toxicity

The discussion of NER and the above clinical, and preclinical, observations highlight the importance of platinum DNA-adducts and the importance of NER repair in the response seen in patients to these agents.

To-date, however, there are no proven methods to stratify patients for platinum based chemotherapy despite the considerable efforts and interest in the use of NER factors as predictive biomarkers, primarily because of the contradictory results from measurements of a single NER factor. It has been hypothesised that this reflects the role of NER as a complex system of multiple interacting proteins, and measurements of a single factor will always have poor sensitivity and specificity with many confounding variables (Bowden 2014). Potentially a functional measurement of adduct levels rather than the level of factors involved, which could be used to measure the changes in adduct frequency over time after drug exposure to measure NER capacity, could obviate these issues. This possible approach - measuring adduct levels rather than NER factor levels - will be discussed in the final section of this introductory chapter.

1.5.1 Methods of measuring platinum-adducts and DNA damage

Until recently, the determination of DNA damage has required the use of techniques that could be considered as either low or high resolution. The techniques classed as low resolution allow total genomic damage to be measured, but are unable to be used to determine DNA damage and repair at the level of the genome or over small ranges of nucleotides. High resolution techniques can be used to measure DNA repair at the single nucleotide or short oligonucleotide levels, but typically can only be used over a very small range of the genome, and so do not allow genome scale hypotheses or conclusions to be generated.

1.5.2 Low-resolution DNA damage detection techniques.

Initial techniques to determine and detect DNA damage used radioactively labelled DNA to assess for the activity of DNA damaging agents. Alkaline sucrose gradient techniques were used to separate labelled strands of genomic DNA by length to determine the presence of different length fragments after damage with x-radiation and alkylating agents (McGrath and Williams 1966). The incorporation of labelled thymine bases into damaged cells can also be used as a gross marker of damage (Setlow and Carrier 1964). This type of assay was first used to demonstrate the activity of NER in mammalian cells (Rasmussen and Painter 1964). Typically only cells in S-phase accumulate radiolabelled nucleotides, indicating on-going DNA replication. After DNA damage with UV radiation short patches of incorporated labelled bases can be detected by autoradiography, first evidence of new sequence synthesis following DNA damage and evidence of NER.

Several gel-based assays have been used to detect and measure DNA damage, again all at low resolution. Pulsed Field Gel Electrophoresis (PFGE) can be used to detect double strand breaks, with the identification of chromosomes as bands on a gel (Contopoulou et al. 1987) and a modification, the Fraction of Activity Released (FAR assay) (Rydberg et al. 1994) has been used to measure the amount of DNA released from a plug of gel containing DNA samples with separation of fragments by PFGE.

Chemical modifications to separate single strand DNA (for example the alkaline unwinding assay, (Elkind and Chang-Liu 1972)) can be used to measure single strand breaks following lesion excision in NER (Erixon and Ahnstrom 1979).

One commonly used assay operating at low resolution to detect DNA damage is the comet assay (Collins 2004). Electrophoresis of DNA products from a single cell separates DNA by the length of fragment. Damaged DNA results in either short or long fragments, with more heavily damaged DNA forming a smear or 'comet' tail. This assay can be adapted to measure single and double strand breaks, crosslinks and base damage, and has been adapted for high throughput screening in 96 well plates (Wood et al. 2010). In a similar type of assay using

circumferential separation of DNA fragments is the HALO assay, using diffusion to determine short and long DNA fragments (Sestili et al. 2006).

The TUNEL assay (terminal deoxynuleotidyl transferase mediated dUTP-biotin nick end labelling) uses a similar system but with fluorescence rather than radiolabelling to determine the proportion of DNA fragments present (Gavrieli et al. 1992).

Other commonly used low resolution techniques include mass spectrometry, which although highly sensitive is expensive and limited by the amount of material that is needed for detection, and immunological methods of DNA damage detection using antibodies against the damage of interest. This includes immuno-slotblot, where damaged DNA is transferred to a membrane and probed with a primary antibody to the damage and identified with an appropriate secondary antibody and has been used for UV damage detection (Perdiz et al. 2000).

1.5.3 High-resolution DNA damage analysis techniques

DNA damaging agents block the action of many DNA polymerases, allowing modifications of the polymerase chain reaction (PCR) to be utilised as an assay to detect DNA damage, demonstrated initially in a study by Govan III in 1990 (Govan et al. 1990). Analysis of the relative reduction in PCR yield between undamaged and damaged DNA regions can give high-resolution information on relative DNA damage at a limited number of sites.

An assay utilising enzymes to cut DNA adjacent to sites of damage and subsequently separation of resulting fragments by electrophoresis can also give a high-resolution view of damage and repair. A method developed in our laboratory by Teng (Teng et al. 1997) uses a T4 phage endonuclease to cut at UV radiation induced CPD damage sites and specific end-damage probes to pick out areas for investigation. Radiolabelling of the longer undamaged and shorter damaged fragments, again by gel electrophoresis, allows identification of areas of damage and repair at high resolution but at only a short 100-200 nucleotide range.

1.5.4 The use of functional measures of DNA damage as clinical tools

The discussion above highlights the role of adduct formation and repair via NER as the main mechanism of action of the platinum agents, through DNA damage in tumour tissue to give response, and through DNA damage in normal tissues to cause toxicity. The benefits of a tool for stratifying patients for treatment are evident. However, so far, no DNA repair assay has been developed into a clinically useful tool, despite the use of the platinum agents in the clinic for over 40 years.

The clinical application of DNA repair and NER factor assays will be discussed in more detail in chapter 4, and, to summarise, the failure of these tools is due to a lack of specificity, sensitivity and predicative power. The currently available assays measure either the level of a single NER repair factor, at risk of low specificity by confounding from the effects of unmeasured repair factors, or use a low resolution approach to measure platinum DNA adducts with a lack of sophistication and sensitivity.

A novel DNA damage and repair analysis technique developed in our laboratory combines the measurement of DNA adducts through comparison of treated and untreated samples by DNA-immunoprecipitation (DIP) of platinum damaged DNA. When combined with microarray analysis (chip) this DIP-chip technique give high-resolution measurement of levels and patterns of platinum-DNA adducts throughout the genome.

The technique will be described in detail in this thesis; the DIP assay in chapter 3 and the chip aspect in chapter 4. It may be possible to translate this assay for use on clinical samples to improve on the sensitivity and specificity of currently available DNA analysis tools, and to develop an assay capable of predicting response and toxicity from platinum based chemotherapy treatment.

1.5.5 Potential applications of human DIP-chip platinum-DNA adduct patterns

The ability to measure and map the genomic location of platinum-DNA adducts at high-resolution with the DIP-chip assay offers a potential mechanism to improve the sensitivity and specificity of currently available platinum-DNA adduct based methods of patient stratification for response and toxicity to platinum chemotherapy. This theme will be discussed in more detail in the introductory sections of chapters 3 and 4.

The ability to map the genomic location of platinum-DNA adducts in human samples has other potential research applications. For instance, the rapid development of DNA sequencing technology has allowed the sequencing of thousands of paired tumour and normal tissue samples, resulting in the identification of thousands of somatic mutations present throughout the genome of germline DNA and cancer tissues (Alexandrov et al. 2013a; Alexandrov et al. 2013b). Analysis of patterns of mutation has revealed the presence of characteristic mutational signatures, associated with particular mutagenic processes. For example, characteristic patterns of mutation associated with UV radiation exposure can be determined in the genome of cells from melanoma samples, and patterns of mutation following treatment with the alkylating chemotherapy agent temozolomide in melanoma patients treated with this agent can also be identified. Of the 21 mutational signatures identified so far over half do not have an obvious underlying mutagenic cause.

The presence of mutational signatures reflects previous carcinogen exposure, and have been proposed as a method for identifying patients at risk of developing malignancies associated with these particular mutational patterns, allowing closer monitoring and screening for these cancers. The DIP-chip assay generated patterns of platinum-DNA adducts could potentially be applied in several ways. For example, in the context of chemotherapy induced mutational signatures, the ability to map the location of platinum-DNA adducts (or other DNA adducts depending on DIP-chip assay modifications) and correlate these patterns with resulting mutational signatures could provide further evidence to support this area of research. Additionally, as a research tool it may be possible

to compare DIP-chip adduct patterns with mutational signatures which have no currently identified mutagenic process, as a method to determine the cause of this phenomenon. Finally, as the DIP-chip adduct pattern is a precursor to the development of mutated DNA sequence, in the same way mutational signatures offer an potential avenue for early identification of malignancy risk the precursor, DIP-chip signatures are an even earlier identification step in this process.

1.6 Aims and objectives of the current study

This thesis focuses on the relationship between oxaliplatin-DNA adducts, DNA repair capacity and the response and toxicity to oxaliplatin based chemotherapy treatment.

The first three results chapters of this thesis will describe the translation of the DIP-chip DNA damage and repair assay into a clinical tool capable of investigating, in high resolution and at genomic scale, the relationship between oxaliplatin-DNA adducts and clinical outcome in clinical samples. This will involve the modification and validation of the DNA-immunoprecipitation technology for use on clinical samples in chapter 3. Chapter 4 will focus on the translation and optimisation of the microarray technique to producing accurate patterns at genomic resolution of platinum-DNA adducts that could potentially be used to stratify patients for treatment.

The development and translation of genomic-scale technologies into a tool for patient stratification is a significant challenge, and especially requires detailed bioinformatic analysis to allow appropriate and accurate data interpretation. Chapter 5 describes the development of new bioinformatic techniques for interpreting DIP-chip microarray outputs. These techniques are a significant advance on previously available bioinformatic tools and, when used in this chapter, shed new light on the interpretation and understanding of platinum DNA adduct patterns in this context, and are essential for the further development of this technique into a clinically applicable tool.

In the second section of this work, presented in chapter 6, evidence will be described linking single nucleotide polymorphisms in the human NER gene *XPF* with the development of oxaliplatin induced peripheral neuropathy (OIPN). This phase of the study will use new information, derived from DNA sequencing of colorectal cancer patients, and develop a functional model of OIPN in *S. cerevisiae*, a model organism ideally suited for genetic manipulation. This experimental system will allow the functional impact of NER gene SNPs to be studied in simplified and controlled conditions. The results described in this phase of the study will demonstrate a functional link between newly identified single nucleotide polymorphisms in NER genes on the development of oxaliplatin toxicity. This evidence is important functional evidence to support the statistically significant findings of the genetic study that discovered the SNPs modelled, as well as providing support to the underlying hypothesis of this thesis - that measuring differences between individuals in their chemotherapy-induced platinum-DNA adduct levels and NER capacity are an important route towards the goal of patient stratification for the platinum chemotherapy agents.

Chapter 2 Materials and Methods

The materials and methods used to conduct the experiments discussed in this thesis are described in this chapter. Where possible, the supplying company and catalogue number of items sourced for use in the experiments described in this thesis is indicated by the abbreviation CN. The majority of the solutions used in this study are described in the relevant text, but occasionally some of the solutions are named but the details of the contents are not fully described. Further information about these solutions can be found in appendix 1.

2.1 The human PBMC DIP-chip assay protocol

The protocol for the DIP-chip assay extensively used in this thesis is described below. The version described here is the final version of the human PBCM DIP-chip assay; it has been modified following the experiments described in chapters 3 to 5, and is the protocol that was used to generate the data presented in these chapters.

2.1.1 Preparation of oxaliplatin and cisplatin

Oxaliplatin and cisplatin was obtained from the chemotherapeutics pharmacy at Velindre Cancer Centre, Cardiff, and was stored at 4°C at a stock concentration of 12.6mM for oxaliplatin and 3.3mM for cisplatin. Commercially obtained oxaliplatin was obtained from Abcam (CN:ab141054) as a powder and stored at 4°C for up to 6 months. The powder was dissolved in DMSO (Sigma CN:D2650) to a concentration of 12.6mM immediately before use.

2.1.2 Blood samples, PBMC isolation and platinum drug treatment protocol

The peripheral blood mononuclear cells (PBMC) used in the experiments in chapters 3 to 5 were harvested from whole blood samples using density gradient centrifugation. Samples were obtained from two healthy volunteers (labelled throughout as patient A or B) or were obtained from donations from the Welsh Blood Service following ethical approval for the use of these tissues for research purposes (application number WBS Ad-hoc 008-12). Samples from the Welsh Blood Service consisted of a single donor unit of 450ml whole blood, following the removal for clinical purposes prior to delivery of the majority of erythrocytes, plasma and platelets. These samples were diluted 50:50 with PBS prior to use. All samples were processed using the following protocol:

1. Whole blood samples were divided into equal aliquots of a maximum of 30ml and gently overlaid onto 15mls of room temperature Histopaque-1077 (Sigma-Aldrich) in 50ml sterile tubes.
2. Samples were centrifuged at 500g for 20 minutes at room temperature in a Heraeus Megafuge 1.0 swinging-head centrifuge with the brake in the off setting.
3. Plasma was carefully removed to within 5mls of the buffy coat interface and discarded. The buffy coat layer was aspirated using a 3ml Pasteur pipette (Fisherbrand CN:FB55349) and transferred to a 15ml sterile tube, before dilution to 15mls final volume in warm PBS (pre-heated to 37°C). Samples were centrifuged at 300g for 5 minutes at room temperature and resuspended in 15mls PBS. The samples were centrifuged for a second time with the same parameters and again the supernatant was removed and discarded.
4. To remove erythrocyte contamination the pellet was briefly resuspended in 5 ml of sterile H_2O for a total of 5 to 10 seconds before being diluted to 15ml with PBS.
5. All samples were centrifuged at 300g for 5 minutes at room temperature before the supernatant was removed and the sample

resuspended in 1 ml of RPMI-1640 supplemented with 1% L-Glutamine and HEPES (Life-Technologies CN:52400-025), 10% foetal calf serum (Invitrogen CN:10106-169) and 1% penicillin and streptomycin (Sigma CN:P0781), pre-warmed to 37°C.

6. The total number of cells were counted using a KOVA Glasstic slide 10 haemocytometer (Fisher CN:22-270-141) and the final cell concentration in the sample was adjusted to $5*10^6$ cells per 200μL of media. Subsequently, 200μL of media containing $5*10^6$ cells were added 4.8ml of media per well in a 6 well plate (Thermo Scientific CN:130184) and incubated in a Heraeus HERA cell incubator or 30 minutes at 37°C and 5% CO_2.
7. After 30 minutes, oxaliplatin or cisplatin was added to a final desired concentration and cells were incubated in a Heraeus HERA cell incubator for 24 hours at 37°C and 5% CO_2.
8. Following incubation, media was aspirated and transferred to a 15 ml sterile tube. All plates were washed twice with 2mls of warm PBS and the surface was gently scraped during each wash using a cell scraper (BD Falcon CN:353085). The PBS from each wash was added to the collected media and plates were checked to ensure cell removal using a Leica DMIL microscope. If large numbers of cells remained a further 2ml PBS wash and repeat scraping step was employed.
9. All samples were centrifuged at 300g for 5 minutes at room temperature and the supernatant was removed and disposed of using appropriate protocols for the disposal of waste contaminated with cytotoxic agents. The cell pellet was resuspended in 200μL room temperature PBS before re-centrifugation at 300g for 5 minutes at room temperature. The supernatant was removed and appropriately disposed of. Cells were re-suspended in 200μL room temperature PBS prior immediate DNA extraction, or stored overnight at -20°C.

2.1.3 DNA extraction

DNA was extracted using the Qiagen DNeasy blood and tissue Kit (Qiagen CN:69506) following the manufacturers protocol for DNA extraction from cell culture material with the following amendments at the start and end of the protocol.

1. Prior to commencing DNA extraction: PBMC pellets are split into two equal samples in a final volume of 200μL PBS prior to DNA extraction.
2. At the end of the protocol: DNA extraction columns are eluted a second time in a volume of 100μL buffer AE. Both the initial 200μL and the second 100μL elution buffer volumes were combined prior to sonication.

2.1.4 DNA Sonication

A Bioruptor Sonicator (Diagenode) was used to fragment the DNA with the following steps:

1. Samples of genomic DNA in 300μL volume of buffer AE from the Qiagen DNeasy Blood and Tissue Kit were transferred to 1.5ml micro-centrifuge tubes.
2. Tubes were placed in a cooling water bath at 4°C before sonication was conducted on the high setting, with cycles of 30 seconds on/30 seconds off for a total of 30 cycles.
3. After 15 cycles were complete the samples were removed, kept on ice, and centrifuged briefly to ensure all DNA was in the maintained in the lower section of the tube. The final 15 cycles were continued with the same settings.
4. Equal sonication between samples was confirmed using a 1.2% TAE gel and a GeneRule 100bp ladder for reference, with a desired mean fragment length of 400 nucleotides.

2.1.5 DNA purification

Sonicated DNA was purified with PureLink PCR Kit (Invitrogen CN:K3100-01) following the manufacturers standard protocol. Elutions volumes of 50μL from the paired samples that were previously split into two volumes at the stage of DNA extraction were recombined following elution.

2.1.6 DNA immunoprecipitation

Sonicated DNA was used as the starting material for the immunoprecipitation of oxaliplatin-DNA adducts with CP9/19 anti cisplatin-modified DNA antibody (Abcam CN:ab103261) with the following protocol:

1. A well-mixed aliquot of 40μL per experimental condition of sheep anti-rat Dynabeads (Invitrogen CN:11035) was washed three times in 500μL of cold (4°C) PBS-BSA solution (0.1%) and resuspended in a final volume of 50μl PBS-BSA (0.1%) per sample to a maximum volume of 500μL. To this, 1.5 μg of CP9/19 antibody per sample was added before incubation at 30°C for 30 minutes at 1300 rpm in a thermomixer (Eppendorf). Following incubation the Dynabead/antibody mix was washed three times in chilled (4°C) PBS-BSA (0.1%) and resuspended in 50μl chilled PBS per sample.
2. PBS-BSA (1%) of 30μl and 6μg of sonicated DNA solution were added to each 50μl aliquot of Dynabead/antibody mix and made up with PBS to a final volume of 300μl.
3. The samples were incubated in a thermomixer for 3 hours at 21°C at 1300 rpm.
4. Following incubation, a series of wash steps were conducted in the following order, with all reagents chilled at 4°C: 1x 500μL FA/SDS buffer, 3 washes in 500μL FA/SDS+1mM NaCl, 1 wash in 500μl LiCl solution, 1 wash with 500μL 1X TE.

5. After complete aspiration and removal of all residual TE, DNA was eluted by the addition of 125μL 1X pronase buffer and incubation at 65°C for 30 minutes at 900 rpm in a thermomixer.
6. All samples were then incubated overnight at 65°C following the addition of 6.25μl 1X pronase (20mg/ml).
7. Control IN samples were produced using 600ng of the pre-immunoprecipitation DNA (a 10th of the IP starting quantity), made up to 100μl in 1X TE and then by the addition of 25μl 5X pronase buffer. The IN samples were incubated alongside the IP sample overnight at 65°C following the addition of 6.25μL pronase.
8. The next day 5μl RNase A (10mg/ml) was added to all IP and IN samples and incubated at 37°C for 1 hour.

2.1.7 Reversal of platinum-DNA adducts

To remove platinum adducts from the DNA, IP and IN sample were incubated at 65°C for 2 hours following the addition of 32.5μL of 1M NaCN, following appropriate precautions for the use and disposal of sodium cyanide.

2.1.8 Phenol/chloroform purification and ethanol precipitation

Following cyanide treatment samples were purified using phenol/chloroform and precipitated under the following conditions:

1. 1 mL of 25:24:1 phenol/chloroform/isoamyl-alcohol mix was added to each sample.
2. Samples were vortexed vigorously and centrifuged at 15,000g for 20 minutes.
3. The upper aqueous phase was carefully removed and retained.

4. An additional phenol:chlorophorm:isoamyl-alcohol purification and a subsequent chloroform only purification were used following the same procedure (repeat steps 1-3).
5. Glycogen 0.5μl (20mg/ml), 18ul of 3M NaAC and 180μl 100% ethanol chilled to -20°C was added to the final aqueous phase.
6. Samples were chilled at -80°C for 45 minutes followed by -20°C for 45 minutes.
7. All samples were centrifuged at 4°C, 20,000 rpm for 20 minutes.
8. The supernatant was removed and discarded before the pellet was washed in 200μl 75% ethanol at -20°C and re-centrifuged at 20,000 rpm 10 minutes at 4°C.
9. The supernatant was removed and the pellet was dried for 5 minutes using an ISS110 SpeedVac system (ThermoSavant) and resuspended in 13.5μl PCR grade H_2O.

2.1.9 Quantitative PCR (qPCR)

The DNA content of the post-immunoprecipitation samples was quantified by qPCR. Samples were amplified using the CFX-connect Real-Time PCR Detection System (Bio-Rad) equipment using the iTAQ universal SYBRgreen supermix, and primers to the 28S genetic loci, with the following parameters used:

1. The qPCR standards used were 'master' standards, pre-diluted in 5 sequential 10-fold dilutions from as mixed sample of untreated IN DNA. Identical standards are used as a quality control measure, to quantify the relative amount of DNA per experimental sample (with respect to the amount in the standards), and to allow accurate comparison between plates.
2. 2.5μl of IP sample was diluted with14μl H_2O
3. IN samples were diluted 500 fold.

4. 5μl of either standards, or diluted IN and IP sample, was mixed with 5μl of SYBRgreen/1% 28S primer mix (10mM concentration) for the PCR reaction, giving a final primer concentration of 500nM in a final well volume of 10μL
5. All qPCR experiments were processed in 96-well plates (Hard-Shell 96 well plates, Bio-Rad) and each sample was processed in triplicate, including for the standard samples. After sealing the plates, each plate was vortexed briefly and centrifuged at 2000 rpm for 2 minutes at room temperature.

2.1.9.1 qPCR programme settings

1. 95°C for 3 minutes
2. 95°C for 10 seconds
3. 60°C for 20 seconds – followed by optical image 4. Go to step 2 x 44 times
5. 95°C for 10 seconds
6. 65°C for 5 seconds
7. Melt curve from 65°C

2.1.9.2 The 28S primer sequence

The sequence of the 28S primers used in this study were:

28S:	Forward:	5′ - CGCAATACGAATGCCCCCG	Tm 61.0 °C
28S	Reverse:	5′ - AGCCGCCTGGATACCGC	Tm 60.5 °C

2.1.9.3 Data analysis and calculation of IP/IN ratio

All data was analysed using Bio-Rad CFX Manager (Version 3.1) and Microsoft Excel in the following process:

1. Prior to analysis, the threshold for CT calculation was adjusted to 500 RFU, ensuring that each sample crosses the threshold during the logarithmic phase of fluorescence increase.
2. Standards were checked to ensure quantification was consistent with previous experiments (depending on the batch of master standards used) and that PCR efficiency was between 80% and 110%
3. Any unknown sample replicate results outside the linear range of the standards or with a threshold cycle value difference greater than 0.5 cycles from the other two replicates were discarded.
4. The DNA quantity relative to the standards was calculated, and exported to Excel.
5. The mean starting quantity from triplicate wells for each of the IP and IN samples was used to calculate an IP/IN ratio for each experimental condition.
6. This value was normalised to the IP/IN ratio from an untreated sample.

2.2 Microarray Work

To process the immunoprecipitated DNA for genome scale microarray analysis (the 'chip' aspect of the assay) the following protocols were used:

2.2.1 DNA amplification

For the amplification of IP and IN samples prior to labelling and hybridisation the GenomePlex WGA2 kit (Sigma CN:WGA2-50RXN) was used according to the manufacturers standard protocol, but with the following amendments:

1. As the DNA is pre-fragmented, the WGA2 fragmentation step is omitted
2. A volume of 10μl from the post immunoprecipitation IP sample is used
3. A volume of 10μL of a 1:10 dilution from the IN sample is used

4. The amplification is conducted according to the manufacturers protocol, except a total of 15 cycles of PCR is used.
5. During the post-amplification purification stage, DNA samples were eluted in 50μl H_2O and concentrated to a final volume of 15μL using an ISS110 SpeedVac system (ThermoSavant)
6. The DNA in each sample was quantified using the Nanodrop-1000 spectrophotometer and adjusted to an equal concentration (which must be >50ng/μL to allow adequate DNA from later microarray hybridisation).

2.2.2 Labelling

The IP and IN samples were differentially fluorescently labelled with Cy5 and Cy3 fluorophore respectively using the BioPrime Total Genomic Labelling Module (Invitrogen) using the following protocol:

1. The IP and IN sample were, pre-adjusted to identical DNA amounts >50ng/μl and in a volume of 10.5μL. 2.5μL of 5mM EDTA and 15μl Cy3 (to the IN sample) or Cy5 (to the IP sample) was added and mixed with pipette before incubation at 95°C for 5 minutes in the dark.
2. Each sample was chilled on ice for a 5 minutes, following which 2μl of Exo-Klenow fragment was added. Samples were incubated for 2 hours at 37°C.
3. The labelled DNA samples were purified using the Invitrogen columns provided with the BioPrime Labelling Module and eluted in a volume of 52μL of elution buffer
4. The efficiency of labelling was assessed using the NanoDrop-1000 Spectrophotometer Microarray Measurement setting.
5. 50μl of paired IP and IN samples were combined together and 12μl of NaAc (3M) and 5μl polyacrylamide (2.5μg/ml) were added to each combined sample.
6. The DNA was precipitated with the addition of 290μl of 100% ethanol

and cooling the samples at -80°C for 10 minutes or -20°C overnight.

7. Each sample was centrifuged at 13000 rpm for 20 minutes at 4°C. The supernatant was removed from each sample and the DNA pellet was washed with 200µl of cold 75% ethanol and centrifuged again at 13000 rpm for 10 minutes at 4°C.
8. The supernatant was removed and the DNA pellet was dried for 5 minutes using the SpeedVac system (Thermo Savant). Dried DNA pellets were stored at -20°C overnight.

2.2.3 Hybridisation and washing

1. Pellets were resuspended in 39µl of water.
2. The following reagents were added to each sample in the following order; 5µl of human Cot-1 DNA (1.0mg/ml) (Invitrogen CN:15279-011), 11µl of Agilent 10x blocking agent and 55µl of Agilent 2 x hybridisation buffer (Agilent CN:5188-5220).
3. The hybridisation mix was incubated at 95°C for 3 minutes then at 37°C for 30 minutes.
4. The total volume of 110µl of each sample was applied to the desired microarray on the glass slide, and firmly sealed with the appropriate cover slip and metal gasket (Agilent CN:G2534-60011), before transfer to a rotating hybridization oven pre-heated to 65°C. Microarrays were incubated for 24 hours.
5. Following hybridisation, slides were removed from the apparatus and washed for 5 minutes in room temperature wash buffer I (consisting of 300ml of 20 x SSPE, 250µl of 20% sarcosine and 700ml of H_2O), followed by a second wash for a further 5 minutes in wash buffer II (3ml of 20 x SSPE, 997ml of H_2O). Wash buffer II was warmed to 31°C before use.

2.2.4 Microarray scanning and data processing

1. All microarrays were scanned with an Agilent G2565BA microarray scanner using the following settings:

Scan region	61 x 21.6 mm
Scan resolution	5µm
5µm scanning mode	Single pass
eXtended dynamic range	Selected
Dye channel	Red & Green
Green PMT	XDR Hi 100%, XDR Lo 10%
Red PMT	XDR Hi 100%, XDR Lo10%

2. The output of the scanner is in the form of a TIFF file that can be processed using Agilent's Feature Extraction Software to extract the data from the microarray output. This data format was then analysed with Sandcastle software. Both of these stages are extensively described in chapter 4, section 4.1.3.5. The Sandcastle software and full documentation is available for download from http://reedlab.cf.ac.uk.

2.3 The materials and methods for *S. cerevisiae* experiments described in chapter 6

2.3.1 Yeast media

The recipes for growth media used in this study are listed in appendix 1. All growth media was made up before use and autoclaved on a liquid cycle. To

produce plates, 2% agar was added prior to autoclaving, and 25ml was poured into sterile petri dish whilst in liquid form and allowed to set.

2.3.2 Storage and growth conditions

The strain, genotype and the origin of the yeast strains used in this thesis are described in chapter 6. The yeast manipulations discussed in this section were conducted in sterile conditions using growth media and equipment sterilised by autoclaving before use.

For the long-term storage of yeast strains, cells were grown to the exponential phase and suspended in a 30% glycerol solution prior to freezing with liquid nitrogen. In this way, all strains were kept in long-term storage at -80°C. Before use, frozen cultures were streaked onto solid media plates and incubated at 30°C in a LEEC compact incubator until grown to stationary phase, typically after 2 days. These plates were then stored at 4°C for short-term use.

For each experiment, single colonies from short-term plates were inoculated in fresh liquid media and maintained in exponential phase. This pre-culture was used to inoculate liquid media, as required, with the necessary dilution calculated to achieve the required cell density for the experiments conducted on the following day. After inoculation all liquid cultures were incubated at 30°C in an Infors HT multitron standard incubator at 180 rpm.

Cell cultures were grown to a density of $2*10^7$ cells/ml before use, calculated using an Improved Neubauer BS7482 cell counting chamber (Hawksley). When an OD_{600} measurement of cell density was required, 1ml of cell culture was used for assessment of optical density at 600nm using a UV-Vis Spectrophotometer (Thermo Scientific)

2.3.3 Gel electrophoresis

Gel electrophoresis was used routinely during this study. Depending on the application, different agarose gel concentrations were prepared. Typically,

1.0% agarose gel was used, whilst a lower concentration of 0.8% agarose was used for isolation of DNA during cloning. Agarose was added to TAE buffer (40mM Tris-Acetate, 1mM ethylenediamine tetraacetic acid (EDTA), pH8.0) and heated until dissolved. The solution was cooled to approximately 50°C and, for every 100ml of solution, 5µl of 10mg/ml ethidium bromide was added. TAE gels were cast in either a Horizon 58 (Life Technologies) or a Mini-Sub cell GT (Bio-Rad). DNA was added to sterile water up to a final volume of 10µl, to which 2µl of 6x MassRuler loading dye (Fermentas) was added. The gel tanks were filled with TAE buffer and samples were loaded into each well, with either GeneRuler 1kb DNA ladder or FastRuler low range DNA ladder (Fermentas) added to the first lane. Gels were run at 10V per cm on a power-pac 200 (BioRad) at room temperature. After electrophoresis, the gel was examined using a BioDoc-It imaging system (UVP) at a wavelength 302nm.

If, following electrophoresis, isolation and purification of specific DNA bands was required the gel was visualised at a wavelength of 365nm to minimise UV damage to the DNA, and the desired band was excised from the gel with a scalpel. The extracted DNA was purified using a PureLink quick gel extraction kit (Invitrogen) according to the manufacturer's instructions.

2.3.4 Polymerase chain reaction (PCR)

All PCR was performed in a total reaction volume of 50µl, using either GoTaq Flexi DNA Polymerase (Promega), Phusion High Fidelity DNA polymerase (NEB) or Q5 Hot Start High Fidelity DNA Polymerase (NEB), and all reactions were run on a PTC-200 PCR machine (MJ Research) using the conditions described in the text and according to the standard protocol for each enzyme, as documented in the manufacturers datasheet. The annealing temperature of each primer was calculated as 5°C lower than the lowest primer 'Tm' and the extension time was calculated according to the respective enzyme data sheet. All PCR products were purified using the PureLink PCR Purification Kit (Invitrogen) according to the manufacturer's instructions and eluted into either water at pH 8.5 or 1X TE buffer.

2.3.5 DNA sequencing

Sequencing reactions were performed using the BigDye terminator v3.1 cycle sequencing kit (Applied Biosystems) using the manufacturers protocol.

Briefly:

1. A PCR reaction was prepared in a 0.5ml polypropylene tube on ice with the following constituents:

Template DNA	20-40ng for 500-1500bp PCR product 5-20ng for plasmid DNA
Primer	1µl at 3.2µM stock
Ready reaction premix	4µl
BigDye sequencing buffer	2µl
Water	Up to a total of 20µl

2. A PCR reaction was performed using a PTC-200 PCR machine (MJ Research) and the following programme:

1. 96.0°C for 1:00 minutes
2. 96.0°C for 0:10 minutes
3. 50°C for 0:05 minutes
4. 60.0°C for 4:00 minutes
5. Go to 2, 25 times
6. End

3. 5µl of EDTA and 60µl of 100% ethanol were added to each reaction. Samples were centrifuged at 13000 rpm for 20 minutes in a Beckman Coulter Microfuge 22R centrifuge at 4°C.
4. After aspirating the supernatant, the pellet was washed with 60µl of 70% ethanol and subsequent centrifugation was performed for 5 minutes with the conditions as described above.
5. The supernatant was removed and pellets were dried for 15 minutes.

6. All samples were analysed by Central Biotechnology Services at Cardiff University.

2.3.6 Cloning

DNA manipulation was performed for cloning purposes as described in chapter 6. The manipulations involved included DNA restriction digestions and DNA ligation. All enzymes and buffers required were obtained from New England Biolabs (NEB).

1. All DNA used was purified into either water or 10mM Tris-HCl pH 8.0, and quantified using a Nanodrop 1000 spectrophotometer (Thermo Scientific) according to manufacturer's instructions.
2. All restriction digestions were conducted in a volume of 40µl. DNA of 1500ng was combined with water to a volume of 34 µl for single enzyme digests, or 32 µl for double enzyme digests. 4 µl of enzyme buffer and 1 µl of each enzyme (at a starting concentration of 10,000 or 20,000 units/ml) was added and the reaction was incubated at 37°C for 90 minutes. DNA was either directly purified with the QIAquick PCR purification kit (Qiagen) to 30µl water or purified following gel electrophoresis as described in section 2.3.3.
3. For DNA ligation and cloning, DNA samples were combined with water to a volume of 18µl, consisting of approximately 200ng of DNA. An insert: vector ratio of 3:1 was used. 2µl of 10x T4 DNA ligase reaction buffer and 1µl of T4 DNA ligase (400,000 cohesive end units/ml; NEB) was added and reactions were incubated at 4°C overnight or at 16°C for 2-3 hours.

2.3.7 Site directed mutagenesis

Site directed mutagenesis was conducted using the Quick-change Lightening SDM kit (Agilent) according to the manufacturers instructions.

Primers used for SDM were as described in the relevant text. Briefly, 10ng of the plasmid template was added to a PCR reaction mixture containing 2x QuikChange Buffer, 1.25 pmol of each primer, 10mM dNTP, 6% QuikChange reagent, and 2.5U of *PfuUltra* HF DNA polymerase. Each sample was made up with H_2O to a final volume of 50μL.

Samples were cycled under the following conditions using a PTC-200 PCR machine (MJ Research):

1. 95°C for 1 minute
2. 95°C for 50 seconds
3. 60°C for 50 seconds
4. 68°C for 1 minute per template Kb
5. Go to 2, 18 times
6. 68°C for 7 minutes
7. Cool to 37°C
8. End

The paternal plasmid strand does not contain the mutation and is *dam* methylated, rendering it vulnerable to digestion by *DpnI* endonuclease that digests methylated DNA. Digestion of the paternal strand was achieved by the addition of 1μL *DpnI* to the post-PCR reaction mix and incubation at 37°C for 1 hour.

2.3.8 *E. coli* transformation

100μl of Library efficient DH5α chemically competent cells (Invitrogen) were transformed with 10μl of plasmid according to the manufacturer's protocol. As the plasmids used in this study all contained ampicillin as a selection marker transformed cells were plated onto LB plates supplemented with 100μg/ml ampicillin and were grown overnight at 37°C. *E. coli* colonies were tested for the transformation of the correct plasmid by colony PCR, using

the technique discussed in section 2.3.10. Positive colonies were grown overnight in LB media supplemented with 100µg/ml ampicillin at 37°C and 225 rpm in a Multitron standard incubation shaker (Infors AG). Following overnight incubation 5ml of culture was purified to 50µl of EB buffer with the Plasmid miniprep kit (Qiagen).

When plasmid preparations were analysed by restriction digestion, half of the reaction volume to that stated above was used (20µl in total) and incubated at 37°C for one hour. Following this, 3µl of 6x MassRuler loading dye (Fermentas) was added and the samples were run on a 1% agarose TAE gel.

2.3.9 Yeast transformation

This protocol was used for the transformation of plasmids into *S. cerevisiae*.

1. 50ml of cell culture was grown overnight to a density of $2*10^7$ cells/ml.
2. Cells were centrifuged at 3000 rpm for 5 minutes at room temperature in an Eppendorf centrifuge 5810R. Cell pellets were washed once in 20ml of sterile water and collected again by centrifugation.
3. Cell pellets were resuspended in 15ml of lithium acetate solution (100mM lithium acetate, 10mM Tris base, 1mM EDTA, pH7.5) and left for one hour at room temperature.
4. Cells were collected again by centrifugation and resuspended in 500µl lithium acetate solution.
5. For each transformation the following was added to a 1.5ml micro-centrifuge tube:

50% Polyethlene glycol 3800, 100mM lithium acetate 10mM Tris base 1mM EDTA, pH7.5	300µl
Cells in lithium acetate solution	50- 100µl
Denatured UltraPure salmon sperm DNA solution 10mg/ml (Invitrogen)	15µl
DNA to be transformed (plasmid ~30ng,)	1-5µl

6. The solutions were incubated on a Mini Labroller rotator (Labnet) at room temperature for 30 minutes and then transferred into a water bath at 42°C for 15 minutes.
7. After the 15-minute incubation was completed the cells were placed on ice for 3 minutes.
8. 1ml of water was added to each tube, the solution was spun at 3000 rpm for 5 minutes in an Eppendorf centrifuge 5414D and the supernatant was subsequently discarded.
9. Cells were resuspended in 100µl of sterile water and plated on selective media plates in either undiluted, 1:10 diluted and 1:100 diluted solutions.
10. These plates were incubated at a 30°C for 72 hours. Visible colonies were identified and transformation was confirmed by colony PCR, DNA sequencing and qPCR

2.3.10 Colony PCR

Colony PCR was performed in both *E. coli* and *S. cerevisiae* to check for the presence of DNA.

1. Selected cell colonies were numbered on the plate for later reference. A sterile pipette tip was used to touch numbered colonies and the tip was placed into 11.5µl of water within a 0.2ml PCR tube. Tips were left for 5 minutes to allow cells to transit into solution, removed and the PCR tubes were closed and micro-waved at full power in a 800W power microwave for 2 minutes. In each tube 0.5µl of forward and reverse primer at an initial concentration of 10µM were added, followed by 12.5µl of 2x ReddyMix PCR master Mix (Thermo Scientific).
2. All reactions were vortex mixed, spun down, and run on a PTC-200 PCR machine (MJ Research) with the following programme:

1. 95.0°C for 4:00 minutes
2. 94.0°C for 0:40 minutes
3. 55.0°C for 0:50 minutes
4. 72.0°C for 1:00 minutes
5. Go to 2. +3 sec/cycle x30 times
6. 72.0°C for 10:00 minutes
7. End

3. To check for the presence of the correct PCR product 7µl of the PCR mix was loaded onto a 1% agarose TAE gel, run under the conditions discussed above.

2.3.11 The preparation of yeast genomic DNA

To obtain large amounts of high quality genomic DNA for sequencing, qPCR and immuno-slot blot experiments the following protocol was used. A maximum of 10^{10} cells can be used and yield typically 300-500 µg of genomic DNA from haploid yeast strains.

1. Cells were collected by centrifugation at 5000 rpm for 5 min and the supernatant was discarded. Cells were washed once with a 5ml Sorbitol-TE solution (comprising 0.9M sorbitol, 0.1M Tris-HCL, pH 8.0, 0.1M EDTA).
2. Cells were centrifuged again and re-suspended in 5ml of Sorbitol with the addition of 1mg/ml zymolyase 20T and 0.28M β-mercaptoethanol (Sigma). Cells were incubated for either 1 hour at 37°C in a shaking incubator, or incubated overnight at 4°C in the dark.
3. Samples were centrifuged at 2000 rpm for 5 minutes. The supernatant was discarded and the cell pellet was re- suspended in 5ml of lysis buffer/PBS 1:1(v/v) solution (see appendix 1). To each 0.3 ml of RNase A (10mg/ml in TE buffer, incubated at 95°C for 10 min, Sigma) was added. Samples were incubated at 37°C for 1 hour with occasional shaking before 0.2ml of Pronase (20mg/ml in TE buffer) was added to each sample and incubated at 37°C for 1 hour and then at 65°C for 1 hour, with occasional shaking.
4. An equal volume of phenol/chloroform 1:1 (v/v) was added to each sample. Tubes were shaken vigorously before being centrifuged at 10,000 rpm for 10 min with a JA-20 rotor. Nucleic acids in the aqueous upper phase were transferred to a new polypropylene tube with a 3 ml plastic Pasteur pipette.
5. To ensure complete deproteinisation a second extraction with phenol/chloroform and a third with chloroform/isoamyl alcohol (24:1) were performed under the same conditions as the first extraction. The absence of protein precipitate at the interphase was indicative of complete deproteinisation.
6. The aqueous phase was transferred to a new tube.
7. To each sample 2 volumes of chilled 100% ethanol (-20°C) was added. Samples were cooled at -80°C for 1 hour or overnight at – 20°C.
8. Precipitated DNA was collected by centrifugation at 10,000 rpm for 15 min at 4°C. DNA pellets were air dried before re-suspension in 1ml of 1x TE.

9. DNA was precipitated again by the addition of 1 volume chilled isopropanol (1 ml). The samples were gently shaken until the DNA became visible in the solution. The DNA precipitate was collected by centrifugation at 12000 rpm for 10 minutes at room temperature and re-suspended in 500μl 1x TE.
10. The concentration of the DNA was measured using a Nanodrop 1000 Spectrophotometer (ThermoScientific).
11. The DNA was stored at 4°C short-term or at -20°C for longer-term storage.

2.3.12 Quantitative PCR

RAD1 gene DNA content was assessed by qPCR, and compared to *RAD16* as a control gene. Samples were amplified using the CFX-connect Real-Time PCR Detection System (Bio-Rad) equipment using the iTAQ universal SYBRgreen supermix, and primers to the both genetic loci as discussed in the relevant results chapter, with the following parameters used:

1. qPCR standards were obtained by 5 sequential 10-fold dilutions of *RAD16* DNA from the wild-type SX46a strain.
2. 5μl of each DNA sample was diluted with 95μl H_2O
3. 5μl of either standard or experimental sample, was mixed with 5μl of SYBRgreen/1% primer mix (10mM concentration) for the PCR reaction, giving a final primer concentration of 500nM and a final well volume of 10μL
4. All qPCR experiments were processed in 96-well plates (Hard-Shell 96 well plates, Bio-Rad) and each sample was processed in triplicate, including for the standards. After sealing the plates, each plate was vortexed briefly and centrifuged at 2000 rpm for 2 minutes at room temperature.

2.3.12.1 qPCR programme settings

1. 95°C for 3 minutes
2. 95°C for 10 seconds
3. 55°C for 20 seconds – followed by optical image 4. Go to step 2 x 44 times
5. 95°C for 10 seconds
6. 65°C for 5 seconds
7. Melt curve from 65°C

2.3.12.2 Data analysis and calculation of IP/IN ratio

All data was analysed using Bio-Rad CFX Manager (Version 3.1) and Microsoft Excel by the following process:

1. Standards were checked to ensure quantification was appropriate, and that the qPCR reaction efficiency was between 80% and 110%.
2. Any unknown sample replicates results outside the linear range of the standards or with a threshold cycle value greater than 0.5 cycles from the remaining two replicates was discarded.
3. The DNA quantity relative to the standards was calculated, and exported to Excel.
4. The mean starting quantity from triplicate wells for each sample was used as the initial DNA quantity
5. This *RAD1* DNA quantity was normalised to the *RAD16* DNA quantity for each experimental sample.

2.4 Yeast survival and DNA damage assays

2.4.1 UV survival analysis

Cells were grown to log-phase in appropriate media, then counted and adjusted to a concentration of 10^7 cells/ml. Samples were serially diluted to a concentration of 2,000 cells/ml. To each 2% agar plate containing appropriate growth media, 200 cells (100μl volume of cell suspension) was plated for UV treatments of <40 J/m^2. For 40-80 J/m^2 doses a total of 2000 cells per plate were used and 20,000 cells per plate were used for UV doses above 80 J/m^2. For all UV doses plates were produced in triplicate.

Each plate was irradiated by exposure of between 1 to 160 J/m^2 of UV light at 254nm from a germicidal lamp at a fluence of 10W/m^2. Following UV irradiation, plates were incubated at 37°C for 48 hours. After this interval colonies were counted, and the mean percentage of surviving cells was calculated for each triplicate set of plates, and adjusted to an untreated plate average as the 100%.

2.4.2 UV damage and repair assay protocol

Yeast cells were collected by centrifugation and resuspended in chilled PBS (at 4°C) to a cell density of $2*10^7$ cells/ml. For UV treatment aliquots of 50ml of resuspended culture was poured into a 15cm diameter clear glass Pyrex dish and exposed to varying dose of UV light at 254nm from a germicidal lamp at a fluence of 10W/m^2. This procedure was repeated until the total volume of all samples had been treated. Yeast cells were subsequently centrifuged, the PBS was removed, and the cells were resuspended in liquid media. Cells were either used for immediate DNA extraction or allowed to repair for varying amounts of time before DNA was extracted.

2.4.3 Immuno-slotblot assay

For the detection of DNA adducts (either CPDs following UV irradiation and oxaliplatin-DNA adducts following oxaliplatin exposure), DNA was obtained following treatment using the protocol described in section 2.3.11.

Samples of DNA were assessed for the presence of DNA adducts using Bio-Dot SF Microfiltration equipment and antibody detection, using the protocol as follows:

1. A double layer of Bio-DOT SF (Bio-Rad) paper was placed on the equipment and overlaid by a layer of Gene Screen Plus Hybridisation transfer membrane, pre-soaked in 0.4M NaOH for 2 minutes.
2. The lid was secured tightly and the outlet was attached to a VacuGene pump (Pharmacia Biotech, UK). A pressure of 60 mbar was applied.
3. 200μL of DNA solution (comprising 200ng DNA, 1X TE buffer and NaOH at a final concentration of 0.4M) was pipetted into each well.
4. Following complete aspiration through to the membrane of the DNA solution, a further 200μL of NaOH was applied to each well and aspirated again to the membrane.
5. The membrane was removed from the equipment and blocked by overnight immersion in a 3% non-fat milk powder/1X TBST solution at 4°C overnight or room temperature for 2 hours.
6. The membrane was placed in a shallow tray and immersed in 3% milk powder/1X TBST solution, with the addition of either 1uL CP9/19 antibody for the detection of platinum adducts or 2uL CPD antibody (Kamiya Biomedical Company, Seattle Anti-Thymine Dimer Clone KTM53), used for the detection of UV damage. The membrane was immersed for 2 hours at room temperature on a platform shaker with gentle agitation.
7. The membrane was washed 3 times in 1X TBST for a total of 10 minutes each wash, before immersion for 1 hour at room temperature in 10mls 3% milk/1X TBST with the addition of 5uL alkaline-phosphatase linked

Goat anti-rat IgG secondary antibody (Sigma-Aldrich A8439) (1:10,000) on a platform shaker.

8. Following secondary antibody exposure, a further 3 washes with TBST were conducted as described previously.
9. 2mls of enhanced Chemo-Florescence (ECF) solution (GE Healthcare) was applied to the membrane at room temperature for 2 minutes. The emission of fluorescence was detected with the Typhoon TRIO Variable Mode imager (Amersham Biosciences).
10. Bands were quantified using ImageJ software (http://imagej.nih.gov/ij/)

2.4.4 Oxaliplatin treatment of yeast strains

2.4.5 Yeast growth analysis following oxaliplatin exposure

Growth curves were assessed with each yeast strain to determine the rate and manner in which the culture grows with or without the addition of oxaliplatin. Yeast cultures of 5ml volume were grown overnight to reach early stationary phase. The cultures were then diluted to OD_{600} of 0.5, and exact OD_{600} values were recorded at hourly intervals. Once cells were growing at a rate of increase in OD_{600} of 0.28 per-hour, for two successive hours, oxaliplatin was added to the media to the desired final concentration. The sequential change in OD_{600} was measured over the next 6 hours.

2.4.6 Oxaliplatin treatment in PBS solution

10mls of yeast culture were grown overnight to reach early stationary phase. The cultures were then diluted to OD_{600} 0.5, and exact OD_{600} value were recorded at hourly intervals. Once cells were growing at a rate of increase of 0.28 OD per-hour for two successive hours cells were collected by centrifugation

and resuspended in room temperature PBS, with or without the addition of oxaliplatin to the desired final concentration. Cells were incubated for three hours at 30°C at 180 rpm. Following the three-hour incubation, cells were collected by centrifugation, washed twice with growth media, and resuspended in appropriate media.

The samples were assessed for the effect of oxaliplatin in two ways:

1. To measure the effect of treatment on growth rate of cells, the OD_{600} was adjusted to 0.8 and measured hourly for the following 12 hours. At OD_{600} of 2 cells were diluted to an OD_{600} of 1 and all subsequent values were adjusted to take into account of this dilution.

2. To determine survival by the effect on colony formation after oxaliplatin exposure, samples were diluted to a concentration of 200 cells per 100μl and plated in triplicate. Plates were incubated at 37°C for 48 hours. Colonies were counted, and the mean percentage of surviving cells was calculated for each triplicate set of plates was used, and adjusted to the untreated average as 100%.

Chapter 3 The translation of a 'DIP-chip' assay, capable of measuring genome wide DNA repair, into a clinical tool to measure oxaliplatin-DNA adducts in patient blood samples

3.1 Introduction

Many studies have demonstrated that the measurement of DNA repair capacity, and in particular the measurement of platinum-DNA adduct levels, could potentially be used to predict response and toxicity from chemotherapy (for examples see section 3.1.1). The findings of these studies are consistent with the central hypothesis of this thesis - that as DNA repair pathways are responsible for the removal of adducts formed by the interaction of platinum agents with DNA, through understanding variations in the functioning of these pathways in individual patients it may be possible to stratify patients for response and toxicity from these agents, and for the risk of OIPN in particular. However, the currently available methods of NER analysis or platinum-DNA adduct measurements give an average measure of the level of platinum-DNA adducts across the whole genome, and, as a result, have limited predictive and prognostic power, as will be outlined in the following discussion (Bowden 2014).

Our laboratory has made several advances in DNA repair assay technology that have resulted in the development of an assay capable of measuring DNA damage and DNA repair capacity at high resolution at a genomic scale (Teng et al. 2010). This DIP-chip technology, based an a combination of DNA immunoprecipitation (DIP) and microarray analysis (chip), has more recently been translated for use in human tissue for the detection of platinum-DNA adducts in cultured cells previously exposed to cisplatin or oxaliplatin (Powell 2014).

It may be possible to use this technique in patient samples to produce high-resolution, genomic scale patterns of DNA damage and repair, in the context of oxaliplatin-induced DNA damage. This approach has the potential to improve the sensitivity and specificity of DNA damage and repair analysis compared to

currently available methods, with the novel ability to detect subtle differences in platinum-DNA adduct patterns at key regions of the genome. It has the potential to be a potent tool for patient stratification.

Several challenges need to be overcome in the translation of this research tool into a clinically usable assay, including adapting the method for use on appropriate clinical samples, ensuring the reliability and reproducibility of the assay, the development of bioinformatic tools and robust pipelines for data analysis, and clinical validation. These steps must be addressed before the assay can be confidently used on precious and limited clinical samples obtained from patients. It is these challenges that will be addressed in this and the following chapters.

Following a discussion of the previous technologies available, the evidence supporting the use of this type of approach to patient stratification and a discussion of the development of the DIP-chip technique, the main focus of this chapter will be on the translation and validation of the DNA immunoprecipitation (DIP) stage of the assay into a tool for reliably and reproducibly measuring levels of platinum-DNA adducts in *ex vivo* oxaliplatin treated human blood samples.

3.1.1 An overview of platinum-DNA adduct studies

Since the introduction into routine clinical practice of cisplatin in the 1970s, and of oxaliplatin in the last decade, the potential benefits of accurate patient stratification for response and toxicity to these agents has been recognised and studied. As early as the 1980s attempts to predict clinical outcome by measuring platinum-DNA adduct levels commenced (Fichtinger-Schepman et al. 1987; Poirier et al. 1982). Initial evidence from these studies is contradictory with conflicting outcomes (Ma et al. 1995), but as the technology available to measure adducts has developed these assays have become more sophisticated and more accurate (Nel et al. 2013; Wang et al. 2011). Broadly, the methods used can be classified as antibody based, radio-nucleotide approaches, analytical chemistry techniques, or can be grouped as functional studies.

Of note, all of the technologies available measure a broad output of total platinum-DNA adducts per unit of DNA, or result in a crude measure of global DNA adduct repair. Until the development of our DIP-chip approach, these techniques have been unable to investigate the subtleties of platinum-DNA adduct patterns at high resolution throughout the genome, with a resulting lack of sensitivity that has hindered progression to a workable clinical assay. It is this key difference that we aim to exploit with our technique.

3.1.2 Antibody based assays

Several antibodies have been raised against platinum-DNA adducts and have been applied in a variety of studies. These typically use enzyme-linked immunosorbent assays (ELISA) and immunohistochemical based techniques.

Examples include a study by Poirier, with the development rabbit anti-serum raised against cisplatin treated calf thymus DNA (Poirier et al. 1982). The resulting antibody was capable of detecting platinum-DNA adducts with high specificity for platinum modified DNA in ELISA assays (Poirier 1982). In a similar study by Fischtinger-Schepman, the group immunised rabbits with cis-Pt(pGpG) or cis-Pt(ApG) to produce an antibody against these specific epitopes (Fichtinger-Schepman et al. 1987). The resulting antibody was able to recognise adducts after ELISA, nuclease digestion of isolated DNA and chromatographic isolation of platinum-dinucleotides, and was tested and able to detect platinum DNA-adducts in DNA extracted from blood samples from 7 patients undergoing cisplatin chemotherapy (Fichtinger-Schepman et al. 1987).

There are several other examples of antibodies raised to platinum treated DNA, including platinum-DNA adduct antibodies developed by Tilby (Tilby et al. 1991), Chao (Chao et al. 1994) and Sundquist (Sundquist et al. 1987). Other examples of anti-platinum antibodies developed for the measurement of platinum-GG adducts are the antibodies Mab R-C18 and Pt-(ApG) R-B3 developed by Liedert (Liedert et al. 2006).

Only one antibody is commercially available and is the antibody used throughout this study, CP 9/19. This antibody was developed by Tilby and used

initially for ELISA studies (Tilby 1999). The epitope was elucidated and confirmed to be the predominant intrastrand platinum-GG adduct by demonstrating the preference for binding of the antibody CP9/19 to guanine-guanine intrastrand crosslinks (Meczes et al. 2005).

Platinum-antibody ELISA based assays have been used in several clinical studies. Examples include an ELISA assay used on samples from 55 ovarian cancer patients undergoing platinum chemotherapy. This study demonstrated a positive correlation between treatment response and adduct levels in leukocyte DNA (Reed et al. 1987). Another early example of a similar clinical study used an ELISA based on an antibody raised to platinum drugs to measure platinum-DNA adducts in peripheral blood leukocytes in several patients with either testicular or ovarian cancer undergoing chemotherapy with carboplatin or cisplatin. This study demonstrated a positive correlation between adduct levels and response to the treatment (Reed et al. 1987, 1988). Several studies have been conducted using a similar approach in a wide variety of clinical samples also showed a positive correlation, including germ cell tumours (Motzer et al. 1994), ovarian cancer (Reed et al. 1990), breast cancers (Gupta-Burt et al. 1993) and a study an unselected wide variety of cancers (Reed et al. 1993). However, other studies using ELISA based techniques failed to show any significant correlation between platinum-DNA adduct levels and response to treatment (e.g. (Fichtinger-Schepman et al. 1987)).

When these early studies were reviewed (Ma et al. 1995) the variable and contradictory outcomes were attributed to differences in sample preparation technique and analysis methods, occasionally resulting in greater than 10 fold differences in adduct levels recorded between similarly treated patients.

Immunohistochemical assays are an alternative antibody-based assay approach, and can be used to directly visualise cellular and nuclear adduct levels. In one example, using antibodies developed by Terheggen ((Terheggen et al. 1991)), buccal cells from patients undergoing platinum chemotherapy for a variety of tumour types were assessed. This study described a positive correlation between adduct levels and response to treatment with carboplatin,

but not with cisplatin (Blommaert et al. 1993). Similar experiments have been conducted in patients with non-small cell lung cancer (NSCLC) receiving cisplatin and carboplatin, suggesting adduct levels may be prognostic in these patients (van de Vaart et al. 2000).

In a contemporary example of the approach, levels of cisplatin adducts were measured by an immunocytochemistry assay of platinum-(GpG) adducts in nuclear DNA of circulating tumour cells treated *ex vivo*, and compared to ERCC1 mRNA expression profiles (Nel et al. 2013). The formation and repair of adducts was assessed over a 24-hour period in a small number of patients with NSCLC undergoing cisplatin chemotherapy. Using multivariate regression analysis cisplatin AUC and maximum detectible adduct levels correlated with a positive response to treatment when compared to patients with stable disease. There was no significant correlation of with ERCC1 expression.

3.1.3 Radio-nucleotide assays

The ^{32}P-post-labelling assay (PLA) was developed by Blommaert (Blommaert and Saris 1995) and subsequently modified by Welters (Welters et al. 1997) and Pluim (Pluim et al. 1999). This assay has been shown to be up to 10 fold more sensitive than ELISA based DNA-adduct assays, and is capable of detecting in the range of one adduct per 10^7 to 10^8 nucleotides (Welters et al. 1997). Advantageously, the PLA requires sample processing steps that render it incompatible with the measurement of platinum-DNA mono-adducts, reducing the risk of cross-specificity. Briefly, DNA is digested and platinum-DNA adducts are separated by cation exchange chromatography, due to positive charge of the platinum ion. After removal of platinum, remaining dinucleotides are labelled with ^{32}P and separated by chromatography, before scintillation counting for quantification (Pluim et al. 1999).

Three clinical studies have used the PLA to analyse platinum-DNA adduct levels and relationship between adducts and the patient response to chemotherapy. In one example, fresh samples from eight testicular cancer patients and head and neck patients were treated *ex vivo* with cisplatin. Higher

adduct levels in *ex vivo* treated biopsy specimens detected by PLA correlated with better clinical response (Welters et al. 1999b). Hoebers (Hoebers et al. 2008; Hoebers et al. 2006) examined adduct levels in 35 patients treated with cisplatin as part of chemo-radiotherapy treatment for head an neck squamous cell cancer, finding a correlation between disease free survival and the platinum-GG adduct levels in the primary tumour and a trend, but no statistical significance, for overall survival. In a further study using PLA, adducts were determined in 63 patient from normal tissue (peripheral blood mononuclear cell (PBMC) and buccal cells) and from 23 biopsies in cisplatin treated patients undergoing chemo-radiation. No correlation was demonstrable between tumour biopsies and normal tissues (Hoebers et al. 2008).

The advantage of PLA is a high sensitivity for detecting the presence of adducts, and that it requires only minimal DNA samples for processing. However, it is labour intensive and requires radioactive reagents, making it somewhat impractical to develop for routine clinical use (Phillips et al. 2000; Zayed et al. 2011).

3.1.4 Gel based platinum-DNA adduct assays

A gel-based assay for the detection of platinum-DNA adducts has been developed, described as the TARDIS assay, an acronym for trapped in agarose DNA immunostaining assay (Cowell et al. 2011). This technique traps a single cell in agarose, then detects adducts using a primary anti-platinum antibody. So far no clinical studies have used this research technique. A modification of the comet assay (discussed in chapter 1 section 1.5.2) has been used on peripheral blood lymphocytes samples from colorectal cancer patients treated with oxaliplatin to measure oxaliplatin-DNA adducts, and some evidence of inter-individual variability could be determined with this approach (Almeida et al. 2006). However, this assay is generally considered qualitative as it does not directly provide information on adduct number (Zayed et al. 2011). It is possible that further modifications of this approach may result in a clinically useful tool (McKenna et al. 2008).

3.1.5 Analytical-chemistry based assays

Broadly, several methods can be considered as a group of analytical chemistry based approaches to determine platinum-DNA adduct levels. Methods include mass spectrometry, which can be used to measure platinum levels per unit of DNA. These approaches are highly sensitive, however they require expensive and specialised equipment limiting the potential clinical utility (Phillips et al. 2000). Variations on this technique include Atomic Absorption Spectroscopy (AAS) (Kloft et al. 1999), LC-ion trap Mass Spectroscopy (Le Pla et al. 2007), Accelerator mass spectrometry (AMS) combined with HPLC purification (Hah et al. 2007) and ICP-MC (Bonetti et al. 1996; Zayed et al. 2011). Atomic absorption spectroscopy measures total platinum, but lacks specificity for routine clinical applications (Henderson et al. 2011). Inductively coupled mass spectrometry has higher sensitivity that could be useful for clinical applications, although results in a general measure of total platinum, not the type or location of adducts (Henderson et al. 2011).

Several clinical studies have used these approaches. The earliest example used AAS to measure platinum ions per unit of DNA extracted from PBMC, with no clinical correlation evident in this study (Bonetti et al. 1996). The measurement of adduct levels using these types of analysis techniques has been proposed as a method of adjusting treatment in several tumour types, including head and neck cancer (Schellens et al. 1996) and non-small cell lung cancer. A study of oxaliplatin response and toxicity used adsorptive stripping voltammetry, a variant of mass spectrometry, to measure oxaliplatin adducts per unit of extracted DNA in a cohort of 27 patients with a variety of solid tumours all treated with oxaliplatin. With increasing adduct levels there was a positive, but not statistically significant, increase in tumour response and toxicity (an OIPN based neuropathy score). Given the low number of patients and heterogeneous tumour types studied, the lack of statistical power is unsurprising (Pieck et al. 2008).

Again the clinical results of analytic chemistry based techniques have been contradictory, especially when measuring adducts generated in surrogate tissues of peripheral blood cells, in germ cell tumours and ovarian cancer. The

differences in outcome reported have been explained by the wide variety of tumour types, clinical samples, investigative tools and preparation methods used in these studies (Bonetti et al. 1996; Motzer et al. 1994; Pieck et al. 2008; Schellens et al. 1996).

3.1.6 Recent functional studies showing a relationship between DNA repair capacity and clinical response to platinum chemotherapy

Rather than measure total adduct levels an alternative approach is to measure functional DNA repair capacity for platinum adducts. Wang used a functional assay in an example of this approach (Wang et al. 2011). PBMC from 591 patients with NSCLC were stored pre-cisplatin chemotherapy. The cells were subsequently cultured and assessed for DNA repair characteristics using a host reactivation assay with benzo[a]pyrine diol epoxide (BPDE) damaged non-replicating recombinant plasmid with a chloramphenicol acetyltransferase (CAT) reporter gene. Higher DNA repair capacity resulted in higher expression of CAT, detected by the radioactivity of products produced from the addition of ^{3}H-radiolabelled acetyl coenzyme A and chloramphenicol. Lower germline DNA repair capacity measured with this assay was associated with improved survival following cisplatin treatment, indicating that DNA repair capacity of peripherally harvested PBMC may be used as a surrogate biomarker for platinum response in tumours.

3.1.7 Surrogate tissue versus tumour tissue for clinical sample analysis

Measuring platinum-DNA levels in tumour tissue is technically difficult and, for a true reflection of treatment effect, requires the extraction of tumour after treatment with inherent risks of additional tumour sampling in the post-chemotherapy period. An alternative would be an additional biopsy pre-treatment to allow adequate cellular material to be gathered and treated *ex vivo*, although the link with the specific tumour microenvironment and treatment conditions is then disturbed. As an indication of the risk of repeat patient sampling in a recent study of image-guided biopsies conducted for research purposes in clinical studies the rate of complications was 5.2% generally, and 17.1% in intra-thoracic and abdomino-pelvic solid organ biopsies (Overman et al. 2013). Particularly during assay development projects these concerns limit the available material for research purposes and the limit the practicality of this approach.

Surrogate tissue, such as blood cells, have significant advantages in reducing these risks. However, there are obvious concerns with using tissue other than the tumour, including that platinum uptake and DNA adduct processing may be different among the different haematopoietic cells commonly used as surrogate tissue (Bracker et al. 2006) and that platinum-DNA adduct levels may be lower at physiological doses in normal tissues than in tumours (Liedert et al. 2006). When using normal tissues as a marker of germline repair capacity or to examine the response of normal tissues, for example in a study of toxicity, these issues are less of a concern.

One promising contemporary approach is using circulating tumour cells (CTC) as a more accurate surrogate of tumour characteristics. A recent study used circulating tumour cells treated *ex vivo* with cisplatin to compare platinum-DNA adduct levels detected with immunocytochemistry with tumour response (Nel et al. 2013). This approach obviates the need for collecting tumour tissue and may be a more reliable surrogate marker. However, in about 1/3 of patients with metastatic cancer CTC are undetectable, most patients have small numbers of CTC, and a count of >100 per 7.5ml of blood is considered extreme and is a

very poor prognostic marker (Alemar and Schuur 2013). Evidently the yield of CTC from a clinical blood sample, and hence usable DNA, is very low.

In many of the currently available assay techniques and in the clinical studies discussed above the use of blood samples and PBMC as a surrogate tissue attests to fact that this is the best and most practical surrogate patient tissue available.

3.1.8 Potential improvements using a high-resolution genome-wide approach

Compared to markers of DNA repair or NER protein levels, in the context of platinum-DNA adduct clinical studies the measurement of total platinum content of cells has been suggested as a more relevant, and hence more reliable, marker of platinum effectiveness and toxicity (Henderson et al. 2011; Kim et al. 2012), since primarily the level of platinum damage represents the effective stimulus leading to downstream effects of apoptosis and cell death.

The studies discussed in this section demonstrate that the approach of measuring platinum-DNA adduct levels, of studying DNA repair in the germline of patients treated with platinum-based chemotherapy, and the issues surrounding the use of tumour, surrogate or CTC tissues are an active area of on-going research, driven by the need to improve the therapeutic profile of the platinum agents. Unfortunately, the results so far are conflicting, and lack robust predictive power for response or toxicity. To-date these remain research tools and no technique or assay has made an impact in the clinic (Bowden 2014).

There is a wealth of genetic data such as genome wide association studies (GWAS) for platinum response and toxicity, as extensively discussed in chapter 6, that suggest that small changes in DNA damage levels or repair capacity at different genes and sections of the genome could be important predictors of response and toxicity. Tools to capture the intricacies of genome-wide DNA damage and repair are potentially more sophisticated ways of characterising the effects of platinum-DNA adducts, and may be more sensitive and specific than a global measures of 'whole genome platinum per unit of DNA' available in the

studies outlined in this chapter. Differential damage and repair in varied sections of the genome, particularly evident in post-mitotic tissues where GG-NER is down regulated whilst TC-NER is not (Nouspikel 2009), are potential mechanisms through which idiosyncratic patient responses to treatment occur, and are not detectable with currently available approaches to DNA adduct measurement. High resolution genome-wide platinum-DNA adduct patterns are detectable in cell culture models with DIP-chip assay technology recently developed in our laboratory (Powell 2014; Teng et al. 2010) and may offer a method to resolve some of these issues.

3.1.9 The development of the DIP-chip assay

The DIP-chip assay used in the experiments discussed in this, and following, chapters involve two main steps, DNA immunoprecipitation (DIP) and microarray (chip) analysis to obtain high-resolution genome scale platinum-DNA adduct profiles. The assay is a development of chromatin immunoprecipitation and microarray (ChIP-chip), a technique that first gained common acceptance as a research tool almost 15 years ago (Iyer et al. 2001; Ren et al. 2000). The ChIP-chip experimental approach has proved useful primarily for deciphering the complexities of transcriptional regulation in the context of chromatin (Lee et al. 2002) and epigenetic regulation through the modification of histones (Pokholok et al. 2005). The technique has been extensively used with *S. cerevisiae* to study the genome-wide location of transcription factor binding sites (Iyer et al. 2001; Lee et al. 2002; Ren et al. 2000) and histone modifications (Kurdistani et al. 2002; Robyr et al. 2002), and in human cells for similar types of studies (Kim et al. 2005; Weinmann et al. 2002; Wells et al. 2003). In the clinical context, ChIP-chip has been used to identify novel disease markers and potential targets in a number of conditions, including cancer (Ordway et al. 2006), cardiac disease (Movassagh et al. 2010) and psychiatric disorders (Tsankova et al. 2007).

3.1.9.1 An overview of ChIP-chip experimental design

The basic design of a ChIP-chip experiment is maintained throughout the applications discussed above. Broadly, cells cultured under experimental conditions are exposed to formaldehyde to fix the amino groups of proteins with the nitrogen atoms of DNA (or other proteins) by crosslinking. Chromatin is extracted and fragmented through sonication or enzymatic means, typically to between 200 to 1000 nucleotides in length. Fragments of chromatin containing the cross-linked protein, or target of interest, are immunoprecipitated by incubation with an antibody to the target (usually a protein or experimentally designed epitope-tagged protein if no appropriate primary antibody is available), and antibody bound fragments are separated from the undesired fraction with an appropriate secondary antibody (e.g. Invitrogen Dynabead system).

To improve the accuracy of the assay results, several methods of internal control and data normalisation have been used (Adriaens et al. 2012). In many experiments, and in our assay DIP-chip, an input (IN) sample is processed alongside each immunoprecipitated (IP) sample. An IN sample is an aliquot of the purified, treated genomic DNA that has been taken from the starting material prior immunoprecipitation. It is subject to all of the steps in the assay except the immunoprecipitation reaction. As such, it represents both the background variation in DNA quantity and initial fragment frequency, and acts as a control for non-biological variability in the assay technique, including in sample handling and differential labelling and PCR amplification efficiency (Adriaens et al. 2012). During quantification of fragments, either by qPCR or microarray, the amount of fragments at each analysed locus in the immunoprecipitated (IP) sample is divided by the number of fragments at that locus in the paired control IN sample, to account for these potential sources of sample variability.

Following reversal of crosslinking, proteinase treatment and purification (resulting in naked DNA), the relative proportion of DNA fragments in the IP and IN samples can be calculated by quantitative PCR (qPCR). This gives an IP/IN ratio - the relative amount of DNA in both samples at the genetic region immunoprecipitated as determined by the qPCR primers used in the reaction.

By comparing the efficiency of immunoprecipitation, using the IP/IN ratio, between different conditions, the relative effect of each condition compared to a reference sample (typically an untreated sample) can be observed.

This phase of the assay is useful in its own right, and has been used in experiments as a ChIP-qPCR assay (e.g. (Hecht et al. 1996)). It is, however, limited in power by the requirement to use one or a small number of pairs of PCR primers to generate information at a limited number of genetic loci. By combining ChIP experiments with microarrays (and more recently with next generation sequencing technology (Marinov et al. 2014; Park 2009)) it is possible to analyse the relationship between IP and IN samples at thousands, or millions, of genetic loci simultaneously.

DNA microarrays consist of short oligonucleotide fragments (probes) printed onto a glass slide, grouped into 'features' –lithographically printed spots of DNA on the slide of identical oligonucleotide probes. The application of differentially fluorescently labelled DNA (typically the IP and IN samples) results in competitive hybridisation of these to complementary probes on the array. The relative quantity of IP and IN samples at each probe can be calculated by the relative fluorescence at each feature. With appropriate bioinformatic processing it is then possible to produce an overview of the relative levels of immunoprecipitation efficiency, which is a reflection of the condition of interest, along the genome at high-resolution (Figure 3.1).

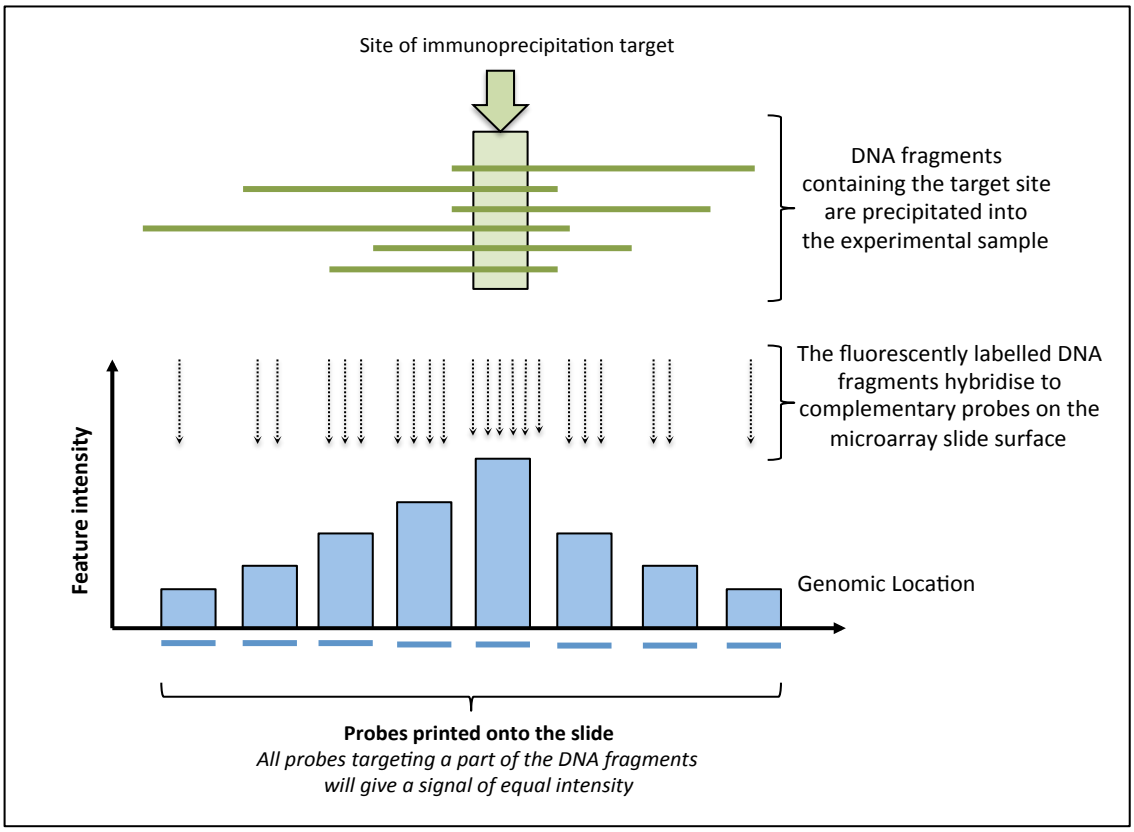


Figure 3.1: An overview of DNA microarray approach to immunoprecipitation coupled with genome-wide analysis.
Hybridisation of fluorescently labelled DNA fragments containing the immunoprecipitated target of interest to oligonucleotide probes on a microarray results in relative changes in fluorescence detected on the microarray surface. After bioinformatic processing the changes in fluorescence can be converted to calculate the location of the site of interest, at high resolution and on a genomic scale (adapted from (Adriaens et al. 2012)).

3.1.9.2 Modifications to ChIP-chip

By altering of the antibody used in the immunoprecipitation phase of the experiment a wide variety of targets can be detected. The assay has been used to detect the genomic location of a host of transcription factors and histone modifications (discussed above) and in a wide variety of species, including yeast (Gasmi et al. 2014), human (Li et al. 2014), bacteria (Rajeev et al. 2014) and mice (Ferrer-Vicens et al. 2014). With the use of specific antibodies it is possible to modify the technique to perform genome-wide analysis on histone methylation in human cells (Kurdistani et al. 2002; Robyr et al. 2002) or to immunoprecipitate methylated DNA (the meDIP-chip technique) to explore mechanisms of transcriptional silencing (Movassagh et al. 2010).

The advent of improved sequencing technology over the past decade has lead to the adoption of DNA sequencing as a viable alternative method for genome-wide analysis, in comparison to microarrays (Johnson et al. 2007; Park 2009). ChIP followed by sequencing (ChIP-seq) offers certain advantages over ChIP-chip, including having no requirement to use organism specific microarrays with a spatially limited number of potential probes, and a reduction in the amount of starting material required following immunoprecipitation. ChIP-seq also has the potential advantage of improved resolution and dynamic range, and an improved signal-to noise ratio compared to ChIP-chip, resulting in increased sensitivity and specificity for binding-site detection. As DNA sequencing technology evolves these advantages are leading to rapid adoption of this technology as an alternative to ChIP-chip (Ho et al. 2011).

Given the advantages of sequencing techniques, the continued use of microarrays for the genomic component of our DIP-chip assay could be questioned. However, there are significant advantages to working with microarrays, particularly in assay development work. These advantages include that the technique is less expensive, is widely available, is robust, is well established with a clear experimental workflow, and the bioinformatic methodologies used when interpreting the large datasets generated have been a studied for over a decade. Whilst sequencing is becoming the optimal technique to analyse ChIP samples, the current experimental and discovery phase of the platform, and the experience our group has in the microarray field, make DIP-chip a more attractive platform for the development of this assay at the current time.

3.1.9.3 The DIP-chip assay

The DIP-chip assay used in this thesis was developed in our laboratory as a modification of ChIP-chip technology, initially through adjustment to detect UV radiation induced CPDs in *S. cerevisiae* with the use of an anti-CPD antibody (Teng et al. 2010). Recently, we have made a further modification to use an anti-

platinum modified DNA antibody to immunoprecipitate cisplatin-induced damage, first in yeast and then in human cell cultures (Powell 2014). An overview of the technique is shown in figure 3.2.

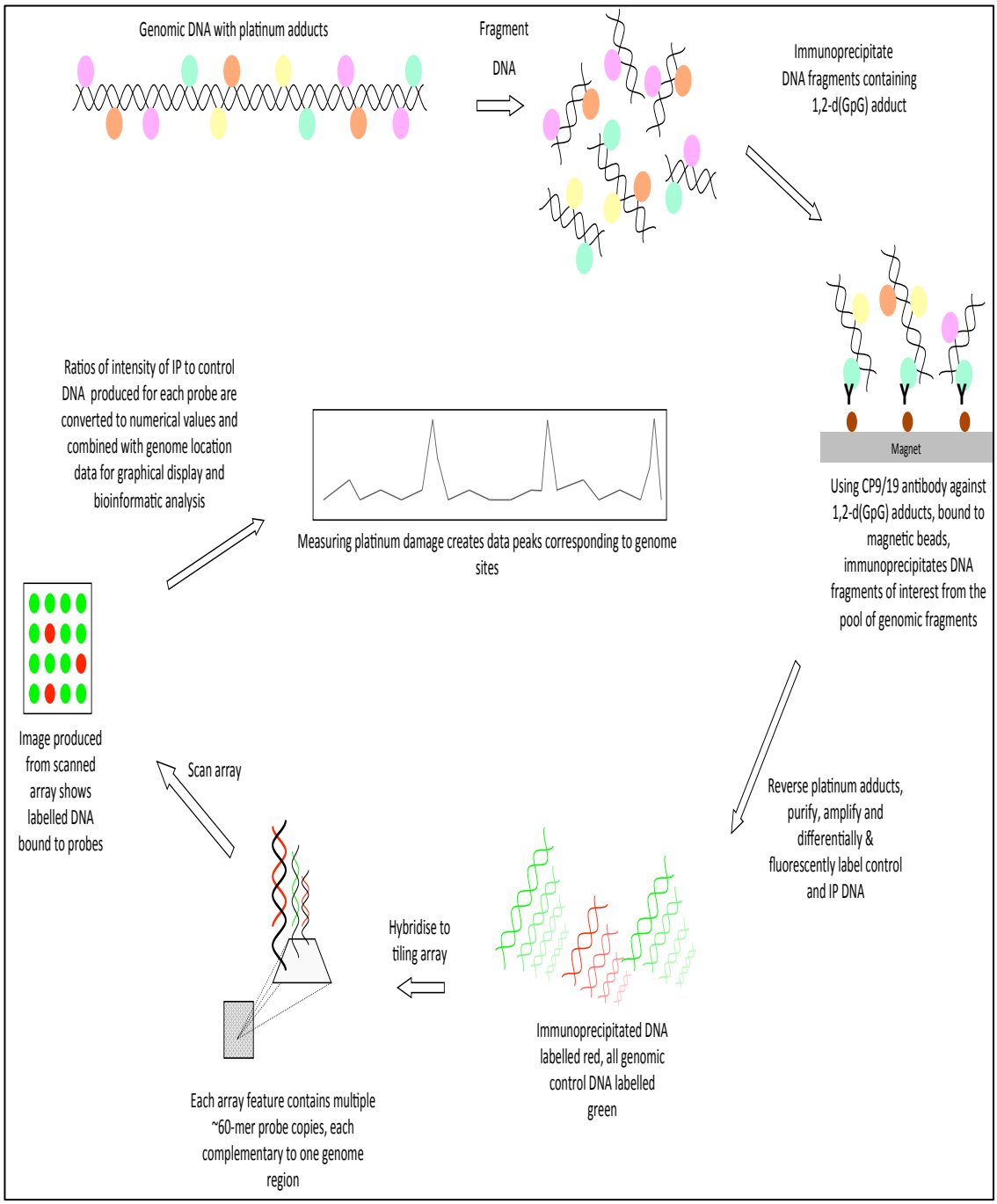


Figure 3.2: The DIP-chip experimental pathway.
Upper left – genomic DNA with platinum-induced DNA adducts following ex vivo exposure of cells to platinum drugs. DNA is sonicated and immunoprecipitated with CP9/19 anti-platinum adduct antibody. Antibody bound fragments are separated with magnetic beads (Invitrogen Dynabead system). Following purification the immunoprecipitated (IP) sample and an input (IN) control sample are amplified, fluorescently labelled, and hybridised to a custom microarray prior to data extraction and bioinformatic processing to generate a high resolution genome scale damage pattern spectrum.

3.1.9.4 The development of the DIP-chip assay for the detection of platinum-DNA adducts

When a novel DNA adduct is detected using an alternative antibody several adjustments to the experimental pathway are required to generate reliable results. The formation of platinum-induced DNA adducts results in conformational changes to the DNA. In this context, compared to protein binding experiments, the requirement for crosslinking of chromatin is removed, and naked DNA can be used during the immunoprecipitation – hence DIP-chip rather than ChIP-chip. Additionally, the presence of platinum-DNA adducts impairs the function of DNA polymerases, and this must be accounted for in the assay protocol. For example, modifications to remove platinum adducts prior to qPCR are required, otherwise a range of PCR efficiency between fragments may occur dependent on the levels of DNA adducts present on each fragment. This could result in different amplification rates during PCR, and as a consequence lead to inaccurate quantification of the DNA concentration of differentially treated samples. The amplification process prior to microarrays is also modified, and will be discussed in detail in the next two chapters.

In this chapter the translation of the DIP phase of the assay from a research tool used in cell cultures into a reliable technique for use in human clinical samples will be discussed. The samples used in the experiments outlined in this chapter will be analysed by qPCR. The following chapters, 4 and 5, will discuss the microarray 'chip' phase of the translation of this tool. With this in mind, two other pieces of information are essential for consideration in this chapter; what is the exact target of the immunoprecipitation reaction, and at which genetic locus is the IP/IN ratio to be calculated by qPCR. In the experiments presented in this thesis the anti-platinum antibody is CP9/19 and the qPCR primers used are from the 28S genetic loci. The details of both will be discussed below.

3.1.9.5 The CP9/19 Antibody

The CP9/19 antibody, initially developed by Tilby has been used to reliably detect adducts from cisplatin and carboplatin modified DNA by ELISA (Tilby et al. 1991). The antibody is sensitive, regardless of whether treatment was *in vivo* or *in vitro*, with low background binding to untreated DNA, and with demonstrated specificity for carboplatin induced DNA adducts in addition to cisplatin adducts to which it was raised. Further experiments in our laboratory have demonstrated that the antibody also detects oxaliplatin induced DNA damage (Powell 2014).

The exact epitope recognised was initially uncertain and was not addressed in the original publication, although the poor antibody recognition to platinum-DNA adducts formed by transplatin (which fails to form the 1,2-d(GpG) intrastrand crosslink) indicated that this lesion is probably a major target. Further evidence to support this was elucidated using a set of restricted oligonucleotide fragments as targets in a competitive ELISA (Meczes et al. 2005). These results demonstrated that the antibody acts by recognition of crosslinks between adjacent guanines.

Employed for ELISA based assays predominantly, CP9/19 has been used to detect platinum-DNA adducts in ovarian cancer cell lines (Mellish et al. 1995), rat thymocyte (Evans et al. 1994), Chinese Hamster Ovary cell lines (De Silva et al. 2002) and in clinical samples of peripheral blood mononuclear cells (Ghazal-Aswad et al. 1999; Peng et al. 1997; Veal et al. 2001). All of these experiments were conducted prior to the development and clinical use of oxaliplatin. As adducts formed by cisplatin and oxaliplatin are the very similar (chapter 1.2.9) the antibody should be applicable for the detection of oxaliplatin DNA damage. This has been confirmed by experiments conducted in our laboratory (Powell 2014).

3.1.9.6 The *28S* locus

Platinum-induced DNA damage is predominantly specific for small dinucleotide sequences, but there is no evidence of a bias for different parts of the genome (see chapter 1 section 1.2.5 - 1.2.6). With this in mind, the selection of the locus for qPCR assessment of fragments enriched by immunoprecipitation is to some extent arbitrary, although the genome-wide approaches used in the subsequent chapters may provide greater information on optimum sites for qPCR fragment quantification. The qPCR locus chosen in this study is that used in the initial development work in our laboratory (Powell 2014) and is the 28S locus, the ribosomal *28s* gene, present in 5 regions on chromosome 13, 14, 15, 21 and 22.

There are advantages to using qPCR primers to a gene with multiple genetic loci, as opposed to primers at a single site. There is the potential for confounding if analysing at a single qPCR locus, from large-scale structural variants, such as deletions, translocation and chromosomal breakages, or from the millions of insertions and deletions ranging from 1 to 10,000 nucleotides occur throughout the human genome (Mills et al. 2011) that could affect the ability of the assay to detect platinum-DNA adducts in different individuals. The absence, or low concentration, of DNA in a DIP-chip sample with the use of a single unique PCR locus may reflect natural genetic variability, rather than differences in platinum-DNA adduct levels between individuals. The multi-locus nature of the ribosomal 28S gene offers a degree of redundancy to counter this, whilst retaining the same sequence characteristics, and hence likely platinum-DNA adduct levels, in the same individual at each specific 28S locus.

3.2 Aims of this chapter

The DIP-chip assay developed in our laboratory has been used to detect patterns of platinum-DNA adduct formation in human cell culture conditions. The assay may be of benefit as a predictive tool for the response and toxicity of platinum agents, as discussed above.

For use in human clinical samples the assay has to be translated to detect adducts on patient samples. It can be used initially as a DIP-qPCR assay to measure adduct levels at the *28S* genetic locus, analogous to a functional approach to platinum-DNA adduct measurements similar to the study on NSCLC patients conducted by Wang (Wang et al. 2011), as discussed in section 3.1.6. If a consistent DIP-qPCR assay output can be obtained showing reproducible measurement of induced platinum-DNA adducts using qPCR, then this would provide confidence that these samples might be used reliably for patient stratification and downstream genomic approaches to measure damage at high resolution in multiple loci using DIP-chip approaches, where the signatures of the response to the drug may be detected.

The experiments described in this chapter will outline the experimental approaches and evidence obtained to translate the methods used in yeast and in human cell cultures into a method applicable to clinical samples in the form of human fresh blood samples, using PBMC as a surrogate tissue to measure *ex vivo* generated adducts to oxaliplatin.

The experiments outlined below are used to generate consistent samples to use in a DIP-qPCR assay and allow further development using genomic technologies outlined in the following chapters. The first stage of this aspect of the project is to demonstrate the ability to affinity-capture oxaliplatin-DNA adducts in *ex vivo* treated PBMC, as confirmed by demonstrating an increasing ratio of DNA fragments between untreated and treated samples using qPCR.

3.2.1 Methods

The methods used in the experiments in this chapter are briefly discussed in each section. For full details of the methods used see chapter 2. Note that the DIP-chip protocol in chapter 2 is the final version, incorporating the protocol modifications and amendments discussed in this chapter.

3.2.2 An overview of the initial DIP-chip workflow

For context, when considering the protocol amendments examined and described in this chapter, the initial DIP-chip workflow developed to detect platinum-DNA adduct patterns in human cell cultures will be described here. It is this protocol that is modified in the experiments described below.

The cells used for DIP-chip human cell culture experiments were AG16409 human dermal fibroblast cells, grown in DMEM media containing 10% foetal calf serum, L-glutamine and 1% penicillin and streptomycin and incubated at 37°C and 5% CO_2. Prior to treatment, cells were grown to 80% confluence on tissue culture plates, media was removed and DMEM was added (without foetal calf serum, l-glutamine or penicillin/streptomycin). Oxaliplatin and cisplatin were added to these cultures to a final concentration of 2.5mM. After 4 hours incubation, cells were treated with trypsin for 30 minutes before the media (containing all cellular material) was transferred to a 15ml sterile tube and centrifuged. The media, containing cytotoxic waste, was removed and safely disposed of. The cell pellet was washed twice in PBS before a final re-suspension in 200μL PBS prior to DNA extraction.

DNA was extracted using the Qiagen DNeasy Blood and Tissue Kit using the manufacturers standard protocol, into a final elution volume of 200μL buffer. Samples were sonicated at 4°C using a Bioruptor Sonicator (Diagenode) for 24 cycles of 30 seconds off/30 seconds on high power setting, resulting in an average fragment length of 200bp.

A total of 6μg sonicated DNA was used as the starting material for the DIP phase of the assay. A mixture of 1.5μg CP9/19 (Abcam) antibody and 40μl sheep anti-rat Dynabeads (Invitrogen) per experimental sample were pre-incubated for 30 minutes, before combining with the 6μg sonicated of DNA per sample. After a 3-hour incubation, the bead mixture was washed several times to remove unbound, or weakly bound, DNA fragments. The final pool of immunoprecipitated DNA fragments were eluted from the antibody/Dynabead sample and treated overnight with pronase, and subsequently with RNase, to remove any residual protein or RNA contamination. This immunoprecipitated sample is the IP sample. An internal control input ('IN') sample containing 600ng of sonicated DNA is treated overnight with pronase, then RNase A, and is processed in parallel during the remainder of the DIP-chip assay

To remove platinum adducts from the DNA fragments, the IP and IN sample were incubated at 65°C for 2 hours in a final concentration of 0.2mM NaCN. The DNA samples were purified by phenol/chloroform purification and ethanol precipitation before a final suspension of the IP and IN fragment pools of DNA fragments in 15μl PCR grade H2O.

DNA in the post immunoprecipitation sample was quantified by qPCR to determine the DNA concentration in paired IP and IN samples between experimental conditions, using a CFX-connect Real-Time PCR Detection System (Bio-Rad), iTAQ universal SYBRgreen supermix, and primers for the 28S genetic locus. The DNA amount in the paired IP and IN samples are used to calculate the IP/IN ratio for each pair of samples - the proportion of DNA fragments that have been immunoprecipitated in each experimental condition, as an indication of the relative adduct levels generated in the DNA for each treatment level. The qPCR result is primarily used as a quality control step for the DIP stage of the assay. It can be used to demonstrate successful immunoprecipitation by comparing the IP/IN ratio in treated and untreated samples, and the IP/IN ratio can be compared between identical samples to give an indication of the assay reproducibility.

Following quantification of DNA amounts in the IP and IN samples at the 28S loci, the samples are prepared for microarray analysis – the 'chip' phase of the assay. The paired IP and IN samples are amplified to generate adequate DNA for microarray hybridisation (a minimum of 500ng). This is performed using whole genome PCR-based amplification (WGA-PCR) using the WGA2 kit (Sigma) using 16 cycles of PCR, and otherwise conducted according to the manufacturers protocol. The amplified fragments are labelled with a different fluorophore in the IN and IP DNA fragments using the BioPrime Total Genomic Labelling System (Invitrogen) following the manufacturers standard protocol. Following this, the labelled DNA fragments are hybridised to a custom DNA microarray (Agilent Technologies G4497A) using the Agilent Technologies aGCH protocol. The DNA fragments hybridise to complementary DNA probes on the microarray slide surface.

Following hybridisation the microarray slides are scanned to reveal the relative proportion of fragments from the IP and IN sample at each cluster of probes. The scanner uses a laser of appropriate wavelength to excite the fluorophore labelled DNA and detect emitted light, in this way it can measure the relative intensity of fluorescence at each probe cluster between the IP and IN samples. The data is processed to generate a plot of the relative IP/IN ratio at each probe position along the section of the genome represented on the microarray slide. This is a measure at each probe position of the relative immunoprecipitation efficacy as the result of the adduct levels at that region of the genome. A typical output from the DIP-chip assay for the detection of cisplatin-DNA adducts in fibroblast cell cultures after treatment with 2500μM cisplatin is shown as an example below (figure 3.3). The IP/IN ratio is plotted on the y-axis, typically as the $\log_2$ of the value to compress the extremes of the possible range of data. In this plot scale, a value of 1 represents equal fragment numbers in the IP and IN samples. On the x-axis the probe position is plotted as a grey dot in the lower axis and the genomic location is plotted on the upper x-axis. This builds up to represent the pattern of adducts present in the sample along the genome.

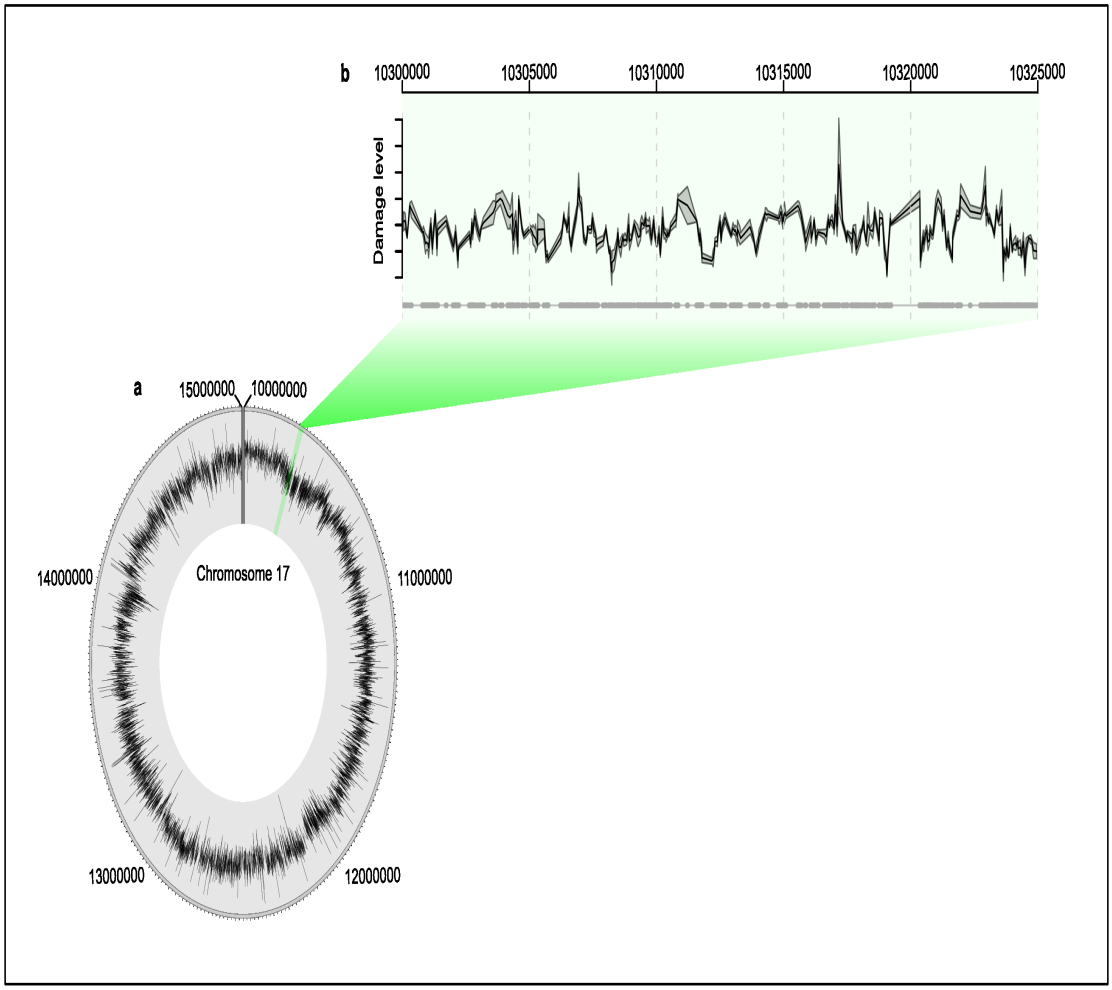


Figure 3.3: A cisplatin-DNA adduct DIP-chip assay output
A DIP-chip assay output from two cisplatin treated dermal fibroblast samples in cell culture. Probes over a 5Mb region of chromosome 17 are present on the custom microarray. Hybridisation of differentially labelled IP and IN DNA samples results changes to the relative fluorescence at each probe detected during scanning, reflecting the relative immunoprecipitation efficacy as a consequence of platinum-DNA adduct levels at each probe site. The IP/IN ratio value at each probe is calculated and plotted at all the probe regions in a circos plot (A), with the genomic position along chromosome 17 as labelled, giving an indication of the adduct pattern along the whole dataset. (B) shows a short section. The log2 IP/IN is plotted on the y-axis as the damage level at each probe position, shown by the grey dot in the lower x-axis. The upper x-axis shows the genomic location of the probes. Two biological repeat samples are shown in B, with a close match between the adduct patterns over this section of the genome (Powell 2014).

3.2.3 Sources of clinical samples used in these studies

For the assay development work presented in this chapter blood samples will be the clinical material used. This reflects the common use of blood samples in platinum-DNA adduct studies, and the ease of obtaining and using these samples. Specifically, blood cells with nuclear DNA are required for adduct analysis - the fraction containing peripheral blood mononuclear cells (PBMC) – which can be separated from the whole blood sample by density gradient centrifugation. The blood samples in this study are obtained from two healthy volunteers in the laboratory (labelled patients A, B) or are obtained from donations from the Welsh Blood Service following ethical approval for the use of these tissues for research purposes (application WBS Ad-hoc 008-12). Samples from the Welsh Blood Service consist of a single donor unit of 450ml whole blood with the majority of erythrocytes, plasma and platelets removed for clinical purposes prior to delivery. As such, they are a concentrated source of PBMC cells and once diluted in phosphate buffered saline can be subject to the same density gradient centrifugation procedure as whole blood samples (see chapter 2, section 2.1).

3.3 Results

3.3.1 Confirmation of successful sample processing and DNA extraction

To confirm successful sample preparation and handling using the protocol for PBMC separation, extraction and treatment in cell culture conditions using the protocol described in chapter 2 section 2.1, five repeat samples were taken from one individual of 20ml blood and PBMC were harvested. The amount collected is shown in figure 3.4 (panel A), and is in the range expected for these types of procedures of 0.8 to 3 million PBMC per ml of whole blood (Njai et al. 2011), confirming the protocol is appropriate for PBMC harvest from these clinical samples. Following separation, these PBMC were cultured in RPMI-1640

without the addition of serum or antibiotics, with or without the addition of oxaliplatin. These conditions were chosen as they are the treatment protocol used for cell culture experiments on human dermal fibroblasts in the development work of the DIP-chip assay (Powell 2014).

Following treatment with oxaliplatin (at variable doses) DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen) following the manufacturers protocol and quantified using a Nanodrop-1000 spectrophotometer. The results, figure 3.4(B) are the average of untreated and treated samples for each of the 5 independent experiments. The average of these two conditions is used as there is no significant difference in untreated and treated sample yield in these experiments.

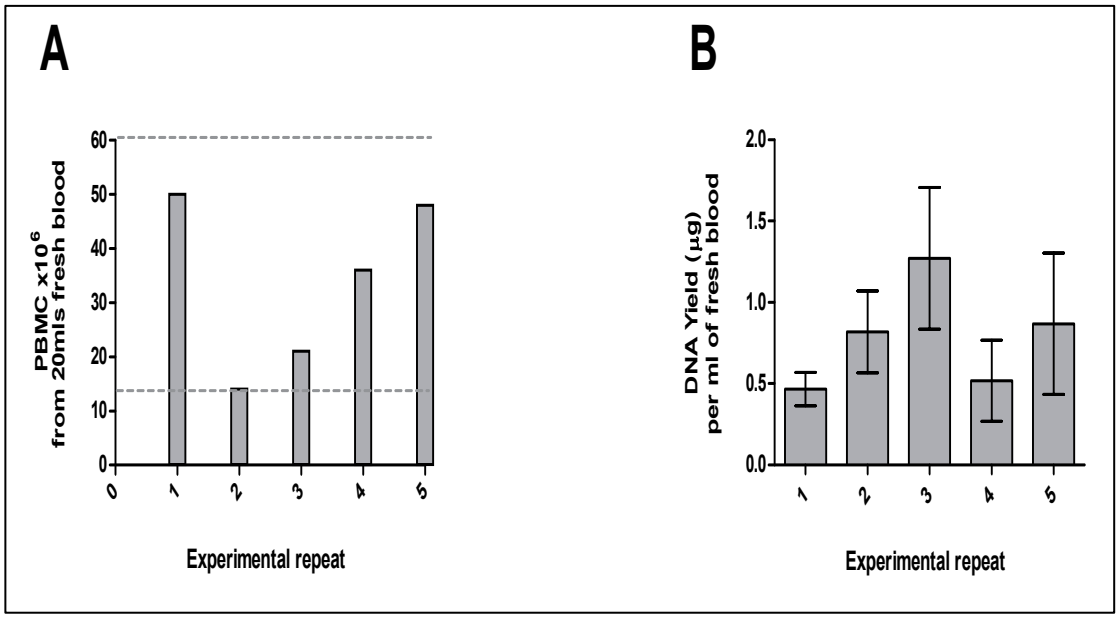


Figure 3.4: Successful PBMC sample processing and DNA extraction.
(A) Following the protocol outlined in chapter 2, section 2.1, PBMC were harvested from 20ml whole blood samples and 5 sequential occasions from the same individual. The yield of PBMC in each of 5 independent experiments is shown. The normal range is demonstrated by the grey dotted lines (Njai et al. 2011).
(B) Following culture with, or without, the addition of oxaliplatin and a variable time to recover after treatment, DNA was extracted using Qiagen DNeasy Blood and tissue kit following the manufacturers protocol. DNA quantity was calculated using a Nanodrop-1000 spectrophotometer. The mean and SEM of 4-6 samples from each of 5 sequential independent experiments is shown.

The PBMC yield per 20mls of fresh blood is within the normal range in each of 5 sequential experiments, indicated by the grey dashed lines, and evidence that the PBMC isolation protocol is suitable (Figure 3.4 A). In a

standard fibroblast cell culture DIP experiment 6.6μg of DNA per sample is required and the average yield of DNA per ml of blood is in the range of 0.47 to 1.27μg over 5 sequential independent experiments. This translates to a minimum blood volume to obtain adequate DNA for immunoprecipitation of 14.4 ml per experimental condition. In view of this, the DIP-chip protocol requires modification to ensure enough DNA is available for downstream processing during the assay and to maximise the DNA yield per ml of blood sample. This would give the greatest scope for development experiments, to maximise the number of experimental conditions that can be studied in each clinical sample, and minimise the volume of blood required from patients when the DIP-chip assay is used in the clinical setting.

Nevertheless, these results indicate that the first step in the translation of the DIP-chip assay is working, and that PBMC can be successfully harvested, treated and DNA extracted for use in the immunoprecipitation reaction.

3.3.2 Analysis of initial DIP protocol for variability

For use as a clinical assay to stratify patients prior to chemotherapy, and for use in complex treatment decisions with significant consequences, the assay must be reliable and reproducible. Consistent results and low variability must be demonstrated before for it to be validated for clinical use, and would give confidence for the sample to be taken on for genome-wide microarray analysis. A commonly used measure of laboratory assay variability for the assessment of quantitative assays is the coefficient of variance (CV), calculated by dividing the standard deviation by the mean of a number of repeat experiments, expressed as a percentage (Reed et al. 2002). This has the advantage of normalising the standard deviation for the magnitude of the mean and allows a comparison of variability between assays. For a typical laboratory assay the standard deviation of repeat measurements on the same sample should be less than 5% of the mean of an assay result, giving a CV of <5%. A CV of 10% or more is considered variable and over 20% is a clinically unreliable assay (Burd 2010). In our case, the DIP-qPCR result is not the final assay output, as the sample will be applied to

microarrays for genome-scale platinum-DNA adduct analysis, rather than used as an assay of the level of adducts at a single locus. This affords an additional opportunity beyond the qPCR stage for further data processing, using the greater scale of genomic data to reduce the variability in the final result (as will be demonstrated in chapter 5, section 5.6). In this context, a high CV result for qPCR from DIP samples does not render the DIP-chip assay non-viable. Nevertheless, the DIP-qPCR output is an important measure of the consistency of the immunoprecipitation 'DIP' step of the assay, and the CV provides a robust and widely used metric to quantify improvements in assay variability that can be potentially be made through modifications to the experimental protocol.

To calculate the CV of the DIP assay a series of independent experiments using blood samples from patient A and B were treated following the protocol and by the method outlined - directly adapted from the human cell culture development experiments. Briefly, following harvesting and washing, cultured PBMC were incubated for 4 hours in serum free media containing oxaliplatin at 1mM, before either immediate DNA extraction or culture in media containing 10% foetal calf serum, penicillin/streptomycin and L-glutamine at 37°C and 5% CO_2 for either 24 or 48 hours before DNA extraction. DNA was sonicated, immunoprecipitated and purified and analysed following the standard fibroblast DIP-chip method outlined above. The results of several repeat experiments are shown in figure 3.5 and the CV is calculated for 5 repeats from patient B in table 3.1. Due to technical failure of the assay with two of the exponential conditions for patient A, insufficient data is available to calculate the CV for this individual, except for at the 24-hour time point.

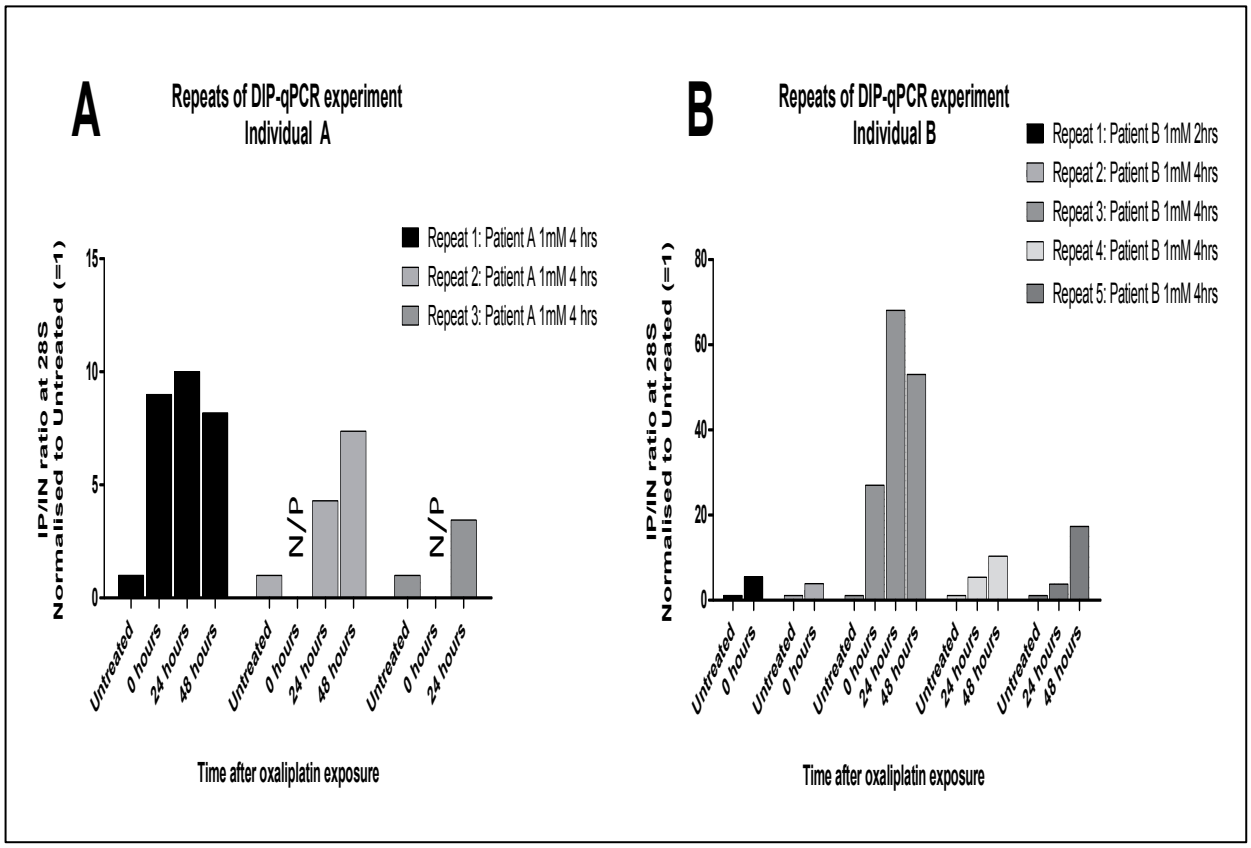


Figure 3.5: Analysis of initial DIP protocol for variability.
Several independent repeat DIP experiments were conducted according to the standard protocol and treated with 1mM oxaliplatin for 4 hours following PBMC harvest and culture (section 4.3.1). The result of the DIP assay is displayed as the relative change in IP/IN ratio calculated by qPCR with primers for 28S loci. The results are normalised to the IP/IN ratio in the untreated sample. 3 independent consecutive repeats from patient A and 5 repeats from patient B are displayed. Due to experimental failure two samples from patient A were not processed (indicated by N/P).

Variability Calculation – 5 repeats of Patient B			
Time post exposure	Mean	Standard Deviation	Coefficient of Variance
Untreated	1	-	-
0 hours	12.6	9.9	78.4%
24 hours	25.6	36.7	143.5%
48 hours	26.7	22.9	85.1%

Table 3.1: PBMC DIP result variability calculation
The IP/IN ratio normalised to the untreated sample for 5 independent consecutive repeats of the DIP assay for patient B is used for this table. The mean, standard deviation and coefficient of variance (CV) are calculated for each time-point.

These results highlight two main features of the use of PBMC samples with this experimental protocol. Firstly, the level of induced damage detectible

by DIP-qPCR when these cells are treated at 1mM oxaliplatin for 4 hours in serum free media is relatively low - the relative enrichment in all but one of the results obtained at time 0 (directly after treatment) is in the order of 5-10 times the amount of DNA immunoprecipitated in the untreated sample. This demonstrates poor discrimination between treated and untreated samples when using this schedule of oxaliplatin treatment. In comparable studies with dermal fibroblasts cultures in the development stages of this work the enrichment immediately after treatment is in the order of 50-100 fold (Powell 2014).

This difference could reflect a relative platinum resistance of PBMC compared to dermal fibroblasts. The dose used in the fibroblast experiments was in the order of 2.5mM, so is significantly higher, although the effect on cell function and cell survival was not assessed. The time scale of exposure is also relatively short, given that the platinum agents initially form mono-adducts, before forming di-adducts over several hours (Chapter 1, section 1.2.6) the true value of an immediate time point post exposure is questionable. It is notable 24 hours after exposure in all experiments a significantly increased enrichment is demonstrated compared to immediately post exposure, although this too is highly variable.

Another obvious finding is that there is significant variability in results generated using this protocol, with the 5 repeats of individual B having a high degree of variation in the outcome, which is also shown in the three repeats with individual A. The calculated CV from the 5 independent repeats from patient B is between 80-140%, indicating that the standard deviation similar in magnitude to the mean, and that the results are highly variable. A typical laboratory assay has a CV of <5%, and 10% is considered a high degree of assay variation (Burd 2010). Although this could be explained by the outlying abnormal result in Patient B repeat 3 (Figure 3.4), Patient A has a CV of 60% on the 24 hour sample (calculation not shown) confirming high assay variability between the same patient repeat samples.

As discussed above, the qPCR quantification of the DIP stage of the assay is intended as a confirmation of the success of immunoprecipitation and as a quality control step, rather than as an output for patient stratification, so a

proportion of the variability seen here will be reduced through the microarray processing and by using properties of the genomic scale of the resulting data to identify and remove experimental noise from the system, as will be discussed in chapter 5. The use of CV, in this context, is to quantify the variability at this stage of the assay and when the assay is conducted using the assay protocol as developed in cell culture system, notably with standardised conditions that are inherently less achievable with clinical samples. Even with these caveats, it is clear that the DIP stage of the assay requires further study to identify areas of high variability and to improve the protocol to reduce variability demonstrated here, and quantified using the assay CV metric.

3.3.3 Identification of assay stages with high variability

The variability seen in the previous experiment could have a biological component – for example due to variability in PBMC responses to oxaliplatin during the several weeks over which the experiment was conducted. However, there is also likely to be a component of technical variation in the assay. To investigate the relative contributions of biological and technical variability on the DIP assay, a single large pooled sample of DNA extracted from oxaliplatin treated PBMC was used for multiple repeat experiments. DNA was obtained from several identical aliquots of PBMC harvested from a single blood donation. The PBMC were treated in identical conditions with three dose levels of oxaliplatin for a 24 hour incubation to increase the generated levels of adducts. Once extracted, DNA from replicates at the same dose level was combined to give a large pool of identical DNA for each dose point.

DNA samples from each oxaliplatin dose (Untreated, 100μM, 500μM and 1mM) was then immunoprecipitated on three separate occasions to determine technical variability, as opposed to biological differences in the starting material. Each replicate was then analysed by qPCR in triplicate, to give nine assay outputs for each dose level. Any variability demonstrated should be due to technical variation in the immunoprecipitation reaction and from the qPCR step, rather than due to biological differences. The contribution of the variability of the assay

for each can then be identified. The results are shown below for the 1000μM sample, which is representative of the results at each dose level and used for illustration of assay variability (figure 3.6)

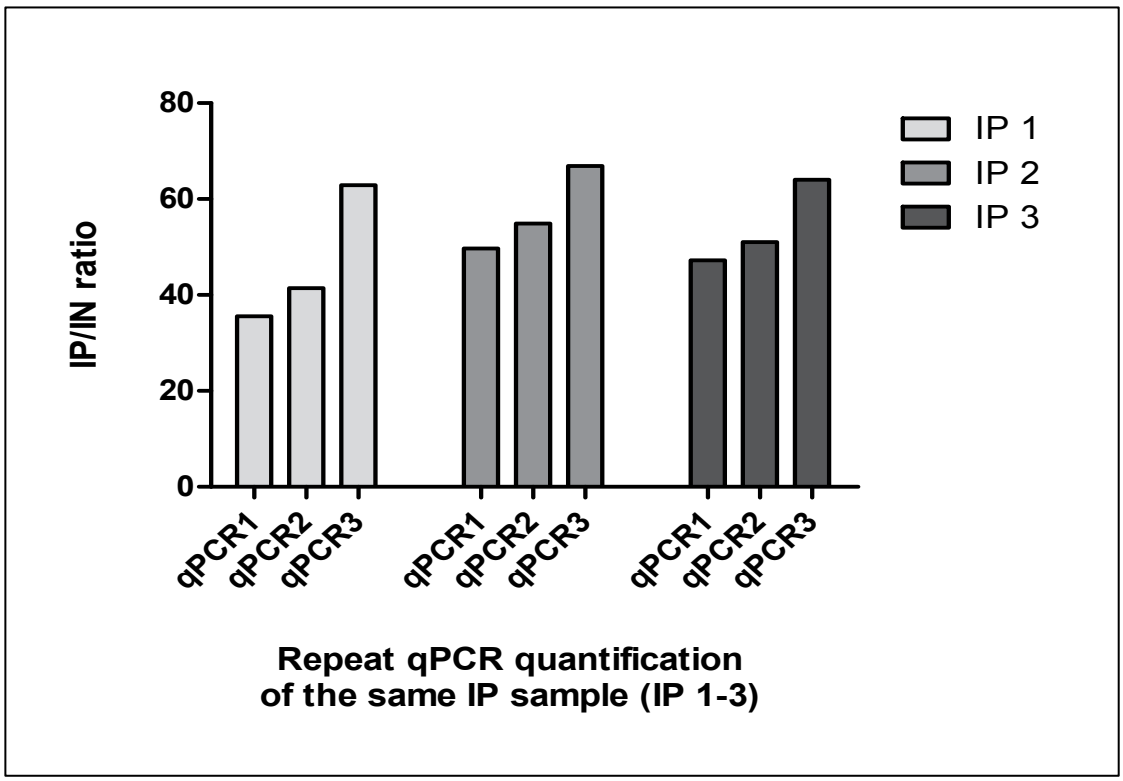


Figure 3.6: Identification of assay stages with high variability.
Pooled DNA samples from batches of identical patient PBMC treated with oxaliplatin at 1000μM doses for 24 hours. DNA is immunoprecipitated on 3 independent occasions (IP 1-3) and each samples is quantified in triplicate with qPCR (qPCR 1-3). The IP/IN ratio normalised to an untreated sample is displayed.

The result shows the assay technical variability. Each immunoprecipitation (IP 1-3) is from the identical post-sonication DNA sample. When processed on the same qPCR reaction (1-3) each is subject to the same qPCR efficiency, and each is normalised to the same on-plate reference standards and untreated sample. During qPCR1 the IP/IN ratio of IP1 is 35, increasing to 50 in IP 2 and 47 in IP 3. This is from an identical DNA sample that has been immunoprecipitated on 3 separate occasions and quantified in the same qPCR reaction. Therefore, the variability predominantly comes from the repeated immunoprecipitation, rather than from the qPCR (which is done on the same run

and intra-qPCR variation is reduced by using an on-plate standards and an on-plate untreated sample).

The same sample (IP 1-3) is also run on 3 sequential qPCR quantifications (qPCR 1 to 3). For each IP sample, the IP/IN ratios are higher on the third qPCR compared to the first – with almost a 2-fold increase in IP sample 1. This indicates high between-plate variability exists with the qPCR quantification of the DIP-chip assay and is also a source of variability in the DIP assay result. Both the immunoprecipitation step and the qPCR step are resulting in technical variation that would result in more consistent DIP output if the variability is decreased.

3.4 Modifications to the experimental protocol

3.4.1 Adjustment to tissue culture and oxaliplatin treatment conditions

The above experiment demonstrates that the experimental protocol requires significant modifications with several goals in mind. Firstly, modifications are required to increase the ability of the immunoprecipitation to discriminate between treatment doses by enhancing the IP/IN ratio, increasing the mean enrichment compared to the experimental error. This can be achieved by a second modification to improve the biological relevance of the assay by using a more biologically and pharmacokinetically representative treatment schedule, reflecting that the dose of oxaliplatin in patients remains constant in ultra-filterable plasma for 24 hours after treatment and that oxaliplatin-DNA di-adducts may take many hours to from the initial mono-adducts induced (chapter 1, section 1.2.5), making the changes detected with a short duration treatment and immediate DNA extraction problematic.

Thirdly, with a longer incubation time the protocol requires modification to add appropriate media supplements of serum and antibiotics. Serum starvation, even over a short period, results in changes to gene expression, metabolism, and induction of autophagy of non-essential cellular components (Chen et al. 2014; Golpour et al. 2014). Consistent use of supplemented media

reduces the potential confounding effects from the use of serum-free media and possible effects of serum starvation on the PBMC population.

3.4.2 Changes to oxaliplatin dose and incubation time

To predict for a response to treatment in DIP-chip clinical samples it could be argued that the *ex vivo* treatment schedule should reflect as far as possible the exposure that the tumour and normal tissues will be exposed to *in vivo*. During chemotherapy treatment, the pharmacokinetic studies discussed in chapter 1, section 1.2.2 show the that the dose of oxaliplatin given to a patient in free filterable plasma is in the order of 3μM to 5μM, and this level remains for 24 hours.

In order to mimic more closely the treatment schedule of patients, and to make any results generated more accurate, a modified treatment regimen was adopted. Once collected by density gradient centrifugation, PBMC were incubated in RPMI-1640 with the addition of 10% foetal calf serum, l-glutamine and penicillin/streptomycin for 30 minutes at 37°C and at 5% CO_2, prior to the addition of cytotoxic agents. Samples were either untreated, or treated with the addition of different doses of cisplatin or oxaliplatin. After 24 hours incubation the drug was removed, the cells washed twice in PBS, and DNA was extracted. Using this approach any information derived from the experiment should more accurately reflect the conditions for treating patients.

Using this new treatment schedule and unchanged extraction, sonication, immunoprecipitation, purification and qPCR downstream processing the IP/IN result generated from these samples, referenced to an untreated sample are shown in figure 3.7.

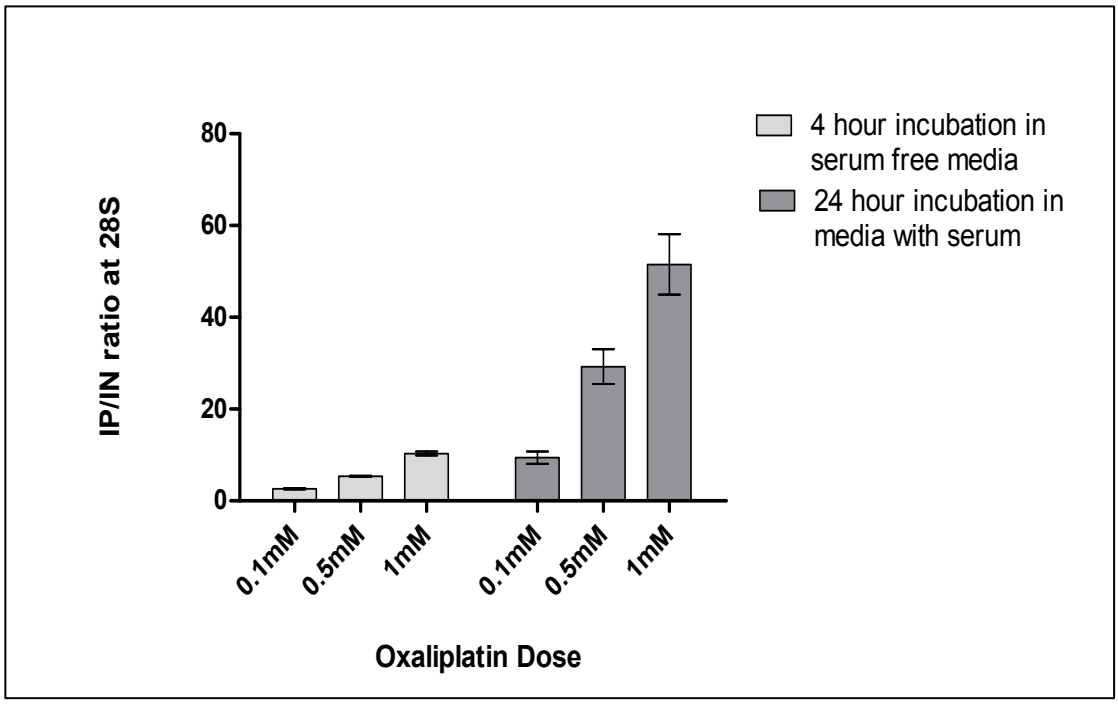


Figure 3.7: Demonstrating the effect of changes to oxaliplatin dose and incubation time. PBMC harvested from the same blood sample are treated with differing doses of oxaliplatin in RPMI-1640 with or without the addition of serum supplements, and for 4 hours duration compared to 24 hours. The IP/IN ratio in each sample is normalised to an untreated sample and the mean and SEM of three independent experiments are shown.

This result demonstrates that the detectable adduct levels (the relative increase in IP/IN ratio as a measure of the increase in immunoprecipitated fragments compared to the baseline IN sample, normalised to an untreated sample) are significantly increased compared to the original 4 hour drug exposure in media without serum, improving the ability of the assay to discriminate between dose levels following treatment.

3.4.3 Determination of cell survival post incubation and treatment

The effect of the PBMC preparation (separation by density gradient centrifugation, washing, and culturing) and of the effect of different exposure to oxaliplatin treatment can be measured by assessing the survival of cells using flow cytometry. At the high doses of 1mM used in this study, 200-300 times the plasma dose experienced in patients during chemotherapy, the platinum-adduct levels and any correlation with response and toxicity could be potentially less

biologically relevant if the experimental process results in significant metabolic effects or high levels of cell death. The addition of propidum iodide (PI) to a cell population fluorescently labels non-viable cells, as viable cells are impermeable to the molecule. The cell population can be sorted in PI positive and negative populations using flow cytometer (in this instance an Accuri C6 flow cytometer was used), giving an accurate indication of cell survival in the PBMC population. Using the treatment schedule outlined in section 3.4.2, including the 24-hour oxaliplatin incubation period, the following cell survival results were obtained (Figure 3.8) and the combined results of three repeated experiments are shown below (Figure 3.9).

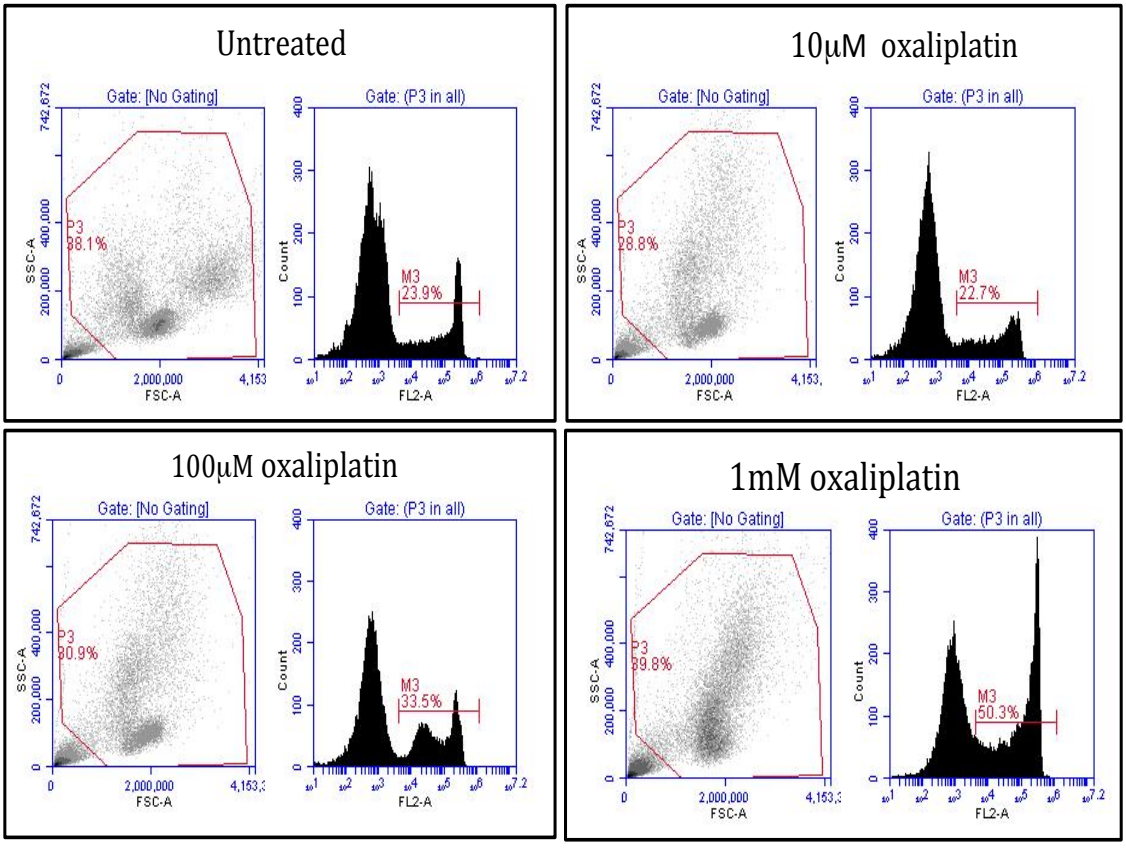


Figure 3.8: PBMC cell survival post incubation and treatment.
Flow cytometry results from analysis of a PBMC population exposed to oxaliplatin in cell culture for 24 hours. The addition of propidum iodide allows identification and quantification of a subpopulation of PBMC that are non-viable. At each dose level the right window shows the proportion of cells that are non-viable and take up PI by the relative size of the right hand peak.

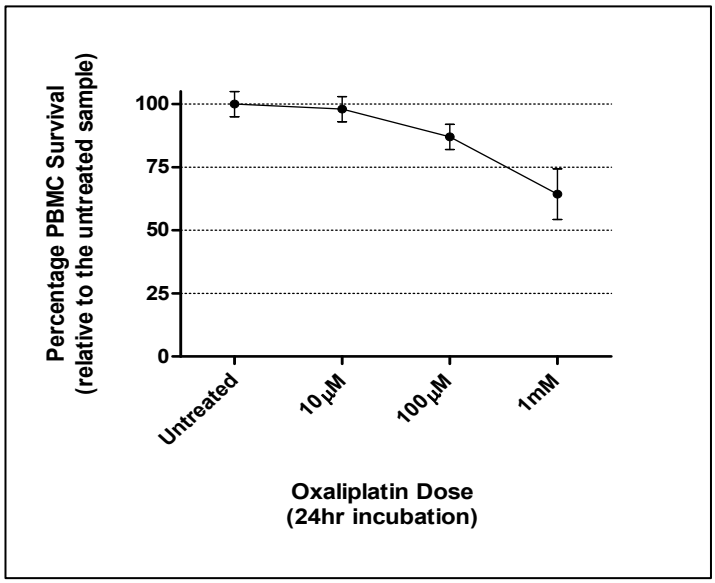


Figure 3.9: The survival of PBMC populations treated with oxaliplatin in cell culture.
The survival of PBMC populations treated with oxaliplatin in cell culture for 24 hours as assessed by flow cytometry (Figure 3.8) is shown relative to the untreated sample. The mean and SEM of three repeats is displayed.

These results demonstrate that about 80% of the cells are viable following processing in both the untreated sample and at the 10μM oxaliplatin dose. Therefore, the plasma concentration of 3-5μM oxaliplatin found in patients during chemotherapy treatment is unlikely to result in increased rates of circulating PBMC *in vivo*.

Increasing the oxaliplatin exposure to 100μM and 1000μM increases the rate of cell death in the PBMC population compared to the untreated and 10μM (and hence compared to the *in vivo* clinical dose). Although this introduces potential differences in the biological behaviour of our *ex vivo* assay PBMC and the *in vivo* patient situation, the higher doses are required to generate sufficient oxaliplatin-DNA adducts to enable differences between the dose points to be reproducibly detected by immunoprecipitation (figure 3.7). As these doses are requires, the effect is unavoidable; the discrepancy is noted, and the effect of higher rates of cell death at the high dose levels will be considered as a potential confounding factor throughout the on-going experiments to optimise the DIP-chip assay for clinical applications.

3.4.4 Standardising supply and storage of oxaliplatin and cisplatin

To maximise reproducibility of the assay prior to use on clinical samples, the treatment with oxaliplatin and cisplatin must be standardised. Initial experiments were conducted on donations of oxaliplatin from the local chemotherapy pharmacy, each of varying age and shelf life. These samples were stored at 4°C, and have been used for several months for initial validation experiments and development studies. To ensure consistency of supply, oxaliplatin and cisplatin were commercially obtained (Abcam), stored as a powder at 4°C, and suspended in DMSO immediately prior to use, as suggested in the accompanying product information.

To confirm the reproducibility, reliability and comparability to previous results obtained with the pharmacy supplied reagents, the level of adducts detectable by treatment with cisplatin and oxaliplatin in both the legacy stock (cisplatin stored for 6 months at 4°C) and the fresh supply of powdered drugs in DMSO (0.4% final concentration) were assessed on a sample of PBMC obtained from the Welsh Blood Service from a single donor. The results are shown below (Figure 3.10).

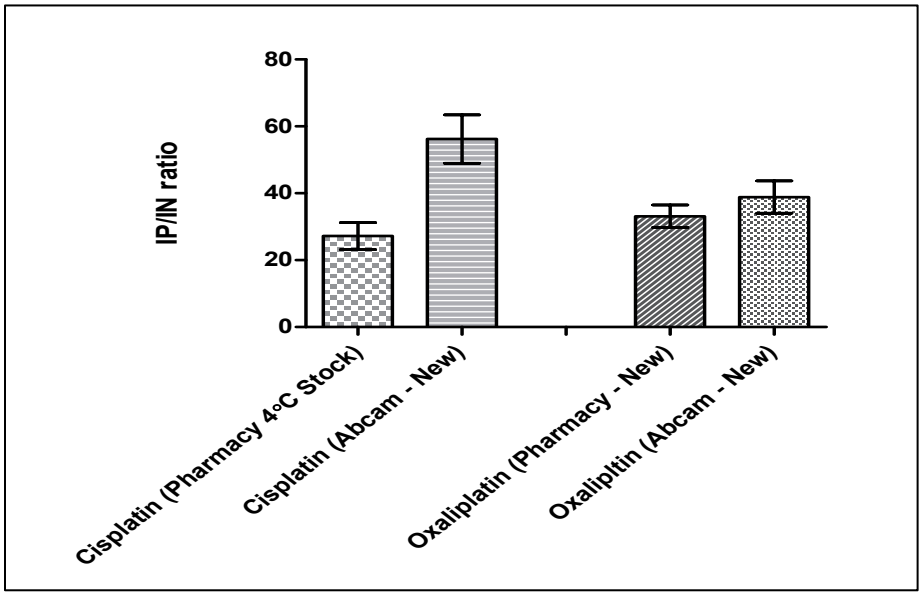


Figure 3.10: Measuring the effect of platinum drug age and supplier.
The adducts generated by local pharmacy and commercially supplied platinum agents was assessed for reproducibility and comparability between suppliers. PBMC from a single donor were treated over three repeat experiments. The mean adduct levels (calculated by the IP/IN ratio normalised to an untreated sample) are shown, with the SEM shown by error bars.

The result of this experiment demonstrates the expected pattern of 2-3 fold increase in adducts with fresh cisplatin compared to fresh samples of oxaliplatin, and comparable adduct generation between the two fresh formulations of oxaliplatin. The older stock of cisplatin generates significantly fewer adducts, highlighting the issue of inconsistent drug batches and an effect of drug storage on the potential variability of assay results, unless fresh drug is made and stored appropriately before each experiment. In the pharmacy, a fresh vial of platinum stock is used for each patient and then discarded. In our case, although all stock was used within the documented shelf life of the vial, as only microliter volumes were required for each experiment a single vial was used multiple times over the course of several months. One each occasion sterile technique was maintained, but the potential for contamination of the vial cannot be discounted. Each time a sterile needle was used to pierce the vial seal to gain access to the drug, so an alternative explanation is the possible breakdown of the seal with multiple needle tracks, resulting in a portal for the entry of air, moisture and contaminants that could affect the quality of the drug over time.

3.4.5 Optimising DNA extraction, condition and sample purity

Downstream applications of DIP-qPCR and DIP-chip require significant amounts of DNA at high purity, free from protein and RNA contamination; over 95% of intra-cellular platinum is bound to macromolecules other than DNA, so contamination is potentially a cause of significant confounding and assay variability. The benefit of improving the DNA yield from limited volumes of whole blood are obvious, and are especially relevant in a clinical assay - maximising the number of experimental conditions that can be studied, minimising patient inconvenience and allowing repeat experiments if a technical failure occurs.

3.4.5.1 Assessment of DNA yield and quality

The experiments described to data use the same PBMC amount per treatment condition (~5 million PBMC). At higher doses of oxaliplatin exposure of 1mM for 24 hours cell survival falls to 50% - with a fixed amount of cells prior to treatment this could result in a reduction in DNA yield with increasing doses of platinum drugs. Using the treatment conditions outlined above, PBMC from a single donor were exposed in three replicate experiments to oxaliplatin. The DNA yield at the end of the DNA extraction step was quantified with a Nanodrop-1000 spectrophotometer and is shown below (Figure 3.11) and the DNA quality was assessed by gel electrophoresis (Figure 3.12).

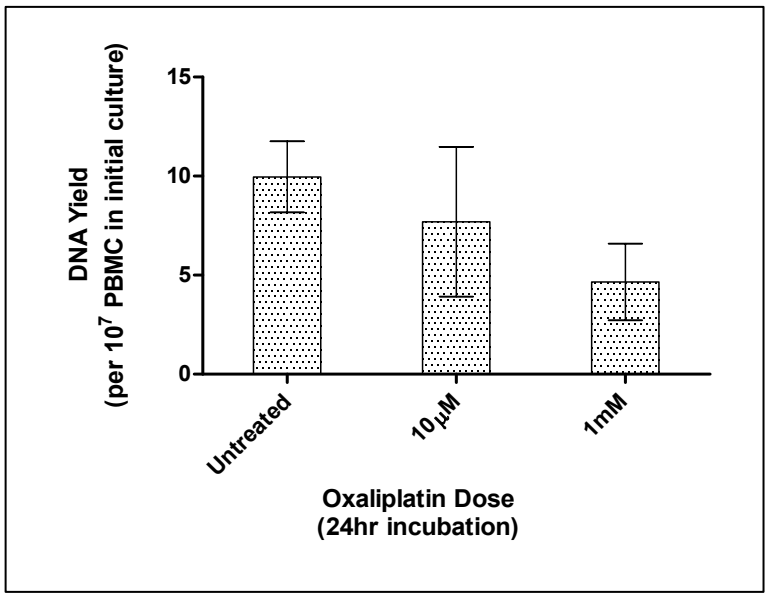


Figure 3.11: Quantifying DNA yield following sample processing and DNA extraction.
PBMC from a single donor were treated with oxaliplatin at varying doses for 24 hours in cell culture prior to DNA extraction. The DNA amount is quantified by Nanodrop-1000 spectrophotometer and adjusted to DNA yield per 107 PBMC

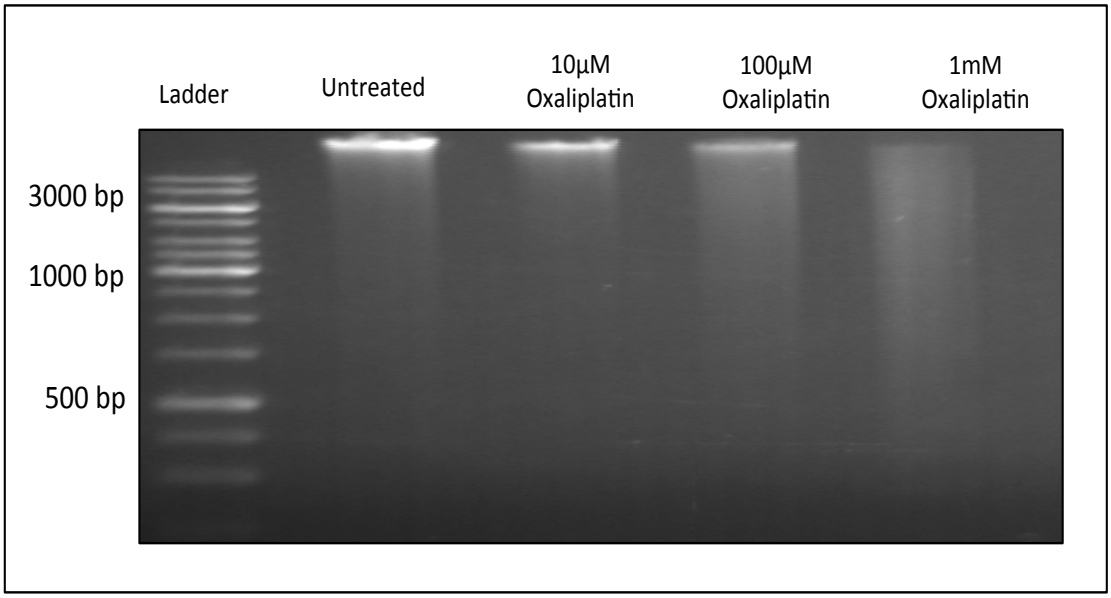


Figure 3.12: Assessing the quality of the extracted DNA.
The DNA extracted from PBMC treated with varying doses of oxaliplatin is run on a 0.8% TAE gel.

These experiments reveal that the lower survival of PBMC with 1mM oxaliplatin is associated with a decrease in extractable DNA following treatment and that the DNA extracted is degraded and of poor quality compared to lower-dose samples. As apoptosis is the primary mechanism of cell death following platinum-DNA adduct formation (see chapter 1 section 1.3) it is the likely explanation for the shearing of genomic DNA seen in this result. The degradation of the extracted DNA at increased oxaliplatin doses could have an impact on the reliability of the DIP assay and suggests that increasing the dose of drug further could have an adverse impact on assay performance.

3.4.5.2 Optimising DNA extraction conditions

The Qiagen DNeasy Blood and Tissue DNA extraction kit protocol describes significant fall in DNA yield if cell quantities above $5*10^6$ cells are used in the extraction process, presumably as a result of a less efficient reaction if the extraction columns are overwhelmed with contaminants and cell debris.

To maximise the DNA available from limited clinical samples, it may be optimal to split the cells into several columns, reducing the inefficiency of the process and maximising DNA yield for each millilitre of initial blood volume. By

splitting the cells into increasing fractions for processing and performing a second elution step, changes in obtainable DNA yield are shown in figure 3.13.

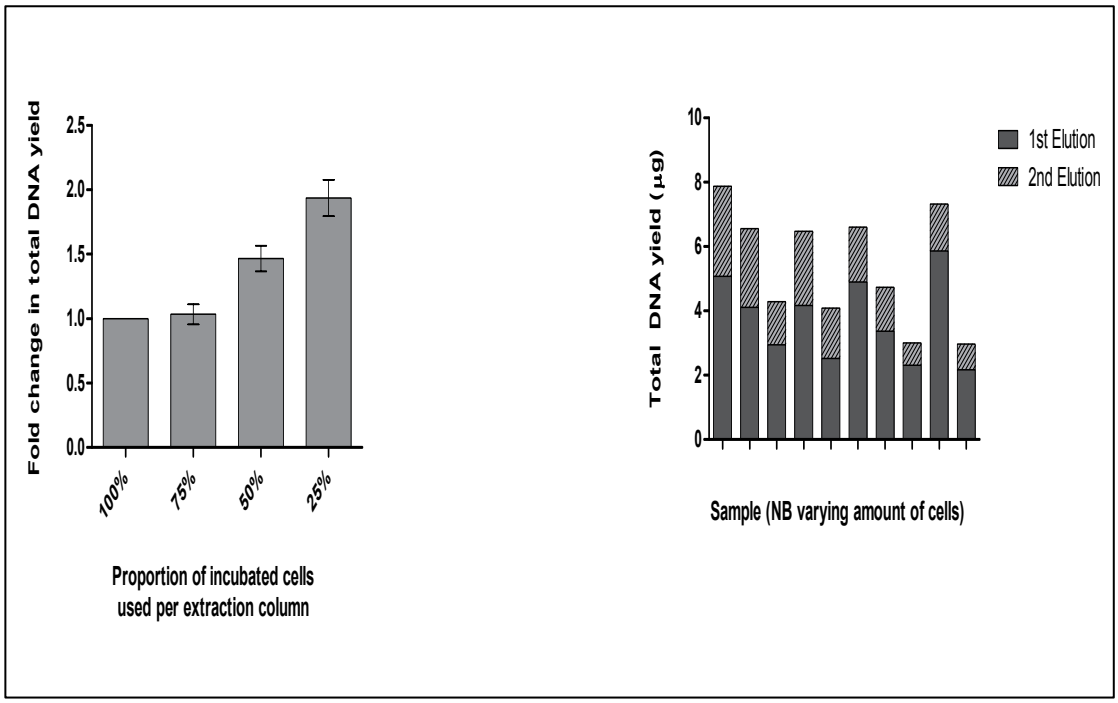


Figure 3.13: Measuring the effect on DNA yield of optimising the DNA extraction process. The results of changes to DNA extraction protocols to optimise DNA yield from limited clinical samples are shown. The left graph reveals that reducing the cell count per extraction column and combining the product significantly increases the possible DNA yield from the process. The right graph shows that a second DNA elution from each purification column adds approximately 50% to the obtainable yield from the Qiagen DNeasy Blood and Tissue Kit.

Splitting the samples into fewer cells per column and a second elution step results in a significantly increased total yield. A second elution in 100uL of buffer AE rather than two elutions in 200uL results in a more concentrated second elution, in keeping with the concentration of the first elution, and ensures that the final concentration is not too dilute for downstream applications. As an example of the effect of this protocol change Figure 3.14 shows the average DNA yield from 4 PBMC extractions of 5 million cells on two occasions pre- and post-protocol changes, demonstrating the benefits of this modification.

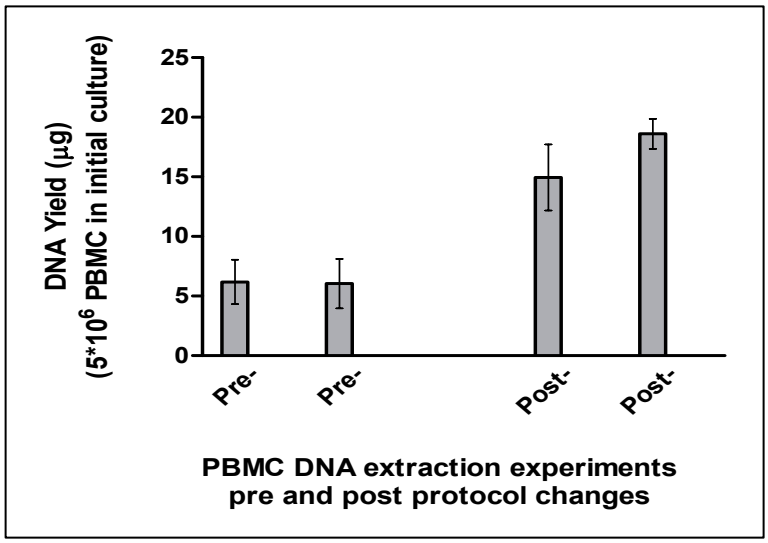


Figure 3.14: Measuring the effect on DNA yield of optimising the DNA extraction process. The DNA yield per 5 million PBMC following the DNA extraction protocol of splitting cell pools 50/50 prior to extraction and eluting the resulting DNA twice with a second volume of 100µL.

3.4.5.3 Improving purification of DNA prior to immunoprecipitation

Over 95% of all cellular platinum is bound to macromolecules other than DNA, resulting in a significant risk of these confounding effects altering downstream processing. For example, protein and RNA contamination into the immunoprecipitation reaction may occur if all protein and RNA is not removed efficiently. DNA extracted from the Qiagen DNeasy Blood and Tissue kit can have significant RNA and protein contamination, as detected during DNA quantification with the Nanodrop-1000 spectrophotometer. The ratio of absorbance at 260nm to 280nm is used to assess purity of DNA with a ratio of 1.8 indicating pure DNA and a ratio of 2 indicating pure RNA. A low ratio indicates protein or other contaminants that typically absorb at, or near 280nm. The ratio of absorbance at 260/230 also indicates sample purity and should be in the range 2 to 2.2 for a pure sample. Again protein is a significant contaminant if the ratio is abnormal.

To measure contamination, 30 consecutive DNA samples extracted using the DNeasy Blood and Tissue Kit (Qiagen) from PBMC treated with oxaliplatin were assessed for purity using the Nanodrop-1000 spectrophotometer, and

purified using PureLink PCR Purification Kit (Invitrogen) prior to reanalysis for purity. The results are shown in figure 3.15.

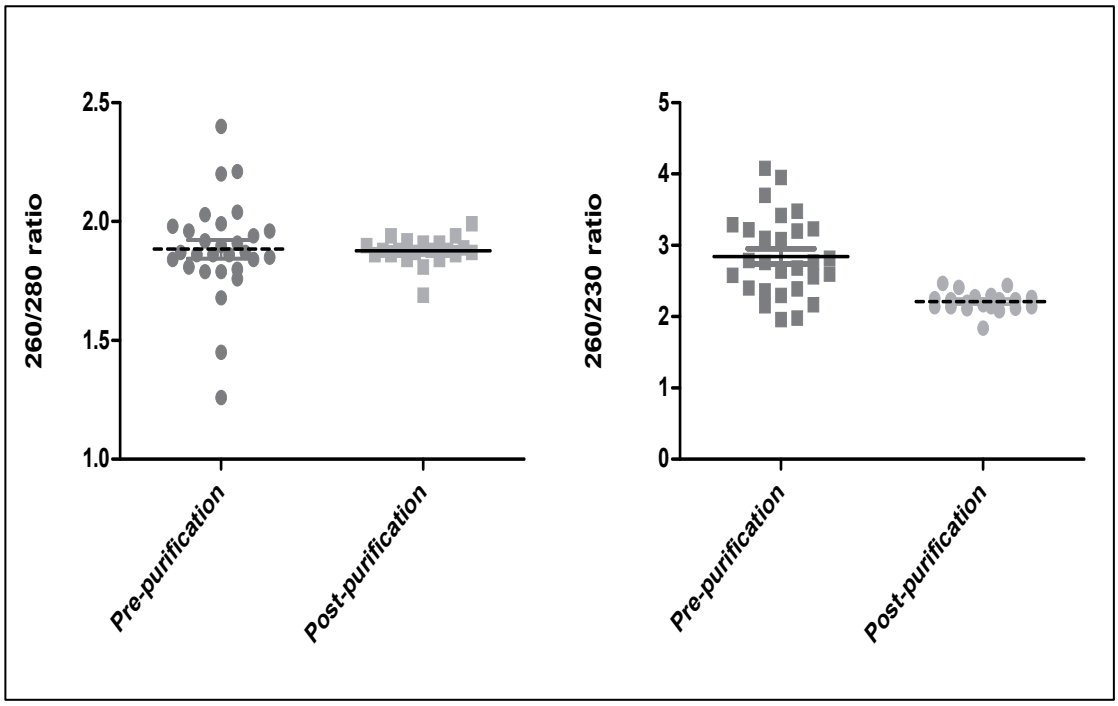


Figure 3.15: Improving DNA quality prior to immunoprecipitation.
30 treated PBMC cultures of 5 million cells were treated and DNA was extracted with the Qiagen DNeasy Blood and Tissue Kit. Protein and RNA contamination was measured with a Nanodrop-1000 spectrophotometer. A 260/280 ratio of 1.8 and 260/230 ranges of 2 to 2.2 are considered pure DNA.

In the pre-purification samples significant contamination can be seen, shown by the wide ranges of 260/280 and 260/230 ratios displayed. Following purification with Pure Link PCR DNA Purification kit (Invitrogen) the ratios are highly homogenous, indicating that the DNA for downstream processes is free of protein and RNA contamination.

3.4.5.4 Optimising the timing of DNA purification

Column purification using the Invitrogen Pure Link PCR DNA purification kit is more efficient with fragmented rather than genomic DNA. To confirm this, 14 DNA samples extracted with the DNeasy Blood and Tissue kit (Qiagen) were purified pre-sonication and 21 samples post sonication, with resulting DNA

amount compared to the starting sample to calculate the amount of DNA lost per procedure. The relative change in yield of DNA due to purification is shown below in figure 3.16.

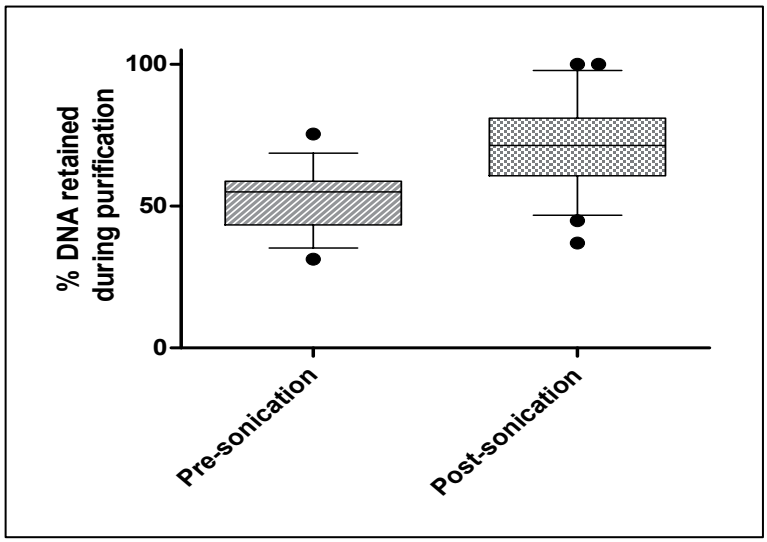


Figure 3.16: Quantifying the effect of purification on DNA yield.
The loss of DNA if purification is added to the protocol pre or post sonication is shown. 14 samples are analysed pre-sonication and 21 samples purified post-sonication. The mean, IQR and outliers are shown.

Clearly, the addition of a purification step post sonication results in pure DNA with improved yield. These modifications result in a good yield (Figure 3.14) of high quality DNA compared to the initial protocol used at the start of these experiments. High quality DNA is then available for sonication and use in the immunoprecipitation reaction.

3.4.6 Optimisation of sonication conditions

Differences in fragment length have a major impact on PCR efficiency, with shorter fragments preferentially amplified compared to longer fragments and with a high risk of introducing bias over several cycles of PCR (Dabney and Meyer 2012; Shagin et al. 1999). Differential DNA fragment length also has a significant effect on microarray probe hybridisation efficiency. In one example a change in fragment size from 45 base pairs to 1480 base pairs resulted in a 14-fold reduction in probe hybridisation and fluorescence detection in one study

(Liu et al. 2007). Evidently, to minimise systematic variability in both the DIP and the DIP-chip assay between sample replicates and between different treatment conditions the sonication phase has to be carefully controlled and highly reproducible between all samples.

During the DIP-chip assay development work a protocol was developed for sonication of naked human DNA using a Diagenode Bioruptor Sonicator system for 30 seconds on/30 seconds off, with a starting DNA amount of 200µl total volume and a concentration of 100ng/µl. The samples are cooled to 4°C and with 4 samples per carousel for each sonication experiment. This results in an average fragment length of 200bp (Powell 2014).

There are several issues with the translation of this phase of the protocol for use in these clinical samples. Optimisation of the yield of DNA extraction results in a final volume of 300µl, rather than 200µL. The extra 50% volume may have an effect on the efficiency and consistency of sonication. The extracted DNA concentration can also vary widely in clinical samples, in contrast to the high concentrations available from batching of large numbers of cell culture experiments. A concentration of between 100ng/µL and 30ng/µL is typical for the PBMC extraction experiments described so far, reflecting the variability in yield as a consequence of treatment effect (Figure 3.11) and purification (Figure 3.15). This variation in DNA concentration could also have an impact on the consistency of sonication. Additionally, in the interests of efficiency, with many samples processed in parallel to reduce error, it is optimal to use 6 tubes rather than 4 per cycle, maximising workflow. The impact of these factors is uncertain and may result in unequal or variable fragment length following sonication.

To measure the effect of these variables on sonication reproducibility and fragment length distribution, an experiment was conducted using 70 samples of DNA extracted using the modification discussed in the above experiments, resulting in varying DNA concentration between 33 ng/µL and 100 ng/µL. The samples are untreated, treated with oxaliplatin, or treated with cisplatin. The number of sonication cycles, number of tubes per sonication, DNA concentration was varied to measure the effects on sonication consistency. Following sonication the samples were analysed using a TapeStation 2200 (Agilent Technologies) a tool for simultaneous gel electrophoresis of multiple samples to

measure and quantify the fragment distribution. Data was recorded for the shortest 5%, followed by the 25%, 50%, 75% and 95% fragment lengths, quantifying the distribution of DNA in each sample with increasing cycles of sonication. The results are analysed to demonstrate the effect of varying DNA concentration, adduct type, cycle number and tubes per reaction in the following figures (Figure 3.17 to 3.19)

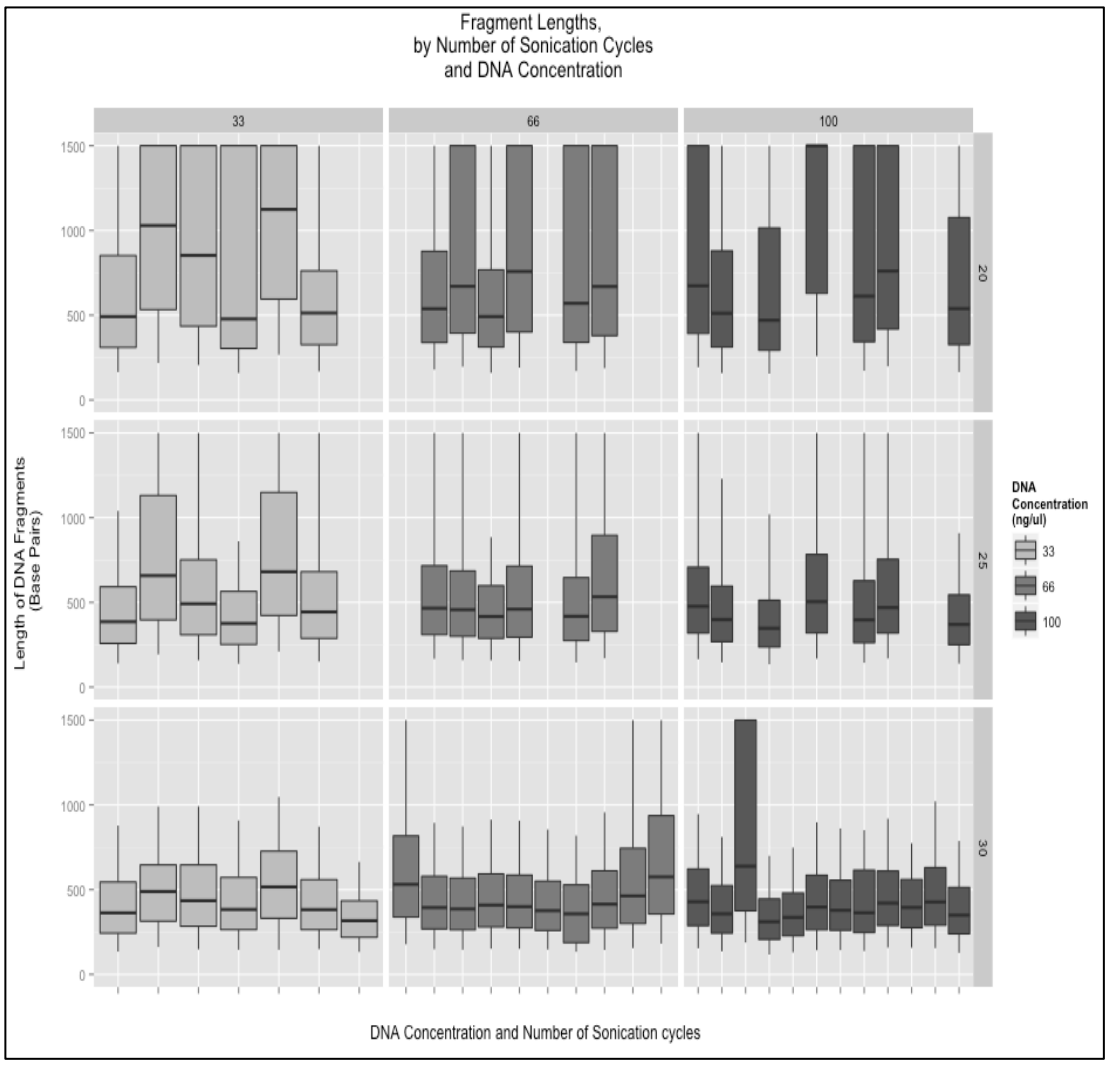


Figure 3.17: Determining the effect of variation in sonication parameters on the consistency of DNA fragmentation.
A demonstration of several DNA fragment profiles following sonication with a Diagenode Bioruptor Sonicator, 30 seconds on/30 seconds off, at 4°C. The mean fragment length, 25% and 75% are shown in the central box and the distribution of the 5% and 95% longest fragments is shown by the whiskers. The increasing number of cycles of 20, 25 and 30 results in a tighter distribution of fragment levels, with the central 50% of the fragments between 300 bp and 500 bp in most samples and 90% of the fragments (5% to 95%) between 100 and 900 bp after 30 cycles. The distribution is more variable with fewer cycles. There is no effect of DNA concentration on sonication performance as fragment distribution profiles do not change with DNA concentration (33ng/µl - light grey, 66ng/µl – grey, 100ng/µl – dark grey)

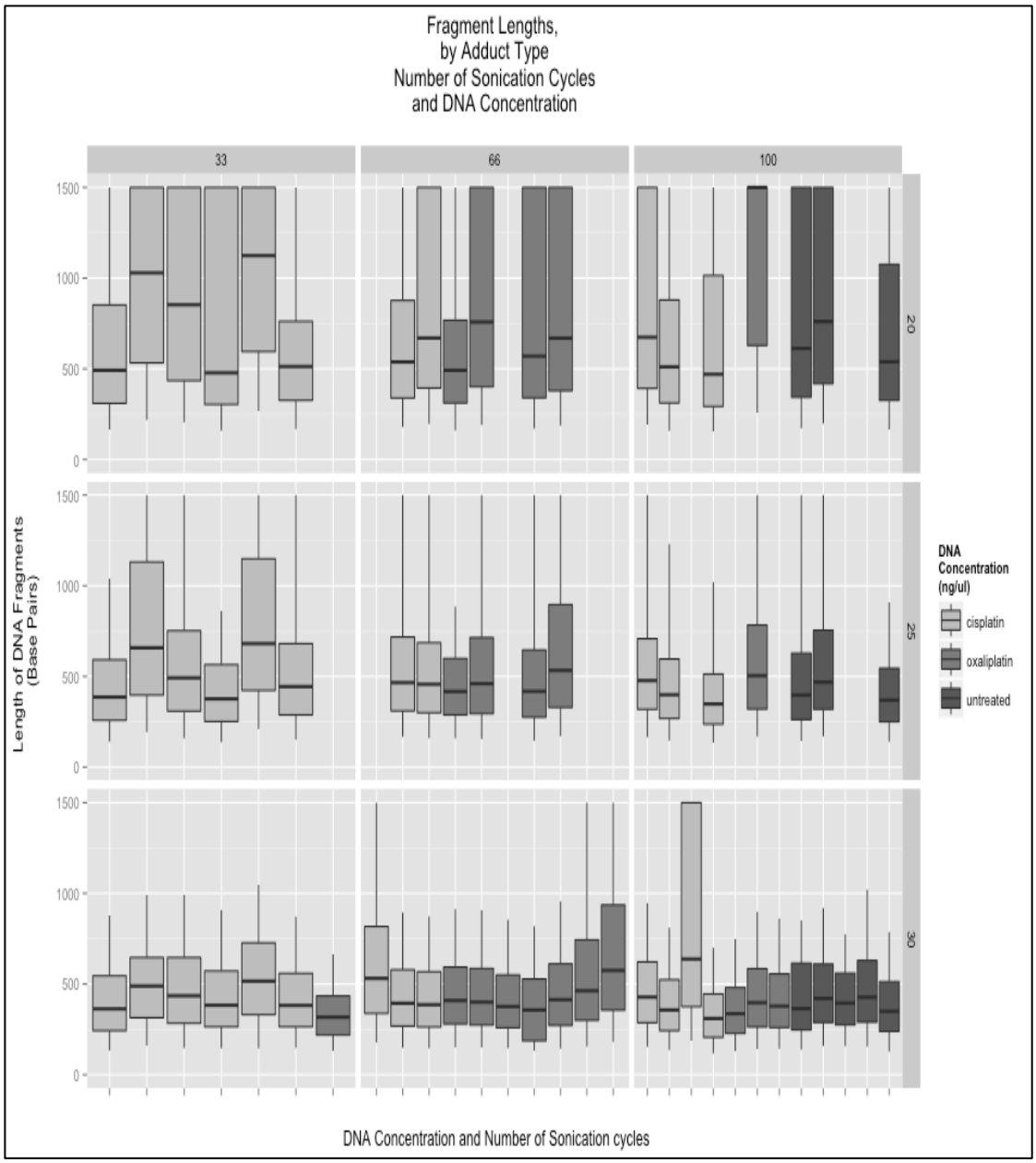


Figure 3.18: Determining the effect of variation in sonication parameters on the consistency of DNA fragmentation.
A demonstration of several DNA fragment profiles following sonication with a Diagenode Bioruptor Sonicator, 30 seconds on/30 seconds off, at 4°C. The mean fragment length, 25% and 75% are shown in the central box and the distribution of the 5% and 95% longest fragments is shown by the whiskers. The increasing number of cycles of 20, 25 and 30 results in a tighter distribution of fragment length, with the central 50% of the fragments by amount concentrated between 300 bp and 500bp in most samples and 90% of the fragments (5% to 95%) between 100 and 900 bp after 30 cycles. The distribution is more variable with fewer cycles. There is no effect of adduct type on sonication performance as fragment distribution profiles do not change with adduct type (cisplatin – light grey, oxaliplatin – grey, untreated – dark grey)

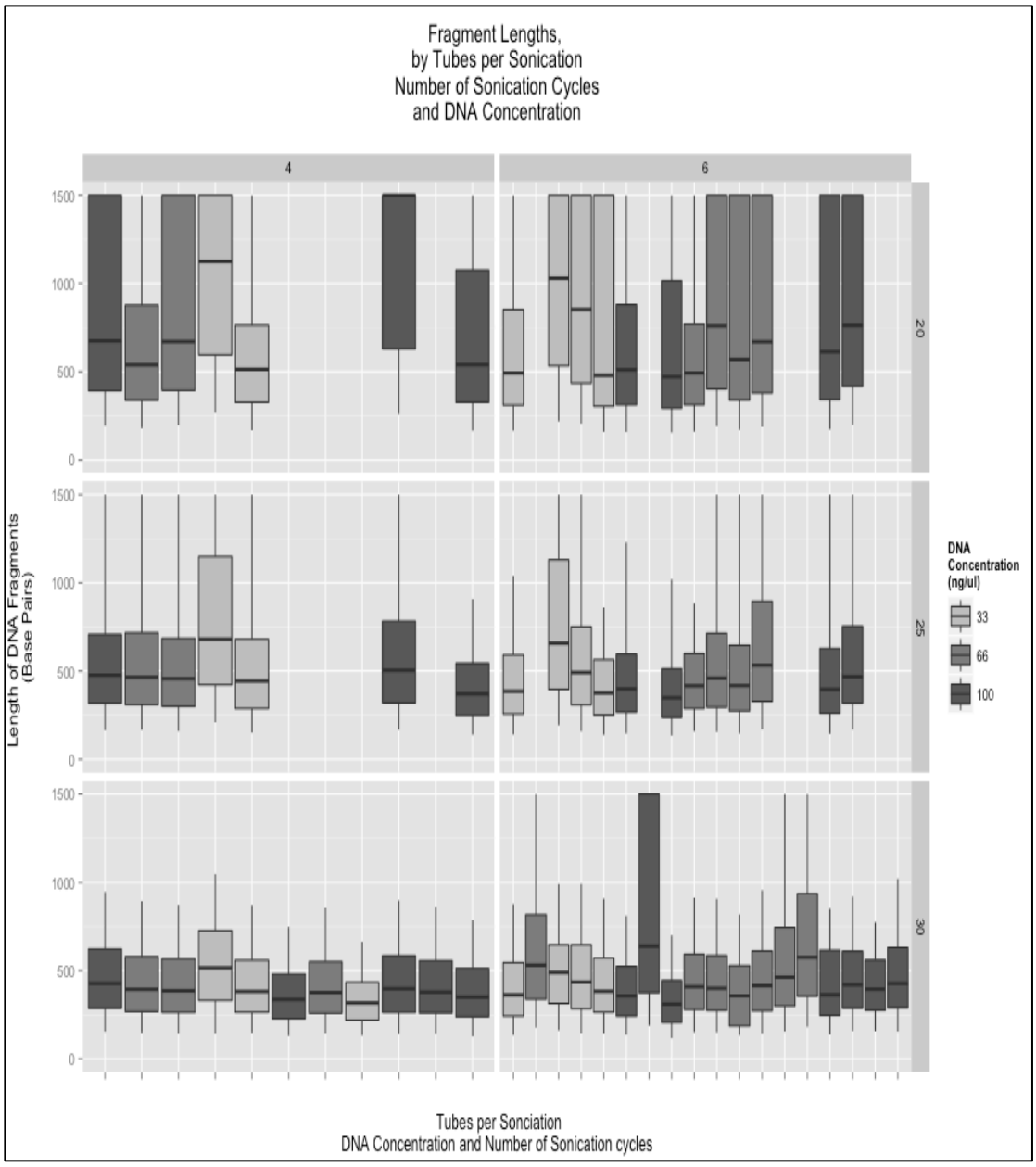


Figure 3.19: Determining the effect of variation in sonication parameters on the consistency of DNA fragmentation.
A demonstration of several DNA fragment profiles following sonication with a Diagenode Bioruptor Sonicator, 30 seconds on/30 seconds off, at 4°C. The mean fragment length, 25% and 75% are shown in the central box and the distribution of the 5% and 95% longest fragments is shown by the whiskers. The increasing number of cycles of 20, 25 and 30 results in a tighter distribution of fragment lengths, with the central 50% of the fragments by amount concentrated between 300 bp and 500bp in most samples and 90% of the fragments (5% to 95%) between 100 and 900 bp after 30 cycles. The distribution is more variable with fewer cycles. There is no effect of the number of tubes per sonication (4 vs. 6) on sonication performance as fragment distribution profiles do not change at 30 cycles, or with DNA concentration (33ng/μl - light grey, 66ng/μl – grey, 100ng/μl – dark grey)

These data, displayed in Figures 3.17 to 3.19 demonstrate that there is high variability in the effect of sonication at 20 and 25 cycles. However, with an

increasing number of sonication cycles to 30 the distribution fragments becomes highly consistent. The median fragment length is 400bp and the central 50% of the fragments are within +/-200bp, and 90% of the fragments between 150 and 900 base pairs. Furthermore, the sonication is consistent and unaffected by DNA concentration (Figure 3.17) adduct type (Figure 3.18) or using 4 or 6 tubes per sonication (Figure 3.19).

The protocol used in cell cultures needs to be modified to 30 cycles from 24, as in these experiments 25 cycles results in significant heterogeneity in fragment distribution. This is likely to be a consequence of increasing the sample volume prior to sonication, as necessary to optimise the DNA yield from precious clinical samples compared to unlimited availability in cell culture models.

3.4.7 Optimisation of immunoprecipitation consistency

3.4.7.1 Calculation of DNA amounts for immunoprecipitation

The amount of DNA typically used for a DIP experiment for detecting DNA damage is 6μg, as determined from the experiments used in the assay development (Powell 2014; Teng et al. 2010). This amount of DNA in the immunoprecipitation reaction results in adequate amounts of precipitated template DNA being available post-immunoprecipitation for downstream applications during the DIP-chip process. If possible, using less DNA per reaction would increase the number of experiments that could be conducted with each clinical sample. However, using less DNA could influence the variability of the assay, and potentially could have a significant impact on reproducibility if it reduces the available template for DNA amplification, required to generate adequate material for microarray analysis during the chip experiment – it may then require further cycles of PCR with increases in the inherent bias of this process.

To explore these effects, an experiment was conducted to assess the effect of reducing the amount of DNA in an immunoprecipitation reaction. The use of

1μg of DNA was chosen to compare with the standard 6μg and the results are shown below (Figure 3.20).

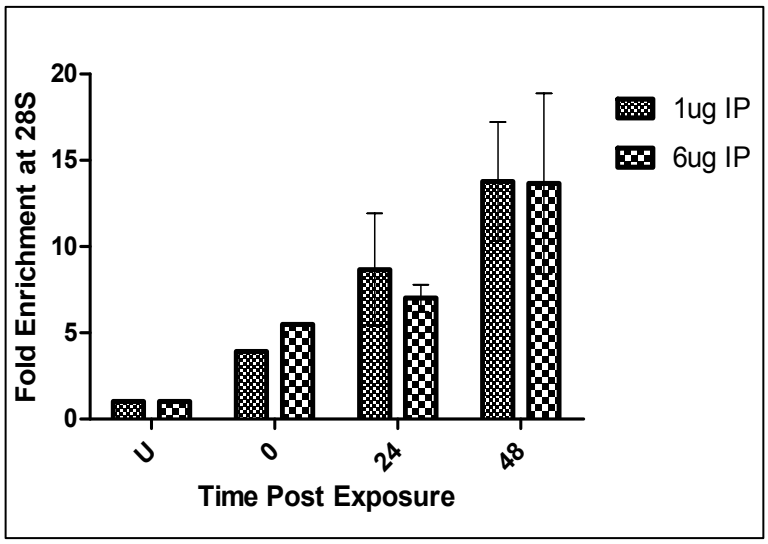


Figure 3.20: Calculation of amount of DNA for reproducible DNA immunoprecipitation. Immunoprecipitation experiments with a different starting DNA amount on the same sample of oxaliplatin treated PBMC DNA, repeated on three independent occasions. The mean and SEM are shown. In each experiment either 1μg or 6μg of DNA is used, with other conditions identical.

Figure 3.20 demonstrates that it is possible to reduce the amount of DNA in the immunoprecipitation step of the DIP assay with no significant effect on the result, indicating that the range of DNA used in this experiment is within the dynamic range of the immunoprecipitation assay. If the Dip-qPCR assay is used as a stand-alone measure of DNA damage and repair, lower amounts of DNA can be used, increasing the number of conditions that can be studied or reducing the amount of sample that is required. As 1 ml of blood produced 0.8-3 million PBMC ((Njai et al. 2011), Figure 3.4), and 5 million PBMC produces approximately 20μg DNA with the protocol amendments discussed (Figure 3.14) a few millilitres of blood may be enough to generate data for several experimental conditions per patient. However, the effect of reducing the DNA output from the immunoprecipitation on downstream amplification and on the reproducibility of microarray is yet to be determined and may be significant.

3.4.7.2 Optimising CP9/19 antibody for immunoprecipitation

The initial developmental work using human dermal fibroblast in cell culture employed a CP9/19 antibody amount of 1.5μg per 6μg of DNA for each immunoprecipitation experiment (Powell 2014). To confirm that this is applicable to the PBMC clinical samples, and to show that changes to the antibody concentration have minimal impact on the variability, an antibody titration was conducted using 6μg DNA for immunoprecipitation at a range of antibody concentrations, as shown below (Figure 3.21)

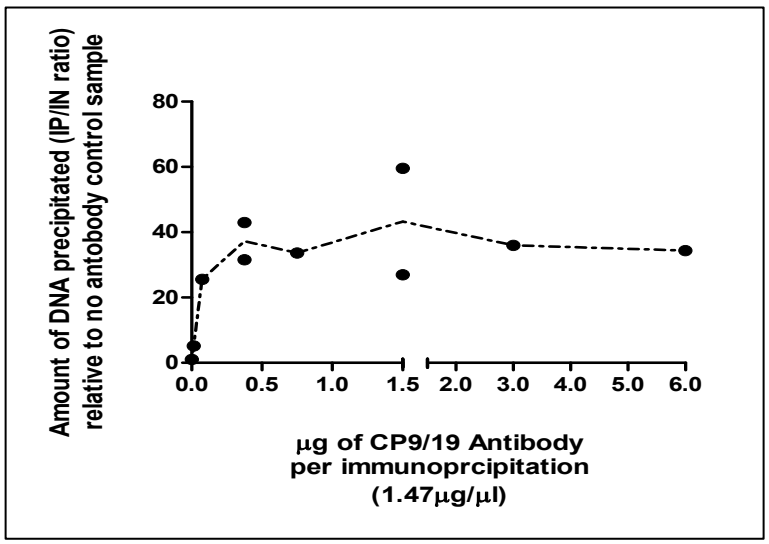


Figure 3.21: CP9/19 antibody titration.
An antibody titration experiment using 6μg DNA per immunoprecipitation and variable amounts of CP9/19 anti platinum antibody. The IP/IN ratio is normalised to a no-antibody sample and a single result or the mean of two repeats is shown. These results are identical to those in the assay development work (Powell 2014) showing a plateau of antibody effect with no improvement above 0.5μg of antibody per reaction.

These results confirm that the previously identified antibody concentration is appropriate for this series of experiments, and the amount used is saturating, indicating that variability seen in the experiment is not a consequence of limiting amounts of antibody. As the titration experiment indicates that a 66% dilution or 400% increase in antibody used would yield a similar IP/IN ratio, the antibody concentration is not contributing to variability in the DIP assay.

3.4.7.3 Effect of DNA concentration on immunoprecipitation consistency

DNA extraction results in varying concentrations of DNA being obtained prior to immunoprecipitation. This is a consequence of using a standard amount of PBMC for cell culture, drug treatment, the effect of the drug on reducing the survival of cells and the subsequent extractable DNA amount and quality at higher doses, as described in this chapter. At the higher dose of 1mM oxaliplatin treatment the amount of material available following extraction and DNA purification can be 50% lower than samples that are untreated or treated with a lower dose of drug.

The amount of DNA is adequate for immunoprecipitation (1μg is as good as 6μg), but the DNA concentration varies, typically between as high as 100ng/μl and as low as 33ng/μl. This has no effect on the reproducibility of sonication. However, when used in the immunoprecipitation stage the different concentration of DNA results in a three fold difference in buffer used in the reaction. This is a potential cause of variability in the assay.

This potential effect was investigated using a pool of DNA extracted from a single donor pool of PBMC supplied by the Welsh Blood Service and treated with oxaliplatin and cisplatin using the standard protocol. The extracted DNA from several samples were pooled and adjusted to a final concentration of 100ng/μl. Serial samples from this batch of DNA were diluted in extraction buffer either to a final concentration of 66ng/μl, or to 33ng/μl. The resulting buffer mix during the immunoprecipitation was therefore either a standard 100ng/μl DNA with approximately 75% PBS and 25% AE (DNA extraction buffer from the Quiagen DNeasy Blood and Tissue Kit), or in the most dilute samples was approximately 90% AE and 10% PBS. The result on the IP/IN ratio (the mean - error bars show SEM) of 3 independent experiments is shown in figure 3.22.

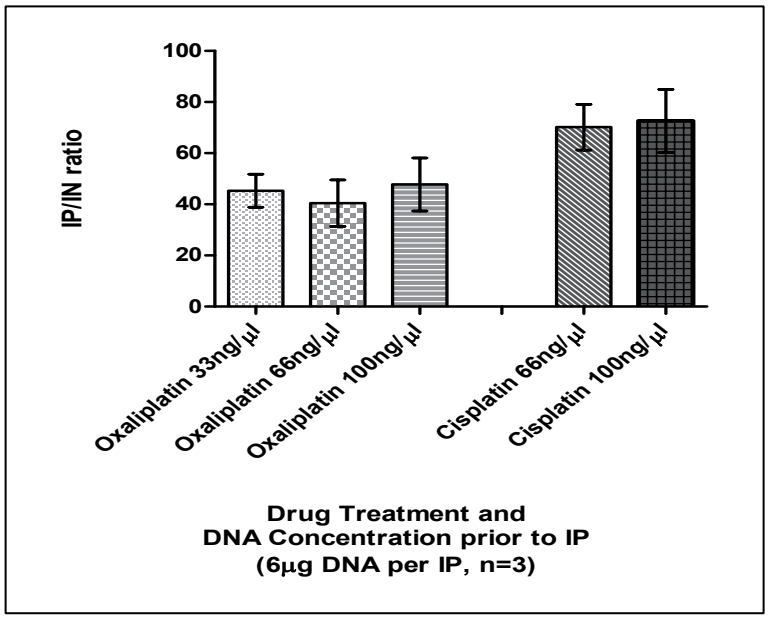


Figure 3.22: The effect of DNA concentration and buffer variation of IP consistency.
A pooled sample of DNA extracted from PBMC from a single donor treated ex vivo with oxaliplatin and cisplatin is subject to immunoprecipitation at varying initial starting concentration. The mean IP/IN ratio normalised to an untreated sample from three independent repeat experiments (the error bars show SEM).

This results show that the outcome of the immunoprecipitation, as measured by the relative enrichment between the untreated and treated samples is consistent despite the significant difference in buffer composition in the reaction mix, dependant on the initial DNA starting concentration. This confirms that the DNA concentration and resulting composition of buffers in the reaction has no effect on the outcome of immunoprecipitation, and samples can be used as prepared from the DNA extraction and sonication without concerns that this step is introducing variability into the final result.

3.4.8 Reducing qPCR variability

For the assay to be used reliably, the results of the immunoprecipitation measured by qPCR and must be reproducible. When used for calculation of the coefficient of variance and for identification of sources of variability (in sections 3.3.3 and 3.3.4) it is clear that the qPCR step is a significant source of assay error, and that results of PCR quantification of identical DNA on different occasions introduces considerable run-to-run variation, leading to variability in the results obtained from the assay.

The qPCR step can be used only as a qualitative confirmation of immunoprecipitation success, if the DNA is later to be analysed by microarray to generate the DIP-chip result rather than as a stand alone measure of adduct damage and repair in a DIP-qPCR assay, reducing the significance of this assay variability. However, even in this circumstance, the qPCR is used to estimate the amount of DNA to be amplified on the microarray, so ideally it should be as consistent as possible to reduce the potential for downstream error and bias.

Concerns regarding qPCR variability when comparing between experiments have been previously identified in the scientific literature (Burns et al. 2005; Rutledge 2004). The widespread use of multi-step protocols with subtle variations at each step introduces inconsistency into qPCR data, and leads to in difficulties in reproducing and comparing studies in which the conclusions are based on qPCR analysis (Bustin 2010). In an effort to standardise the reporting of qPCR data a set of guidelines to allow reproduction and comparison between experiments have been published (Bustin et al. 2009). These guidelines, the Minimum Information for Publication of Quantitative Real-Time PCR Experiments, "MIQE" guidelines, have been cited in the scientific literature over 2450 times by July 2014. The issue of run-to-run variability is discussed and comparable samples on each plate are suggested for use as 'inter-run calibrators'.

To determine the magnitude of run-to-run variability and to develop inter-run calibrators to allow accurate comparisons to be made between qPCR results, an experiment was conducted to develop a set of common 'master

standards' to compared between qPCR runs. These standards were produced from several pooled input samples from untreated PBMC DNA that had been subject to all of the processes involved in the DIP assay. Initially, to assess the reproducibility of the qPCR and the validity of this approach the same standards were run on 10 sequential qPCR assays, using primers of varying batch and several different batches of SYBR green reagent, to mimic the conditions that would be used in the clinical use of the DIP-chip assay.

Initial analysis of the variability in the threshold cycles of the standards is shown below (Figure 3.23 A and B), demonstrating that the quantification of identical DNA samples is variable between repeated experiments, at first glance indicating that the qPCR process of this assay is subject to considerable variability. Each sample crosses the threshold for quantification with a 95% confidence interval of (approximately) +/- 2 PCR cycles, meaning that there is occasionally a 4-cycle difference in par conditions with identical DNA.

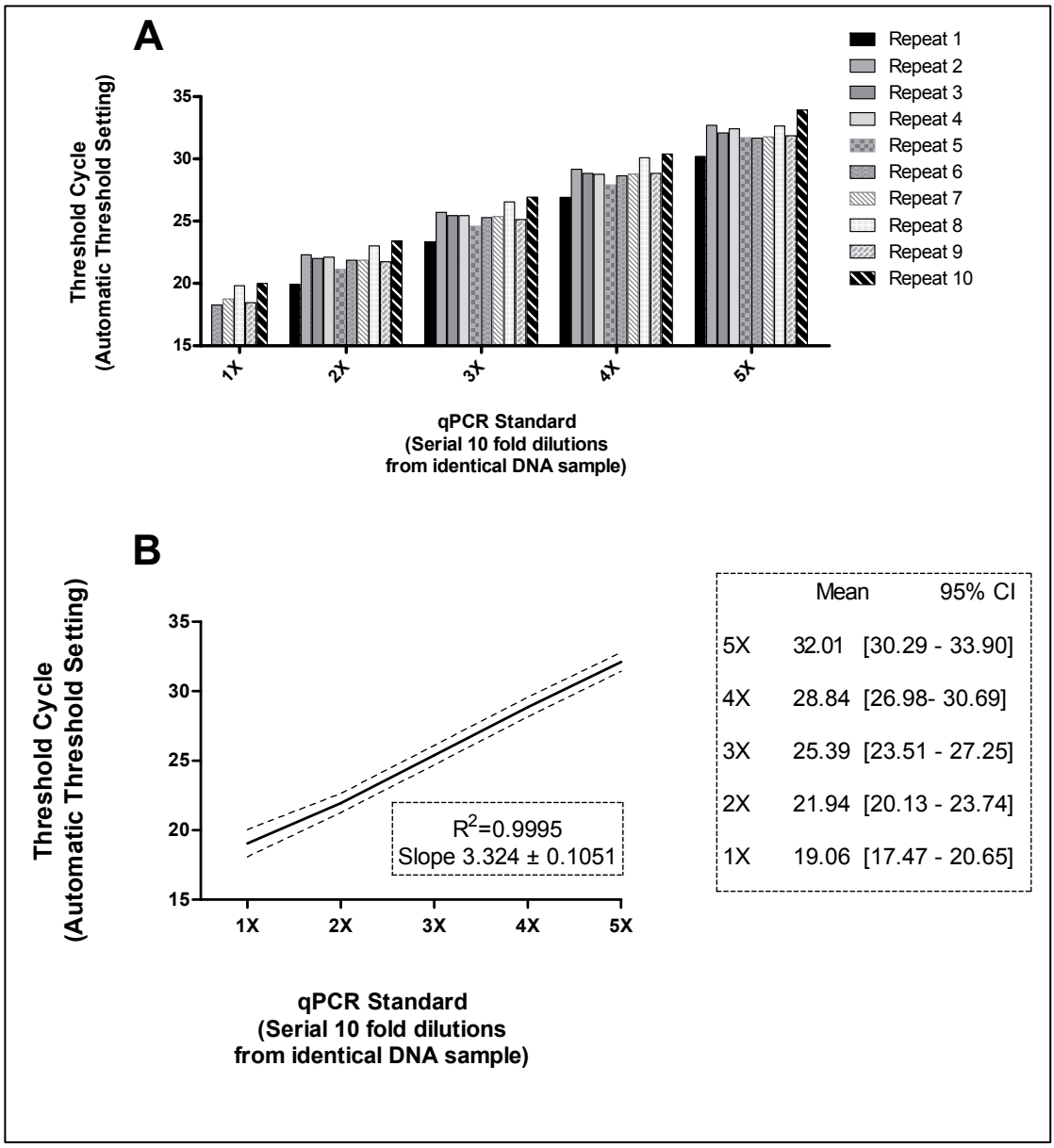


Figure 3.23: Determination of qPCR variability.
10 consecutive qPCR experimental quantifications of identical DNA standards was performed using varied batches of 28S primers and SYBR green reagent. The threshold cycle of each standard as calculated by the automatic software (Bio-Rad-CFX connect) is shown in (A) and the mean and 95% confidence interval of these repeats in (B).

This inherent variability in DNA quantification of identical DNA standards potentially occurs as a consequence of automatic variable threshold setting by the qPCR analysis software (Bio-Rad CFX Connect), during which the threshold for DNA quantification is set automatically at a value which is 10 times the baseline florescence detected during the run.

To remove this variability an arbitrary figure of 500 RFU was used to manually set a fluorescence threshold. In this circumstance each experiment is

subject to the same criteria for successful amplification and subsequent DNA quantification. In each of the 10 runs in this analysis this arbitrary threshold is higher than the automated generated threshold, and is in the logarithmic phase of fluorescence increase for all of the standards - it is therefore an appropriate measure to use for manual threshold setting. Using a common threshold of 500 RFU the result of the 10 repeats of the 5 standards were re-analysed and is shown in figure 3.24 A and B.

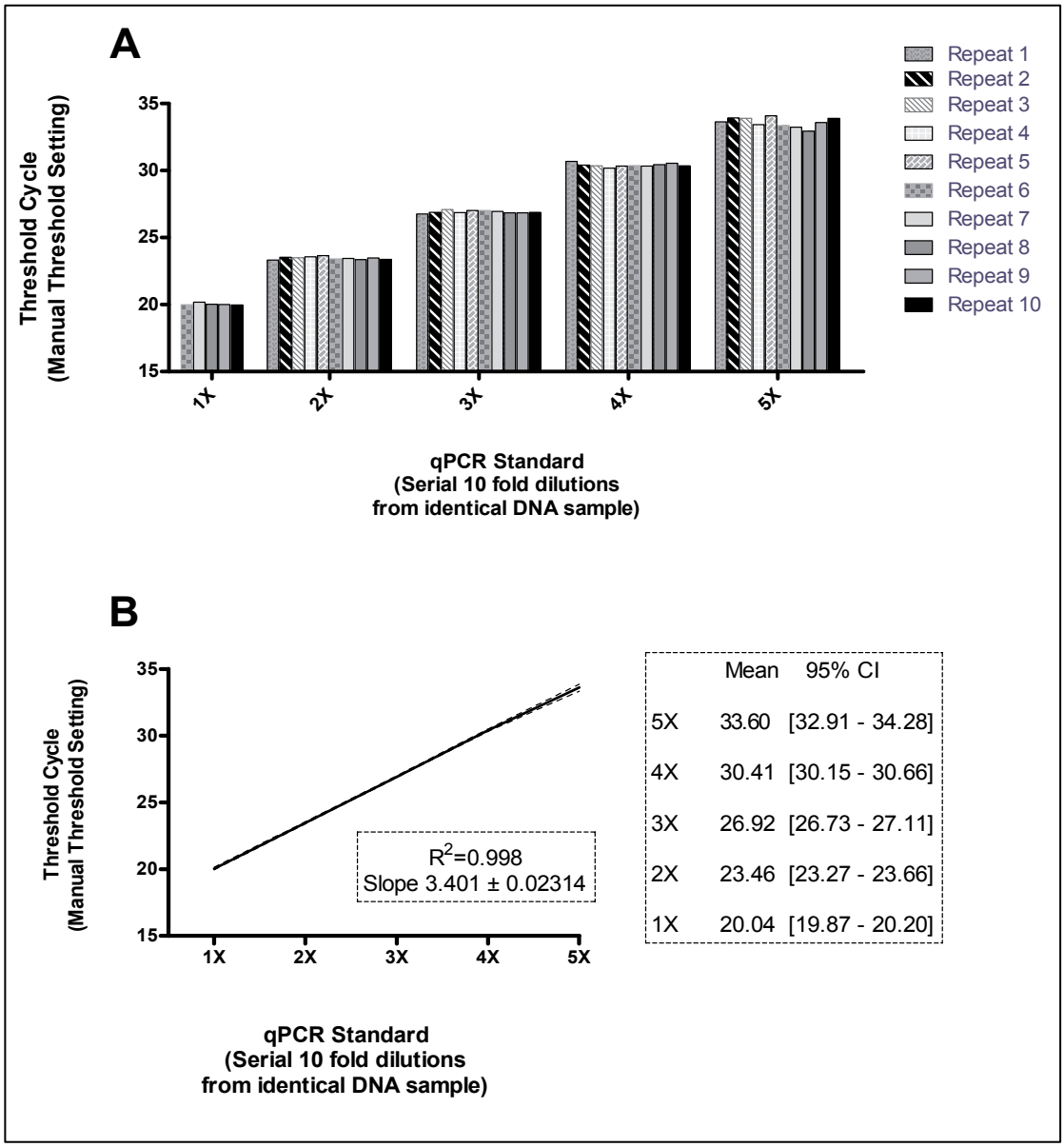


Figure 3.24: The effect of manual threshold setting on qPCR variability
10 consecutive qPCR experimental quantifications of identical DNA standards were performed using varied batches of 28S primers and SYBR green reagent. The threshold for each experiment was manually adjusted to 500 RFU and the threshold cycle for each repeat is shown in (A) and the mean and 95% confidence interval of these repeats in (B).

By plotting the mean and 95% confidence interval for the threshold cycle for each standard in the 10 repeats before and after manual threshold setting, the effect of automatic versus manual threshold setting can be observed. With this adjustment it can be demonstrated that even with different runs, primer batches and SYBR green reagents the 50 standards are quantified highly consistently. Using this approach the variability (95% CI) is reduced to a fraction of a cycle for standards 1 to 4, indicating that the same amount of DNA is quantified with high precision between runs and that cross-run comparisons of unknown experimental samples would be valid

This experiment demonstrates that by using common qPCR standards and manually setting a arbitrary threshold the qPCR quantification step of this assay can be made highly consistent and allow accurate comparison between samples between runs – essential if this assay is to be used clinically. As the acceptable range of standards is defined, they can also be used to reassure that the qPCR has progressed satisfactorily and as a marker of success or failure. This experiment highlights the importance of standardising qPCR protocols and analysis for between experimental comparisons, both in this assay and for the published qPCR literature in general. If subtle differences, such as the variable automatic threshold set by the analysis software, can introduce noise into an experiment, this emphasises the need for the MIQE guideline approach to ensure all parameters are published to allow reproduction of important experimental findings

To demonstrate the effect of manual vs. automatic threshold setting on the result of the DIP-qPCR, rather than on the quantification of the qPCR standards, a sample of DNA from immunoprecipitation of PBMC extracted DNA was quantified on 4 consecutive occasions using the master standards and variable compared to arbitrary 500 RFU threshold setting (Figure 3.25).

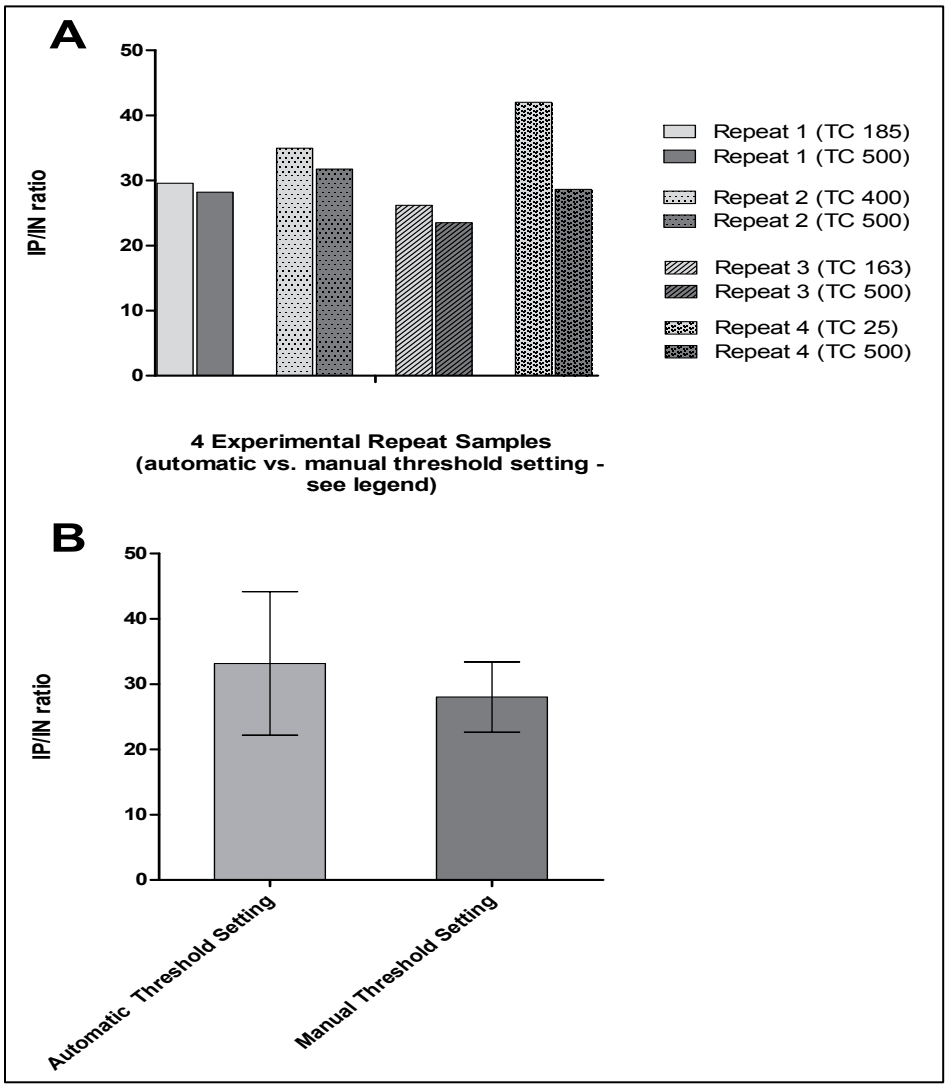


Figure 3.25: Demonstrating the effect of manual threshold setting on DIP-qPCR assay variability.
Identical DIP-qPCR samples are analysed on 4 independent occasions using master qPCR standards. The data is analysed by automatic threshold and using a manual set threshold of 500RFU. (A) shows the changes to the IP/IN ratio calculated with reference to an untreated sample for 4 repeat experiments (light grey – automatic setting vs. dark grey – manual threshold setting). (B) is the mean and 95% confidence interval of the 4 repeat samples processed using automatic threshold calculation (light grey) and manual threshold setting (dark grey).

This result highlights the importance of manual threshold setting and using master qPCR standards. Using the manual setting approach the 4 repeats are more closely matched (graph A) and the 95% confidence intervals are narrower than with automatic threshold setting (graph B). To reduce qPCR variability this method will be used for the PBMC DIP-chip assay.

3.5 Confirmation of the reduction in assay variability with protocol amendments

Initial attempts at using the DIP-qPCR assay to measure induced oxaliplatin-DNA adducts in multiple independent repeat experiments on donated blood samples from human volunteers revealed considerable heterogeneity in the results, with a CV calculated as between 78% and 143% for 5 repeat DIP-qPCR samples from patient B (section 3.3.3). The heterogeneity shown in initial use of the protocol rendered the experiment unsuitable for use in a clinical study with single attempts at obtaining samples for analysis, especially if the clinical study is then used to predict the requirement or necessity of chemotherapy. As a consequence, the effect of several modifications to the assay protocol were investigated and implemented.

To measure the effect of the modified protocol, 3 independent experiments were repeated with PBMC from patient B. They were retested to measure induced adduct levels at a 1mM oxaliplatin dose, comparable to the experiment in the initial CV calculations. A further experiment using a sample of oxaliplatin treated PBMC from a single donor, treated at the same time with the DNA samples extracted and stored in parallel, was also processed in parallel in a series of replicates, and repeated on 3 independent occasions. These two sets of samples were used to recalculate the assay CV obtained with the amended assay protocol (Figure 3.26 and Figure 3.27)

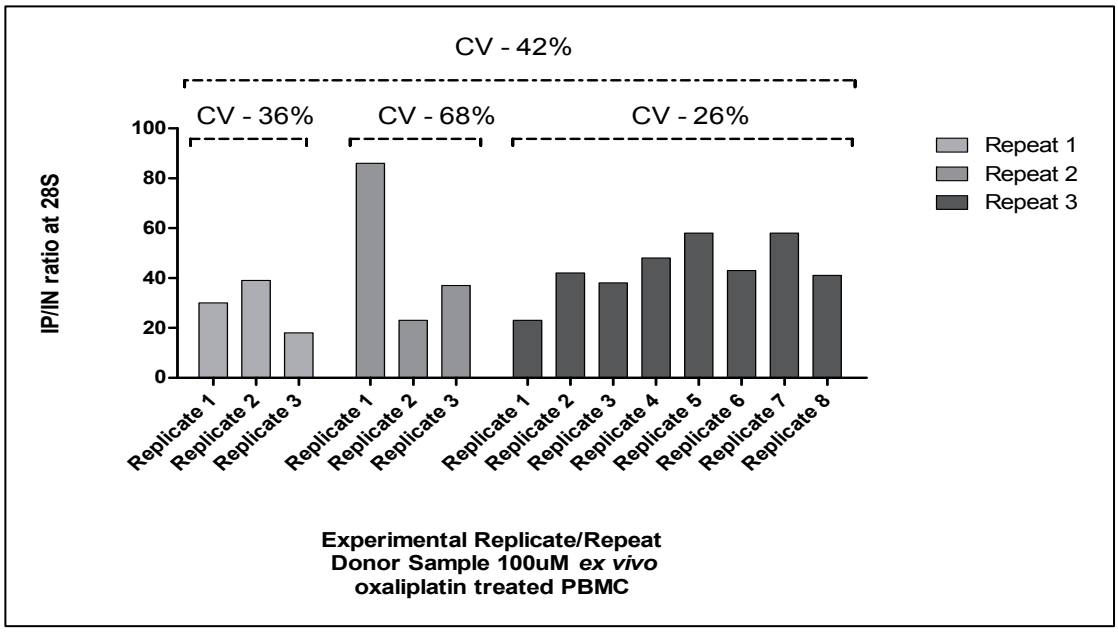


Figure 3.26: The effect of DIP assay protocol amendments on assay variability.
PBMC from a single donor were treated with 100μM oxaliplatin in parallel, and DNA was extracted. Samples were processed in separate experimental replicates on three independent occasions. The IP/IN ratio, normalised to an untreated sample, was calculated by qPCR. Overall the CV is 42% with all results, and between 26% and 68% with a series of replicates run during the same immunoprecipitation.

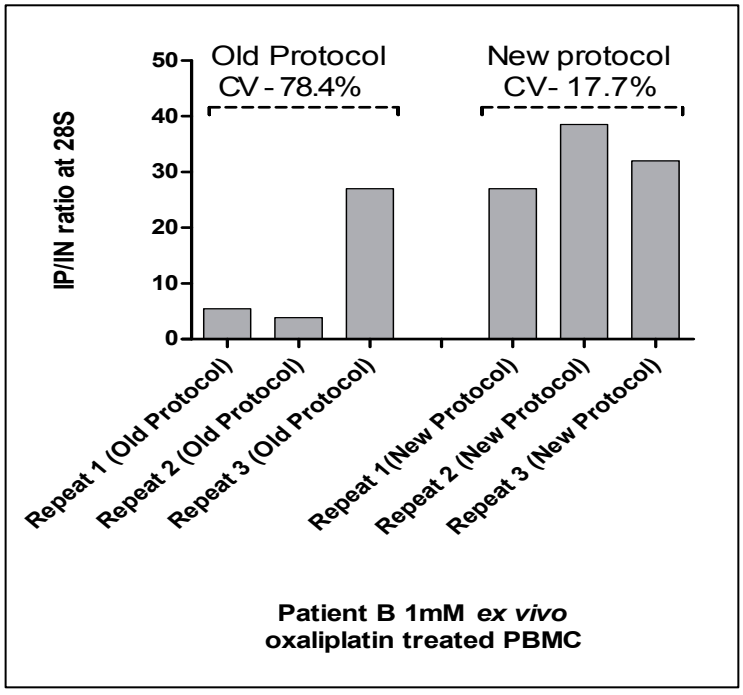


Figure 3.27: A comparison the effect of old and new DIP-qPCR assay protocols on assay variability.
A direct comparison between the old DIP-qPCR and the modified human PBMC DIP-qPCR protocol is shown. PBMC from patient B were treated with 1mM oxaliplatin in three independent repeats using the original protocol, with a coefficient of variance of 78.4%. 3 independent repeat experiments using the modified protocol result in CV of 17.7% (comparable data between old and new protocol for patient A was not available due to technical failure of two samples with the old protocol)

These result show that the DIP-qPCR with the modified human DIP PBMC protocol is much more reproducible for single use in clinical samples, with a CV of 17.7% to 42%, compared to 78.4% to 143% in several experiments with the initial protocol. The direct comparison between experiments (Figure 3.27) is only available for patient B, as insufficient data is available to calculate the CV for a comparable dose point for patient A due to technical failure – see section 3.3.3. This result confirms the value of the experiments and modifications developed as reported in this chapter. These changes significantly reduce the variability seen in the measurement of induced oxaliplatin-DNA adducts in clinical samples used in this study, as demonstrated by a significant reduction in the coefficient of variance of the assay in these experiments.

3.6 Summary of the DIP-chip protocol amendments determined in the experiments described in this chapter

In summary, several changes to the DIP-chip experimental protocol have been examined and collectively result in an improvement in the reproducibility of the DIP phase of the assay from a CV of 78.4% to a CV of 17.7%. The amendments required to the DIP-chip protocol are listed below, and have been used for the human PBMC DIP-chip protocol described in the methods chapter 2, section 2.1.

- A protocol for the separation of PBMC has been developed (section 3.3.1)

- Modification of the oxaliplatin dose and exposure time is required. This results in higher immunoprecipitation efficiency and better discrimination between dose levels (section 3.4.2)

- The need for careful storage and making fresh drug with each experiment was determined (section 3.4.4)

- The requirement to split PBMC samples into two equal aliquots for DNA extraction was demonstrated. Eluting extracted DNA twice in a second volume of 100μL was also beneficial to increase DNA yield (section 3.4.5.2)

- 30 cycles of sonication rather than 24 is necessary as a consequence of the increase in extracted DNA sample volume (section 3.4.6)

- The requirement for a second purification step for DNA post-sonication was confirmed (sections 3.4.5.3 and 3.4.5.4)

- Each qPCR requires the use of a set of master standards and manual threshold setting at 500RFU to ensure run-to-run comparability and consistency

3.7 Discussion

The formation of platinum-DNA adducts and removal of adducts by DNA repair, particularly nucleotide excision repair (NER) are the key events governing the response and toxicity in patients of the platinum family of chemotherapy agents (Bowden 2014). As such, many studies have investigated the link between induced platinum-DNA adduct levels, adduct repair, and the clinical response from treatment, as discussed in several sections of this thesis. A large retrospective study demonstrating a correlation between response to platinum chemotherapy and germline DNA repair capacity in peripheral blood cells in patients with non-small cell lung cancer is the strongest indication that these approaches could be successful (Olaussen et al. 2012; Wang et al. 2011). So far, however, the results of many of these types of studies are contradictory, conflicting or underpowered to demonstrate a significant correlation between these factors. A single measure of a NER factor (either protein or mRNA levels), overall platinum-DNA adduct level, or a single measure of adduct repair lacks

specificity and sensitivity, limiting the clinical utility of these types of assay (Bowden 2014; Olaussen et al. 2012; Wang et al. 2011). The DIP-chip assay offers scope through analysis of platinum-DNA adduct levels and patterns at high resolution in sections of the genome to improve on the predictive power of these techniques.

The DIP-chip assay can be considered in two phases – DNA immunoprecipitation (DIP), confirmed by qPCR analysis of platinum-DNA adduct levels, followed by genome-resolution analysis with microarrays (chip). It is the former that is the main concern of this chapter, although the validity of the microarray analysis is based on the assumption that the DIP stage is robust, reliable and reproducible.

The experiments described in this chapter demonstrate that, with the protocol modifications described, it is possible to treat PBMC with oxaliplatin *ex vivo* and obtain high quality DNA in sufficient amounts for immunoprecipitation and downstream applications. Given the limitations on clinical samples from patients undergoing chemotherapy that this assay is intended for use with, these modifications and amendments to the DNA extraction protocol are an important improvement on this initial DIP-chip protocol.

This study also demonstrates that the DIP protocol, as initially developed in human cell culture models, results in a high variability when used repeatedly on the same human clinical PBMC samples. The experiments discussed in this chapter describe and document the process through which the DIP assay is deconstructed into its component parts, and how each stage can be assessed and amended to reduce the overall variation in final result. A reduction in coefficient of variance, a well-used laboratory marker of assay quality, from 78% to 17% with the protocol modifications described demonstrates the benefits of this stepwise approach. This improvement confirms the DIP phase of the assay is producing relatively reproducible pools of DNA fragments in the IP and IN samples following immunoprecipitation and processing.

For use as a single locus platinum-DNA adduct level assay ideally a CV of <10% would be achieved before a form of DIP-qPCR assay could be used on patient samples. This, however, is not the intended use of these DIP samples, as they will be applied to microarrays for genome-scale platinum-DNA adduct

analysis. The large datasets generated through this approach can be used to identify and remove the effect of variability in the assay due to experimental noise, potentially making further improvements in consistency in the DIP phase of the assay unnecessary – this concept will be explored further in chapter 5.

3.8 Conclusions

Using the experiments outlined in this chapter, the DIP assay protocol has been thoroughly examined, amended and optimised for use in human clinical blood samples, with a significant improvement in the validity of the results generated by this technique. Although further improvements are desirable to further reduce the variability if used as a stand-alone DIP-qPCR based assay, this is not the intention of this project. It is unlikely that a single genetic locus DIP-qPCR approach would offer much benefit above the 'platinum-adduct per unit DNA' approach used in many of the clinical studies already conducted.

In view of this, the additional strength of the experiments outlined in this chapter are that by reducing the DIP variability they improve the validity of the high-sensitivity platinum-DNA adduct patterns generated when the improved DIP protocol is coupled with microarray based genome-scale techniques. These experiments will be the focus of the following chapters.

Chapter 4 Microarray analysis of DIP-chip samples

4.1 Introduction

The modified DNA immunoprecipitation (DIP) assay, discussed in chapter 3, results in reproducible production of samples of fragmented DNA, when extracted from clinical samples - in this case *ex vivo* treated peripheral blood mononuclear cells (PBMC). These fragments contain a reasonably consistent level of oxaliplatin induced DNA adducts in samples that are treated with the same dose of oxaliplatin, as measured by qPCR at the 28S loci and compared to a non-immunoprecipitated 'input' reference and an untreated sample.

In itself, this DIP-qPCR assay could be a useful tool to study levels of induced platinum-DNA adducts. It could potentially be used as an assay to predict response or toxicity to platinum drugs, in a similar approach to the technique used by Wang (2011), when measuring the germline DNA repair capacity of peripheral blood mononuclear cells of non-small cell lung cancer patients undergoing cisplatin chemotherapy (Wang et al. 2011), as the output resulting from the DIP-qPCR analysis is a quantification of induced platinum-DNA adducts at a single, or limited number of genetic loci. It is also a similar quantification to the 'platinum-DNA adducts per unit of DNA' result produced by several other assays that have been used to measure platinum-DNA damage, as discussed in the introductory discussion in the previous chapter, section 3.1. Unfortunately, these assays are of not powerful enough for clinical use, as they lack sensitivity and specificity, resulting in a limited capability to discriminate between different clinical outcomes (Bowden 2014).

One reason for the lack of predictive power of 'platinum-DNA adducts per unit of DNA' assays may be variation in the presence of platinum-DNA adducts at different regions of the genome. A variable level of induction of platinum-DNA adducts *in vivo* has been demonstrated in human cell culture models using our DIP-chip technique (Powell 2014). Additionally, in different tissues nucleotide excision repair (NER) operates at variable rates at distinct sections of the genome (Furuta et al. 2002; Nouspikel 2009). As a consequence of variable platinum-DNA damage induction, combined with differences in NER rate, the

level of platinum-DNA adducts potentially varies significantly throughout the genome during and in the period after chemotherapy treatment. Crude assays of 'platinum-DNA adducts per unit' are an average measure of adduct levels, and are unable to detect differences in induced adduct patterns and repair rates at specific regions. In this context, patterns of platinum-DNA adducts and platinum-DNA adduct repair at key genomic regions (for example, the genes which are most actively transcribed) may be a more sensitive and specific marker for platinum drug response or toxicity than the single measure of overall genomic platinum-DNA adducts currently available.

There is, therefore, the possibility that an approach to measure patterns of platinum-DNA adducts at high-resolution throughout the genome could be a significant improvement in patient stratification technology compared to single locus or single measure assays. With the modified human PBMC DIP protocol developed in chapter 3 coupled to microarray chip analysis techniques, high-resolution genomic-scale information for DNA damage patterns can be obtained. By analysing the immunoprecipitated (IP) and input (IN) control sample generated by the DIP assay with microarray technology it is possible to compare the relative amounts of DNA fragments in paired IP and IN samples at thousands of genetic loci simultaneously, resulting in a genomic pattern of the levels of platinum-DNA adducts (for an example see figure 4.2).

The experiments discussed in this chapter used custom designed DNA microarrays to measure the relative fragment frequency at 44,000 loci over a 5Mb region of chromosome 17. These parameters are chosen as they have been established as the standard human microarray parameters used in our laboratory. The assay characteristics and performance has been determined in previous experiments using the DIP-chip technique to measure cisplatin and oxaliplatin damage in cell culture models (Powell 2014), and demonstrate high-resolution DNA damage patterns.

The techniques, results and discussions presented both in this chapter and the next chapter (chapter 5) can be considered as one section of this thesis, although each chapter has a different emphasis. As discussed extensively in chapter 3, the translation of a cell culture tool into an assay to analyse human clinical samples requires in-depth investigation, optimisation and modification

of the laboratory workflow and chip protocol to maximise the reproducibility and reliability of the technique for these types of human tissues. Central to this are experiments to determine the ability the DIP-chip assay to generate reproducible and accurate platinum-DNA adduct patterns in repeated samples taken from the same individual, when treated under the same experimental conditions – an issue of minimising variability in the assay. This will be the main focus of the experiments described in this chapter.

Once accurate and reproducible adduct patterns have been produced by the DIP-chip assay, the next issue will be the detection of a biological signal – the differences in platinum-DNA adduct patterns that occur between samples taken from different individuals, when treated under the same experimental conditions. This is the ultimate role of the DIP-chip assay in the context of this thesis and for use in future clinical studies. Determining the capability of the DIP-chip assay to detect inter-individual differences in platinum-induced DNA adduct patterns requires the development of new analytical and bioinformatic tools to detect and report characteristics of biological signals. Additionally, this effort requires experimental evidence, described in this chapter, to quantify and minimise noise in the assay. Along with bioinformatic tools developed to improve the analysis of DIP-chip assay development experiments and for quality control purposes, the bioinformatic methods for the detection of inter-individual differences will be presented in the final section of the following chapter.

4.1.1 Microarray technology

Microarrays have been used as a widely accepted research tool to generate genomic data from ChIP-chip experiments for over a decade (Ren et al. 2000). As discussed in chapter 3, section 3.1.9, examples of this include many types of related experimental approaches to measure protein-DNA binding (Iyer et al. 2001), histone modifications (Kurdistani et al. 2002), and to detect the genomic location of methylated DNA (Weber et al. 2007). Microarrays have also been widely used in other research techniques, including use for gene expression profiling experiments, for the detection of alternative RNA splicing, the study of

microRNA, and for comparative genomic hybridisation and genotyping. A review published in 2006 provides a broad summary of the range of applications to which this technology has been applied (Hoheisel 2006).

Next-generation DNA sequencing technologies are increasingly providing an alternative platform to microarray-based analysis, including for ChIP-chip (Furey 2012; Ho et al. 2011). However, the well-established laboratory workflow, available bioinformatic techniques for data processing and the more manageable size of the genomic-scale data generated make microarrays an ideal platform for development studies and in this effort to translate the DIP-chip assay for use on clinical samples. Once any issues surrounding the translation of the DIP-chip assay for clinical use are addressed and a stable and reliable assay pipeline has been developed, the data generated by the DIP-chip assay can provide a baseline and a model for the future development of DIP-sequencing in this context, especially in light of the rapid and continuing advances in DNA sequencing that are currently occurring (Bahassi and Stambrook 2014).

4.1.2 Microarray structure and layout

In simple terms, a DNA microarray consists of thousands of unique oligonucleotide probes lithographically printed onto a glass slide. Groups of identical oligonucleotides are adjacently situated to form 'features', so called as these groups of probes, when hybridised to a fluorescently labelled DNA sample, can be scanned and detected as an individual genomic 'feature' for analysis. In the experiments discussed in this thesis the microarrays used are custom designed '4 x 44K' feature arrays obtained from Agilent Technologies Inc. As such, they are made up of 4 microarrays per slide, each with 45,219 individual features, the majority of which are available to generate experimental data (Technologies 2008). The remaining features are present as controls and for localisation and identification during scanning (Figure 4.1).

The application of two pools of DNA to the microarray slide, each with a different fluorescent label, under appropriate conditions results in hybridisation between complementary fragments to the corresponding feature(s). The

relative hybridisation between fragments from each DNA sample can be detected using a fluorescent laser scanner to excite, and subsequently detect, the relative fluorescence present after hybridisation at each microarray feature. Information on the spatial location and the fluorescence intensity can then be linked back to the feature's genomic location and to the known probe sequence. In this way, the relative proportions of fragments in the two DNA samples applied can be mapped back to the sequence of the DNA region of interest, identifying the amounts of material in each pooled sample at multiple loci along the genome.

In this DIP-chip assay, the two samples applied to the microarray are the paired IP and IN sample from one experimental condition per microarray. Microarray interrogation consequently generates a numerical value for the IP/IN ratio. As previously discussed, the IP/IN ratio is a measure of the immunoprecipitation efficiency, and is a method of quantifying, normalising and comparing platinum-DNA adduct levels at different loci. At each feature analysed, a higher IP/IN value indicates that a greater proportion of fragments are immunoprecipitated during the DIP relative to the input control, indicating that there is a higher level of platinum-DNA adducts at this location. Mapping the IP/IN ratio at each genomic location of every feature allows the generation of a high-resolution platinum-DNA adduct patterns along the genome (Figure 4.2)

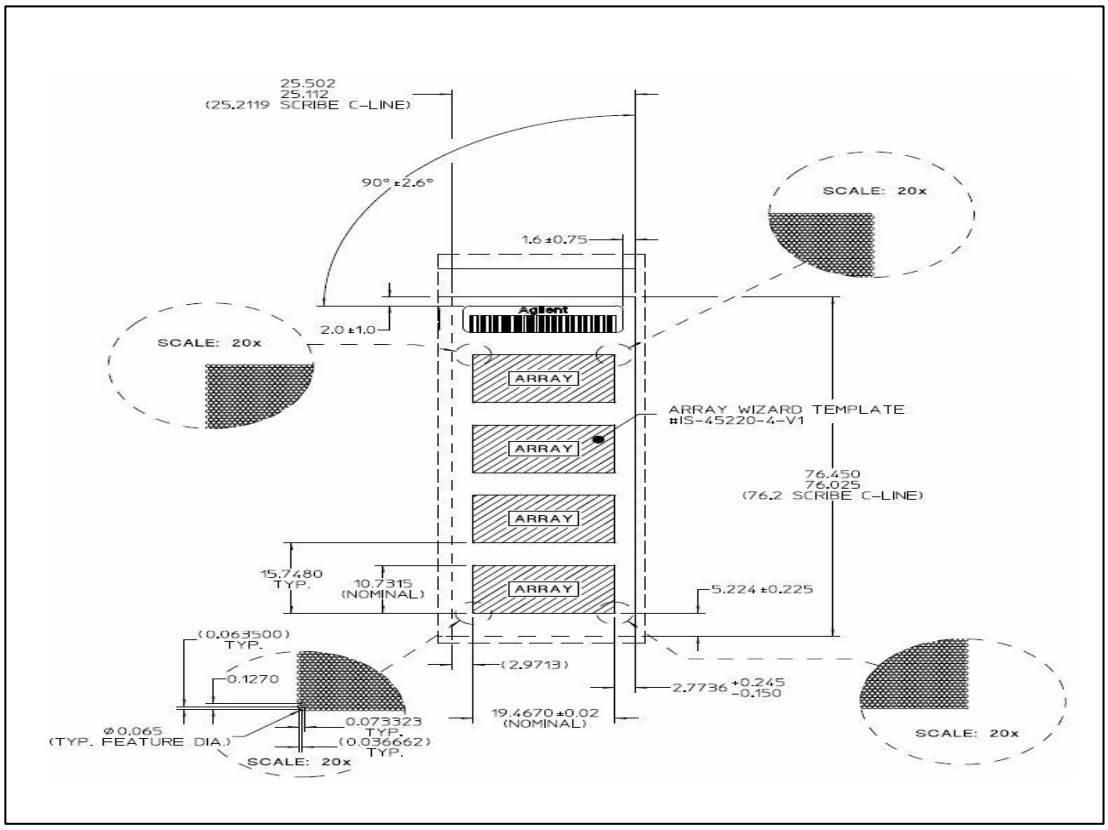


Figure 4.1: The layout of an Agilent 4x 44k microarray.
The layout of an Agilent 4x 44k microarray (Agilent Technologies Inc. 2007) (http://www.chem.agilent.com/Library/technicaloverviews/Public/G450290001_MicroarrayFormat.pdf)

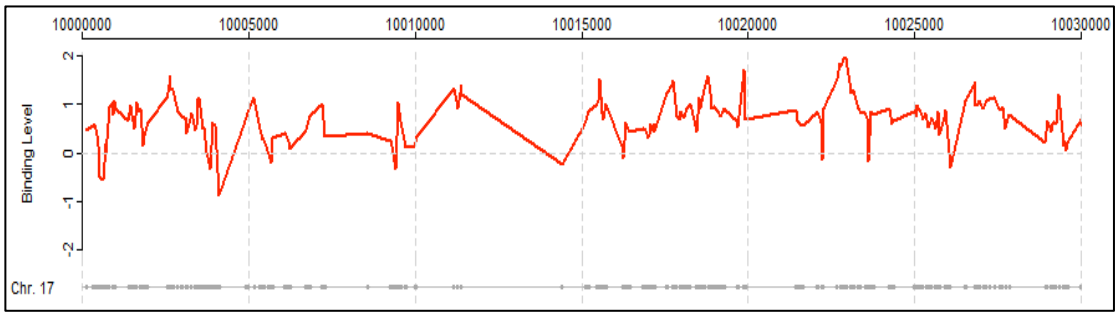


Figure 4.2: An example of a oxaliplatin-DNA adduct pattern plot.
The log2 IP/IN ratio is plotted on the y-axis against the location of each probe along the x-axis. The genomic location of each probe is shown on the lower x-axis as a grey dot, and the position along chromosome 17 is displayed on the upper x-axis. Peaks and troughs can be seen corresponding to putative areas of increased and decreased oxaliplatin-DNA adduct formation.
The use of the log2 of the IP/IN ratio for the y-axis allows easier visualisation when there is a large difference between the IP and IN fluorescence values, and allows equal representation of reciprocal ratios (i.e. when the probe IP value is twice or half of the IN value the distance from the centre of the data will be equal, but with opposing direction). A log2 binding level of 0 equates to an IP/IN ratio of 1, as $2^0=1$.

4.1.3 Overview of the 'chip' experiment

The experimental workflow of the DIP-chip assay can be conceptualised in two distinct phases; an initial DNA immunoprecipitation (DIP), followed by genomic scale analysis using microarray technology (chip). The initial DNA immunoprecipitation was extensively discussed in chapter 3. For each experimental condition the DIP results in two pools of DNA fragments, the immunoprecipitated (IP) sample and the control input (IN) sample. The IN sample is taken from the pre-immunoprecipitated pool of purified, sonicated DNA and represents the background variation in DNA quality and initial fragment frequency. As it is processed in parallel to the IP sample (except for the immunoprecipitation stage) it also acts as an internal control for non-biological variability in the assay technique, especially in sample handling, DNA labelling and in PCR amplification steps.

In the chip aspect of the assay there are four main laboratory preparation phases (figure 4.3): DNA amplification, sample labelling, microarray hybridisation, and scanning and data extraction. There is a fifth step - the extracted data then require bioinformatic processing to generate and interpret platinum-DNA adduct patterns.

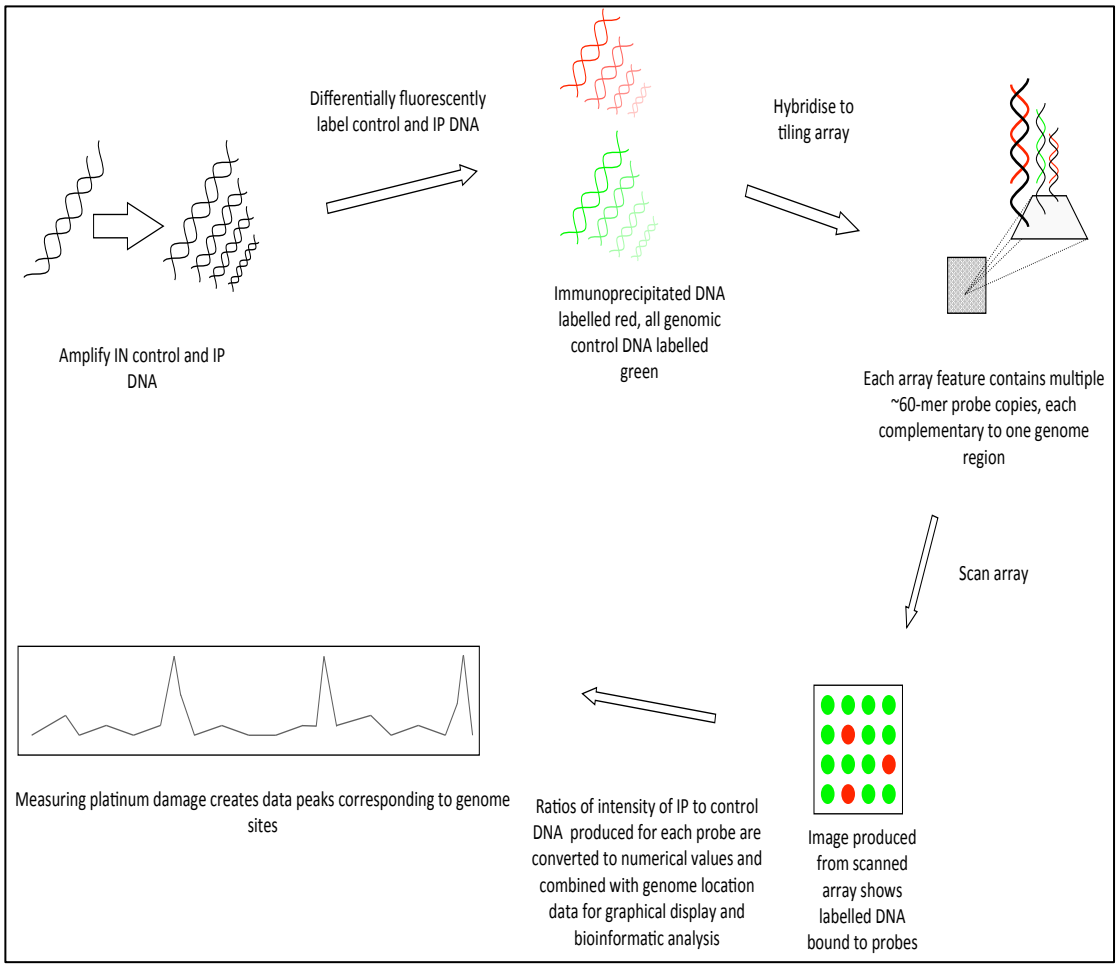


Figure 4.3: The chip phase of the DIP-chip assay experimental pathway.
Following the immunoprecipitation purification the immunoprecipitated (IP) sample and an input (IN) control sample are amplified, fluorescently labelled, and hybridised to a custom microarray prior to data extraction and bioinformatic processing to generate a high resolution genome scale damage pattern trace

4.1.3.1 DNA amplification

To prepare the IP and IN DNA fragment pools for microarrays analysis sufficient DNA must be available. For an Agilent Technologies 4x microarray format, each array requires a minimum of 500ng of DNA per sample. Although the IN sample occasionally contains enough DNA (depending on the experimental conditions), as a result of the efficiency and process of immunoprecipitation the IP sample DNA amount is typically 10-100 fold lower, so the recovered material following immunoprecipitation is insufficient to use directly on the array. Since the IP sample must be amplified, we dilute the IN sample to a similar concentration in order to use the same amplification process

and parameters on both the IP and the diluted IN samples. In this way the IP and IN samples have, as much as is possible, been processed in the same fashion, with the intention of maintaining the IN as an internal control.

Several methods of amplification have been used for this purpose in published ChIP-chip experiments. The two typical ways of amplifying DNA in this context are though ligation-mediated PCR (LM-PCR), as described by (Mueller and Wold 1989) or through whole genome (WGA-PCR) amplification (O'Geen et al. 2006).

As used extensively in our laboratory, LM-PCR requires initial DNA manipulation to ensure all DNA fragments have a blunt end, using T4 DNA polymerase (NEB) to act as a 3' to 5' endonuclease and to end-fill 3' recessed nucleotides. Purified fragments are then incubated with T4 DNA ligase (NEB) and a common linker DNA sequence. Ligation to the DNA fragments of the common linker subsequently allows PCR amplification using primers to the common sequence. This process is effective to adequately amplify the DNA, and variations of LM-PCR have been widely used in many ChIP-chip experiments (O'Geen et al. 2006).

Whole genome amplification (WGA-PCR) is also widely used to amplify DNA prior to microarray (O'Geen et al. 2006). In the WGA process typically used in our laboratory the PCR based amplification method uses a random priming approach – a primer with a random 3' sequence and defined 5' sequence of DNA is annealed randomly to the DNA in the mixed pool of DNA fragments. This is followed by a PCR amplification using primers to the defined 5' random primer sequence. The WGA-PCR kit used in these experiments and in the DIP-chip assay development is the WGA2 kit (Sigma UK). The WGA2 kit a two stage PCR, with proprietary universal primers in an initial brief PCR reaction with the pool of fragmented DNA, resulting in PCR-amplifiable DNA library fragments flanked by universal priming sites. Primers to the universal flanking sequence are then used in a second, more extensive PCR amplification (Sigma).

WGA-PCR has been compared directly to LM-PCR when used to amplify ChIP-chip DNA prior to microarray, in an experiment investigating the genomic location Oct4 protein-DNA binding sites in the human genome (O'Geen et al. 2006). A comparison of the technologies in this circumstance showed that WGA-

PCR produced superior signal to noise ratios following hybridisation compared to LM-PCR.

Of particular importance, when translating this technique for use in detecting subtle differences between experimental conditions and between individuals as in this thesis, is that PCR-based DNA amplification for library preparation for sequencing or microarray analysis is known to be a potential source of significant experimental bias (Dabney and Meyer 2012; Hasmats et al. 2014; Quail et al. 2012; van Dijk et al. 2014), and WGA-PCR amplification bias due to preferential annealing of universal or random primers with DNA fragments of differing characteristics is a complication that can have statistically significant effects on experimental results (Pinard et al. 2006). For example, in one study of WGA-PCR based amplification bias, copy number at individual loci following amplification varied between 100-10,000 fold (Dean et al. 2002) and during WGA-PCR based amplification of heterozygous alleles for pre-implantation genetic diagnosis, preferential WGA-PCR amplification of one allele compared to the other has been described (Paunio et al. 1996). Evidently, the risk of amplification biases in the experiments discussed in this chapter must be addressed.

4.1.3.2 DNA labelling

Once sufficient DNA has been amplified and purified the samples are quantified by spectrophotometry using a Nanodrop-1000 spectrophotometer. After adjustment to identical DNA amounts in the paired IP and IN samples each is labelled with a fluorescent marker. In our DIP-chip assay this is done using the BioPrime Total Genomic DNA Labelling System (Invitrogen) following the manufacturers protocol. Typically, Alexa-Fluor 5 is used for the IP sample and Alexa-Fluor 3 for the IN sample. The labelled IN sample then produces green fluorescence at 555nm excitation and 565 nm emission, and the Alexa-Fluor 5 labelled IP sample produces red fluorescence at 650nm/670nm excitation and emission respectively. Again, labelling requires a DNA amplification reaction, using the Exo-Klenow fragment (the N-terminal fragment of DNA polymerase I)

to incorporate the labelled nucleotides into newly synthesized DNA fragments, resulting in an eight-fold increase in DNA quantity during the reaction (Invitrogen 2014). Following labelling, both samples are purified and the labelling efficiency and quantity of labelled DNA in each sample can be quantified using the Nanodrop-1000 spectrophotometer.

4.1.3.3 Microarray hybridisation

Prior to hybridisation the IN and IP samples are combined and precipitated into a single DNA pellet, before re-suspension in purified water. The sample is mixed with appropriate buffers, denatured, cooled to 37°C and then applied to a microarray chamber. The oligonucleotide covered glass slide forms the roof of the small chamber. Each individual microarray is sealed and the chamber is gently rotated for 24 hours at 65°C to allow free circulation of the sample over the slide surface, giving maximum opportunity for hybridisation to complementary probes. At the end of the hybridisation period the slide is washed to remove non-specific and weakly bound DNA fragments, the slide is gently dried, kept clean and covered from light in the period prior to scanning.

4.1.3.4 Microarray scanning and processing

The microarray slide is then scanned, using an Agilent Microarray Scanner (model G2505B). Laser light of the desired wavelength is used to excite the relevant fluorophore and any emitted light is measured to a 5-micron resolution. Increasing hybridisation of greater amounts of one fluorophore labelled DNA compared to the paired sample gives a brighter fluorescence of that colour.

4.1.3.5 Initial data extraction and processing

The TIFF file output from the microarray scanner, giving the relative colour and spatial location of each feature, is analysed using Agilent's Feature Extraction (FE) software. In basic terms, each feature is identified by comparison of the spatial location of the feature with respect to a grid aligned to the position of reference features printed on the array. The relative fluorescence is calculated, giving a conversion of the fluorescence intensity at each feature into a numerical value, linked to the location, feature position and sequence information. An additional computation is made by the FE software to calculate each feature intensity value subtracted for the background intensity emanating from the fluorescence of surrounding features. The background corrected IP and IN values are used for the experiments described in this thesis. This dataset forms the FE software output in the format of a tab delimited text file, containing the vital feature identification and location information, feature intensity values, and background subtracted feature intensity data, in addition to multiple fields of technical information.

A second bioinformatic package, Sandcastle, is used to interpret the text file containing the data generated by the FE software. Sandcastle (Software for the Analysis and Normalisation of Data from ChIP-chip AssayS of Two or more Linked Experiments) is a package of tools developed in our laboratory (Bennett 2013), designed for use in R, the programming environment (R Development Core Team 2014). Functions in the Sandcastle package can be used to import the relevant information from the text file into *R* and to ensure the data generated from the microarray meets quality standards. The data can be normalised (if possible and applicable) between replicate data sets, and the package can be used to generate plots to visualise patterns of data and to calculate relationships between microarray datasets.

4.1.3.6 Sandcastle outputs for DIP-chip microarray data analysis

Three analysis tools and data outputs from Sandcastle in particular are relevant to the DIP-chip experiments conducted in this thesis.

4.1.3.6.1 Adduct pattern plots

The pattern of platinum-DNA adducts along the genome can be plotted in what will be referred to as an adduct pattern plot. An example is shown below in figure 4.4. On the x-axis the genomic location is identified above, and the location of the individual probe position to which the IP/IN ratio corresponds is identified below by a grey circle. The adduct pattern is generated by plotting the $\log_2$ ratio of the background corrected IP/IN ratio at each probe position along the section of the genome examined.

On the y-axis the use of the $\log_2$ for the IP/IN ratio allows visualisation of probe IP/IN values when there are large differences between probe IP and IN values, as a consequence of large differences between the IP and IN sample probe hybridisation and fluorescence. In addition, the use of a $\log_2$ scale allows equal visual representation of reciprocal ratios - for example, ratios generated if the IP value is either double or half of the IN value (i.e. if considering two probes, at one the IP value = a*IN value, and at the other probe the IN value = (1/a)*IP value, the distance of each plotted point from the central point of the pattern plot data will be the same, although in opposing direction. For reference, as $2^0=1$, a $\log_2$ binding level of 0 equates to equal probe IP and IN fluorescence intensity values, resulting in an IP/IN ratio of 1. This corresponds to approximately equal fragments in the IP and IN samples at that probe region.

An equal amount of fragments in the IP and IN samples is not present in the pre-amplification DIP samples, because much more DNA is present in the IN sample than the IP sample - only ~1% of fragments are immunoprecipitated (discussed later in section 4.4.5 and table 4.1). Only after DNA amplification and labelling is the total DNA content of the IP and IN samples equal, allowing determination with the microarray of the relative hybridisation at each probe.

By amplifying all IP samples and adjusting to the same final DNA concentration pre-hybridisation, information on the relative immunoprecipitation efficiency between IP samples, previously acquired in the qPCR stage of the DIP phase of the assay and dependent on the overall level of induced adducts, is lost. What remains is the relationship between fragments in the IP and IN sample – the pattern of peaks and troughs – that can be compared between repeat samples from the same individual (an issue of assay noise) and between samples different individuals treated under the same conditions (an issue of assay signal).

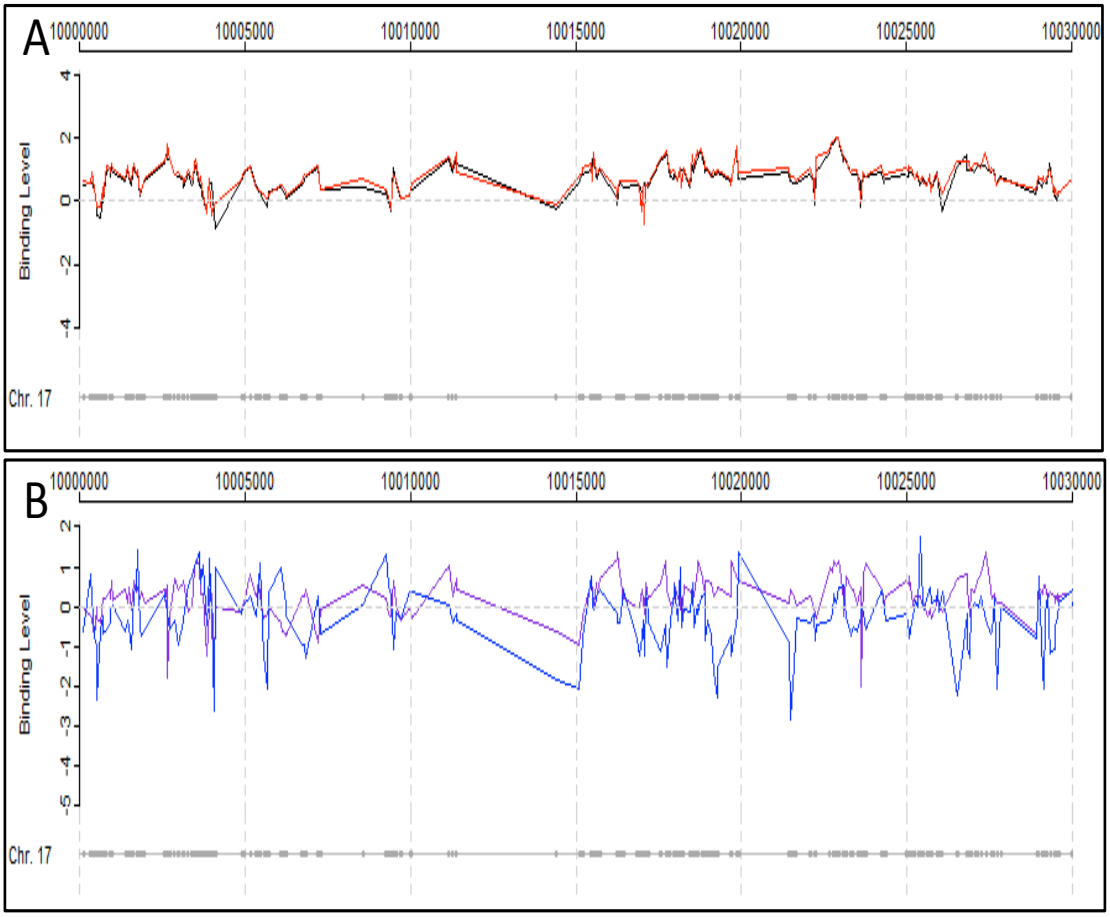


Figure 4.4: Example oxaliplatin-DNA adduct pattern plots.
Oxaliplatin-DNA adduct pattern plots from four experimentally generated microarray DIP-chip datasets over a short section of chromosome 17 are shown. Datasets 1 (black) and 2 (red) are compared in the upper panel (A) and datasets 3 (blue) and 4 (purple) are compared in the lower panel (B). This figure demonstrates how the adduct pattern profiles can be visualised with a adduct pattern plot. It is obvious in this extreme example to identify the two matching oxaliplatin-DNA adduct pattern datasets (1 and 2 in panel A) and two poorly matching datasets (3 and 4 in panel B).

4.1.3.6.2 Spearman's correlation coefficient

The second metric available in R, and used extensively in this thesis, is the Spearman's correlation coefficient, a non-parametric measure of statistical dependence between two (microarray) datasets. The non-parametric statistic is considered more statistically robust and used in preference to the related Pearson's correlation, a parametric statistic, as the distribution and parameters of the experimental datasets and are not normally distributed and cannot be assumed. In most instances, however, the Spearman's and Pearson's correlations are very similar for the DIP-chip datasets used in this thesis (data not shown).

In the context used in this thesis, the Spearman's correlation is a single quantification of how closely two independent microarray adduct pattern datasets match one-another. In each dataset the probes are ranked in order of (log_2) IP/IN ratio value. The correlation between probe rankings are then compared, with a Spearman's correlation of 1 if, in both datasets, the rank orders are identical, -1 if the rankings are exactly inverse, and 0 if there is no association between rank orders (and hence adduct patterns) between datasets. Spearman's correlation values are, in many instances in the following discussion, displayed in a matrix of correlation values between several experimental conditions with colour coding to give a broad overview of the strength of association between several different samples.

An example of the relationship between the adduct pattern plot and Spearman's correlation in this context can be demonstrated using 4 experimentally generated oxaliplatin-DNA adduct pattern datasets (arbitrarily labelled 1 to 4). An oxaliplatin-DNA adduct pattern plot of two of the experimentally generated datasets, 1 and 2, with patterns that match well is displayed above in figure 4.4 (A), and two poorly matching datasets, 3 and 4, in figure 4.4 (B).

In figure 4.5 the probe log_2 IP/IN values from each pair of experimental microarray datasets are plotted against one another in a density scatter plot, with darker blue colour representing a higher density of data-points from the

~44,000 features represented in each dataset. In 4.5 (A) the closely matching adduct pattern, datasets 1 and 2, are plotted, and in 4.5 (B) data from the poorly matching datasets 3 and 4 is displayed. In figure 4.5 (A), where there is close matching of the underlying oxaliplatin-DNA adduct patterns, the data points lie predominantly on the line y=x, i.e. at each probe in dataset 1 the IP/IN value closely matches that in dataset 2. The close match of the data is summarised by the Spearman's correlation of 0.93.

Compare this with figure 4.5 (B), in which the IP/IN values at each probe from the poorly matching oxaliplatin-DNA adduct pattern datasets 3 and 4 are plotted. There is a poor relationship between the IP/IN values at each probe, and the data shows no obvious relationship along the line y=x. This information - the lack of relationship between IP/IN values at each probe in each dataset - can be condensed into the single Spearman's correlation value of 0.07.

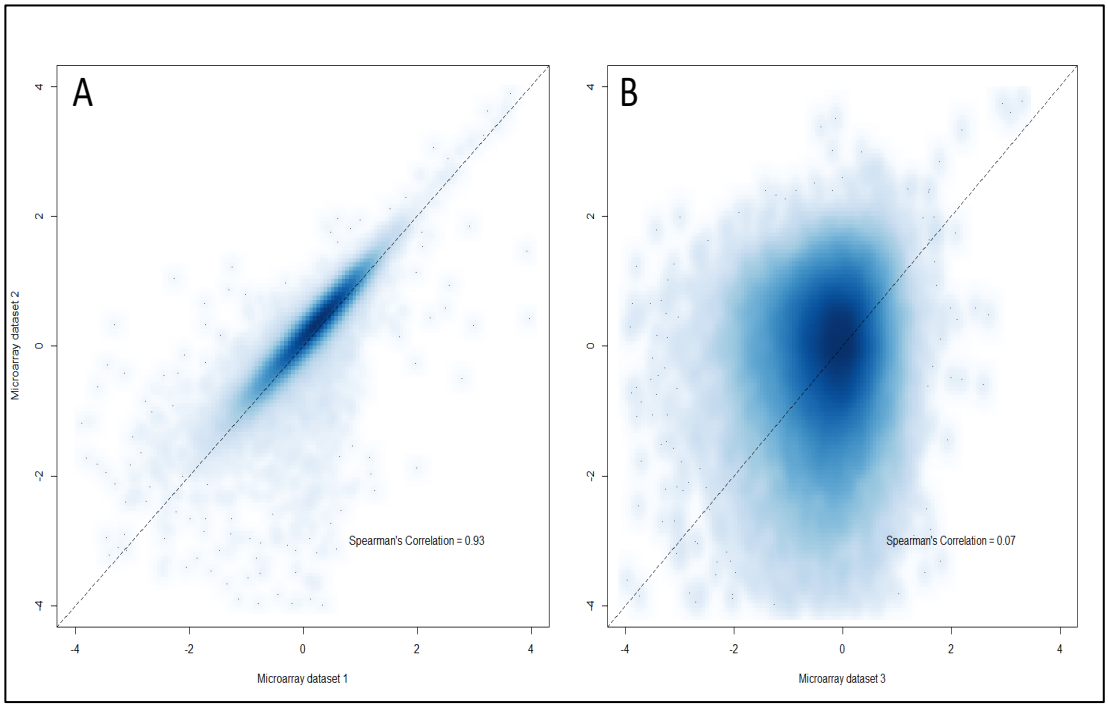


Figure 4.5: Example density scatter plots.
The four experimentally generated oxaliplatin-DNA adduct pattern microarray datasets used in figure 5.4 are compared in density scatter plots in this figure. In A, the log2 IP/IN value at each probe is plotted in two closely matching datasets (1 and 2, also shown in figure 5.4A). Panel B shows a plot of the IP/IN values of datasets 3 and 4 (corresponding to poorly matching patterns and the datasets in figure 5.4 B). The closely matching datasets 1 and 2 (in plot A) cluster along the line y=x, and have a Spearman's correlation of 0.93. The poorly matching datasets 3 and 4 in panel B demonstrate no obvious relationship, in agreement with Spearman's correlation value of 0.07 and the poorly matching adduct patterns in figure 5.4B

4.1.3.6.3 Mathematically generated predicted adduct profiles

A third Sandcastle tool, used extensively in the following sections, is the ability to generate a theoretical dataset representing a predicted platinum-DNA adduct pattern. As discussed in chapter 1, section 1.2.6, both cisplatin and oxaliplatin form platinum-DNA intrastrand adducts at the same ratio between different combinations of dinucleotides – 65% between adjacent guanines (GpG), 25% between adjacent adenine-guanine dinucleotides (ApG), and 5-10% between two guanines separated by an unspecified third nucleotide (GpNpG) (Kozelka 2009). Using this information it is possible to mathematically estimate the relative probability of adduct formation at each probe location, based on the predicted affinity of platinum-DNA adduct for the sequence of each probe and the adjacent DNA sequence.

Notably, the Sandcastle platinum predicted profile is based only on DNA sequence information. The studies that elucidated the dinucleotide specificity of cisplatin were based on studies in naked, purified DNA and may not fully reflect the adduct profile that occurs *in vivo* (chapter 1 section 1.2.6). For example, the model does not take account of the effect of higher order chromatin structure on adduct formation, or the preference for the formation of adducts at nucleosome linker regions compared to the nucleosome core (see chapter 1, section 1.2.8). As such, the model is an approximation of a true platinum-DNA adduct pattern.

Even with these caveats, the Sandcastle platinum prediction profile appears to correlate positively with platinum adduct patterns experimentally generated with *in vivo* cell culture models, with a Spearman's correlation coefficient of 0.76 between predicted and experimental microarray data (figure 4.6 (Powell 2014)). In the context of this thesis, the mathematically generated platinum predicted profile dataset will be compared to experimentally generated microarray datasets in the adduct pattern plot and by using the Spearman's correlation to give an single metric of how experimental data matches this independent reference data.

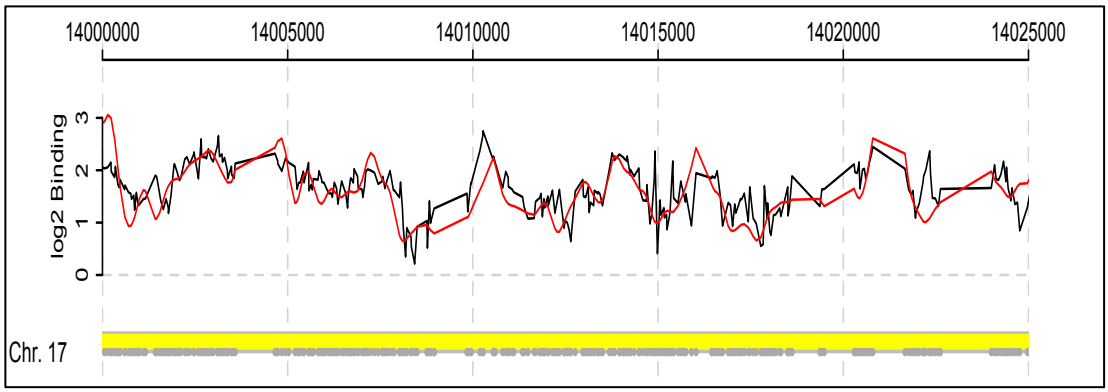


Figure 4.6: An example cisplatin-DNA adduct pattern plot and predicted adduct profile. An adduct pattern plot showing an experimentally generated cisplatin adduct profile (black) over a short section of chromosome 17. The theoretical 'platinum predication profile', mathematically generated to reflect the preference for the formation of platinum-DNA adducts at specific dinucleotide sequences is also displayed (red). The Spearman's correlation between experimental and theoretic datasets is 0.76 in this section, confirming that the patterns match closely (Data from (Powell 2014)).

4.2 Aims of this chapter

The DNA damage microarray technique developed in our laboratory has been translated, initially from a research tool to investigate protein binding in yeast, into a DIP-chip method of detecting DNA damage at high resolution through sections of the genome from UV irradiation in yeast (Teng et al. 2010). More recent work has adapted the technique into a method that can be used for detecting platinum-DNA adducts in cultured human cells (Powell 2014).

The experiments described in the previous chapter demonstrate several steps that were required to translate the DNA immunoprecipitation stage of the DIP-chip assay into a tool capable of reliably measuring oxaliplatin-DNA adducts in *ex vivo* oxaliplatin treated clinical samples at a limited number of 28S loci by qPCR. This chapter concerns the next step in this process - the translation of the full DIP-chip technique into a single-use assay for the generation and analysis of patterns of platinum-DNA adducts in the same human clinical samples. This requires a thorough study of the performance of the assay in this new tissue, and the modification and optimisation of the assay to generate platinum-DNA adduct profiles in a reliable and reproducible output, a critical step in adapting this technology for use in future clinical studies.

The distinction between two specific applications of the DIP-chip assay during development and validation studies is central to this and the next chapter. Firstly, experiments are required to assess the ability of the DIP-chip assay to reproducibly generate oxaliplatin-DNA adduct patterns from clinical samples taken repeatedly from the same individual and treated in the same experimental conditions – an issue of assay 'noise'. This differs from developing and determining the capability of the assay, and associated bioinformatic analysis tools, to detect 'signals' by comparing oxaliplatin-DNA adduct patterns generated between two different individuals under the same experimental conditions. Detection of biological signals with the DIP-chip assay is the ultimate aim of this thesis, and is essential for future clinical studies.

The ability to reliably detect a biological signal - to discriminate between adduct patterns from different individuals - is partially dependent on the noise in the assay. The assay must reproducibly generate adduct patterns in repeated samples from the same individual (i.e. have a low noise profile) before comparisons between different individuals can be trusted and used to guide treatment decisions. The concept of the relationship between signal and noise in the DIP-chip assay – the signal to noise ratio – is central to this and the next chapter.

The specific aims of this chapter are to translate the chip assay for use in analysing clinical samples, and to explore the experimental and laboratory issues that arise from this transition from a cultured cell model, particularly when repeatedly generating oxaliplatin-DNA adduct patterns from the same individual. As will be shown, the transfer of this technology into clinical samples results in significant variability between adduct patterns from the same individual, leading to experimental and technical challenges that need to be overcome before the technique can be used reliably in the clinical setting.

The experiments and results described in this chapter are also extensively referred to on many occasions in chapter 5 for two main reasons. Firstly, during the laboratory phase of the translation of this technique it will become evident that new bioinformatic tools are required to analyse the data in more depth than are currently available in the standard laboratory ChIP/DIP-chip Sandcastle analysis package. In combination with experimental approaches used below,

chapter 5 describes and documents the development of new bioinformatic approaches required to resolve some of the challenges that arise in the experiments conducted in this chapter.

Secondly, in chapter 5, bioinformatic tools will be developed to detect signals, the differences between PBMC from different individuals treated in the same experimental conditions – the goal of this assay. This relies on the concept of the signal to noise ratio in the DIP-chip assay, and so in many instances the developments and discussions conducted during this phase of the thesis, to be presented in chapter 5, also refer back extensively to the experimental development and validation evidence presented in this chapter.

4.2.1 Methods

The methods and laboratory protocols used for the microarray phase of the DIP-chip assay are described in chapter 2, section 2.2. When applicable, a brief overview will follow at stages in the subsequent discussion.

4.2.2 Overview of 'chip' assay laboratory workflow

The experiments conducted in this chapter continue to use the methods and techniques used extensively in chapter 3 to generate DIP immunoprecipitated DNA. Further processing of the post-qPCR IP and IN samples for amplification, DNA labelling, microarray hybridisation and data analysis are the main focus of this chapter.

Briefly, the following experimental approach was used. Peripheral blood mononuclear cells (PBMC) samples used in these experiments were harvested from blood samples donated by two volunteers (Patient A and B) or from blood samples obtained from the Welsh Blood Service after appropriate ethical permissions were obtained (reference WBS Ad-Hoc-008-12). Following venesection, PBMC were harvested, washed, counted and subsequently 5 million cells per experimental condition were cultured at 37°C, 5% CO_2 for 30 minutes in RPMI 1640 media containing serum, L-glutamine and antibiotic, before exposure

to oxaliplatin for 24 hours by the addition of small volumes of concentrated drug to the culture medium. The doses of oxaliplatin used were chosen following detailed review in the experiments in the previous chapter, and covered a dose range of 10μM (considered a 'physiological' dose point), 100μM and 1000μM, acknowledging the potential for an increased rate of adduct formation, cell death and DNA degradation previously demonstrated at these doses, as described in the previous chapter, section 3.4.5. After 24 hours incubation the cells were harvested, washed twice in PBS and re-suspended in 200uL of PBS prior to DNA extraction using DNeasy Blood and Tissue Kit (Quiagen) using the manufacturers protocol. Extracted DNA samples were processed by sonication, immunoprecipitation, purification and quantification by qPCR, with the conditions as outlined in the previous chapter, and described in the methods chapter 2, section 2.1.

Independent repeat experiments were performed twice for the untreated and 10μM oxaliplatin treated samples. The treated samples at 100μM and 1000μM dose of oxaliplatin were performed three times as fully independent repeat experiments.

The microarray aspect of the DIP-chip assay requires four stages of laboratory processing of the paired immunoprecipitated samples (IP) and the input control sample (IN). Firstly, amplification to (at least) the DNA amount required for downstream processing using whole genome amplification PCR (WGA-PCR) based amplification using the WGA2 Kit (Sigma) and associated protocol (Sigma 2014). Secondly, the amplified DNA was labelled using the BioPrime Total Genome Labelling Kit (Invitrogen), using the manufacturers protocol. In the third and fourth stages the labelled DNA was hybridised to the microarray slide using the Agilent aCGH kit and associated microarray protocol (Technologies 2008), and the slide was scanned for data extraction. Finally, all data was processed following the procedure explained above in section 4.1.3.5.

In addition to each experiment being repeated as complete independent biological repeats, at different stages of the assay several samples were run in parallel as 'technical replicates' allowing further detailed investigation and distinction between the biological and technical variability at each stage of the chip phase of the DIP-chip assay.

4.3 Results

4.3.1 Initial analysis of oxaliplatin PBMC microarrays

At each dose point (untreated, 10μM, 100μM and 1000μM), and for patients A and B, several independent repeat DIP-chip experiments were conducted. Microarrays were processed and data was analysed as discussed above. Oxaliplatin-DNA adduct pattern plots (Figure 5.7 and 5.8) and a matrix of Spearman's correlations between each sample was generated (Figure 5.9).

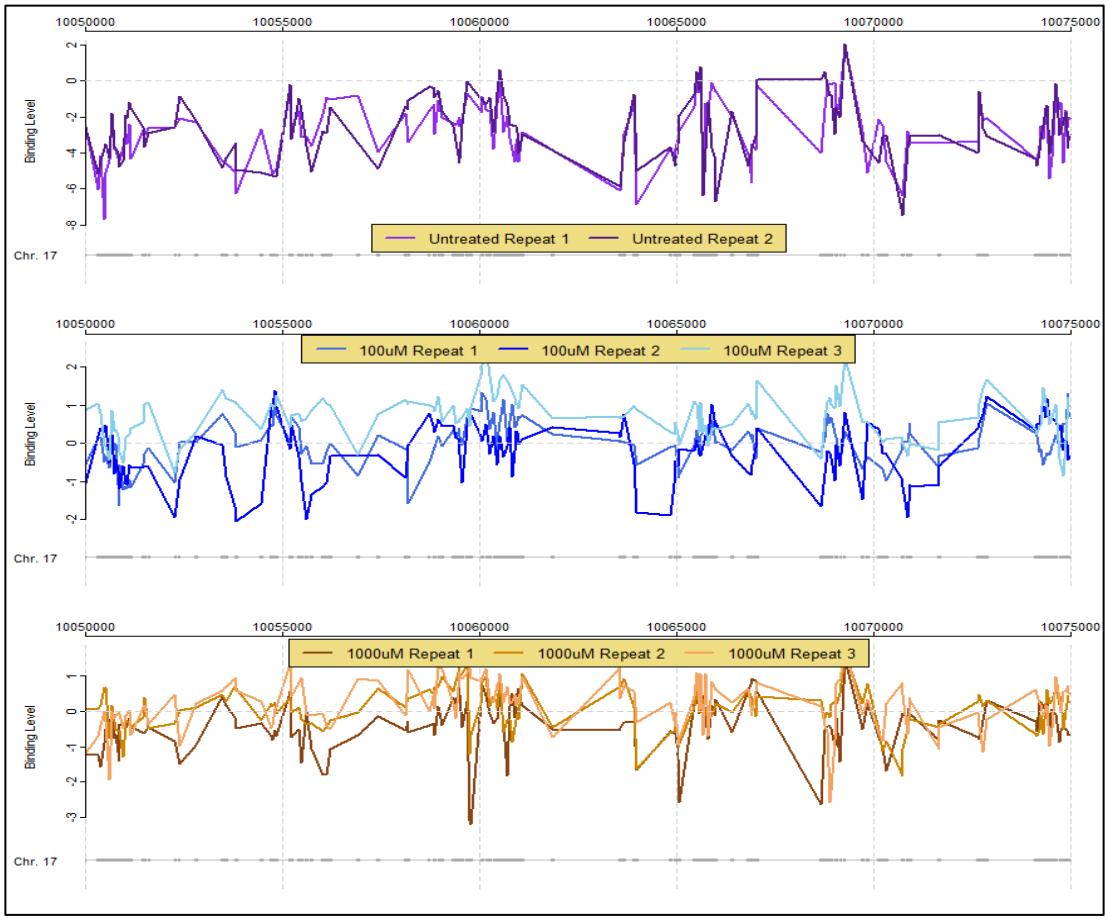


Figure 4.7: A series of oxaliplatin-DNA adduct pattern plots of several oxaliplatin treated PBMC independent biological repeat samples from patient A.
The top section shows the IP/IN ratio patterns generated in the untreated samples (purple). The 100μM (middle - blue), and 1000μM (bottom - orange) samples are also presented. The 'trace' generated from each independent repeat experiment is shown separately in each panel, as described in the legend.

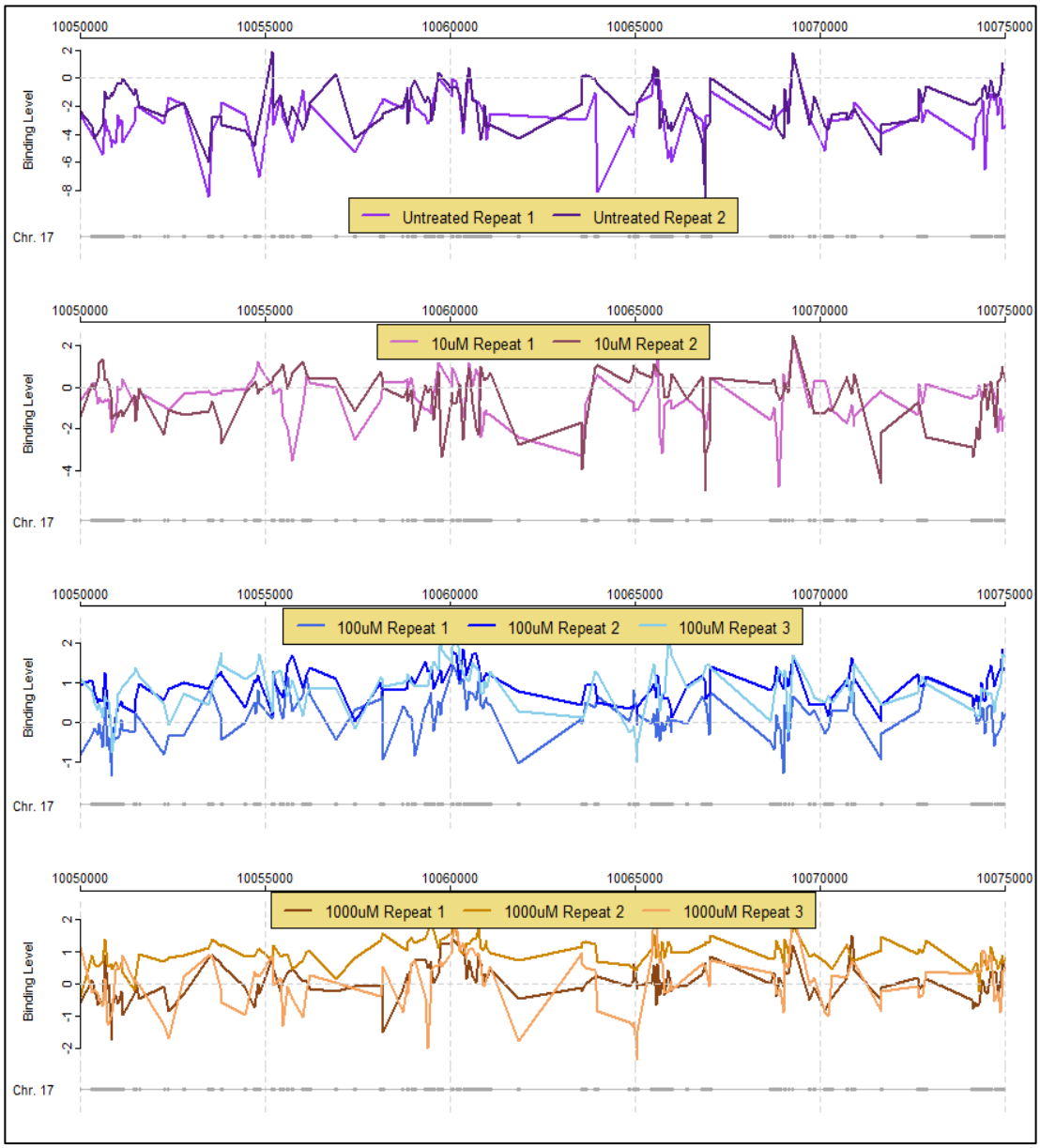


Figure 4.8: A series of oxaliplatin-DNA adduct pattern plots of several oxaliplatin treated PBMC independent biological repeat samples from patient B.
The top section displays the IP/IN ratio patterns generated in the untreated samples (purple). In each lower panel the 10μM (pink), 100μM (blue), and 1000μM (orange) samples are shown respectively. The trace generated from each independent repeat experiment is shown in each panel, as described in the legend.

The oxaliplatin-DNA adduct pattern plots presented in figure 4.7 and 4.8, generated from the DIP-chip experiment, reveal underlying variability in the patterns generated by the assay when experiments are repeated in what should be identical samples and conditions. This is not unexpected when translating this technology from carefully controlled cell culture models into human clinical samples for the first time. In all of the oxaliplatin-DNA adduct pattern plots

(Figures 4.7 and 4.8) the pattern varies over the area of the genome plotted, with variation in profile at each dose level and in both individuals.

Notably, a pattern can be detected in the untreated sample, reflecting the fact that some DNA is immunoprecipitated by the CP9/19 antibody and has been amplified sufficiently by WGA-PCR to allow labelling and application to the microarray. Note that the centre of the data meets the y-axis at approximately -4, as opposed to 0 in the treated samples. When considering the adduct pattern troughs, a value of $\log_2$ -8 correspond to an IP/IN ratio of 0.004, so at this probe the relative fluorescence intensity of the IP to the IN sample is ~0.4%, indicating that minimal DNA is present at this probe in the IP sample. The midpoint of the probe values is approximately -4 on a $\log_2$ scale, corresponding to an IP/IN ratio of 0.06 and only occasionally is the fluorescence intensity in the IP and IN samples equal, resulting in a $\log_2$ binding level of 0. This indicates that even though a pattern can be seen, very little DNA corresponding to this 5Mb section of the genome is present in the IP sample and almost the entire pool of DNA fragments hybridised to each probe is from the IN sample.

In the treated 10μM, 100μM and 1000μM samples the midpoint of the plot is approximately set at 0 on the y-axis, indicating equal fluorescence intensity at these probes in the IP and IN sample. With these experimental conditions, the comparison is then between the pattern of peaks and troughs, indicating areas of the genome where oxaliplatin-DNA adducts are more, or less prevalent in the treated DNA samples.

The presence of variability between adduct patterns from the same individual during independent repeat experiments, and at each experimental condition, is confirmed by reviewing the matrix of Spearman's correlation values (figure 4.9). This format allows a visual comparison of all of the processed clinical samples, at all of the different dose levels in both volunteers, and for each repeat.

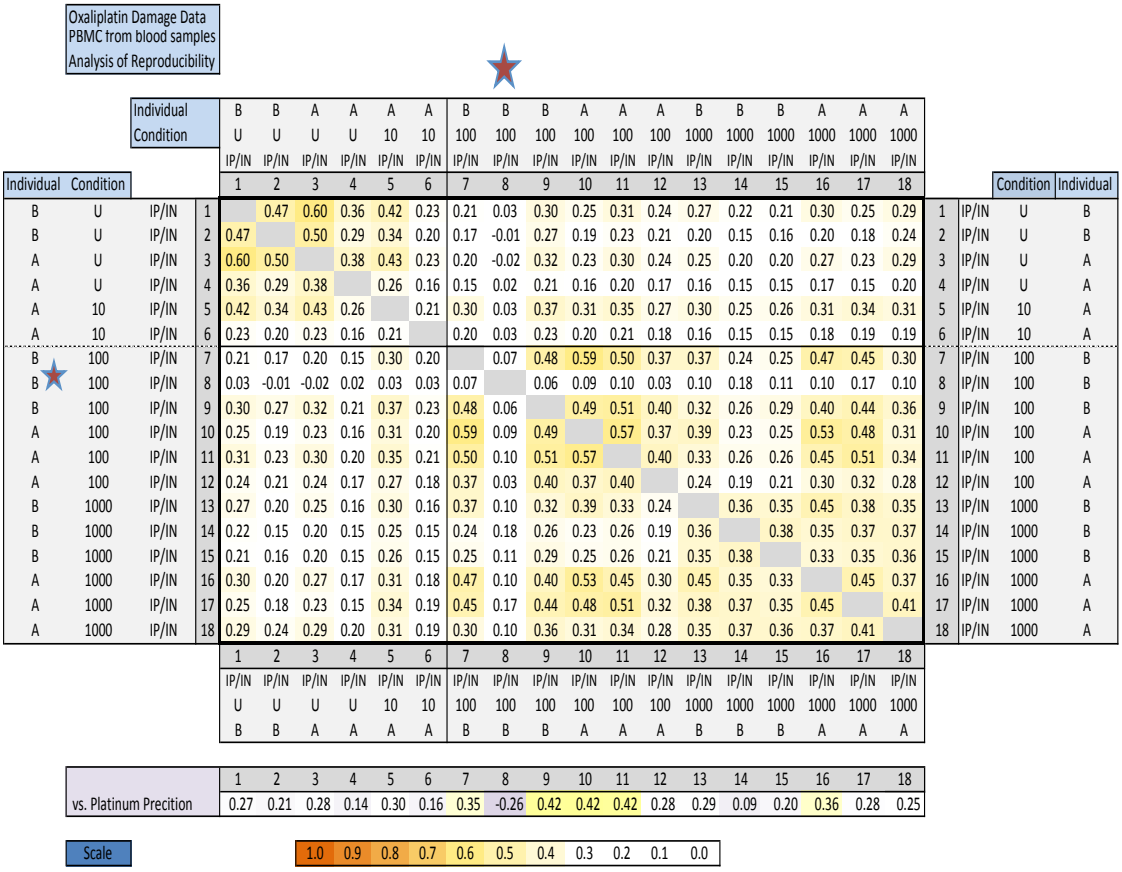


Figure 4.9: A correlation matrix between several repeat DIP-chip experiment datasets
Several independent DIP-chip experiments were conducted in two individuals (Patient A and B) and at up to 4 dose points - untreated, 10μM, 100μM and 1000μM. Oxaliplatin-DNA adduct patterns datasets have been generated for each condition (Figures 5.7 and 5.8). The degree of similarity between each pattern in each condition is displayed here, using the Spearman's correlation between each sample as a marker of pattern similarity. A correlation value of 1 indicates a identical oxaliplatin-DNA adduct patterns and a correlation value of 0 indicates no relationship between samples. The colour scale and correlation value of each sample with an independent, mathematically generated platinum prediction profile is displayed underneath. The red star indicates a sample that has failed for undetermined technical reasons (Individual B, 100μM oxaliplatin, repeat 2).

The benefit of a correlation matrix as displayed is that it allows the visualisation of patterns from an analysis of a large number of datasets that would not be apparent if a series of numerical values was simply listed. Several patterns are evident in the data as generated. Firstly, the untreated samples (grouped together in the top left section) correlate reasonably well with other untreated and 10μM samples – typical values of 0.4-0.6 – indicating a greater degree of similarity in pattern than when an untreated sample is compared to the treated samples, as shown by the lower correlation values in the top right and bottom left sections of the matrix, typically of 0.1-0.3. The central and

bottom right section contains the correlation values between microarray adduct patterns generated in treated samples, which have higher correlation values between one-another (e.g. Spearman's correlation values of 0.45-0.6 between 100μM samples and 0.4-0.5 between 100μM and 1000μM samples) than with the untreated samples (values of 0.1 to 0.3). This suggests that the DIP-chip technique is able to detect a difference in the patterns generated between the higher dose oxaliplatin treated samples and the low dose or untreated samples.

The matrix format also allows rapid identification of array datasets that may have developed technical problems. In this instance, a single experiment appears to have failed, the sample indicated by the red star (Patient B, 100μM oxaliplatin, repeat 2). In this sample the adduct pattern does not correlate well with any other samples (0.0-0.2). Also of note, the 100μM samples correlate with each other (0.4-0.6) and with the predicted profile (~0.4) better than the highest dose 1000μM samples (0.3-0.4). This may reflect the degradation and poor DNA quality of the initial DNA, as demonstrated in experiments shown in chapter 3, section 3.4.5.1.

The correlation between the oxaliplatin-DNA adduct patterns generated in the experimental datasets with the mathematically generated platinum adduct predicted profile (described in section 4.1.3.6.3) is shown in the lower bar, with higher correlation values appearing darker yellow in colour. The highest correlation value achieved is with several of the 100uM samples, and lower correlation occurs with the untreated and very highly treated samples, again potentially reflecting the poor quality of the DNA used in the supra-physiological doses 1000μM oxaliplatin samples.

Several initial conclusions and hypotheses can be drawn from this first analysis of human PBMC data. The higher correlation at 100μM between independent repeats and with the predicted profile suggests that there is an oxaliplatin-DNA adduct treatment 'signal' that can be detected with the DIP-chip experiment, as would be expected. There is, however, significant noise resulting in relatively high variability between patterns generated in the same individual in this series of independent biological repeats (Figures 4.7 and 4.8).

It appears that this technique can detect an oxaliplatin-DNA pattern in the blood from an individual patient, with the potential that this signal could be used in to attempt to predict clinical outcome. It may be possible to improve the chances of reliably detecting the DNA damage signal if the assay could be improved to reduce the 'noise' and improve the reproducibility. A higher 'signal-to-noise' ratio would be beneficial if the assay is to be used as a single-stop assay to generate an accurate platinum-DNA adduct pattern that could be analysed as a predictive marker of biological effect and response or toxicity to chemotherapy agents.

4.3.2 Using mean adduct pattern datasets as a method to reduce variability

Significant variability in the oxaliplatin-DNA patterns generated occurs in repeat samples from the same individual. It may be that calculating the mean IP/IN value at each probe position from several independent biological repeat datasets can reduce background noise present in the assay, and could potentially improve the correlation of the experimental patterns generated with the platinum predicted profile and with other clinical samples. To test this hypothesis, the mean IP/IN value at each probe location from each treatment condition in for each individual was calculated to produce a mean oxaliplatin-DNA adduct pattern at each dose level, with the standard error of the repeat samples also calculated, displayed in figure 4.10. The mean adduct pattern only is plotted in figure 4.11.

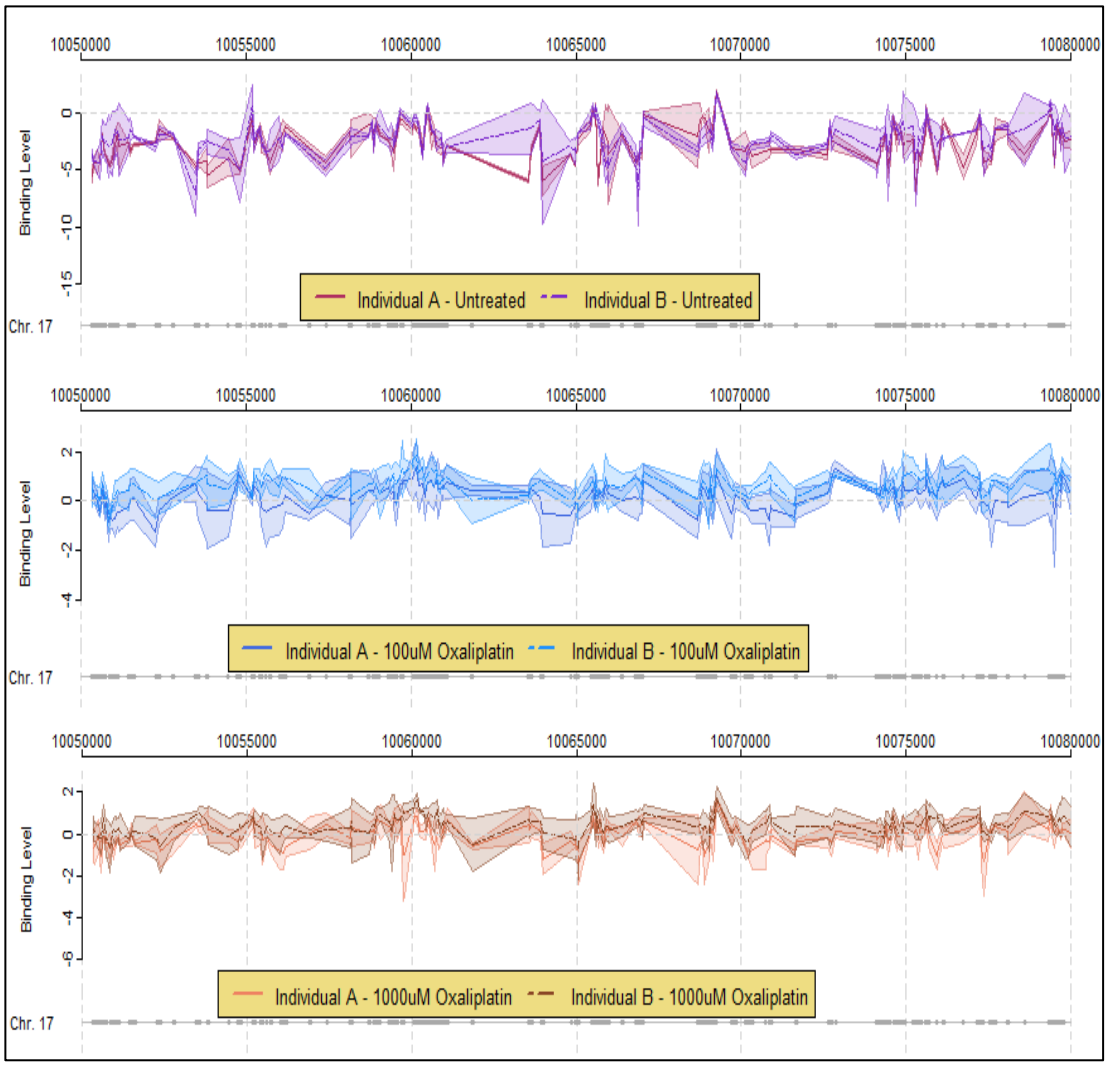


Figure 4.10: The mean adduct pattern and standard error of several independent DIP-chip experiments
The mean oxaliplatin-DNA adduct pattern (solid line) and standard error (shaded area) from a series of oxaliplatin-DNA adduct pattern plots of several oxaliplatin treated PBMC independent biological repeat samples from patient A and B is shown. This plot displays the log2 IP/IN ratio patterns generated in the untreated samples in the top plot, in the middle plot the 100μM (blue) samples, and the lower plot shows the 1000μM (orange) samples.

As can be seen, the standard error of all of the comparisons between patients A and B overlaps in the central distribution of each trace. It is not possible to differentiate the pattern between patient A and B under these conditions. However the mean trace, a similar pattern in between individuals is more apparent and is demonstrated in figure 4.11.

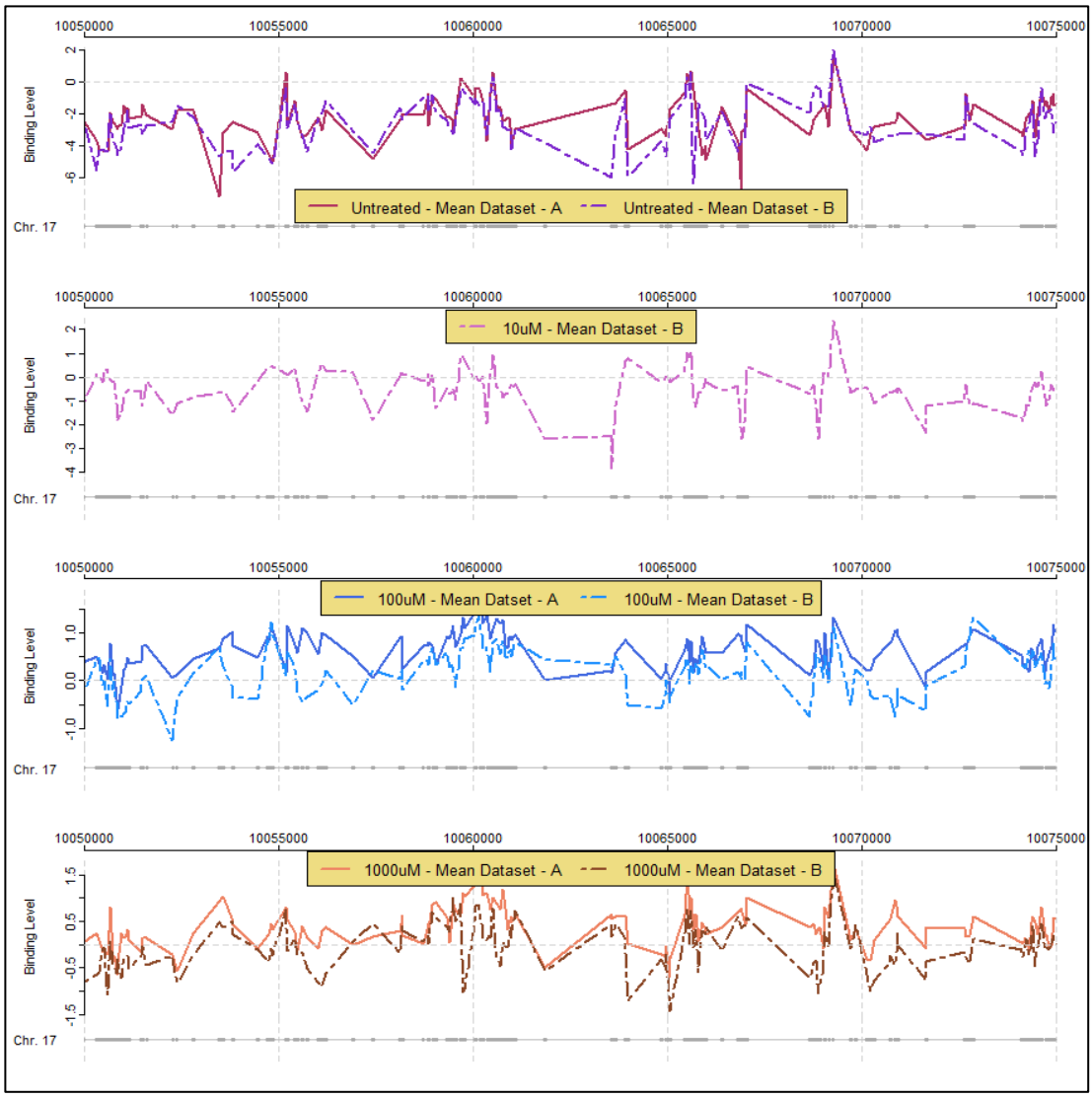


Figure 4.11: Adduct pattern plots with mean datasets:
The mean oxaliplatin-DNA adduct pattern (solid line) from a series of oxaliplatin treated PBMC independent biological repeat samples from patient A and B is shown. The top section displays the mean IP/IN ratio patterns generated in the untreated samples. In the middle panels 10μM (pink) 100μM (blue), and 1000μM (orange) oxaliplatin treated samples are shown respectively.

A correlation matrix generated using the mean probe value data, figure 4.12, reveals that the mean of the independent samples follow a similar overall relationship as previously demonstrated in the correlation matrix from individual repeats in figure 4.9, in this instance with untreated samples correlating better together (0.63) than with the treated samples (a Spearman's correlation of 0.3-0.4).

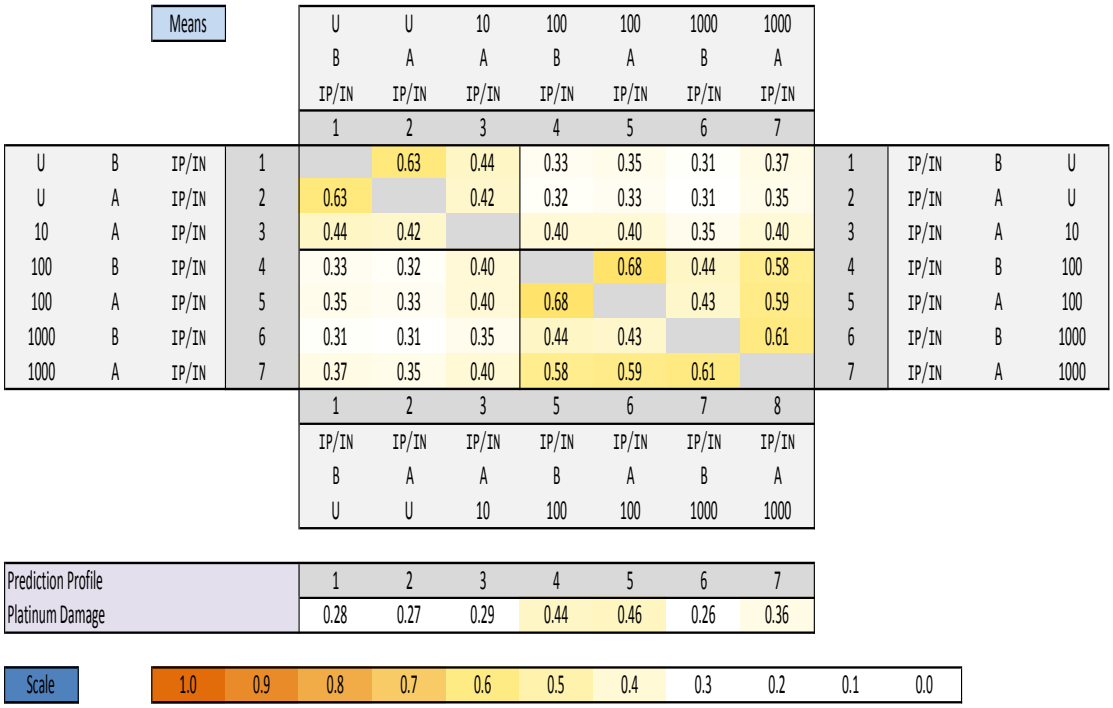


Figure 4.12: Correlations between mean probe value datasets.
A matrix of Spearman's correlation values comparing oxaliplatin-DNA adduct patterns generated from mean probe IP/IN values from independent biological repeats at each dose level of oxaliplatin treated PBMC samples from individuals A and B. The correlation between the mean damage pattern and the platinum predicted profile is also shown in the lowest bar.

Using this 'mean dataset' approach the similarities between patterns from DIP-chip samples treated under the same experimental conditions between individuals are greater, and the discrimination between treated and untreated adduct patterns is also better. For example, untreated patterns between individuals match 0.63 (as opposed to 0.3-0.6 with individual samples – figure 4.9) and 100μM sample correlate 0.69 between patient A and B as opposed to 0.37-0.59 with individual samples (figure 4.9). In the top right section the mean treated datasets compared to the mean untreated datasets correlate 0.3-0.4, indicating the difference between untreated and treated patterns is more robust and more consistent than with individual samples. Additionally, the correlation values between mean treated samples and the platinum predicted profiles are higher than with individual samples (a range of 0.26 to 0.46 (figure 4.12) vs. 0.25-0.42 (figure 4.9)).

With independent repeat samples the pattern from any one sample overlaps with patterns from repeat samples (figure 4.10) indicating the presence of high assay noise, resulting in inability to discriminate pattern 'signals' between individuals. However, when the mean probe values only are plotted (figure 4.11) these patterns demonstrate more consistent correlation between individual A and B at each dose level. These results indicate that when developing tools to identify pattern difference signals between patients A and B, using a mean dataset approach will be more reliable and less influenced by assay noise than comparing single DIP-chip assay datasets.

In the present form, if the variability in patterns between the same individual remains, the assay would generate inconsistent single experiment patterns when used, unless repeated sampling from patients was employed. This would be impractical, especially if the assay is to be used in future to analyse other clinical tissues such as biopsy samples. To refine the assay it may be possible to further optimise the laboratory protocols. The effect of protocol modifications and of experimental methods to reduce the noise further by improving the technical performance of the assay will be examined in the remaining experiments presented in this chapter.

4.3.3 Determining the assay stages contributing to DIP-chip experimental variability

In the multi-step DIP-chip assay there are several stages of the protocol that could be responsible for introducing the variability, demonstrated above, into the experimentally generated microarray results. As shown in many of the experiments described in chapter 3, the variability of the immunoprecipitation phase of the assay can be minimised by careful handling and processing of samples within experimentally well-defined parameters, resulting in improved consistency in IP/IN ratio compared to the IP/IN ratio in untreated samples, albeit at limited genetic loci as described. Because of the consistent results seen with the improved DIP protocol the variability seen in the microarray datasets is likely to be produced predominantly during the processing downstream of the immunoprecipitation sample and input control sample during the microarray preparation stages.

To estimate the noise generated at each stage of the assay an experiment using a series of technical replicates was conducted. These samples consist of aliquots of PBMC-extracted DNA immunoprecipitated in parallel before pooling and dividing into equal identical samples prior to processing at sequential stages during the microarray preparation. As each stage in the experiment has the potential to add noise to the final result, the microarray datasets generated from this experiment are analysed from the microarray stage in reverse, to quantify increases in variability between identical technical repeats with increasing number of assay stages. This approach identifies and quantifies the variability introduced at each stage of the assay.

Four sets of paired technical replicates were used either pre-immunoprecipitation, pre-amplification, pre-labelling, pre-hybridisation, and identical replicates were run on the same and different arrays to quantify intra- and inter-array variability. A comparison of the oxaliplatin-DNA adduct patterns at each stage of the assay and Spearman's correlation values between technical replicates are shown in figure 4.13.

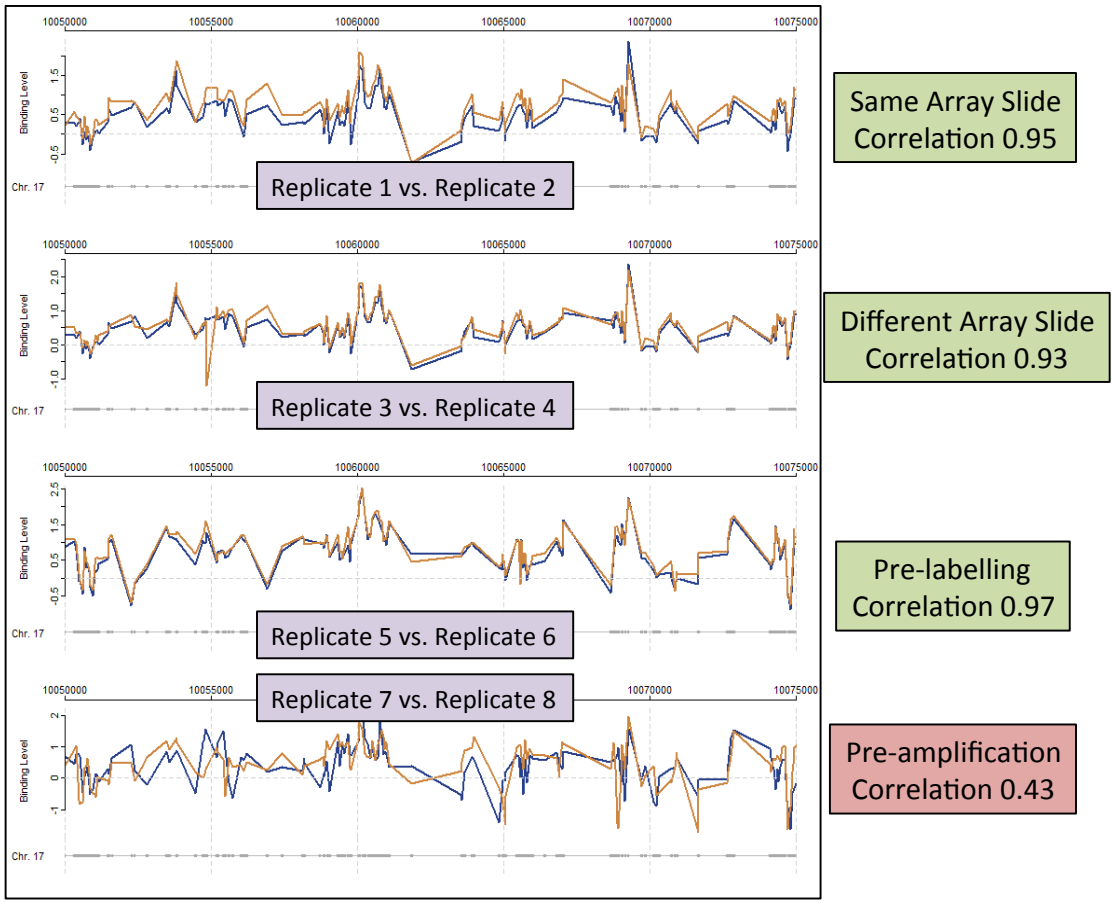


Figure 4.13: A comparison of pairs of oxaliplatin-DNA adduct patterns generated from identical technical replicate samples of oxaliplatin treated PBMC.
Sample variability is compared when a pool of two DIP-chip samples is separated onto two microarrays on different slides (top panel), between replicates run on microarrays on the same array (second panel), between replicates combined from 2 DIP samples and divided at the labelling stage (third panel) and between replicates compared at the amplification stage (bottom panel). The damage pattern and overall Spearman's correlation value between patterns is shown, with a green box for high correlation and a red box for lower correlation between replicates.

This data demonstrates that technical replicate DIP-chip samples, pooled and divided prior to each stage of the microarray preparation to ensure the fragment population is identical, when run on the same or different microarrays, show very high levels of reproducibility. This result also indicates that amplification during DNA labelling, the labelling reproducibility and the inter- and intra-array reproducibility is very consistent.

Identical technical replicates divided into equal samples prior to the amplification stage, however, have a much lower reproducibility, as shown by the reduction in Spearman's correlation value from >0.93 to 0.4 between these datasets (lower panel figure 4.13). These values are as low as samples which are true independent biological repeats (as in figure 4.9), suggesting that the

majority of the variability introduced into this microarray phase of the assay is a consequence of the whole-genome PCR (WGA-PCR) amplification process required to ensure adequate DNA is available for labelling and hybridisation.

The biological signal present in the assay is transmitted through differences in the proportion of DNA fragments between experimental conditions resulting from the affinity of the DIP antibody for the target, in this case platinum-DNA adducts. It follows then, that a fundamental requirement of the DIP-chip assay is to faithfully maintain the relationship between DNA fragment amounts and distribution in each IP and IN sample throughout the multi-step assay process. Otherwise, the specific biological information that potentially exists between samples (i.e. that should result in specific inter-patient platinum-DNA adduct patterns) could be masked by experimentally generated differences in fragment pools, resulting in noise or systematic bias. As discussed above in section 4.1.3.1, it is well recognised that PCR based amplification techniques have significant potential to introduce noise and bias during library preparation by differential amplification of fragments depending on subtle differences in random or universal primer annealing and by the preferential amplification of different fragments due to small differences physical or chemical properties between fragments (including GC content, fragment length and annealing and denaturing temperature). Small differences in fragment amplification characteristics ultimately have a large impact during the exponential amplification process. The result of this experiment (figure 4.13) indicates that this well-recognised phenomenon is an important issue in the DIP-chip assay used here.

4.3.4 Experimental measures to improve DIP-chip assay reproducibly

As discussed, the issue of PCR-based amplification noise and bias is well recognised and a significant issue for immunoprecipitation microarray or sequencing based experiments, in which the proportion of fragments generated by immunoprecipitation (and in the IN sample for correct data normalisation) must be maintained during all stages of the assay, and are especially vulnerable to bias if subtle variation exist between fragments are multiplied through the exponential kinetics of PCR amplification. Several factors affecting the fidelity of PCR reactions, and hence the reproducibility of WGA-PCR, have been documented, and include the polymerase enzyme used in the reaction, the initial DNA template amount, the number of PCR cycles used during amplification, the temperature ramp rate of the thermo cycler, the GC content of the fragments amplified and the fragment length (Aird et al. 2011; Han et al. 2012; Hasmats et al. 2014; Pinard et al. 2006; Quail et al. 2012; van Dijk et al. 2014).

Notably, the DIP-chip assay was developed in *S. cerevisiae* and human cell culture models where these technical issues were less significant in these systems, and the technical reproducibility of DNA damage patterns produced by the assay was high, especially when used to investigate the genomic location of protein binding. In this situation, the vast majority of fragments are immunoprecipitated at discrete sites with very little DNA immunoprecipitated at other locations in the rest of the genome – for example at some protein binding sites there can be a >30,000 fold difference in IP/IN ratio compared to non-protein binding regions of the genome (Teng et al. 2010). In these circumstances, noise generated by small differences in relative fragment amplification when the immunoprecipitated fragments are exclusively at binding sites (at a ratio of 30,000 to 1) would not result in loss of such a strong biological signal in the assay. The high signal-to-noise ratio results in good assay reproducibility metrics. Conversely in the human platinum-adduct DIP-chip experiments the peak-to-trough difference in the treated samples is in the order of 2 to 4 fold for oxaliplatin and cisplatin (e.g. figures 4.7 for oxaliplatin and figure 4.6 for cisplatin) presumably reflecting the widespread induction of platinum-DNA adducts throughout the genome. Under these conditions subtle

differences in fragment WGA-PCR amplification characteristics would result in a much higher noise to signal ratio, with the consequences outlined in the results so far in this chapter.

It is only through the experiments discussed here, of the translation of this technique into human clinical samples, that issues of reproducibility, unique to this context, have become apparent. In the following section, the effect of WGA-PCR related variables on the reproducibility of human DIP-chip PBMC experiments will be explored, with the intention of minimising the noise and bias introduced into the assay during DNA amplification.

4.3.5 Measuring the effect of DNA template on WGA-PCR amplification noise and bias

During the DIP-chip experiment the IN and IP sample are amplified in parallel. Because immunoprecipitation is by nature selective for the DNA fragments of interest, only a fraction of the original DNA fragments are retained in the PBMC post-immunoprecipitation sample (typically 0.1-0.2% - table 4.1). Following immunoprecipitation, purification and qPCR, differing absolute amounts of DNA template is therefore available for amplification in each sample, and can be calculated following sample quantification by qPCR (table 4.1). If the difference in template concentration has a significant effect on the reproducibility of the WGA-PCR stage, it may be possible to demonstrate amplification differences experimentally in these samples. The absolute amounts of DNA in each sample can also be calculated and compared to the human cell culture development experiments conducted to ascertain whether in template variability resulting in PCR noise is specific to the PBMC context.

Oxaliplatin treated PBMC Sample Dose	DNA amount in IP Sample as a percentage of paired IN sample (required fold dilution of IN to match IP concentration) Repeat 1	DNA amount in IP Sample as a percentage of paired IN sample (required fold dilution of IN to match IP concentration) Repeat 2	DNA amount in IP Sample as a percentage of paired IN sample (required fold dilution of IN to match IP concentration) Repeat 3
Untreated	0.01% (~15,000 fold)	0.01% (~15,000 fold)	0.01% (~15,000 fold)
100μM	0.19% (~500 fold)	0.12% (~850 fold)	0.09% (~1000 fold)
1000μM	0.37% (~250 fold)	0.22% (~450 fold)	0.21% (~480 fold)
Oxaliplatin treated dermal fibroblasts	**Repeat 1**	**Repeat 2**	**Repeat 3**
Untreated	0.03 (~3000 fold)		
2500μM	5.4 (~18 fold)	4.4 (~22 fold)	4.8 (~20 fold)

Table 4.1: The relative amounts of DNA in paired IP and IN samples processed through the DIP reaction and quantified by qPCR.
In three independent experiments the amount of DNA in the IP samples is shown as a percentage of the IN sample. The relative dilution of the IN sample that would be required to match the concentration in the IP sample is displayed below each percentage figure. The volumes of each sample are the same post immunoprecipitation. Results from oxaliplatin treated PBMC and dermal fibroblast cell cultures (treated with 2500μM oxaliplatin) are shown in the lower section.

The template concentration in the post-immunoprecipitated IP sample available for WGA-PCR in the PBMC experiments is 100 to 1000 fold lower than the IN samples (table 4.1). The difference in concentration between samples is greater than in assay-development experiments conducted with dermal fibroblasts, where the available template in the IP sample for WGA-PCR amplification was higher, reflecting the 25 fold increased dose used and may also reflect the platinum sensitivity of the cell line.

When attempting to develop a clinical assay based on the correlation between surrogate tissue derived oxaliplatin-DNA adduct patterns and the idiosyncratic response of patients to 2-5μM dose of oxaliplatin (the typical ultra filterable plasma dose during treatment), a relevant dose of chemotherapy is required for *ex vivo* treatment to make the results as biologically relevant as

possible – a dose point of 1000 times that seen in patients used during assay development was rejected on these grounds. Notably, at similar oxaliplatin doses and treatment schedule (4 hours in serum free media) in human PBMC cell the IP/IN ratio obtained was relatively low (chapter 3, section 3.4.2) and at a 1mM dose 50% cell death (chapter 3, figure 3.8) and DNA degradation occurred (chapter 3, figure 3.11) making this an unviable treatment schedule in human clinical samples.

The effect of template quantity on the quality of the WGA-PCR amplification output can be demonstrated by comparison of DNA fragment profiles by gel electrophoresis using the Agilent TapeStation 2200 to perform electrophoresis and fragment distribution analysis on post-immunoprecipitation IP and IN samples of DNA fragments between 50 and 1200bp in length (figure 4.14)

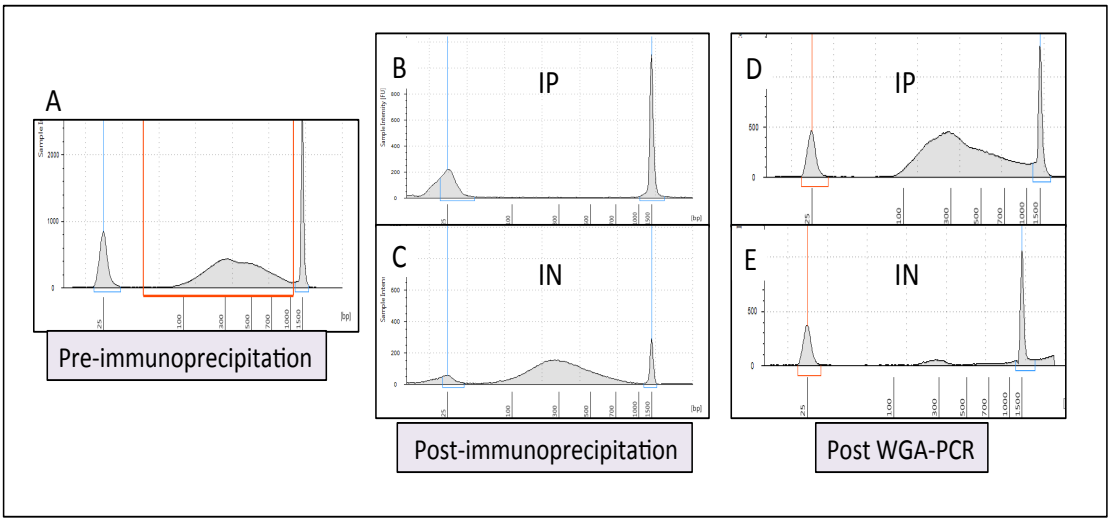


Figure 4.14: Analysis of sample fragment distribution during the DIP-chip experiment
The profile of DNA fragments in an oxaliplatin treated PBMC pre- and post-immunoprecipitation and amplification samples during an example DIP-chip oxaliplatin treated PBMC experiment is analysed with TapeStation 2200 (Agilent Technologies). The left window (A) shows the DNA fragment profile following sonication and before immunoprecipitation. The central windows (B and C) show the DNA fragment distribution following the immunoprecipitation stage. The IP sample (B) is not detectible using this technique as the quantity of DNA in the sample is below the detectible threshold. The fragment profiles following WGA-PCR are shown in the right hand panels D and E.

With electrophoresis-based analysis (figure 4.14) of the DNA fragments present in the IN and IP samples pre-immunoprecipitation (A), pre- WGA-PCR (B

and C) and post- WGA-PCR (D and E) it is possible to demonstrate a significant change in fragment distribution caused by the WGA-PCR process in the IN sample. It is clear that longer fragments have been amplified in preference to shorter fragments, skewing the distribution profile to the right (E). This has not occurred in the IP sample (D) compared to the pre-immunoprecipitation sample (A), although because no fragments can be detected at the lower limit of detection in the pre-WGA-PCR IP sample (B) it is not possible to confirm whether the distribution post-WGA is due to the immunoprecipitation or the amplification process.

To determine if these findings are due to template concentration by gel electrophoresis, an experiment was conducted taking a single pre-amplification IN sample and diluting this sample in serial 10 fold dilutions. In this way, the diluted IN sample concentration includes the typical concentration of IPs samples usually a 100-1000 fold lower concentration. Each diluted sample was amplified through the WGA-PCR and for visualisation run on a 0.8% TAE gel (figure 4.15). The typical IN and IP pre-WGA-PCR sample concentrations are indicated below by a green and red star respectively.

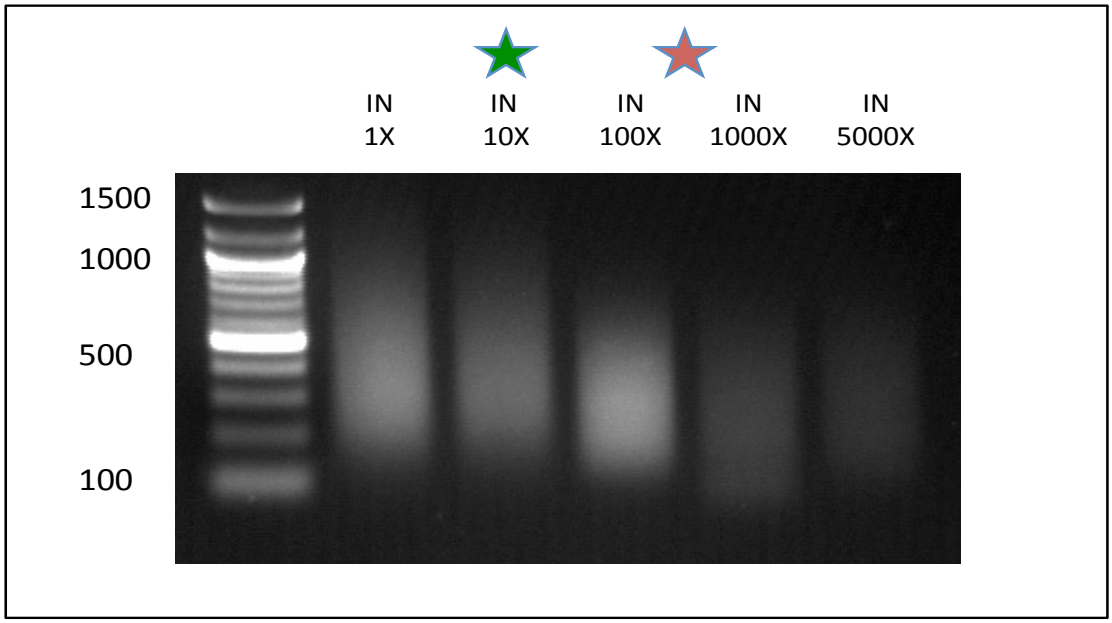


Figure 4.15: Determining the effect of sample concentration on fragment profile post WGA-amplification.
A 0.8% TAE gel comprising a DNA ladder and the same IN post-immunoprecipitation sample serially diluted (dilution factor shown above each lane) and WGA-PCR amplified. The typical IN and paired IP sample concentrations for WGA-PCR are indicated by the green and red starts respectively

Figure 4.15 demonstrates that the template concentration has a significant effect on the resulting fragment profile following WGA-PCR amplification, presumably by influencing which fragments are preferentially amplified. With each sequential 10-fold dilution the profile of fragments changes, with a progressive reduction of the median and range of fragments visible on the electrophoresis gel. This implies that the IN and IP samples in the PBMC samples are subject to differential amplification, as the staring template DNA concentrations in these samples are several fold different. As the IN is intended to be a faithful reflection of the relative amounts of fragments in the initial starting sample, and the difference in the IP and IN samples at the microarray stage are presumably due only to the fragment-selective immunoprecipitation step, there appears to be a significant concern that template-dependent amplification bias may introduce variability into the assay. The effect of this potential source of bias on the microarray output will be examined in the next section.

4.3.6 Examining the effect of WGA template concentration of IN samples on array correlation values

The above experiments (Figures 4.13-4.15) demonstrate that variability in DIP-chip generated oxaliplatin-DNA adduct patterns, present in the repeat analysis of oxaliplatin treated PBMC samples, is primarily due to the WGA-PCR process. This variability could potentially be generated through noise introduced in either, or both, of the IN and IP samples, as during data analysis they are combined to give the IP/IN ratio at each probe used to generate the adduct pattern.

As demonstrated above (Figure 4.15), the WGA-PCR template concentration has a significant impact on the distribution of fragments generated by amplification, although it is yet to be shown if these changes have a significant impact on the microarray variability. Intuitively this would be expected to have an effect, as the IN fragment profile and distribution is generated and used specifically to normalise variability in the IP probe values due to underlying

variation in DNA fragment frequency in the initial pre-immunoprecipitation sample and to correct for experimental variability. Inadvertently experimentally altering the distribution of DNA fragments in either the IN or IP samples would presumably have a differential effect on the IP/IN value at each probe in different samples, particularly if the background frequency and distribution of IN fragments at each probe has been altered compared to the IP (as the template differential is in the order of 100-1000 fold). Another uncertainty is differential amplification of IP fragments dependant on the treatment condition and resulting immunoprecipitation efficiency, as the IP/IN ratio in 10μM samples and 1mM samples is roughly 10 fold different (Chapter 3 figure 3.6) – a difference in template large enough to result in differential amplification and a different pool of fragments produced by the WGA-PCR process in these two IP samples (Figure 4.15)

To reduce unintentionally introduced template-concentration dependent amplification bias, it follows that the template concentration must be the same in both IP and IN samples prior to amplification. If an IP sample with much higher concentration could be generated, sufficient to match the IN sample concentration, this would be optimal. It may be possible for the IP concentration to be increased, possibly by increasing the drug concentration of the DNA damaging agent, by improving the antibody affinity for damaged DNA, or by using cells that are more sensitive to platinum drugs as a surrogate tissue. The degraded DNA seen in 1000μM samples compared to 100μM samples (chapter 3, figure 3.11) suggests that in PBMC increasing the drug dose, the simplest option, may not be viable.

Another possible option to equalise any amplification bias between IP and IN samples is to decrease the IN sample template concentration prior to WGA-PCR to ensure any template dependent amplification bias is matched between IN and IP samples, and therefore normalised by the IP/IN ratio used in adduct pattern generation.

To see if using the same concentration of template in the WGA-PCR reaction would eliminate this potential source of bias, microarray experiments

were conducted using several parallel immunoprecipitated IP samples which were combined and equally divided prior to amplification, with the intention of producing identical technical IP replicate samples prior to WGA-PCR. An IN sample was diluted either 10 fold (the typical dilution used) or diluted 500 fold to match the IP template concentration as calculated by qPCR in these samples. Following WGA-PCR amplification these technical replicate samples were analysed by gel electrophoresis and using the TapeStation 2200 (Agilent Technologies) prior to the labelling step of the assay (figure 4.16).

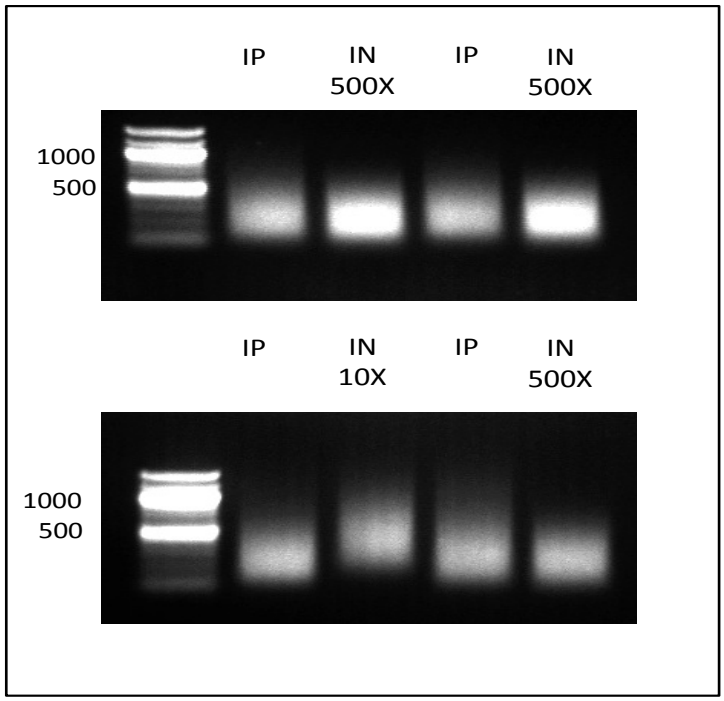


Figure 4.16: Determining the effect of sample concentration on fragment profile post WGA-amplification.
Several IP and IN samples were immunoprecipitated before being combined and split to produce pools of identical technical IP and IN replicates. The IN samples are diluted either 10X (as per the initial DIP-chip protocol) or 500X to maintain the same concentration as the IP sample (as previously quantified with qPCR).

The gel electrophoresis in figure 4.16 again shows the relationship between pre-WGA-PCR DNA template concentration and the fragment distribution in the amplified material. All of the IP samples have the same fragment distribution following amplification, as do the IN samples which have been pre-diluted to match the IP concentration, as calculated by qPCR. The IN sample which was diluted 10 fold to give a 10ng/μl DNA concentration prior to

amplification (our standard DIP-chip protocol) has a significantly longer fragment median and range. These results again confirm that different pools of DNA fragments are produced by WGA-PCR amplification depending on the concentration of starting material, a potential source of system noise and systematic bias.

As all the IP samples are identical replicates, and the IN samples differ only by the amplification-dependent change in fragment profiles, the paired IP and IN samples were labelled, hybridised and analysed on standard microarray protocols. The oxaliplatin-DNA adduct patterns generated in three pairs of samples (2 IP/IN pairs with 500X dilution and 1 IP/IN sample pair with a 10X IN dilution) are compared in figure 4.17. The Spearman's correlation value between the patterns is included in the legend.

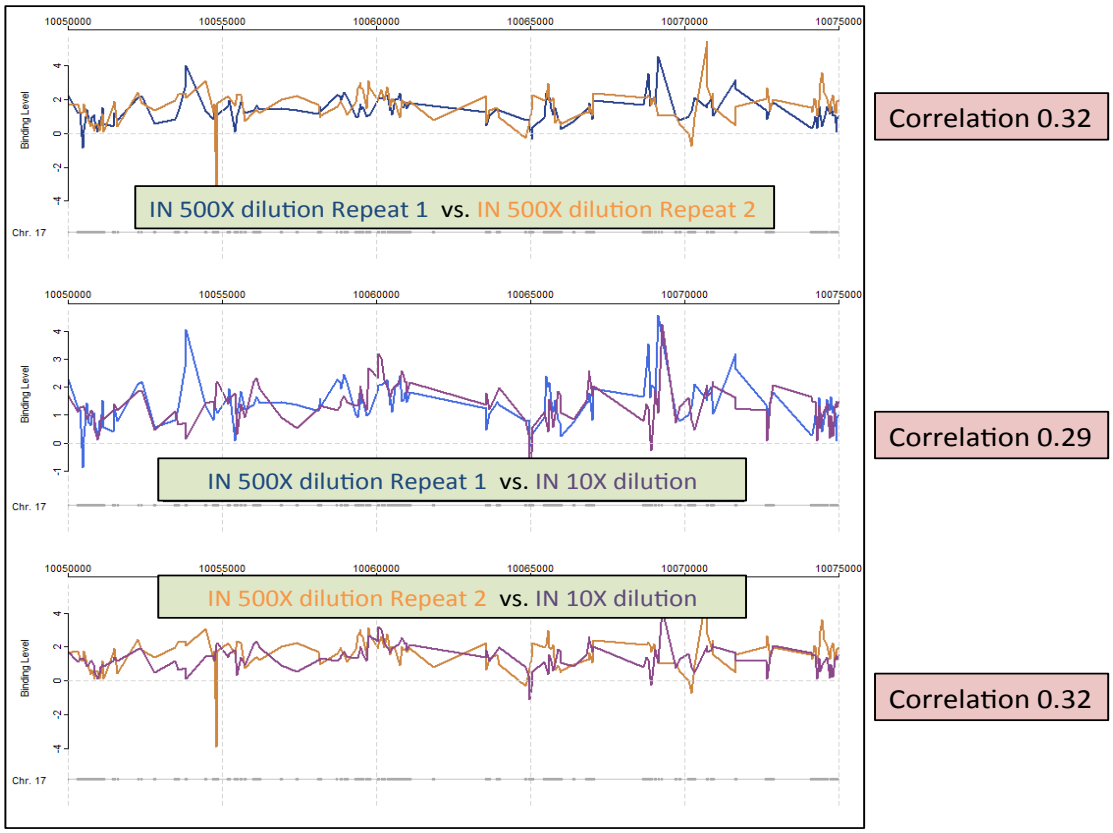


Figure 4.17: Determining the effect of sample concentration pre-WGA on adduct pattern plot.
An oxaliplatin-DNA adduct pattern plot comparison of oxaliplatin-DNA adduct patterns generated from 100μM oxaliplatin treated PBMC samples with identical IP samples and differentially amplified IN samples. Two patterns have been generated from IN samples diluted 500X to match the pre WGA-PCR DNA concentration, and one pattern is generated using standard 10X IN dilution prior to WGA.

On reviewing the Spearman's correlation coefficients in this figure 4.17, between the oxaliplatin-DNA adduct patterns generated when the samples with variable IN amplification are run on microarrays, the patterns generated are highly variable and fail to correlate well with one another (Correlations values of ~0.3). This does not improve when the IN sample template concentration is decreased to match the IP concentration (middle and lower panels compared to upper panel). This indicates that changing the IN concentration to match the IP template concentration does not improve the reproducibility of the microarray by normalising any bias introduced by template dependent amplification. It may be, however, that this potentially bias-reducing effect is masked by other sources of noise in the microarray data, specifically within the IP channel rather than within just the IN channel. With the currently available bioinformatic analysis techniques it is impossible to identify where, or why, this problem is occurring, signalling the need for further analysis tools to inspect the data in greater depth to resolve this issue, rather than to perform further experimental permutations to explore this problem further.

4.3.7 Modification of IP template amount and adjustment of PCR cycles

A reduction in the IN template concentration prior to WGA-PCR amplification results in a better match of fragment distribution with the IP sample (figure 4.16) and, more importantly, with the original pre-immunoprecipitation DNA sample (figure 4.14), but still results in a low correlation between identical technical replicates run through the later stages of the DIP-chip assay (figure 4.17). The solution may be to increase the template of the IP sample prior to amplification, leading to more reproducible damage patterns than reducing the IN concentration.

With this in mind, an experiment was conducted to modify the amount of DNA template for WGA-PCR in the IP sample by increasing the amount of DNA used in the immunoprecipitation reaction, and combining immunoprecipitated samples prior to amplification. This approach was used to enhance the IP sample DNA concentration rather than increasing the drug exposure of the PBMC

in culture because as previously demonstrated, increasing the dose of drug (from 100μM to 1000μM) does increase adduct levels and improve IP/IN enrichment at qPCR (chapter 3 figure 3.6), but results in higher cell death and poor quality degraded DNA (chapter 3, figures 3.7 and 3.11), and is associated with a reduction in damage pattern reproducibility (figure 4.11). The use of drug doses of increasing orders of magnitude higher than used in patients undergoing chemotherapy (typically 3-5μM) also further reduces the potential biological relevance of the output of the assay.

To measure the effect of increasing IP template concentration, allowing a reduction in PCR cycles to generate adequate WGA-PCR DNA for labelling and microarray hybridisation, a single post-immunoprecipitation IN sample was diluted into several identical aliquots at a concentration of to 10 ng/μL prior to WGA-PCR were used for the IN channel. IP samples were generated by immunoprecipitation with either 3μg or 6μg of sonicated DNA, aiming to increase the absolute DNA yield of the IP procedure, or with two pooled post-immunoprecipitation 6μg samples for the highest template concentration. Fidelity of PCR amplification is known to decrease with increasing cycle numbers (van Dijk et al. 2014), so with each 2-fold increase in template amount the cycle number was able to be decreased, whilst still generating enough DNA for downstream processing, further reducing the potential for bias. Oxaliplatin-DNA adduct pattern plots and the corresponding Spearman's correlation coefficient demonstrating the reproducibility between paired technical replicate IP samples at each template level are displayed in figure 4.18.

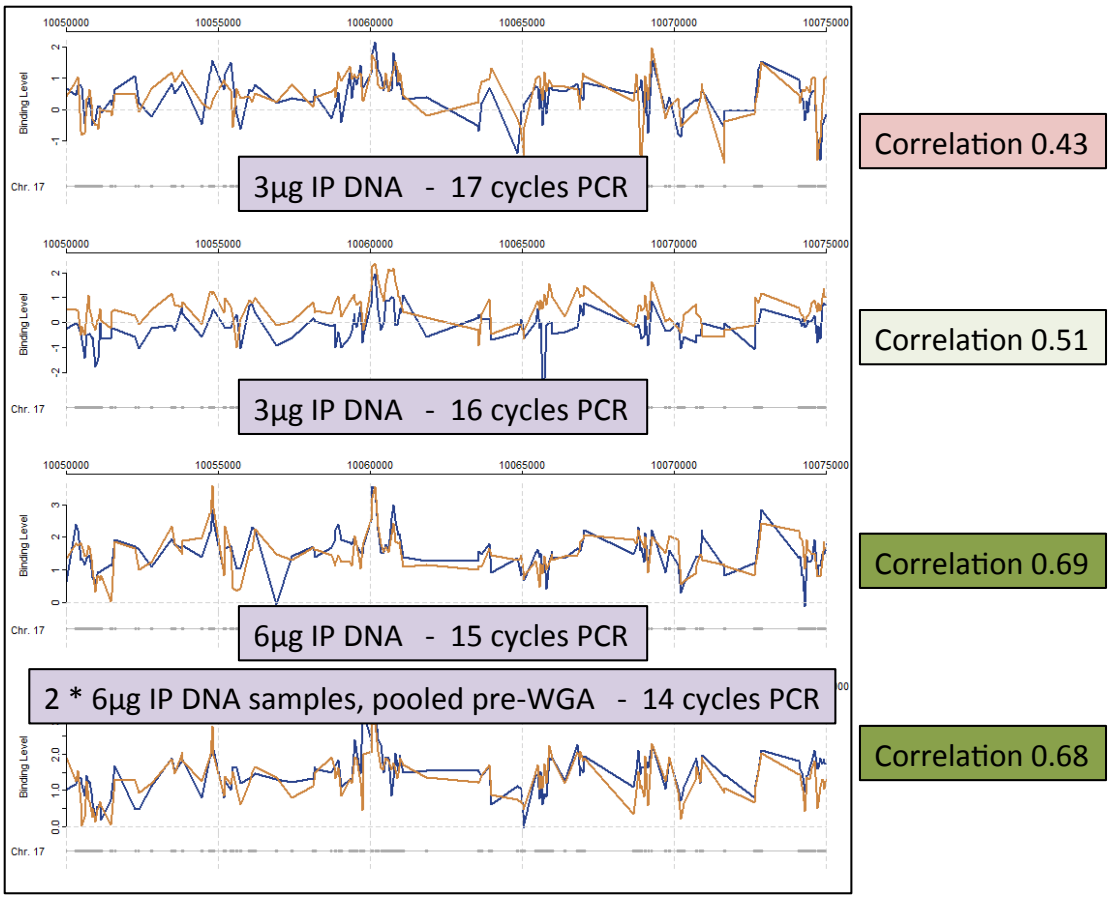


Figure 4.18: The effect of PCR conditions and template concentration on DIP-chip reproducibility.
An oxaliplatin-DNA adduct pattern plot comparison of oxaliplatin-DNA adduct patterns generated from 100µM oxaliplatin treated PBMC samples with identical IN samples and IP samples of increasing concentration prior to WGA-PCR. With each lower panel the IP template available for WGA-PCR is increased and the PCR cycles are decreased (see each panel legend). The Spearman's correlation between replicate samples is shown on the right with an increasing change from red to green to highlight increasing value (arbitrary scale).

This result (Figure 4.18) demonstrates that an increase in reproducibility between technical replicates occurs with increasing template amounts and decreasing PCR cycle number during WGA-PCR (increasing from 0.43 to ~0.7), suggesting that the a significant component of the poor assay reproducibility lies with the amplification of low template IP samples, rather than with the IN samples, in which 10-100 fold more DNA is available as a template for WGA-PCR.

The amplification process can be made more reproducible by reducing the numbers of PCR cycles from 17 to 15 in this example. Although it is not clear from this experiment whether the enhanced reproducibility occurs from increasing template or reducing PCR cycles, the yield from the experiment is at the lower limit of DNA required for labelling and hybridisation (500ng), so it is

not possible to reduce PCR cycles further in lower template samples without correspondingly requiring a doubling in the initial template DNA amount. Even using pooled samples and 6μg of DNA in the original immunoprecipitation stage of the experiment (bottom panel figure 4.18), significant variability still occurs between what should be identical samples, and is still well short of the 95%-100% reproducibility seen between technical replicates when run post WGA-PCR (figure 4.13).

Somewhat limiting the use of the approach of further boosting the DNA in the IP sample to further improve assay reproducibility is that 12μg of DNA is close to, or in some instances greater, than the maximum amount of DNA extracted from each PBMC clinical sample in the initial DIP phase of the assay (chapter 3, figure 3.13). Whilst this experiment highlights the role of enhancing the IP sample in improving assay reproducibility, it is not a complete solution to this problem, both because 12μg of DNA for immunoprecipitation still results in significant assay variability between what should be identical post-immunoprecipitation DNA, and also because of the physical limits to the DNA available for the assay from PBMC samples.

This result, however, confirms that measures to improve the technical performance of the WGA-PCR amplification of the IP sample may be the key to resolving the issue of poor reproducibility of the assay in this context. In combination with this approach, it may be also possible to apply bioinformatic tools to the large datasets generated by the DIP-chip to minimise the effect of noise introduced during the processing of the assay and improve that capacity of the assay to detect oxaliplatin-induced adduct signals between individuals.

4.4 Summary

The key results described in this chapter are summarised as follows:

- Using the DIP-chip assay it is possible to detect differences in generated patterns between untreated and oxaliplatin treated PBMC samples (Figures 4.7-4.9)

- Significant variability occurs between oxaliplatin treated single PBMC samples from the same individual (Figure 4.10)

- Mean adduct pattern data may be more accurate than single samples for later use in developing tools to detect biological signals (Figures 4.11-4.12)

- The single sample variability is primarily due to noise introduced during WGA-PCR (Figure 4.13)

- A large difference in the DNA template available for WGA-PCR occurs between the IP and IN sample, due to the nature of the immunoprecipitation process. This difference is greater in the PBMC samples than in fibroblast cell culture models (table 4.1)

- Over the range of DNA concentrations present in the pre-amplification IP and IN samples, a different profile of fragments is produced by WGA-PCR amplification (Figures 4.14-4.16)

- Reducing the IN sample concentration to match the IP sample concentration to equalise this potential bias does not improve correlations between replicate samples (Figure 4.17)

- Increasing the IP sample template concentration for WGA-PCR (closer to that in the IN sample), and reducing the PCR cycles for WGA-PCR does improve correlation between replicate DIP-chip samples.

In summary, the experimental techniques and bioinformatic analysis conducted, using the tools available and demonstrated in this chapter, indicate that the WGA-PCR amplification stage can introduce significant variability into the output of the DIP-chip assay (figure 4.13), if not carefully controlled. This effect appears to be dependent on the PCR template concentration (figures 4.14-4.16), and is a particularly important issue in the context of the low adduct amounts and resulting post-immunoprecipitation DNA yield seen with human PBMC samples (see table 4.1).

The effect of WGA-PCR template-dependent changes in the distribution of DNA fragments in the IN sample compared to the pre-amplification and starting material (figure 4.14) is a concern, given the fundamental role of this sample in normalising, at every probe, the IP value in the processed immunoprecipitated sample. This could be a significant source of the noise introduced into the DIP-chip assay patterns. Two approaches have been taken to ameliorate this potential source of variability. First, in the most experimentally simple approach, the IN sample is diluted to match the IP sample concentration. However, this does not improve the correlation between patterns produced by replicate samples (figure 4.17). The alternative is to boost the DNA in the IP sample, which, in combination with reducing the PCR cycles, does improve the correlation between replicate samples (figure 4.18). The effect of reducing the PCR cycles without increasing the IP sample template concentration is impossible to independently determine, as without template improvements this modification results in insufficient DNA for microarray hybridisation.

The results of the experiments presented in this chapter indicate that measures to improve the technical performance of the DIP assay and WGA-PCR amplification, specifically to focused on the IP sample, are the main areas of assay development that should improve the reproducibility of adduct patterns generated by the DIP-chip assay, as can be determined by assessing variability in

patterns from repeated samples obtained from the same individual. There could also be an alternative to this purely laboratory-based approach. It may be possible to apply bioinformatic tools to the large datasets generated by the DIP-chip to address and reduce the contribution of noise to the overall signal to noise ratio, and in this way improve the capacity of the assay to detect oxaliplatin-induced adduct signals between individuals. A combination of both laboratory and bioinformatic approaches is more likely to be successful.

One drawback with the current experimental approach to assay improvement is that the analysis tools available are not able to give a clear picture of the independent contribution of the separate IN and IP channels in the generation of variability in the assay, primarily because the only bioinformatic metric available is the measure of the IP/IN ratio and the correlation between IP/IN value patterns. As an further example of the limitations of the tools available for DIP-chip analysis, in the 18 repeat oxaliplatin treated PBMC samples displayed in figure 4.9, a sample from patient B, 100μM repeat 2 - indicated by a red star, fails to match any of the other samples. With the current Sandcastle tools for the analysis for DIP-chip datasets it is possible to detect that this experiment failed, but it is not possible to determine in which of the IP or IN samples (or both) the technical problem has arose, or at which stage of the assay failure occurred. These are fundamental questions that require appropriate tools to answer, as this type of experimental failure could provide useful information for assay modification and optimisation. This highlights the limitations of the current microarray analysis techniques available.

Inherent in the goal of this project is the development a model experimental pathway to allow the translation, optimisation and validation of this DIP-chip assay into other clinical tissues, such as tumour or biopsy samples. It is likely that during future projects similar issue with DIP-chip assay variability in different tissue types will manifest. The experimental approach outlined here, and in the previous chapter, can be used as a model pathway for determining the relative contribution of different stages of the multi-step assay and for optimising the assay for different situations.

One bottleneck, evident in this discussion, is therefore developing appropriate bioinformatic analysis tools to further assess, optimise and validate the DIP-chip assay. With this in mind, of critical importance is the development of novel tools capable of accurately measuring the independent effects of experimental modifications made to the IN and IP samples. These developments will be one of the two main aspects of the following chapter.

Finally, it has been demonstrated in this chapter that mean adduct pattern datasets generated as an average pattern from repeat DIP-chip experiments are more reproducible than single samples. With on-going developments in assay technique both single sample and average adduct patterns should become even more accurate. As the mean adduct patterns generated between PBMC samples from two different individuals match relatively well already (correlations of 0.68 – figure 4.12), rather than to improve the correlation between these patterns further, it is appropriate to use these datasets to develop tools to detect and determine the significance of adduct pattern differences between individuals – to detect biological signal - the ultimate aim of this project. The development of tools for this analysis will finally enable the use of the DIP-chip assay in clinical studies. This phase of the project will be discussed in the second section of the following chapter.

Chapter 5 The development of bioinformatic tools for the in-depth analysis of DIP-chip microarray data

The results and discussion of the previous chapter reveal that the optimisation of the DIP-chip technology to generate oxaliplatin-DNA damage patterns from human clinical samples poses significant experimental and analytical challenges, and demonstrates the advances made in these areas to date. Further efforts to optimise and enhance the reproducibility of the technique could improve the assay. The initial analysis indicates that the problem with relatively low reproducibility of microarray adduct profiles evaluated over multiple independent repeat experiments (Figures 4.7 to 4.9) appears to be primarily a consequence of WGA-PCR generated noise and bias introduced during the PCR-based amplification process (Figure 4.13). This is shown to be a consequence of the lower template in the IP sample than in the IN sample (Table 4.1 and figures 4.15-4.16), because the poor reproducibility between identical technical replicate samples is partially corrected by improving the template amount for WGA-PCR in the IP sample (Figure 4.18).

The results of the experiments described in chapters 3 and 4 indicate that the lower reproducibility of the PBMC oxaliplatin-DNA adduct patterns generated (a correlation value, at best, of 0.4-0.6) compares poorly to DIP-chip microarrays from pairs of repeat samples of dermal fibroblasts treated as cultured cells with a dose of 2500μM oxaliplatin during the assay development (with a correlation between two repeat samples of up to 0.74 (Powell 2014)). In view of the problem associated with the template-dependent PCR issue, this difference in reproducibility could be related to the relatively low amount of DNA present in the PBMC IP samples (~0.2% of the IN sample DNA amount) compared to the IP samples from dermal fibroblast experiments (~5% of the IN sample DNA amount) prior to the WGA-PCR amplification (Table 4.1). Significantly, the IN samples contain similar amounts of DNA in the dermal fibroblast and PBMC assays.

This highlights a central issue in the translation of the DIP-chip assay from one cell type to another. Unlike the dermal fibroblasts, PBMC samples have

a relatively low IP/IN ratio with an equivalent treatment schedule to that used in dermal fibroblast cell culture work (Chapter 3, figure 3.6, and (Powell 2014)) and the sensitivity of PBMC to platinum drugs appears to limit the dose of oxaliplatin to which the cells can be exposed. This is demonstrated by high rates of cell death (chapter 3 Figures 3.7-3.8) and deterioration the quality of extracted DNA at a 1mM oxaliplatin dose level (Chapter 3, figure 3.11). Additionally, the highest doses of oxaliplatin result in a worse correlation between adduct patterns (at 1mM a Spearman's correlation of 0.4-0.5 between repeats) compared to 100μM dose levels (a correlation value of 0.5-0.6 between repeats) (Figure 4.9). In an assay developed to predict clinical response by treating surrogate tissue in clinically relevant conditions in order to generate meaningful results, a dose of 100μM is still roughly 30 fold higher than the level used in patients (Chapter 1, section 1.2.1), compared to 300 times the patient dose with 1000μM and 1000 times the dose used to treat patients at 2500μM. At very high doses the biological relevance and the applicability of any generated platinum-DNA adduct patterns to the clinical setting may be less relevant.

In the context of this information, it is not feasible to modify the PBMC DIP-chip assay by increasing the oxaliplatin dose further to raise the level of adducts generated per DNA fragment, with the intention to improve the proportion of DNA in the IP sample following immunoprecipitation. Instead, technical improvements in the assay are required to improve the reproducibility, using the adduct levels and samples generated at the current oxaliplatin dose levels. For future translational projects, a useful application of the DIP-chip assay would be to use blood samples taken from patients during, or in the period after, chemotherapy – a true *in vivo* measure of oxaliplatin-DNA damage. For this type of application of the assay, or in the translation of the assay to other clinical tissues, technical improvements are also desirable to further develop the capacity of the reliable detection of DNA adduct patterns at even lower levels of drug exposure, or in tissues that are more sensitive, or more platinum resistant, than the PBMC samples used here.

It is also apparent that the analysis tools currently available for the interpretation of platinum-DNA adduct patterns (a visual inspection of adduct patterns and a single metric of a Spearman's correlation between probe IP/IN

ratios) are inadequate to permit a thorough understanding of the technical issues faced in the transfer of this technology. It is, therefore, necessary to develop more sophisticated analysis tools to interpret the available experimentally generated datasets in more detail.

A more sophisticated DIP-chip data analysis tool may also be useful as a quality control measure and to examine at which stage, and in which of the IP or IN samples, experimental failures occur, notably in the sample that failed in Figure 4.9 (Patient B, 100μM oxaliplatin treatment, repeat 2). This particular oxaliplatin-DNA adduct pattern fails to correlate with any of the other adduct patterns. The currently available analysis tools are unable to determine whether the issue lies with the IP or IN channel in the experiment, or possibly with a combination of both of these. The ability to compare platinum-DNA adduct pattern correlations at different sections of the same paired microarrays would also allow the determination of the impact of other factors on assay variability, importantly in the context of PCR noise such as the GC content of sections of the genome, which are well known to have an impact on PCR reproducibility.

So far the experiments in this thesis have focussed on reducing system 'noise' by reducing the differences between datasets generated by DIP-chip in repeated sampling of the same patients. In potential future applications of the DIP-chip assay, for example examining differences in adduct patterns between 'responder' and 'non-responder' cohorts of patients, a method of rapidly and accurately identifying the signal, the inter-individual differences between DIP-chip microarray datasets, is required.

The tools to facilitate these types of detailed DIP-chip data analysis will be developed in this chapter and used to analyse the experimental DIP-chip platinum-DNA adduct datasets presented in the previous section.

5.1 The examination of microarray datasets to enable a 'single channel analysis' of IP and IN samples

The current version of Sandcastle (version 1.0), the microarray data analysis software developed in our laboratory (Bennett 2013), allows the

calculation of, and graphical plotting of, the IP value corrected for the background fragment levels via the IN sample at each probe. This corrects each IP probe value for the relative level of DNA fragments present in the starting material, and, to a degree, for experimental variability as the paired IP and IN samples are processed throughout the assay in parallel. As the IP/IN ratio value is the proportion of fragments in the IP sample relative to the background, it is a measure of the relative immunoprecipitation efficiency between probes, and hence the relative amount of adducts at each loci. It is by plotting the (log_2) ratio of the IP to IN probe values at each probe location that the platinum-DNA adduct pattern is generated.

Sandcastle imports data to the R statistical package from the text file dataset generated by the Agilent Feature Extraction Software during the analysis of scanned microarray slides (see section 4.1.3.4 for a detailed description of this process). The data imported at each probe location is the fluorescence intensity at each feature/probe location of the background subtracted red and green channels, corresponding to the IP and IN values at each probe respectively. This data on the fluorescence intensity has been adjusted by the feature extraction software to account for background and adjacent probe fluorescence intensity, giving an accurate reading for each individual probe by removing the influence of adjacent probe fluorescence. During the import process, from the feature extraction output file to R by Sandcastle, the (log_2) ratio of the BGredsubsignal (the background adjusted IP values) and BGgreensubsignal (the background adjusted IN values) are calculated and retained in R as the log_2 IP/IN ratio.

It is, therefore, possible to manipulate the data by manually editing the feature extraction generated text file datasets to adjust each probe 'BGgreensubsignal' value to a new value of 1. Once this modified dataset is imported by Sandcastle to R, the 'ratios' value, used in the subsequent Sandcastle functions, becomes BGredsubsignal/1, altering the ratios value to the initial IP probe value. Correspondingly, by modifying each BGredsubsignal to a value of 1 the ratios values for the imported R data become 1/BGgreensubsignal, or 1/IN probe value. By further modifying the BGgreensubsignal to its reciprocal value in the initial data file, once imported the ratios value and associated sandcastle plotting and analysis tools give the probe values for the IN channel only. Three

microarray datasets of fluorescence intensity values are generated from the initial data, at each probe position – the initial data of the IP/IN values, the IP intensity values only, and the IN intensity values only respectively.

Following this modification these three datasets can be plotted simultaneously (figure 5.1). The platinum-DNA adduct profile (black) can be plotted, as in all previous adduct pattern plots, but now along with the probe intensity values for the IN (green) and IP (red) single channels (Figure 5.1 - upper plot). Once the BGgreensubsignal is corrected by modification the reciprocal value, the IP and IN probe values are juxtaposed, allowing a visual comparison of the two probe value patterns along the datasets (figure 5.1 - lower plot)

These modifications to the procedure for loading the data allow a new analysis of the data – referred to from this point as a 'single channel analysis', or SCA. This can identify the separate influence of the IN and IP values in the experiment, more clearly demonstrating why some experimentally generated datasets fail to correlate with one-another, and why some of the experimental protocol modifications discussed in the section above failed to improve the assay reliability (e.g. diluting the IN sample to match the IP sample pre-amplification - section 4.4.6).

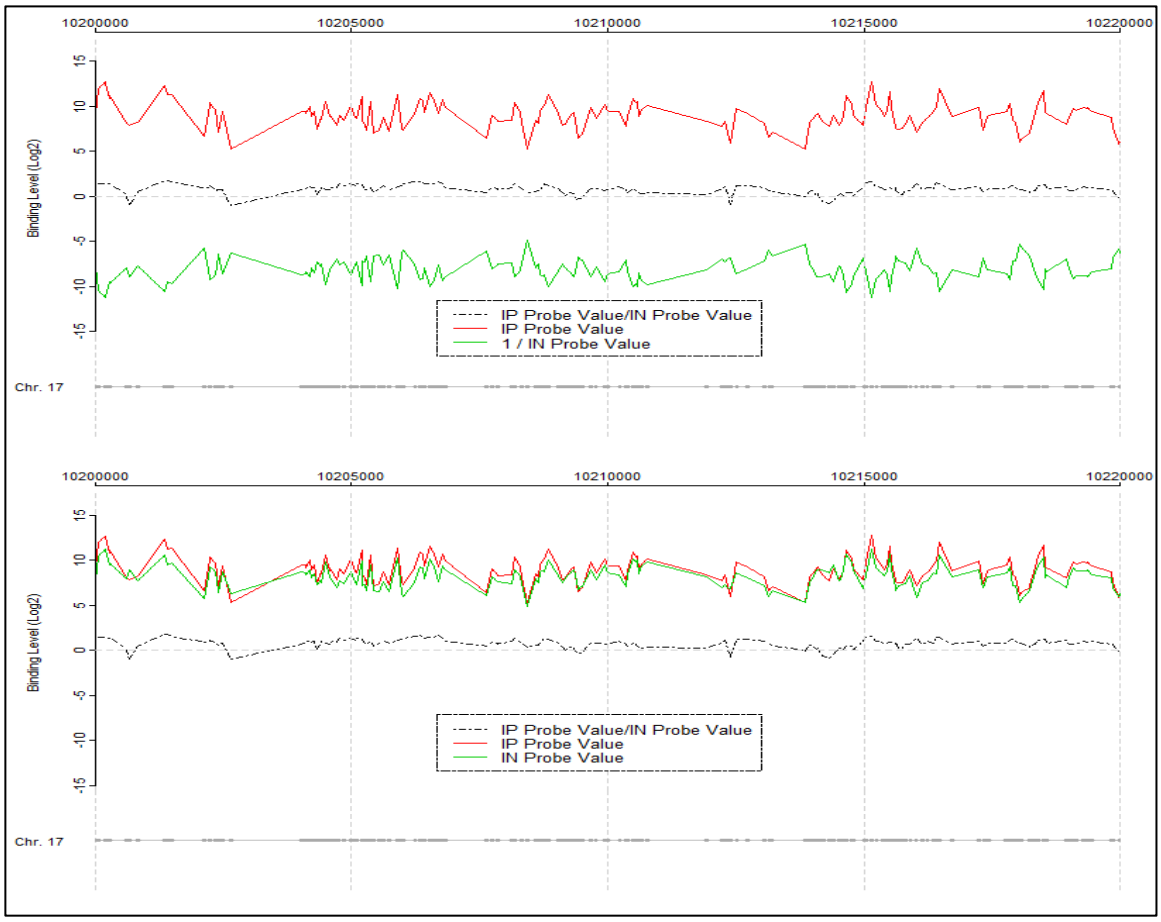


Figure 5.1: An example single channel analysis (SCA) plot.
An example single channel analysis (SCA) plot of modified microarray data from a 100μM oxaliplatin treated PBMC sample. In the top panel the initial modification is demonstrated with the oxaliplatin-DNA adduct pattern present (black), the single IN channel data (green) and the single IP channel (red) displayed. This demonstrates the relationship between the three datasets after initial data modification. The lower panel displays the same result after the IN data is further adjusted to modify the BGgreensubsignal value to the reciprocal of the BGgreensubsignal value, allowing easier comparison of the IN and IP patterns.

5.1.1 An example of the benefit of Single Channel Analysis

During the generation of DIP-chip data from oxaliplatin-treated PBMC from individuals A and B (Section 4.4.1) one of the independent repeats of the PBMC microarrays of patient B (repeat 2 at the 100μM dose level) displays an obviously different pattern to repeat 1 and 3, with a correlation value of 0.07 between repeat 2 and either of repeat 1 or 3. The generated adduct patterns from repeats 1 and 3 appear typical for this experiment, with a Spearman's correlation value of 0.48 between repeats 1 and 3 and 0.5-0.6 between other 100μM samples (Figures 4.9 and again in figure 5.2), indicating that the problem

lies with repeat sample 2. Using the standard Sandcastle analysis tools it is not possible to identify at which stage, or in which 'channel', the second experiment failed, only that the patterns do not match well. The oxaliplatin-DNA adduct patterns and corresponding correlation values are calculated using standard Sandcastle tools for figure 5.2.

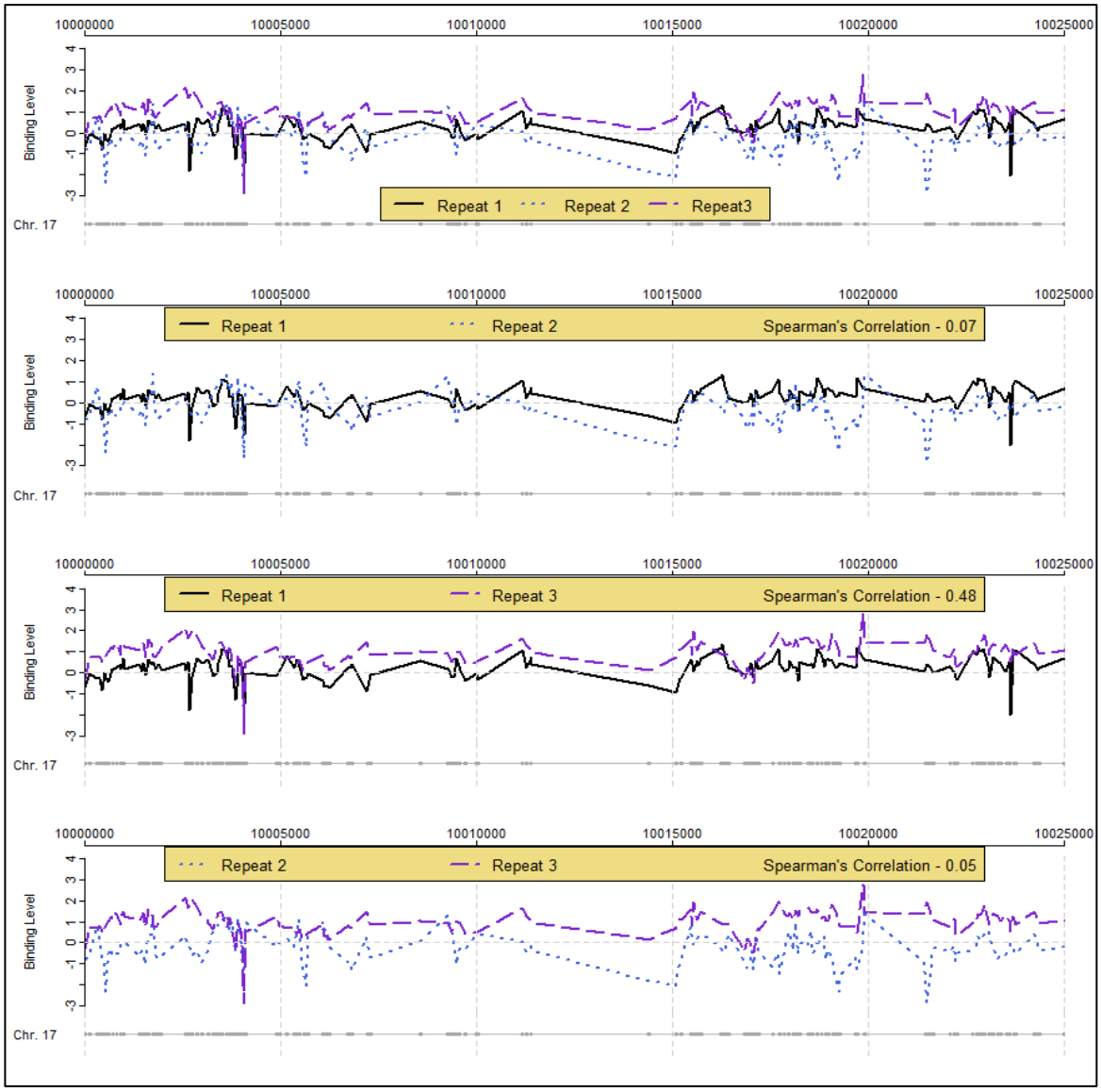


Figure 5.2: Oxaliplatin-DNA adduct pattern plots from three independent repeat experiments.
Oxaliplatin-DNA adduct pattern plots have been generated from three independent biological repeats of 100μM oxaliplatin treated PBMC from individual B, and all three patterns are shown in the top panel. The three lower panels show sequential comparisons between pairs of the repeat datasets and the Spearman's correlation value between the patterns. Repeat 2 has a correlation with repeat 1 of 0.07 and with repeat 3 of 0.05 (Second top and lowest panel). In the third panel (labelled repeat 1 vs. repeat 3) the adduct patterns generated in repeats 1 and 3 match more closely (a correlation of 0.48).

The result displayed in figure 5.2 demonstrates that adduct patterns in repeats 1 and 3 (the plot in the third panel labelled Repeat 1 vs. Repeat 3) correlate most closely, as shown by the Spearman's correlation coefficient of 0.48, and the visually similar appearing traces on the adduct pattern plot, whilst repeat 2 fails to match the patterns on the two subsequent pairwise plots (panels 2 and 4) and has a very low Spearman's correlation value (0.07 and 0.05). This identifies a discrepancy with this particular experiment, but does not demonstrate at which stage or in which of the IP or IN samples the problem has occurred.

After modifying the three 100μM oxaliplatin repeat microarray datasets with the data alterations as described above to allow single channel analysis it is then possible to plot and calculate the Spearman's correlation values between the separate IN and IP single channels. These single channel IN (green) and IP (red) patterns are plotted in figure 5.3 for the same data.

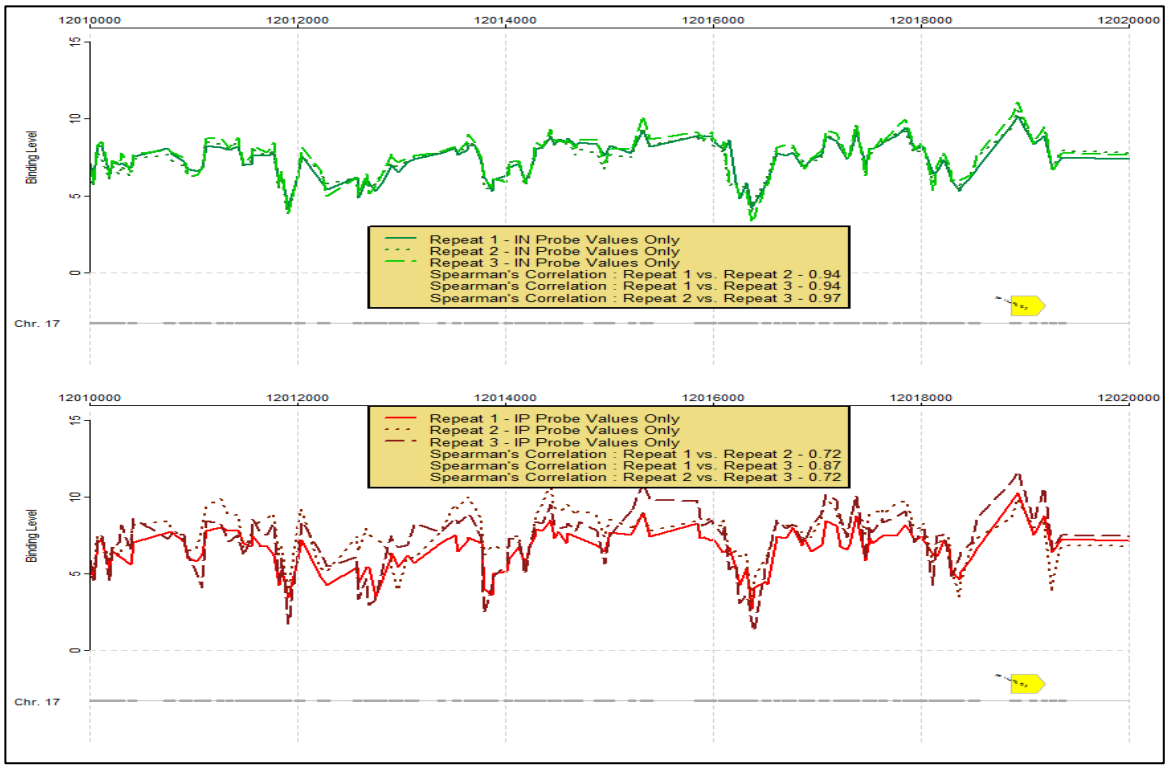


Figure 5.3: SCA plot of three repeat experiments.
A single channel analysis plot of the IN and IP single channel patterns from 3 independent repeats of 100uM oxaliplatin treated PBMC from individual B. The IN channels of the 3 repeats are shown in the upper panel (green) and the IP channels in the lower panel (red). The correlation values between each repeat in each group of IN or IP channels is shown in the legend.

Using the single channel analysis approach it is clear in figure 5.3 that the IN samples (green) have a very high degree of reproducibility, demonstrated by a high correlation value (Spearman's correlation between patterns of >0.94), and with no visible difference evident between the patterns generated and plotted from the data. This is a high degree of consistency, considering that 44,000 data points are involved and the samples were processed independently over the course of a six-week period. The IP channel (bottom panel, red) shows a more variable pattern in the second sample, compared to the first and third repeats which match one-another better, in keeping with the Spearman's correlation values displayed of 0.87 between repeat 1 and 3 and 0.72 between either pattern 1 or 3 when compared to pattern 2.

Additionally, a 'single channel analysis' matrix (an SCA matrix) of Spearman's correlation values between the IP/IN patterns (the oxaliplatin-DNA adduct pattern) and between single IP and IN channels can been generated, and is displayed in figure 5.4. The oxaliplatin-DNA adduct patterns (the IP/IN ratio patterns) are shown the left top box (yellow), the IN single channel patterns are in the central box (green), and the correlation between IP channels are displayed in the right lower box (red).

		Rpt	1	2	3	1	2	3	1	2	3	Rpt	
	Rpt	Type	IP/IN	IP/IN	IP/IN	IN	IN	IN	IP	IP	IP	Type	Rpt
Patient B 100uM Oxaliplatin	1	IP/IN		0.07	0.48							IP/IN	1
Patient B 100uM Oxaliplatin	2	IP/IN	0.07		0.06							IP/IN	2
Patient B 100uM Oxaliplatin	3	IP/IN	0.48	0.06								IP/IN	3
Patient B 100uM Oxaliplatin	1	IN					0.94	0.94				IN	1
Patient B 100uM Oxaliplatin	2	IN				0.94		0.97				IN	2
Patient B 100uM Oxaliplatin	3	IN				0.94	0.97					IN	3
Patient B 100uM Oxaliplatin	1	IP								0.72	0.87	IP	1
Patient B 100uM Oxaliplatin	2	IP							0.72		0.72	IP	2
Patient B 100uM Oxaliplatin	3	IP							0.87	0.72		IP	3

Scale	IP/IN	vs.	IP/IN
	IN	vs.	IN
	IP	vs.	IP

1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0
1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0
1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0

Figure 5.4: A SCA matrix of three independent repeat experiments.
A single channel analysis (SCA) matrix of Spearman's correlation values between oxaliplatin-DNA adduct (IP/IN) patterns (yellow), single IN channels (green) and single IP channels (red) from 3 independent repeats of 100uM oxaliplatin treated PBMC from individual B. The higher the correlation value the darker the box colour, as indicated in the key below the main plot.

This matrix approach allows a visual overview of all of the correlation values between samples, a useful approach when many samples are compared simultaneously and the relative colour intensity allows a rapid appreciation of the numerical similarities and differences between pairs and groups of samples.

The SCA matrix demonstrates that patterns generated from the single channel IN data match very well, as shown by correlation values of 0.94-0.97. This indicates that IN samples are highly reproducible in these experiments. The IP patterns match less well (correlation values of 0.72 to 0.87), and significantly lower in pairwise comparisons involving sample 2 (correlations of 0.72) than between samples 1 and 3 (correlation 0.87). The decrease in correlation between IP sample 2 compared to between 1 and 3 is therefore likely to be responsible for the failure of the IP/IN adduct pattern to match the other repeat samples (correlations of 0.05 and 0.07), and indicates that there has been problems with the IP experiment in sample 2, whereas the IN sample appears to have been purified, amplified, labelled and hybridised appropriately. A small reduction in IP channel reproducibility, when subsequently used to calculate the IP/IN ratio, results in a small difference that is multiplied when the data is used at 44,000 data points to generate an IP/IN pattern. As the IN channels are highly reproducibly, it must be this small variation in IP channel consistency that is sufficient, through the introduction of subtle variability at each of 44,000 values, to introduce significant noise into the IP/IN adduct pattern to result in a dramatic fall in correlation between IP/IN adduct patterns.

This example demonstrates the use of SCA for data interpretation and for quality control assessment of DIP-chip data. Prior to this development it was possible only to identify a failed DIP-chip assay. It is now possible to identify in which sample the issue occurred using SCA, information that can be compared to laboratory records to identify assay failings. As SCA can identify subtle differences in single samples, it is potentially a useful tool in assay development and validation experiments, as will be shown in subsequent sections.

5.1.2 Extracting latent information from microarrays: The development of a custom R functions for single channel analysis to produce SCA plots and SCA correlation matrices

The above example of single channel analysis (SCA) demonstrates the power of this approach to allow a more detailed interpretation of DIP-chip microarray data, specifically in determining the influence of changes in IN and IP samples on the ultimate platinum-DNA adduct pattern. Application of the SCA approach to microarray data analysis will be applied to the experimental attempts to improve the DIP-chip reproducibility, conducted and described in the previous chapter, in order to accurately determine the effect of experimental and analytical changes on either, or both, of the single IP and IN channels. This approach may shed new light on why the experimental changes intended to improve assay reproducibility met with only limited success, as described in chapter 4.

In order to perform this more detailed analysis, and to develop the SCA process as a tool for future studies, the SCA technique was automated by the development of a series of R functions capable of taking initial standard DIP-chip microarray datasets and to rapidly generate two graphical outputs - SCA pattern plots of single channel data (as in Figure 5.3) and SCA correlation matrices (as in figure 5.4) - allowing visualisation and comparisons between adduct patterns (IP/IN ratios), IN and IP single channels simultaneously.

In the discussion and description of the R functions below, the following terminology is used:

Function: An R function is a discrete package of R script that takes a series of inputs (in the form of arguments), and returns an output. The output is typically derived from a series of calculations and sub-functions, such as graphical plot commands, specified in the body of the function. In essence, an R function binds together a series of R programming steps, simplifying potentially lengthy analysis and graphical operations into a consistent single integrated

procedure. In the discussion below the names of individual functions are noted in bold font.

***Argument*:** An argument provides information as an input into an R function, and is required to generate the output from the function. Arguments can be essential for the generation of function output (i.e. which datasets to use for the analysis) or can be optional (i.e. specifying the text size of the graphical output). Arguments can be pre-set to a default value if optional. In the description below arguments when named are in italic font.

Vector: In the context used below, a vector is an object that can contain information in the R environment. In the description below vectors when named are underlined.

For each of the five functions described below, the full R script code is included in appendices 2 to 4, and an electronic version is included in the electronic appendix.

5.1.2.1 SCA ancillary functions

The goal of the work as described here is the development of custom R functions that will allow automated data entry, analysis and graphical outputs, generating a single channel analysis correlation matrix and plot from DIP-chip microarray datasets. This allows the procedure to be consistently and easily performed by any member of the laboratory team, using a simple line of R code to call the function and to specify the arguments specifically required for each analysis.

Firstly, the data must be entered in the correct format for SCA, and is then analysed to generate correlation matrices, which can subsequently be plotted as SCA adduct-pattern plots or as correlation tables. This process is conducted through a set of three SCA ancillary functions, named **SCAdataload, corsCalc**

and **corsPredCalc.** These three SCA ancillary functions are available in appendix 2, which lists the source code required for these preparatory functions.

The function **SCAdataLoad** is run first. This function uses pre-set procedures available in Sandcastle (version 1.0) to extract data from the text file output from Agilent's feature extraction software, generated from the required microarray datasets in the working directory. The function stores information from these datasets to the vector SCAdata, in the order of: 1) adduct pattern data (for all datasets present sequentially) followed by 2) the probe IN values, followed by 3) the probe IP values. For example, for three microarray datasets A, B and C the data is stored in the order A-IP/IN ratios, B-IP/IN ratios, C-IP/IN ratios, A-IN values, B-IN values C-IN values, A-IP value, B IP values, C IP values.

The function **SCAdataLoad** can be modified by specifying the optional argument '*all=FALSE*', allowing the specification of individual microarray data files to load for analysis and resulting in the generation of an on-screen request for the selection of the files to be loaded by identification of the dataset by the order they appear in the working directory (i.e. only the 1st, 3rd and 7th datasets can be used for SCA analysis by entering these numeric identifiers when requested).

The ancillary functions **corsCalc** and **corsPredCalc** are then used to calculate the Spearman's correlation between the datasets that have been loaded by **SCAdataLoad.** The **corsCalc** function calculates the correlation between all datasets and between each single channel, and the output must be stored in the vector 'cors'. For the prediction to be accurately generated, the prediction profiles are required to be loaded to R with Sandcastle in the vector predPlat for the platinum profile and the vector predUV for the UV prediction profile, if these are required for the analysis. The function **corsPredCalc** can then be run to calculate the correlation between predicted profiles and microarray single channel data. The output must be stored to the vector 'corspred'.

5.1.2.2 SCA plot and correlation matrix functions

Once the ancillary functions have been run the data can be used to generate a SCA plot (using the function **SCAplot**) or an SCA correlation matrix (with the function **corsPlot**) (full scripts for these two functions are available in appendix 3 and 4).

To plot a SCA plot (as in figure 5.3 and figure 5.5) the function **SCAplot** is used. This function uses the data loaded with preparatory function **SCAdataLoad** to plot the platinum-DNA adduct patterns from each dataset in the upper plot, the IN values in the middle plot (patterns displayed in green) and the IP values in the lower plot (patterns displayed in red).

The default region plotted is chromosome 17, nucleotides 10050000 to 10075000, and can be adjusted with the optional arguments '*start*' and '*end*'. The arguments *legends*, *labels*, *highlight* and *datatohighlight* can also be used. If the argument *legends=FALSE* is used this will remove the data legends. If *labels=TRUE* is used the name of each dataset will be prompted for, and subsequently displayed alongside the plot. To highlight a dataset the argument *highlights* can be set *highlights=TRUE* and a numerical value (n) used for *datatohighlight=n* to identify which dataset to highlight. (see figure 5.5 for an example)

The function **corPlot** is used to generate and label an SCA correlation matrix of Spearman's correlation values in the format used in figure 5.4 and figure 5.6 A and B. This standardised format will be used for all SCA correlation matrices. For a breakdown of the format see figure 5.6B.

The IP/IN (the adduct pattern) correlations are presented in the upper left (the analysis labelled A in figure 5.6B), the correlation between IN single channels in the central green box (the analysis labelled B in figure 5.6B), and correlation between IP single channel in the left-lower red box (the analysis labelled C in figure 5.6B). Correlations between IN and IP single channels can be seen in the central lower box (the analysis labelled X in figure 5.6B) and between IP/IN ratios and IN and IP single channels in the left hand middle and lower box respectively (the analysis labelled Y and Z in figure 5.6B). In all cases the

strength of the correlation is graded by increasing darkness of colour as shown by the key to the right hand of the plot. The colour scale used for each key is adjusted to reflect the typical range of values seen with the SCA correlation matrix. The single IN and IP channel correlations are typically in the range of 0.85-1. The IP/IN patterns are derived from the IN and IP single channels, and as such the error in this channel is due to a combination of the error in each single channel, resulting in a greater variability in correlation values in the IP/IN pattern compared to the single channels, and this is reflected in differences in the key scales. For example, a colour gradient begins at above a correlation of 0.5 is used in the IP/IN region, but in the single channel IN or IP section a colour scale above values of 0.85 is used. The correlation between experimentally generated samples and the mathematically generated predicted profiles, if used, are shown below the main matrix with the colour gradient key as applied.

The optional arguments *prediction=TRUE* can be used to add the correlation values of the experimentally generated data to the mathematically generated prediction profile (using the arguments *UV=TRUE* and/or *platinum=TRUE* to select the prediction profile to use). The arguments '*boxtext*' and '*FigSize*' can be numerically set to scale the correlation value text size and the figure size respectively, useful for visualisation purposes (the default is 1). The argument *labels=TRUE* will result in a call for defined label names for each sample. For ease of display the key can be removed using the argument *key=FALSE*.

An example of the output plots generated by this **SCAplot** and **corPlot** functions are shown in figure 5.5 and 5.6 respectively, and are created with the DIP-chip datasets used for the example single channel analysis from in the previous section 5.1.1.

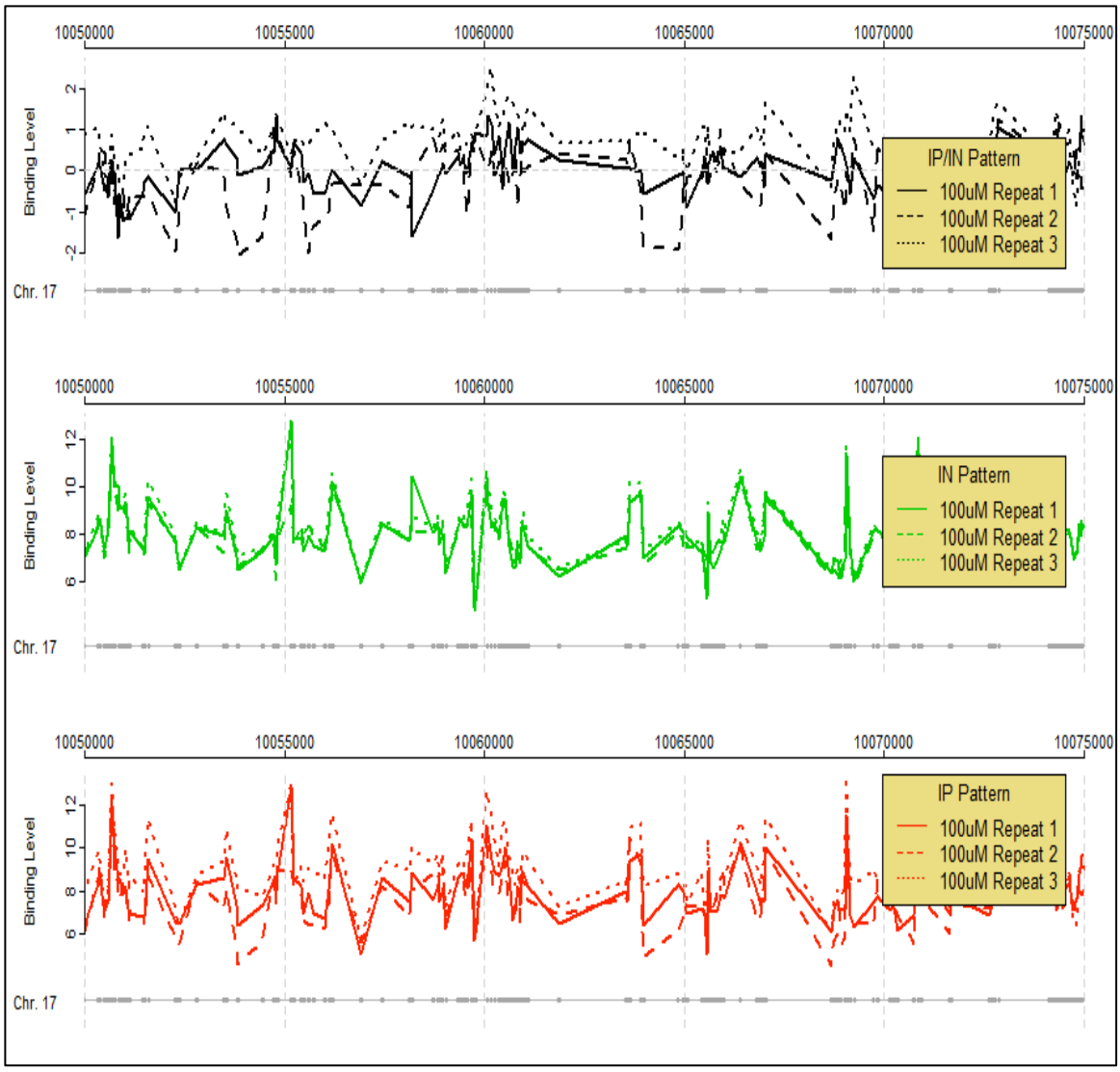


Figure 5.5: An example plot generated by the SCAplot R function.
An example plot generated by the SCAplot R function developed to allow efficient single channel analysis of DIP-chip microarray datasets. Three repeats of oxaliplatin-DNA adduct pattern microarray datasets from 100μM oxaliplatin treated PBMC from patient B are shown. The top plot is of three platinum-DNA adduct patterns, the middle plot is the single IN channels (green) and the bottom plot is the single channel IP patterns (red) generated from the three datasets.

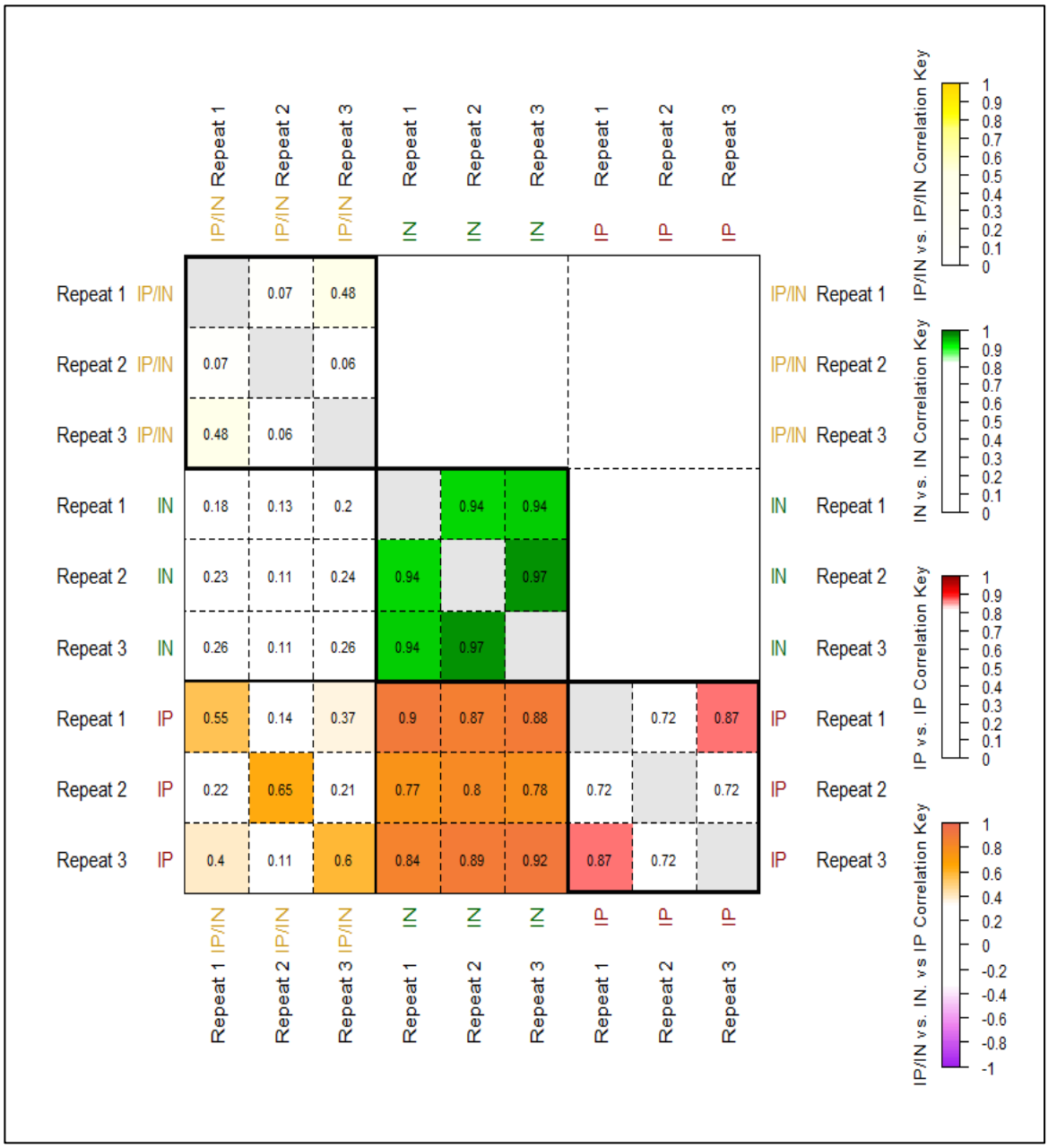


Figure 5.6: (A) An example of the SCA matrix produced by the corPlot function.
An example of the SCA matrix produced by the corPlot function developed to enable rapid single channel analysis of DIP-chip microarray datasets. This figure is a recreation of figure 5.4, using three repeat 100μM oxaliplatin treated PBMC sample datasets to compare the IP/IN ratios (upper left box – darker yellow with increasing Spearman's correlation), the single channel IN samples (central box – darker green for higher Spearman's correlations) and lower left box for the single channel IP relationships (darker red for higher Spearman's correlations). The correlation between patterns generated between IP and IN can be seen in the central lower box, and between IP/IN ratios and the single channel II or IP patterns in the left centre and lower boxes respectively

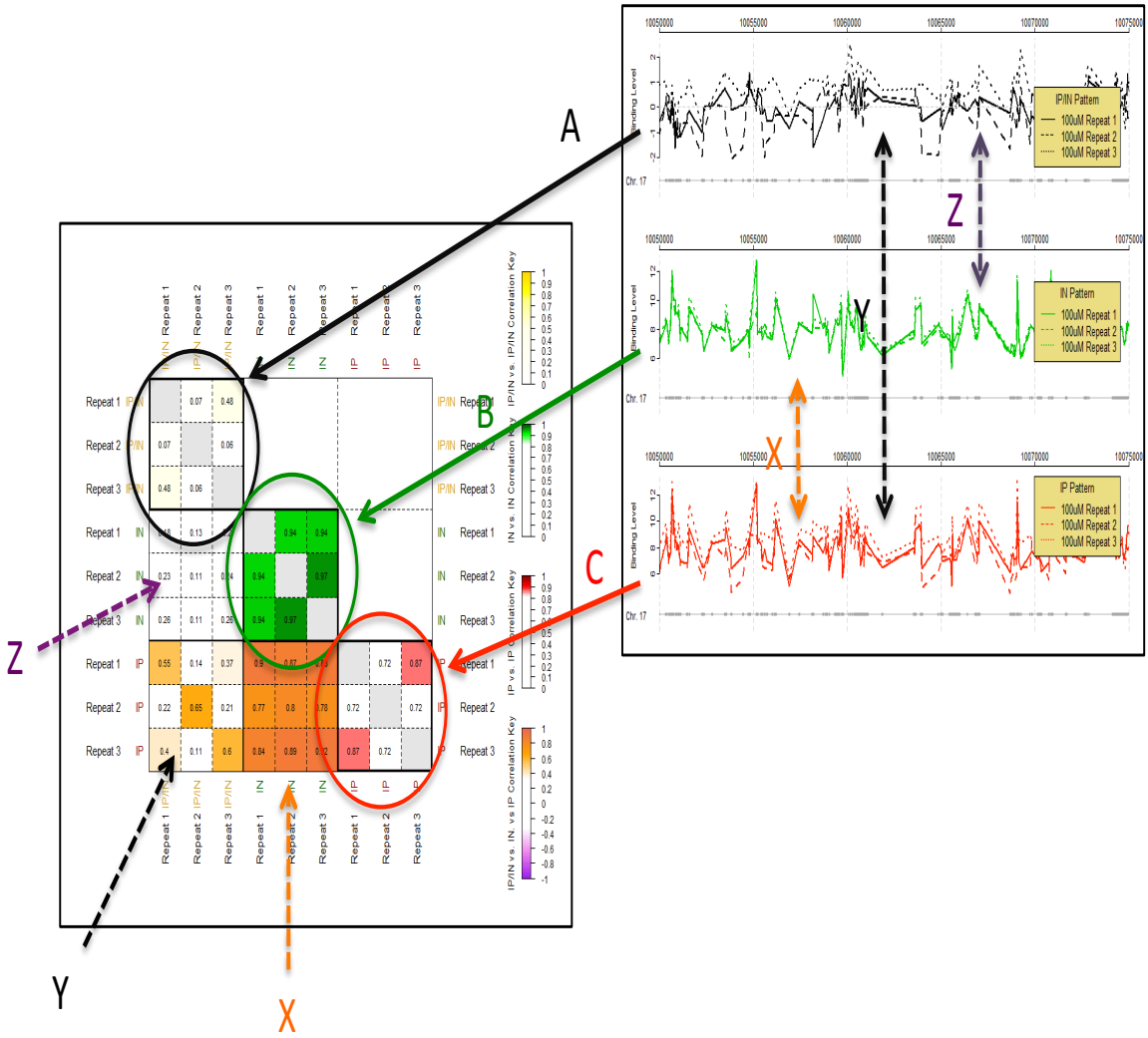


Figure 5.6 (B): An example of the SCA matrix produced by the corPlot function

This figure shows the relationship between the SCAplot (right) and SCA correlation matrix (left) generated by the SCAplot and corPlot functions and presented in figures 5.5 and 5.6A above.

The three key relationships are demonstrated and are between:

A) IP/IN adduct patterns in the top of the SCAplot (black) and the correlations between these patterns (analysis A) presented in the top left of the correlation matrix.

B) The relationship between IN single channels (green) as plotted in the SCAplot central plot is shown in analysis B and presented in the central box of the correlation matrix.

C) The relationship between single channels is shown in red in the lower plot of the SCAplot and the lower right box of the correlation matrix.

Three further relationships are also shown here:

X) Analysis X shows the relationship between single IP and IN channels, and is displayed in the central lower box of the correlation matrix.

Y) The relationship between the IP single channels and the IP/IN adduct pattern is shown by Y and is displayed in the lower left box.

Z) The relationship between IN single channel patterns and the IP/IN adduct pattern is shown by Z and displayed in the left central box.

5.2 Single channel analysis of oxaliplatin-PBMC microarray datasets

In this section the SCA method will be used to re-analyse the DIP-chip microarray datasets generated in the experiments described in the previous chapter. Firstly, in section 5.2.1, this analysis will be conducted to examine the relationship between single IN and IP channels in human PBMC experimental repeat samples to determine the reason for low reproducibility between biological repeat samples. In sections 5.2.2 to section 5.2.4 the SCA method will then be applied to technical replicate experiments to further identify the cause of poor reproducibility and to demonstrate in which channel experimental modifications to improve the protocol are acting. Finally, in section 5.2.5, the SCA method will be applied to dermal fibroblast datasets generated in cell culture conditions during the DIP-chip assay development. This analysis will determine the level of single channel reproducibility that is required to generate reproducible DIP-chip datasets in clinical samples.

5.2.1 A single channel analysis of oxaliplatin-treated PBMC DIP-chip datasets

In section 4.3.1, DIP-chip assays to generate 18 microarray datasets were performed. The DIP-chip experiments discussed were conducted on oxaliplatin treated PBMC from patients A and B at 4 dose levels (untreated, 10μM, 100μM and 1000μM), with correlations between the resulting microarray data and oxaliplatin-DNA adduct patterns at best of 0.6 between repeated experiments in the same experimental condition (inter-sample correlations are displayed in figure 4.9).

In the following SCA analysis it is intended to explore in more detail the correlations and patterns evident between the oxaliplatin-DNA adduct patterns and the single IN and IP channels, with the intent of identifying in which channel and in which samples noise is being introduced into the assay. SCA results in a large matrix of correlation values (figure 5.7), a format that is less useful for visualising the individual correlation between samples, but more useful for

obtaining an overview of the relationships between the various samples and channels. To allow this to be viewed adequately, the data is included in the electronic appendix so more information or a higher resolution image can be viewed. The correlation matrix has been transferred from the R function to an excel output in this instance solely to allow presentation of this large dataset and matrix in the format required for this thesis.

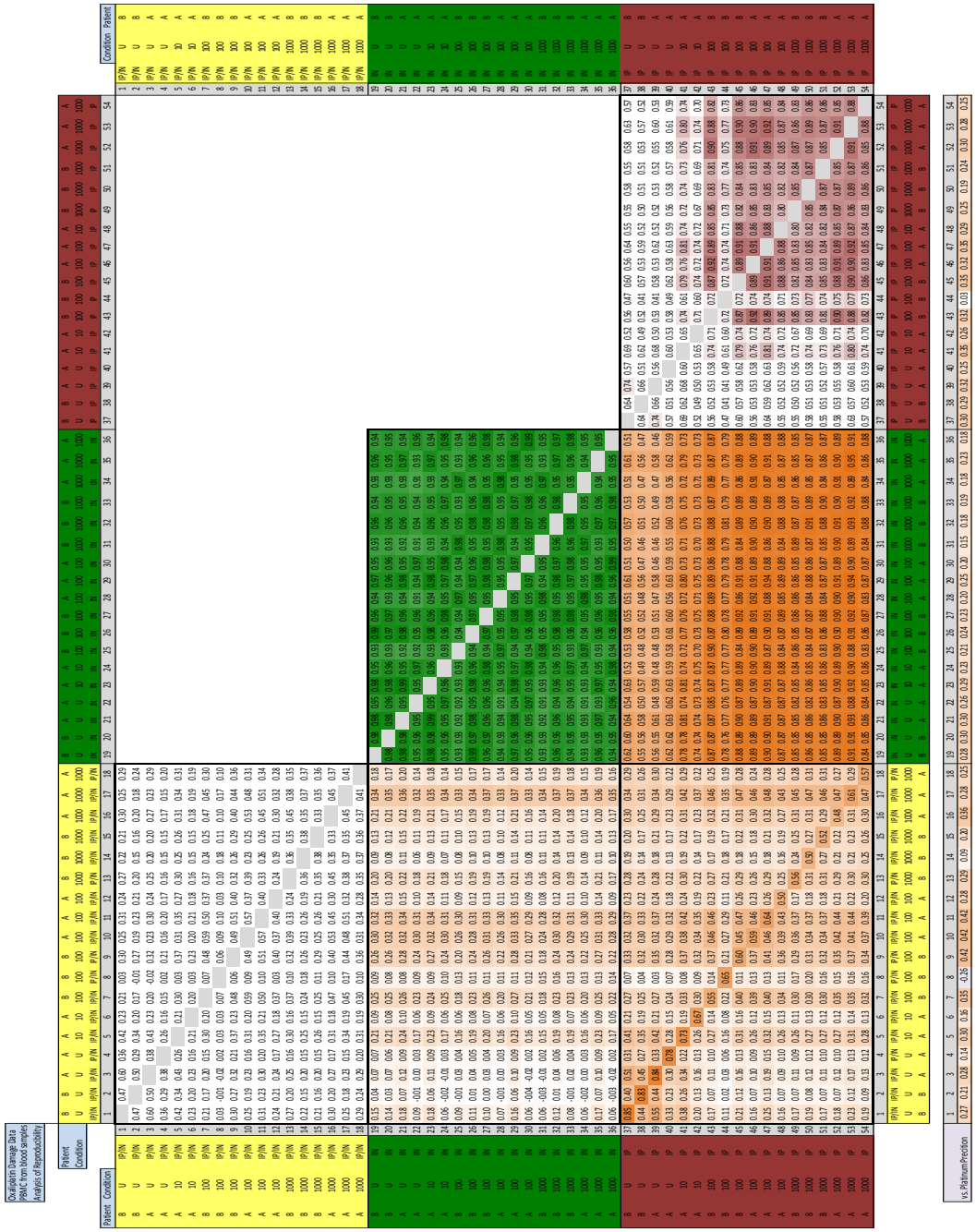


Figure 5.7: An SCA matrix of Spearman's correlations values between 18 DIP-chip oxaliplatin-treated PBMC datasets.
Individual small boxes are coloured to indicate the strength of the correlation between patterns indicated by the colour in the key below the plot. The 6 larger, bold outlined boxes

are used to give an overview of the relationships between IP/IN adduct patterns (upper right) and patterns of single channel IN samples (green) and IP channels (red). The left upper box displays the correlation between IP/IN adduct patterns (all between 0.0 and 0.6) and previously documented in figure 4.9. The central green box shows the relationships between the IN single channels, and all correlations are >0.9. The right lower red box displays the relationships between the patterns in the IP channels for each of 18 samples. The correlation between samples increases with increasing oxaliplatin dose. The left central and lower boxes show the correlation between IP/IN adduct pattern and single channel IN and IP respectively (correlations typically 0.0-0.4 IN/IP vs. IN and 0.0-0.6 IP/IN vs. IP). The central lower box shows a high correlation ~0.9 between IP and IN single channel patterns in samples treated with a higher dose.

In each small box in the matrix is the Spearman's correlation co-efficient between the pattern generated in the samples and channels labelled along the columns and rows. In each column the sample data is arranged in the order IP/IN ratio (i.e. the IP/IN pattern representing the platinum-DNA adduct pattern (coloured yellow)), IN single channel for each sample (coloured green) then the IP single channel for each sample (coloured red), arranged sequentially by individual PBMC donor then by increasing oxaliplatin dose for each sample followed by experiment repeat number.

Several overall patterns can be inferred from visualising the data in this fashion by breaking the matrix into four sections to analyse: the yellow box showing the correlations between IP/IN patterns, the green central box displaying correlations between IN single channel patterns, the red lower left IP single channel pattern correlation box, and finally the boxes comparing between channels and adduct patterns in the left central, left lower and central lower boxes.

In the top left box is the IP/IN relationships (i.e. between the platinum-DNA adduct patterns) between all 18 samples, with clear boxes below a cut-off of a correlation value of 0.7 and yellow boxes showing values closer to 1, a 'perfect' match between patterns. As no boxes are coloured, the correlation between any of the samples is <0.7, as previously shown for this data in chapter 4, figure 4.9.

The central large IN channel box (dark green) shows the correlation values between IN channels for each of the 18 datasets, with values less than 0.7 in clear boxes and with the samples darker in green colour the closer the value to 1. It is clear from the homogenous dark green box that all of the IN channels

correlate with one another highly, i.e. they have a very similar pattern in each experiment, with correlation values between each sample of 0.93-0.99.

The right lower IP channel large box (red) shows the inter-IP channel correlations, with values below 0.7 in colourless boxes, and increasing colour red boxes as the values increase from 0.7 to 1. The lower right aspect of the IP box demonstrates improving correlation between the IP channels for these samples, corresponding to an increase in oxaliplatin dose, with the highest (darkest red) correlations between 100uM samples. The untreated and 10uM samples have correlation values of 0.5-0.7, lower than the 100uM and 1000uM samples, corresponding to more variability between patterns at these doses.

This analysis confirms the hypothesis, outlined in the final discussion section of chapter 4, that the failure of the patient A and patient B PBMC repeat DIP-chip adduct profiles to match well between individual repeat experiments under the same conditions is due to a failure of the IP channel to be reliably reproduced. The IN channels all correlate highly (values >0.92), indicating that the purification, amplification, labelling and array hybridisation are, in these samples, highly reproducible. In the IP channel there are clear increases in the assay reproducibility with increasing doses of drug, until at the highest 1000μM dose there is a slight decrease in reproducibility (0.85-0.9) in the IP sample compared to the 100uM dose level (>0.9). This may be accounted for by the decrease in DNA quality demonstrated by the degraded and smeared DNA samples extracted from these samples and shown by gel electrophoresis in the chapter 3, figure 3.11.

This analysis indicates that there is potential for increasing the reproducibility of the datasets, by modification and optimisation of the workflow to increase the reproducibility of the IP sample. This step is important to improve the reproducibility of adduct patterns in these types of clinical samples. Because of the drop in reproducibility at 1000μM compared to 100μM doses (a fall in correlation of adduct patterns from 0.5-0.6 to 0.4-0.5 and IP patterns from >0.9 to 0.85-0.9) it appears that an improvement in reproducibility cannot simply be achieved by increasing the dose of drug, as this results in degradation of the DNA, increased apoptosis of cells during *ex vivo* treatment, and arguably

reduces the biological relevance of any subsequent findings. As previously discussed, central to improving the reproducibility is improving the technical performance of the assay, now demonstrated to be failing because of variability in the IP channel.

5.2.2 Examining IP amplification fidelity using single channel analysis

In this section, SCA analysis will be conducted on microarray datasets generated in the previous chapter to determine the effect of protocol changes with a focus on the reproducibility of the single IP channel.

Previous analysis in the experimental results documented in the last chapter, section 4.3.6, demonstrated that reducing the IN concentration prior to amplification to match the concentration of the IP template prior to WGA-PCR reduces the reproducibility of the damage profile (figure 4.17). Further experiments (section 4.3.7) indicate that increasing the IP template amount, with a corresponding reduction in PCR cycles, improves the reproducibility of the assay (figure 4.18). In this section these datasets will be re-analysed using single channel analysis to attempt to demonstrate exactly how these changes cause variability in the resulting adduct profiles, and to attempt to confirm the hypothesis that poor reproducibility is a result of failure of the IP channel only.

In the experiment in section 4.3.3, identical technical replicates were run in parallel, and divided into either pre-amplification, pre-labelling or pre-hybridisation, and on the same or different arrays to measure the inter- and intra-array reproducibility, demonstrating poor reproducibility in identical IP and IN samples divided pre-amplification (correlation of 0.43 between adduct pattern) compared to post-WGA-PCR amplification (correlation between all other technical replicates of >0.93) (Figure 4.13). For this single channel re-analysis, an SCA matrix and SCA plot of the experimental data previously presented in section 4.3.3 is shown below in figure 5.8 and 5.9. For clarity, only the three key boxes of the IP/IN pattern correlation (upper left), single IN channel correlations (central) and single IP channels (lower right) are shown. The full matrix is available in the electronic appendix and in higher resolution.

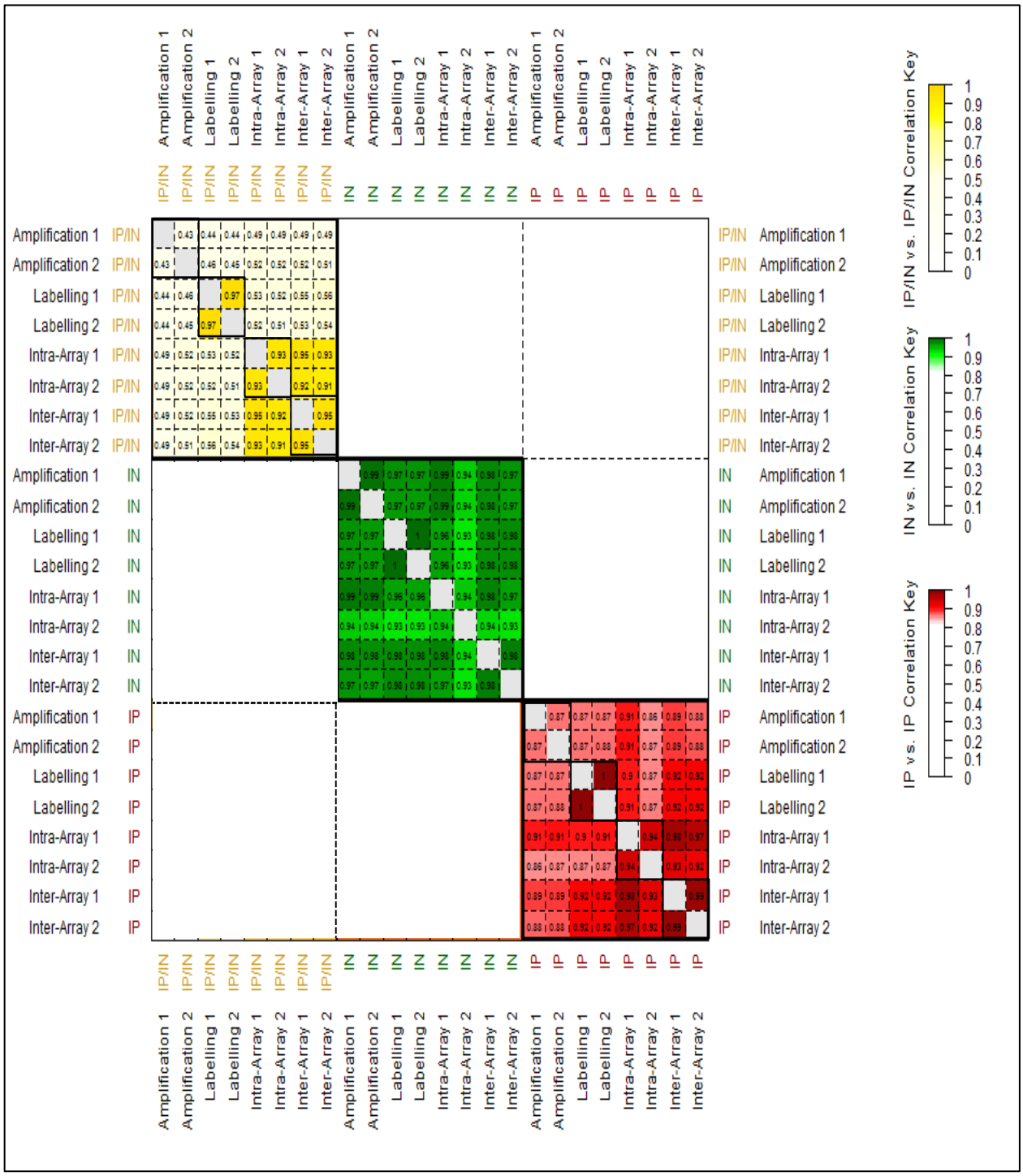


Figure 5.8: AN SCA matrix from 4 pairs of technical replicates at different stages of the assay.

An SCA matrix showing correlation values between IP/IN ratios (platinum-DNA adduct patterns) (top left yellow box), single IN channels (central green box) and IP single channels (lower right red box). Samples are from 4 pairs of technical replicate oxaliplatin treated PBMC samples divided immediately pre–amplification, pre-labelling, or between microarrays on the same slide (intra-array) and between microarrays on separate slides (inter-array). The smaller individual boxes indicate the individual correlations between replicate pairs.

The top-left large box of figure 5.8 displays the correlation between IP/IN adduct profile correlations, as previously discussed in section 4.3.3, showing a relatively poor correlation (0.43) between technical replicates if run in parallel pre-amplification, and a high correlation between replicates if run post-amplification (a correlation between paired replicate samples of >0.93).

In the single channel boxes (green and red) the reason for this can be immediately discerned. The IN channel, shown in the centre box, with darker green being boxes with values closer to 1 and below 0.7 as colourless boxes, demonstrates that the stage of the assay at which the replicates are divided has no effect on the IN correlation values, all of which are above 0.94. However, the IP values (shown in the right bottom box with a scale of below 0.7 as colourless, increasing to 1 as darker red) demonstrate that replicates run in parallel from pre-amplification have a lower correlation of 0.87, where as pairs of replicates run pre-labelling, or when investigating inter- and intra-array hybridisation, show high correlation between samples (>0.94), confirming that the noise in the damage profile repeats is generated through the amplification of the IP sample.

These findings also can be demonstrated by using an SCA plot to show the profiles as seen in figure 5.9, with the top panel showing the oxaliplatin-DNA adduct pattern of three of the samples – a pre-amplification sample (blue), intra-array sample (orange) and inter-array sample (brown). The IN and IP single channel plots for each of these three samples are shown in the panel below in green (IN) and red (IP). The most variable sample is the pre-amplification pattern (blue), one of the replicates which has been split pre-amplification is dotted whereas the other two samples (which correlate well) are the solid lines in each plot. It is clear that the pattern variation as shown by the correlation co-efficient corresponds to the pattern changes in the pre-amplification compared to the post-amplification samples, and again demonstrates that the IN is highly reproducible in all of these conditions, and that it is the IP that varies dependent on the amplification process, resulting in a very different damage profiles in the pre-amplification replicates despite identical samples being used.

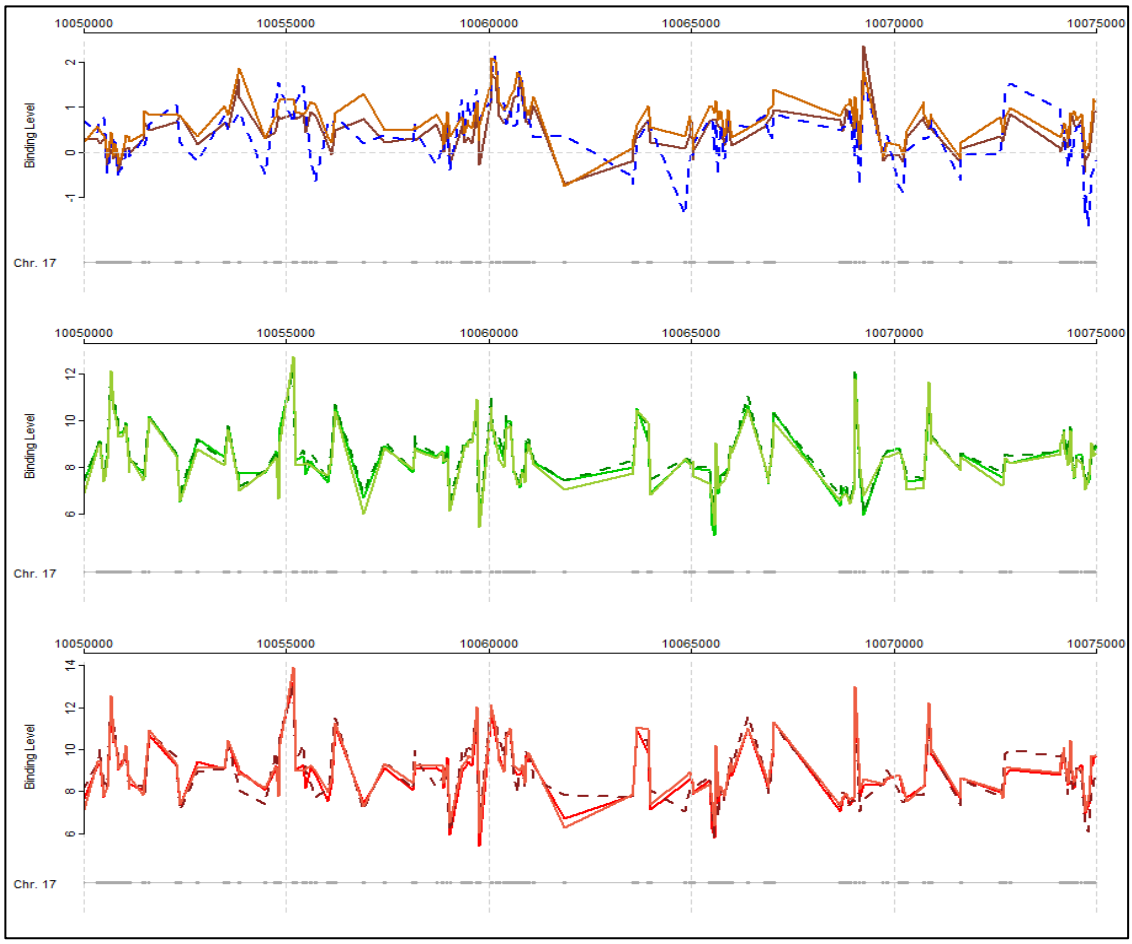


Figure 5.9: A single channel analysis plot of technical replicate samples from oxaliplatin treated PBMC DNA.
Samples are run in parallel pre-amplification (dotted lines) or post-amplification (solid lines). The upper panel shows three damage patterns, with a pre-amplification replicate in blue, a pre-labelling replicate in orange and a pre-hybridisation replicate in brown. The middle and lower panel shows the single channel IN (green) and IP (red).

5.2.3 Single channel analysis of DIP-chip data generated during changes to IP template amount and WGA-PCR cycle number

Using the same SCA approach, re-analysis of the effect on microarray reproducibility of changes made to IP WGA-PCR template concentrations and to PCR cycle number (discussed in section 4.3.7 and figure 4.18) can be performed to show the power of SCA for use in assay development experiments. A SCA correlation matrix showing the IP/IN analysis, IN single channel and IP single channel correlation is displayed, in the format previously used, in figure 5.10.

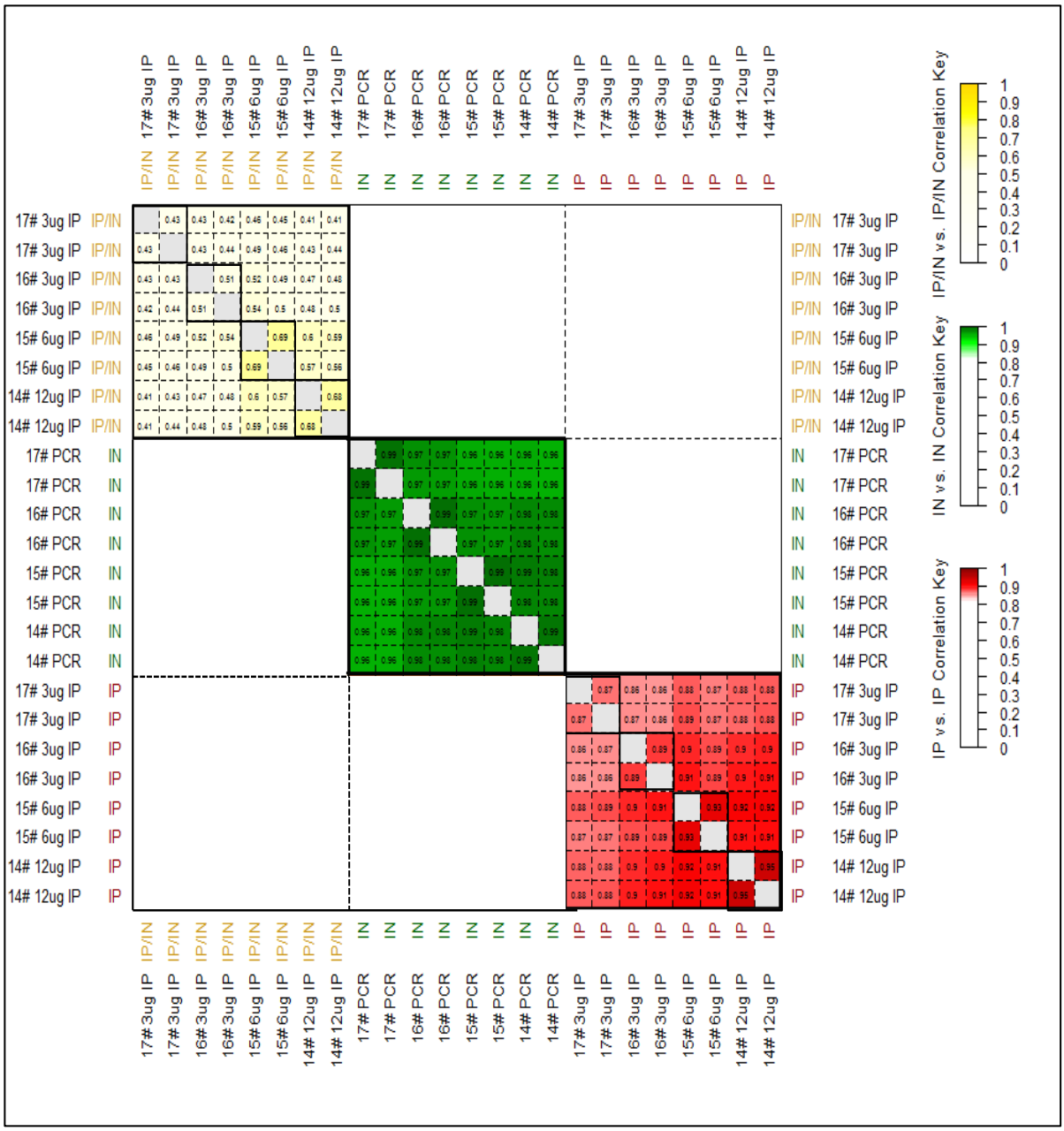


Figure 5.10: An SCA matrix demonstrating the effect of changing PCR conditions on IP, IN and IP/IN sample correlations.
An SCA matrix of correlation values between platinum-DNA adduct patterns (IP/IN ratios) (top left yellow box), single IN channels (central green box) and IP single channels (lower right red box). Samples are from 4 pairs of technical replicate oxaliplatin treated PBMC samples. Each pair of replicates was amplified with a different template of DNA pre-immunoprecipitation and with varied number of PCR cycles, indicated by the labels. The correlation values between replicate pairs are highlighted by the thicker outline between paired samples and coloured according to the key on the right of the plot.

This result confirms that the changes to IP DNA concentration and PCR cycles improves the correlation between technical replicates, as demonstrated by the change in paired IP channel correlations (lower right red box) from 0.87 with 3µg template/17 cycles WGA-PCR, sequentially through 0.89, 0.93 and 0.95

between replicates amplified with 12μg IP template and 14 PCR cycles. The IN correlation values are unaffected by these changes, as the central green IN box is homogenously green with all sample patterns matching with a correlation >0.95. This suggests that the improvements in IP pattern correlations results in the improvements in oxaliplatin-DNA adduct patterns in the upper left yellow box, with improvements in replicate correlations from 0.43 to 0.7 with the increase in IP template and decrease in WGA-PCR cycle number.

5.2.4 SCA re-analysis of microarray experiments determining effect of decreased IN template concentration

In section 4.3.6 and figure 4.17 the effect of diluting the IN samples to match the IP template concentration was experimentally determined. Previous experiments had demonstrated that the fragment pool produced by WGA-PCR was different at the concentrations of template DNA used in the IN and IP samples (figure 4.15). The intention of this experiment was to use the same template concentration in both IP and IN sample to remove any bias by correcting WGA-PCR induced fragment changes later in the analysis. Because the adduct pattern is generated from the ratio of IP to IN feature intensity, proportional to the fragment distribution, if both pools of fragments are equally biased by the WGA-PCR process the ratio calculation would potentially remove this effect from the final calculated adduct pattern. To examine the effect of these experimental changes on the IP and IN samples a SCA was performed (figure 5.11). Only the IP/IN, single IP and IN channels are shown. For the full matrix see the electronic appendix.

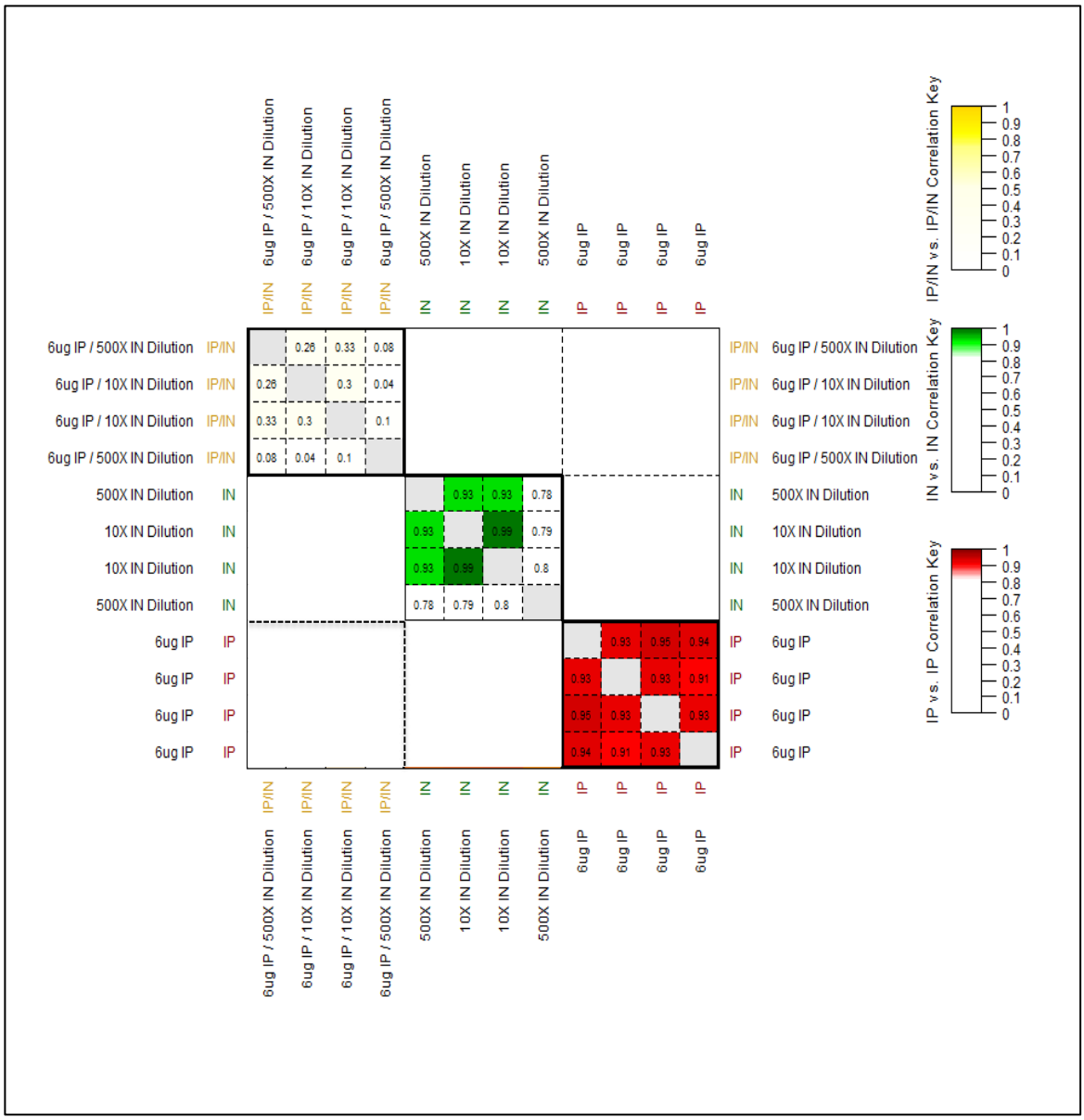


Figure 5.11: An SCA matrix demonstrating the effect of changing sample concentration pre-amplification on between IP, IN and IP/IN sample correlations.
An SCA matrix of correlation values between platinum-DNA adduct patterns (top left yellow box), single IN channels (central green box) and IP single channels (lower right red box). Samples are from 4 technical replicate oxaliplatin treated PBMC samples. In each sample the IP is identical – 4 aliquots of 6ug DIP-chip samples were pooled post-immunoprecipitation then divided into 4 equal samples for WGA-PCR. A single IN sample was used for the IN replicates, either diluted 10x for two of the samples or 500X for the other two samples. DNA was amplified and all pairs of IP and IN samples were labelled and hybridised and processed on the same microarray during the same experiment.

Despite dilution of the IN concentration to reflect the IP template concentration, resulting in a better match of pre-amplification IN and post-amplification fragment profile (the experiment shown in chapter 4, figure 4.16), dilution of the IN template to match the IP concentration results in more variable

IN samples than seen previously (shown in the central green IN channel large box of figure 5.11), demonstrating a fall in IN correlation to 0.93 with a 500-fold template dilution pre-WGA-PCR, compared to 0.99 with a standard 10X dilution, and resulting in more variable adduct patterns (upper left box). Also, using the SCA technique it is clear in this example sample 4 failed in the IN channel, and this is detected by a significant fall in the 4th IN sample correlation with the corresponding IN samples.

5.2.5 The identification of optimal target single IP channel correlation values

The dermal fibroblast platinum-DNA adduct pattern datasets generated during assay development (Powell 2014) can be re-analysed by SCA to give an indication of the target single IP channel reproducibility required to generate less variable oxaliplatin-DNA adduct patterns in PBMC samples. The datasets used are of two oxaliplatin adduct DIP-chip experiments with two paired biological replicates in each experiment.

The resulting oxaliplatin adduct pattern profiles from the fibroblast experiments are more reproducible than the PBMC datasets, with correlation values of 0.75 between biological replicates (Powell 2014), significantly higher than the best value of 0.6 with PBMC experiments. Using single channel analysis of the successful fibroblast experiment datasets may give an indication of to what extent an IP sample must be amplified to give greater adduct pattern reproducibility. The single channel analysis matrix is shown below in figure 5.12.

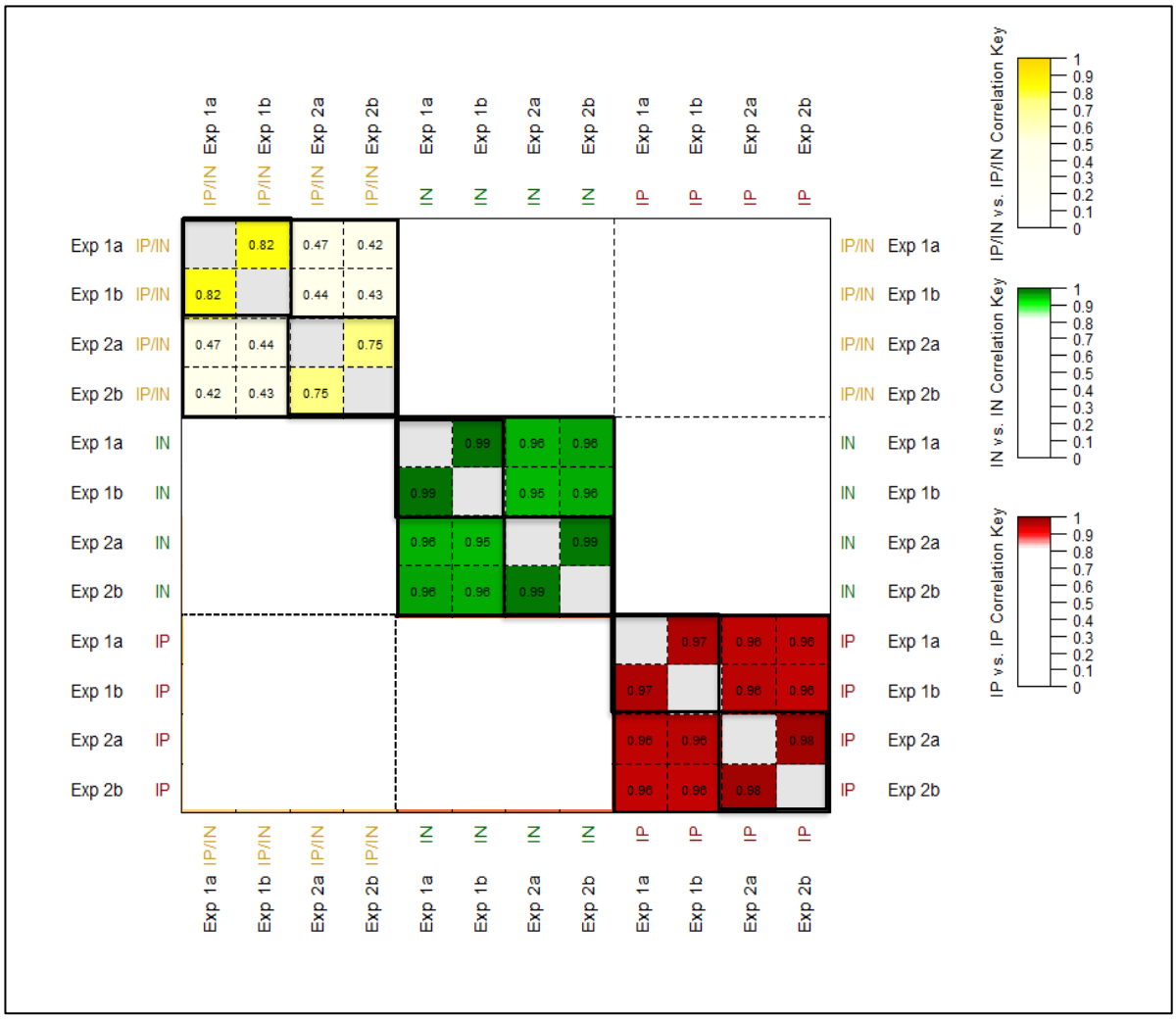


Figure 5.12: A SCA correlation matrix comparing 4 experimentally generated oxaliplatin-DNA DIP-chip datasets generated in dermal fibroblasts
A SCA correlation matrix comparing 4 experimentally generated oxaliplatin-DNA DIP-chip datasets generated in dermal fibroblasts (data from Powell 2014). The left upper box shows the correlation between IP/IN - adduct patterns - in the four samples. The central green box shows the correlation between IN single channels and the lower right red box shows the correlation between IP single channels. The strength of correlation is indicated by the box colour, as per the right hand key.

Once again, the amplification of the IN sample is highly reproducible, as shown by the consistently dark green colour of the central large box. Notably the IP box in the bottom right (as before, darker red indicates higher correlation values) shows values of 0.95 - 0.98. The combination of these two highly reproducible channels evidently results from the better correlation of the corresponding adduct patterns (upper left box – correlations between paired experimental repeats of 0.75 and 0.82). As a marker for the PBMC experiments, this level of single channel IP reproducibility is therefore required to match adduct patterns at this level of correlation.

5.2.6 Summary of SCA analysis

In summary, the single channel analysis method developed here appears to be a very useful tool when applied to the analysis of platinum DIP-chip microarray datasets. It gives more information than the previously available adduct pattern plot and adduct pattern correlation value. Using SCA to perform a detailed analysis of experimentally generated platinum DIP-chip microarray datasets confirms that the poor reproducibility of the platinum-DNA adduct patterns obtained from oxaliplatin treated PBMC biological repeats is a consequence of a decrease in reproducibility of the IP channel, due to an increase in amplification dependent noise generated at the low template levels seen in PBMCs treated with oxaliplatin. This is demonstrated by the greater variability of the IP channel compared to the IN channel. Analysis of previously successful datasets from dermal fibroblast cell culture models suggests that to achieve a reproducible damage profile in the PBMC samples IP correlations of >0.95 are required.

5.3 Examination of DNA GC content and the effect on DIP-chip microarray reproducibility

It is well established that one of the key variables governing the success, fidelity and relative yield of PCR amplification is the GC content of the fragment to be amplified (Dabney and Meyer 2012; Hasmats et al. 2014; van Dijk et al. 2014). In the 5Mb region of chromosome 17 used to generate the platinum DIP-chip adduct patterns in this thesis the average GC content of the fragments to be amplified varies from as low as 36% or as high as 46% when looked at in windows of 100,000 bases (figure 5.13).

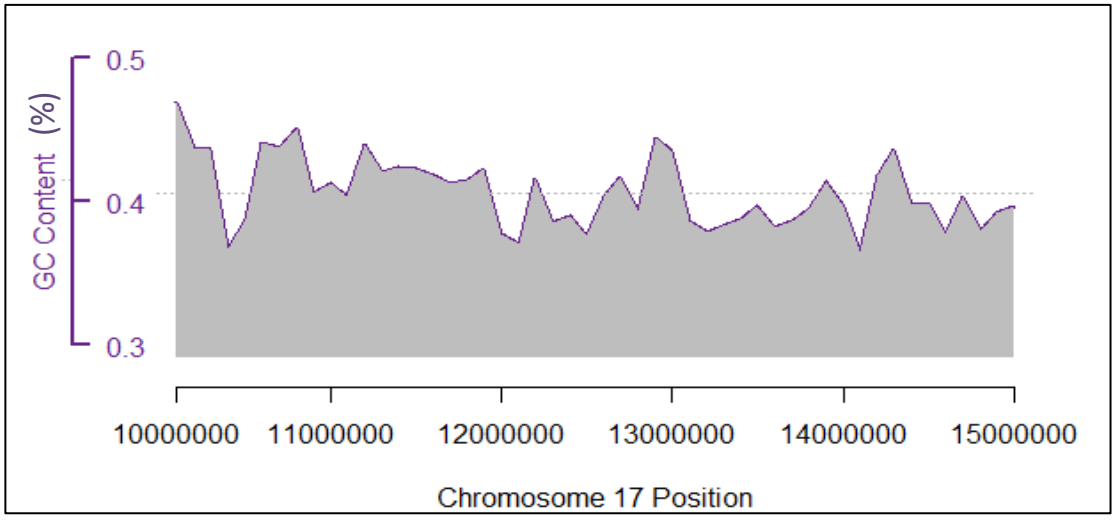


Figure 5.13: The average GC content of 100,000 base pair windows along the 5Mb of chromosome 17 used in this series of DIP-chip experiments.
Data calculated using R package BSgenome with genome reference sequence hg19. (http://www.bioconductor.org/packages/release/bioc/html/BSgenome.html)

The yield of PCR reactions with DNA polymerases commonly used for this type of experiment are all highly dependent on GC content (figure 5.14). The data for the GC-dependent PCR yield for the proprietary DNA polymerase used for the WGA2 Kit (Sigma) is unknown.

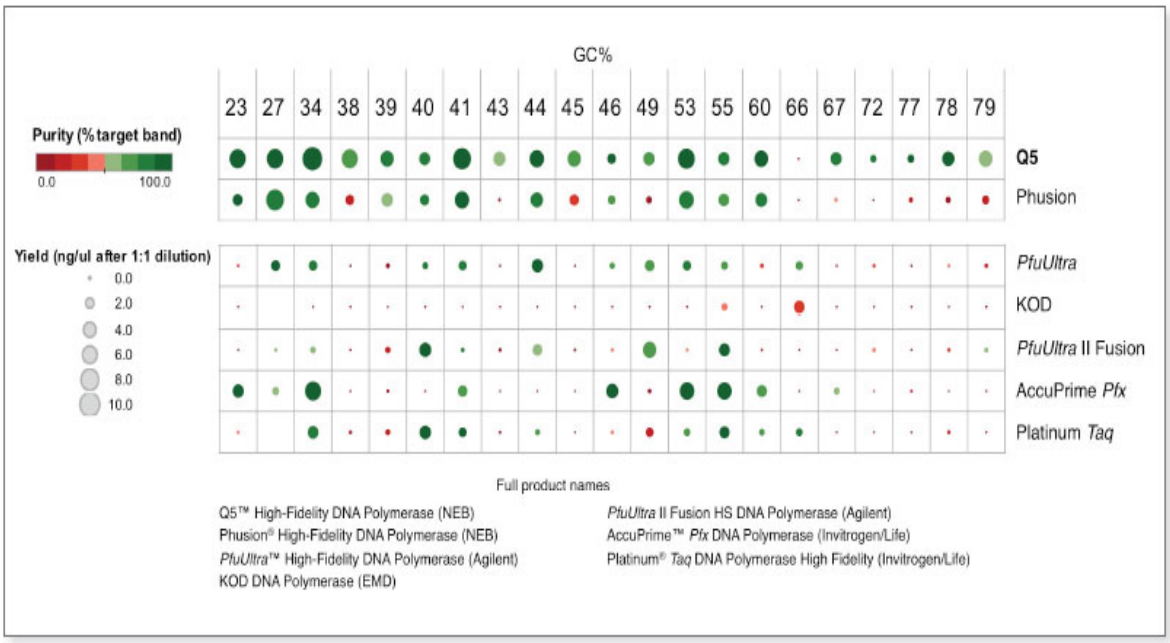


Figure 5.14: PCR DNA yield by GC content and enzyme.
The relative yield of PCR product is dependent on the DNA polymerase used and the GC content of the DNA template for many commercially available DNA polymerases. For each template GC value different polymerases result in different DNA yields (shown by the area of the accompanying circle). (http://www.neb.uk.com/product_overview/Q5.asp)

The DIP-chip assay is dependent on the faithful amplification of the relative proportion of DNA fragments which are generated by immunoprecipitation, and this information is corrected during analysis by the IN sample, also representing a faithful copy of the relative proportions of DNA fragments in the pre-immunoprecipitation sample. If the relative amount of DNA fragments generated through WGA-PCR of each fragment varies dependent on the fragment GC content this could have a significant bearing on the ultimate platinum-DNA adduct pattern produced, as the DIP-chip assay is dependent on the maintenance of the fragment relationship and the accurate detection of subtle (at most ~4 fold) differences in the relative proportion of 200bp-600bp DNA fragments along a 5MB stretch of chromosome 17 in two samples, processed over a ten-day laboratory protocol.

If the variation in PCR amplification, dependent on GC content as illustrated in figure 5.14, occurs with the polymerase used in this assay, it may contribute considerably to variability in the final adduct pattern generated. At the extreme, it could result in IP/IN patterns based on the characteristics of the

DNA polymerase used, rather than based on the subtle biological signal the assay is intended to detect.

Additionally, as platinum adducts form at the highest proportion at adjacent GG adducts (~65% of adducts) (Kozelka 2009), the proportion of damaged fragments in the post-immunoprecipitated sample could be over represented in high GC (and hence GG rich) fragments, resulting a combination of both differential template concentration and GC content–related polymerase bias. Although the IN sample may be expected to correct for some of this potential bias, the fragment profile will not be GC dependent in the IN sample, but will differ depending on GC content in the IP sample, because of the bias for immunoprecipitation of GC rich fragments. In this way the GC content has the potential to affect multiple aspects of the functioning of this assay when used to detect platinum-DNA adducts. It may be possible to measure the relative magnitude of these effects, if present, by comparing any GC/GG content effects between platinum and UV (i.e. a non-GG immunoprecipitation dependent) adduct samples.

5.3.1 The development of a custom R function for the 'Along Genome Correlation 'of paired DIP-chip microarray datasets

To enable to study the effects of GC and GG content on the reproducibility of the oxaliplatin treated PBMC sample data, and on oxaliplatin and UV treated fibroblast cell culture sample data, it was necessary to develop an analysis tool to examine the reproducibility of datasets along the 5Mb of chromosome 17, allowing the comparison of variability between datasets with changes in GG and GC content.

To perform this analysis a custom set of two R functions were developed – the along genome correlation 'AGC' functions **AGCcalc** and **AGCplot**. For the R script and source code see appendix 5 and 6 respectively). The **AGCcalc** R function generates a matrix of comparisons of paired DIP-chip array datasets at matched coordinates along the chromosome. This is achieved by specifying the start probe coordinate (*s*), end probe coordinate (*e*) and the number of sections

in which the data is to be analysed (*n*), and dividing the data into *(e-s)/n* windows. In each window the Spearman's correlation coefficient between paired datasets is compared by calculation of the correlation between the IP/IN values at each probe in both datasets in the specific window. Additionally, at each window the DNA sequence is obtained (based on genome reference sequence hg19) between the start and end coordinate for each window, and the percentage GC or GG content for each window is calculated and stored.

Using the second function, **AGCplot**, the between dataset window correlation values can then be plotted for each window along chromosome 17 from co-ordinate *s* to co-ordinate *e* at *n* data-points, referenced to the left hand y-axis. Specifying the argument *GC=TRUE* (the default) results in each window GC content being plotted in the lower section of the graph and referenced to the right hand y-axis. The function prompts for the title and legend details, and the final plot is customisable using several arguments to modify the look of the final plot (see appendix 6).

5.3.2 Analysis of oxaliplatin-treated PBMC DIP-chip datasets with the AGC function

An AGC plot is shown in figure 5.15 for the along genome comparison (AGC) of mean probe value datasets experimentally generated from 100μM oxaliplatin treated PBMC from patients A and B. At each of 50 windows the probe IP/IN values are compared using the Spearman's correlation, and are plotted for that specific window. The mean correlation between whole datasets is plotted by a single horizontal line.

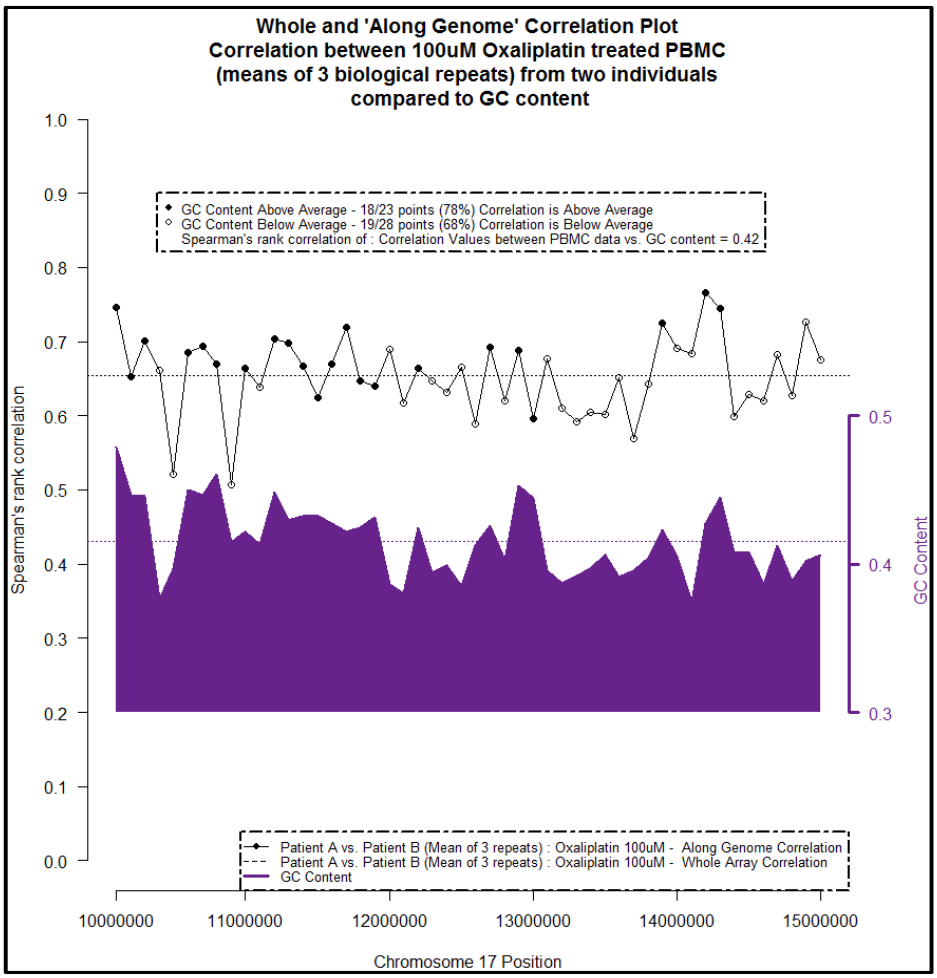


Figure 5.15: An example Along Genome Correlation (AGC) plot.
This plot is showing the correlation between two experimentally generated microarray datasets at 50 windows along 5Mb of chromosome 17. The panel shows the correlation between datasets (calculated from the mean probe values of three independent repeats of 100uM oxaliplatin treated PBMC samples) from individual A compared to individual B. This data is displayed by the connecting line and solid points for each window. The average correlation value between the whole microarray data from both datasets is displayed by the horizontal dotted line.

In figure 5.15 the correlation values between paired DIP-chip microarray datasets along the genome are plotted, and the mean correlation value for the whole microarray datasets are also demonstrated by the horizontal line. Clearly, the correlation values vary along the chromosome, with some windows showing the adduct patterns at this region are more closely matched than at other sections, as indicated by the higher window correlation value. The GC content for each window is plotted below (purple), and is referenced to the right hand scale. To aid visualisation, in the window correlation trace open and closed

points are used to display whether in each window if the GC content is above average (closed point) or below average (open point) for the whole dataset.

A second Spearman's correlation value and plot can then be generated to examine the relationship between the 'primary' window correlation value (the Spearman's correlation between the two microarray datasets at each window) and the GC content of the window GG (figure 5.16).

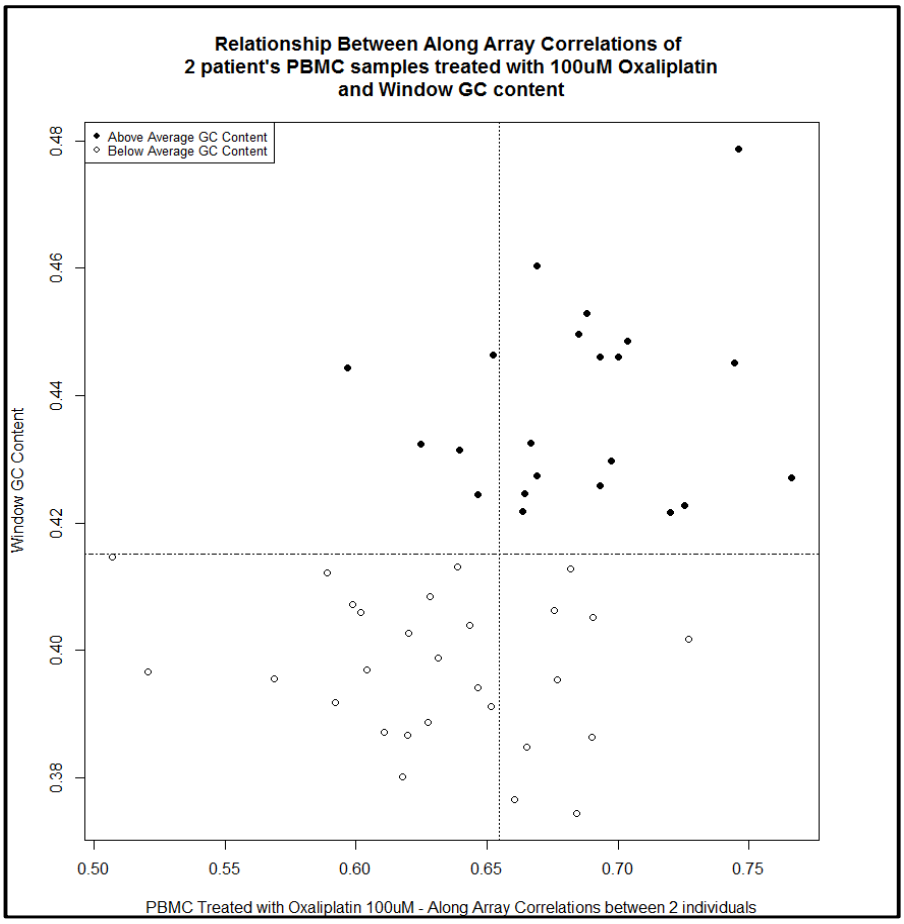


Figure 5.16: An example secondary AGC correlation plot.
This panel shows the secondary correlation between the window between dataset correlation values and the window GC content. The window microarray correlation is plotted against the window GC content. The mean value for each variable is plotted by the vertical and horizontal lines respectively. Open points are used in windows with below average GC content and closed points for windows with above average GC content.

Of the 50 windows analysed along the 5Mb of chromosome 17 used in this analysis there is a positive relationship between higher window GC content and improved correlation between microarray datasets. In the 22 windows with GC content above average the correlation between the two array datasets is above average in 18 (78%) compared to the 28 windows with a GC content

below average, in which 19/28 correlation values between repeats are below average (68%) (figure 5.15). The secondary correlation of the GC content and array window reproducibility is 0.42, indicating that GC content is an important component in the reproducibility between the datasets.

As an additional way of displaying this data, in the secondary plot (figure 5.16) all of the closed points show the windows with above average GC content. 18/23 (78%) of these points are in the upper right quadrant, indicating that these are the windows with above average reproducibility, compared to only 5/23 closed points (22%) are in the upper left quadrant. This indicates that there is a tendency towards more reproducible microarray data in higher GC content fragments.

This relationship could be due to less reproducible WGA-PCR in lower GC content DNA fragments. Alternatively, it could occur because higher GC content regions of the genome contain higher amounts of platinum adducts (as previously discussed 65% of platinum adducts form at adjacent guanine nucleotides (Kozelka 2009). DNA fragments with platinum adducts are preferentially immunoprecipitated, as the epitope of the CP9/19 antibody used is the platinum-GG DNA adduct (Meczes et al. 2005), resulting in higher proportion of high GG DNA fragments in the pre-amplification sample, and therefore greater template amounts available for amplification by WGA-PCR. It has been established in the previous section that higher DNA template concentration for WGA-PCR results in a less variable DNA fragment distribution post-amplification, and better correlation between microarray datasets.

Mathematically, the guanine content of the human genome is 21%, so the percentage of adjacent GG dinucleotides should be ~4.4% (i.e. the probability of a pair of dinucleotides being adjacent GGs is 0.21*0.21). The window GC content should be an accurate reflection of the window GG content unless there is bias away from GG dinucleotides in the windows used over the 5Mb section of chromosome 17 in this analysis. For example, the sequences GCGCGCGCGCGC and GGCCGGCCGGCC have the same GC 100% content, but the GG content is 0% in the first and 50% in the second. To confirm that the GC and GG content of this section of the human genome are directly correlated, window GC and GG

contents were calculated and plotted at 50 windows over 5Mb of chromosome 17 (figure 5.17).

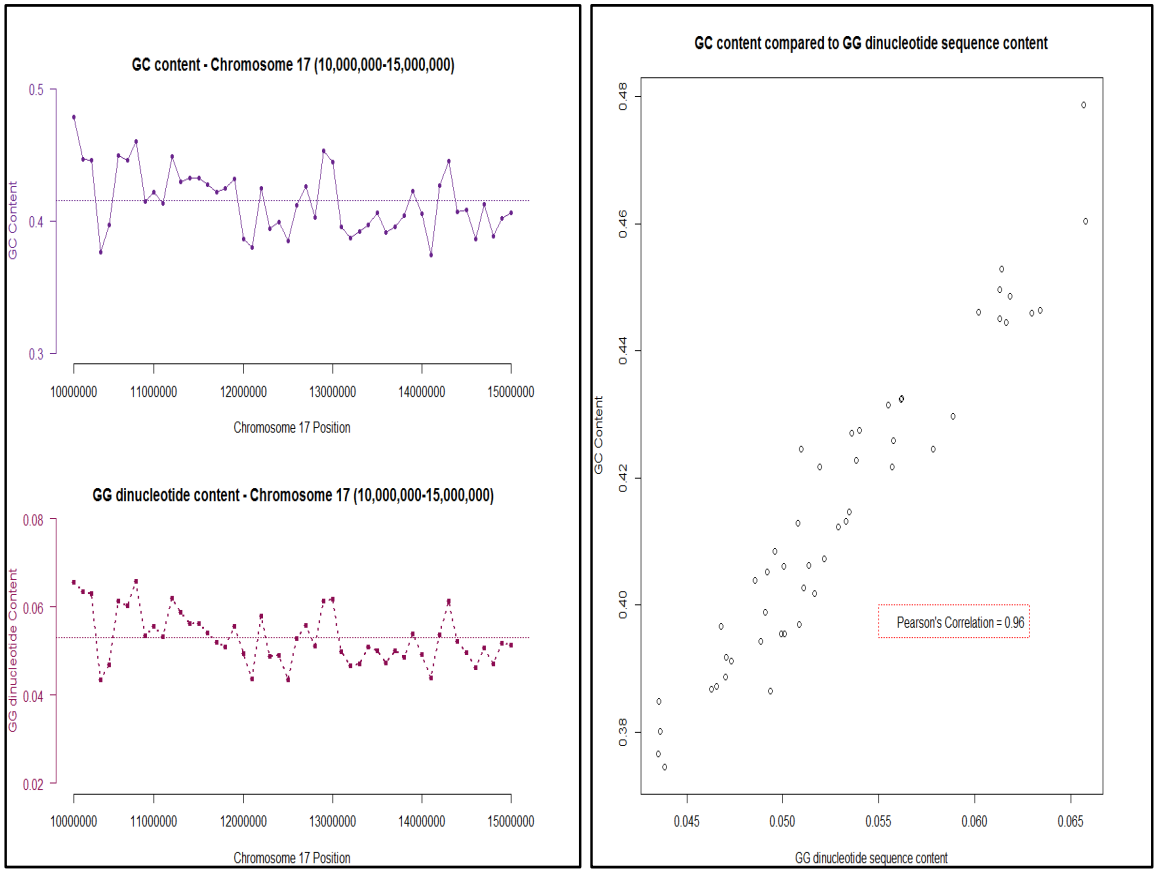


Figure 5.17: A comparison of window GC and GG content.
The left-hand plot shows the window GC content (above) and GG content (below) along 5 Mb of chromosome 17 (nucleotides 10,000,000 to 15,000,000), divided into 50 windows. The pattern of higher and lower than average GC content matches the window GG content, in a ratio of ~10:1. The correlation between each pattern is shown in the right hand panel, with a correlation between window GG and GC content of 0.96.

The result indicates that window GC and GG content are proportional in this region of the genome, suggesting that higher GC content fragments will be over represented in platinum treated DIP samples because of the preference for platinum-DNA adduct formation at GG dinucleotides. As demonstrated extensively in this previous section, higher DNA template concentrations for WGA-PCR result in more reproducible microarray data, particularly in the IP single channel – suggesting this could be the mechanism between better microarray correlation at high GC content regions of the genome.

5.3.3 Comparison of AGC plots from platinum-DNA adduct DIP-chip samples and UV-DNA adduct DIP-chip samples

Figures 5.15 and 5.16 demonstrate that the PBMC DIP-chip microarray datasets are more reproducible at regions of the genome with an above average GC content. This mechanism behind this observation could be a relationship solely between GC content and PCR reproducibility, or could potentially be through an effect of changes in WGA-PCR template concentration affecting microarray reproducibility. As the epitope for the anti-platinum antibody used in the DIP-chip is GG di-adducts, higher GC content fragments are likely to be immunoprecipitated compared to DNA fragments with a low GC content.

The relative contribution of these two effects on microarray reproducibility can be assessed by comparing the 'along genome correlation' window values between paired platinum treated and paired UV treated datasets. The UV treated microarray datasets in this analysis were generated during the DIP-chip assay development study (Powell 2014). Dermal fibroblasts were treated with UV as a source of DNA adducts, forming cyclobutane-pyrimidine (CPD) dimers in DNA, as immunoprecipitated as the DNA-adduct target for the DIP-chip assay. Approximately 68% of CPDs occur at adjacent thymidine nucleotides and the antibody used in this form of the DIP-chip assay was for this CPD subtype (the same antibody was used in chapter 6, section 6.6.1 for immuno-slotblot detection of UV damage). The chip stage of the assay (amplification, labelling, hybridisation, scanning and data analysis) are identical to that used with this platinum-DNA adduct DIP-chip, including the same WGA-PCR amplification and PCR conditions. As any fragment bias from immunoprecipitation should be for thymidine rich fragments in UV induced DNA adduct, as opposed to guanine rich fragments with platinum-DNA adduct immunoprecipitation, if the improvement in window array reproducibility demonstrated at regions of higher GC content is due to a PCR-template effect, because of an increase in adducts for immunoprecipitation at GC rich fragments, this should occur in an inverse relationship between paired platinum-DNA adduct samples and paired UV treated DNA samples. If the effect is due to GC

effect on PCR reproducibility it should be similar in both UV and platinum microarray paired datasets.

To examine for this effect, the 100μM oxaliplatin PBMC datasets generated from the mean probe values from three independent repeats was compared between individuals A and B. Two UV damage microarray datasets were also compared by AGC in the same way. Both AGC plots were aligned to show the similarities and differences between window reproducibility, and compared to window GC content. The result is shown in figure 5.18 A and B.

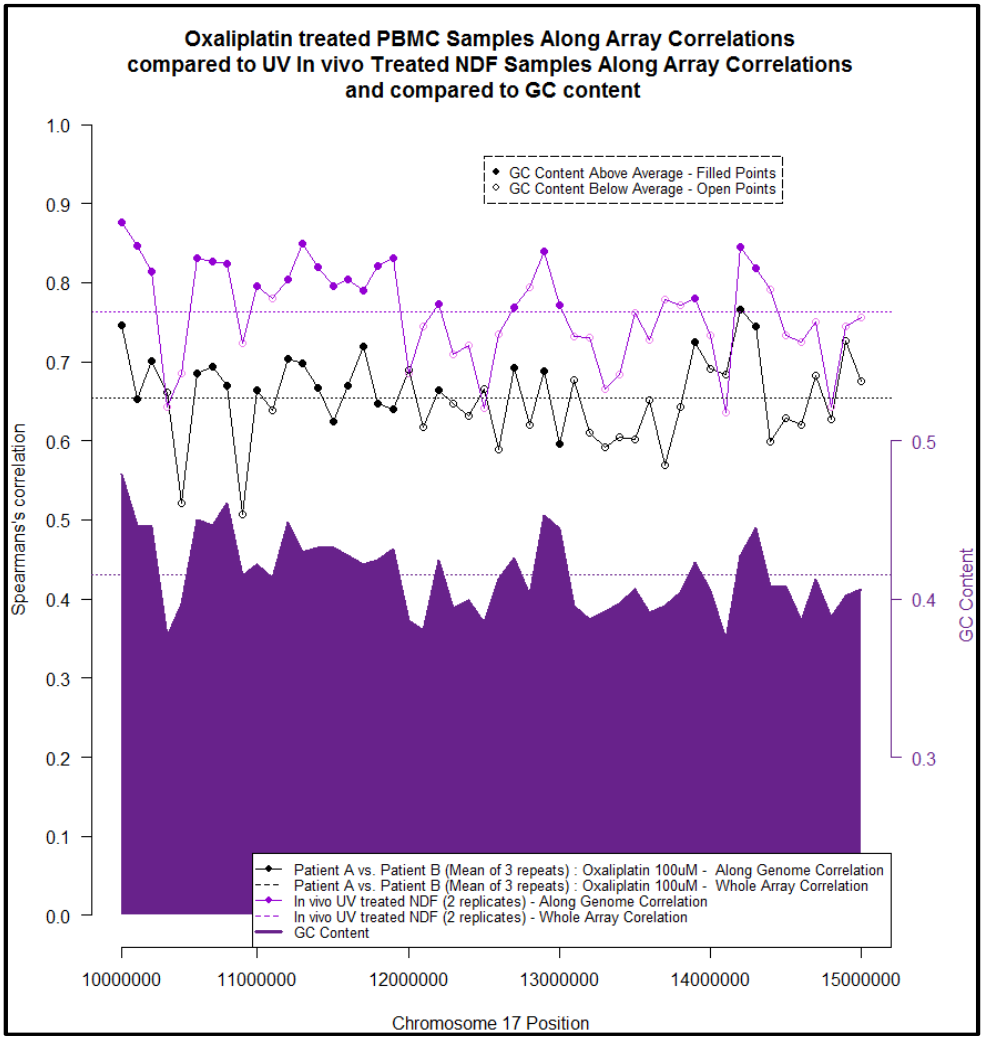


Figure 5.18 (A): A 50 window AGC plot of dermal fibroblast cell lines treated with UV irradiation compared to PBMC samples treated with oxaliplatin.
A 50 window AGC plot of dermal fibroblast cell lines treated with UV irradiation and immunoprecipitated by the DIP-chip assay with an anti-TT CPD antibody (lilac points) is compared to an along genome correlation plot of oxaliplatin-DNA adduct microarray datasets (black points) generated from the mean values from three independent repeat experiments in two individuals. The GC content of each of 50 windows is presented in the lower section of the plot and referenced to the right hand y-axis. The UV and oxaliplatin microarray window correlation values are plotted with solid points in higher than average GC content windows and open points in lower than average GC content windows.

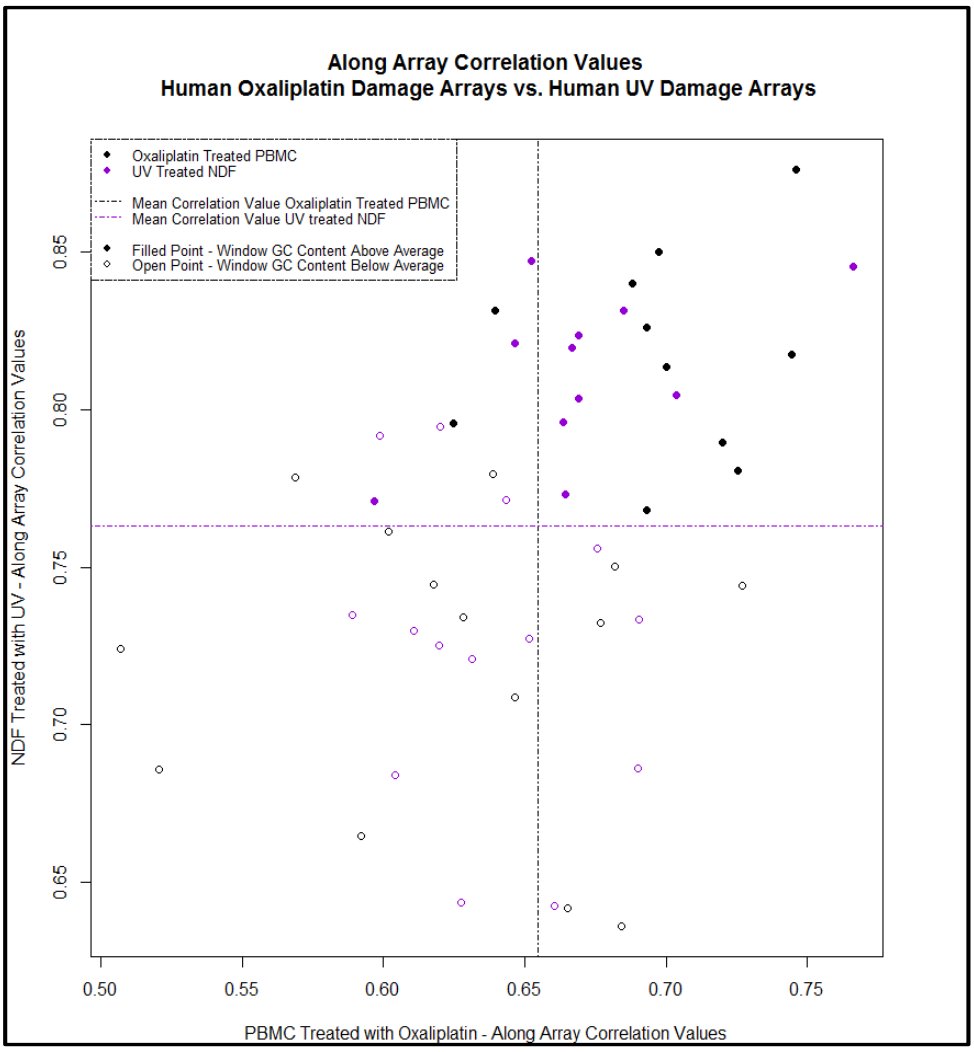


Figure 5.18 (B): The window correlations between the paired UV (purple) and oxaliplatin (black) DIP-chip microarray datasets are plotted against one another.
The average correlation of all probe values of between each pair of datasets is shown by the horizontal purple line (UV datasets) and black vertical line (oxaliplatin datasets). The window GC content is shown by the closed points if window GC content is above average and open points if the window GC content is below average

Both the UV adduct DIP-chip datasets (lilac) and oxaliplatin adduct DIP-chip datasets (black) show a better correlation between paired datasets in windows with higher than average GC content (Figure 5.18A and 5.18B). In 5.18B the distribution of window correlation vales for each dataset comparison is plotted and compared. All of the windows with an above average window correlation in both dataset comparisons have an above average GC content, as shown by the exclusive presence of closed points in the upper right corner of this plot, indicating that high GC content is associated with less variability in the adduct patterns produced from the DIP-chip assay.

This result indicates the GC effect on DIP-chip sample reproducibility is a consequence of more reproducible amplification in GC rich regions. If these differences were due to a PCR-template effect on reproducibility as a result of preferential immunoprecipitation of GC rich fragments in the platinum DIP-chip and TT rich i.e. low GC fragments in the UV DIP-chip inverse patterns should be seen between AGC plots of paired UV DIP-chip and platinum DIP-chip samples. As the pattern of GC related amplification fidelity occurs similarly in both UV DIP-chip and platinum DIP-chip samples this is not the predominant cause of this result.

5.3.4 Summary of AGC analysis

The potential explanation for these results presented above, following along genome correlation (AGC) analysis of paired microarray datasets, is that increased 'noise' introduced by the WGA-PCR amplification process occurs at regions with lower GC content, contributing to a reduction in array reproducibility when compared to higher GC content sections along the chromosome. This is demonstrated in both the UV DIP-chip and platinum DIP-chip samples. As this effect occurs in both UV and platinum damaged datasets it is a function of the WGA-PCR process, rather than a template dependent effect resulting from differences in DNA concentration as a consequence of the profile of adducts immunoprecipitated.

This evidence leads to potential methods to improve the reproducibility of PBMC DIP-chip microarray datasets. One method is through modification and optimisation of WGA-PCR conditions to improve the fidelity of amplification of low GC content regions, resulting in an overall increase in reproducibility. Further experiments to examine the effects of PCR modifications is outside of the scope of the work presented in this thesis, but could be conducted during future projects involving DIP-chip assay development.

Alternatively, if AGC analysis is used to identify area of the genome with characteristics that result in poor reproducibility of the DIP-chip assay as a

consequence of WGA-PCR, then these regions could legitimately be excluded from downstream analysis, specifically when determining differences between adduct-DNA patterns from clinical samples. This could be justified by the analysis presented above, as higher rates of adduct-DNA pattern variation due to technical factors may reduce the reliability of any pattern differences detected in these high variability genomic regions during downstream analysis.

5.4 Determining the effect of DNA polymerases on adduct patterns produced by the DIP-chip assay

As outlined above the reproducibility of the DIP-chip assay generated platinum-DNA adduct patterns is critically dependent on the WGA-PCR amplification step of the assay. Both the amount of DNA template available in the sample for amplification and the GC content of the regions amplified are a significant factor in introducing noise into the assay. An additional factor to be considered is the role of the particular DNA polymerase used for the PCR, in light of the evidence that many DNA polymerases produce a DNA yield that varies depending on the characteristics of the template, especially the GC content (Figure 5.14).

The DNA polymerase available with the WGA2 kit (Sigma) is a proprietary polymerase. Several variations to the WGA2 kit are available. One version, the WGA1 kit is identical except no polymerase comes with the product, allowing the use of alternative DNA polymerases whilst maintaining the universal priming system and maintaining the use of proprietary amplification buffers.

A WGA4 variant is also available. This differs in the fragmentation protocol (not used as our DNA is pre-fragmented) and is suggested for amplification of genomic DNA from a single cell for use on picogram quantities of DNA. The main difference in the WGA2 kit and WGA4 kit protocol is the use of 25 cycle of PCR with WGA4 rather than 14 to 17 cycles with WGA 1 and 2. As the template DNA in the IP sample is low, the WGA4 kit potentially could be used to improve the IP sample amplification fidelity, however, the downside of the WGA4 kit is a documented allelic dropout rate of 30% - because of the low DNA

template ~30% of the DNA does not amplify correctly, making it less useful for the DIP-chip assay.

To consider the effect of changes to the DNA polymerase on adduct patterns generated by DIP-chip, three identical IP samples were used and amplified with the standard WGA2 protocol. One post-DIP IN sample was used to generate 4 WGA-PCR IN samples. In one sample the IN was used unamplified. A further 3 aliquots of 1μL of the IN sample were diluted 10 fold to produce three identical IN samples for amplification. One of these samples was amplified with the standard WGA2 kit, and one with the WGA1 kit with the addition of Q5 high fidelity DNA polymerase (NEB). Additionally, a further IN sample was processed with the WGA2 kit with a paired IP sample amplified with the WGA4 kit.

This experiment allows a comparison on single IN channel patterns produced in the unamplified sample compared to the WGA2 polymerase and Q5 polymerase. Following amplification each IN sample was paired with an IP sample and processed to generate the platinum-DNA adduct pattern dataset, which was subsequently analysed by single channel analysis (Figure 5.19).

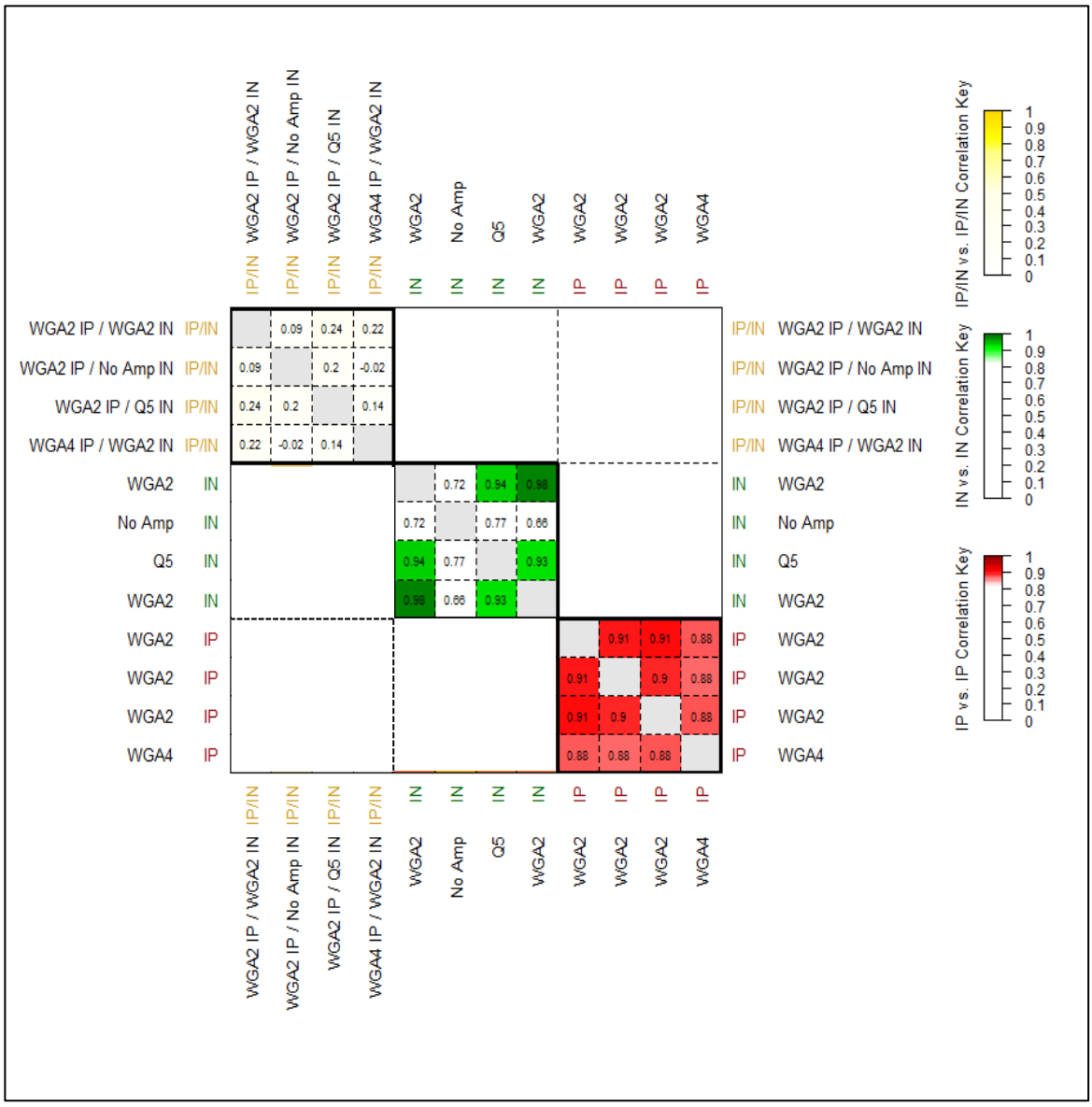


Figure 5.19: A single channel analysis plot of DNA produced by a DIP experiment using a variety of DNA polymerases during amplification.
IP samples were pooled and divided to generate identical DNA pools prior to amplification of the IP samples. A single post-DIP IN sample was used to generate one undiluted sample and 3 identical IN sample replicates. All samples were then amplified using a variety of WGA-PCR kits and DNA polymerases (see labels for details). The correlation between each pattern generated is shown in each small box. These are grouped into larger, thick outlined boxes to show the relationship between groups of IP/IN adduct patterns (top left), between groups of IN single channels (middle, green) and between groups of IP single channel patterns (bottom left – red).

The relationships between the patterns generated and displayed in the central, green box are key to comparing the effect of different polymerases. This box shows the relationship between IN patterns in the unamplified sample and in the samples derived from this DNA and amplified with the WGA2 polymerase or Q5 polymerase. Firstly, the pattern in the unamplified sample is different

from the patterns in the sample when amplified by WGA polymerase or Q5 (correlation 0.66-0.77). This is evidence that the pattern generated by the amplification is different to the underlying pattern of fragments in the pre-amplification sample. The amplified samples are more closely matched (0.93-0.94) between WGA2 amplified IN DNA and Q5 IN DNA. A fall in single IN channel correlations to 0.93 is enough to result in high variability in the IP/IN generated adduct pattern (as shown by differences in the previous figure 5.8-5.10), and by the fact that correlation of single channel IN patterns of 0.99 is needed to generate reproducible adduct patterns in the dermal fibroblast cell culture system (e.g. figure 5.12). The WGA2 patterns generated by the two pools of DNA amplified by WGA2 polymerase correlate 0.98, as previously demonstrated, but the patterns of probe values in these samples are both different to the unamplified sample.

The three identical IP sample patterns from DNA amplified with the WGA2 kit all correlate with each other reasonably well (~0.90) and slightly better with each other than with the WGA4 amplified DNA (~0.88), suggesting that this variant of WGA-PCR may potentially have a use in amplification of low amounts of DNA, but that this needs to be optimised and assessed with further experiments.

The variability generated by the differences in IN single channels and IP single channels results in low correlations between the adduct patterns derived from these paired single channel patterns (IP/IN box, top right, all correlations <0.3).

This result is evidence that the WGA-PCR amplification process is producing amplified DNA that is different to the pre-amplified sample (compare the unamplified to amplified IN single channels – correlations of ~0.8) and that the enzyme used in the amplification results in a different pattern of amplified DNA (compare single channel IN correlations between Q5 and WGA Polymerase – correlations of ~0.93). This is further evidence of the role of WGA-PCR DNA amplification in the generation variable and of PCR-dependent adduct patterns by the DIP-chip assay. In future, it may be that using unamplified IN samples recues the potential for PCR induced bias in the assay and improves the overall

microarray dataset reproducibility. Experiments to determine if this approach offers benefits above the current amplification of the IN sample are beyond the scope of this thesis but may be investigated in future assay development projects.

5.5 Summary

The analytical tools developed in the section above in this chapter are methods of extracting latent information from microarray datasets. The single channel analysis (SCA) and along genome correlation (AGC) approaches can be used as quality control steps, to confirm the DIP-chip experiment has been successfully conducted. Additionally, these tools have demonstrated in the work presented so far in this chapter that they are valuable analytical tools for DIP-chip assay development work. The application of these techniques to the microarray datasets generated in the previous chapter have clearly demonstrated issue with the introduction of variability into the DIP-chip assay through the WGA-PCR step, and have highlighted potential areas for further experimental projects to improve the assay further, including by optimising the PCR to maximise the IP single channel reproducibility between experimental repeats, and to optimise the PCR reaction to improve the fidelity of amplification at low GC content regions.

The AGC analysis also highlights that microarray datasets derived at specific regions of the genome have higher variability than other regions, and, in the absence of experimental measures to reduce this variability, to improve the validity of the data it may be prudent to exclude these areas from downstream analysis.

Finally, the use of unamplified and amplified IN samples demonstrates that the WGA-PCR process does, to a degree, alter the pattern of fragments produced during amplification, and this does have an impact on the adduct-DNA pattern produced. This is one rationale for the use of the IN sample to normalise these effects in the IP sample during the generation of the adduct pattern. The benefit and implication of using unamplified IN samples to generate the final

IP/IN adduct pattern has yet to be explored and is outside of the scope of the experiments presented in this thesis.

All of the experiments and tools developed to this point focus on reducing noise in the DIP-chip assay, reducing variation between adduct-DNA patterns when the same sample is run repeatedly, with the intent to increase the validity of inter-individual differences in adduct patterns detected in downstream analysis. In the final section of this chapter we will now turn to the development of tools to allow the detected of inter-individual difference in DIP-chip platinum-DNA adduct patterns – the ultimate goal of this project.

5.6 The development of an R function to detect, display and compare regions of difference between DIP-chip generated adduct patterns derived from different individuals.

The DIP-chip microarray datasets used in these analyses each contain approximately 30,000 to 40,000 data points. So far the focus of this thesis has been on improving the signal–to-noise ratio in the DIP-chip assay by decreasing the background noise. This has been achieved by modifying the protocol to reduce the differences between DIP-chip generated adduct datasets in samples of identical DNA, as demonstrated with the new bioinformatic tools developed in this chapter. In the final section of this chapter the effort will be on detecting the signal - differences *between* adduct patterns *between* individuals - by the development of a new analysis tool to be used with the DIP-chip assay.

The goal of the DIP-chip assay is to detect biological differences in adduct patterns between individuals, as a potential marker of how an individual's phenotypic response to the drug may differ as a consequence of this signal. Therefore, the development of a function to identify biological differences between adduct patterns generated between individuals will be necessary for the DIP-chip assay to work in this context.

When two experimental conditions are compared over several independent repeat experiments the mean of several repeats is similar in adduct patterns. However, the IP/IN value at each data point varies considerably (Chapter 4 section 4.3.1-4.3.2). With 40,000 data points, variation in probe value can be influenced and introduced by the experimental technique, by true biological signal, and by differences between probe values occurring due to chance variation. The true biological signal needs to be identified from differences in the datasets due to background experimental and chance variation.

Two issues require consideration in the development of a tool for identifying and concisely displaying regions of putative biological significance. Firstly, a statistically robust method of identifying biological differences between datasets and generating the information in a set of genomic location and magnitude values identifying where this variation occurs is required. Secondly,

this list of locations of interest and magnitude of difference between datasets needs to be utilised and plotted, alongside supporting information with respect to the statistical validity of the approach.

5.6.1 Statistical considerations

There are several potential methods of identifying statistically significant differences between datasets. During the development of the analysis function described below, an outlier detection method is employed to identify regions of the genome with the largest differences between probe values paired between datasets. This may be a valid approach if the key differences between drug response between individuals are determined by the probes that vary the most widely between two individuals. Additionally, much of the noise in the system (the smaller magnitude variation between datasets) is filtered out by concentrating on the data points with the largest variation between individuals.

Whilst useful in the development of this R function, an outlier approach to identifying differences between datasets has limitations, including that ranking each paired probe independently of genomic position results in the loss of information from the genomic context, local relationships and patterns of the surrounding probes. This approach is also has the limitation inherent in identifying a set of probes with the largest difference between datasets - potentially it also identifies probes at which high variation between probes is due to experimental failure at this location (and resulting in larger differences between datasets), rather than through the presence of true biological difference.

A potential alternative approach is to use t-tests on probes paired between groups of datasets to identify a list of probes with value differences of significance determined by a pre-set statistically cut-off, rather than identifying the most extreme differences between pairs of probes. This approach is in development in our laboratory, however further refinement of the technique is currently limited by the necessity to have several relevant datasets available to optimise the method. This shortcoming will be addressed through a pilot study

currently in progress to obtain the microarray data required to optimise this approach (as discussed in chapter 7, section 7.2).

A modification of the t-test approach utilises a 'sliding window' method to perform a t-test to identify statistically significant differences between probe values in two groups of datasets at different sections along the genome, combining the statistical rigour of a t-test and preserving the information available from the influence of genomic position and local probe context. This approach to identifying statistically significant differences between microarray datasets at different windows along the genome has been developed in our laboratory and validated for ChIP-chip microarray datasets (Bennett 2013) and has recently been peer-reviewed and accepted for publication. This method can be adapted for the analysis of DIP-chip datasets and is under development. Again further preliminary data is needed to refine this approach, but once optimised this has the potential to be a statistically rigorous way of identifying patterns of differences between DIP-chip microarray datasets.

In the R data analysis function developed in this section the main emphasis is on visually displaying and streamlining the process for analysing the output of probes identified as having biological significance, and can be modified to use any of these statistical methods to generate a list of probes of interest. To summarise, the optimal approach to generate the list of probes of difference between datasets is a matter of current debate, requires further microarray datasets to refine, and is subject to on-going work along the lines discussed above by the laboratory bioinformatic and statistical team.

5.6.2 Outlier Detection

In this section, one of several potential methods of generating a list of probes of possible significance is demonstrated, and is used as the basis for the development of the R functions required – essential for streamlining the analysis and visualisation process. For this example, probes of interest are identified by examining the data for probes with the greatest magnitude of difference between datasets.

As an example of this approach, two 100μM oxaliplatin treated PBMC microarray datasets generated and used throughout this series of experiments between patient A and B will be compared to detect probes with the largest value difference between datasets, either positive or negative. If the IP/IN ratio value used to generate the adduct pattern at each probe is subtracted in one dataset from the other the spread of the 'subtracted values' approximates a normal distribution (Figure 5.37). The tails of the plot of subtracted values indicates a potential population of 'outlier' probes, locations at which the probe values have a larger difference (either positive or negative) between datasets. A key requirement of the DIP-chip analysis is to identify which of the probes with the greatest difference between datasets vary because of the tail of the normally distributed experimental noise, and which probes in this pool vary because of biological differences between datasets.

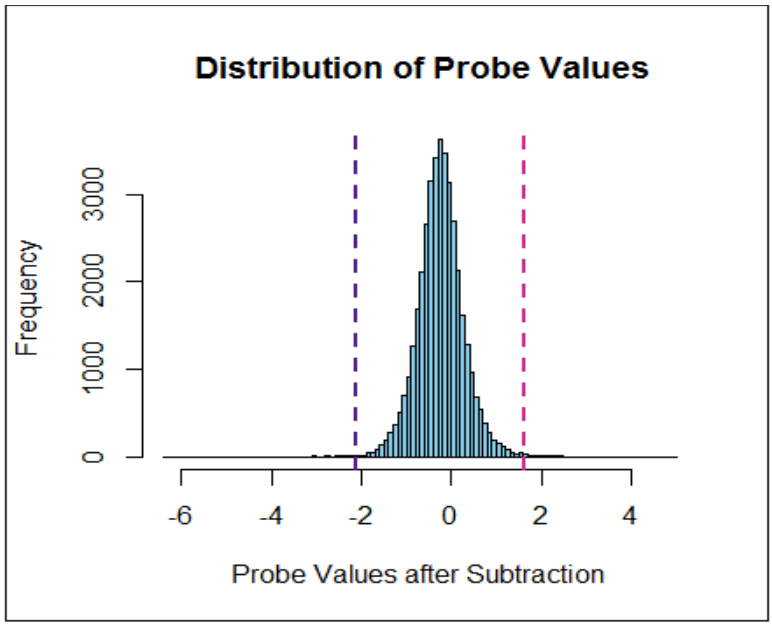


Figure 5.20: Plotting the distribution of subtracted probe values.
The IP/IN probe values from one 100μM oxaliplatin treated PBMC dataset generated from patient B is subtracted from a similar dataset from patient A. The difference between probe values (A-B) approximates a normal distribution. The tail of the distribution indicates where the probe values (and hence adduct patterns) have a greater difference between individuals. Arbitrary purple and pink dotted horizontal lines are added to indicate potential outlier probes.

The identification of the particular outlier probes in this population - the probes with the most different values between datasets – is then an issue of statistical interpretation of the significance of different IP/IN values at probes

within this population. Outliers in large datasets can, somewhat subjectively, be defined as "an observation (or subset of observations) that appears to be inconsistent with the remainder of that set of data" (Barnett and Lewis 1994)

As discussed above, a method is required to identify a population of probes with values in this population that differ between datasets due to normal variation in the assay, and a population of probes that have significantly different values between datasets due to a potential biological signal. The techniques used in our laboratory include t-tests and R statistical functions and packages used to identify outlier populations. A t-test approach was initially performed to analyse outliers between the datasets generated from individual A and B, but was not able to detect statistically significant outliers between these two datasets. For this thesis, and for the critical development of custom R functions to automate data analysis and processes for the visualisation of the result of DIP-chip comparative analysis, an outlier detection methodology used previously in our laboratory (Bennett 2013; Powell 2014) – the R extremevalues package - will be employed.

For a detailed description of the statistical methods behind this package see an associated reference (Loo 2010). As explained by the author in the summary "the extreme values package estimates the underlying data distribution by regression of the observed values on their estimated QQ plot position using a model cumulative distribution function… Having obtained a description of the bulk distribution… we determine the values above which less than a certain number of observations are expected, given the total number of observations and the fitted distribution."

In essence, normally distributed data forms a straight line on a QQ plot when the observed data is compared to a theoretical predicted (and in this case normally distributed) dataset. The greater the variation of subtracted value data points to this line, the more likely the difference in value between probe between two datasets is due to biological signal rather than normally distributed experimental error. The extremevalues package assesses the data under these assumptions and generates a QQplot, identifying likely outlier probes as probes as those that vary most from this distribution, shown in red at the left and right extremes of a QQ plot (Figure 5.21). The software package generates a list of

these probes that are likely to be outliers identified by this method. By manually searching the outlier probe list generated these can be linked back to the original Sandcastle adduct pattern plots. The limitations inherent in this approach have been discussed in section 5.6.1.

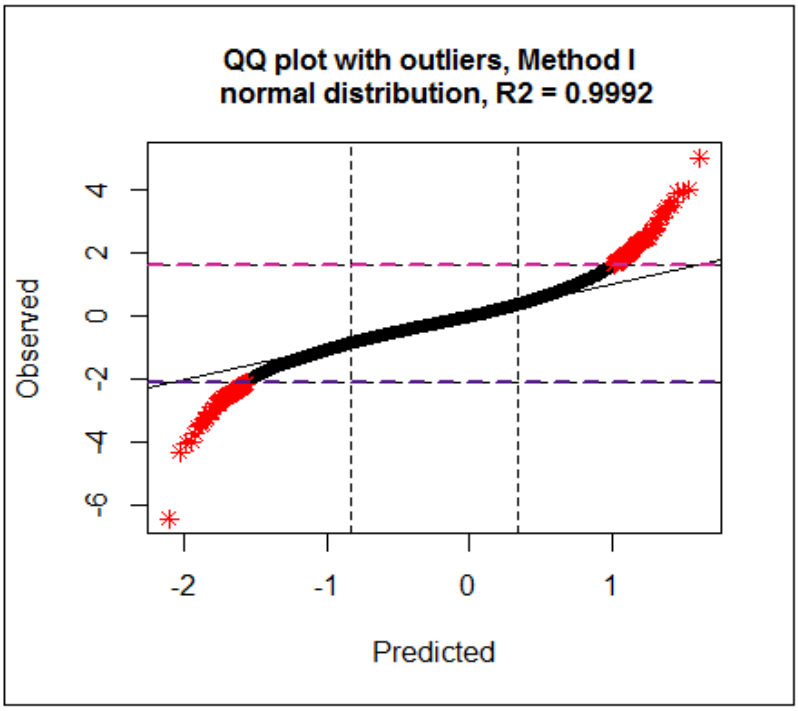


Figure 5.21: A QQ plot output form the extremevalues R package, used to compare two sets of experimentally generated oxaliplatin treated PBMC microarray datasets.
The variation in probe values between the two datasets approximates a normal distribution, resulting in close approximation of the data to a straight line on the QQ plot, when the observed dataset is compared to a predicted dataset if the data is normally distributed. At the left and right extremes of the plot probes that do not approximate the normal distribution above a threshold determined by the underlying spread of the data (shown by the purple and pink lines) are identified in red.

5.6.3 The development of the outlierDataPlot R function

For analysis of DIP-chip data a function is required to identify significant outlier probes between experimentally generated datasets and to plot these outlier regions for comparisons between experimental conditions or between individuals. This is functionality is unavailable in Sandcastle. The function **outlierDataPlot** was developed to perform this analysis (appendix 7), and at its core uses the extremevalues R package to generate lists of probes that are likely to be different between datasets because of biological signal rather than

underlying noise in the assay, before displaying these in an variety of plots to aid interpretation and providing a list of outlier probes for further analysis.

Other methods of identifying statistically significant outlier probes between platinum adduct DIP-chip datasets are available and the development of these is underway in our laboratory, including using t-tests (Dr M Bennett – unpublished work). The extremevalues package is used in this instance for the reasons outlined above, but the **outlierDataPlot** R function can be modified easily to use alternative methods of generating outlier probe lists.

One initiated, the **outlierDataPlot** function requires the identification of the numerical location of the datasets to be compared from the vector data. The first two arguments are reserved for these identifiers.

An example plot generated from the analysis of 100μM oxaliplatin treated PBMC samples from individual A compared to individual B is shown below (figure 5.22). Once run, the function generates the outlier probe list and plots three graphics. Firstly, in the top left the QQ plot and outlier probe distribution is plotted. Secondly, in the top right the distribution of the subtracted dataset is displayed to give an overview of the spread and distribution of the data over the ~44,000 data points.

The third plot, the lower plot, shows the position of the outlier probes along the section of the genome analysed, and with a colour scale for the magnitude of the difference between the two datasets compared indicated by the colour key. The outlier can be negative or positive (i.e. the difference between the two adduct patterns is either significantly higher in one than the other, or significantly lower). The central data, between the purple and pink dotted lines – not shown, would contain the remaining non-outlier data points, with normally distributed differences between datasets of low magnitude and likely due to experimental variation rather than biological differences. A list of outlier probes, subtracted probe values and the genomic location can be obtained from the **outlierDataPlot** function using the argument *csv=TRUE*.

This tool can now be used to rapidly assess for variations between clinical samples from patients in the DIP-chip profiles generated and, additionally,

removes much of the normally distributed variation between samples presumed to be due to experimentally generated noise. In this example the outlier probes are well spread over the section of the genome studied (figure 5.22), however on visual inspection there appears to be a group of outlier probes at approximately 10,500,000, possibly indicating an area of greater inter-individual variability between adduct patterns in this region that could be examined in more detail and at greater resolution.

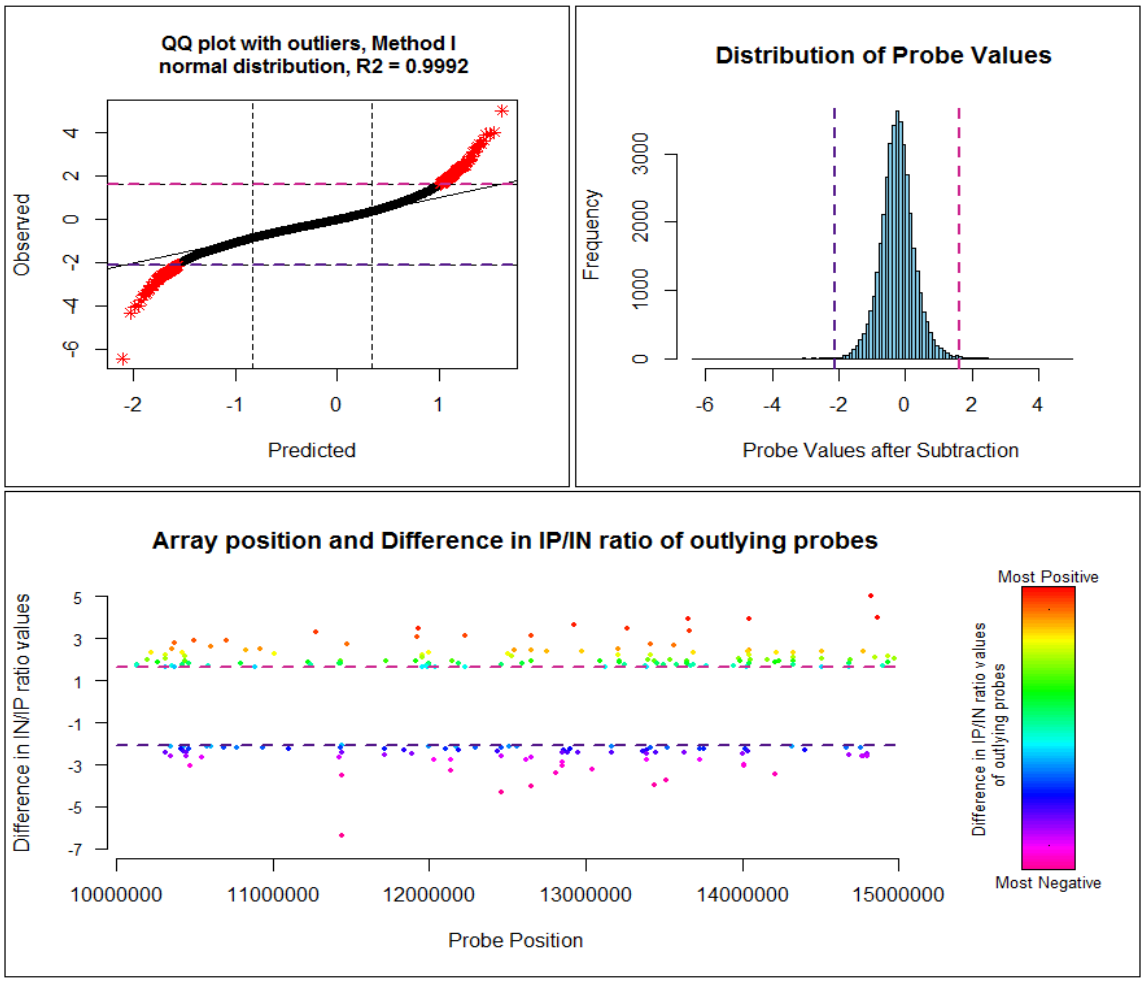


Figure 5.22: A plot generated by the outlierDataPlot R function.
Two experimentally generated oxaliplatin treated PBMC DIP-chip microarray datasets are compared to identify significant outlying probes in this example. The top left plot shows the QQ plot between samples, with the majority of data points approximating the normal distribution and a tail of outlying points at the left and right extremes identified as outliers by the increase distance between these points and the normal line. The right upper plot shows the spread of probe IP/IN value when one dataset is subtracted from the other and the two tails where significant outliers are likely to be found (the pink and purple cut-off lines) correspond to the QQ plot and have been generated by the extremevalues package. The lower plot shows the genomic location of identified outlier probes and the magnitude of the difference between the two datasets of the value of the outlier probe (corresponding to the right hand key).

5.6.4 Development of the outlierDataMultiPlot R function

To compare between multiple datasets the **outlierDataMultiPlot** function was developed (for the full script see appendix 8). This allows the simultaneous generation of lists and plots of outlier probes, calculated again by the extremevalues package and modifiable for alternative statistical approaches if desired. The number of datasets to compare is specified in the first argument and subsequent to this the programme will prompt the numerical identifier of the datasets to compare from the vector 'data'.

The genomic locations and magnitude of subtracted values at each outlier probe for each dataset can be overlaid on the same plot using the argument *plotsame=TRUE*, or in separate plots using the argument *plotsame=FALSE*. Tiles for each plot will be requested by the function if the argument *title=TRUE* is used, and again a csv file of all of the outlying probes and genomic locations can be generated and labelled by using the arguments *csv=TRUE* and specifying a file name with the argument *CSVtitle*.

An example of the potential use of the **outlierDataMultiPlot** function is for the comparisons of 100μM oxaliplatin treated PBMC DIP-chip datasets between patient A and B. As previously discussed, (chapter 4 section 4.3.2) the mean datasets between three repeats matching well (correlation 0.69) but for any one for any two of the repeats the adduct patterns are more variable (correlation 0.4-0.6).

The question to be answered in this example analysis is whether the outliers between repeats experiments are at the same probe location when comparing the independent repeat patterns from one individual against the mean of the other individual as a reference pattern. In this way, the mean of the repeats from individual A is used as a reference to compared the variability in detected outlier position in each individual dataset for patient B. The **outlerDataMultiPlot** produces three genomic plots showing the location of outliers for each of these three comparisons (figure 5.23)

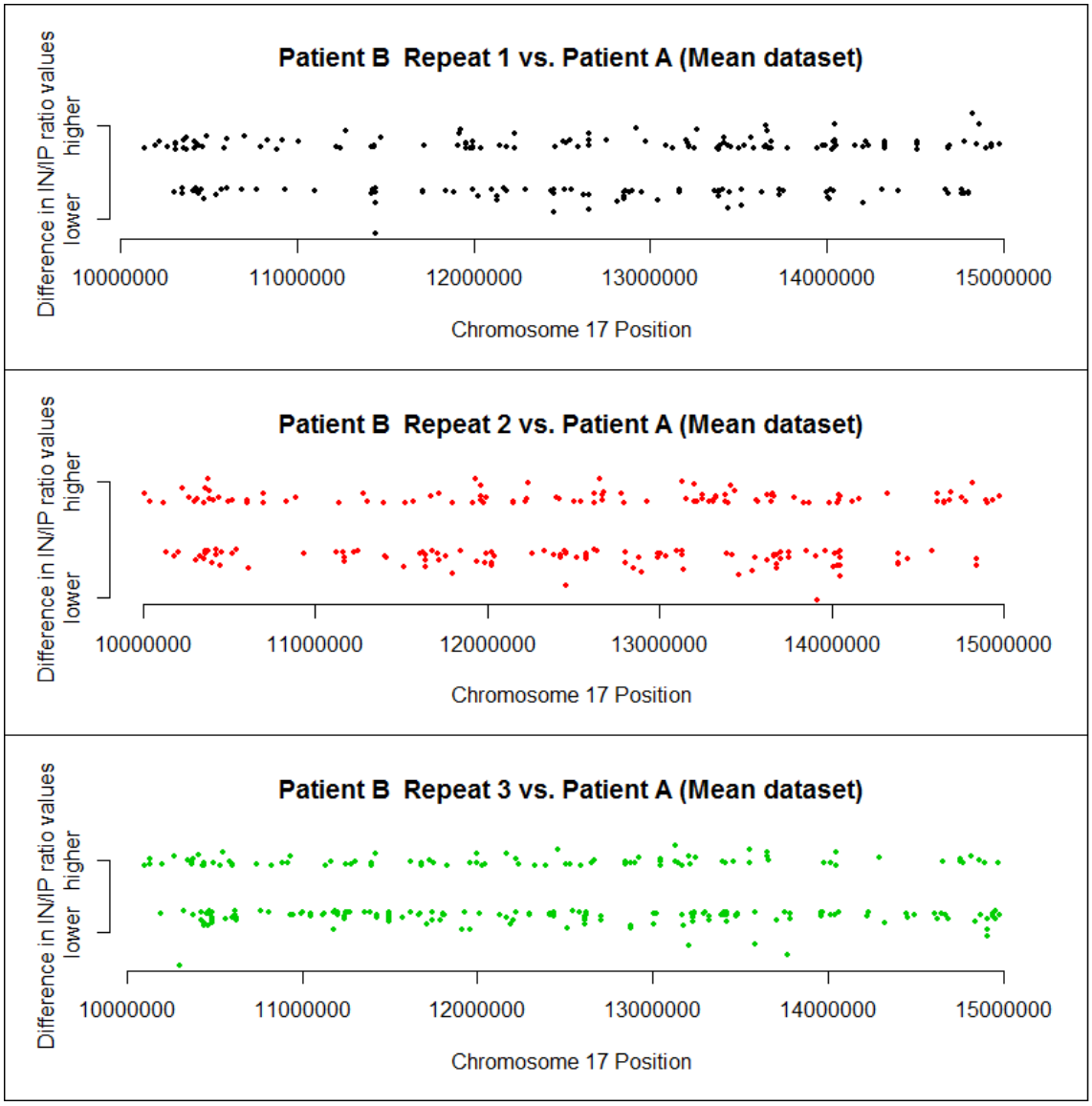


Figure 5.23: An outlierDataMultiPlot output.
An outlierDataMultiPlot output showing the genomic location of outlier probes in the comparison of Dip-chip microarray datasets between a reference sample (in this case the mean of three repeats of oxaliplatin treated PBMC from patient A) and three biological repeat experimental generated data from individual B.

In this example, outliers appear to be relatively evenly distributed along the genome. This function can be used to compare multiple outlier plots simultaneously, allowing easier visual identification in patterns within and between microarray datasets. In samples with higher signal and lower noise this should be a valuable analysis tool.

5.6.5 Summary

The functions presented in this section (5.5) are capable of detecting, plotting and listing individual probes, from DIP-chip microarray datasets, that differ *between* individuals, with the level of significance determined by either t-tests or with outlier detection package. As discussed, the optimal statistical method used to identify probes with differences between datasets due to true biological signal is yet to be determined, and is a current area of on-going research by our laboratory bioinformatics team. Potential limitations of the outlier detection approach used above include that by the process of identifying probes with the greatest magnitude of difference between datasets, regions of potential experimental failure may be additionally be detected (as failure in one experiment may result in large differences between these failed and a non-failed paired dataset). Additionally, this technique may result in a loss of information from the genomic and local probe context by identifying probes only by the greatest difference between datasets. As discussed in section 5.6.2 a 'sliding window' based t-test approach may offer a robust statistical approach and retain the genomic and local probe context during analysis. At the time of writing these bioinformatic developments are incomplete, and a detailed discussion of the current debate as to the relative merits of different statistical approaches to identify differences between the datasets are beyond the scope and limitations of this thesis.

However, despite these limitations, this capability is now sufficiently developed with the tools described here, using the currently available t-tests or extreme outlier approaches (and more refined approaches as they are developed in future), to allow the assay to be applied to clinical samples from cohorts of patients with a different response or toxicity from platinum based chemotherapy, in order to identify patterns and probes that differ between individuals and cohorts. Using the assay on cohorts of patients with valid phenotypic differences will, in addition, allow the further refinement and validation of several candidate statistical approaches to more precisely identify differences between microarray datasets. In this way, the PBMC DIP-chip assay and analytical tools developed here can now be used as a clinical assay.

5.7 Discussion and Conclusions

The aim of this project is to continue the translation of a cell culture based laboratory assay, with limited data analysis tools, into a single-use clinical assay, utilising the DIP-chip technique to generate platinum-DNA adduct profiles that can be used to predict for clinical outcome from platinum treatment. As a clinical assay, the DIP-chip assay would be run prior to first exposure to chemotherapy, through a single blood test that could rapidly and accurately be analysed to generate an idiosyncratic adduct pattern. The adduct pattern, or patterns of outlier probes with validated predictive significance, could then be analysed to estimate the chances of clinical outcome of interest, and be used to guide the treatment choices.

One limitation of the DIP-chip assay for this clinical application is the use of PBMC as a surrogate tissue, reflecting aspects of cell specific and germline DNA damage and repair capacity rather than that the specific DNA damage and repair profile of tumour tissue. However, as discussed in section 3.1.7, repeated biopsies to obtain tumour tissue are impractical and hazardous, and despite the limitations PBMC are commonly used as an easily accessible surrogate tissue in similar DNA damage and repair based clinical assays of tumour response and toxicity, as extensively discussed in the introductory sections to chapters 3, 4 and 5. To determine the impact of this effect it may be possible in future to compare platinum-induced adduct patterns in paired PBMC and tumour tissue.

The experiments conducted in human PBMC cells in this, and the two proceeding chapters, and the results of the analysis conducted in these chapters, are the first time this technique has been used in clinical samples. The work conducted and described highlights the significant experimental and bioinformatic challenges involved in this process and represents several important steps in adapting the DIP-chip technology towards our goal.

The methods and experiments outlined in chapter 3, demonstrating the adaptation of the DIP-chip technique to PBMC samples, as analysed at a single genetic locus, has identified that with protocol modifications and amendments it is possible to generate reproducible immunoprecipitated samples. The experiments subsequently conducted in chapter 4 focused on the development

and optimisation of the high-resolution genomic scale application of the DIP-chip assay. Several experiments were conducted to enhance the reproducibility of the assay following identification of WGA-PCR as a step that introduces variability into the final platinum-DNA adduct pattern generated by the technique. High levels of variability potentially limit the clinical utility of the assay, which is required to generate reproducible patterns of induced DNA damage when repeated on the same individual - an essential output to allow reliable comparisons between individuals.

Several protocol amendments described have resulted in improvements in adduct pattern reproducibility, for example by increasing the IP sample DNA concentration or decreasing the PCR cycles (Chapter 4, section 4.3.7). However, to assess the effects of small changes in DIP-chip assay protocol is a slow process, as each assay takes almost two weeks to perform, limiting the rate of experimental effort to measure the effect of subtle protocol amendments in multiple repeat experiments. Additionally, given the limited information available from the available analysis tools applicable to DIP-chip assay, it was not possible to make rapid or further significant progress on this front.

The bioinformatic techniques for use with the DIP-chip assay developed and demonstrated in this chapter are a significant step forward for the data interpretation, assay development and quality control of DIP-chip experiments.

The development of the single channel analysis functionality (SCA - section 5.2) allows the in-depth interpretation of DIP-chip microarray datasets to identify the effects of subtle protocol amendments during optimisation work, for example by demonstrating the effect on the IP sample of PCR template changes (figure 5.10). SCA is also a useful quality control metric to confirm successful completion of the DIP-chip assay, by the demonstration appropriate IN and IP sample correlations (e.g. identifying an IP correlation of >0.95, section 5.2.5) and to identify the presence and stages at which any technical failures have occurred (for example see figure 5. 6).

The along genome correlation functionality (AGC, section 5.3) gives the ability to interpret DIP-chip datasets in different windows along the data, and has proved useful in identifying low DNA GC content as a source of WGA-PCR introduced noise into the assay. Previous studies have demonstrated that

additives to PCR samples and modification of reaction conditions can reduce GC content related differences in PCR efficiency (Aird et al. 2011). The AGC function will be of benefit for this type of DIP-chip development work. For example, with this new tool it is now be possible to test and demonstrate the effect of the addition of PCR additives or modification of PCR conditions to improve the amplification of low GC regions.

The experiments and developments discussed so far have focussed on the reduction of noise to improve the signal-to-noise ratio in the DIP-chip assay. The final group of tools developed in this chapter are tools to identify biological signals amongst the large microarray datasets generated by this technique. Within the limitations of the statistical approaches currently available to identify differences between datasets, as extensively discussed, the **outlierDataPlot** functions developed in this chapter streamline what was previously a multistep process requiring the use of several R functions to generate a list of potential outlier probes. The **outlierDataPlot** functions provide several graphical outputs for quality control (the QQplot and histogram of subtracted data) and several metrics for outlier probe analysis, including overlaid and separate plots showing the genomic position of outlier probes and magnitude of difference between datasets, and with the **outlierDataMulitPlot** function provides the ability to simultaneously visually compare multiple outlier probes generated from several paired DIP-chip datasets and provides lists of data for further analyses.

There are significant limitations to the statistical approaches applied to identify difference between datasets, including the loss of genomic and local probe context if whole array or outlier approaches to identify differences between datasets are employed. By focussing on maximal differences between microarray datasets the outlier detection methodology used also identifies potential regions of experimental failure, rather than just true biological signal. We are in the process of modifying an accepted method of ChIP-chip analysis to develop a t-test based approach used on sections of the datasets to retain the local probe context, an approach that should lead to a more statistically robust method of identifying regions of difference between two microarray DIP-chip datasets. The refinement of these approaches requires the availability of DIP-chip datasets with underlying biological differences, which will be obtained

through a small-scale clinical studies, the details of which are discussed further in chapter 7.

In summary, the experiments and developments described in this, and the proceeding two chapters, are an important component of the on-going efforts to develop the DIP-chip assay into a clinically useful tool. The work presented here is critical experimental evidence to support the protocol modifications that reduce the noise in the DIP assay (chapter 3) and the chip aspect of the assay (chapter 4), improving the signal-to-noise ratio and potential clinical applicability of the assay. The bioinformatic tools developed and documented in this chapter have demonstrated their worth in the re-analysis of DIP-chip microarray datasets, and provide vital additional analytical capabilities, essential for the translation of this DIP-chip assay for patient clinical samples.

To date, the assay has been used on healthy volunteers, potentially exaggerating the effect of noise on the signal-to-noise ratio of the assay, as the biological signal from healthy volunteers is likely to be low. Alongside on-going efforts to improve the reproducibility of the assay through protocol modifications, particularly of the WGA-PCR, one of the critical phases of the development of this assay is to test the system in situations where any biological signal is likely to be strongly present, either through the development of model systems, as described in the next chapter, or through clinical studies in cohorts of patients with, or without, the outcome of interest. Only in this way will it be possible to refine the statistical approaches to identify the underlying biological signal, and to determine the influence of signal strength on the signal-to-noise ratio and on the clinical performance of the DIP-chip assay.

Chapter 6 The development of *S. cerevisiae* models to examine the functional impact of single nucleotide polymorphisms associated with oxaliplatin induced peripheral neuropathy in colorectal cancer patients.

6.1 Introduction

Oxaliplatin chemotherapy, a treatment used to significantly improve cure rates in patients with colorectal cancer, often results in oxaliplatin induced peripheral neuropathy (OIPN). In some patients this can be a permanent and disabling side effect. At present there is no way to stratify patients to predict those who will benefit from treatment or to reduce the risk of this potentially devastating complication.

There may be a role for the DIP-chip assay technology in stratifying patients for the risk of developing OIPN. This condition has been linked to defects in DNA repair capacity, particularly in the nucleotide excision repair (NER) pathway, through genome wide association mechanistic and functional studies (see (Cavaletti et al. 2011; Dzagnidze et al. 2007; Ta et al. 2006) respectively). The conclusions from many of these studies are compatible with the underlying hypothesis of this thesis - that as DNA repair pathways are responsible for the removal of adducts formed by the interaction of platinum agents with DNA, through understanding variations in the functioning of these pathways in individual patients it may be possible to stratify patients for response and toxicity from these agents, and for the risk of OIPN in particular.

Novel single nucleotide polymorphisms (SNPs) have been recently identified at Cardiff University from a cohort of colorectal cancer patients who developed OIPN during treatment (West 2013). These polymorphisms are in the gene *XPF*, involved in both the NER and interstrand crosslink (ICL) repair pathways, and may have a mechanistic effect on DNA repair and consequently on the risk of developing OIPN

The main focus of this chapter is on the development of genetic models that will be used to determine whether these SNPs have an impact on DNA repair, through assessment of UV damage as a paradigm for NER, and on the result of oxaliplatin sensitivity assays. This study will use cloning techniques to genetically manipulate *S. cerevisiae* to model analogous SNPs in the homologous yeast NER gene *RAD1*, and will attempt to demonstrate the effect of these polymorphisms on oxaliplatin toxicity in this classical genetic and functional model system. If a link between subtle perturbations in DNA-repair capacity and OIPN can be demonstrated this will provide further evidence to support the use of the DIP-chip assay in future clinical studies in this context. A robust laboratory model of OIPN would also be an invaluable tool for the development of our DNA repair based assays for patient stratification in this condition.

6.1.1 Current biomarkers of response and toxicity to platinum agents

The platinum family of chemotherapy drugs are an essential component of the majority of chemotherapy regimes, but come with significant complications and adverse effects, and can occasionally be fatal in a patient otherwise cured of their initial malignant disease (Reed 2006). In light of this, biomarker studies with the aim of finding a mechanism to predict response and toxicity to these drugs have been a research priority, especially in tumour types in which platinum agents form the main line of effective treatment, notably in lung cancer (Bowden 2014).

There are a multitude of potential biomarkers that are candidates to have a role in stratification for response and toxicity to platinum drugs. Of those proposed, many are at early stages of development - typically in preclinical stages of identification and validation. A few have progressed through to large-scale clinical trials to attempt to demonstrate evidence of clinically meaningful predictive or prognostic power. In general, however, very few have progressed to this stage, and in the main the clinical results, usually of retrospective studies and analyses, have been inconsistent (Besse et al. 2013; Bowden 2014).

As discussed in chapter 1, section 1.4.1.6, there is significant evidence that platinum-DNA adducts are the lethal damage caused by exposure to platinum agents (Martin et al. 2008) and, as such, the repair pathways are leading candidates in the search for clinically meaningful biomarkers. An example of this, and a leading candidate that has progressed through to validation through large scale in clinical trials, is tumour ERCC1 status (Bowden 2014; Cobo et al. 2007).

The ERCC1 protein has an important role in NER, acting in a heterodimeric endonuclease complex with XPF to excise 5' of the pre-incision complex (see chapter 1 section 1.4 and (Marteijn et al. 2014)). High levels of ERCC1 protein expression have been demonstrated to be a positive prognostic factor in untreated surgically resected NSCLC (Simon et al. 2007) and ERCC1 expression, obtained by measuring protein or mRNA levels, has been implicated in resistance to platinum agents in lung, ovarian and gastric malignancies (Li et al. 2010; Tiseo et al. 2013). The most common method of assessment of ERCC1 status is through the measurement of relative protein levels measured by immunohistochemistry, for example used in the clinical trial by Cobo *et al.* as a marker for the response to platinum agents in non-small cell lung cancer (Cobo et al. 2007). ERCC1 status is currently being assessed as a predictive biomarker for cisplatin response in several lung cancer treatment clinical trials (Bowden 2014).

Although several studies have revealed a possible relationship between changes in DNA repair capacity and the response of patients to platinum chemotherapy, especially in the treatment of lung cancer, it is so far unclear whether the best biological markers are at the level of genetic polymorphisms, mRNA levels or immunohistochemistry for protein expression (Peters et al. 2014).

One recent publication regarding this issue describes a lack of specificity of the antibodies commonly used in immunohistochemical analysis of ERCC1 levels for specific isoforms of the protein (Friboulet et al. 2013). In the case of ERCC1, five isoforms exist, and the protein level of only one of the possible five isoforms correlates with cisplatin response. Because of strong homology among the five isoforms, the commonly available and frequently used antibodies for

immunohistochemical detection of this protein are non-isoform specific and, as a consequence, both functional and non-functional isoforms are detected by immunohistochemistry. In some circumstances a tumour may selectively overexpress non-functional isoforms of ERCC1 and be characterised by immunohistochemistry as ERCC1 positive, but have no increase in functional NER capacity, reducing the association between immunohistochemical determined ERCC1 level and the outcome from treatment. Additionally this study discovered evidence that the epitope for the most commonly used anti-ERCC1 antibody (8F1) may have changed over the period these antibodies have been regularly used for clinical studies, with discordance in detectable ERCC1 expression between the same samples re-tested over a 5-year period (Friboulet et al. 2013). This could introduce unexpected variability in staining between pre-clinical and related clinical validation studies.

As an alternative approach, several groups have attempted to identify predicative biomarkers at the genetic level. Many studies, usually genome wide association studies (GWAS), have identified single nucleotide polymorphisms that are statistically associated both with response and toxicity to platinum agents (Wheeler et al. 2011) especially in lung cancer (examples include(Ren et al. 2012; Shiraishi et al. 2010; Tan et al. 2011), gastric cancer (Park et al. 2011) and pancreatic cancer (Okazaki et al. 2008). As an indication of the importance of the DNA repair pathways in this area, these studies highlight the potential biomarker role of DNA repair associated genes, particularly of SNPs in ERCC1, XPD, XRCC1 and XRCC3, as predictive markers of patient outcome (Tiseo et al. 2013).

Despite these advances, progress to a clinically applicable marker of platinum response or toxicity is hindered by the fact that the results in the literature are not consistent, mainly relate to cisplatin, and typically are from retrospective single-arm studies (Tiseo et al. 2013). To-date, none of these putative associations has been translated into a tool that has made an impact in the clinic, and the optimal pathway for the rapid translation of this type of pre-treatment genetic screening assay into the clinic is uncertain (Houtsma et al. 2010).

6.1.2 Oxaliplatin Induced Peripheral Neuropathy - OIPN

The natural history and molecular mechanism of the development of OIPN is extensively discussed in chapter 1, section 1.4.4. The serious long term impact of OIPN and the lack of available diagnostic tools and treatments makes OIPN an ideal area for the development of predicative biomarkers enabling the stratification of treatment and risk reduction of this adverse event (Travis et al. 2014).

6.1.3 Proposed biomarkers of OIPN

Several studies have focussed on polymorphisms in DNA repair and platinum-drug detoxification pathways that could be used to predict oxaliplatin response and OIPN toxicity (Boige et al. 2010; Custodio et al. 2014; Inada et al. 2010; Kanai et al. 2010). Examples include studies examining SNPs in a host of genes related to oxaliplatin response and toxicity, including GSTP1 and GSTT1, members of the glutathione S-transferase pathway that sequester platinum agents in the cytoplasm (Cortejoso et al. 2013; Goekkurt et al. 2009b; Inada et al. 2010).

SNPs in several DNA repair genes involved in platinum-DNA adduct repair (see chapter 1, section 1.4) have been studied, including the polymorphism Asn118Asn in ERCC1, correlated to protein expression level and toxicity from oxaliplatin in colorectal cancer patients (Zhou et al. 2010), as have SNPs in ERCC2 and XRCC1 (Boige et al. 2010). In a study by Cortejoso (2013), SNPs in ERCC1 have also been associated with reduced risk of toxicity from treatment (Cortejoso et al. 2013). Other genetic studies have shown contradictory results from investigating the same SNPs, particularly in ERCC1 (reviewed in (Henriette Tanja et al. 2009)) and different groups have not found a strong link with ERCC1 and OIPN (Chua et al. 2009; Cortejoso et al. 2013{Inada, 2010 #2474), for several reasons, as will be discussed below.

In the multitude of published GWAS and genetic studies of chemotherapy-induced neuropathy (extensively reviewed in(Cavaletti et al. 2011)) many

candidate genetic polymorphisms have been proposed as potentially having an impact on toxicity and response, particularly to oxaliplatin. These studies are typically single centre studies investigating statistical associations in a selected handful of genes considered to be potentially correlated to outcome. The currently available data is conflicting, inconclusive, and insufficient to justify the use of these markers in clinical practice (Cavaletti et al. 2011; Custodio et al. 2014). In addition to the conflicting data from many of the statistically significant GWAS and genetic studies, very few groups provide any functional experimental evidence to support their genetic findings.

One example of an alternative functional approach to OIPN biomarker detection is a study by Pieck (Pieck et al. 2008). The kinetics of oxaliplatin-DNA adduct formation and repair was investigated in peripheral blood cells from 27 patients undergoing chemotherapy with oxaliplatin for a variety of solid tumours. Oxaliplatin-DNA adducts were measured by using atomic absorption spectroscopy to detect platinum molecules bound to extracted genomic DNA, giving a platinum-per unit of DNA measurement for samples from peripheral blood mononuclear cells (PBMC) taken from patients after treatment. In the small number of patients tested (with a heterogeneous treatment schedule and in a wide variety of solid tumours) a trend towards higher oxaliplatin-DNA adduct levels in patients with more OIPN was described, but perhaps not surprisingly given the small sample size and the relatively crude assay this was not statistically significant (Pieck et al. 2008).

Attempts have also been made to develop clinical predictive risk markers for OIPN (Argyriou et al. 2013; Renn et al. 2011; Velasco et al. 2014) but have been hampered by a lack of agreement on a consistent grading system for OIPN. One of several grading systems is commonly used for clinical trials (Cavaletti et al. 2013), making comparisons between studies using different scales difficult to interpret, and to-date no consistent tool for neuropathy grading has been adopted in routine practice (Argyriou et al. 2014).

To-date no biomarker or risk stratification tool has been developed to accurately stratify patients for the risk of OIPN. A recent study, lead by

Professor. J. Cheadle at Cardiff University, has identified new candidate SNPs in DNA repair genes that could be associated with the risk of developing OIPN (West 2013). Given the conflicting results and the lack of supporting mechanistic work in the publications discussed above, our approach, outlined in the experiments described in this chapter, is to develop additional functional and mechanistic experimental evidence to support the statistically significant findings described below.

6.2 Candidate biomarkers of OIPN recently discovered in the COIN trial: OIPN associated SNPs in XPF

Recently, the laboratory of Professor J. Cheadle at Cardiff University has used an enriched population of patients with OIPN to find candidate genetic markers that may predict for the development of this adverse effect. A brief outline of the protocol used by the Cheadle laboratory to generate this data is discussed below.

6.2.1 Methods

Genomic DNA was obtained from stored blood samples from the Cardiff University-based, UK national 2445 patient COIN trial investigating the benefit of continuous compared to intermittent oxaliplatin and fluoropyrimidine palliative chemotherapy for advanced colorectal cancer and the effect of the addition of anti-EGFR cetuximab to standard chemotherapy (Adams et al. 2011; Maughan et al. 2011). All patients received oxaliplatin and fluoropyrimidine treatment for the first 12 weeks, with a randomisation for the addition of cetuximab therapy (Figure 6.1).

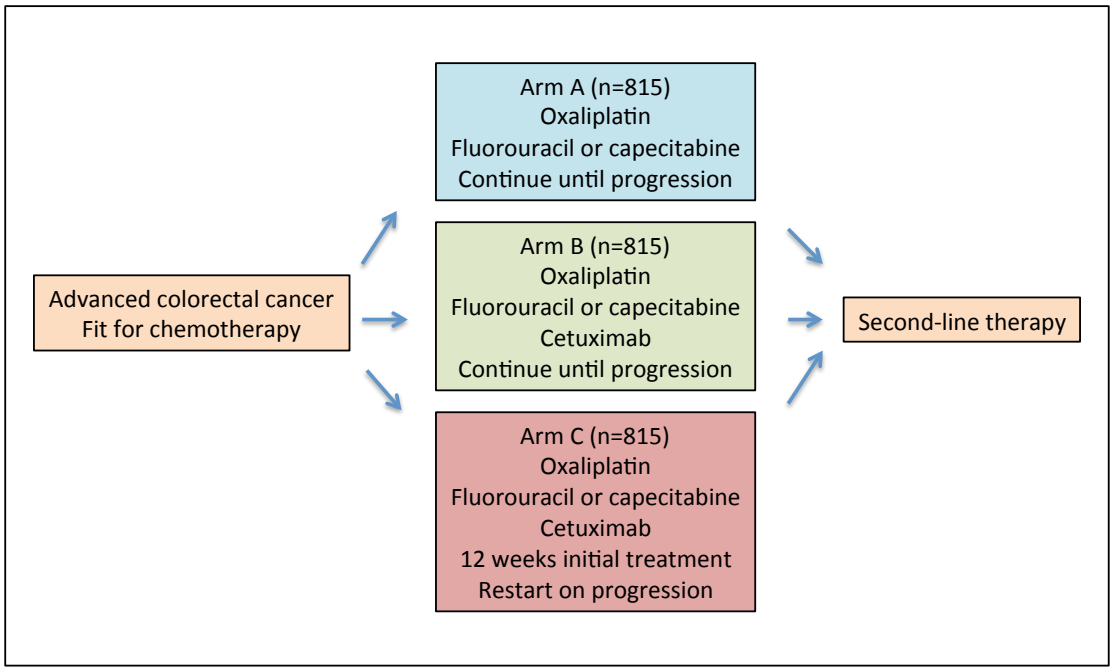


Figure 6.1: The UK COIN trial design (from (Adams et al. 2011))

During an initial phase of the this study DNA from the 10 patients identified as developing the most severe OIPN (at least CTC grade 3 or 4) (figure 6.2) during the first 12 weeks of treatment was exome re-sequenced.

NEUROLOGY Page 4 of 5

		Grade				
Adverse Event	**Short Name**	**1**	**2**	**3**	**4**	**5**
Neuropathy: sensory	Neuropathy-sensory	Asymptomatic; loss of deep tendon reflexes or paresthesia (including tingling) but not interfering with function	Sensory alteration or paresthesia (including tingling), interfering with function, but not interfering with ADL	Sensory alteration or paresthesia interfering with ADL	Disabling	Death

REMARK: Cranial nerve sensory neuropathy is graded as Neuropathy: cranial – *Select*.

Figure 6.2: NCI-CTC Common Toxicity Criteria for Adverse Advents Version 3:
Sensory Neuropathy Criteria (http://ctep.cancer.gov) (ADL- activities of daily living)

The exome resequencing data obtained was filtered using a pathway approach, concentrating on a list of 104 genes that potentially have a role in the development of OIPN, identified through literature reviews. The data was examined for the presence of rare SNPs, stop-gain mutations and truncating indels. The genes studied predominantly involved DNA damage response and signalling pathways (n=32) and DNA repair pathways (n=32), and also included

genes involved in drug influx (n=4), oxaliplatin trafficking (n=3), drug detoxification (n=7), oxalate metabolism (n=2), sequestration (n=3), and drug efflux (n=7). Although patients with a history of pre-existing neuropathy were excluded from the COIN trial, the main genetic causes of peripheral neuropathy were excluded in all 10 patients in this sub-study by sequencing the 28 genes associated with rare hereditary neuropathies in each patient. No causative mutations, polymorphisms or truncating indels in this selection of neuropathy-associated genes was identified.

6.2.2 Study results

A single patient with severe OIPN carried a novel stop gain (a polymorphism resulting in the change of a codon to a stop-codon) - Ser613X in exon 9 of *XPF*. Sanger sequencing of an independent PCR product verified the presence of the novel SNP and mutation in the second *XPF* allele was excluded by direct sequence analysis of the entire open reading frame and flanking intronic sequences. This patient had no past medical history of skin cancers or related diseases, and no history of ataxia, memory loss or muscle weakness. The patient did not clinically have XP and was a carrier of a single mutant *XPF* allele. This raises the possibility that haploinsufficiency for a mutant *XPF* allele is sufficient to induce OIPN upon exposure to oxaliplatin.

To explore further the potential role of SNPs in *XPF* as a biomarker of OIPN, the open reading frame and flanking sequences of the gene were sequenced in an additional 54 patients in the COIN cohort who developed CTC grade 3 or 4 OIPN during treatment.

Using this approach, 5 non-synonymous variants were identified and referenced against known SNPs using the dbSNP database - http://www.ncbi.nlm.nih.gov/SNP/. The variant Pro379Ser was found in 3 patients with OIPN and was previously documented in dbSNP (rs1799802). Arg415Gln was found in 9 patients and in dbSNP (rs1800067), His466Gln in a single patient and not in dbSNP, Arg576Thr was present in 1 patient and in

dbSNP (rs1800068) and Glu875Gly was identified in 4 patients and in dbSNP (rs1800124).

An *in silico* analysis of the potential of these variants to affect XPF protein function was conducted using the online tool Align-GVDG (http://agvgd.iarc.fr/agvgd_input.php, (Tavtigian et al. 2006)). The variants Pro379Ser, Arg576Thr and Glu875Gly were predicted to interfere with function, whilst the variant Arg415Gln was not.

To examine the prevalence of these variants in the COIN population all 2186 available sample from individual patients in the COIN study were genotyped for the 4 *XPF* gene variants previously listed in dbSNP (Pro379Ser, Arg415Gln, Arg576Thr and Glu875Gly) with an overall genotyping success rate of 98.1%. The frequencies of individual variants, grouped by the putative functional effects, in patients with, and without, OIPN were compared (Table 6.1). Each 'functional' variant was found more frequently in cases with OIPN as compared to those without, however, none were individually significantly over-represented. In combination more patients with OIPN carried a potentially functionally impaired variant (7/63, 11.11%) compared to patients without OIPN (90/1762, 5.1%). This gives an overall odds ratio for the presence of one of the functionally impaired variants in a COIN study OIPN case, compared to a non-OIPN case, of 2.32 [95% CI 1.02-5.24]. Considering the presence of a function variant in an OIPN case, compared to a non-OIPN case, the statistical significance was assessed using Chi square (and Fishers exact test for cells with a value of <5), giving a X^2 of 4.36 and a *P* value of 0.037. In contrast the variant, Arg415Gln, which was predicted as unlikely to interfere with function, was found in a similar proportion of patients with and without OIPN (Table 6.1).

Non-synonymous and stopgain variants associated with OIPN						
Variant	**rs Number**	**Frequency in OIPN patients**	**Frequency in non-OIPN patients**	**Odds Ratio (95% CI)**	**X^2**	***p***
Pro379Ser	rs1799802	3/63 (4.8%)	27/1763 (1.5%)	-	-	0.082
Arg576Thr	rs1800068	1/63 (1.6%)	4/1762 (0.2%)	-	-	0.161
Ser613X	-	1/63 (1.6%)	-	-	-	-
Glu875Gly	rs1800124	4/63 (6.4%)	60/1763 (3.4%)	-	-	0.176
Total		**7/63 (11.1%)**	**90/1762 (5.1%)**	**2.32 (1.02-5.24)**	**4.36**	**0.037**
Variants unlikely to affect function						
Arg415Gln	Rs1800067	**9/63 (14.3%)**	**260/1754 (14.8%)**	-	**0.014**	**0.91**

Table 6.1: Variants in *XPF* identified in the study by Cheadle (West 2013).
The frequency of each SNP in OIPN cases and controls, and the statistical significance is displayed. The four variants with a likely functional role (defined using align-GVGD) when combined are significantly more likely to occur in OIPN cases compared to controls (OR 2.32). The non-functional Arg415Gln is equally represented in cases and controls. The Chi-square (X^2) test was used to test significance (or Fishers exact test was used if cell value was below 5). The significance is shown by the respective P value (P), given along with the odds ratio (OR) with 95% confidence intervals.

6.2.3 Study summary: a rationale for mechanistic experiments

With a conservative odds ratio of 1.8 it would need a trial greater than four times the size of COIN (n=252 cases with OIPN and 7048 cases without OIPN) to have sufficient power to replicate these findings at >75% power at P=0.05 – a consequence of the relative rarity of the *XPF* alleles identified in the study described here. Additionally, a detailed *in silico*-based analysis of the variants on function is also not feasible as there is limited data available on the crystal structure of *XPF*. Given these constraints, in addition to the rationale for functional experiments to support these types of genetic studies discussed above (section 6.1.3), functional and mechanistic experimental evidence is required in order to demonstrate whether a causative effect of these novel SNPs on DNA repair exists, and hence a possible link with the risk of developing OIPN.

6.3 Potential mechanism of OIPN susceptibility as a result of XPF SNPs - The use of the budding yeast *S. cerevisiae* as a genetic model

The budding yeast *S. cerevisiae* has been a widely used model organism in genetic research for several decades and has been described as the 'premier model for eukaryotic cell biology' (Botstein and Fink 2011). This is especially true in the field of NER research, in which the proteins and pathways of NER in *S. cerevisiae* and humans are highly conserved (Reed 2011), and for gene-protein-function association experiments (Botstein and Fink 2011; Dunham and Fowler 2013). In an era with a wealth of genetic sequencing information derived from human tissues and tumours, defining which genetic variants result in significant changes to phenotype is often a rate-limiting step (Dunham and Fowler 2013). To overcome this, *S. cerevisiae* has long been extensively used in many studies to recreate genetic variants in homologous genes, including experiments identifying the association of colorectal cancer risk with the alleles of the MMR gene *MSH2* (Gammie et al. 2007) and the relationship between variants in *MTO1*, detected by high throughput sequencing, and the cardiac condition hypertrophic cardiomyopathy (Ghezzi et al. 2012).

Yeast models have several advantages over their mammalian counterparts. The doubling time of yeast is approximately 90 minutes, compared to several days for most human cell cultures, allowing for more rapid cell production and experimental throughput, and the small, well-characterised genome is a manageable 12MB, with many tried-and-tested tools available for accurate genetic manipulation. The key advantage is gene targeting due to extraordinary homologous recombination efficiency (Dunham and Fowler 2013).

To simplify this study, two of the SNPs identified by the Cheadle laboratory will be taken forward into our *S. cerevisiae* genetic model. Ser613X, the stop-gain mutation, will be modelled, as it is novel, likely to have a more pronounced phenotype and is potentially easier to detect. Of the five other non-synonymous mutations identified Pro379Ser occurs at a highly conserved region of the yeast *XPF* homolog *RAD1* and will be used to model the effect of the variants on cell survival and DNA repair.

6.3.1 An introduction to *S. cerevisiae RAD1* and *RAD10*, and the mammalian homologs *XPF* and *ERCC1*

Initially, the details of much of the mechanisms of human DNA repair were elucidated in genetic studies using *S. cerevisiae*, an approach possible because of the high degree of conservation between these systems in eukaryotes. Early experiments demonstrated that induced mutations leading to deletion of one or more of a group several genes result in extreme sensitivity to UV radiation. This group of genes, named RAD for 'radiation sensitive', include the regions subsequently identified as *RAD1* and *RAD10*, indicating an important role for these genes in repair of UV damage. Subsequent screening of yeast genomic libraries for the complementation of UV radiation sensitivity of *RAD1* and *RAD10* mutants identified the sequence of the genes concerned. These were later identified to be highly homologous to the mammalian orthologs *XPF* and *ERCC1* (reviewed in (Friedberg 2005)).

With *in vitro* yeast two-hybrid experiments it was subsequently demonstrated that, in the absence of damaged DNA, Rad1 and Rad10 protein form a stable heterodimeric complex with high affinity and a long half-life of >15 hours (Bailly et al. 1992; Bardwell et al. 1992). In a purified protein system the endonuclease activity of Rad1 and Rad10 requires the combination of both proteins; when applied individually no activity can be detected (Sung et al. 1993; Tomkinson et al. 1993). The Rad1-Rad10 complex has significant affinity for junctions between double stranded DNA and single stranded DNA tails with 3'ends (Bardwell et al. 1994), which indicates its role in NER as the equivalent to ERCC1-XPF, cleaving 5' of the pre-incision complex at the junction of single stranded and double stranded DNA (Marteijn et al. 2014), making it one of the two junction specific endonucleases involved in excision of the damage containing oligonucleotide during NER.

The protein structures of Rad1 and Rad10 are shown below (Figure 6.3). The Rad1 protein includes a Rad10 protein-binding domain between amino acids 809 and 997, and conversely the Rad1 binding domain of Rad10 is located between amino acids 90 and 210 (Bardwell et al. 1993; Bardwell et al. 1992). These interaction domains are hydrophobic and the strength of the complex can

be increased at a higher ionic strength (Bardwell et al. 1993). The association between proteins is mediated by a leucine-isolucine zipper motif between amino acids 931 and 966 in the Rad10 binding domain of Rad1 (Prakash et al. 1993), and a helix-turn-helix motif typical of protein-protein interactions have been demonstrated in the Rad1 binding domain of Rad10 (Reynolds et al. 1985; van Duin et al. 1986).

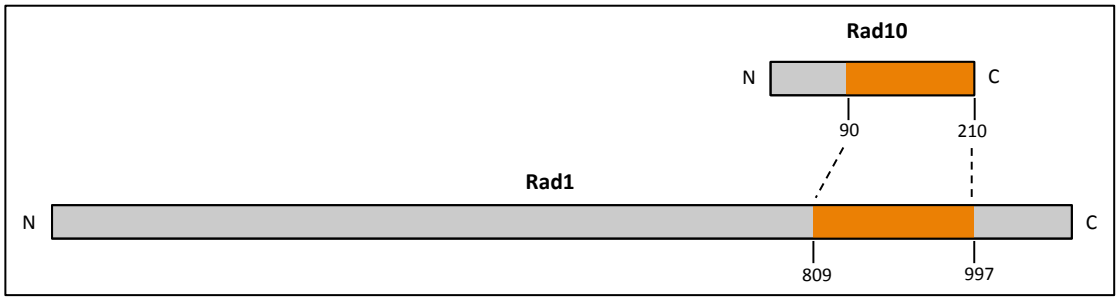


Figure 6.3: The interaction domains of the Rad1 and Rad10 proteins.
The regions (orange) between amino acids 90 and 210 of Rad10 and 809 and 997 of Rad1 are necessary for the heterodimerisation and interaction of these proteins (Adapted from (Friedberg 2005))

6.3.1.1 The *XPF* gene and XPF protein –the *RAD1* human homolog

The human orthologs of the *S. cerevisiae RAD1* gene specifically corrects phenotypes in cells from patents of XP complementation group F and was identified from its homology with the yeast *RAD1* gene (Sijbers et al. 1996). All patients in XP complementation group F have been shown to have mutation in the *XPF* gene (Matsumura et al. 1998; Sijbers et al. 1996). The open reading frame of *XPF* encodes a 916 amino acid polypeptide of molecular weight 104 kDa orthologous to Rad1 protein of *S. cerevisiae* (Brookman et al. 1996).

6.3.1.2 *ERCC1*- the *RAD10* human homolog

The cDNA of the *RAD10* human homolog *ERCC1* encodes a polypeptide of 297 amino acids and of 32.5kDa (van Duin et al. 1986). However, similar to yeast Rad1-Rad10, its role in NER requires interaction of ERCC1 protein in a

heterodimeric complex with XPF. In the absence of *XPF*, the human *ERCC1* gene does not compliment XP cell lines that are defective in NER (van Duin et al. 1989).

6.3.1.3 ERCC1-XPF also form a functional heterodimeric complex

The initial suggestion that the ERCC1 and XPF proteins form a heterodimer complex was evidenced by the lack of complementation when mutant cell extracts are mixed *in vitro* (Biggerstaff et al. 1993; Reardon et al. 1993; van Vuuren et al. 1993) and were subsequently shown to be active in a heterodimeric complex (Brookman et al. 1996; Sijbers et al. 1996). ERCC1 and XFP are unstable in the absence of the other protein, as can be demonstrated by low levels present of ERCC1 protein in *XPF* mutant cells in the presence of normal ERCC1 mRNA levels (Biggerstaff et al. 1993; Yagi et al. 1997), indicating breakdown of the protein after synthesis.

Analogous to Rad1-Rad10, ERCC1-XPF is a structure specific endonuclease and can cut at junctions between double and single stranded DNA where the single strand moves 5' to 3' away from the junction (de Laat et al. 1998; Sijbers et al. 1996), allowing the protein to act as a site specific endonuclease 5' of unwound DNA during NER (Bardwell et al. 1994; Mu et al. 1996; Sijbers et al. 1996).

The XPF-ERCC1 heterodimer and is crucially involved in NER, but also in a number of other DNA repair pathways with a possible role in platinum-DNA repair. Its other functions are in ICL repair and in double strand break repair, especially in the function of replication dependent ICL repair (Kirschner and Melton 2010).

6.3.1.4 Protein domain structure of the ERCC1-XPF heterodimer

The ERCC1-XPF protein family (including Rad1 and Rad10) are highly conserved in eukaryotes. Neither protein has orthologs in *E. coli* or other eubacteria. Instead, XPF appears to be related to archeal helicases (Aravind et al.

1999; Sgouros et al. 1999). In archaeobacteria the protein equivalent of XPF only is present and functions as a homodimer (Aravind et al. 1999; Gaillard and Wood 2001), indicating that ERCC1 is likely to have evolved from gene duplication of XPF in the eukaryotic lineage (Gaillard and Wood 2001).

An overview of the protein domains of XPF and ERCC1 are shown below (Figure 6.4)

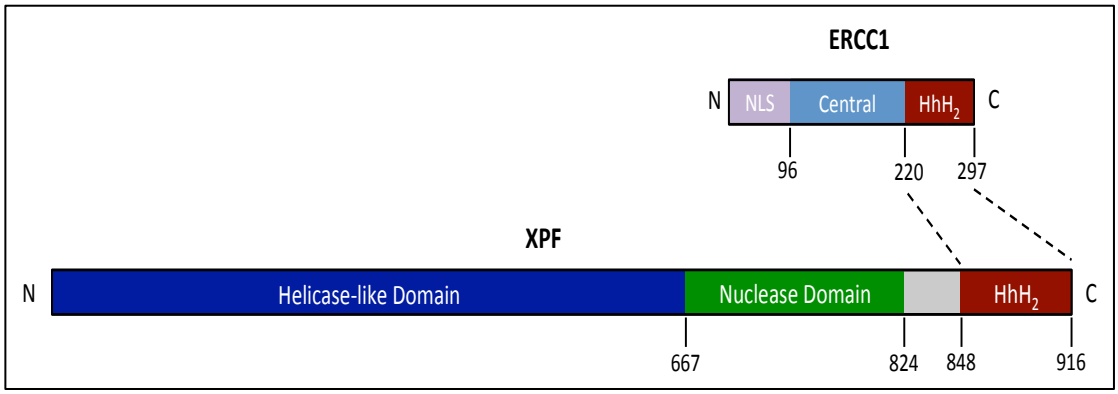


Figure 6.4: The domain structure of XPF and ERCC1.
The active nuclease domain of XPF is indicated in green and the H2H2 site of interaction between ERCC1 and XPF is shown in dark red. ERCC1 interacts with XPA through the central domain. NLS – nuclear localisation signature (Adapted from (McNeil and Melton 2012))

Alignments of ERCC1 and XPF orthologs from several eukaryotes reveal areas of significant similarity at the N- terminal and C-terminal thirds, separated by a poorly conserved central region. The C-terminus of both proteins is highly conserved and contains helix-hairpin-helix motifs (HhH_2) (Gaillard and Wood 2001; Sijbers et al. 1996). These regions are the interacting domains between XPF and ERCC1.

The HhH_2 domains are similar and are highly conserved and XPF adopts a canonical HhH_2 folded structure (Tripsianes et al. 2007). The two proteins dimerise at the HhH_2 domains to form a stable heterodimer and the dimerised H_2H domains form independent binding sites to complex with single strand DNA junctions (Tripsianes et al. 2007).

Mammalian XPF has nuclease, C-terminal HhH_2 domains and a large N-terminal helicase-like domain (Figure 3.3). Despite the structural similarity XPF notably does not function as a helicase (Sgouros et al. 1999). The structure is

consistent with the function of XPF as an ATP-independent nuclease rather than an ATP-dependent helicase - the similarities between XPF and helicase structure may reflect the necessity to bind DNA as many helicase enzymes bind to the junction between dsDNA and RNA tail with a 3'end, similar to the function of XPF to bind the double strand-single strand DNA junction in a 5' to 3' direction prior to cleavage at double strand/ single strand junctions (de Laat et al. 1998; Yao et al. 1997).

The nuclease region 667 to 824, contains approximately 90 amino acids from 670-740 that are highly conserved in the XPF and archeal enzyme families, and is not present in ERCC1. Using an affinity cleavage assay the active site of nuclease catalysis by ERCC1-XPF was located to residues 670-740 of XPF (Enzlin and Scharer 2002). Mutation of several specific residues, including Asp 687, Asp 715, Lys 727 and Asp731 have identified amino acids in the this region that are key to the nuclease function of XPF (McNeil and Melton 2012).

The ERCC1 protein differs from XPF. Aside from the C-terminal HhH_2 domain, it consists of a catalytically inactive central domain responsible for DNA binding, and unlike XPF has no helicase domain (Figure 6.4).

6.3.1.5 Specific function of XPF and ERCC1 subunits

In mammalian ERCC1-XPF the XPF sub-unit contributes the nuclease activity whilst the helicase related N-terminal domain probably contributes to DNA binding. ERCC1 has no active nuclease site but retains the HhH domain for active binding of XPF and to DNA. As ERCC1 is catalytically inactive, its role is in mediating DNA and XPA interactions through a v-shaped groove in the central domain. This region binds to single stranded DNA with a strong preference for 5' overhangs, resulting in a degree of structural specificity of the ERCC1-XPF complex (McNeil and Melton 2012).

6.3.1.6 The role of ERCC1-XPF in NER

As discussed previously (Chapter 1 section 1.4) NER involves the sequential action of approximately 30 proteins in a series of stepwise and complex interactions (See chapter 1, section 1.4, Figure 1.12). The role of ERCC1-XPF was initially elucidated by *in vitro* studies that demonstrated that the incision step requires 6 proteins to be present *in vitro*, namely XPC/RAD23B, XPA, RPA, TFIIH, XPG and XPF-ERCC1 (Aboussekhra et al. 1995) and the protein-protein interaction of these individual pathway components is fundamental for NER to proceed (Park and Choi 2006). Once DNA damage is recognised by stalled RNA polymerase II at transcription forks (TC-NER sub-pathway) or through XPC-RAD23B at structurally distorted DNA (via the GG-NER sub-pathway) the common phase of the repair mechanism occurs, consisting of recruitment of TFIIH, comprising key subunits XPB and XPD directional helicases which unwind DNA through an ATP-dependent helicase activity (Park and Choi 2006). Recruitment of XPA and RPA proceed to stabilise the unwound complex and recruit the structure specific endonucleases XPG and ERCC1-XPF to incise the damaged strand either side of the damage. XPA recognises the helical kink at the unwinding of DNA duplex and acts as the framework for the binding of TFIIH, RPA and ERCC1-XPF.

The function of the ERCC1-XPF heterodimer is thought to be mediated by ERCC1-XPA interaction and XPF-RPA interaction, resulting in excision of the single strand at the double-single strand 5' unwinding junction (Friedberg 2005; Marteijn et al. 2014; McNeil and Melton 2012).

6.3.1.7 ERCC1-XPF in DSB repair

Double strand breaks are induced by ionizing radiation, free radicals and chemotherapy agents and can be repaired by either homologous recombination (HR) or by non-homologous end joining (NHEJ). Rad1 and Rad10, the orthologues of ERCC1 and XPF, were used initially to demonstrate the importance of this complex in DSB repair (Schiestl and Prakash 1990).

ERCC1-XPF mutant mammalian cells are sensitive to DSBs (Ahmad 2008) (Ahmad et al. 2008) and the NHEJ pathways for DSB repair are attenuated (Al-Minawi et al. 2009; Niedernhofer et al. 2001; Sargent et al. 1997). The main function appears the be the role of ERCC1-XPF in removing non-homologous 3' single stranded flaps at the broken end of double strand breaks before they are repaired (Ahmad et al. 2008) in single strand annealing (SSA) sub pathway of HR and in micro-homology mediated end joining (MMEJ) sub pathway of NHEJ (Ahmad et al. 2008; Ma et al. 2003; McVey and Lee 2008).

6.3.1.8 ERCC1-XPF in ICL repair

Removal of inter-stand cross-links (ICLs) is essential to cellular survival. Crosslinks can be formed through several mechanisms, including through the action of the platinum agents, as well as other chemotherapy drugs including mitomycin C, and agents such as psoralens (Wood 2010). As discussed previously (Chapter 1, section 1.4.2.1), the linking of opposite strands of DNA prevents essential cellular functions of transcription and DNA replication. Evidence for the importance of this pathway is that mammalian NER mutants are sensitive to cross-linking agents and XPF-ERCC1 mutants are particularly sensitive, indicating an especially important role of the heterodimer in this pathway (Wood 2010).

Although the full and precise mechanism of ICL repair is not fully elucidated, (see chapter 1, section 1.4.2.1) XPF-ERCC1 features in a dependent step and has been shown to be required for both S-phase dependent and independent ICL repair (Bhagwat et al. 2009; Niedernhofer et al. 2004).

6.3.1.9 Non-repair roles of ERCC1-XPF

The ERCC1-XPF heterodimer is also involved in telomere maintenance and interacts with TRF2, a telomere associated protein. It has also has been shown to have a role in mitotic progression (McNeil and Melton 2012). Given

the mechanism of platinum damage and mechanism of OIPN, these non-DNA repair interactions are unlikely to be related to this condition.

XPF has also been shown to participate in the modification to chromatin structure during transcription, in a role independent of its function in the repair of DNA damage (Le May et al. 2012).

6.3.2 Summary – the potential effects of SNPs on XPF protein structure and function

The variants identified in the genetic study discussed above and taken forward into this investigation, Ser613X and Pro379Ser, both potentially have an effect on the structure and function of XPF. As described in the previous section ERCC1-XPF is involved in several molecular pathways, primarily DNA repair, but with several additional roles in other non-repair pathways. Using a yeast model system it is possible to examine the effect of mutations on many of these pathways, including NER, DSB repair and ICL repair. As RAD1, and subsequently XPF, were discovered through the profound effect of their absence on NER, it is likely that the central role these proteins play in this pathway will give the highest chances of determining if functional effects result from the presence of these variants. For this reason, the effect of the Pro379Ser and Ser613X variants will primarily be assessed thought the impact of these mutations on NER, with the opportunity for future studies (beyond the scope and time constraints of this project) to examine the impact of these changes on other DNA repair, and non-repair, molecular pathways.

6.4 Aims of this chapter

This chapter will investigate whether it is possible to demonstrate a functional role of single nucleotide polymorphisms recently identified from a cohort of colorectal cancer patients who developed OIPN during chemotherapy. These polymorphisms are located in the NER gene XPF, involved in both the NER and interstrand crosslink repair (ICL) pathways, and may have a functional impact on NER, and consequently on the risk of developing OIPN.

The functional effects of the recently identified XPF SNPs will be investigated in an *S. cerevisiae* model system. The appropriate mutations will be introduced into the *XPF* homolog *RAD1* and their effect will be assessed on UV-induced DNA damage and repair (as a marker of NER capacity) and oxaliplatin sensitivity.

This study will be conducted in the following stages. Firstly, a plasmid containing a selectable marker and wild type (WT) *RAD1* will be constructed. After demonstrating appropriate function by complementation of UV resistance in a *rad1Δ* strain, mutations in p*RAD1* will be introduced into the plasmid construct and transformed into *rad1Δ* strains. These models will then be assessed for UV and oxaliplatin sensitivity compared to control strains to determine if there is a functional effect in these OIPN-associated variants.

6.5 Materials and Methods

For a full description of the materials and methods used see chapter 2, section 2.3-2.4. A brief outline will be described at each phase in the following discussion

6.5.1 Strain Tables

Strain (haploid)	Genotype	Source
SX46a Wild-Type	*Mat a, ade2, his3-372, trp 289, ura3-52*	EUROSCARF
SX46a *rad1*Δ	*Mat a, rad1::URA3,* *ade2, his3-372, trp-289, ura3-52*	EUROSCARF
SX46a *rad1*Δ pRS313-*RAD1*		This Study
SX46a *rad1*Δ pRS313-*RAD1*-P469S		This Study
SX46a *rad1*Δ pRS313-*RAD1*-S747X		This Study

Table 6.2: The *S. cerevisiae* strain identifier, genotype and origin of the five strains used for this study

6.5.2 Conservation of amino acid sequence between XPF and RAD1

The site of mutations to be inserted into the *RAD1* gene to enable modelling of the XPF SNPs was identified by an alignment comparison between the amino acid sequence of the homologs XPF and RAD1 using the online tool Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) with protein sequences obtained from NCBI (http://www.ncbi.nlm.nih.gov/protein). To increase the validity of the alignment, the XPF homolog Rad16 from the yeast *Saccharomyces Pombe* was also aligned with both XPF and Rad1 protein sequences. The alignment is shown below (figure 6.5) with the Rad1 amino acid sequence also displayed for reference (figure 6.6). The alignment demonstrates the conservation of sequence around position 379 in XPF, with a proline residue present at XPF position 379, as shown by the green star, corresponding to Rad1 position 469 (figure 6.6). The XPF Ser613 alignment demonstrates serine conserved at XPF position 613, indicated by the pink star, and corresponding to Rad1 position 747 (figure 6.6).

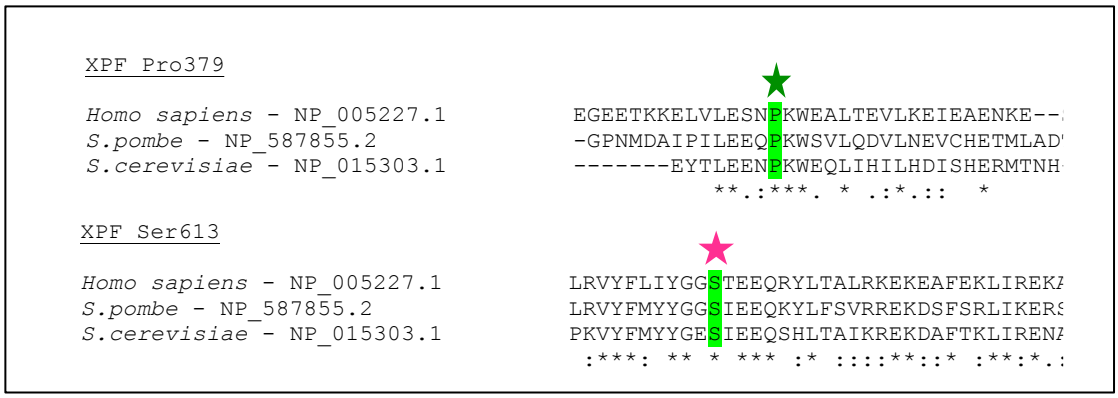


Figure 6.5: Alignment of residues in *XPF (Homo Sapiens),* Rad 16 *(S. pombe)* and Rad1 (*S. cerevisiae)* associated with OIPN in the COIN trial patient cohort.
Pro379 and alignment with XPF and Rad1 is shown by the green band and green star. The alignment of Ser613 in XPF with Rad1 in S. cerevisiae is shown by the green band and pink star.

```
   1  MSQLFYQGDS DDELQEELTR QTTQASQSSK IKNEDEPDDS NHLNEVENED
  51  SKVLDDDAVL YPLIPNEPDD IETSKPNIND IRPVDIQLTL PLPFQQKVVE
 101  NSLITEDALI IMGKGLGLLD IVANLLHVLA TPTSINGQLK RALVLVLNAK
 151  PIDNVRIKEA LEELSWFSNT GKDDDDTAVE SDDELFERPF NVVTADSLSI
 201  EKRRKLYISG GILSITSRIL IVDLLSGIVH PNRVTGMLVL NADSLRHNSN
 251  ESFILEIYRS KNTWGFIKAF SEAPETFVME FSPLRTKMKE LRLKNVLLWP
 301  RFRVEVSSCL NATNKTSHNK VIEVKVSLTN SMSQIQFGLM ECLKKCIAEL
 351  SRKNPELALD WWNMENVLDI NFIRSIDSVM VPNWHRISYE SKQLVKDIRF
 401  LRHLLKMLVT SDAVDFFGEI QLSLDANKPS VSRKYSESPW LLVDEAQLVI
 451  SYAKKRIFYK NEYTLEENPK WEQLIHILHD ISHERMTNHL QGPTLVACSD
 501  NLTCLELAKV LNASNKKRGV RQVLLNKLKW YRKQREETKK LVKEVQSQDT
 551  FPENATLNVS STFSKEQVTT KRRRTRGASQ VAAVEKLRNA GTNVDMEVVF
 601  EDHKLSEEIK KGSGDDLDDG QEENAANDSK IFEIQEQENE ILIDDGDAEF
 651  DNGELEYVGD LPQHITTHFN KDLWAEHCNE YEYVDRQDEI LISTFKSLND
 701  NCSLQEMMPS YIIMFEPDIS FIRQIEVYKA IVKDLQPKVY FMYYGESIEE
 751  QSHLTAIKRE KDAFTKLIRE NANLSHHFET NEDLSHYKNL AERKLKLSKL
 801  RKSNTRNAGG QQGFHNLTQD VVIVDTREFN ASLPGLLYRY GIRVIPCMLT
 851  VGDYVITPDI CLERKSISDL IGSLQNNRLA NQCKKMLKYY AYPTLLIEFD
 901  EGQSFSLEPF SERRNYKNKD ISTVHPISSK LSQDEIQLKL AKLVLRFPTL
 951  KIIWSSSPLQ TVNIILELKL GREQPDPSNA VILGTNKVRS DFNSTAKGLK
1001  DGDNESKFKR LLNVPGVSKI DYFNLRKKIK SFNKLQKLSW NEINELINDE
1051  DLTDRIYYFL RTEKEEQEQE STDENLESPG KTTDDNALHD HHNDVPEAPV
1101  *
```

Figure 6.6: The amino acid sequence of *S. cerevisiae* Rad1 protein.
The position of the amino acids Pro469 and Ser747 are identified as desired sites for mutation for this study and are indicated by the green and pink stars respectively.

To model the two mutations XPF of Pro379Ser and Ser613X in the homologous *RAD1* gene, strains with the analogous mutations in the RAD1 protein of Pro469Ser and Ser747X must be constructed. Accounting for the altered codon composition between species, the simplest mutation required to model the changes in the human context is the substitution of proline (codon TCC) to serine (codon TTC) with the *RAD1* gene mutation C1406T (figure 6.7) and modification serine (codon AGT) to a stop codon (TGA) at positions A2240T and T2242A (figure 6.8), resulting in the desired amino acid and protein sequence substitutions

XPF Amino Acid Sequence	**L**	**E**	**S**	**N**	**P**	**K**	**W**	**E**	**A**

RAD1 Amino Acid Sequence	**L**	**E**	**E**	**N**	**P**	**K**	**W**	**E**	**Q**
Position	465	466	467	468	469	470	471	472	473
Initial DNA Sequence	TTA	GAA	GAA	AAT	CCA	AAA	TGG	GAA	CAA
Mutated DNA Sequence	TTA	GAA	GAA	AAT	TCA	AAA	TGG	GAA	CAA
Pro469Ser Variant Amino Acid Sequence	**L**	**E**	**E**	**N**	**S**	**K**	**W**	**E**	**Q**

Figure 6.7: The DNA sequence changes required that result in the Pro469Ser Rad1 variant. The proline amino acid at 469 (green box) is changed to a serine by the conversion of the cytosine at DNA position 1406 to a thymine.

XPF Amino Acid Sequence	**I**	**Y**	**G**	**G**	**S**	**T**	**E**	**E**	**Q**

RAD1 Amino Acid Sequence	**Y**	**Y**	**G**	**E**	**S**	**I**	**E**	**E**	**Q**
Position	743	744	745	746	747	748	749	750	751
Initial DNA Sequence	TAC	TAC	GGT	GAA	AGT	ATT	GAA	GAG	CAA
Mutated DNA sequence	TAC	TAC	GGT	GAA	TGA	ATT	GAA	GAG	CAA
Ser747X Variant Amino Acid Sequence	**Y**	**Y**	**G**	**E**	**STOP**	**I**	**E**	**E**	**Q**

Figure 6.8: The DNA sequence changes required that result in the Ser747X Rad 1 variant. The proline amino acid at 469 (green box) is changed to a serine by the conversion of the cytosine at DNA position 1406 to a thymine.

6.5.3 Overview of strain construction – cloning strategy

To investigate the functional effects of the recently identified *XPF* SNPs in an *S. cerevisiae* model the appropriate mutations need to be introduced into the homolog *RAD1.* The initial step is constructing plasmids containing a selectable marker (in this instance histidine) and wild type (WT) *RAD1* that can then be introduced into *rad1*Δ strains.

6.5.3.1 Preparation of *RAD1* DNA with modified restriction sites for cloning

A 4058bp product containing *RAD1* and 500 nucleotide promoter region was amplified from BY4742 genomic DNA template by PCR using a 30-mer upstream primer, modified to introduce a proximal BamH1 restriction site 500 nucleotides 3' of the ORF, and 36-nucleotide primer modified to introduce a Sal1 restriction site 5' of the stop codon. The primers used and the nucleotides altered to introduce the required restriction sites are highlighted and underlined in orange (BamH1) for the forward RAD1-BAMH1-U primer and purple (Sal1) for the RAD1-Sal1-L reverse primer.

PCR amplification was conducted using the high fidelity Phusion DNA polymerase (NEB) and PTJ2000 thermocycler using the following programme:

1) 95°C	2 minutes
2) 95°C	30 seconds
3) 55°C	30seconds
4) 70°C	60 seconds
5) Go to 2	15 times
6) 70°C	5 minutes

The resulting 4085-nucleotide PCR product was confirmed by gel electrophoresis following purification with PureLink PCR column purification kit (Invitrogen) (figure 6.9).

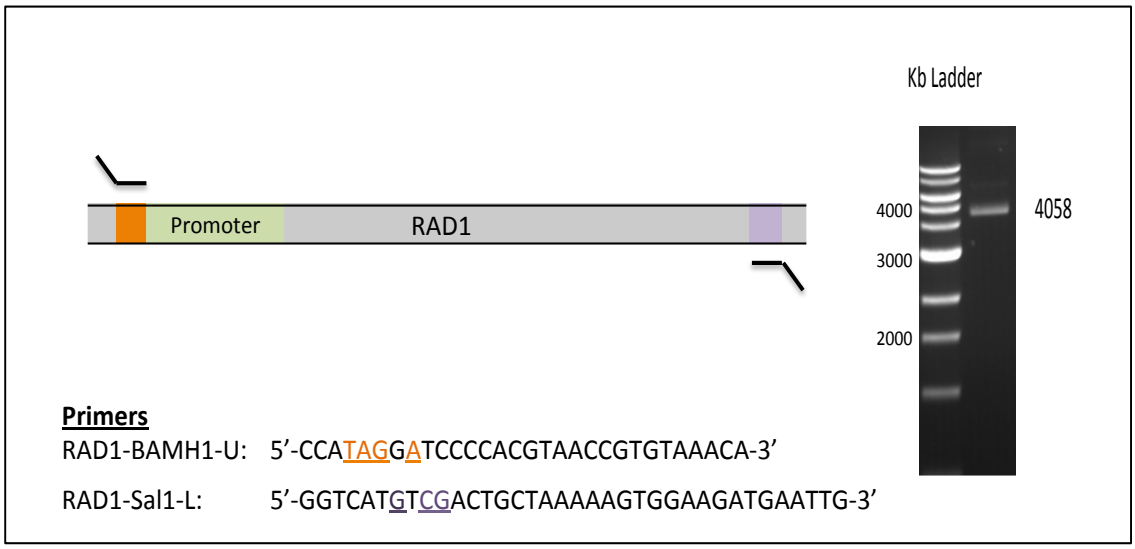


Figure 6.9: Introduction of restriction enzyme sites into *RAD1*
The 4058 kb region of genomic BY4742 DNA is indicated by the diagram in grey, with the forward RAD1-BAMH1-U primer position and reverse RAD1-Sal1-L primer position and sequence as indicated. The nucleotides modified from the initial *RAD1* sequence are indicated in orange to introduce a BamH1 restriction site proximal to 500bp of *RAD1* to include the promoter region. The reverse RAD1-Sal1-L primer is modified from the initial *RAD*1 with the nucleotides highlighted in purple to introduce a distal Sal1 restriction site, approximately 100bp 3' of the *RAD*1 gene (purple box). Following PCR a 4058 bp product can be identified on a 1% TAE gel.

6.5.3.2 Ligation to pJET1.2 plasmid vector

The 4058 nucleotide PCR product was blunt-end ligated to pJET1.2 (Thermo Scientific) plasmid vector and isolated using restriction endonucleases BamH1 and Sal1 (NEB), confirming appropriate ligation and generating correctly orientated overhanging ends, as confirmed by analysis of length of the products of restriction enzyme digestion (Figure 6.10).

For further details of the plasmid pRS313 genes and sequence see http://www.snapgene.com/resources/plasmid_files/basic_cloning_vectors/pJET1.2/)

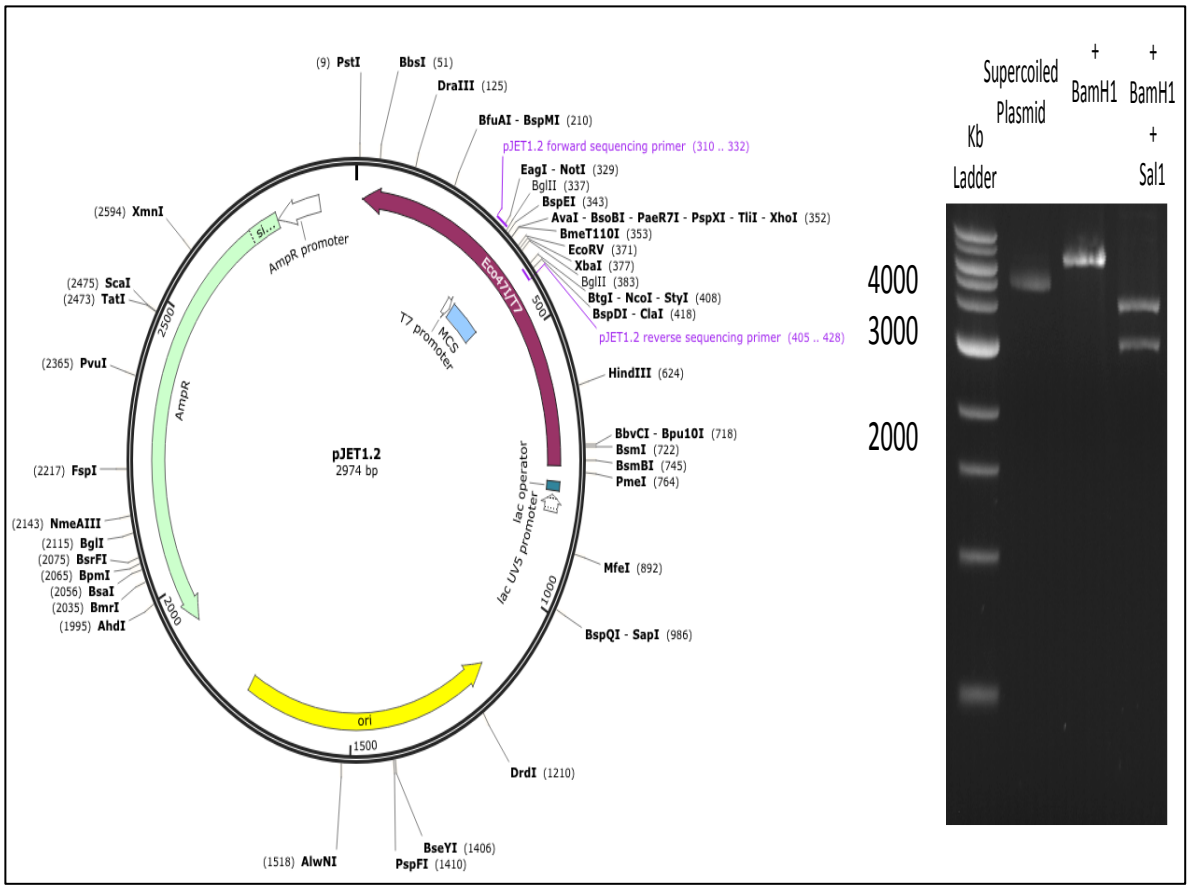


Figure 6.10: Blunt end ligation of *RAD1* into plasmid pJET1.2.
The 4058 bp *RAD*1 product is blunt-end ligated to 2974 bp pJET1.2 vector. Following ligation, digestion with a single BamH1 restriction enzyme results in a 7kb product when run on a 1% TAE gel, and digestion with both BamH1 and Sal1 returns the 4058 bp fragment and the 2974 bp pJET1.2 vector, a demonstrated by the gel image.
(pJET1.2 image obtained from
http://www.snapgene.com/resources/plasmid_files/basic_cloning_vectors/pJET1.2/)

6.5.3.3 Ligation into pRS313 plasmid vector

The PCR product from restriction digestion with BamH1 and Sal1 from the pJet1.2-*RAD1* plasmid was isolated and purified by gel purification. The purified product was subsequently ligated to plasmid pRS313 and correct ligation was confirmed by restriction enzyme digestion, resulting in two products of the appropriate size (figure 6.11).

For further details of the plasmid pRS313 genes and sequence see www.snapgene.com/resources/plasmid_files/yeast_plasmids/pRS313/

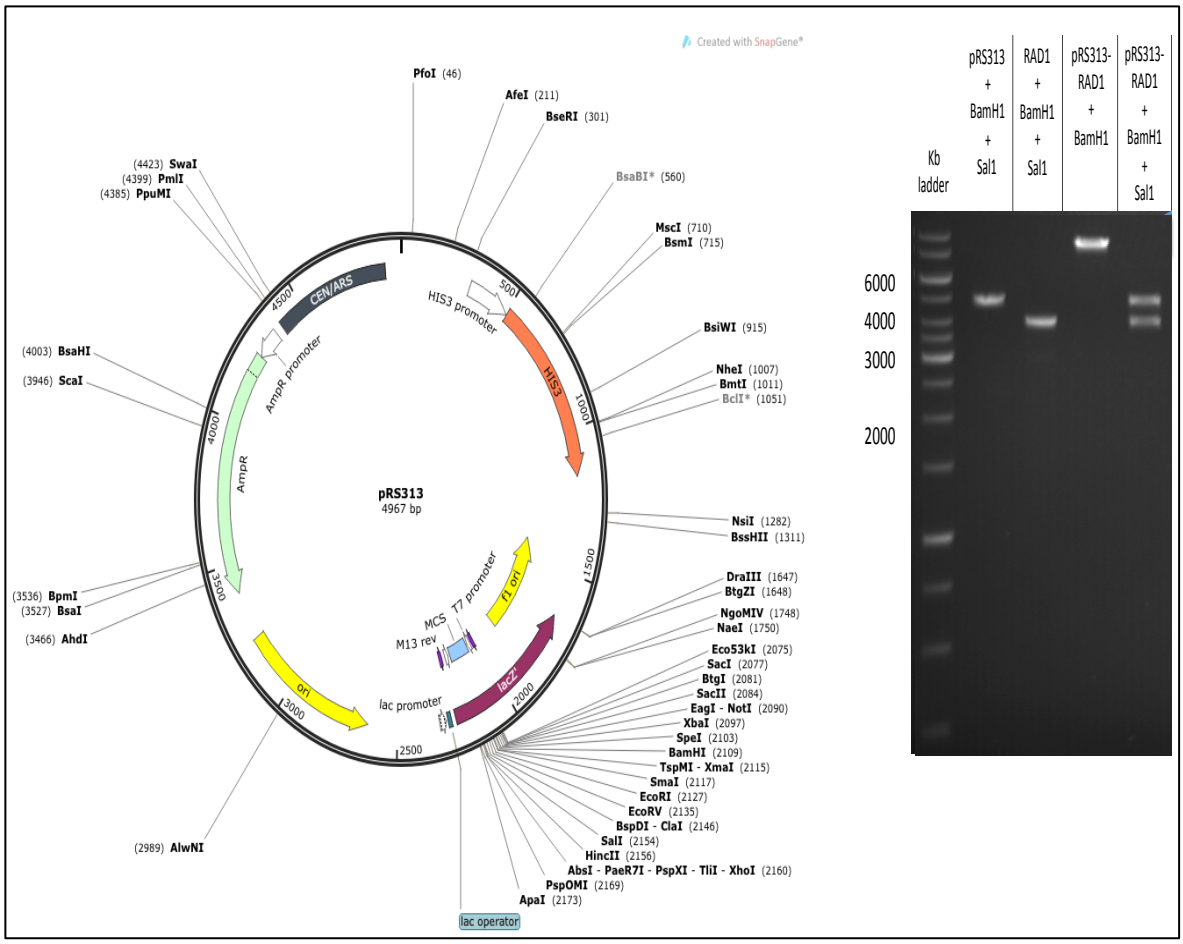


Figure 6.11: Insertion of *RAD1* into vector pRS313.
The 4058 bp fragment is ligated into the 4967 bp plasmid pRS313 at the BamH1 and Sal1 restriction sites. To confirm ligation, the plasmid and insert can be identified as a single band after single restriction enzyme digestion and gel electrophoresis of pRS313-RAD1, or a double band of the correct sizes after digestion with BamH1 and Sal1.
(pRS313 image from
https://www.snapgene.com/resources/plasmid_files/yeast_plasmids/pRS313)

6.5.3.4 *E. coli* transformation by pRS313-*RAD1*

To generate the quantity of plasmid necessary for the experiments outlined in this study, ultra-competent DH5α *E. coli* (Invitrogen) were transformed by the pRS313-*RAD1* plasmid, using the manufacturers standard protocol, and were allowed to grow overnight at 37°C in a rotatory incubator at 180 rpm prior to plasmid extraction using PureLink Quick Plasmid MiniPrep Kit (Invitrogen) plasmid using the manufacturers standard protocol.

6.5.3.5 Transformation of SX46a *rad1*Δby pRS313-*RAD1*

SX46a *rad1*Δ strain was transformed by the pRS313-*RAD1* plasmid following the TRAFO protocol ((Gietz and Woods 2002) and http://home.cc.umanitoba.ca/~gietz/method.html). As the pRS313 plasmid contains a selectable histidine genetic marker successful transformants were selected by the ability of colonies to grow on histidine deficient (His-) plates and in His- growth media.

To confirm successful transformation, the presence of *RAD1* DNA in a genomic BY4742 DNA control, the SX46a *rad1*Δ pRS313-RAD1 strain and the absence of *RAD1* DNA in the SX46a *rad1*Δ strain is demonstrated by colony PCR of a 1500 nucleotide product of the central section of *RAD1* in Figure 6.12, using the colony PCR method outlined in chapter 2, section 2.3.10 to extract DNA, and the primers Check-F and Check-R listed below

Check-F	5'-TCAGGTCAATTGACTCGGTGATGG-3'
Check-R	5'-GGCTAATCTGTTATTCTGTAATGACCCA-3'

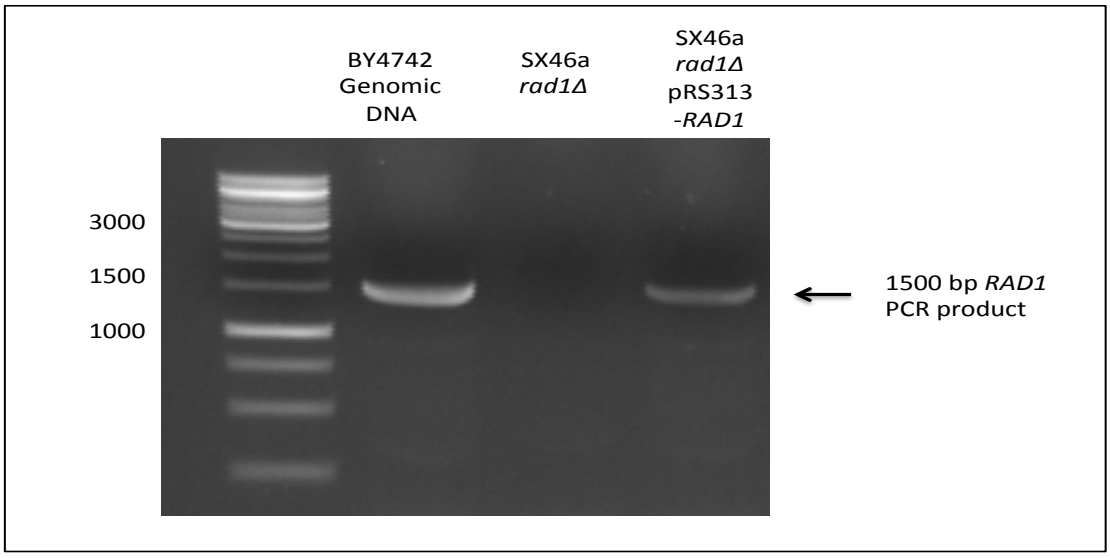


Figure 6.12: Identification of the presence of *RAD1* DNA using colony PCR.
The presence of a 1500 base pair section of *RAD*1 sequence can be identified by PCR of BY4742 genomic DNA control and in DNA from the newly created SX46a *rad1*Δ pRS313-*RAD*1 strain, using the PCR conditions described in chapter 2, section 2.3.10 and the primers Check-F and Check-R (see appendix 9). No PCR product is identifiable in the SX46a *rad1*Δ strain

6.5.3.6 Complementation of UV resistance of *rad1*Δ strain transformed by pRS313-*RAD1*

To demonstrate that the pRS313-*RAD1* plasmid is effective in restoring UV resistance of the SX46a *rad1*Δ strain to wild-type levels, these three strains were assessed using a UV sensitivity assay (Figure 6.13).

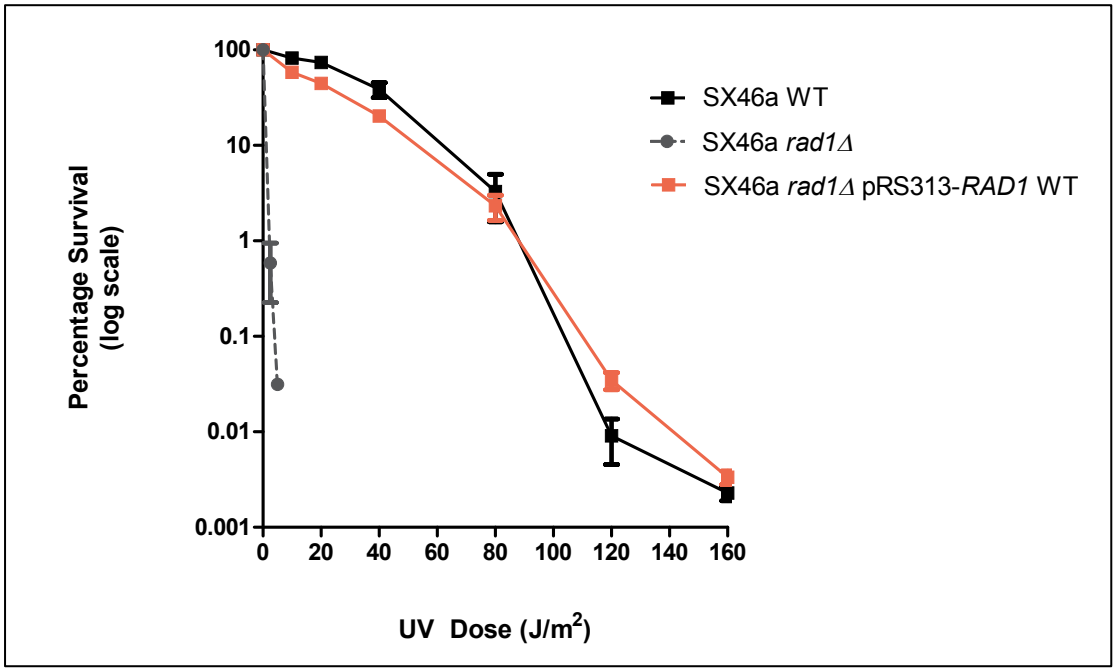


Figure 6.13: UV sensitivity analysis of yeast strains.
The UV sensitivity of SX46a wildtype (WT) (black), SX46a *rad1*Δ (grey) and SX46a *rad1*Δ pRS313-*RAD*1 (orange) is shown in this figure. Colonies of WT and the plasmid strain show similar UV resistance whilst the *rad1*Δ strain is highly UV sensitive.

As can be seen the SX46a *rad1*Δ pRS313-*RAD1* strain UV resistance is complemented to near wildtype levels by the transformation of mutant strain by the plasmid containing *RAD1* sequence and it's promoters, in contrast to the extreme sensitivity of the SX46a *rad1*Δ strain.

6.5.4 Site directed mutagenesis of pRS313-*RAD1* to create the variants Pro469S and Ser747X

Mutation P469S (C1406T in *RAD1*) and S747X: (A2240T and T2242A in *RAD1*) of the pRS313-*RAD1* plasmid, the homologous mutations in *RAD1* to the P379S and S613X mutations in XPF, was introduced using QuikChange Lightening Site Directed Mutagenesis Kit, using the manufacturers protocol (Revision B) and the following mutagenic primers, with the position of mutated bases indicated in red.

P469S-Forward	5'-TTTAGAAGAAAATTCAAAATGGGAACAA-3'
P469S-Reverse	5'-TTGTTCCCATTTTGAATTTTCTTCTAAA-3'

S747X Forward	5'-ATGTACTACGGTGAATGAATTGAAGAGCAAAGTCA-3'
S747X Reverse	5'-TGACTTTGCTCTTCAATTCATTCACCGTAGTACAT-3'

6.5.4.1 Confirmation of successful *RAD1* mutation by sequencing

The initial construct pRS313-*RAD1*, and the two variants generated by site directed mutagenesis were sequenced using BigDye Terminator 3.1 Sequencing Kit (LifeTechnologies). Sequencing reactions were conducted through Cardiff University School of Medicine Central Biotechnology Service.

In addition, to exclude the generation of inadvertent mutations in the plasmid RAD1 and promoter sequence through site directed mutagenesis, sequential primers were used to generate 700bp fragments to fully cover the cloned gene, promoters and downstream regions. The sequence, priming site and orientation of each primer used in the experiment are listed in appendix 9.

Sequencing information was viewed using 4Peaks (http://4peaks.en.softonic.com/mac) and multiple sequences from each primer were assembled using DNA Baser V 3.5.3 (http://www.dnabaser.com), confirming only the desired mutations had been introduced. The data from this sequencing experiment and assembly is included in the electronic appendix 9.

6.5.4.2 Transformation of *E. coli*, plasmid amplification, extraction and purification

DH5α *E.* coli (Invitrogen) were transformed with the pRS313-*RAD1*-P469S and pRS313-RAD1-S747X plasmids following the manufacturers standard protocol, and grown overnight at 37°C in a rotatory incubator at 180 rpm prior to plasmid extraction using PureLink Quick Plasmid MiniPrep Kit (Invitrogen) plasmid using the manufacturers standard protocol.

6.5.4.3 Transformation of SX46a *rad1*Δ by mutant pRS313-*RAD1* plasmids

Following confirmation of appropriate mutation through sequencing of the entire promoter and *RAD1* gene, the yeast strain SX46a Δrad1 was transformed with both mutated plasmids in two experiments using the TRAFO protocol ((Gietz and Woods 2002) and http://home.cc.umanitoba.ca/~gietz/method.html). Selective His- media and plates were used for selection of transformed colonies. Again, the presence of *RAD1* was confirmed by colony PCR using CHECK-F and CHECK-R primers (see 6.5.3.6), and by sequencing of the colony PCR product to ensure the desired mutation was *in situ* and in the correctly labelled strain.

6.5.4.3.1 Confirmation of transformation and of appropriate pRS313-*RAD1* plasmid copy number

Successfully grown colonies were grown in His- media, and DNA was subsequently extracted using a phenol/chloroform extraction protocol (see Chapter 2, Section 2.3.11). PCR of DNA from each strain was conducted using primers for RAD1 and primers for RAD16 as a control to examine for the presence of *RAD1* DNA, using the protocol and primers described in chapter 2, section 2.3.10. The PCR products were resolved by gel electrophoresis (figure 6.14).

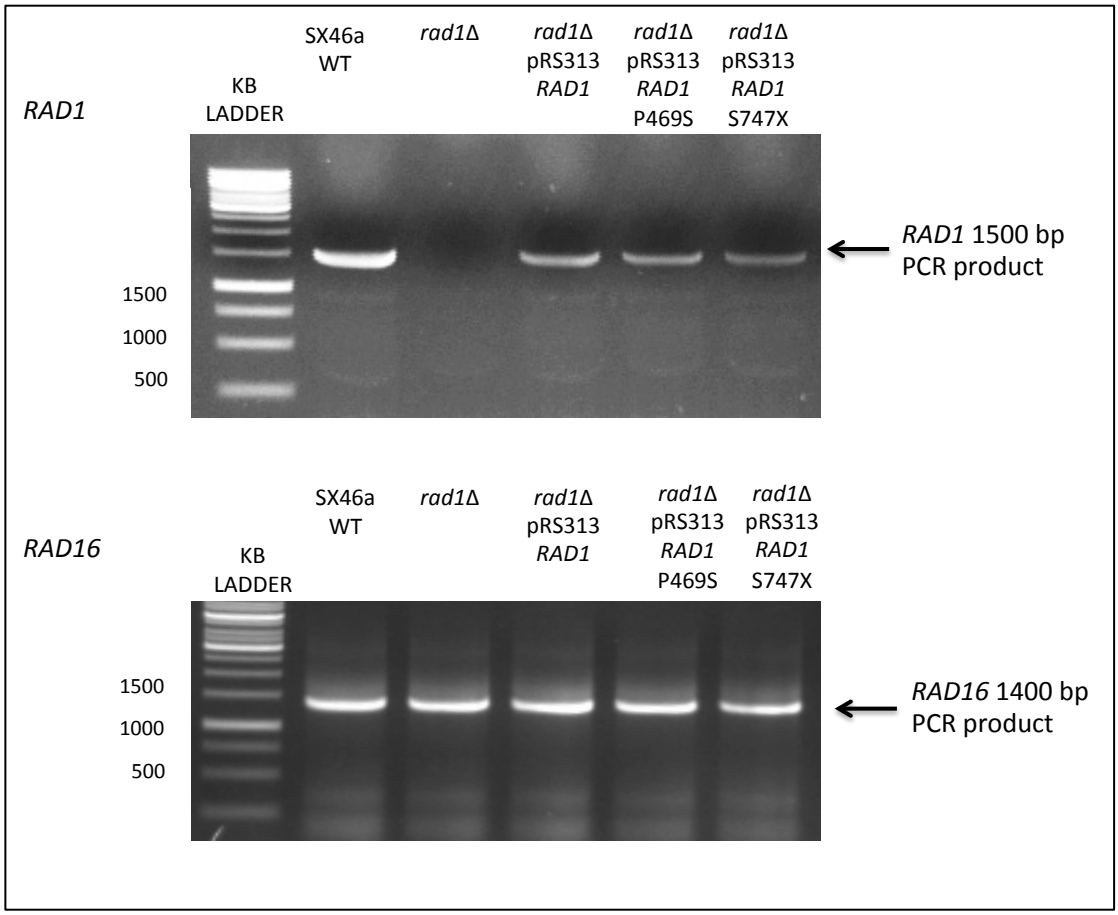


Figure 6.14: Gel electrophoresis of colony PCR products.
DNA extracted from each of 5 strains is subject to PCR with primers for *RAD*1 and *RAD*16 as a control gene. Each strain shows successful PCR in the lower panel. In the upper panel no PCR product is present in the *rad1*Δ strain. The plasmid strains each demonstrate the presence of *RAD1*.

The primers used are as follows:

Check-F	5'- TCAGGTCAATTGACTCGGTGATGG
Check-R	5'-GGCTAATCTGTTATTCTGTAATGACCCA
Rad16-F	5'- TGTTTTTGGCAGACGAAATGGGTATGGGTGCGACCATCCAACTAT
Rad16-R	5'-GCTGAGGAGCCCATTGAATCTAAAGCTCACGCTAAGTTCTGTCGT

This result demonstrates that, compared to *RAD16* reference gene, *RAD1* DNA is now present in the transformed SX46a *rad1*Δ strains, demonstrating the presence the pRS313-*RAD1* plasmid in each of the transformed strains.

The pRS313 series of plasmid vectors are centromeric, and should be maintained at a single copy number per cell following transformation (Clarke and Carbon 1980; Sikorski and Hieter 1989). To confirm this and ensure no differential in *RAD1* gene dosing is present that could affect subsequent results, the relative amount of *RAD1* DNA per strain was quantified using qPCR following the protocol in chapter 2, section 2.3.12, and normalised to the actin gene DNA content in each sample. The results are shown in figure 6.15.

The primers used were as follows:

RAD1 Forward	5'-AAATGCCGCAAACGATTCAA-3'
RAD1 Reverse	5'-CAAATCCTTCACTATGGCCTTA-3'
ACT1 Forward	5'-GTTTGGAATCTGCCGGTATT-3'
ACT1 Reverse	5'-TACCACCGGACATAACGATG-3'

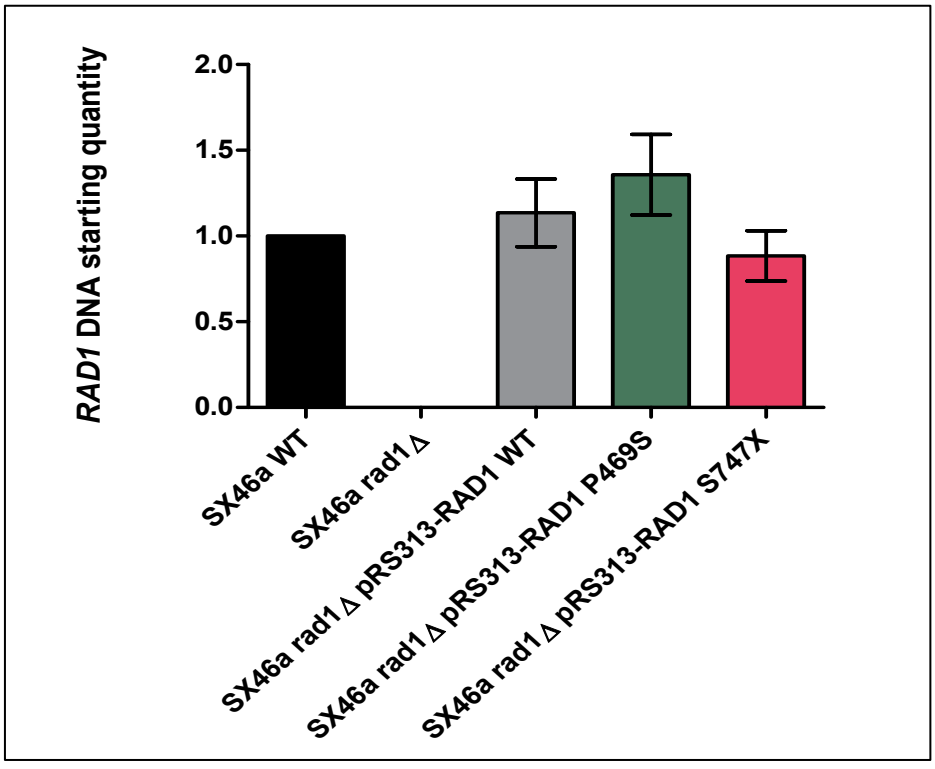


Figure 6.15: The relative amount of *RAD1* DNA quantified by qPCR in each strain used in this study.
The relative amount of *RAD*1 DNA normalised to the amount of actin DNA is shown, and normalised to the SX46a WT *RAD*1 DNA content. The mean and SEM of 5 repeats for the three plasmid containing strains is shown.

These results conform the successful transformation of the SX46a *rad1*Δ strain by each plasmid (the wildtype *RAD1* and both mutated *RAD1* plasmids) and the retention of a single copy of each plasmid per cell. The data also confirms that *RAD1* is undetectable in the pre-transformation SX46a *rad1*Δ strain.

6.5.4.4 Determining the effect of *RAD1* mutations and plasmid transformation on cell growth

To determine the effect of the addition of the plasmid with a wild type or mutated *RAD1* gene on the growth of cells in media, the average growth rate of the strains was measured by spectrometry. To maintain selective pressure to ensure the continued presence of the pRS313-*RAD1* plasmid the transformed strains are required to be grown in selective dropout media. This contains containing all of the necessary nutrients and amino acids required for cell growth except histidine. The SX46a wildtype and SX46a *rad1*Δ strains are unable to grow in histidine deficient selective growth media so are not directly compared with the plasmid containing strains in this experiment.

Briefly, the growth rate was calculated by growing cells to log phase and diluting to an OD_{600} of 0.5, calculated using a spectrophotometer (Thermo Scientific Evolution 600 UV-Vis Spectrophotometer). Subsequent hourly changes in OD_{600} were measured over the following 6 hours in three independent experiments and the mean change in OD_{600} is shown below (Figure 6.16).

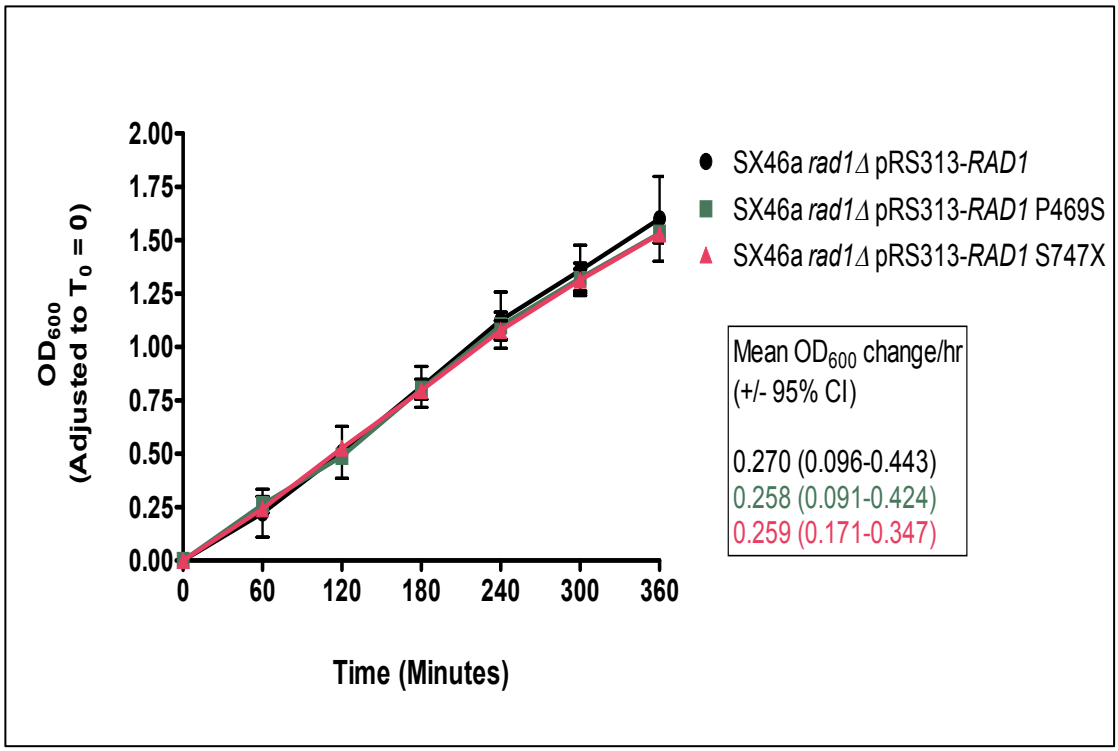


Figure 6.16: **The growth rate of each strain containing a histidine selectable plasmid and grown in his- selective media.**
The OD600 is calculated hourly and the data adjusted to show an increase in OD from a staring value. The mean and 95% confidence intervals calculated from 4 independent repeat experiments are shown. In the right lower inset box the mean hourly increase in OD600 is displayed with the 95% confidence intervals (CI) for each strain, as indicated by colour with reference to the legend.

As can be seen, the growth rate is identical in all three strains, indicating the presence of the plasmid, and the additional *RAD1* mutations, has no significant effect on the health of these strains in standard growth conditions.

6.5.4.5 Summary of experiments to validate the pRS313-*RAD1* model

To complete the validation of the model system constructed in these experiments, ideally protein levels of each plasmid strain would be confirmed by western blotting. However, no specific antibody to RAD1 is available to allow this quantification – only non-specific products are seen with the two commercially available S. *cerevisiae* RAD1 antibodies (data not shown). In the absence of data on RAD1 protein levels, the sequencing and qPCR experiments described above demonstrate that the *RAD1* gene and promoters are present with the same sequence and with one copy of promoter and *RAD1* gene per cell

in each strain. Crucially, the transformation of SX46a *rad1*Δ strain with the pRS313-*RAD1* plasmid results in complementation of UV resistance to near wild-type levels, which is strong evidence that the plasmid model system, as constructed, is robust. The only difference identified in these experiments between each of the three strains containing the pRS313-*RAD1* plasmid is a single nucleotide in the P469S variant and two nucleotides in the Ser747X variant. Experiments to determine the functional effect of these point mutations will be the focus of the following section.

6.6 Determination of UV phenotype of newly created *RAD1* yeast strains

Exposure of cells to ultraviolet irradiation results in the formation of cyclobutane-pyrimidine dimers (CPDs) in DNA, which are repaired by NER. Hence, assessment of colony survival following DNA damage induction with UV radiation can serve as a marker of the strains ability to repair DNA damage, and so act as a measure of NER efficiency.

The UV sensitivity of the five strains was assessed using the method outlined in chapter 2, section 2.4.1. Briefly, a fixed number of cells were spread evenly on agar plates prior to being irradiated with UV-C radiation at a dose of between 1 and 160 J/m^2. After incubation at 30°C in the dark for 72 hours the visible colonies formed on each plate were counted. In an untreated plate the plating efficiency is typically 100% with this type of experiment in *S. cerevisiae,* but to account for any variation in cell counts in biological repeat experiments the results were normalised against the average number of colonies formed on the untreated plate. Each experiment was conducted in triplicate and the mean result was used, and experiments were repeated on three independent biological repeats (figure 6.17).

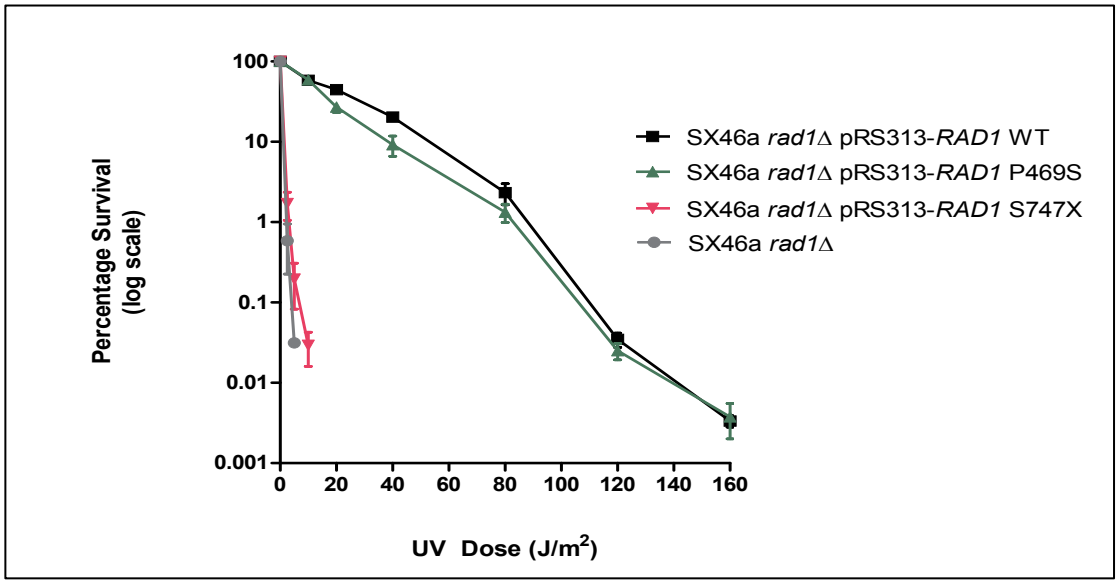


Figure 6.17: The results of UV irradiation colony survival assay.
The average number of visible colonies formed 72 hours after UV irradiation of a standard number of cells per plate, normalised to the average of the untreated plate is shown. All experiments were conducted in triplicate and in four independent biological repeats. The mean and SEM is displayed by the error bars.

The results of this UV survival assay demonstrate slightly lower levels, of survival for each dose of UV irradiation in the mutant *RAD1* strain SX46a *rad1*Δ pRS313-*RAD1*-P469S compared to the *RAD1* wildtype strain SX46a *rad1*Δ pRS313-*RAD1*. This demonstrates that the mutation results in an increase in UV sensitivity, possibly due to impaired NER capacity in this point mutant strain.

The stopgain mutant Ser747X demonstrates extreme UV sensitivity, with identical levels of colony formation following UV irradiation to the SX46a *rad1*Δ strain, indicating that NER is not functional in this mutant strain, despite the presence of the gene and promoters as previously demonstrated by sequencing, colony PCR and qPCR. This result is consistent with the hypothesis that the stop-gain mutant should produce a non-functional RAD1 protein, and is as sensitive, and hence as deficient in NER, as a strain completely lacking the *RAD1* gene.

6.6.1 Measurement of UV induced DNA damage and repair capacity using an immuno-slotblot assay

To demonstrate if a difference in UV induced CPD damage and repair occurs in strains an experiment was conducted using an immuno-slotblot method to measure CPD damage and repair rates. For details of the method see chapter 2, section 2.4.2-2.4.3. Briefly, cells were grown to log phase and centrifuged and re-suspended in PBS. The suspension was irradiated with UV at a dose of $50J/m^{2,}$, chosen as this was the UV radiation dose which resulted in the greatest difference in survival between the wildtype and P469S mutant in the previous study (figure 6.17) . Following irradiation cells were suspended in YPD medium and either harvested immediately or allowed to repair for one or two hours, before cells were chilled to 4°C to prevent further enzymatic activity and DNA was subsequently extracted.

For each experimental condition DNA was transferred to a GeneScreen Plus Hybridisation transfer membrane using Bio-Dot SF Microfiltration equipment. The level of UV damage was measured using an antibody raised against CPDs and probed using an alkaline phosphatase linked secondary

antibody. The binding of secondary antibody was detected with an enhanced chemi-fluorescent (ECF) dye and developed on a Typhoon TRIO Variable Mode imager (Amersham Biosciences). The relative band intensity is calculated using Image J (http://imagej.nih.gov/ij/). An example figure of the immuno-slotblot output (figure 6.18) and the results of UV DNA damage and repair in each strain from the average independent experiments are shown in figure 6.19.

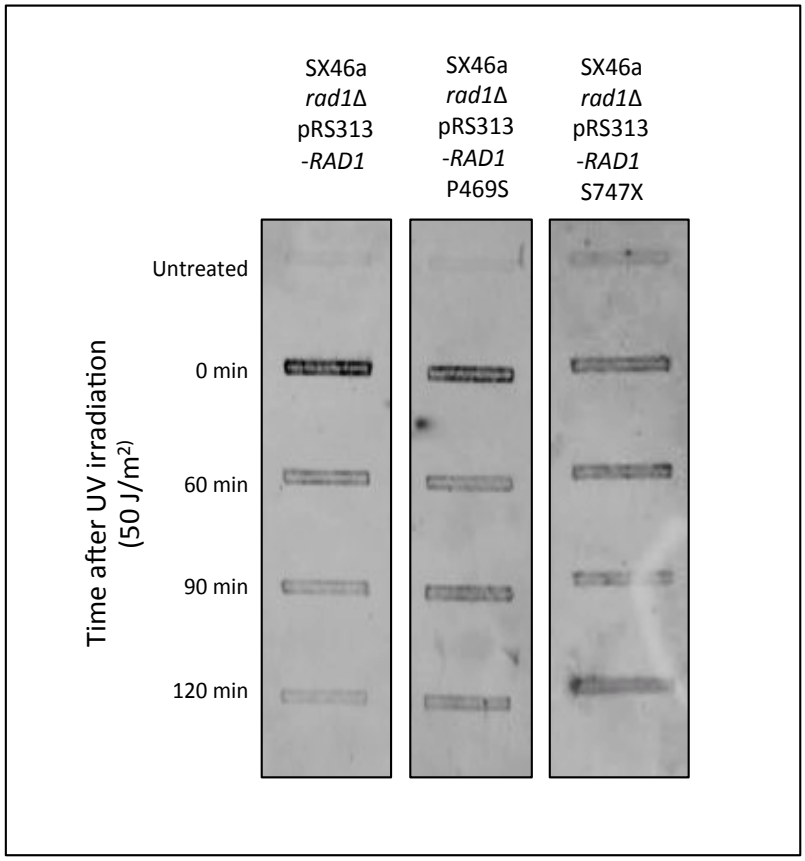


Figure 6.18: **An example of a UV damage and repair immuno-slotblot.**
This image is a composite of 3 separate experiments on 3 different pRS313-*RAD*1 plasmid containing strains, showing the levels of UV damage remaining immediately after exposure (0 minutes) and over a 120 minute recovery period (time 0 to 120 minutes).

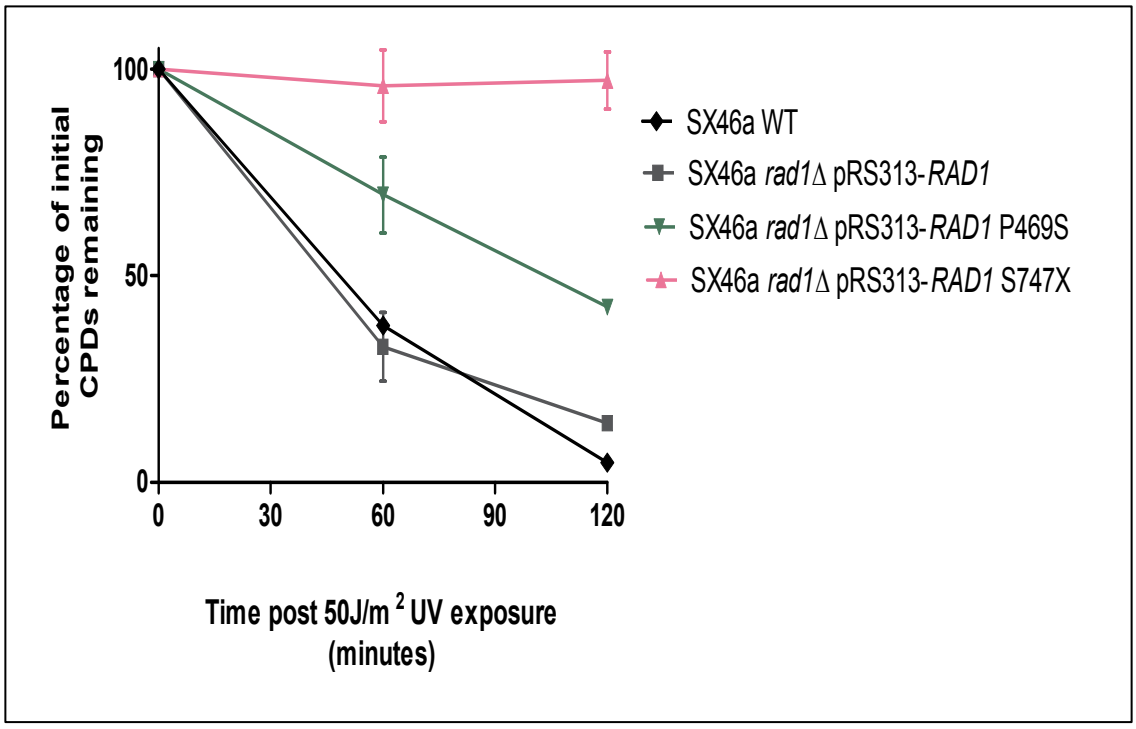


Figure 6.19: Combined result of several UV immuno-slotblot experiments.
Combined result from UV immuno-slotblot experiments on the yeast strains indicated in the legend, demonstrating changes in CPD-DNA adduct levels detectible during a 120-minute period after UV radiation exposure. Data points are from 1-3 independent repeats and the mean and standard error is shown.

The results (figure 6.19) demonstrate a typical immuno-slotblot readout from experiments on each strain as indicated. High levels of damage are present and can be detected immediately after UV exposure (0 minutes). The band intensity decreases with increasing repair time in the left- hand and central panels. At 120 minutes the band is less intense in the left-hand panel, the SX4a *rad1*Δ pRS313-*RAD1* wildtype strain, compared to the central panel showing the SX4a *rad1*Δ pRS313-*RAD1*-P469S strain. This is evidence that the P469S point mutation reduces the NER capacity of the strain. The right-hand panel shows the S747X mutant strain and the band intensity is similar at all time points post UV exposure, indicating that no CPD repair is occurring in this strain over the period studied in this example. This is evidence of a complete loss of NER capacity in this mutant strain.

This experiment was conducted in a number of independent biological repeats (for raw datasets see the accompanying electronic appendix). The result of each experiment is combined to show the mean (and SEM) percentage of CPDs remaining at each time point for the strains in figure 6.20, normalised to the CPD

adducts present immediately after exposure. Clearly, the stop-gain mutant fails to repair DNA damage over the time course of the experiment, and the rate of DNA repair is lower in the P469S mutant than the wildtype strains. The SX46a (black) and SX46a *rad1*Δ pRS313-*RAD1* wildtype (grey) strains show identical levels of CPD repair, which is further evidence that the model plasmid system constructed here is valid.

These results are consistent with the UV survival result (Figure 6.17) and together these experiments demonstrate that the P469S mutant has impaired levels of NER compared the wild-type RAD1 strains, resulting in impaired survival after UV treatment. The stopgain S747X mutant shows no significant DNA repair after two hours, corresponding to the extreme UV sensitivity phenotype demonstrated by the survival assay.

These results reflect the UV induced of NER capacity of the strains examined. The next phase of this study is to test the strains following oxaliplatin treatment to assess the effect of this drug on NER.

6.7 Determination of oxaliplatin phenotype of newly created *RAD1* yeast strains

In the previous section, 6.6, the UV treatment of yeast cells was easily accomplished using a brief exposure to UV light and returning the cells to growth media or plates to measure the effect on colony forming ability of the strains. A short exposure to UV radiation is sufficient to rapidly induce DNA adducts. The treatment with chemical agents is more complex for several reasons. Chemical agents are used in solution, and the effect of the solution on the cell phenotype and on the drug efficacy must be determined. Oxaliplatin also forms di-adducts over many hours after drug exposure, so for a period after treatment both the formation and removal of adducts will be occurring simultaneously. This complicates the measurement of adduct levels, and data generated during this ill-defined period is more difficult to interpret compared to the measurement of UV induced DNA adducts.

In order to determine the conditions for treating yeast with oxaliplatin experiments to establish treatment parameters that result differential sensitivity of the strains to oxaliplatin were conducted. In the first instance, this was by measuring the effect of adding different doses of drug to the growth of cells in standard histidine deficient selective yeast culture media. Logarithmically growing cells growing at the same rate (a doubling time of 1 hour 50 minutes) were diluted to OD_{600} 0.8 and concentrated oxaliplatin was added at several dose levels. The OD_{600} was measured hourly in each strain over the following 8 hours. The effects of the addition of oxaliplatin to the growth of SX46a *rad1*Δ pRS313-RAD1, and the P469S and S747X mutants can be seen in figure 6.20.

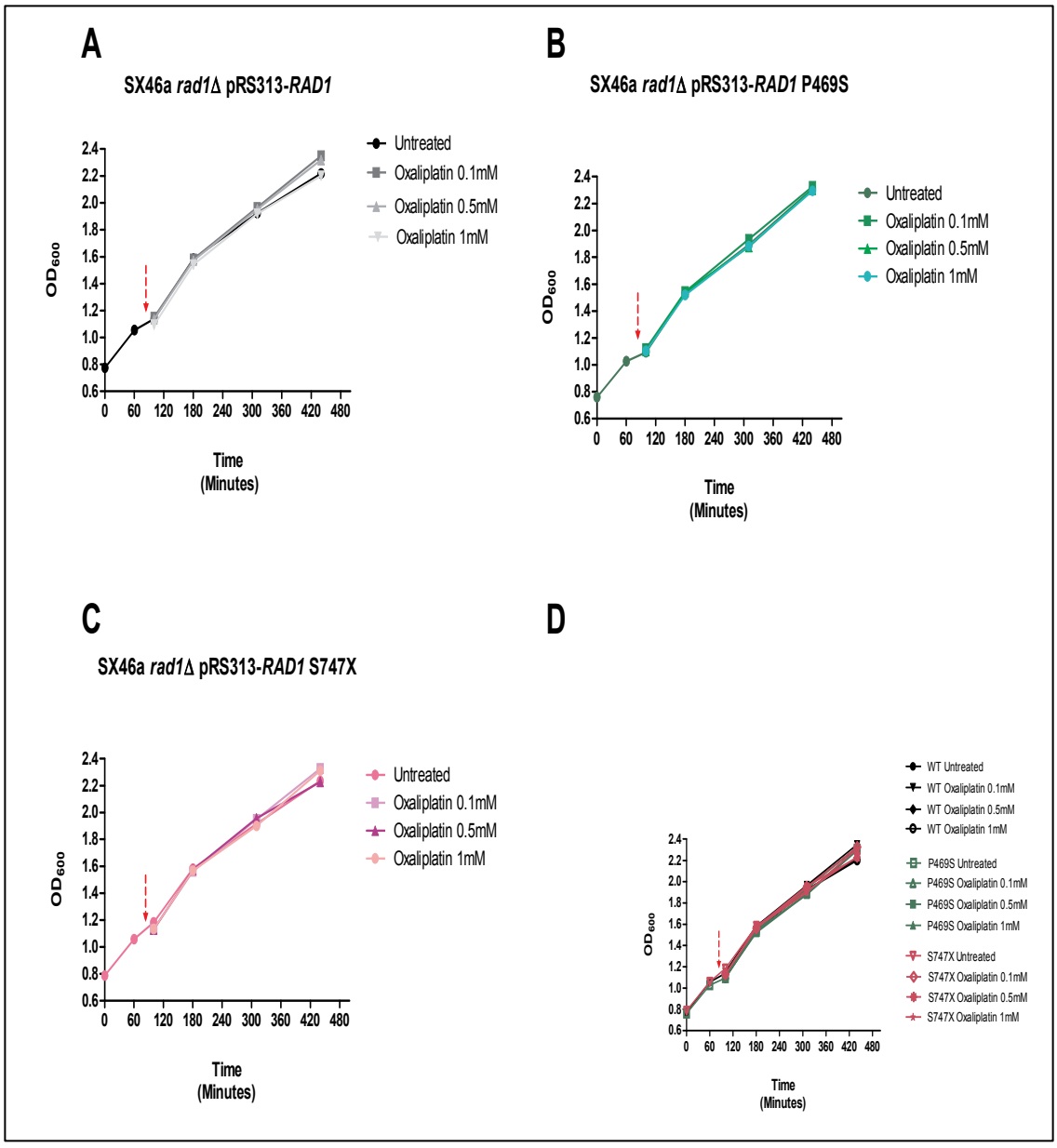


Figure 6.20: The effect of the addition of oxaliplatin on yeast growth rates.
This figure shows the effect of the addition of oxaliplatin to a final concentration of 100μM, 500μM and 1mM to cells growing for 90 minutes in selective histidine deficient growth media. The addition of the drug indicated by a red arrow. Each strain is shown in a separate figure (A-C) and all strains are displayed in figure D to allow comparison

These results indicate that the doses of oxaliplatin used in this experiment have no effect on the rate of growth in the 6 hours following administration, under these treatment conditions. It may be that the media neutralises the oxaliplatin; plasma proteins bind oxaliplatin irreversibly when administered to patients (as discussed in chapter 1, section 1.2.1), and the protein rich media used for yeast growth may sequester some of the oxaliplatin,

making the effective dose lower. It also may take several hours from the addition of oxaliplatin to the growth media for oxaliplatin-DNA mono-adducts and subsequent di-adducts to form, and for the cell to mount a DNA damage response. It could also be that the doses used are insufficient, although the strains in this experiment are exposed to oxaliplatin doses that are 30-300 fold more than are used during chemotherapy in humans.

6.7.1 Modification of treatment conditions: Incubation of cells with oxaliplatin in PBS

The proteins-rich media use in the previous experiment may sequester the oxaliplatin. This may result in a lower exposure of cells to drug and may result in unwanted variability between experimental conditions. To reduce this potential effect and to ensure adequate chloride concentration for the formation of reactive oxaliplatin intermediates cells were resuspended in phosphate buffered saline (PBS) prior to the addition of oxaliplatin.

Briefly, cells were grown to the log phase in His- minimal media until growing at a constant OD_{600} increase per hour and at a constant doubling time of 1 hour 50 minutes. The cells were washed twice in PBS and incubated at 30°C and 180 rpm whilst suspended in PBS solution. Oxaliplatin was added to the samples to the desired final concentration. After three hours incubation the cells were washed twice and re-suspended in histidine deficient selective media before incubation at 30°C at 180rpm. The OD_{600} was calculated and adjusted to 1 at the start of incubation, and all further hourly measurements for the next 12 hours are normalised to the initial OD measurement. At OD_{600} of 2 the cells were diluted back to OD_{600} of 1 and future results normalised to take account of this adjustment. The same results are reported twice, firstly grouped by strain to show the effect of increasing doses of drug (figure 6.21), and, for easier comparison between strains, secondly grouped in relation to dose (figure 6.22).

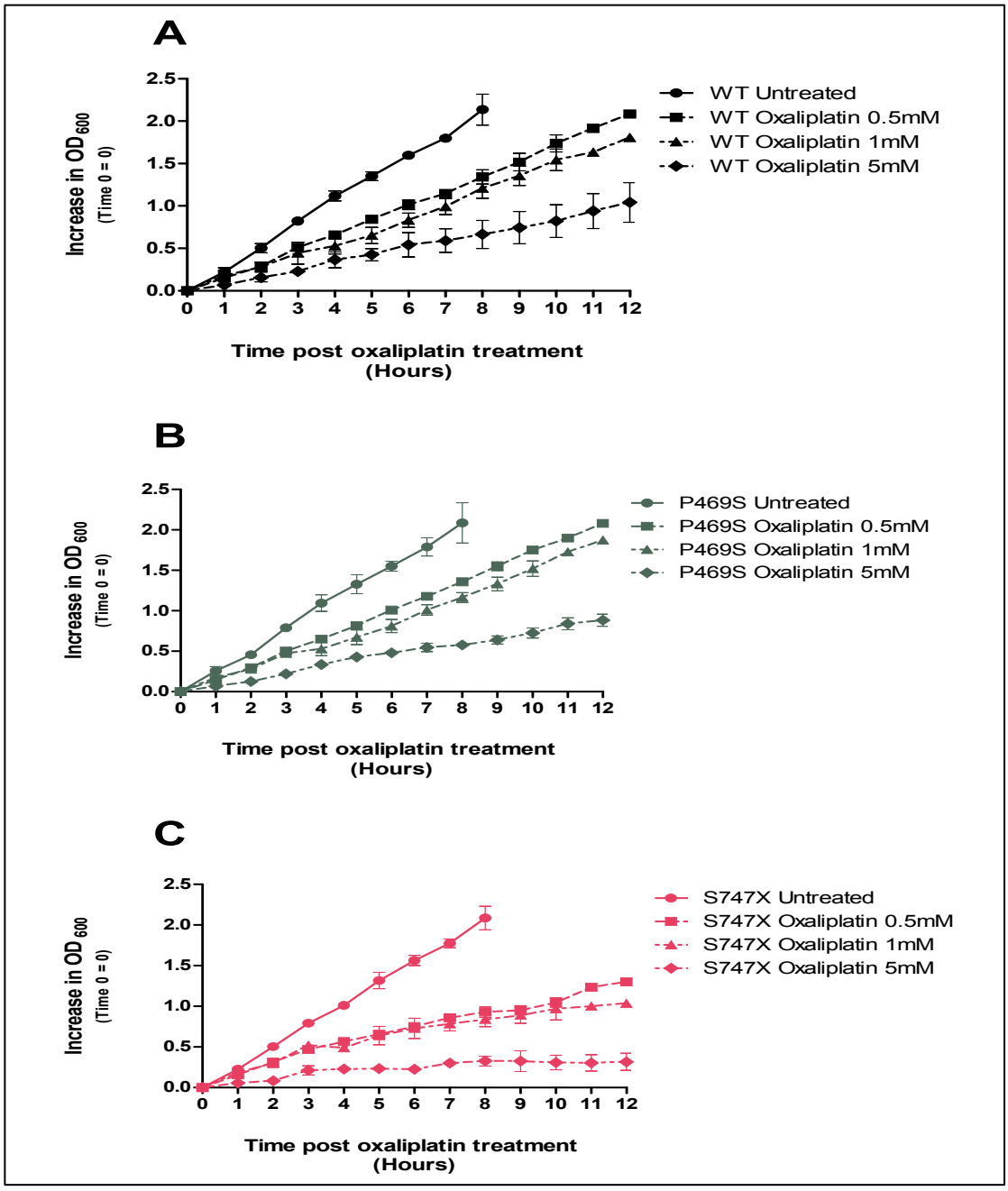


Figure 6.21: The effect of prior oxaliplatin/PBS incubation on yeast strain growth rates. Each strain is exposed to increasing doses of oxaliplatin suspended in PBS for 3 hours before washing and incubation in growth media. The increase in OD600 over the 12 hours following drug exposure is displayed in each panel (A to C), normalised to the initial staring OD600 and adjusted to 0 at t=0. The mean of three independent biological repeats and the standard error is shown.

Figure 6.21 shows the effect of oxaliplatin on each of the three plasmid-containing strains. As can be seen in the results, each untreated strain grows at a constant rate of 0.28 OD_{600} per hour (a doubling time of 1 hour 50 minutes), indicating that the incubation for three hours in PBS has had no adverse effect on

growth characteristics. In each of the three strains the effect on growth of the addition of increasing doses of oxaliplatin can be seen. A sequential decrease in growth rate with increasing exposure to oxaliplatin occurs and is more marked in the S747X mutant than the wildtype and P493S mutant.

To allow and easier visual comparison of the effect of oxaliplatin exposure between the three strains the same result is displayed plotted by dose (Figure 6.22).

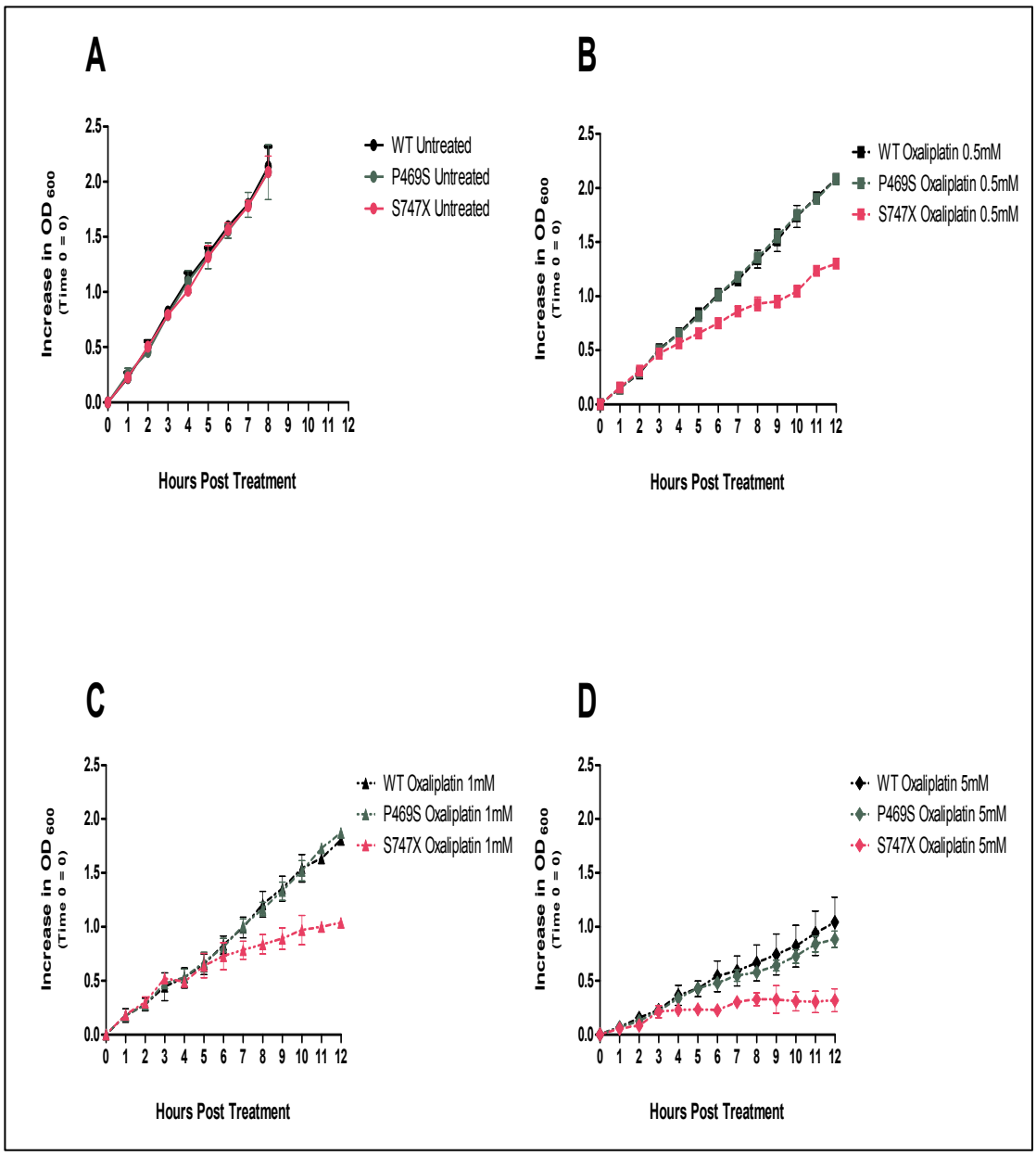


Figure 6.22: The results from figure 6.21 displayed by dose rather than by strain.
This is to allow comparison of the effect of the addition of oxaliplatin between strains. Each panel (A to D) displayed the effect on subsequent growth of an increased exposure of the strains to oxaliplatin.

At each dose level (A to D) the relative effect of oxaliplatin on the growth between each strain is evident. In the untreated samples (A) the growth rate is identical at a 0.28 increase in OD_{600} per hour, corresponding to a doubling time of 1 hour 50 minutes. This indicates that a 3-hour incubation in PBS does not have a detrimental effect on cell growth rate – when retuned to growth media the cells immediately begin to grow at the rate at which they were growing before the PBS incubation.

With exposure to increasing doses of oxaliplatin the inhibition of future cell growth is identical between the wildtype and P469S mutant plasmid containing strains, although at the 5mM dose level a slight trend towards a lower growth rate is seen in the P469S mutant strain. The effect of oxaliplatin on the stopgain S747X mutant at each dose level is much greater, with no evidence of an increase in cell growth at 5mM in the 12 hours following oxaliplatin exposure.

This result, using these specific oxaliplatin treatment parameters, clearly demonstrates an effect of oxaliplatin exposure on the three plasmid-containing strains, and that this effect is greater in the S747X variant than the wildtype or P469S variant. It is, however, not clear from this result exactly what this data represents. The assay measures the change in optical density of 1ml volume sample of yeast cells, a robust and widely used proxy for the number of cells present in the sample (Myers et al. 2013). Over time after oxaliplatin exposure the optical density increases, and increases at a slower rate with higher doses of drug. This could reflect the action of oxaliplatin on all of the cells in the media, for example an inhibitory effect on cell division following platinum-DNA adduct formation. It could also reflect the survival of a specific proportion of cells after drug exposure, and this population goes on to result in the detected increases in optical density following oxaliplatin exposure.

The treatment parameters established here result in detectible differences between strain phenotypes and will therefore be used in an alternative assay to measure cell survival following oxaliplatin exposure.

6.7.2 Measuring survival following exposure of plasmid-containing yeast strains to oxaliplatin

An alternative survival assay was conducted using a similar method to the UV survival assay outlined above (6.5.1) and using the method described in chapter 2, section 2.4.6. Briefly, cells growing in log phase were washed, and subsequently incubated in PBS containing increasing doses of oxaliplatin for 3 hours at 30°C with rotation at 180 rpm. Following treatment, cells were washed twice and re-suspended in PBS, and serially diluted. 200 cells per plate were plated in triplicate onto agar plates and were incubated for 72 hours at 30°C. The number of visible colonies on each plate was counted and normalised to the average number of cells on the untreated plates. The experiment was conducted on 3 independent occasions and the results are shown below (Figure 6.23).

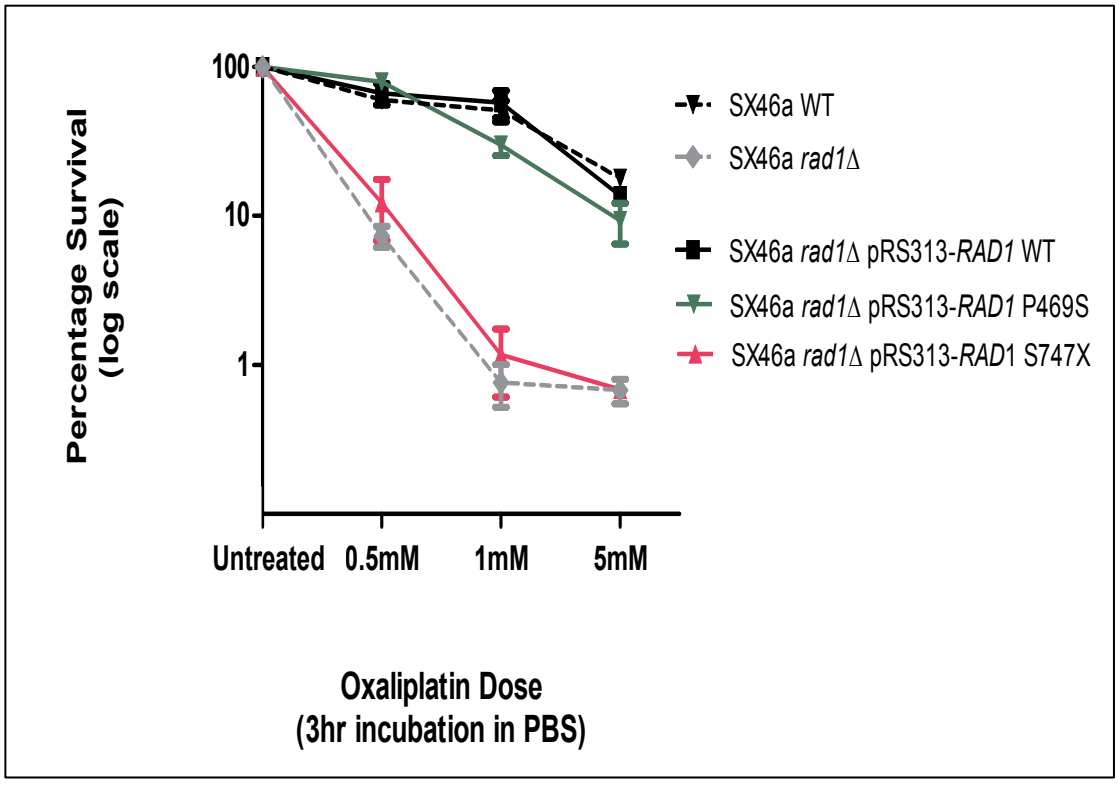


Figure 6.23: An oxaliplatin survival assay.
Colonies formed 72 hours following oxaliplatin exposure were counted and normalised to the average of the untreated plate. Experiments were conducted in triplicate and on 3 or 4 independent occasions. The mean and SEM is shown.

The results of this experiment in figure 6.24 demonstrate a difference in survival following oxaliplatin exposure in all 3 plasmid-containing strains. Firstly, the survival in the SX46a wildtype and SX46a *rad1*Δ pRS313-*RAD1* wildtype is the same, again supporting the validity of the model.

Following exposure to oxaliplatin the P469S point mutation results in a decrease in survival compared to wildtype, and the S747X variant results in a much lower survival, similar to the SX46a *rad1*Δ strain with complete absence of the *RAD1* gene.

6.8 Discussion and Conclusions

Oxaliplatin induced peripheral neuropathy is a severe side effect of oxaliplatin chemotherapy treatment that can result in long-term numbness and painful paraesthesia in patients often cured of their malignant disease. At moderate doses the development of OIPN is idiosyncratic, and to-date there is no way of stratifying patients for the risk of developing this toxicity. It is made more of an issue by the lack of any effective treatment and by the persistence or, at best, slow resolution of symptoms over many years (Argyriou et al. 2014). When oxaliplatin is used in the adjuvant setting, in over 10,000, patients per year in the UK, the benefit is in the order of 5% to 10% improvement in cure rate (i.e. a number-needed-to-treat of 10 to 20) so the majority of patients at risk of OIPN gain no benefit from the addition of oxaliplatin treatment to surgery and fluoropyrimidine adjuvant chemotherapy (Andre et al. 2009).

The need to find accurate biomarkers to stratify patients for the risk of OIPN is urgent. So far, many genomic studies have identified statistically significant potential genetic markers of risk, but these studies are hampered by inconsistent and conflicting results, and a lack of mechanistic or functional evidence to support their conclusions (Cavaletti et al. 2011; Custodio et al. 2014).

A recent study by Professor J Cheadle at Cardiff University using patient samples from the UK national COIN trial (Maughan et al. 2011) has identified several potential biomarkers markers of OIPN risk by studying DNA

polymorphisms in the DNA repair gene *XPF*. The aim of this study is to provide supporting functional and mechanistic information in a robust model system to support these findings.

The results of the experiments conducted and described in this chapter outline the steps taken to develop a *S. cerevisiae* model to determine whether two of the SNPs found in the *XPF* gene of patients who have developed OIPN could be mimicked and have a mechanistic effect on DNA repair and on sensitivity to oxaliplatin DNA damage.

The evidence presented demonstrates the effect of these genetic variants on DNA repair and survival following DNA damage. The sequencing confirmation after site-directed mutagenesis, demonstration of appropriate copy number of plasmids, and the identical growth rate of the constructed strains indicate that, as a baseline, these strains have similar characteristics and phenotype with respect to these parameters.

After UV treatment the model demonstrates that the mutations in *RAD1* do result in a functional defect in NER through the survival assay results and immuno-slotblot assessment of NER capacity.

The *RAD1* stopgain variant Ser747X results in complete failure to repair (UV induced) CPD or photo products, resulting in an extreme UV sensitive phenotype. When exposed to oxaliplatin this strain also shows a highly sensitive phenotype, manifested in significantly reduced colony survival and highly impaired growth in the 12 hours after drug exposure.

The stain SX46a *rad1*Δ pRS313-*RAD1*-P469S also demonstrates a significant difference when compared to the SX46a wildtype and the SX46a *rad1*Δ pRS313-*RAD1* complemented strain. The UV sensitivity shown in the UV survival assay is increased, although this effect is subtle. The reduction in survival after UV exposure is demonstrated to be due to the lower rate of repair of UV induced CPD adducts when compared to the wildtype strains.

The sensitivity to oxaliplatin of this strain compared to the wildtype is statistically significant, but subtle. When the effect of the mutation on the growth rate following oxaliplatin exposure is evaluated there is little difference at lower doses when compared to the wildtype strains although at a high dose

there appears to be a trend towards slower growth rates in the P469S mutant. When exposed to increasing doses of oxaliplatin the P469S mutant strain displays a significant but small decrease in survival, consistent with the effect of impairment in DNA repair previously demonstrated.

The findings of these experiments indicate that modelling the XPF variants leads to a functional defect in cell growth and repair after DNA damage. In theory, a stopgain variant S613X in XPF should result in the complete loss of function of the protein from this allele, with translation of mRNA terminated prematurely and resulting in a truncated protein missing the C-terminal nuclease and ERCC1 binding domains. A complete loss of function is demonstrated in our study by showing a phenotype similar to that of a total *RAD1* deletion strain. In a patient with this mutation the effect on XPF protein level may not translate to an XP phenotype if the heterozygous allele produces adequate XPF protein to maintain a certain NER capacity. In keeping with XP as an autosomal recessive disease, in a series of patients with XP due to mutations in XPF described by (Ahmad et al. 2010) none of the patients with the disease had a wildtype *XPF* allele – all had mutations in both alleles. The father of one of the patients, a carrier with one wildtype allele and one with a variant (R799W), showed no clinical evidence of XP, normal functioning unscheduled DNA synthesis and no evidence of an increase in UV sensitivity. Contrary to this, the fact that no mutation could be found on the other *XPF* allele in our patient with the S613X variant indicates that this is a potential example of haplo-insufficiency. However, the patient did not clinically have XP. In the diploid heterozygous context it may be that the S613X mutation results in lower levels of XPF protein, only manifesting the OIPN phenotype under the severe DNA damage induced by chemotherapy. However, only a limited number of 104 genes were examined for the presence of genetic variants (see section 6.2.1) and an alternative explanation is the presence of mutations elsewhere in the chromosome or genome that contribute to this patient's phenotype, but that have been filtered out, and hence not detected, by the type of pathway-based approach used in this investigation.

The variant P379S has been documented in patients with mild XP (Ahmad et al. 2010; Berneburg et al. 2000) and has been hypothesised to result in protein misfolding and loss of nuclear localisation signalling (Ahmad et al. 2010). The consequence appear to be changes to the cellular allocation of XPF, with a reduction in nuclear localisation of XPF from 93% in controls to 73% in a patient with single allele mutated to P379S, and a resulting increase in non-nuclear localisation (using immunofluorescence) from 7% to 27% (Ahmad et al. 2010). A reduction in availability in the nucleus of essential NER factors could potentially be detected in our model as a reduction in NER activity, although this would likely be subtle in the P379S strain than in the stop-gain mutant. The subtle decrease in NER capacity demonstrated in our haploid *S. cerevisiae* model may be absent in the heterozygote diploid human context, but again may be enough to result in OIPN when exposed to chemotherapy. Unfortunately, as no effective RAD1 antibody is available, it is not possible to probe for differences in RAD1 cellular location in our *S. cerevisiae* model.

A limitation of the work as described is the inability to demonstrate that functional RAD1 is produced in each strain in similar amounts, as it is not technically possible with the commercially available RAD1 antibodies to probe for this protein. The DNA quantification qPCR experiment suggests strongly that only a single copy of each plasmid is present per cell, and as the promoters and expression of the protein in the plasmids used in the SX46a *rad1Δ* strain transformation are identical (as confirmed by sequencing) it is likely that there is no difference between strains.

In conclusion, the experiments outlined in this chapter demonstrate robust evidence of a functional link between the presence of SNPs identified in the *XPF* gene of patients who have developed severe OIPN and NER capacity, and on the sensitivity to oxaliplatin. These results indicate that these SNPs are functionally linked to the development of OIPN and may be good candidate biomarkers for future genetic pre-chemotherapy screening. This work also lends further weight to the hypothesis that functional measurements of DNA repair capacity are a useful tool that could potentially be used to stratify patients for the response and toxicity to platinum chemotherapy agents; an idea explored in all of the chapters of this thesis.

Chapter 7 Discussion, future directions and thesis conclusion

7.1 Thesis summary

In summary, the key findings and developments outlined in this thesis are as follows:

- The translation of a DIP-chip assay from use in cell culture into an assay capable of detecting platinum-DNA adduct patterns in clinical samples, requiring the following steps:

    - The development of standardised and robust treatment protocols for the *ex vivo* treatment of PBMC from clinical samples.
    - The optimisation of assay experimental protocols to ensure and demonstrate reproducible DNA immunoprecipitation.
    - The optimisation of protocols for the production of reproducible genome scale microarray chip platinum-DNA patterns from DIP samples.

- The development of enhanced DIP-chip bioinformatic analysis tools for:

    - The extraction of latent information from DIP-chip microarrays: used as a tool for enhanced QC of DIP-chip data and for detailed analysis and interpretation of assay-development experiments.
    - For determining reproducibly of microarray data along the genome – used to determine the effect of assay protocol modifications and for the identification of regions of DIP-chip microarray data that is highly variable, and which can be excluded from inter-individual analysis
    - Determining, plotting and listing regions of significant differences between inter-individual DIP-chip platinum-DNA adduct patterns.

- An experimental study confirming and supporting the importance of the link between germline DNA repair gene polymorphisms and platinum toxicity experienced by patients during platinum-based chemotherapy.

7.2 Discussion

This thesis primarily concerns the development of methods and tools that can be used for stratification of patients for response and toxicity to platinum-based chemotherapy, with an emphasis on oxaliplatin induced peripheral neuropathy in the context of the treatment of colorectal cancer.

As discussed extensively in the introductory section of chapters 3, 4 and 6, robust evidence exists that several measurable aspects of DNA repair capacity are important determinants of the response and toxicity that occurs in patients treated with platinum agents. Many studies that have been conducted to examine this relationship have been cited in these chapter introductions, and several of these studies have determined a correlation between these factors. As described in chapter 3, these studies can be grouped by the different approaches taken to investigate this concept, including grouping into 'genetic studies' - studies examining for the presence or absence of single nucleotide polymorphisms in repair genes, into studies that use measurements of germline DNA repair capacity, studies examining levels of expressed DNA repair proteins, and studies measuring the levels of DNA adducts in samples from patients previously treated with platinum agents.

On the strength of these results, the concept of measuring features of the DNA repair pathway as a potential determinant of platinum response and toxicity has been adopted by the clinical community, and has progressed to use in large-scale clinical studies to stratify patients for the response and toxicity of platinum agents (for reviews see (Cavaletti et al. 2011) and (Bowden 2014)). As discussed, the currently available approaches to measure platinum-DNA adducts when applied to clinical studies can be characterised, in general, as studies utilising a single measurement of a single feature of DNA repair capacity, either at the genetic level by assessing the presence or absence of polymorphisms, at

the protein levels by measuring the presence or absence of a single DNA repair protein (such as ERCC1), or at the functional level by determining the total level of platinum-DNA adducts per unit of DNA. To date, however, the widespread adoption of these types of stratification tools has been limited by the lack of predictive power that has been demonstrated in clinical studies using these approaches (Bowden 2014). However, the general concept of these pre-clinical and clinical studies is support for the rationale of the experiments and approaches explored and described this thesis.

Whilst there are several limitations to the current technique and aspects of the DIP-chip assay to be improved, once optimised and refined it is our contention that the DIP-chip assay, as used extensively throughout this thesis, has three key advantages compared to the currently available approaches to measure platinum-DNA adducts described above.

Firstly, a DIP-chip assay is a functional assay, and measures platinum-DNA adducts directly, rather than measuring surrogate markers (such a repair gene SNPs) or measuring markers several steps removed from the final common endpoint of DNA damage, such as protein expression levels. As platinum-DNA adducts are the direct insult leading to the DNA damage response, apoptosis or necrosis, and ultimately govern the response and toxicity of these agents, measurement directly of adducts rather than 'upstream' and surrogate markers is less subject to potential bias and confounding, and is therefore more likely to be a reliable marker for determining platinum response and toxicity.

One significant limitation of the assay is the requirement to use surrogate tissue of PBMC instead of tumour tissue. A PBMC-based assay may be accurate in the assessment of germline DNA damage and repair as a measure of toxicity of chemotherapy on normal tissues, rather than when assessing the heterogeneous tumour repair profile for tumour response, and may also have a use in the assessment of response of normal tissue with similar post-differentiated DNA repair patterns such as neuronal tissue in the assessment of OIPN. Despite the limitations, PBMC are commonly used for this type of assay (see section 3.1.7) as a surrogate tissue and reduce the need for potential hazardous biopsies in a research study. It may be possible to determine the effect of similarities and

differences between platinum-induced DNA adduct profiles in PBMC and paired tumour tissue in future research projects using the DIP-chip assay.

A second strength of the DIP- chip assay is the ability to measure the level of induced DNA adducts at multiple genetic loci simultaneously. As demonstrated previously in our laboratory in fibroblast cell culture models (Powell 2014) and in many occasions in the experiments and results described in chapter 4 and 5, platinum-DNA adducts form at different levels at different regions of the genome. Rather than the single average measures of DNA damage or repair used in previous clinical studies (Bowden 2014; Wang et al. 2011), the ability to determine the level of platinum-DNA adducts at high resolution and a genomic scale gives in the DIP-chip assay an important advantage for detecting and measuring the distribution of platinum-DNA adducts. Additionally, as it is recognised that NER occurs at different rates in different section of the genome in different tissues, especially neuronal tissue, a key determinant of peripheral neuropathy, and in PBMC, the surrogate tissue used in these experiments (Nouspikel 2009), the DIP-chip assay is particularly appropriate for investigating oxaliplatin induced peripheral neuropathy.

A third key strength of the DIP-chip assay is the genomic-scale approach and genomic-scale datasets generated by this technology, in comparison to the data generated from a single measure assay. Rather than a single output (for example high or low ERCC1 protein expression, or high or low average adduct levels per unit of DNA) the genomic approach allows a non-biased and un-filtered ability to determine differences in patterns through the representative section of the genome examined. Single or limited measure approaches filter information by concentrating on one particular marker and ignoring information in other genetic regions or in other proteins, reducing the power, applicability and predictive ability of these types of assay.

There are significant limitations of the DIP-chip assay to overcome. As discussed above, the use of PBMC as a surrogate tissue is a potential source of inaccuracy. The use of PBMC is a common approach in this type of assay, and reduces the need for repeat tissue biopsy, especially during assay development (see section 3.1.7). One future project under consideration is to induce platinum DNA damage in paired tumour and PBMC tissue from the same individuals to

compare the DIP-chip patterns generated to determine similarities and differences between the induced DNA damage patterns in these different tissue types. This may resolve or quantify the impact of this potential limitation of the PBMC DIP-chip assay.

The accuracy and applicability of the DIP-chip assay is also limited by the poor reproducibility of the assay, and the resolution of this issue is a primary focus of chapters 3, 4 and 5. To reliably compare inter-individual differences the assay must be capable of generating relatively reproducible DNA damage patterns in repeated samples from the same individual. This has partially been addressed through modifying the DIP-chip protocol, and by using this approach the experimental variability has been demonstrably reduced. Additionally, by filtering the datasets to identify only statistically significant differences between datasets, primarily by using the bioinformatic tools developed in chapter 5, the impact of this variability can also potentially be reduced. The choice of statistical method to achieve this filtering requires refinement, and the outlier detection technique used in chapter 5 has several limitations, including the identification of outliers which may result from experimental error rather than biological signal and the loss of genomic and local probe context by focusing on individual probes comparisons between datasets. Further developments in the bioinformatic analysis of datasets discussed in section 5.6.2, particularly the use of a sliding window t-test, may resolve some of these issues.

If the limitations of the assay can be overcome, by combining a functional measure of platinum-DNA damage with a genome-scale, un-biased and non-filtered approach to measuring patterns of adduct levels, and using genomic scale data and bioinformatic tools to reveal differences in patterns between individuals, the DIP-chip assay has many potential advantages compared to the previously available technology used in this field of research.

In this context, the key developments described in this thesis is the translation of cell culture model analysis DNA immunoprecipitation and microarray (DIP-chip) techniques into a robust method for measuring the distribution of induced oxaliplatin-DNA adducts at high resolution and at a

genomic scale in human clinical samples. This thesis also includes the development of the necessary bioinformatic analysis tools for the quality control and the analysis of inter-individual differences in DIP-chip generated platinum-DNA adduct patterns. This was attained over the course of experiments and developments outlined and described over three chapters. Chapter 3 described the experimental development and optimisation of DNA immunoprecipitation, the first stage of the DIP chip assay, for use in human clinical samples. Chapter 4 dealt with issues that arise from the translation of thus technique into the 'chip' genomic scale analysis phase of the assay. For the assay to function properly, for further development work, for quality control, and to ultimately detect differences in platinum-DNA adduct signatures the bioinformatic tools required were developed in chapter 5.

Additionally, this thesis includes a functional study that demonstrates the influence that subtle perturbations in DNA repair capacity can have on the outcome of chemotherapy treatment in colorectal cancer patients. This aspect of the study provides functional evidence to support the initial genetic evidence – a study that identified a statistical link between DNA repair gene polymorphisms and oxaliplatin toxicity (West 2013). The findings of this chapter are support for the central hypothesis of this thesis; that by measuring changes in DNA repair capacity, and specifically in the level and patterns of induced platinum-DNA adducts, it may be possible to stratify patients for the response and toxicity that occurs during treatment with platinum based chemotherapy. The three *S. cerevisiae* models developed during this study are of a normal, moderate and extreme oxaliplatin-induced peripheral neuropathy (OIPN) phenotype, and will have an important role in the further development and validation of DIP-chip based assays of OIPN in future projects in our laboratory.

7.3 Pathway for clinical development

The DIP-chip assay protocols and bioinformatic analysis tools presented here, despite the limitations and need for refinements discussed above, represent a clinically usable assay, from initial venesection of 10-20mls of fresh

blood to the generation and analysis of the pattern of induced platinum-DNA adducts at high resolution over a 5Mb section of the genome. The analysis tools developed here, within the limitations discussed, allow the determination of differences between patterns generated in different individuals and between cohorts with different responses and toxicities to platinum-based chemotherapy treatment.

There are many steps still required to reliable use the assay in clinical studies and to validate the assay for use in these contexts. A three stage process of biomarker development, as suggested by Hall, would be a valuable model to follow to develop and validate the assay towards these goals, and will be described below (Hall et al. 2007).

Following the pre-clinic development described in this thesis the first phase of clinical use must be a feasibility and small-scale pilot study. The assay needs to be tested in the clinical environment to assess the practicality of using the technique with the current protocol, and to understand the difficulties that may occur when taking samples in the clinical setting. For example, one issue may be optimising the timing and transport of venesection samples – issues that do not arise in the laboratory-based development phase.

In such a pilot study it would be prudent to use two groups of patients of known extreme phenotypic response to platinum agents to increase the chance of detecting differences between cohorts. For example, platinum-based chemotherapy for stage IV lung cancer has approximately a 50% response rate (Cobo et al. 2007). Taking a cohort of 10 patient who responded to treatment and 10 who progressed on treatment would assess the feasibility of taking clinical samples, and potentially allow the initial identification of putative patterns of adducts that are different between the two cohorts. This type of exploratory study would be hypothesis generating; producing putative genomic signatures of chemotherapy response, and would allow the further refinement of the assay and the process by which blood samples are taken, transported to the laboratory, and the development and refinement of the statistical approaches needed to robustly analyse the microarray data generated. As discussed in chapter 5, the identification of characteristic patterns and the refinement of the

statistical approaches and cut-offs required to detect this putative difference are critical to the future development of this assay in this context.

The putative genomic signature identified in the exploratory study would then require confirmation in a larger and statistically robust second phase study. The magnitude of differences between the datasets identified as a candidate genomic signature in the pilot project would allow power calculation of sample sizes to allow rigorous independent statistical validation of the hypothesised patterns. This type of study would require the relationship to be investigated in a similar population and context as used in the initial pilot project. With adequate samples numbers.

It may be that an iterative process is required between the two phases of clinical projects, with a small hypothesis generating study and a larger 'confirmation of relationship' study required over several iterations to refine the signature under consideration and a the magnitude of differences detected. The requirement for obtaining genomic scale data may be reduced if the region of the genome is identified as containing a putative signature exists over a small number of genetic loci. In this circumstance it may be possible to detect these differences between samples at a small number of loci using only DIP-qPCR, dramatically reducing the complexity and cost of the assay.

If a hypothesised signature can be confirmed in a larger, statistically robust second-phase study the next stage would be to estimate the clinical relevance of the assay, principally by testing in a prospective fashion that the marker improves patient or treatment selection compared to currently available approaches. This could be conducted in a similar manner to the testing of putative ERCC1 biomarker in a 2007 lung cancer trial (Cobo et al. 2007) or similar to the on-going UK FOCUS 4 trial assessing the impact of several different biomarkers in predicting the outcome of treatment for colorectal cancer (MRC 2015). This type of large prospective study would be required to provide robust evidence for the use of a predictive biomarker in the selection of treatment for routine use in clinical patients.

7.4 Potential clinical applications

The evidence presented in this thesis demonstrates that the technique is ready for testing and refinement through clinical application, following the pathway discussed in the previous section and the refinement and resolution of the limitations described. The context and potential uses of the assay in clinical studies requires careful consideration, as the development pathway outlined above is context specific and would require the ultimate clinical application to be in mind at the outset; validation and development of putative genomic signatures needs to be conducted with the ultimate use of the assay in mind. In view of this, several example clinical scenarios where a validated and prospectively tested predictive biomarker assay for platinum response or toxicity could be employed are discussed below.

One example of a potential clinical application is to examine for differences in induced oxaliplatin-DNA adduct patterns in cohorts of colorectal cancer patients previously treated with adjuvant oxaliplatin-base chemotherapy, comparing the induced patterns in a cohort with and without OIPN during treatment. Evidence presented in the functional study described in this thesis suggests some patients who develop severe OIPN may have defects in DNA repair capacity, supporting this approach. If a signature of platinum-DNA adducts identifies patients at higher risk of severe OIPN, this would give the treating clinician several advantages, including to allow identification of patients who would benefit from closer monitoring for the development of OIPN, modification of the chemotherapy drugs used to a less neurotoxic equivalent, and more effective testing of preventative measures in a high risk cohort.

An alternative potential clinical study is to examine platinum-DNA adduct patterns in patients treated with cisplatin for metastatic or surgically resected non-small cell lung cancer (NSCLC). Cisplatin in NSCLC has approximately a 40% response rate (Cobo et al. 2007). A 20ml blood sample prior to chemotherapy could be taken and platinum-DNA adduct patterns generated. The tools developed in this thesis could be used to compare the adduct patterns generated

between responders and non-responders, potentially identifying platinum-DNA adducts at different regions of the genome that correlate more closely with response and toxicity than the 'average' platinum-DNA adduct levels, or average levels of DNA repair that have previously demonstrated significant, albeit limited, predictive potential in this context (Hutchinson 2011; Wang et al. 2011)

Other clinical tumour sites and treatments that could benefit from this approach to patient stratification include the treatment of malignancies of the head and neck. For example, many patients are treated with curative intent with a combination of chemotherapy and radiotherapy, with the addition of cisplatin chemotherapy during radiotherapy offering a 5-10% additional survival benefit, depending on tumour type, stage and patient age (Pignon et al. 2009). The addition of cisplatin significantly increases the side effect profile of treatment (Bernier et al. 2004; Cooper et al. 2004), although an alternative treatment with monoclonal antibody treatment cetuximab offers a less toxic, but much more expensive alternative (Bonner et al. 2006), which currently can only be used if cisplatin is contraindicated (NICE 2008). The DIP-chip assay could be applied to pre-treatment blood samples and generated platinum-DNA adduct patterns compared between responders and non-responders, or between patients who developed severe toxicities. If a characteristic signature of platinum non-response or platinum-related severe toxicity could be identified, it could potentially be used pre-treatment to determine if patients should receive concomitant chemotherapy with cisplatin or cetuximab during curative radiotherapy treatment.

Alternative lines of clinically relevant experiments could also be pursued. One example is study to compare the generation of platinum-DNA adduct patterns in surrogate tissues (such as PBMC or buccal cells) compared to tumour samples. This could be done in the context of oxaliplatin use as a treatment prior to the resection of liver or lung metastasis, or in lung cancer comparing tissue taken during curative surgery of primary lung tumours, comparing the germline DNA and tumour DNA platinum adduct patterns, and compared to the response and toxicity from adjuvant cisplatin chemotherapy.

These are just a few examples of the important clinical role for which an accurate method of identifying patients at high risk of toxicity or low risk of response would be a valuable clinical tool.

7.5 Other potential applications of the DIP-chip assay

The DNA-damage adduct patterns generated by this technique have many applications for basic scientific research purposes. One example is in the study of causes of mutation spectra recently identified during next-generation sequencing of tumour and germline DNA (Alexandrov et al. 2013a; Alexandrov et al. 2013b; Nik-Zainal et al. 2012). Patterns of mutations have been identified which characterise a specific underlying mutational processes - for example smokers have a particular spectrum of mutation, as do patients previously treated for cancer with temozolomide or patient heavily exposed to UV damage who develop melanoma. Of the characteristic mutation spectra identified, in several examples the specific underlying cause remains undetermined.

One potential application of the DIP-chip assay technology is the generation of induced DNA damage patterns, comparing the sites of elevated DNA-adduct formation with the mutation spectra that occur. This could be used in patients previously exposed to platinum-treatment. For example, many young patients with testicular cancer are treated curatively with cisplatin, but adducts can be identified decades after treatment completed (Gietema et al. 2000), and, by some estimates, there is a 40% increased risk of second malignancy in the decades following chemotherapy treatment for this condition (Fung et al. 2013). Comparison of DIP-chip cisplatin-DNA adduct patterns with the spectrum of mutations present in germline and tumour DNA in this cohort of patients may shed light on this process, and could again be potentially used as a stratification tool to identify patients at increased risk of second malignancy following curative chemotherapy.

Additional applications of the work presented in this thesis include the use of the bioinformatic tools developed in other similar experimental systems, and these tools are applied in the laboratory for a quality control measure for DNA micro-arrays, to assess whether the IN and IP samples have correctly amplified and processed compared to other experimental samples, not just for platinum-DIP chip samples, but in a variety of different contexts including protein-binding analysis and histone acetylation DIP-chip studies.

Finally, inherent in the goal of this project is the development a model experimental pathway to allow the translation, optimisation and validation of this DIP-chip assay into other clinical tissues, such as tumour or biopsy samples. It is likely that during future projects similar issues surrounding DIP-chip assay variability in different tissue types will manifest. The experimental approach outlined in this thesis can be used as a model pathway for determining the relative contribution of different stages of the multi-step assay and for optimising the DIP-chip assay for different situations and different clinical material.

7.6 Conclusions

Subtle differences in DNA repair gene function are associated with outcome from chemotherapy, for example in the form of OIPN as demonstrated in the functional study presented in this thesis. This work lends weight to the hypothesis that functional measurements of DNA repair are potentially a useful tool that could be used to stratify patients for the response and toxicity to platinum chemotherapy agents; an idea explored in all of the chapters of this thesis. Evidence from 'single measure' adduct studies, as described above, have demonstrated that stratification of patients for response and toxicity to platinum agents by measuring features of germline DNA repair capacity is possible, although with a limited predictive power.

By measuring induced platinum-DNA adduct levels at high resolution and at a genomic-scale, and by developing and applying unbiased genomic scale

analysis techniques, we have developed a technology with robust experimental protocols and analysis pipelines, with the potential to overcome these technological limitations. This thesis documents and describes the development of this technology for use in human clinical samples and includes robust protocols to perform the assay and analyse the result, as well as providing a pathway for the refinement and future validation of the technique if applied to other clinical tissues. The DIP-chip assay, in the form and context as described here, has opened the door to many clinical and research applications, and several potential studies that will be conducted in our laboratory in the coming months and years.

Chapter 8 Appendicies

8.1 Appendix 1: Media and Solutions

8.1.1 Yeast Media

YPD (400 ml)
4 g Bacto Yeast Extract
8 g Bacto Peptone
8 g Glucose
Made up to 400 ml with H_2O
YPD plates were obtained by supplementing above medium with 8 g Bacto Agar

Minimal Media (400 ml)
2.68 g Yeast Nitrogen Base w/o Amino Acids 8 g Glucose
Made up to 400 ml with H_2O
For plates supplement with 2% Bacto Agar

Dosage of Amino Acids added:

Adenine	40 µg/ml
Leucine	60 µg/ml
Uracil	20 µg/ml
Tryptophan	40 µg/ml
Lysine	40 µg/ml
Histidine	40 µg/ml

8.1.2 Solutions

0.5 M EDTA (pH 8.0)

$EDTA \cdot Na_2 \cdot 2H_2O$	186.1g
H_2O	800ml

Stir on a magnetic stirrer.
Adjust the pH to 8.0 with NaOH (~22 g of NaOH pellets).
Add H_2O to make 1 L and sterilise by autoclaving.

1 M Tris

Tris base	121.1 g
H_2O	800 ml

Adjust the pH to the desired value by adding concentrated HCl.
Add H_2O to make 1 L.

10 x TE Buffer (400 ml)
40 ml 1 M Tris-HCl, pH 7.5
8 ml 0.5 M EDTA, pH 8.0
352 ml of H_2O

3 M Sodium acetate (pH 5.2) (400 ml)

Sodium acetate - $3H_2O$	163.24 g
H_2O	300 ml

Adjust the pH to 5.2 with acetic acid.
Adjust the volume to 400 ml with H_2O.
Filter to sterilise.

5M NaCL (400 ml)

Dissolve 116.9 g of NaCl in 350 ml of H_2O
Adjust the volume to 400 ml with H_2O.
Sterilise by autoclaving.

10% SDS (Sodium dodecyl sulfate) (1 L)

Dissolve 100g of SDS in 800 ml of distilled H_2O.
Add distilled H_2O to make a total volume of 1 L.

20% SDS (Sodium dodecyl sulfate) (500 ml)

Dissolve 100 g of SDS in 350 ml of distilled H_2O.
Add distilled H_2O to make a total volume of 500 ml.

PBS (1 L)

NaCL	8g
KCl	0.2g
Na_2HPO_4	1.44 g
KH_2PO_4	0.24g
H_2O	800ml

Adjust the pH to 7.4.
Add H_2O to 1 L.
Sterilise by autoclaving.

Sorbitol TE (1L) (Kept in cold room at 4°C)

Sorbitol	165g
Tris . HCl (pH8.0)	100ml
EDTA	200ml

Add 500 ml H_2O to dissolve the sorbitol.
Adjust the final volume to 1L.

DNA Lysis Buffer (1 L)

Urea	240g
NaCl	11.69g
CDTA	5g

SDS 5g
1 M Tris-HCl (pH 8.0) 100ml
Add 700 ml of H_2O to dissolve the chemicals
Adjust the final volume to 1L.

1 M DTT (Dithiothreitol)
DTT 3.09 g
Dissolved in 20 ml of 0.01 M sodium acetate (pH 5.2)
Sterilise by filtration.
Do not autoclave solutions containing DTT.

8.1.3 Solutions for electrophoresis

50 x TAE (Tris-acetate) (1 L)

Tris base	242g
Sodium Acetate $\cdot 3\ H_2O$	136g
0.5 M EDTA	200ml

Adjust to pH 7.2 with acetic acid.
Add H_2O to make 1 L.

10 x TBE (1 L)

Tris base	106g
Boric acid	55g
EDTA	8.3g

Add H_2O to 1 L.

8.1.4 Solutions for ChIP

FA/SDS Buffer

HEPES KOH pH 7.5	50mM
NaCl	150mM
EDTA	1mM
Triton X 100	1%
Deoxycholate Na	0.1%
SDS	0.1%
PMSF *	1mM

*Add just before use as the activity decreases in H20 solution The NaCl concentration can be adjusted to 500mM.

5 x Pronase Buffer (100 ml)

1 M Tris pH7.5	12.5mM
0.5 M EDTA	5ml
10% SDS	25ml

Add H_2O to make 100 ml.

LiCl Buffer (500 ml)

1 M Tris pH 8.0	5ml
5 M LiCl	25ml
0.5 M EDTA	1ml
NP40	2.5ml
Deoxycholate Na	25ml

Add H_20 to make 500 ml.

8.2 Appendix 2: SCA Ancillary Functions Scripts

---

1) SCAdataload function

Version 2.0 11/09/14
Role: Loads data for SCA plot and corPlot functions
Requirements: Sandcastle 1.0
Output to vector SCAdata

Example
SCAdata<-SCAdataLoad
SCAdata<-SCAdataLoad

```
SCAdataLoad<-function(all=TRUE){
  dataratio<-loadArrayFiles(chromosomeIDs=c(1:22,"X","Y"))
  dataIN<-loadArrayFiles(chromosomeIDs=c(1:22,"X","Y"),
     redValues=NULL)
  dataIP<-loadArrayFiles(chromosomeIDs=c(1:22,"X","Y"),
     greenValues=NULL)

if(all){
    SCAdata<-c(dataratio,dataIN,dataIP)}

else{
     z<-readline("How many array files to load for SCA analysis?")
     z<-as.numeric(z)
     mat<-matrix(nrow=2,ncol=z)
     for (n in 1:z) mat[1,n]<-n
     print("Enter datasets in order to load")
     for (n in 1:z) {

     print(n)
     mat[2,n]<-as.numeric(readline("Enter dataset..."))}
     for (n in 1:z) for (m in 1:2) mat[m,n]<-as.numeric(mat[m,n])
     SCAdata<-dataratio[,(mat[2,1])]
     for (n in 1:(z-1)) SCAdata<-c(SCAdata,dataratio[,(mat[2,(n
          +1)])])
     for (n in 1:z) SCAdata<-c(SCAdata,dataIN[,(mat[2,n])])
     for (n in 1:z) SCAdata<-c(SCAdata,dataIP[,(mat[2,n])])
    return(SCAdata)}
}#end function
```

---

2) corsCalc function

Version 1.0 11/06/14
Role: Calculates correlation matrix table from SCAdata to use in SCAplot and corsPlot
Requirements: Vector SCAdata from SCAdataload

Output to vector cors

Example
cors<-corsCalc()

```
corsCalc<-function(){
  cors<-matrix(ncol=(ncol(SCAdata)), nrow=(ncol(SCAdata)))
  for (n in 1:ncol(SCAdata)) for (m in 1:ncol(SCAdata)) cors[m,n]<-
cor.test(SCAdata[,n]$ratios,SCAdata[,m]$ratios, method="spearman")
$estimate
  return(cors)}#end of function
```

---

3) corsPredCalc function

Version 1.0 11/06/14
Role: Calculates matrix of correlations of SCAdata with prediction profiles
Requirements: Sandcastle
Preloaded prediction profiles to vector predPlat
for platinum prediction profile and/or vector predUV
for UV prediction profiles
Output to vector corspred

Example
corspred<-corsPredCalc()

```
corsPredCalc<-function(){
  r=2
  corspred<-matrix(ncol=(ncol(SCAdata)), nrow=r)
  for (n in 1:ncol(SCAdata)) corspred[1,n]<-cor.test(SCAdata[,n]
$ratios,predPlat$ratios, method="spearman")$estimate
  for (n in 1:ncol(SCAdata)) corspred[r,n]<-cor.test(SCAdata[,n]
$ratios,predUV$ratios, method="spearman")$estimate
  return(corspred)}
```

8.3 Appendix 3: SCAplot function

Version 1.1, 110614
Role: Plots data in vector SCAdata, separated by channels
Requirements: Sandcastle package.
Human genome annotation to vector anno.human
SCA ancillary function SCAdataload to vector SCAdata

Examples:

SCAplot()
SCAplot(end=10080000,legends=FALSE)
SCAplot(11000000,10060000)
SCAplot(highlight=T, labels=T, datatohighlight=5)

```
SCAplot<-function(start=10050000,end=10075000, legends=TRUE,
labels=FALSE, highlight=FALSE, datatohighlight=FALSE){

     if ((end-start)>50000) stop(" Resolution too low >50000
     nucleotides")
     if ((start>end)) stop("End position must be after start
     position")

a<-(ncol(SCAdata)/3)
dth<-datatohighlight
dthcol="purple"

#add text for labels if needed
if(labels){
mat<-matrix(nrow=1,ncol=a)
print("Enter dataset labels in order of use")
for (n in 1:a) {
print(n)
mat[1,n]<-readline("Enter label...")}}
x<-matrix(ncol=5,c(17,17,17,start,start,start,end,end,end,1,(a+1),
((2*a)+1),a,(2*a),(3*a)))

if(!highlight){
    genomePlot(SCAdata[,c((1:a),((a+1):(2*a)),(((2*a)+1):(3*a)))],
               anno.human,chr=17,ylim.constant=FALSE,
               multi=TRUE,toPlot=x, lwd=2,
               cols=list(rep(1,a),rep(3,a),rep(2,a)),
               ltys=list(seq(1:a)))

if (legends){

if (!labels){
    legend(cex=0.8,x="topright",box.lwd=1, box.lty=1, box.col=1,
     bg="lightgoldenrod",
     ncol=1,title="IP/IN Pattern",
     legend=seq(1:a),
     col=rep(1,a),
     lty=seq(1:a))

legend(cex=0.8,x="right",box.lwd=1, box.lty=1, box.col=1,
```

```
     bg="lightgoldenrod",
     ncol=1,title="IN Pattern",
     legend=seq(1:a),
     col=rep(3,a),
     lty=seq(1:a))

legend(cex=0.8,x="bottomright",box.lwd=1, box.lty=1,
     box.col=1, bg="lightgoldenrod",
     ncol=1,title="IP Pattern",
     legend=seq(1:a),
     col=rep(2,a),
     lty=seq(1:a))}

if(labels){

legend(cex=0.8,x="topright",box.lwd=1, box.lty=1, box.col=1,
     bg="lightgoldenrod",
     ncol=1,title="IP/IN Pattern",
     legend=mat[1,],
     col=rep(1,a),
     lty=seq(1:a))

legend(cex=0.8,x="right",box.lwd=1, box.lty=1, box.col=1,
     bg="lightgoldenrod",
     ncol=1,title="IN Pattern",
     legend=mat[1,],
     col=rep(3,a),
     lty=seq(1:a))

legend(cex=0.8,x="bottomright",box.lwd=1, box.lty=1,
     box.col=1, bg="lightgoldenrod",
      ncol=1,title="IP Pattern",
      legend=mat[1,],
      col=rep(2,a),
      lty=seq(1:a))}}}

if(highlight){
    genomePlot(SCAdata[,c((1:a),((a+1):(2*a)),(((2*a)+1):(3*a)))],
          anno.human,chr=17,ylim.constant=FALSE,
          multi=TRUE,toPlot=x, lwd=2,
          cols=list(c(rep(1,(dth-1)),dthcol,rep(1,(a-
dth))),c(rep(3,(dth-1)),dthcol,rep(3,(a-dth))),c(rep(2,
(dth-1)),dthcol,rep(2,(a-dth)))),
          ltys=list(seq(1:a)))

if (legends){

if(!labels){

legend(cex=0.8,x="topright",box.lwd=1, box.lty=1, box.col=1,
          bg="lightgoldenrod",
          ncol=1,title="IP/IN Pattern",
          legend=seq(1:a),
          col=c(rep(1,(dth-1)),dthcol,rep(1,(a-dth))),
          lty=seq(1:a))

legend(cex=0.8,x="right",box.lwd=1, box.lty=1, box.col=1,
          bg="lightgoldenrod",
          ncol=1,title="IN Pattern",
          legend=seq(1:a),
          col=c(rep(3,(dth-1)),dthcol,rep(3,(a-dth))),
          lty=seq(1:a))
```

```
legend(cex=0.8,x="bottomright",box.lwd=1,
     box.lty=1,box.col=1,bg="lightgoldenrod",
      ncol=1,title="IP Pattern",
      legend=seq(1:a),
      col=c(rep(2,(dth-1)),dthcol,rep(2,(a-dth))),
      lty=seq(1:a))}}

if(labels){

legend(cex=0.8,x="topright",box.lwd=1, box.lty=1, box.col=1,
     bg="lightgoldenrod",
     ncol=1,title="IP/IN Pattern",
     legend=mat[1,],
     col=c(rep(1,(dth-1)),dthcol,rep(1,(a-dth))),
     lty=seq(1:a))

legend(cex=0.8,x="right",box.lwd=1, box.lty=1, box.col=1,
     bg="lightgoldenrod",
     ncol=1,title="IN Pattern",
     legend=mat[1,],
     col=c(rep(3,(dth-1)),dthcol,rep(3,(a-dth))),
     lty=seq(1:a))

legend(cex=0.8,x="bottomright",box.lwd=1, box.lty=1,box.col=1,
     bg="lightgoldenrod",
     ncol=1,title="IP Pattern",
     legend=mat[1,],
     col=c(rep(2,(dth-1)),dthcol,rep(2,(a-dth))),
      lty=seq(1:a))}}

} #end of function
```

8.4 Appendix 4: The corPlot Function

Version 1.2 11/09/14
Role: Plots SCA matrix
Requirements: Data in vector SCAdata from SCAdataLoad function
Correlations calculated by corsCalc in vector cors
If prediction profile required the prediction correlations in vector corspred from function corsPredCalc

Examples
corPlot(labels=F,FigSize=1)
corPlot(labels=T, FigSize=0.7)

```
corPlot<-function(boxtext=TRUE, prediction=FALSE, UV=FALSE,
platinum=FALSE, labels=FALSE, FigSize=1, key=TRUE){

#check if prediction required then has been specified and logical
arguments are correct
if (!is.logical(boxtext) |!is.logical(prediction) | !is.logical(UV) |
!is.logical(platinum)) stop ("Invalid TRUE/FALSEargument present")
if ((prediction) & ((!UV)&(!platinum))) stop ("Prediction pattern is
selected.  Choose either/both UV or platinum as TRUE and ensure
correct matrix from CorsPredCal() is complete")

#set up numbers
s<-(ncol(cors)/3)       #need to use to remove data from cor table
for individual colour boxes
  boxtextsize<-((1-(0.07*s))*FigSize)
  axistextsize<-((1.2-(0.05*s))*FigSize)

#add text for labels if needed
    if(labels){
      same<-readline("Use same labels for IP/IN, IN and IP? - Enter 1
if yes or 2 if no")
    if ((same!=1)&(same!=2)) stop ("Must be 1 or 2")
    if (same==1){
      mat<-matrix(nrow=2,ncol=s)
      for (n in 1:s) mat[1,n]<-n
      print("Enter dataset labels in order of use")

for (n in 1:s) {
      print(n)
      mat[2,n]<-readline("Enter label...")}}

if (same==2){
      mat<-matrix(nrow=2,ncol=(3*s))
      for (n in 1:(3*s)) mat[1,n]<-n
      print("Enter dataset labels in order of use")
      for (n in 1:(3*s)) {
      print(n)
            mat[2,n]<-readline("Enter label...")}}
    }

#set up color palettes
```

```
palINIP<-colorRampPalette(c("white","yellow","gold"))(n=199)
breaksINIP=c(seq(0,0.5,length=10),seq(0.51,0.75,length=40),seq(0.75,1
,length=150))

palIN<-
colorRampPalette(c("white","white","green","darkgreen"))(n=199)
breaksIN=c(seq(0,0.5,length=5),seq(0.51,0.75,length=10),seq(0.76,1,le
ngth=185))

palIP<-colorRampPalette(c("white","white","red","darkred"))(n=199)
breaksIP=c(seq(0,0.5,length=5),seq(0.51,0.75,length=10),seq(0.76,1,le
ngth=185))

palrest<-
colorRampPalette(c("purple","violet","white","white","white","orange
","coral2"))(n=199)
breaksrest=c(seq(-1,-0.8,length=20),seq(-
0.8,0.8,length=160),seq(0.8,1,length=20))

#calculate plot specific correlation values and rotate for image
corst<-t(cors)[,nrow(cors):1]

#calculate individual value sets to use for overlying plots
#only 1s for image
corst1only<-corst
corst1only[corst1only<0.999]<-NA

#remove 1vs1 for later tables
corst[corst>0.999]<-NA
#IP/IN values only
corstIPIN<-corst
corstIPIN[1:(3*s),1:(2*s)]<-NA
corstIPIN[(s+1):(s*3),(2*s+1):(3*s)]<-NA
#IN values only
corstIN<-corst
corstIN[1:s,1:(3*s)]<-NA
corstIN[(2*s+1):(3*s),1:(3*s)]<-NA
corstIN[1:(3*s),1:s]<-NA
corstIN[1:(3*s),(2*s+1):(3*s)]<-NA
#IP values only
corstIP<-corst
corstIP[1:(3*s),(s+1):(3*s)]<-NA
corstIP[1:(2*s),1:(s)]<-NA
#rest values only
corstrest<-corst
corstrest[(s+1):(3*s),(s+1):(3*s)]<-NA
corstrest[1:s,(2*s+1):(3*s)]<-NA
corstrest[(2*s+1):(3*s),1:s]<-NA

#plot individual sections, overlying each other on same grid
#set plot area and adjust from center by figure size
if (!prediction) par(fig=c((0.425-(0.375*FigSize)),
(0.425+(0.375*FigSize)),(0.5-(0.4*FigSize)),(0.5+(0.4*FigSize))))
else {par(fig=c((0.425-(0.375*FigSize)),(0.425+(0.375*FigSize)),(0.6-
(0.4*FigSize)),(0.6+(0.4*FigSize))))

axistextsize<-(axistextsize-0.2)}
 #plot 1s only as grey
image(x=seq(dim(corst1only)[2]),y=seq(dim(corst1only)[1]),z=corst1onl
y, axes=FALSE,xlab="",ylab="",useRaster=TRUE, col="grey90")

#plot IP/IN
par(new=TRUE)
```

```
image(x=seq(dim(corstIPIN)[2]),y=seq(dim(corstIPIN)[1]),z=corstIPIN,
axes=FALSE,xlab="",ylab="", useRaster=TRUE, col=palINIP,
breaks=breaksINIP)

if (boxtext)
text(expand.grid(x=seq(dim(corstIPIN)[2]),y=seq(dim(corstIPIN)[1])),l
abels=round(c(corstIPIN),2),cex=boxtextsize)

#plotIN
par(new=TRUE)
image(x=seq(dim(corstIN)[2]),y=seq(dim(corstIN)[1]),z=corstIN,
axes=FALSE,xlab="",ylab="",useRaster=TRUE, col=palIN,
breaks=breaksIN)

if (boxtext)
text(expand.grid(x=seq(dim(corstIN)[2]),y=seq(dim(corstIN)[1])),label
s=round(c(corstIN),
2),cex=boxtextsize)

#plotIP
par(new=TRUE)
image(x=seq(dim(corstIP)[2]),y=seq(dim(corstIP)[1]),z=corstIP,
axes=FALSE,xlab="",ylab="", useRaster=TRUE, col=palIP,
breaks=breaksIP)

if (boxtext)
text(expand.grid(x=seq(dim(corstIP)
[2]),y=seq(dim(corstIP)[1])),labels=round(c(corstIP),
2),cex=boxtextsize)

#plotrest
par(new=TRUE)
image(x=seq(dim(corstrest)[2]),y=seq(dim(corstrest)[1]),z=corstrest,
axes=FALSE,xlab="",ylab="",useRaster=TRUE, col=palrest,
breaks=breaksrest)

if (boxtext) text(expand.grid(x=seq(dim(corstrest)
[2]),y=seq(dim(corstrest)[1])),labels=round(c(corstrest),
2),cex=boxtextsize)

#plot axes
#if no labels needed
  if (!labels){ linewd=1.5
axis(2, at=c(seq((3*s),1,-
1)),las=1,lwd=1,lab=c(rep(1:s),rep(1:s),rep(1:s)),
cex.axis=axistextsize,
line=linewd, tck=0, lty=0)
axis(1,
at=c(seq(1,(3*s),1,)),las=2,lwd=1,lab=c(rep(1:s),rep(1:s),rep(1:s)),
cex.axis=axistextsize,
line=linewd, tck=0, lty=0)
axis(4, at=c(seq((3*s),1,-
1)),las=1,lwd=1,lab=c(rep(1:s),rep(1:s),rep(1:s)),
cex.axis=axistextsize,
line=linewd, tck=0, lty=0)
axis(3,
at=c(seq(1,(3*s),1,)),las=2,lwd=1,lab=c(rep(1:s),rep(1:s),rep(1:s)),
cex.axis=axistextsize,
line=linewd, tck=0, lty=0)}

#add option to label individual lanes
  else {linewd=1.5
    if(same==1){
```

```
      #if same label for IP/IN, IN or IP needed
      for (n in 1:s) {
            axis(1, at=c(n,(s+n),(2*s+n)),
      lab=c(mat[2,n],mat[2,n],mat[2,n]), las=2,
cex.axis=axistextsize,
      line=linewd, lty=0)
      axis(2, at=c((3*s+1-n),(2*s+1-n),(s+1-n)),
      lab=c(mat[2,n],mat[2,n],mat[2,n]),las=1, cex.axis=axistextsize,
      line=linewd, lty=0)
            axis(3, at=c(n,(s+n),(2*s+n)),
      lab=c(mat[2,n],mat[2,n],mat[2,n]),las=2, cex.axis=axistextsize,
      line=linewd, lty=0)
            axis(4, at=c((3*s+1-n),(2*s+1-n),(s+1-n)),
      lab=c(mat[2,n],mat[2,n],mat[2,n]),las=1, cex.axis=axistextsize,
      line=linewd, lty=0)
      }}

else{
      #if individual labels for IP/IN, IN and IPs needed
      for (n in 1:s) {
        axis(1, at=c(n,(s+n),(2*s+n)), lab=c(mat[2,n],mat[2,(s
+n)],mat[2,(n+2*s)]), las=2, cex.axis=axistextsize, line=linewd,
lty=0)
        axis(2, at=c((3*s+1-n),(2*s+1-n),(s+1-n)),
lab=c(mat[2,n],mat[2,(s+n)],mat[2,(2*s+n)]),las=1,
cex.axis=axistextsize, line=linewd, lty=0)
        axis(3, at=c(n,(s+n),(2*s+n)), lab=c(mat[2,n],mat[2,(s
+n)],mat[2,(n+2*s)]),las=2, cex.axis=axistextsize, line=linewd,
lty=0)
        axis(4, at=c((3*s+1-n),(2*s+1-n),(s+1-n)),
lab=c(mat[2,n],mat[2,s+n],mat[2,n+2*s]),las=1,
cex.axis=axistextsize, line=linewd, lty=0)
}}}
  #add IP/IN, IN and IP inner coloured labels to plot
  linewd<-(-0.5)
  if (s<11) {
    for (n in 1:s)
    {axis(1, at=n, lab=("IP/IN"), col.axis="goldenrod3",
line=linewd, las=2, lty=0, cex.axis=axistextsize)
     axis(1, at=s+n, lab=("IN"), col.axis="darkgreen", line=linewd,
tck=0, las=2, lty=0, cex.axis=axistextsize)
     axis(1, at=(2*s+n), lab=("IP"), col.axis="darkred",
line=linewd, tck=0, las=2, lty=0, cex.axis=axistextsize)
     axis(2, at=2*s+n, lab=("IP/IN"), col.axis="goldenrod3",
line=linewd, tck=0, las=2, lty=0, cex.axis=axistextsize)
     axis(2, at=s+n, lab=("IN"), col.axis="darkgreen", line=linewd,
tck=0, las=1, lty=0, cex.axis=axistextsize)
     axis(2, at=n, lab=("IP"), col.axis="darkred", line=linewd,
tck=0, las=1, lty=0, cex.axis=axistextsize)
     axis(3, at=n, lab=("IP/IN"), col.axis="goldenrod3",
line=linewd, tck=0, las=2, lty=0, cex.axis=axistextsize)
     axis(3, at=s+n, lab=("IN"), col.axis="darkgreen", line=linewd,
tck=0, las=2, lty=0, cex.axis=axistextsize)
     axis(3, at=(2*s+n), lab=("IP"), col.axis="darkred",
line=linewd, tck=0, las=2, lty=0, cex.axis=axistextsize)
     axis(4, at=2*s+n, lab=("IP/IN"), col.axis="goldenrod3",
line=linewd, tck=0, las=2, lty=0, cex.axis=axistextsize)
     axis(4, at=s+n, lab=("IN"), col.axis="darkgreen", line=linewd,
tck=0, las=1, lty=0, cex.axis=axistextsize)
     axis(4, at=n, lab=("IP"), col.axis="darkred", line=linewd,
tck=0, las=1, lty=0, cex.axis=axistextsize)}}
  else {
```

```
    #just add one IP/IN IN and IP label if too many datasets to avoid
crowding
    axis(1, at=(s/2), lab=("IP/IN"), col.axis="goldenrod3",
line=linewd, tck=0, las=2, lty=0, cex.axis=axistextsize)
    axis(1, at=1.5*s, lab=("IN"), col.axis="darkgreen", line=linewd,
tck=0, las=2, lty=0, cex.axis=axistextsize)
    axis(1, at=2.5*s, lab=("IP"), col.axis="darkred", line=linewd,
tck=0, las=2, lty=0, cex.axis=axistextsize)
    axis(2, at=2.5*s, lab=("IP/IN"), col.axis="goldenrod3",
line=linewd, tck=0, las=2, lty=0, cex.axis=axistextsize)
    axis(2, at=1.5*s, lab=("IN"), col.axis="darkgreen", line=linewd,
tck=0, las=1, lty=0, cex.axis=axistextsize)
    axis(2, at=(s/2), lab=("IP"), col.axis="darkred", line=linewd,
tck=0, las=1, lty=0, cex.axis=axistextsize)
    axis(3, at=(s/2), lab=("IP/IN"), col.axis="goldenrod3",
line=linewd, tck=0, las=2, lty=0, cex.axis=axistextsize)
    axis(3, at=1.5*s, lab=("IN"), col.axis="darkgreen", line=linewd,
tck=0, las=2, lty=0, cex.axis=axistextsize)
    axis(3, at=2.5*s, lab=("IP"), col.axis="darkred", line=linewd,
tck=0, las=2, lty=0, cex.axis=axistextsize)
    axis(4, at=2.5*s, lab=("IP/IN"), col.axis="goldenrod3",
line=linewd, tck=0, las=2, lty=0, cex.axis=axistextsize)
    axis(4, at=1.5*s, lab=("IN"), col.axis="darkgreen", line=linewd,
tck=0, las=1, lty=0, cex.axis=axistextsize)
    axis(4, at=(s/2), lab=("IP"), col.axis="darkred", line=linewd,
tck=0, las=1, lty=0, cex.axis=axistextsize)
}
  #fill in lines around specific boxes
  box()
  segments((s+0.5),(s+0.5),(s+0.5),(3*s+0.5), lwd=3)
  segments(0,(2*s+0.5),(2*s+0.5),(2*s+0.5), lwd=3)
  segments(0.52,(2*s+0.5),0.52,(3*s+0.5), lwd=3)
  segments(0.5,(3*s+0.48),(s+0.5),(3*s+0.48), lwd=3)
  segments((s+0.5),(s+0.5),(3*s+0.5),(s+0.5), lwd=3)
  segments((2*s+0.5),0,(2*s+0.5),(2*s+0.5), lwd=3)
  segments((2*s+0.5),0.52,(3*s+0.5),0.52, lwd=3)
  segments((3*s+0.48),0.5,(3*s+0.48),(s+0.5), lwd=3)
  segments((s+0.5),(s+0.5),(s+0.5),0, lwd=2)
  segments(0.5,(s+0.5),(s+0.5),(s+0.5), lwd=2)
  abline(h=2*s+0.5, lty=2)
  abline(v=2*s+0.5, lty=2)
  for (n in 1:s){
    abline (h=(n+0.5), v=(n+0.5),lty=2)
    segments((s+n+0.5),(2*s+0.5),(s+n+0.5),0, lty=2)
    segments((2*s+n+0.5),(s+0.5),(2*s+n+0.5),0, lty=2)
    segments(0,(s+n+0.5),(2*s+0.5),(s+n+0.5), lty=2)
    segments(0,(2*s+n+0.5),(s+0.5),(2*s+n+0.5), lty=2)
}
  #draw prediction profiles under main figure
  if (prediction){
    #If both needed
    if((platinum)&(UV)){
      corstpred<-t(corspred)[,nrow(corspred):1]
      nrow(corstpred)->r
      if (FigSize==1) par(fig=c(0.05,0.8,0.05,0.3),new=TRUE) else
par(fig=c((0.4-(0.34*FigSize)),(0.4+(0.4*FigSize)),0.05,0.3),
new=TRUE)
      image(x=seq(1,(nrow(corstpred)),1),y=seq(1,2,1),z=corstpred,
ylab="", xlab="",col=palrest, breaks=breaksrest, axes=FALSE)
      text(expand.grid(x=seq(1,(nrow(corstpred)),
1),y=seq(1,2,1)),labels=round(c(corstpred),2), cex=0.5)
      box(lwd=2)
      abline(h=1.5, lty=2)
```

```
      for (n in 1:2) abline(v=(n*(r/3)+0.5), lwd=2)
      for (n in 1:r) abline(v=n+0.5, lty=2)
      mtext(1, line=2, text="Correlation with Predicted Profiles",
cex=0.75)
      axis(4, at=c(1,2), lab=c("vs. \nUV", "vs.\nPlatinum"), las=2,
tck=0, cex.axis=0.75)
    }
    #if only platinum
    if((platinum)&(!UV)){
      corstpred<-t(corspred)[,nrow(corspred):1]
      nrow(corstpred)->r
      corstpred[,1]<-NA
      if (FigSize==1) par(fig=c(0.05,0.8,0.01,0.25),new=TRUE) else
par(fig=c((0.4-(0.34*FigSize)),(0.4+(0.4*FigSize)),0.01,0.25),
new=TRUE)
      image(x=seq(1,(nrow(corstpred)),1),y=seq(1,2,1),z=corstpred,
ylab="", xlab="",col=palrest, breaks=breaksrest, axes=FALSE)
      text(expand.grid(x=seq(1,(nrow(corstpred)),
1),y=seq(1,2,1)),labels=round(c(corstpred),2), cex=0.5)
      abline(h=1.5, lwd=2)
      abline(h=2.48,lwd=2)
      segments(0.52,1.5,0.52,2.5,lwd=2)
      for (n in 1:4) segments((((n-1)*s)+0.5),1.5,(((n-1)*s)+0.5),
2.5, lwd=2)
      for (n in 1:r) segments((n+0.5),1.5,(n+0.5),2.5, lty=2)
      mtext(1, line=1, text="Correlation with Platinum Predicted
Profile", cex=0.75)
      axis(4, at=2, lab=c("vs.\nPlatinum"), las=2, tck=0,
cex.axis=0.75)
    }
    #if only UV
    if((!platinum)&(UV)){
      corstpred<-t(corspred)[,nrow(corspred):1]
      nrow(corstpred)->r
      corstpred[,2]<-NA
      if (FigSize==1) par(fig=c(0.05,0.8,0.05,0.3),new=TRUE) else
par(fig=c((0.4-(0.34*FigSize)),(0.4+(0.4*FigSize)),0.05,0.3),
new=TRUE)
      image(x=seq(1,r,1),y=seq(1,2,1),z=corstpred, ylab="",
xlab="",col=palrest, breaks=breaksrest, axes=FALSE)
      text(expand.grid(x=seq(1,(nrow(corstpred)),
1),y=seq(1,2,1)),labels=round(c(corstpred),2), cex=0.5)
      abline(h=0.5, lwd=2)
      abline(h=1.5,lwd=2)
      segments(0.52,0.5,0.52,1.5,lwd=2)
      for (n in 1:4) segments((((n-1)*s)+0.5),0.5,(((n-1)*s)+0.5),
1.5, lwd=2)
      for (n in 1:r) segments((n+0.5),0.5,(n+0.5),1.5, lty=2)
      mtext(1, line=3, text="Correlation with UV Predicted Profile",
cex=0.75)
      axis(4, at=1, lab=c("vs.\nUV "), las=2, tck=0, cex.axis=0.75)
}}
  #draw key
   if (key){
  # set up key specific breaks guide
  brkINIP=c(seq(0,50,length=10),seq(51,75,length=40),
(seq(75,100,length=150)))
  brkIN=c(seq(0,50,length=5),seq(51,75,length=10),
(seq(75,100,length=185)))
  brkIP=c(seq(0,50,length=5),seq(51,75,length=10),
(seq(75,100,length=185)))
  brkrest=c(seq(0,10,length=20),seq(10,90,length=160),
(seq(90,100,length=20)))
```

```
  #set up matrix for key plots
  z=matrix(1:100,nrow=1)
  x=1
  y=seq(0,1,length=101)
  p=matrix(1:100,nrow=1)
  q=1
  s=seq(-1,1,length=101)
  #draw scale forIP/IN
  par(fig=c(0.8,0.95,0.65,1), new=TRUE)
  image(x,y,z,col=palINIP,breaks=brkINIP,axes=FALSE,ylab="",xlab="")
  mtext("IP/IN vs. IP/IN Correlation Key",2,line=0.5, cex=0.8)
  axis(4, at=c(seq(0.0,1.0,0.1)), lab=c(seq(0.0,1.0,0.1)), las=2,
cex.axis=0.75, tck=-0.5)
  box()
  #draw scale for IN
  par(fig=c(0.8,0.95,0.45,0.8), new=TRUE)
  image(x,y,z,col=palIN,breaks=brkIN,axes=FALSE, ylab="", xlab="")
  mtext("IN vs. IN Correlation Key",2,line=0.5, cex=0.8)
  axis(4, at=c(seq(0.0,1.0,0.1)), lab=c(seq(0.0,1.0,0.1)), las=2,
cex.axis=0.75,tck=-0.5)
  box()
  #draw scale for IP
  par(fig=c(0.8,0.95,0.25,0.6), new=TRUE)
  image(x,y,z,col=palIP,breaks=brkIP,axes=FALSE,ylab="",xlab="")
  mtext("IP vs. IP Correlation Key",2,line=0.5, cex=0.8)
  axis(4, at=c(seq(0.0,1.0,0.1)), lab=c(seq(0.0,1.0,0.1)), las=2,
cex.axis=0.75,tck=-0.5)
  box()
  #draw scale for rest of correlations
  if(!prediction){
  par(fig=c(0.8,0.95,0,0.4), new=TRUE)
  image(q,s,p,col=palrest,
breaks=brkrest,axes=FALSE,ylab="",xlab="")
  mtext("IP/IN vs. IN. vs IP Correlation Key",2,line=0.5, cex=0.8)
  axis(4, at=c(seq(-1,1.0,0.2)), lab=c(seq(-1,1.0,0.2)), las=2,
cex.axis=0.75,tck=-0.5)
box()} else {
  par(fig=c(0.8,0.95,0,0.4), new=TRUE)
  image(q,s,p,col=palrest,
breaks=brkrest,axes=FALSE,ylab="",xlab="")
  mtext("IP/IN vs. IN. vs IP vs. Prediction\n Correlation Key",
2,line=0.5, cex=0.8)
  axis(4, at=c(seq(-1,1.0,0.2)), lab=c(seq(-1,1.0,0.2)), las=2,
cex.axis=0.75,tck=-0.5)
box()} }
}#End of function
```

8.5 Appendix 5: The AGCcalc function

Version 1.1
Role: Generates correlation between datasets at specified windows. Calculates window GC content.
Requirements: DIP-chip microarray data in vector dataAGC
Output stored to vector AGCmatrix

Example:
AGCmatrix<-AGCcalc()

```
AGCcalc<-
function(start=10000000,end=15000000,segments=50,comparisons=1){

#choose which datasets in dataAGC to compare
     tocompare<-matrix(ncol=3,nrow=comparisons)
     for (n in 1:comparisons) tocompare[n,1]<-n
     print("Enter datasets to copmare in order of desired comparisons
          i.e. 1a vs 1b, 2a vs 2b...")
     for (n in 1:comparisons){
     print("For comparison")
     print(n)
     tocompare[n,2]<-as.numeric(readline("Enter dataset
     'a' (identified using numerical order in vector 'dataAGC')..."))
     tocompare[n,3]<-as.numeric(readline("Enter dataset
     'b' (identified using numerical order in vector 'dataAGC')..."))
    }

#caluclate GC content at each window
    #check necessary packages and obtain if absent
    require(Biostrings) #load package if required
    require(BSgenome) #load package if required
    require(BSgenome.Hsapiens.UCSC.hg19)

#set up matrix for to contain window data
    AGCmatrix<-matrix(nrow=comparisons+1,ncol=segments)
    winsize=((end-start)/segments)

#calculate GC content of window and add to matrix seqmat

for (n in 1:segments){
     winstart=start+((n-1)*winsize)
     winend=start+n*(winsize)
     chr17s<-getSeq(Hsapiens,"chr17",start=winstart,end=winend-1)
     L<-length(chr17s)
     p1<-"G"
     p2<-"C"
     G<-countPattern(p1,chr17s)
     C<-countPattern(p2,chr17s)
     AGCmatrix[(comparisons+1),n]<-(G+C)/L}

#calculate correlation between datasets at each window

for (m in 1:segments) for (n in 1:comparisons) AGCmatrix[n,m]<-
cor.test(dataAGC[,(tocompare[n,2])][dataAGC[,(tocompare[n,2])]
$coordinates[,1]==17 & dataAGC[,(tocompare[n,2])]$coordinates[,
2]>=((start-((end-start)/segments))+(m*((end-start)/segments))) &
```

```
dataAGC[,(tocompare[n,2])]$coordinates[,3]<=(start+(m*((end-start)/
segments))),1]$ratios,dataAGC[,(tocompare[n,3])][dataAGC[,
(tocompare[n,3])]$coordinates[,1]==17 & dataAGC[,(tocompare[n,3])]
$coordinates[,2]>=((start-((end-start)/segments))+(m*((end-start)/
segments))) & dataAGC[,(tocompare[n,3])]$coordinates[,3]<=(start+
(m*((end-start)/segments))),1]$ratios, method="spearman")$estimate
    return(AGCmatrix)
}#end of function
```

8.6 Appendix 6: The AGCplot function

Version 1.1 15/09/14
Role: Plots AGC plots
Requirements: Data in vector AGCmatrix generated by AGCcalc function
Arguments start, end, segments and comparisons must be the same as used to generate data with the AGCcalc function

Examples:
ACGplot()
AGCplot(repositionGC=0.1,title=F,titlelines=1,GC=T,legend=T,legcex=0.8,conditional=F)

```
AGCplot<-function(start=10000000,end=15000000,segments=50,
comparisons=1,GC=TRUE,conditional=TRUE,repositionGC=0,title=FALSE,
titlelines=1,legend=FALSE,legpos="bottomright",legcex=1){

options(scipen=999)#remove scientific notation from axes
lowpoints=19
if(conditional)lowpoints<-1

#plots
#plot comparions 1a vs 1b as inital plot

par(mar=c(4,4,4,4))
  plot(as.numeric(AGCmatrix[1,]),
     ,type="o",lty=1, lwd=1,
     ylab="",xlab="",
     ylim=c(0,1),
     axes=FALSE,
     col=1,
     pch=ifelse((AGCmatrix[(comparisons+1),]>mean(AGCmatrix[(comparis
     ons+1),])),19,lowpoints))

for (n in 1:comparisons)
abline(h=mean(AGCmatrix[n,]),lty=3,col=n,lwd=1)

#overlie plots from additional comparisons
if (comparisons>1) {for (n in 1:(comparisons-1)) points(corseq[(n
     +1),], lty=1, col=(n+1),
     pch=ifelse((corseq[(comparisons+1),]>mean(corseq[(comparisons
     +1),])),19,lowpoints))}

if (comparisons>1) {for (n in 1:(comparisons-1)) lines(corseq[(n
     +1),], lty=1, col=(n+1))}

#add axis and labels
#x-axis
axis(1, at=c(1,(0.2*segments),(0.4*segments),(0.6*segments),
(0.8*segments),segments), lab=c(start,(start+(0.2*(end-start))),
(start+(0.4*(end-start))),(start+(0.6*(end-start))),(start
+(0.8*(end-start))),end))
mtext("Chromosome 17 Position", side=1, line=3)

#add left y axis
axis(2, at=c(0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1), las=2)
```

```
mtext("Spearman's rank correlation", side=2, line=3, col="black")

#add GC content if argument GC=TRUE is used
legcex=1){  if(GC){
    par(new=TRUE) #set par to overlay previous plot
    #plot window GC content
    plot(as.numeric(AGCmatrix[(comparisons+1),]),
     type="l",col="darkorchid4",lty=1,lwd=2,ylim=c((0.2+reposition
     GC),(0.7+repositionGC)),ylab="",xlab="",axes=FALSE)

for (n in 1:(segments-1)) polygon(c(n,n,(n+1),(n
+1)),c(0.3,AGCmatrix[(comparisons+1),n],AGCmatrix[(comparisons+1),(n
+1)],0.3), col="grey", border=NA)

#add average line
    abline(h=mean(AGCmatrix[comparisons+1,]), col="grey",lty=3)
    #add axis - right Y axis
    axis(4, col="darkorchid4", col.axis="darkorchid4",
     at=c(0.3,0.4,0.5), las=2, lwd=2)
     mtext("GC Content", side=4, line=3, col="darkorchid4", adj=(0.4-
     repositionGC), lwd=2)}

#add user defined title if argument title=TRUE is used, and for
titlelines number of lines

if(title){
    for (n in 1:titlelines){
      print(c("For line",n))
      usertitle<-readline("Please enter plot title .....")
      title(main=usertitle,line=(titlelines-n))
}}

#add legend if argument legend=TRUE is used
  if(legend){
    legmatrix<-matrix(nrow=2*comparisons,ncol=1)
    legmatrixpoints<-matrix(nrow=2*comparisons,ncol=4)
    for(n in 1:comparisons){
      print(c("Add legend details for comparison",n))
      legmatrix[(2*n-1),1]<-readline("Please enter comparison
         details ...")
      legmatrix[(2*n),1]<-paste(legmatrix[(2*n-1),1],"- Mean Dataset
         Correlation")
      legmatrix[(2*n-1),1]<-paste(legmatrix[(2*n-1),1],"- Window
         Correlation Value")
      legmatrixpoints[c((2*n-1),2*n),1]<-n

     legmatrixpoints[(2*n-1),2]<-19
     legmatrixpoints[(2*n),2]<-NA
     legmatrixpoints[(2*n-1),3]<-1
     legmatrixpoints[(2*n),3]<-3
     legmatrixpoints[c((2*n-1),2*n),4]<-2

legGCmatrix<-matrix(nrow=2, ncol=1)
      legGCmatrixpoints<-matrix(nrow=2,ncol=4)
      legGCmatrix[1,]<-"Window GC Content"
      legGCmatrix[2,]<-"Mean GC Content"
      legGCmatrixpoints[1,]<-c(8,NA,1,4)
      legGCmatrixpoints[2,]<-c(8,NA,3,2)

legmatrix<-rbind(legmatrix,legGCmatrix)
legmatrixpoints<-rbind(legmatrixpoints,legGCmatrixpoints)

legend(cex=legcex,x=legpos,box.lwd=2, box.lty=6,
```

```
          legend=(legmatrix[,1]),
          col=(legmatrixpoints[,1]),
          pch=(legmatrixpoints[,2]),
          lty=(legmatrixpoints[,3]),
          lwd=(legmatrixpoints[,4]))}}

}#end of function
```

8.7 Appendix 7: The outlierDataPlot function

Version 1.1: 11/06/14
Role: Plots outlier probes locations between two DIP-chip datasets and QC plots
Requirements: Datasets in vector data

Examples
outlierDataPlots(1,2,sameaxis=FALSE)
outlierDataPlots(1,2,csv=TRUE)

```
outlierDataPlots<-function(dataset1=FALSE, dataset2=FALSE,
sameaxis=FALSE, csv=FALSE, CSVtitle="outlierprobes.csv"){

#remove arguments if inconsistent
if(!dataset1) stop ("Need identifying value for dataset1 for
comparison")
if(!dataset2) stop ("Need identifying value for dataset2 for
comparison")
if(!is.logical(sameaxis)) stop ("Waterfall plot must be either on
same axis (sameaxis=TRUE) or separate axis (sameaxis=FALSE)")
if(!is.logical(csv)) stop ("Specify whether to produce csv file of
outlier probes cvs=TRUE or csv=FALSE")

#subtract dataset2 from dataset1 to give difference of IP/IN ratio
#values at each probe
subvalues<-data[,dataset1]-data[,dataset2]

#remove NAs (non matching probes) as cannot be put through outlier
#detection programme
sv<-subvalues[!is.na(subvalues$ratios[,1]),]

#get extreme values package, run to generate list of outliers, and
#programme calculated upper and lower limits
  require(extremevalues)
  outliers<-
getOutliers(sv$ratios[,1],method="I",distribution="normal")
  outlierlist<-c(outliers$iRight,outliers$iLeft)
  limits<-matrix(as.numeric(outliers$limit))

#calculations for rainfall plot
#calculating for if both postive and negative subtracted values to
#be displayed on same axis with all values negative
  outlierprobematrix<-matrix(unlist(outlierlist))

proberanktable<-matrix(ncol=3, nrow=dim(outlierprobematrix))
proberanktable[,1]<-outlierprobematrix
proberanktable[,2]<-((sv[outlierlist,1]$coordinates[,
2]+sv[outlierlist,1]$coordinates[,3])/2)
proberanktable[,3]<-sv[outlierlist,1]$ratios

colnames(proberanktable)<-
c("outlierprobe","midprobecoord","subtracted value")
proberanktable_posandneg<-proberanktable[order(-proberanktable[,
3]
proberanktable_all<-proberanktable
proberanktable_all[,3]<-(sqrt(proberanktable[,3]*proberanktable[,
3]))
```

```
proberanktable_all[,3]<-(proberanktable_all[,3]-
(2*proberanktable_all[,3]))
proberanktable_all<-proberanktable_all[order(-proberanktable_all[,
3]),]

#plots
#top left - qqplot with outliers
par(fig=c(0.01,0.4975,0.505,0.99
outlierPlot(sv$ratios[,1],outliers,mode="qq")
abline(h=limits[1,1], col="purple4", lty=2, lwd=2)
abline(h=limits[2,1], col="maroon3", lty=2, lwd=2)
box(which="figure")


# top right - hstogram with upper and lower limits
par(fig=c(0.5025,0.99,0.505,0.99), new=TRUE)
hist(sv, col="skyblue", breaks=100,
       main=title(" Distribution of Probe Values"),
       xlab="Probe Values after Subtraction")
  abline(v=limits[1,1], lty=2, col="purple4", lwd=2)
  abline(v=limits[2,1], lty=2, col="maroon3", lwd=2
  box(which="figure)

#draw lower box for plot and key
par(fig=c(0.01,0.99,0.01,0.5), new=TRUE)
box(which="figure")
#make either single waterfall or split by pos/neg outliers
#if on same axis (argument sameaxis=TRUE)
#bottom row - rainfall plot

if(sameaxis){
    par(fig=c(0.01,0.85,0.01,0.5), new=TRUE)
    plot(proberanktable_all[,2],proberanktable_all[,3], pch=16,
     cex=0.5,
     col=rainbow(1.1*(nrow(outlierprobematrix))),
     xlab="Probe Position", ylab="Difference in IN/IP ratio
     values",
     main=title("Array position and Difference in IP/IN ratio of
     outlying probes \n (Probe values calculated as distance from \n
     mean of all subtracted probe values)",
     cex.main=1, font.main=1),
     axes=FALSE)
axis(1, at=c(seq(10000000,15000000,1000000)))
axis(2, at=c(seq((ceiling(max(proberanktable_all[,3]))),
(ceiling(min(proberanktable_all[,3]))-1),-1)),
lab=c(seq((-ceiling(max(proberanktable_all[,3]))),(-
(ceiling(min(proberanktable_all[,3]))-1)),1)),
         las=2,cex.axis=0.8)

#set up matrix for key plots
    z=matrix(100:1,nrow=1)
    x=1
    y=seq(0,1,length=101)
    #draw key
    par(fig=c(0.8,0.98,0,0.5), new=TRUE)
image(x,y,z,col=rainbow(dim(outlierprobematrix)),axes=FALSE,ylab="",
xlab="", ylim=c(0.1,1))
axis(1, at=1,lab="Maximum", cex.axis=0.8,tck=0,line=-1)
axis(3,at=1,lab="Minimum",cex.axis=0.8,tck=0,line=-1)
mtext("Difference in IP/IN ratio values \n of outlying probes",
2,line=0.5, cex=0.8)
box()
```

```
}

#if all data plotted above and below central data
  if (!sameaxis){
    #bottom row - rainfall up and down pos and neg
    par(fig=c(0.01,0.85,0.01,0.5), new=TRUE)
    plot(proberanktable_posandneg[,2],proberanktable_posandneg[,3],
     pch=16, cex=0.5,
     col=rainbow(1.1*(nrow(proberanktable_posandneg))),
     xlab="Probe Position", ylab="Difference in IN/IP ratio
     values",main=title("Array position and Difference in IP/IN ratio
of
     outlying probes"),
     axes=FALSE,
     ylim=c((ceiling(min(proberanktable_posandneg[,3]))-1),
     (ceiling(max(proberanktable_posandneg[,3])))))
     segments(10000000,limits[2,1],15000000,limits[2,1], lty=2,
     lwd=2, col="maroon3")
     segments(10000000,limits[1,1],15000000,limits[1,1], lty=2,
     lwd=2, col="purple4")
     axissize<-max(proberanktable_posandneg[,3])-
     min(proberanktable_posandneg[,3])

if(axissize>10) steps=2 else steps=1
     axis(1, at=c(seq(10000000,15000000,1000000)))
     axis(2, at=c(seq((ceiling(min(proberanktable_posandneg[,3]))-1),
     (ceiling(max(proberanktable_posandneg[,3]))),steps)),
     ,las=2, cex.axis=0.8,
     lab=c(seq((ceiling(min(proberanktable_posandneg[,3]))-1),
     (ceiling(max(proberanktable_posandneg[,3]))),steps)))

#set up matrix for key plots
    z=matrix(100:1,nrow=1)
    x=1
    y=seq(0,1,length=101)
    #draw key
    par(fig=c(0.8,0.98,0,0.5), new=TRUE)
image(x,y,z,col=rainbow(dim(outlierprobematrix)),axes=FALSE,ylab="",
xlab="", ylim=c(0.1,1))
    axis(1, at=1,lab="Most Negative", cex.axis=0.8,tck=0,line=-1)
    axis(3,at=1,lab="Most Positive",cex.axis=0.8,tck=0,line=-1)
    mtext("Difference in IP/IN ratio values \n of outlying probes",
2,line=0.5, cex=0.8)
    box()}
  if(csv){
    write.csv(proberanktable_posandneg, file=CSVtitle)}
  } #end function
```

8.8 Appendix 8: The outlierDataMultiPlot function

Version 1.0 14/09/14
Role: Compare outlier probe positions between datasets

Example:
outlierDataMultiPlot(3, plotsame=F, title=T)

```
outlierDataMultiPlot<-function(comparisons, plotsame=TRUE,
title=FALSE, csv=FALSE, CSVtitle="outlierprobes.csv"){

require(extremevalues)
options(scipen=999)
#choose which datasets in data to compare

tocompare<-matrix(ncol=5,nrow=comparisons)
  for (n in 1:comparisons) tocompare[n,1]<-n
  print("Enter datasets to copmare in order of desired comparisons
     i.e. 1a vs 1b, 2a vs 2b...")
  for (n in 1:comparisons){
  print("For comparison")
  print(n)
    tocompare[n,2]<-as.numeric(readline("Enter dataset
'a' (identified using numerical order in vector 'data')..."))
    tocompare[n,3]<-as.numeric(readline("Enter dataset
'b' (identified using numerical order in vector 'data')..."))

    tmpsv<-data[,(tocompare[n,2])]-data[,(tocompare[n,3])]
    tmpsvna<-tmpsv[!is.na(tmpsv$ratios[,1]),]
    tmpoutliers<-getOutliers(tmpsvna$ratios[,1],method="I",
     distribution="normal")
    tmpoutlierlist<-c(tmpoutliers$iRight,tmpoutliers$iLeft)
    tolm<-matrix(unlist(tmpoutlierlist))
    tocompare[n,5]<-nrow(tolm)
}

#make large matrix of 1)outlier probe, 2) midpoint 3)subtracted
€probe value for each comparison
    proberanktable<-matrix(ncol=3,nrow=(max(tocompare[,5])))
    masterproberanktable<-matrix(ncol=3,nrow=(max(tocompare[,5])))
    for (n in 1:comparisons){
      tmpsv<-data[,(tocompare[n,2])]-data[,(tocompare[n,3])]
      tmpsvna<-tmpsv[!is.na(tmpsv$ratios[,1]),]
      tmpoutliers<-getOutliers(tmpsvna$ratios[,1],method="I",
         distribution="normal")
      tmpoutlierlist<-c(tmpoutliers$iRight,tmpoutliers$iLeft)
      tolm<-matrix(unlist(tmpoutlierlist))
      proberanktable[(1:(nrow(tolm))),1]<-tolm
      proberanktable[(1:(nrow(tolm))),2]<-((tmpsvna[tmpoutlierlist,
     1]$coordinates[,2]+tmpsvna[tmpoutlierlist,1]$coordinates[,3])/2)
      proberanktable[(1:(nrow(tolm))),3]<-tmpsvna[tmpoutlierlist,
     1]$ratios

if(n<2){masterproberanktable<-proberanktable} else
masterproberanktable<-cbind(masterproberanktable,proberanktable)
}

#plots
```

```
  if(plotsame){
par(fig=c(0,0.9,0,1),new=FALSE)
plot(masterproberanktable[,2],masterproberanktable[,3], pch=16,
         cex=0.5,
         #ylim=c(min(masterproberanktable[,c(3*(seq(1,n,
         1)))]),max(masterproberanktable[,c(3*(seq(1,n,1)))])),
         xlab="Probe Position", ylab="Difference in IN/IP ratio
         values")

for (n in 1:(comparisons-1)){
    par(new=TRUE)
    plot(masterproberanktable[,(3*n+2)],masterproberanktable[,3*n
         +3],
         #ylim=c(min(masterproberanktable[,c(3*(seq(1,n,
         1)))]),max(masterproberanktable[,c(3*(seq(1,n,1)))])),
         pch=16, cex=0.5, col=n+1, axes=FALSE, xlab="", ylab="")
         }}

if(!plotsame){
    titlemat<-matrix(nrow=comparisons,ncol=1)
    if(title){
        for (n in 1:comparisons){
          print("For plot")
          print(n)
          titlemat[n,]<-readline("Enter title....")
}}
    maxmat<-matrix(apply(masterproberanktable,2,max,na.rm=TRUE))
    minmat<-matrix(apply(masterproberanktable,2,min,na.rm=TRUE))

par(fig=c(0,0.9,(1-1/comparisons),1),new=FALSE)
     plot(masterproberanktable[,2],masterproberanktable[,3], pch=16,
     cex=0.5,
     axes=FALSE,xlab="",ylab="",
     main=title(titlemat[1,]))
     axis(1)
     axis(2,at=c((maxmat[3,1]/1.3),(minmat[3,1]/
     1.3)),lab=c("higher","lower"))
     mtext("Chromosome 17 Position", side=1, line=3)
     mtext("Difference in IN/IP ratio values",side=2, line=2)
     box(which="figure")

for (n in 1:(comparisons-1)){
     par(fig=c(0,0.9,(1-((n+1)*(1/comparisons))),(1-(n*(1/
     comparisons)))),new=TRUE)

     plot(masterproberanktable[,(3*n+2)],masterproberanktable[,3*n
          +3],
          axes=FALSE,xlab="",ylab="",
          pch=16, cex=0.5, col=n+1,
          main=title(titlemat[(n+1),]))
axis(1)
axis(2,at=c((maxmat[3*n,1]/1.2),(minmat[3*n,1]/
1.2)),lab=c("higher","lower"))
mtext("Chromosome 17 Position", side=1, line=3)
mtext("Difference in IN/IP ratio values",side=2, line=2)
box(which="figure")
}}
  if(csv){
    write.csv(masterproberanktable, file=CSVtitle)}
} #end function
```

8.9 Appendix 9: *RAD1* Primers

RAD1 position	Strand	Name	Sequence (5' to 3')
-544 to -514	(US)	BamH1	CCATAGGATCCCCAAGTAACCGTGATAAACA
97 to 118	(US)	n69-US	AAATGAAGATGAACCCGACGAC
119 to 98	(LS)	n69-LS	GTCGTCGGGTTCATCTTCATTT
673 to 694	(US)	n672-US	CTTATCCGGCATTGTTCACCCA
1118 to 1142	(US)	Check-F	TCAGGTCAATTGACTCGGTGATGG
1150 to 1171	(US)	Rad1-F	AACTGGCACCGAATTTCTTATGAA
1393 to 1420	(US)	rad1(P379S)-F	TTTAGAAGAAAATTCAAAATGGGAACAA
1420 to 1393	(LS)	rad1(P379S)-R	TTGTTCCCATTTTGAATTTTCTTCTAAA
1870 to 1879	(US)	Rad1(S613X)-Seq	AAATGCCGCAAACGATTCAA
1879 to 1870	(LS)	Rad1(P379S)-Seq	TTGAATCGTTTGCGGCATTT
2185 to 2206	(US)	n2184 US	TAAGGCCATAGTGAAGGATTTG
2206 to 2185	(LS)	n2184 LS	CAAATCCTTCACTATGGCCTTA
2224 to 2259	(US)	rad1(S613X)-F	ATGTACTACGGTGAATGAATTGAAGAGCAAAGTCA
2259 to 2224	(LS)	rad1(S613X)-R	TGACTTTGCTCTTCAATTCATTCACCGTAGTACAT
2562 to 2538	(LS)	Rad1-R	TAATCGCCGACTGTCAACATACAA
2641 to 2614	(LS)	Check-R	GGCTAATCTGTTATTCTGTAATGACCCA
2783 to 2801	(US)	n2781-US	TCAAGCAAGTTATCCCAGGATG
3560 to 3515	(LS)	Rad1-Sal1-L	GGTCATGTCGACTGCTAAAAAGTGGAAGATGAATTG

8.10 Appendix 10: Content of the Electronic Appendix

Chapter 3

A folder labelled Chapter 3, containing an Excel spread sheet with raw data for each figure, separated by workbook.

Chapter 4

A folder labelled Chapter 4, containing Sandcastle version 1.0 R package, and microarray datasets separated for each relevant figure

Chapter 5

A folder labelled Chapter 5 containing relevant microarray datasets separated for each figure and a sub-folder containing custom R scripts as described in the chapter and the above appendices. An additional subfolder contains several chapter 5 figures in higher resolution.

Chapter 6

A folder labelled Chapter 6, contain an Excel spread sheet with all raw data for the relevant figures, separated by workbook. Additionally, sequencing data for the entirety of each plasmid is included in the sub-folder labelled sequencing data.

References

Aboussekhra, A. et al. 1995. Mammalian DNA nucleotide excision repair reconstituted with purified protein components. *Cell* 80(6), pp. 859-868.

Adams, R. A. et al. 2011. Intermittent versus continuous oxaliplatin and fluoropyrimidine combination chemotherapy for first-line treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. *Lancet Oncol* 12(7), pp. 642-653.

Adelsberger, H. et al. 2000. The chemotherapeutic oxaliplatin alters voltage-gated Na(+) channel kinetics on rat sensory neurons. *European journal of pharmacology* 406(1), pp. 25-32.

Adriaens, M. E. et al. 2012. An evaluation of two-channel ChIP-on-chip and DNA methylation microarray normalization strategies. *BMC Genomics* 13, p. 42.

Ahmad, A. et al. 2010. Mislocalization of XPF-ERCC1 nuclease contributes to reduced DNA repair in XP-F patients. *PLoS genetics* 6(3), p. e1000871.

Ahmad, A. et al. 2008. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. *Molecular and cellular biology* 28(16), pp. 5082-5092.

Aird, D. et al. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol* 12(2), p. R18.

Akaboshi, M. et al. 1992. The number of platinum atoms binding to DNA, RNA and protein molecules of HeLa cells treated with cisplatin at its mean lethal concentration. *Jpn J Cancer Res* 83(5), pp. 522-526.

Akaboshi, M. et al. 1994. Binding characteristics of (-)-(R)-2-aminomethylpyrrolidine(1,1-cyclobutanedicarboxylato)-2-platin um(II) to DNA, RNA and protein molecules in HeLa cells and its lethal effect: comparison with cis- and trans-diamminedichloroplatinums(II). *Jpn J Cancer Res* 85(1), pp. 106-111.

Al-Minawi, A. Z. et al. 2009. The ERCC1/XPF endonuclease is required for completion of homologous recombination at DNA replication forks stalled by inter-strand cross-links. *Nucleic Acids Res* 37(19), pp. 6400-6413.

Alcindor, T. and Beauger, N. 2011. Oxaliplatin: a review in the era of molecularly targeted therapy. *Current oncology* 18(1), pp. 18-25.

Alejandro, L. M. et al. 2013. Predicting acute and persistent neuropathy associated with oxaliplatin. *Am J Clin Oncol* 36(4), pp. 331-337.

Alemar, J. and Schuur, E. R. 2013. Progress in using circulating tumor cell information to improve metastatic breast cancer therapy. *J Oncol* 2013, p. 702732.

Alexandrov, L. B. et al. 2013a. Signatures of mutational processes in human cancer. *Nature*.

Alexandrov, L. B. et al. 2013b. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep* 3(1), pp. 246-259.

Ali, B. H. 2010. Amelioration of oxaliplatin neurotoxicity by drugs in humans and experimental animals: a minireview of recent literature. *Basic & Clinical Pharmacology & Toxicology* 106(4), pp. 272-279.

Allain, P. et al. 2000. Early biotransformations of oxaliplatin after its intravenous administration to cancer patients. *Drug metabolism and disposition: the biological fate of chemicals* 28(11), pp. 1379-1384.

Allegra, C. J. et al. 2009. Initial safety report of NSABP C-08: A randomized phase III study of modified FOLFOX6 with or without bevacizumab for the adjuvant treatment of patients with stage II or III colon cancer. *J Clin Oncol* 27(20), pp. 3385-3390.

Almeida, G. M. et al. 2006. Detection of oxaliplatin-induced DNA crosslinks in vitro and in cancer patients using the alkaline comet assay. *DNA Repair (Amst)* 5(2), pp. 219-225.

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. *Journal of Clinical Oncology* 27(19), pp. 3109-3116.

Anindya, R. et al. 2010. A ubiquitin-binding domain in Cockayne syndrome B required for transcription-coupled nucleotide excision repair. *Mol Cell* 38(5), pp. 637-648.

Araki, M. et al. 2001. Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair. *J Biol Chem* 276(22), pp. 18665-18672.

Aravind, L. et al. 1999. Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res* 27(5), pp. 1223-1242.

Argyriou, A. A. et al. 2012a. Chemotherapy-induced peripheral neurotoxicity (CIPN): An update. *Crit Rev Oncol Hematol* 82(1), pp. 51-77.

Argyriou, A. A. et al. 2013. Clinical pattern and associations of oxaliplatin acute neurotoxicity: A prospective study in 170 patients with colorectal cancer. *Cancer* 119(2), pp. 438-444.

Argyriou, A. A. et al. 2014. Chemotherapy-induced peripheral neuropathy in adults: a comprehensive update of the literature. *Cancer Manag Res* 6, pp. 135-147.

Argyriou, A. A. et al. 2007. The usefulness of nerve conduction studies in objectively assessing oxaliplatin-induced peripheral neuropathy. *Oncologist* 12(11), pp. 1371-1372; author reply 1372-1373.

Argyriou, A. A. et al. 2008. A review on oxaliplatin-induced peripheral nerve damage. *Cancer Treatment Reviews* 34(4), pp. 368-377.

Argyriou, A. A. et al. 2012b. Peripheral neurotoxicity of oxaliplatin in combination with 5-fluorouracil (FOLFOX) or capecitabine (XELOX): a prospective evaluation of 150 colorectal cancer patients. *Ann Oncol* 23(12), pp. 3116-3122.

Arnesano, F. et al. 2013. An Updated View of Cisplatin Transport. *European Journal of Inorganic Chemistry* 2013(15), pp. 2701-2711.

Attard, C. L. et al. 2010. Cost-effectiveness of oxaliplatin in the adjuvant treatment of colon cancer in Canada. *Current oncology* 17(1), pp. 17-24.

Bahassi, E. M. and Stambrook, P. J. 2014. Next-generation sequencing technologies: breaking the sound barrier of human genetics. *Mutagenesis* 29(5), pp. 303-310.

Bailly, V. et al. 1992. Specific complex formation between proteins encoded by the yeast DNA repair and recombination genes RAD1 and RAD10. *Proc Natl Acad Sci U S A* 89(17), pp. 8273-8277.

Balmer, C. E. et al. 2001. Who wants second-line, palliative chemotherapy? *Psychooncology* 10(5), pp. 410-418.

Bancroft, D. P. et al. 1990. ^{195}PtNMR Kinetic and Mechanistic Studies of cis-and trans-Diamminedichloroplatinum(II) Binding to DNA. *J Am Chem Soc* (112), pp. 6860-6871.

Bardwell, A. J. et al. 1993. Yeast DNA recombination and repair proteins Rad1 and Rad10 constitute a complex in vivo mediated by localized hydrophobic domains. *Mol Microbiol* 8(6), pp. 1177-1188.

Bardwell, A. J. et al. 1994. Specific cleavage of model recombination and repair intermediates by the yeast Rad1-Rad10 DNA endonuclease. *Science* 265(5181), pp. 2082-2085.

Bardwell, L. et al. 1992. Stable and specific association between the yeast recombination and DNA repair proteins RAD1 and RAD10 in vitro. *Mol Cell Biol* 12(7), pp. 3041-3049.

Barnett, V. and Lewis, T. 1994. *Outliers in statistical data*. 3rd Edition ed. Chichester: John Wiley and Sons, p. 584.

Bellacosa, A. 2001. Functional interactions and signaling properties of mammalian DNA mismatch repair proteins. *Cell Death Differ* 8(11), pp. 1076-1092.

Bennett, M. R. 2013. *Integrative analysis of chip-on-chip datasets in Saccharomyces cerevisiae.* PhD, Cardiff University.

Berger, A. et al. 2004. Clinical characteristics and economic costs of patients with painful neuropathic disorders. *J Pain* 5(3), pp. 143-149.

Berneburg, M. et al. 2000. The cancer-free phenotype in trichothiodystrophy is unrelated to its repair defect. *Cancer Res* 60(2), pp. 431-438.

Bernier, J. et al. 2004. Postoperative irradiation with or without concomitant chemotherapy for locally advanced head and neck cancer. *N Engl J Med* 350(19), pp. 1945-1952.

Besse, B. et al. 2013. ERCC1 and RRM1: ready for prime time? *J Clin Oncol* 31(8), pp. 1050-1060.

Bhagwat, N. et al. 2009. XPF-ERCC1 participates in the Fanconi anemia pathway of cross-link repair. *Mol Cell Biol* 29(24), pp. 6427-6437.

Biggerstaff, M. et al. 1993. Co-correction of the ERCC1, ERCC4 and xeroderma pigmentosum group F DNA repair defects in vitro. *The EMBO journal* 12(9), pp. 3685-3692.

Binks, S. P. and Dobrota, M. 1990. Kinetics and mechanism of uptake of platinum-based pharmaceuticals by the rat small intestine. *Biochem Pharmacol* 40(6), pp. 1329-1336.

Blanc, C. et al. 2000. Caspase-3 is essential for procaspase-9 processing and cisplatin-induced apoptosis of MCF-7 breast cancer cells. *Cancer Res* 60(16), pp. 4386-4390.

Blommaert, F. A. et al. 1998. The formation and repair of cisplatin-DNA adducts in wild-type and cisplatin-resistant L1210 cells: comparison of immunocytochemical determination with detection in isolated DNA. *Chem Biol Interact* 108(3), pp. 209-225.

Blommaert, F. A. et al. 1993. Drug-induced DNA modification in buccal cells of cancer patients receiving carboplatin and cisplatin combination chemotherapy, as determined by an immunocytochemical method: interindividual variation and correlation with disease response. *Cancer research* 53(23), pp. 5669-5675.

Blommaert, F. A. and Saris, C. P. 1995. Detection of platinum-DNA adducts by 32P-postlabelling. *Nucleic Acids Research* 23(8), pp. 1300-1306.

Bodenner, D. L. et al. 1986. Effect of diethyldithiocarbamate on cis-diamminedichloroplatinum(II)-induced cytotoxicity, DNA cross-linking, and gamma-glutamyl transpeptidase inhibition. *Cancer Res* 46(6), pp. 2745-2750.

Boffetta, P. et al. 1998. Cisplatin-DNA adducts and protein-bound platinum in blood of testicular cancer patients. *Anticancer Drugs* 9(2), pp. 125-129.

Boige, V. et al. 2010. Pharmacogenetic assessment of toxicity and outcome in patients with metastatic colorectal cancer treated with LV5FU2, FOLFOX, and FOLFIRI: FFCD 2000-05. *J Clin Oncol* 28(15), pp. 2556-2564.

Bonetti, A. et al. 1996. Inductively coupled plasma mass spectroscopy quantitation of platinum-DNA adducts in peripheral blood leukocytes of patients receiving cisplatin- or carboplatin-based chemotherapy. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2(11), pp. 1829-1835.

Bonner, J. et al. 2006. Radiotherapy plus Cetuximab for Squamous-Cell Carcinoma of the Head and Neck. *N Engl J Med* 354(6), pp. 567-578.

Botstein, D. and Fink, G. R. 2011. Yeast: an experimental organism for 21st Century biology. *Genetics* 189(3), pp. 695-704.

Bowden, N. A. 2014. Nucleotide excision repair: why is it not used to predict response to platinum-based chemotherapy? *Cancer Lett* 346(2), pp. 163-171.

Bracker, T. U. et al. 2006. Stringent regulation of DNA repair during human hematopoietic differentiation: a gene expression and functional analysis. *Stem Cells* 24(3), pp. 722-730.

Brookman, K. W. et al. 1996. ERCC4 (XPF) encodes a human nucleotide excision repair protein with eukaryotic recombination homologs. *Mol Cell Biol* 16(11), pp. 6553-6562.

Brouwer, J. et al. 1981. Base-pair substitution hotspots in GAG and GCG nucleotide sequences in Escherichia coli K-12 induced by cis-diamminedichloroplatinum (II). *Proc Natl Acad Sci U S A* 78(11), pp. 7010-7014.

Brouwers, E. E. et al. 2009. Persistent neuropathy after treatment with cisplatin and oxaliplatin. *Acta oncologica* 48(6), pp. 832-841.

Bruhn, S. L. et al. 1990. Biological Processing of DNA modified by Platinum Compounds. *Prog Inorg Chem* (38), pp. 477-516.

Bubley, G. J. et al. 1996. Effect of DNA conformation on cisplatin adduct formation. *Biochem Pharmacol* 51(5), pp. 717-721.

Burchenal, J. H. et al. 1979. Rationale for development of platinum analogs. *Cancer Treat Rep* 63(9-10), pp. 1493-1498.

Burd, E. M. 2010. Validation of laboratory-developed molecular assays for infectious diseases. *Clin Microbiol Rev* 23(3), pp. 550-576.

Burger, H. et al. 2010. Differential transport of platinum compounds by the human organic cation transporter hOCT2 (hSLC22A2). *Br J Pharmacol* 159(4), pp. 898-908.

Burns, M. J. et al. 2005. Standardisation of data from real-time quantitative PCR methods - evaluation of outliers and comparison of calibration curves. *BMC Biotechnol* 5, p. 31.

Burz, C. et al. 2009. Clinical and pharmacokinetics study of oxaliplatin in colon cancer patients. *Journal of gastrointestinal and liver diseases : JGLD* 18(1), pp. 39-43.

Bustin, S. A. 2010. Why the need for qPCR publication guidelines?--The case for MIQE. *Methods* 50(4), pp. 217-226.

Bustin, S. A. et al. 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical chemistry* 55(4), pp. 611-622.

Cassidy, J. et al. 2004. XELOX (capecitabine plus oxaliplatin): active first-line therapy for patients with metastatic colorectal cancer. *J Clin Oncol* 22(11), pp. 2084-2091.

Cavaletti, G. et al. 2011. Chemotherapy-induced peripheral neurotoxicity in the era of pharmacogenomics. *Lancet Oncol* 12(12), pp. 1151-1161.

Cavaletti, G. et al. 2013. The chemotherapy-induced peripheral neuropathy outcome measures standardization study: from consensus to the first validity and reliability findings. *Ann Oncol* 24(2), pp. 454-462.

Cavaletti, G. et al. 2001. Effects of different schedules of oxaliplatin treatment on the peripheral nervous system of the rat. *Eur J Cancer* 37(18), pp. 2457-2463.

Chaney, S. G. et al. 2005. Recognition and processing of cisplatin- and oxaliplatin-DNA adducts. *Crit Rev Oncol Hematol* 53(1), pp. 3-11.

Chaney, S. G. and Sancar, A. 1996. DNA repair: enzymatic mechanisms and relevance to drug response. *Journal of the National Cancer Institute* 88(19), pp. 1346-1360.

Chao, C. C. et al. 1994. Use of a monoclonal antibody to detect DNA damage caused by the anticancer drug cis-diamminedichloroplatinum (II) in vivo and in vitro. *FEBS letters* 354(1), pp. 103-109.

Chen, R. et al. 2014. The general amino acid control pathway regulates mTOR and autophagy during serum/glutamine starvation. *J Cell Biol* 206(2), pp. 173-182.

Chijiwa, S. et al. 2010. Polymerization by DNA polymerase eta is blocked by cis-diamminedichloroplatinum(II) 1,3-d(GpTpG) cross-link: implications for cytotoxic effects in nucleotide excision repair-negative tumor cells. *Carcinogenesis* 31(3), pp. 388-393.

Christians, F. C. and Hanawalt, P. C. 1992. Inhibition of transcription and strand-specific DNA repair by alpha-amanitin in Chinese hamster ovary cells. *Mutat Res* 274(2), pp. 93-101.

Chu, G. and Chang, E. 1988. Xeroderma pigmentosum group E cells lack a nuclear factor that binds to damaged DNA. *Science* 242(4878), pp. 564-567.

Chua, W. et al. 2009. Molecular markers of response and toxicity to FOLFOX chemotherapy in metastatic colorectal cancer. *British Journal of Cancer* 101(6), pp. 998-1004.

Ciarimboli, G. et al. 2005. Cisplatin nephrotoxicity is critically mediated via the human organic cation transporter 2. *Am J Pathol* 167(6), pp. 1477-1484.

Clarke, L. and Carbon, J. 1980. Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature* 287(5782), pp. 504-509.

Cobo, M. et al. 2007. Customizing cisplatin based on quantitative excision repair cross-complementing 1 mRNA expression: a phase III trial in non-small-cell lung cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 25(19), pp. 2747-2754.

Collins, A. R. 2004. The comet assay for DNA damage and repair: principles, applications, and limitations. *Mol Biotechnol* 26(3), pp. 249-261.

Connors, T. A. et al. 1972. New platinum complexes with anti-tumour activity. *Chem Biol Interact* 5(6), pp. 415-424.

Contopoulou, C. R. et al. 1987. Analysis of DNA double strand breakage and repair using orthogonal field alternation gel electrophoresis. *Yeast* 3(2), pp. 71-76.

Cooper, J. et al. 2004. Postoperative Concurrent Radiotherapy and Chemotherapy for High-Risk Squamous-Cell Carcinoma of the Head and Neck. *N Engl J Med* 350(19), pp. 1937-1944.

Cortejoso, L. et al. 2013. Differential toxicity biomarkers for irinotecan- and oxaliplatin-containing chemotherapy in colorectal cancer. *Cancer chemotherapy and pharmacology* 71(6), pp. 1463-1472.

Coste, F. et al. 1999. Crystal structure of a double-stranded DNA containing a cisplatin interstrand cross-link at 1.63 A resolution: hydration at the platinated site. *Nucleic Acids Res* 27(8), pp. 1837-1846.

Cowell, I. G. et al. 2011. An overview of the visualisation and quantitation of low and high MW DNA adducts using the trapped in agarose DNA immunostaining (TARDIS) assay. *Mutagenesis* 26(2), pp. 253-260.

CRUK. 2014a. *Bowel Cancer Incidence Statistics* [Online]. Available at: http://www.cancerresearchuk.org/cancer-info/cancerstats/types/bowel/incidence/uk-bowel-cancer-incidence-statistics [Accessed: 22/07/2014 2014].

CRUK. 2014b. *Cancer Survival For Common Cancers* [Online]. Available at: http://www.cancerresearchuk.org/cancer-info/cancerstats/survival/common-cancers/ [Accessed: 22/07/2014 2014].

Cunningham, D. and Starling, N. 2007. Adjuvant chemotherapy of colorectal cancer. *The Lancet* 370(9604), pp. 1980-1981.

Custodio, A. et al. 2014. Pharmacogenetic predictors of severe peripheral neuropathy in colon cancer patients treated with oxaliplatin-based adjuvant chemotherapy: a GEMCAD group study. *Ann Oncol* 25(2), pp. 398-403.

Dabholkar, M. et al. 1994. Messenger RNA levels of XPAC and ERCC1 in ovarian cancer tissue correlate with response to platinum-based chemotherapy. *The Journal of clinical investigation* 94(2), pp. 703-708.

Dabney, J. and Meyer, M. 2012. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* 52(2), pp. 87-94.

de Gramont, A. et al. 2000. Leucovorin and fluorouracil with or without oxaliplatin as first-line treatment in advanced colorectal cancer. *J Clin Oncol* 18(16), pp. 2938-2947.

de Gramont, A. et al. 1997. Oxaliplatin with high-dose leucovorin and 5-fluorouracil 48-hour continuous infusion in pretreated metastatic colorectal cancer. *Eur J Cancer* 33(2), pp. 214-219.

de Laat, W. L. et al. 1998. DNA structural elements required for ERCC1-XPF endonuclease activity. *J Biol Chem* 273(14), pp. 7835-7842.

De Silva, I. U. et al. 2002. Defects in interstrand cross-link uncoupling do not account for the extreme sensitivity of ERCC1 and XPF cells to cisplatin. *Nucleic Acids Res* 30(17), pp. 3848-3856.

Dean, F. B. et al. 2002. Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci U S A* 99(8), pp. 5261-5266.

Deans, A. J. and West, S. C. 2011. DNA interstrand crosslink repair and cancer. *Nature reviews. Cancer* 11(7), pp. 467-480.

Di Francesco, A. M. et al. 2002. Cellular and molecular aspects of drugs of the future: oxaliplatin. *Cell Mol Life Sci* 59(11), pp. 1914-1927.

Diaz-Rubio, E. et al. 1998. Oxaliplatin as single agent in previously untreated colorectal carcinoma patients: a phase II multicentric study. *Ann Oncol* 9(1), pp. 105-108.

Dijt, F. J. et al. 1988. Formation and repair of cisplatin-induced adducts to DNA in cultured normal and repair-deficient human fibroblasts. *Cancer Res* 48(21), pp. 6058-6062.

Donahue, B. A. et al. 1996. Effects of aminofluorene and acetylaminofluorene DNA adducts on transcriptional elongation by RNA polymerase II. *J Biol Chem* 271(18), pp. 10588-10594.

Dornish, J. M. and Pettersen, E. O. 1990. Requirement of a reactive aldehyde moiety for aldehyde-mediated protection against cis-dichlorodiammineplatinum-induced cell inactivation. *Biochem Pharmacol* 39(2), pp. 309-318.

Dunham, M. J. and Fowler, D. M. 2013. Contemporary, yeast-based approaches to understanding human genetic variation. *Curr Opin Genet Dev* 23(6), pp. 658-664.

Dzagnidze, A. et al. 2007. Repair Capacity for Platinum-DNA Adducts Determines the Severity of Cisplatin-Induced Peripheral Neuropathy. *Journal of Neuroscience* 27(35), pp. 9451-9457.

Eastman, A. 1986. Reevaluation of interaction of cis-dichloro(ethylenediamine)platinum(II) with DNA. *Biochemistry* 25(13), pp. 3912-3915.

Eastman, A. 1987. The formation, isolation and characterization of DNA adducts produced by anticancer platinum complexes. *Pharmacol Ther* 34(2), pp. 155-166.

Eastman, A. et al. 1988. Characterization of bifunctional adducts produced in DNA by trans-diamminedichloroplatinum(II). *Chem Biol Interact* 67(1-2), pp. 71-80.

Ehrsson, H. et al. 2002. Pharmacokinetics of oxaliplatin in humans. *Medical oncology* 19(4), pp. 261-265.

Eisen, J. A. et al. 1995. Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. *Nucleic Acids Res* 23(14), pp. 2715-2723.

Elkind, M. M. and Chang-Liu, C. M. 1972. Repair of a DNA complex from x-irradiated Chinese hamster cells. *Int J Radiat Biol Relat Stud Phys Chem Med* 22(1), pp. 75-90.

Enoiu, M. et al. 2012. Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis. *Nucleic Acids Res* 40(18), pp. 8953-8964.

Enzlin, J. H. and Scharer, O. D. 2002. The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif. *The EMBO journal* 21(8), pp. 2045-2053.

Erixon, K. and Ahnstrom, G. 1979. Single-strand breaks in DNA during repair of UV-induced damage in normal human and xeroderma pigmentosum cells as determined by alkaline DNA unwinding and hydroxylapatite chromatography: effects of hydroxyurea, 5-fluorodeoxyuridine and 1-beta-D-arabinofuranosylcytosine on the kinetics of repair. *Mutat Res* 59(2), pp. 257-271.

Evans, D. L. et al. 1994. Differential sensitivity to the induction of apoptosis by cisplatin in proliferating and quiescent immature rat thymocytes is independent of the levels of drug accumulation and DNA adduct formation. *Cancer Res* 54(6), pp. 1596-1603.

Extra, J. M. et al. 1990. Phase I study of oxaliplatin in patients with advanced cancer. *Cancer chemotherapy and pharmacology* 25(4), pp. 299-303.

Extra, J. M. et al. 1998. Pharmacokinetics and safety profile of oxaliplatin. *Semin Oncol* 25(2 Suppl 5), pp. 13-22.

Fagbemi, A. F. et al. 2011. Regulation of endonuclease activity in human nucleotide excision repair. *DNA Repair (Amst)* 10(7), pp. 722-729.

Felsenfeld, G. and Groudine, M. 2003. Controlling the double helix. *Nature* 421(6921), pp. 448-453.

Ferrer-Vicens, I. et al. 2014. In vivo genome-wide binding of Id2 to E2F4 target genes as part of a reversible program in mice liver. *Cell Mol Life Sci* 71(18), pp. 3583-3597.

Fichtinger-Schepman, A. M. et al. 1985. Adducts of the antitumor drug cis-diamminedichloroplatinum(II) with DNA: formation, identification, and quantitation. *Biochemistry* 24(3), pp. 707-713.

Fichtinger-Schepman, A. M. et al. 1990. Kinetics of the formation and removal of cisplatin-DNA adducts in blood cells and tumor tissue of cancer patients receiving chemotherapy: comparison with in vitro adduct formation. *Cancer research* 50(24), pp. 7887-7894.

Fichtinger-Schepman, A. M. et al. 1987. cis-Diamminedichloroplatinum(II)-induced DNA adducts in peripheral leukocytes from seven cancer patients: quantitative immunochemical detection of the adduct induction and removal after a single dose of cis-diamminedichloroplatinum(II). *Cancer research* 47(11), pp. 3000-3004.

Fink, D. et al. 1996. The role of DNA mismatch repair in platinum drug resistance. *Cancer Res* 56(21), pp. 4881-4886.

Foka, M. and Paoletti, J. 1986. Interaction of cis-diamminedichloroplatinum (II) to chromatin. Specificity of the drug distribution. *Biochem Pharmacol* 35(19), pp. 3283-3291.

Friboulet, L. et al. 2013. ERCC1 isoform expression and DNA repair in non-small-cell lung cancer. *N Engl J Med* 368(12), pp. 1101-1110.

Friedberg, E., Walker, GC, Siede, W, Wood, RD, Schultz, RA, Ellenberger, T 2005. *DNA Repair And Mutagenesis*. 2nd edition ed. ASM Press.

Fulda, S. et al. 1998. Chemosensitivity of solid tumor cells in vitro is related to activation of the CD95 system. *Int J Cancer* 76(1), pp. 105-114.

Fung, C. et al. 2013. Solid tumors after chemotherapy or surgery for testicular nonseminoma: a population-based study. *J Clin Oncol* 31(30), pp. 3807-3814.

Furey, T. S. 2012. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nature reviews. Genetics* 13(12), pp. 840-852.

Furuta, T. et al. 2002. Transcription-coupled nucleotide excision repair as a determinant of cisplatin sensitivity of human cells. *Cancer research* 62(17), pp. 4899-4902.

Gaillard, P. H. and Wood, R. D. 2001. Activity of individual ERCC1 and XPF subunits in DNA nucleotide excision repair. *Nucleic Acids Res* 29(4), pp. 872-879.

Gale, G. R. et al. 1974. Antileukemic properties of dichloro(1,2-diaminocyclohexane)platinum(II). *Res Commun Chem Pathol Pharmacol* 7(3), pp. 529-538.

Galea, A. M. and Murray, V. 2002. The interaction of cisplatin and analogues with DNA in reconstituted chromatin. *Biochim Biophys Acta* 1579(2-3), pp. 142-152.

Gammie, A. E. et al. 2007. Functional characterization of pathogenic human MSH2 missense mutations in Saccharomyces cerevisiae. *Genetics* 177(2), pp. 707-721.

Gasmi, N. et al. 2014. The Switch from Fermentation to Respiration in Saccharomyces cerevisiae Is Regulated by the Ert1 Transcriptional Activator/Repressor. *Genetics*.

Gately, D. P. and Howell, S. B. 1993. Cellular accumulation of the anticancer agent cisplatin: a review. *Br J Cancer* 67(6), pp. 1171-1176.

Gavrieli, Y. et al. 1992. Identification of programmed cell death in situ via specific labeling of nuclear DNA fragmentation. *J Cell Biol* 119(3), pp. 493-501.

Gelasco, A. and Lippard, S. J. 1998. NMR solution structure of a DNA dodecamer duplex containing a cis-diammineplatinum(II) d(GpG) intrastrand cross-link, the major adduct of the anticancer drug cisplatin. *Biochemistry* 37(26), pp. 9230-9239.

Ghazal-Aswad, S. et al. 1999. Pharmacokinetically guided dose escalation of carboplatin in epithelial ovarian cancer: effect on drug-plasma AUC and peripheral blood drug-DNA adduct levels. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 10(3), pp. 329-334.

Ghezzi, D. et al. 2012. Mutations of the mitochondrial-tRNA modifier MTO1 cause hypertrophic cardiomyopathy and lactic acidosis. *Am J Hum Genet* 90(6), pp. 1079-1087.

Gietema, J. A. et al. 2000. Circulating plasma platinum more than 10 years after cisplatin treatment for testicular cancer. *Lancet* 355(9209), pp. 1075-1076.

Gietz, R. D. and Woods, R. A. 2002. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol* 350, pp. 87-96.

Gillet, L. C. and Scharer, O. D. 2006. Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chemical reviews* 106(2), pp. 253-276.

Glendenning, J. L. et al. 2010. Long-term neurologic and peripheral vascular toxicity after chemotherapy treatment of testicular cancer. *Cancer* 116(10), pp. 2322-2331.

Goekkurt, E. et al. 2009a. Pharmacogenetic analyses of a phase III trial in metastatic gastroesophageal adenocarcinoma with fluorouracil and leucovorin plus either oxaliplatin or cisplatin: a study of the arbeitsgemeinschaft internistische onkologie. *J Clin Oncol* 27(17), pp. 2863-2873.

Goekkurt, E. et al. 2009b. Pharmacogenetic analyses of hematotoxicity in advanced gastric cancer patients receiving biweekly fluorouracil, leucovorin,

oxaliplatin and docetaxel (FLOT): a translational study of the Arbeitsgemeinschaft Internistische Onkologie (AIO). *Ann Oncol* 20(3), pp. 481-485.

Goldberg, R. M. et al. 2006. Pooled analysis of safety and efficacy of oxaliplatin plus fluorouracil/leucovorin administered bimonthly in elderly patients with colorectal cancer. *J Clin Oncol* 24(25), pp. 4085-4091.

Golpour, M. et al. 2014. Human fibroblast switches to anaerobic metabolic pathway in response to serum starvation: a mimic of warburg effect. *Int J Mol Cell Med* 3(2), pp. 74-80.

Gonzalez, V. M. et al. 2001. Is cisplatin-induced cell death always produced by apoptosis? *Mol Pharmacol* 59(4), pp. 657-663.

Govan, H. L., 3rd et al. 1990. Fine-mapping of DNA damage and repair in specific genomic segments. *Nucleic Acids Res* 18(13), pp. 3823-3830.

Graham, M. A. et al. 2000. Clinical pharmacokinetics of oxaliplatin: a critical review. *Clinical cancer research : an official journal of the American Association for Cancer Research* 6(4), pp. 1205-1218.

Gregg, R. W. et al. 1992. Cisplatin neurotoxicity: the relationship between dosage, time, and platinum concentration in neurologic tissues, and morphologic evidence of toxicity. *J Clin Oncol* 10(5), pp. 795-803.

Groisman, R. et al. 2006. CSA-dependent degradation of CSB by the ubiquitin-proteasome pathway establishes a link between complementation factors of the Cockayne syndrome. *Genes Dev* 20(11), pp. 1429-1434.

Groisman, R. et al. 2003. The ubiquitin ligase activity in the DDB2 and CSA complexes is differentially regulated by the COP9 signalosome in response to DNA damage. *Cell* 113(3), pp. 357-367.

Gupta-Burt, S. et al. 1993. Relationship between patient response in ovarian and breast cancer and platinum drug-DNA adduct formation. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2(3), pp. 229-234.

Hah, S. S. et al. 2007. Characterization of oxaliplatin-DNA adduct formation in DNA and differentiation of cancer cell drug sensitivity at microdose concentrations. *Chemical research in toxicology* 20(12), pp. 1745-1751.

Hall, J. A. et al. 2007. An exploration into study design for biomarker identification: issues and recommendations. *Cancer Genomics Proteomics* 4(3), pp. 111-119.

Han, T. et al. 2012. Characterization of whole genome amplified (WGA) DNA for use in genotyping assay development. *BMC Genomics* 13, p. 217.

Hanawalt, P. C. and Spivak, G. 2008. Transcription-coupled DNA repair: two decades of progress and surprises. *Nat Rev Mol Cell Biol* 9(12), pp. 958-970.

Hansson, J. et al. 1991. DNA excision repair in cell extracts from human cell lines exhibiting hypersensitivity to DNA-damaging agents. *Cancer Res* 51(13), pp. 3384-3390.

Hartmann, J. T. and Lipp, H. P. 2003. Toxicity of platinum compounds. *Expert Opin Pharmacother* 4(6), pp. 889-901.

Hasmats, J. et al. 2014. Assessment of whole genome amplification for sequence capture and massively parallel sequencing. *PLoS One* 9(1), p. e84785.

Hayes, J. and Scovell, W. M. 1991. cis-diamminedichloroplatinum (II) modified chromatin and nucleosomal core particle. *Biochim Biophys Acta* 1089(3), pp. 377-385.

Hecht, A. et al. 1996. Spreading of transcriptional repressor SIR3 from telomeric heterochromatin. *Nature* 383(6595), pp. 92-96.

Henderson, P. T. et al. 2011. A microdosing approach for characterizing formation and repair of carboplatin-DNA monoadducts and chemoresistance. *International journal of cancer. Journal international du cancer* 129(6), pp. 1425-1434.

Hendriks, G. et al. 2010. Transcription-dependent cytosine deamination is a novel mechanism in ultraviolet light-induced mutagenesis. *Curr Biol* 20(2), pp. 170-175.

Henriette Tanja, L. et al. 2009. Pharmacogenetics in chemotherapy of colorectal cancer. *Best Practice & Research Clinical Gastroenterology* 23(2), pp. 257-273.

Hess, M. T. et al. 1997. Bipartite substrate discrimination by human nucleotide excision repair. *Proceedings of the National Academy of Sciences of the United States of America* 94(13), pp. 6664-6669.

Hill, J. M. et al. 1975. Clinical studies of Platinum Coordination compounds in the treatment of various malignant diseases. *Cancer Chemother Rep* 59(3), pp. 647-659.

Hilton, D. A. et al. 2007. Complications following sural and peroneal nerve biopsies. *J Neurol Neurosurg Psychiatry* 78(11), pp. 1271-1272.

Ho, J. W. et al. 2011. ChIP-chip versus ChIP-seq: lessons for experimental design and data analysis. *BMC genomics* 12, p. 134.

Hoebers, F. J. et al. 2008. Cisplatin-DNA adduct formation in patients treated with cisplatin-based chemoradiation: lack of correlation between normal tissues and primary tumor. *Cancer chemotherapy and pharmacology* 61(6), pp. 1075-1081.

Hoebers, F. J. et al. 2006. Prediction of treatment outcome by cisplatin-DNA adduct formation in patients with stage III/IV head and neck squamous cell carcinoma, treated by concurrent cisplatin-radiation (RADPLAT). *International journal of cancer. Journal international du cancer* 119(4), pp. 750-756.

Hoheisel, J. D. 2006. Microarray technology: beyond transcript profiling and genotype analysis. *Nat Rev Genet* 7(3), pp. 200-210.

Holzer, A. K. et al. 2006. Contribution of the major copper influx transporter CTR1 to the cellular accumulation of cisplatin, carboplatin, and oxaliplatin. *Molecular pharmacology* 70(4), pp. 1390-1394.

Holzer, A. K. et al. 2004. The copper influx transporter human copper transport protein 1 regulates the uptake of cisplatin in human ovarian carcinoma cells. *Molecular pharmacology* 66(4), pp. 817-823.

Horacek, P. and Drobnik, J. 1971. Interaction of cis-dichlorodiammineplatinum (II) with DNA. *Biochim Biophys Acta* 254(2), pp. 341-347.

Houtsma, D. et al. 2010. Pharmacogenetics in oncology: a promising field. *Current Pharmaceutical Design* 16(2), pp. 155-163.

Howell, S. B. et al. 2010. Copper transporters and the cellular pharmacology of the platinum-containing cancer drugs. *Molecular pharmacology* 77(6), pp. 887-894.

Hromas, R. A. et al. 1987. Decreased cisplatin uptake by resistant L1210 leukemia cells. *Cancer letters* 36(2), pp. 197-201.

Huang, Y. and Li, L. 2013. DNA crosslinking damage and cancer - a tale of friend and foe. *Transl Cancer Res* 2(3), pp. 144-154.

Hutchinson, L. 2011. Chemotherapy: DNA repair predicts survival in patients with NSCLC. *Nature reviews. Clinical oncology* 8(11), p. 627.

Hyka-Nouspikel, N. et al. 2011. Circulating human B lymphocytes are deficient in nucleotide excision repair and accumulate mutations upon proliferation. *Blood* 117(23), pp. 6277-6286.

Ibrahim, A. et al. 2004. FDA drug approval summaries: oxaliplatin. *The Oncologist* 9(1), pp. 8-12.

Imamura, T. et al. 2001. Interaction with p53 enhances binding of cisplatin-modified DNA by high mobility group 1 protein. *J Biol Chem* 276(10), pp. 7534-7540.

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.

Invitrogen. 2014. *BioPrime Total Genomic Labelling kit* [Online]. Available at: http://www.lifetechnologies.com/uk/en/home/life-science/gene-expression-analysis-genotyping/genotyping-genomic-profiling/genomic-profilling-acgh/bioprime-reg--total-genomic-labeling-system.html [Accessed: 08/09/2014

Ishida, S. et al. 2002. Uptake of the anticancer drug cisplatin mediated by the copper transporter Ctr1 in yeast and mammals. *Proceedings of the National Academy of Sciences of the United States of America* 99(22), pp. 14298-14302.

Iyer, V. R. et al. 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* 409(6819), pp. 533-538.

Jamieson, E. R. and Lippard, S. J. 1999. Structure, Recognition, and Processing of Cisplatin-DNA Adducts. *Chemical reviews* 99(9), pp. 2467-2498.

Jamieson, S. M. et al. 2005. Oxaliplatin causes selective atrophy of a subpopulation of dorsal root ganglion neurons without inducing cell loss. *Cancer chemotherapy and pharmacology* 56(4), pp. 391-399.

Janicijevic, A. et al. 2003. DNA bending by the human damage recognition complex XPC-HR23B. *DNA Repair (Amst)* 2(3), pp. 325-336.

Jemal, A. et al. 2011. Global cancer statistics. *CA Cancer J Clin* 61(2), pp. 69-90.

Jerremalm, E. et al. 2009. New insights into the biotransformation and pharmacokinetics of oxaliplatin. *Journal of pharmaceutical sciences* 98(11), pp. 3879-3885.

Johnson, W. E. et al. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8(1), pp. 118-127.

Jun, S. H. et al. 2006. DNA mismatch repair system. Classical and fresh roles. *FEBS J* 273(8), pp. 1609-1619.

Jung, Y. and Lippard, S. J. 2007. Direct cellular responses to platinum-induced DNA damage. *Chemical reviews* 107(5), pp. 1387-1407.

Kanai, M. et al. 2010. Associations between glutathione S-transferase pi Ile105Val and glyoxylate aminotransferase Pro11Leu and Ile340Met polymorphisms and early-onset oxaliplatin-induced neuropathy. *Cancer Epidemiol* 34(2), pp. 189-193.

Kartalou, M. and Essigmann, J. M. 2001a. Mechanisms of resistance to cisplatin. *Mutation research* 478(1-2), pp. 23-43.

Kartalou, M. and Essigmann, J. M. 2001b. Recognition of cisplatin adducts by cellular proteins. *Mutat Res* 478(1-2), pp. 1-21.

Katano, K. et al. 2002. Acquisition of resistance to cisplatin is accompanied by changes in the cellular pharmacology of copper. *Cancer Res* 62(22), pp. 6559-6565.

Kauffman, G. et al. 2010. Michele Peyrone (1813-1883), DIscoverer of Cisplatin. *Platinum Metals Review* 54(4), pp. 250-256.

Kautio, A. L. et al. 2011. Burden of chemotherapy-induced neuropathy--a cross-sectional study. *Support Care Cancer* 19(12), pp. 1991-1996.

Kim, E. S. et al. 2012. Tissue platinum concentration and tumor response in non-small-cell lung cancer. *J Clin Oncol* 30(27), pp. 3345-3352.

Kim, T. H. et al. 2005. A high-resolution map of active promoters in the human genome. *Nature* 436(7052), pp. 876-880.

Kirschner, K. and Melton, D. W. 2010. Multiple roles of the ERCC1-XPF endonuclease in DNA repair and resistance to anticancer drugs. *Anticancer research* 30(9), pp. 3223-3232.

Kloft, C. et al. 1999. Development and application of a simple assay to quantify cellular adducts of platinum complexes with DNA. *Pharmaceutical research* 16(3), pp. 470-473.

Koberle, B. et al. 1999. Defective repair of cisplatin-induced DNA damage caused by reduced XPA protein in testicular germ cell tumours. *Curr Biol* 9(5), pp. 273-276.

Kojima, H. et al. 1998. Abrogation of mitochondrial cytochrome c release and caspase-3 activation in acquired multidrug resistance. *J Biol Chem* 273(27), pp. 16647-16650.

Komatsu, M. et al. 2000. Copper-transporting P-type adenosine triphosphatase (ATP7B) is associated with cisplatin resistance. *Cancer Res* 60(5), pp. 1312-1316.

Kozelka, J. 2009. Molecular origin of the sequence-dependent kinetics of reactions between cisplatin derivatives and DNA. *Inorganica Chimica Acta* 362(3), pp. 651-668.

Krishnan, A. V. et al. 2006. Oxaliplatin and axonal Na+ channel function in vivo. *Clin Cancer Res* 12(15), pp. 4481-4484.

Kuebler, J. P. et al. 2007. Oxaliplatin combined with weekly bolus fluorouracil and leucovorin as surgical adjuvant chemotherapy for stage II and III colon cancer: results from NSABP C-07. *J Clin Oncol* 25(16), pp. 2198-2204.

Kurdistani, S. K. et al. 2002. Genome-wide binding map of the histone deacetylase Rpd3 in yeast. *Nat Genet* 31(3), pp. 248-254.

Laine, J. P. and Egly, J. M. 2006. Initiation of DNA repair mediated by a stalled RNA polymerase IIO. *The EMBO journal* 25(2), pp. 387-397.

Land, S. R. et al. 2007. Neurotoxicity from oxaliplatin combined with weekly bolus fluorouracil and leucovorin as surgical adjuvant chemotherapy for stage II and III colon cancer: NSABP C-07. *J Clin Oncol* 25(16), pp. 2205-2211.

Le May, N. et al. 2012. XPG and XPF endonucleases trigger chromatin looping and DNA demethylation for accurate expression of activated genes. *Mol Cell* 47(4), pp. 622-632.

Le Pla, R. C. et al. 2007. Development of a liquid chromatography-electrospray ionization tandem mass spectrometry method for detecting oxaliplatin-DNA intrastrand cross-links in biological samples. *Chem Res Toxicol* 20(8), pp. 1177-1182.

Lee, T. I. et al. 2002. Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science* 298(5594), pp. 799-804.

Lehky, T. J. et al. 2004. Oxaliplatin-induced neurotoxicity: acute hyperexcitability and chronic neuropathy. *Muscle & Nerve* 29(3), pp. 387-392.

Levi, F. et al. 1993. Oxaliplatin activity against metastatic colorectal cancer. A phase II study of 5-day continuous venous infusion at circadian rhythm modulated rate. *Eur J Cancer* 29A(9), pp. 1280-1284.

Li, F. et al. 2010. Association between polymorphisms of ERCC1 and XPD and clinical response to platinum-based chemotherapy in advanced non-small cell lung cancer. *American journal of clinical oncology* 33(5), pp. 489-494.

Li, G. M. 2014. New insights and challenges in mismatch repair: Getting over the chromatin hurdle. *DNA Repair (Amst)*.

Li, Y. et al. 2014. The Sex-Determining Factors SRY and SOX9 Regulate Similar Target Genes and Promote Testis Cord Formation during Testicular Differentiation. *Cell Rep* 8(3), pp. 723-733.

Licht, C. L. et al. 2003. Cockayne syndrome group B cellular and biochemical functions. *Am J Hum Genet* 73(6), pp. 1217-1239.

Lieberthal, W. et al. 1996. Mechanisms of death induced by cisplatin in proximal tubular epithelial cells: apoptosis vs. necrosis. *Am J Physiol* 270(4 Pt 2), pp. F700-708.

Liedert, B. et al. 2006. Adduct-specific monoclonal antibodies for the measurement of cisplatin-induced DNA lesions in individual cell nuclei. *Nucleic Acids Research* 34(6), p. e47.

Lin, X. et al. 2002. The copper transporter CTR1 regulates cisplatin uptake in Saccharomyces cerevisiae. *Molecular pharmacology* 62(5), pp. 1154-1159.

Liu, W. T. et al. 2007. Effects of target length on the hybridization efficiency and specificity of rRNA-based oligonucleotide microarrays. *Appl Environ Microbiol* 73(1), pp. 73-82.

Ljungman, M. and Zhang, F. 1996. Blockage of RNA polymerase as a possible trigger for u.v. light-induced apoptosis. *Oncogene* 13(4), pp. 823-831.

Loo, v. d. 2010. *Distribution based outlier detection in univariate data*. The Hague: Statistics Netherlands.

Luo, F. R. et al. 1998. Cytotoxicity, cellular uptake, and cellular biotransformations of oxaliplatin in human colon carcinoma cells. *Oncol Res* 10(11-12), pp. 595-603.

Luo, F. R. et al. 1999. High-performance liquid chromatographic separation of the biotransformation products of oxaliplatin. *J Chromatogr B Biomed Sci Appl* 724(2), pp. 345-356.

Ma, J. et al. 1995. Current sample handling methods for measurement of platinum-DNA adducts in leucocytes in man lead to discrepant results in DNA adduct levels and DNA repair. *British Journal of Cancer* 71(3), pp. 512-517.

Ma, J. L. et al. 2003. Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. *Mol Cell Biol* 23(23), pp. 8820-8828.

Machover, D. et al. 1996. Two consecutive phase II studies of oxaliplatin (L-OHP) for treatment of patients with advanced colorectal carcinoma who were resistant to previous treatment with fluoropyrimidines. *Ann Oncol* 7(1), pp. 95-98.

Maillard, O. et al. 2007. An aromatic sensor with aversion to damaged strands confers versatility to DNA repair. *PLoS Biol* 5(4), p. e79.

Mangiameli, A. et al. 2002. [Erythropoietin and cisplatin-induced neuropathies in cancer patients]. *Clin Ter* 153(3), pp. 177-180.

Mani, S. et al. 2002. Oxaliplatin: a review of evolving concepts. *Cancer Invest* 20(2), pp. 246-263.

Marietta, C. and Brooks, P. J. 2007. Transcriptional bypass of bulky DNA lesions causes new mutant RNA transcripts in human cells. *EMBO Rep* 8(4), pp. 388-393.

Marinov, G. K. et al. 2014. Large-scale quality analysis of published ChIP-seq data. *G3 (Bethesda)* 4(2), pp. 209-223.

Marsischky, G. T. et al. 1996. Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair. *Genes Dev* 10(4), pp. 407-420.

Marteijn, J. A. et al. 2014. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat Rev Mol Cell Biol* 15(7), pp. 465-481.

Martin, L. P. et al. 2008. Platinum resistance: the role of DNA repair pathways. *Clin Cancer Res* 14(5), pp. 1291-1295.

Martinez-Balibrea, E. et al. 2009. Increased levels of copper efflux transporter ATP7B are associated with poor outcome in colorectal cancer patients receiving oxaliplatin-based chemotherapy. *Int J Cancer* 124(12), pp. 2905-2910.

Mathe, G. et al. 1989. Oxalato-platinum or 1-OHP, a third-generation platinum complex: an experimental and clinical appraisal and preliminary comparison with cis-platinum and carboplatinum. *Biomed Pharmacother* 43(4), pp. 237-250.

Matsumura, Y. et al. 1998. Characterization of molecular defects in xeroderma pigmentosum group F in relation to its clinically mild symptoms. *Hum Mol Genet* 7(6), pp. 969-974.

Matsuyama, R. et al. 2006. Why do patients choose chemotherapy near the end of life? A review of the perspective of those facing death from cancer. *J Clin Oncol* 24(21), pp. 3490-3496.

Maughan, T. S. et al. 2011. Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. *Lancet* 377(9783), pp. 2103-2114.

McDonald, E. S. et al. 2005. Cisplatin preferentially binds to DNA in dorsal root ganglion neurons in vitro and in vivo: a potential mechanism for neurotoxicity. *Neurobiology of disease* 18(2), pp. 305-313.

McGrath, R. A. and Williams, R. W. 1966. Reconstruction in vivo of irradiated Escherichia coli deoxyribonucleic acid; the rejoining of broken pieces. *Nature* 212(5061), pp. 534-535.

McKenna, D. J. et al. 2008. Potential use of the comet assay in the clinical management of cancer. *Mutagenesis* 23(3), pp. 183-190.

McNeil, E. M. and Melton, D. W. 2012. DNA repair endonuclease ERCC1-XPF as a novel therapeutic target to overcome chemoresistance in cancer therapy. *Nucleic Acids Res* 40(20), pp. 9990-10004.

McVey, M. and Lee, S. E. 2008. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet* 24(11), pp. 529-538.

Meczes, E. L. et al. 2005. Specific adducts recognised by a monoclonal antibody against cisplatin-modified DNA. *Biochemical Pharmacology* 70(12), pp. 1717-1725.

Meijer, C. et al. 1999. Cisplatin-induced DNA-platination in experimental dorsal root ganglia neuronopathy. *Neurotoxicology* 20(6), pp. 883-887.

Mellish, K. J. et al. 1995. DNA-binding properties of novel cis- and trans platinum-based anticancer agents in 2 human ovarian carcinoma cell lines. *Int J Cancer* 62(6), pp. 717-723.

Mende, M. et al. 2013. Patients' perspectives on palliative chemotherapy of colorectal and non--colorectal cancer: a prospective study in a chemotherapy-experienced population. *BMC Cancer* 13, p. 66.

Millard, J. T. and Wilkes, E. E. 2000. cis- and trans-diamminedichloroplatinum(II) interstrand cross-linking of a defined sequence nucleosomal core particle. *Biochemistry* 39(51), pp. 16046-16055.

Mills, R. E. et al. 2011. Natural genetic variation caused by small insertions and deletions in the human genome. *Genome Res* 21(6), pp. 830-839.

Miyashita, H. et al. 2003. Expression of copper-transporting P-type adenosine triphosphatase (ATP7B) as a chemoresistance marker in human oral squamous cell carcinoma treated with cisplatin. *Oral Oncol* 39(2), pp. 157-162.

Motzer, R. J. et al. 1994. Platinum-DNA adducts assayed in leukocytes of patients with germ cell tumors measured by atomic absorbance spectrometry and enzyme-linked immunosorbent assay. *Cancer* 73(11), pp. 2843-2852.

Movassagh, M. et al. 2010. Differential DNA methylation correlates with differential expression of angiogenic factors in human heart failure. *PLoS One* 5(1), p. e8564.

MRC. 2015. *FOCUS4 Trial Website* [Online]. Available at: http://www.focus4trial.org [Accessed: 01/06/2015 2015].

Mu, D. et al. 1996. Reaction mechanism of human DNA repair excision nuclease. *J Biol Chem* 271(14), pp. 8285-8294.

Mueller, P. R. and Wold, B. 1989. In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* 246(4931), pp. 780-786.

Myers, J. A. et al. 2013. Improving accuracy of cell and chromophore concentration measurements using optical density. *BMC Biophys* 6(1), p. 4.

Nakayama, K. et al. 2002. Copper-transporting P-type adenosine triphosphatase (ATP7B) as a cisplatin based chemoresistance marker in ovarian carcinoma: comparative analysis with expression of MDR1, MRP1, MRP2, LRP and BCRP. *Int J Cancer* 101(5), pp. 488-495.

Nakayama, K. et al. 2004. Prognostic value of the Cu-transporting ATPase in ovarian carcinoma patients receiving cisplatin-based chemotherapy. *Clin Cancer Res* 10(8), pp. 2804-2811.

Nance, M. A. and Berry, S. A. 1992. Cockayne syndrome: review of 140 cases. *Am J Med Genet* 42(1), pp. 68-84.

NCCN. 2012. *NCCN Colon Cancer Guidelines* [Online]. Available at: [Accessed: 30/04/12
2012].

NCI. 2009. *Common Toxicity Criteria for Adverse Events (CTCAE)* [Online]. Available at: http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf [Accessed: 01/11/2014

Nehme, A. et al. 1997. Differential induction of c-Jun NH2-terminal kinase and c-Abl kinase in DNA mismatch repair-proficient and -deficient cells exposed to cisplatin. *Cancer Res* 57(15), pp. 3253-3257.

Nehme, A. et al. 1999. Induction of JNK and c-Abl signalling by cisplatin and oxaliplatin in mismatch repair-proficient and -deficient cells. *Br J Cancer* 79(7-8), pp. 1104-1110.

Nel, I. et al. 2013. Formation and repair kinetics of Pt-(GpG) DNA adducts in extracted circulating tumour cells and response to platinum treatment. *Br J Cancer*.

NICE. 2008. *NICE technology appraisals [TA145]: Cetuximab for the treatment of locally advanced squamous cell cancer of the head and neck* [Online]. Available at: http://www.nice.org.uk/guidance/ta145 [Accessed: 01/10/2014 2014].

NICE 2011. The diagnosis and management of colorectal cancer: full guideline (November 2011). *National Institute for Health and Clinical Excellence*.

Niedernhofer, L. J. et al. 2001. The structure-specific endonuclease Ercc1-Xpf is required for targeted gene replacement in embryonic stem cells. *The EMBO journal* 20(22), pp. 6540-6549.

Niedernhofer, L. J. et al. 2004. The structure-specific endonuclease Ercc1-Xpf is required to resolve DNA interstrand cross-link-induced double-strand breaks. *Mol Cell Biol* 24(13), pp. 5776-5787.

Nik-Zainal, S. et al. 2012. Mutational processes molding the genomes of 21 breast cancers. *Cell* 149(5), pp. 979-993.

Njai, H. F. et al. 2011. Setting Up a Standardized Peripheral Blood Mononuclear Cells Processing Laboratory to Support Multi-center HIV/AIDS Vaccine and Intervention Trials. *Laboratory Medicine* 42(12), pp. 711-718.

Nouspikel, T. 2007. DNA repair in differentiated cells: some new answers to old questions. *Neuroscience* 145(4), pp. 1213-1221.

Nouspikel, T. 2009. DNA repair in mammalian cells : Nucleotide excision repair: variations on versatility. *Cell Mol Life Sci* 66(6), pp. 994-1009.

Nouspikel, T. 2011. Attenuated nucleotide excision repair leads to mutagenesis in cancer cells. *Future Oncol* 7(12), pp. 1361-1363.

Nouspikel, T. and Hanawalt, P. C. 2000. Terminally differentiated human neurons repair transcribed genes but display attenuated global DNA repair and modulation of repair gene expression. *Mol Cell Biol* 20(5), pp. 1562-1570.

Nouspikel, T. and Hanawalt, P. C. 2002. DNA repair in terminally differentiated cells. *DNA Repair (Amst)* 1(1), pp. 59-75.

O'Connell, J. B. et al. 2004. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. *Journal of the National Cancer Institute* 96(19), pp. 1420-1425.

O'Connor, A. B. 2009. Neuropathic pain: quality-of-life impact, costs and cost effectiveness of therapy. *Pharmacoeconomics* 27(2), pp. 95-112.

O'Dwyer, P. J. et al. 2000. Clinical pharmacokinetics and administration of established platinum drugs. *Drugs* 59 Suppl 4, pp. 19-27.

O'Geen, H. et al. 2006. Comparison of sample preparation methods for ChIP-chip assays. *BioTechniques* 41(5), pp. 577-580.

Okazaki, T. et al. 2008. Single-nucleotide polymorphisms of DNA damage response genes are associated with overall survival in patients with pancreatic cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* 14(7), pp. 2042-2048.

Olaussen, K. A. et al. 2012. DNA repair capacity in circulating lymphocytes and influence on platinum effect in tumor cells. *J Clin Oncol* 30(13), pp. 1567-1568; author reply 1568-1569.

Ordway, J. M. et al. 2006. Comprehensive DNA methylation profiling in a human cancer genome identifies novel epigenetic targets. *Carcinogenesis* 27(12), pp. 2409-2423.

Overman, M. J. et al. 2013. Use of research biopsies in clinical trials: are risks and benefits adequately discussed? *J Clin Oncol* 31(1), pp. 17-22.

Page, J. D. et al. 1990. Effect of the diaminocyclohexane carrier ligand on platinum adduct formation, repair, and lethality. *Biochemistry* 29(4), pp. 1016-1024.

Pandey, P. et al. 1996. Activation of p38 mitogen-activated protein kinase by c-Abl-dependent and -independent mechanisms. *J Biol Chem* 271(39), pp. 23775-23779.

Park, C. J. and Choi, B. S. 2006. The protein shuffle. Sequential interactions among components of the human nucleotide excision repair pathway. *FEBS J* 273(8), pp. 1600-1608.

Park, P. J. 2009. ChIP-seq: advantages and challenges of a maturing technology. *Nature reviews. Genetics* 10(10), pp. 669-680.

Park, S. B. et al. 2009. Oxaliplatin-induced neurotoxicity: changes in axonal excitability precede development of neuropathy. *Brain* 132(Pt 10), pp. 2712-2723.

Park, S. R. et al. 2011. CYP2A6 and ERCC1 polymorphisms correlate with efficacy of S-1 plus cisplatin in metastatic gastric cancer patients. *British Journal of Cancer* 104(7), pp. 1126-1134.

Paunio, T. et al. 1996. Preimplantation diagnosis by whole-genome amplification, PCR amplification, and solid-phase minisequencing of blastomere DNA. *Clin Chem* 42(9), pp. 1382-1390.

Pendyala, L. and Creaven, P. J. 1993. In vitro cytotoxicity, protein binding, red blood cell partitioning, and biotransformation of oxaliplatin. *Cancer Res* 53(24), pp. 5970-5976.

Peng, B. et al. 1997. Platinum-DNA adduct formation in leucocytes of children in relation to pharmacokinetics after cisplatin and carboplatin therapy. *British Journal of Cancer* 76(11), pp. 1466-1473.

Peng, M. et al. 2014. Crosstalk between BRCA-Fanconi anemia and mismatch repair pathways prevents MSH2-dependent aberrant DNA damage responses. *EMBO J* 33(15), pp. 1698-1712.

Perdiz, D. et al. 2000. Distribution and repair of bipyrimidine photoproducts in solar UV-irradiated mammalian cells. Possible role of Dewar photoproducts in solar mutagenesis. *J Biol Chem* 275(35), pp. 26732-26742.

Peters, G. J. et al. 2014. Predictive role of repair enzymes in the efficacy of Cisplatin combinations in pancreatic and lung cancer. *Anticancer Res* 34(1), pp. 435-442.

Petersen, L. N. et al. 1996. Increased gene specific repair of cisplatin induced interstrand crosslinks in cisplatin resistant cell lines, and studies on carrier ligand specificity. *Carcinogenesis* 17(12), pp. 2597-2602.

Peyrone, M. 1844. Ueber die Einwirkung des Ammoniaks auf Platinchlorur. *Ann. Chem. Pharm* 51(1).

Phillips, D. H. et al. 2000. Methods of DNA adduct determination and their application to testing compounds for genotoxicity. *Environmental and molecular mutagenesis* 35(3), pp. 222-233.

Pieck, A. C. et al. 2008. Oxaliplatin-DNA adduct formation in white blood cells of cancer patients. *British Journal of Cancer* 98(12), pp. 1959-1965.

Pignon, J.-P. et al. 2009. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): An update on 93 randomised trials and 17,346 patients. *Radiotherapy and Oncology* 92(1), pp. 4-14.

Pike, C. T. et al. 2012. Healthcare costs and workloss burden of patients with chemotherapy-associated peripheral neuropathy in breast, ovarian, head and neck, and nonsmall cell lung cancer. *Chemotherapy research and practice* 2012, p. 913848.

Pinard, R. et al. 2006. Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics* 7, p. 216.

Plasencia, C. et al. 2006. Expression analysis of genes involved in oxaliplatin response and development of oxaliplatin-resistant HT29 colon cancer cells. *International journal of oncology* 29(1), pp. 225-235.

Pluim, D. et al. 1999. 32P-postlabeling assay for the quantification of the major platinum-DNA adducts. *Analytical biochemistry* 275(1), pp. 30-38.

Poirier, M. C. et al. 1982. Antibodies elicited against cis-diamminedichloroplatinum(II)-modified DNA are specific for cis-diamminedichloroplatinum(II)-DNA adducts formed in vivo and in vitro. *Proceedings of the National Academy of Sciences of the United States of America* 79(21), pp. 6443-6447.

Pokholok, D. K. et al. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122(4), pp. 517-527.

Poklar, N. et al. 1996. Influence of cisplatin intrastrand crosslinking on the conformation, thermal stability, and energetics of a 20-mer DNA duplex. *Proc Natl Acad Sci U S A* 93(15), pp. 7606-7611.

Popoff, S. C. et al. 1987. Repair of plasmid DNA damaged in vitro with cis- or trans-diamminedichloroplatinum(II) in Escherichia coli. *Mutat Res* 183(2), pp. 129-137.

Powell, J. 2014. *Measuring DNA damage and associated epigenetic changes genome-wide in cells following exposure to platinum analogue chemotherapeutic drugs.* Cardiff University.

Prakash, S. et al. 1993. DNA repair genes and proteins of Saccharomyces cerevisiae. *Annu Rev Genet* 27, pp. 33-70.

Quail, M. A. et al. 2012. Optimal enzymes for amplifying sequencing libraries. *Nat Methods* 9(1), pp. 10-11.

R Development Core Team 2014. R: A language and environment for statistical computing. R Foundation for
Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rabik, C. and Dolan, M. 2007. Molecular mechanisms of resistance and toxicity associated with platinating agents☆. *Cancer Treatment Reviews* 33(1), pp. 9-23.

Rademakers, S. et al. 2003. Xeroderma pigmentosum group A protein loads as a separate factor onto DNA lesions. *Mol Cell Biol* 23(16), pp. 5755-5767.

Rajeev, L. et al. 2014. DNA-affinity-purified Chip (DAP-chip) Method to Determine Gene Targets for Bacterial Two component Regulatory Systems. *J Vis Exp* (89).

Ramanathan, R. K. et al. 2010. Incidence and evolution of oxaliplatin-induced peripheral sensory neuropathy in diabetic patients with colorectal cancer: a pooled analysis of three phase III studies. *Ann Oncol* 21(4), pp. 754-758.

Rasmussen, R. E. and Painter, R. B. 1964. Evidence for Repair of Ultra-Violet Damaged Deoxyribonucleic Acid in Cultured Mammalian Cells. *Nature* 203, pp. 1360-1362.

Raymond, E. et al. 1998a. Oxaliplatin: a review of preclinical and clinical studies. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 9(10), pp. 1053-1071.

Raymond, E. et al. 2002. Cellular and molecular pharmacology of oxaliplatin. *Mol Cancer Ther* 1(3), pp. 227-235.

Raymond, E. et al. 1998b. Oxaliplatin: mechanism of action and antineoplastic activity. *Semin Oncol* 25(2 Suppl 5), pp. 4-12.

Reardon, J. T. et al. 1993. Excision repair in man and the molecular basis of xeroderma pigmentosum syndrome. *Cold Spring Harb Symp Quant Biol* 58, pp. 605-617.

Reardon, J. T. et al. 1999. Efficient nucleotide excision repair of cisplatin, oxaliplatin, and Bis-aceto-ammine-dichloro-cyclohexylamine-platinum(IV) (JM216) platinum intrastrand DNA diadducts. *Cancer research* 59(16), pp. 3968-3971.

Reed, E. 2006. ERCC1 measurements in clinical oncology. *N Engl J Med* 355(10), pp. 1054-1055.

Reed, E. et al. 1990. Evaluation of platinum-DNA adduct levels relative to known prognostic variables in a cohort of ovarian cancer patients. *Cancer research* 50(8), pp. 2256-2260.

Reed, E. et al. 1987. Platinum-DNA adducts in leukocyte DNA correlate with disease response in ovarian cancer patients receiving platinum-based chemotherapy. *Proceedings of the National Academy of Sciences of the United States of America* 84(14), pp. 5024-5028.

Reed, E. et al. 1988. The measurement of cisplatin-DNA adduct levels in testicular cancer patients. *Carcinogenesis* 9(10), pp. 1909-1911.

Reed, E. et al. 1993. Platinum-DNA adduct in leukocyte DNA of a cohort of 49 patients with 24 different types of malignancies. *Cancer research* 53(16), pp. 3694-3699.

Reed, G. F. et al. 2002. Use of coefficient of variation in assessing variability of quantitative assays. *Clin Diagn Lab Immunol* 9(6), pp. 1235-1239.

Reed, S. H. 2005. Nucleotide excision repair in chromatin: the shape of things to come. *DNA Repair (Amst)* 4(8), pp. 909-918.

Reed, S. H. 2011. Nucleotide excision repair in chromatin: damage removal at the drop of a HAT. *DNA Repair (Amst)* 10(7), pp. 734-742.

Ren, B. et al. 2000. Genome-wide location and function of DNA binding proteins. *Science* 290(5500), pp. 2306-2309.

Ren, S. et al. 2012. Association between polymorphisms of DNA repair genes and survival of advanced NSCLC patients treated with platinum-based chemotherapy. *Lung cancer* 75(1), pp. 102-109.

Renn, C. L. et al. 2011. Multimodal assessment of painful peripheral neuropathy induced by chronic oxaliplatin-based chemotherapy in mice. *Molecular pain* 7, p. 29.

Reslova, S. 1971. The induction of lysogenic strains of Escherichia coli by cis-dichloro-diammineplatinum (II). *Chem Biol Interact* 4(1), pp. 66-70.

Reynolds, P. et al. 1985. Nucleotide sequence of the RAD10 gene of Saccharomyces cerevisiae. *EMBO J* 4(13A), pp. 3549-3552.

Riedl, T. et al. 2003. The comings and goings of nucleotide excision repair factors on damaged DNA. *EMBO J* 22(19), pp. 5293-5303.

Robyr, D. et al. 2002. Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases. *Cell* 109(4), pp. 437-446.

Roos, W. P. and Kaina, B. 2013. DNA damage-induced cell death: from specific DNA lesions to the DNA damage response and apoptosis. *Cancer Lett* 332(2), pp. 237-248.

Rosenberg, B. et al. 1965. Inhibition of Cell Division in Escherichia Coli by Electrolysis Products from a Platinum Electrode. *Nature* 205, pp. 698-699.

Rutledge, R. G. 2004. Sigmoidal curve-fitting redefines quantitative real-time PCR with the prospective of developing automated high-throughput applications. *Nucleic Acids Res* 32(22), p. e178.

Rydberg, B. et al. 1994. DNA double-strand breaks induced by high-energy neon and iron ions in human fibroblasts. I. Pulsed-field gel electrophoresis method. *Radiat Res* 139(2), pp. 133-141.

Safaei, R. 2006. Role of copper transporters in the uptake and efflux of platinum containing drugs. *Cancer letters* 234(1), pp. 34-39.

Safaei, R. and Howell, S. B. 2005. Copper transporters regulate the cellular pharmacology and sensitivity to Pt drugs. *Crit Rev Oncol Hematol* 53(1), pp. 13-23.

Sameer, A. S. et al. 2014. Mismatch repair pathway: molecules, functions, and role in colorectal carcinogenesis. *Eur J Cancer Prev*.

Samimi, G. et al. 2004. Increased expression of the copper efflux transporter ATP7A mediates resistance to cisplatin, carboplatin, and oxaliplatin in ovarian cancer cells. *Clin Cancer Res* 10(14), pp. 4661-4669.

Sargent, R. G. et al. 1997. Recombination-dependent deletion formation in mammalian cells deficient in the nucleotide excision repair gene ERCC1. *Proc Natl Acad Sci U S A* 94(24), pp. 13122-13127.

Saris, C. P. et al. 1996. In vitro formation of DNA adducts by cisplatin, lobaplatin and oxaliplatin in calf thymus DNA in solution and in cultured human cells. *Carcinogenesis* 17(12), pp. 2763-2769.

Sarker, A. H. et al. 2005. Recognition of RNA polymerase II and transcription bubbles by XPG, CSB, and TFIIH: insights for transcription-coupled repair and Cockayne Syndrome. *Mol Cell* 20(2), pp. 187-198.

Schellens, J. H. et al. 1996. Relationship between the exposure to cisplatin, DNA-adduct formation in leucocytes and tumour response in patients with solid tumours. *British Journal of Cancer* 73(12), pp. 1569-1575.

Schiestl, R. H. and Prakash, S. 1990. RAD10, an excision repair gene of Saccharomyces cerevisiae, is involved in the RAD1 pathway of mitotic recombination. *Mol Cell Biol* 10(6), pp. 2485-2491.

Schmidt, W. and Chaney, S. G. 1993. Role of carrier ligand in platinum resistance of human carcinoma cell lines. *Cancer Res* 53(4), pp. 799-805.

Selvakumaran, M. et al. 2003. Enhanced cisplatin cytotoxicity by disturbing the nucleotide excision repair pathway in ovarian cancer cell lines. *Cancer Res* 63(6), pp. 1311-1316.

Sereno, M. et al. 2014. Oxaliplatin induced-neuropathy in digestive tumors. *Crit Rev Oncol Hematol* 89(1), pp. 166-178.

Sestili, P. et al. 2006. The fast halo assay: an improved method to quantify genomic DNA strand breakage at the single-cell level. *Mutat Res* 607(2), pp. 205-214.

Setlow, R. B. and Carrier, W. L. 1964. The Disappearance of Thymine Dimers from DNA: An Error-Correcting Mechanism. *Proc Natl Acad Sci U S A* 51, pp. 226-231.

Seymour, M. et al. 2007. Different strategies of sequential and combination chemotherapy for patients with poor prognosis advanced colorectal cancer (MRC FOCUS): a randomised controlled trial. *The Lancet* 370(9582), pp. 143-152.

Sgouros, J. et al. 1999. A relationship betweena DNA-repair/recombination nuclease family and archaeal helicases. *Trends Biochem Sci* 24(3), pp. 95-97.

Shagin, D. A. et al. 1999. Regulation of average length of complex PCR product. *Nucleic Acids Res* 27(18), p. e23.

Shiraishi, K. et al. 2010. Association of DNA repair gene polymorphisms with response to platinum-based doublet chemotherapy in patients with non-small-cell lung cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 28(33), pp. 4945-4952.

Siddik, Z. H. 2003. Cisplatin: mode of cytotoxic action and molecular basis of resistance. *Oncogene* 22(47), pp. 7265-7279.

Sigma. *Whole Genome Amplifcation Advisor* [Online]. Available at: http://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma-Aldrich/Brochure/1/wga_advisor.pdf [Accessed: 08/09/2014 2014].

Sijbers, A. M. et al. 1996. Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell* 86(5), pp. 811-822.

Sikorski, R. S. and Hieter, P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics* 122(1), pp. 19-27.

Simon, G. R. et al. 2007. Nuclear excision repair-based personalized therapy for non-small cell lung cancer: from hypothesis to reality. *The international journal of biochemistry & cell biology* 39(7-8), pp. 1318-1328.

Sorenson, C. M. and Eastman, A. 1988. Influence of cis-diamminedichloroplatinum(II) on DNA synthesis and cell cycle progression in excision repair proficient and deficient Chinese hamster ovary cells. *Cancer Res* 48(23), pp. 6703-6707.

Spanswick, V. J. et al. 2012. Evidence for different mechanisms of 'unhooking' for melphalan and cisplatin-induced DNA interstrand cross-links in vitro and in clinical acquired resistant tumour samples. *BMC Cancer* 12, p. 436.

Strumberg, D. et al. 2002. Evaluation of long-term toxicity in patients after cisplatin-based chemotherapy for non-seminomatous testicular cancer. *Ann Oncol* 13(2), pp. 229-236.

Sugasawa, K. et al. 2001. A multistep damage recognition mechanism for global genomic nucleotide excision repair. *Genes Dev* 15(5), pp. 507-521.

Sugasawa, K. et al. 2005. UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex. *Cell* 121(3), pp. 387-400.

Sundquist, W. I. et al. 1987. Monoclonal antibodies to DNA modified with cis- or trans-diamminedichloroplatinum(II). *Proceedings of the National Academy of Sciences of the United States of America* 84(23), pp. 8225-8229.

Sung, P. et al. 1993. Purification and characterization of the Saccharomyces cerevisiae RAD1/RAD10 endonuclease. *J Biol Chem* 268(35), pp. 26391-26399.

Sweder, K. S. and Hanawalt, P. C. 1992. Preferential repair of cyclobutane pyrimidine dimers in the transcribed strand of a gene in yeast chromosomes and plasmids is dependent on transcription. *Proceedings of the National Academy of Sciences of the United States of America* 89(22), pp. 10696-10700.

Ta, L. E. et al. 2006. Neurotoxicity of oxaliplatin and cisplatin for dorsal root ganglion neurons correlates with platinum-DNA binding. *Neurotoxicology* 27(6), pp. 992-1002.

Tan, X. L. et al. 2011. Genetic variation predicting cisplatin cytotoxicity associated with overall survival in lung cancer patients receiving platinum-based

chemotherapy. *Clinical cancer research : an official journal of the American Association for Cancer Research* 17(17), pp. 5801-5811.

Tang, M. S. et al. 1989. Quantification of aminofluorene adduct formation and repair in defined DNA sequences in mammalian cells using the UVRABC nuclease. *J Biol Chem* 264(24), pp. 14455-14462.

Tapias, A. et al. 2004. Ordered conformational changes in damaged DNA induced by nucleotide excision repair factors. *J Biol Chem* 279(18), pp. 19074-19083.

Tavtigian, S. V. et al. 2006. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet* 43(4), pp. 295-305.

Technologies, A. 2008. *Agilent Oligonucleotide Array-Based CGH for Genomic DNA Analysis* [Online]. Available at: http://www.chem-agilent.com/pdf/G4410-90020.pdf [Accessed: 08/09/2014 2014].

Teng, Y. et al. 2010. A novel method for the genome-wide high resolution analysis of DNA damage. *Nucleic Acids Research*.

Teng, Y. et al. 1997. Excision repair at the level of the nucleotide in the Saccharomyces cerevisiae MFA2 gene: mapping of where enhanced repair in the transcribed strand begins or ends and identification of only a partial rad16 requisite for repairing upstream control sequences. *J Mol Biol* 267(2), pp. 324-337.

Terheggen, P. M. et al. 1991. Antibodies against cisplatin-modified DNA and cisplatin-modified (di)nucleotides. *Cancer chemotherapy and pharmacology* 28(3), pp. 185-191.

Tilby, M. J. 1999. Measuring DNA adducts by immunoassay (ELISA). *Methods in molecular medicine* 28, pp. 121-128.

Tilby, M. J. et al. 1991. Sensitive detection of DNA modifications induced by cisplatin and carboplatin in vitro and in vivo using a monoclonal antibody. *Cancer research* 51(1), pp. 123-129.

Tiseo, M. et al. 2013. ERCC1/BRCA1 expression and gene polymorphisms as prognostic and predictive factors in advanced NSCLC treated with or without cisplatin. *Br J Cancer*.

Todd, R. C. and Lippard, S. J. 2009. Inhibition of transcription by platinum antitumor compounds. *Metallomics : integrated biometal science* 1(4), pp. 280-291.

Tomkinson, A. E. et al. 1993. Yeast DNA repair and recombination proteins Rad1 and Rad10 constitute a single-stranded-DNA endonuclease. *Nature* 362(6423), pp. 860-862.

Tornaletti, S. and Hanawalt, P. C. 1999. Effect of DNA lesions on transcription elongation. *Biochimie* 81(1-2), pp. 139-146.

Travis, L. B. et al. 2014. Chemotherapy-induced peripheral neurotoxicity and ototoxicity: new paradigms for translational genomics. *J Natl Cancer Inst* 106(5).

Tripsianes, K. et al. 2007. Analysis of the XPA and ssDNA-binding surfaces on the central domain of human ERCC1 reveals evidence for subfunctionalization. *Nucleic Acids Res* 35(17), pp. 5789-5798.

Tsankova, N. et al. 2007. Epigenetic regulation in psychiatric disorders. *Nat Rev Neurosci* 8(5), pp. 355-367.

Vaisman, A. et al. 1999. Effect of DNA polymerases and high mobility group protein 1 on the carrier ligand specificity for translesion synthesis past platinum-DNA adducts. *Biochemistry* 38(34), pp. 11026-11039.

Vaisman, A. et al. 1998. The role of hMLH1, hMSH3, and hMSH6 defects in cisplatin and oxaliplatin resistance: correlation with replicative bypass of platinum-DNA adducts. *Cancer Res* 58(16), pp. 3579-3585.

van de Vaart, P. J. et al. 2000. DNA-adduct levels as a predictor of outcome for NSCLC patients receiving daily cisplatin and radiotherapy. *International journal of cancer. Journal international du cancer* 89(2), pp. 160-166.

van Dijk, E. L. et al. 2014. Library preparation methods for next-generation sequencing: Tone down the bias. *Exp Cell Res* 322(1), pp. 12-20.

van Duin, M. et al. 1986. Molecular characterization of the human excision repair gene ERCC-1: cDNA cloning and amino acid homology with the yeast DNA repair gene RAD10. *Cell* 44(6), pp. 913-923.

van Duin, M. et al. 1989. The cloned human DNA excision repair gene ERCC-1 fails to correct xeroderma pigmentosum complementation groups A through I. *Mutat Res* 217(2), pp. 83-92.

van Vuuren, A. J. et al. 1993. Evidence for a repair enzyme complex involving ERCC1 and complementing activities of ERCC4, ERCC11 and xeroderma pigmentosum group F. *The EMBO journal* 12(9), pp. 3693-3701.

Vasquez, K. M. and Legerski, R. J. 2010. DNA interstrand crosslinks: repair, cell signaling, and therapeutic implications. *Environmental and molecular mutagenesis* 51(6), pp. 491-492.

Veal, G. J. et al. 2001. Influence of cellular factors and pharmacokinetics on the formation of platinum-DNA adducts in leukocytes of children receiving cisplatin therapy. *Clin Cancer Res* 7(8), pp. 2205-2212.

Velasco, R. et al. 2014. Early predictors of oxaliplatin-induced cumulative neuropathy in colorectal cancer patients. *J Neurol Neurosurg Psychiatry* 85(4), pp. 392-398.

Vermeulen, W. and Fousteri, M. 2013. Mammalian transcription-coupled excision repair. *Cold Spring Harb Perspect Biol* 5(8), p. a012625.

Volker, M. et al. 2001. Sequential assembly of the nucleotide excision repair factors in vivo. *Mol Cell* 8(1), pp. 213-224.

Wakasugi, M. et al. 2002. DDB accumulates at DNA damage sites immediately after UV irradiation and directly stimulates nucleotide excision repair. *J Biol Chem* 277(3), pp. 1637-1640.

Wakasugi, M. and Sancar, A. 1998. Assembly, subunit composition, and footprint of human DNA repair excision nuclease. *Proc Natl Acad Sci U S A* 95(12), pp. 6669-6674.

Walker, A. S. et al. 2014. Future directions for the early detection of colorectal cancer recurrence. *J Cancer* 5(4), pp. 272-280.

Wang, D. et al. 2003. Nucleotide excision repair from site-specifically platinum-modified nucleosomes. *Biochemistry* 42(22), pp. 6747-6753.

Wang, D. and Lippard, S. J. 2005. Cellular processing of platinum anticancer drugs. *Nature Reviews Drug Discovery* 4(4), pp. 307-320.

Wang, L. E. et al. 2011. DNA repair capacity in peripheral lymphocytes predicts survival of patients with non-small-cell lung cancer treated with first-line platinum-based chemotherapy. *J Clin Oncol* 29(31), pp. 4121-4128.

Wang, Z. et al. 1993. Nucleotide-excision repair of DNA in cell-free extracts of the yeast Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* 90(11), pp. 4907-4911.

Weber, M. et al. 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* 39(4), pp. 457-466.

Weickhardt, A. et al. 2011. Oxaliplatin-induced neuropathy in colorectal cancer. *Journal of oncology* 2011, p. 201593.

Weinmann, A. S. et al. 2002. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev* 16(2), pp. 235-244.

Wells, J. et al. 2003. Identification of novel pRb binding sites using CpG microarrays suggests that E2F recruits pRb to specific genomic sites during S phase. *Oncogene* 22(10), pp. 1445-1460.

Welsh, C. et al. 2004. Reduced levels of XPA, ERCC1 and XPF DNA repair proteins in testis tumor cell lines. *Int J Cancer* 110(3), pp. 352-361.

Welters, M. J. et al. 1999a. The potential of plantinum-DNA adduct determination in ex vivo treated tumor fragments for the prediction of sensitivity to cisplatin chemotherapy. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 10(1), pp. 97-103.

Welters, M. J. et al. 1999b. Pharmacodynamics of cisplatin in human head and neck cancer: correlation between platinum content, DNA adduct levels and drug sensitivity in vitro and in vivo. *British Journal of Cancer* 79(1), pp. 82-88.

Welters, M. J. et al. 1997. Improved 32P-postlabelling assay for the quantification of the major platinum-DNA adducts. *Carcinogenesis* 18(9), pp. 1767-1774.

West, H. 2013. *Investigating the Genetics and Pharmacogenetics of Bowel Cancer.* Cardiff University.

Wheeler, H. E. et al. 2011. Genome-wide meta-analysis identifies variants associated with platinating agent susceptibility across populations. *Pharmacogenomics J*.

Wong, E. and Giandomenico, C. M. 1999. Current status of platinum-based antitumor drugs. *Chemical reviews* 99(9), pp. 2451-2466.

Wood, D. K. et al. 2010. Single cell trapping and DNA damage analysis using microwell arrays. *Proc Natl Acad Sci U S A* 107(22), pp. 10008-10013.

Wood, R. D. 2010. Mammalian nucleotide excision repair proteins and interstrand crosslink repair. *Environmental and molecular mutagenesis* 51(6), pp. 520-526.

Woynarowski, J. M. et al. 1998. Sequence- and region-specificity of oxaliplatin adducts in naked and cellular DNA. *Molecular pharmacology* 54(5), pp. 770-777.

Woynarowski, J. M. et al. 2000. Oxaliplatin-induced damage of cellular DNA. *Molecular pharmacology* 58(5), pp. 920-927.

Wozniak, K. and Blasiak, J. 2002. Recognition and repair of DNA-cisplatin adducts. *Acta Biochim Pol* 49(3), pp. 583-596.

Wynne, P. et al. 2007. Enhanced repair of DNA interstrand crosslinking in ovarian cancer cells from patients following treatment with platinum-based chemotherapy. *Br J Cancer* 97(7), pp. 927-933.

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(1), pp. 41-44.

Yao, N. et al. 1997. Structure of the hepatitis C virus RNA helicase domain. *Nat Struct Biol* 4(6), pp. 463-467.

Yuan, F. et al. 2004. Evidence for involvement of HMGB1 protein in human DNA mismatch repair. *J Biol Chem* 279(20), pp. 20935-20940.

Zayed, A. et al. 2011. Speciation of oxaliplatin adducts with DNA nucleotides. *Metallomics : integrated biometal science* 3(10), pp. 991-1000.

Zhang, S. et al. 2006. Organic cation transporters are determinants of oxaliplatin cytotoxicity. *Cancer Res* 66(17), pp. 8847-8857.

Zhou, C. et al. 2010. Predictive effects of ERCC1 and XRCC3 SNP on efficacy of platinum-based chemotherapy in advanced NSCLC patients. *Japanese journal of clinical oncology* 40(10), pp. 954-960.

Zwelling, L. A. et al. 1978. Kinetics of formation and disappearance of a DNA cross-linking effect in mouse leukemia L1210 cells treated with cis- and trans-diamminedichloroplatinum(II). *Cancer Res* 38(6), pp. 1762-1768.