Synthetic lethality and functional study of DNA repair defects in ERCC1-deficient non-small-cell lung cancer

Sophie Postel-Vinay

UNIVERSITÉ PARIS SUD
Comprendre le monde, construire l'avenir®

ICR The Institute of Cancer Research

GUSTAVE ROUSSY CANCER CAMPUS GRAND PARIS

Inserm

ÉCOLE DOCTORALE de Cancérologie 418

Laboratoire U981 INSERM
Biomarqueurs prédictifs et nouvelles stratégies moléculaires en thérapeutique anticancéreuse

Gene Function Laboratory - The Institute of Cancer Research

DISCIPLINE: ONCOLOGIE

THÈSE DE DOCTORAT

soutenue le 16 Décembre 2013

par

Sophie POSTEL-VINAY

Synthetic lethality and functional study of DNA repair defects in ERCC1-deficient non-small-cell lung cancer

Directeur de thèse :	Pr Fabrice ANDRE	Institut Gustave Roussy – Unité 981 INSERM
Co-directeur de thèse :	Dr Chris LORD	The Institute of Cancer Research – Gene function

<u>Composition du jury</u> :

Président du jury :	Pr Alexander EGGERMONT	Institut Gustave Roussy - directeur
Rapporteurs :	Dr Olivier DELATTRE	Institut Curie – Unité 830 INSERM
	Pr Fabrice BARLESI	Assistance Publique des Hôpitaux de Marseille
Examinateurs :	Pr Jean-Charles SORIA	Institut Gustave Roussy – Unité 981 INSERM
	Pr Alan ASHWORTH	The Institute of Cancer Research – director

Abstract

Excision Repair Cross-Complementation group 1 (ERCC1) is a DNA repair enzyme that is frequently deficient in non-small cell lung cancer (NSCLC). Although low ERCC1 expression correlates with platinum sensitivity, the clinical effectiveness of platinum therapy is limited - mainly by toxicities and occurrence of resistance - highlighting the need for alternative treatment strategies. In addition, the lack of a reliable assay evaluating ERCC1 functionality in the clinical setting currently precludes personalising therapy based on ERCC1 status.

To discover new synthetic lethality-based therapeutic strategies for ERCC1-defective tumours, high-throughput drug and siRNA screens in an isogenic NSCLC model of ERCC1 deficiency were performed. This approach identified multiple clinical poly(ADP-ribose) polymerase 1 and 2 (PARP1/2) inhibitors such as olaparib (AZD-2281), niraparib (MK-4827) and BMN 673 as being selective for ERCC1 deficiency. The mechanism underlying ERCC1-selective effects was dissected by studying molecular biomarkers of tumour cell response, and revealed that: (i) ERCC1-deficient cells displayed a significant delay in double-strand break repair associated with a profound and prolonged G_2/M arrest following PARP1/2 inhibitor treatment; (ii) ERCC1 isoform 202, which has recently been shown to mediate platinum sensitivity, also modulated PARP1/2 sensitivity; (iii) ERCC1-deficiency was epistatic with homologous recombination deficiency, although ERCC1-deficient cells did not display a defect in RAD51 foci formation. This suggests that ERCC1 might be required to process PARP1/2 inhibitor induced DNA lesions prior to DNA strand invasion; and (iv) PARP1 silencing restored PARP1/2 inhibitor resistance in ERCC1-deficient cells but had no effect in ERCC1-proficient cells, supporting the hypothesis that PARP1 might be required for the ERCC1 selectivity of PARP1/2 inhibitors. This study indicated that PARP1/2 inhibitors as a monotherapy could represent a novel therapeutic strategy for NSCLC patients with ERCC1-deficient tumours, and a clinical protocol is being written to evaluate this hypothesis.

To investigate whether a surrogate biomarker of ERCC1 functionality could be developed, four parallel approaches were undertaken in the ERCC1-isogenic NSCLC model: (i) UV irradiation, to evaluate the Nucleotide Excision Repair (NER) pathway; (ii) whole exome sequencing, to look for an ERCC1-associated genomic scar at the DNA level; (iii) transcriptomic analysis, to investigate changes at the RNA expression level; and (iv) SILAC (Stable Isotope Labeling by Amino acids in Cell culture) analysis, to compare proteomic profiles between ERCC1-proficient and ERCC1-deficient cells. These approaches allowed the identification of putative genomic signature and potential metabolic surrogate biomarkers - guanine deaminase (GDA) and nicotinamide phosphoribosyltransferase (NAMPT). Further validation and mechanistic investigations of these latter preliminary observations are warranted.

Keywords: Non-small-cell lung cancer (NSCLC); ERCC1; PARP1; synthetic lethality; DNA repair; surrogate biomarker

Résumé

Excision Repair Cross-Complementation group 1 (ERCC1) est une enzyme de réparation de l'ADN fréquemment déficiente dans le cancer bronchique non-à-petites cellules. Bien qu'une expression faible d'ERCC1 soit prédictive de réponse aux sels de platine, l'efficacité des chimiothérapies à base de platine est limitée par leur toxicité et l'apparition de résistance, justifiant la nécessité de stratégies thérapeutiques alternatives. Par ailleurs, l'absence de test compagnon diagnostic permettant d'évaluer la fonctionnalité d'ERCC1 dans la pratique clinique empêche actuellement toute thérapie personnalisée basée sur le statut ERCC1.

Afin d'identifier de nouvelles stratégies thérapeutiques pour les tumeurs ERCC1-déficientes en exploitant le concept de létalité synthétique, des screens à haut-débit , utilisant des composés pharmaceutiques ou par ARN interférence, ont été réalisés dans un modèle isogénique de CBNPC déficient en ERCC1. Cette approche a permis d'identifier plusieurs inhibiteurs de poly(ADP-ribose) polymerase 1 et 2 (PARP1/2), tels l'opalarib (AZD2281), le niraparib (MK-24827) et BMN 673 comme sélectifs pour les cellules ERCC1-déficientes. Les mécanismes sous-tendant cette sensibilité sélective ont été étudiés, et les résultats suivants ont été mis en évidence : (i) les cellules ERCC1-déficientes présentent un blocage prolongé en phase G_2/M après exposition à l'olaparib ; (ii) l'isoforme 202 d'ERCC1, dont le rôle a été récemment mis en évidence dans la résistance aux sels de platine, module également la sensibilité aux inhibiteurs de PARP ; (iii) la déficience en ERCC1 est épistatique avec les défauts de recombinaison homologue (RH), malgré une capacité normale des cellules ERCC1-déficientes à former des foyers RAD51 ; ceci suggère qu'ERCC1 pourrait intervenir dans la réparation d'une lésion de l'ADN induite par l'inhibiteur de PARP1/2 en amont de l'invasion du brin d'ADN lors de la RH ; (iv) l'inhibition de l'expression de PARP1 par ARN interférence permet de restaurer la résistance aux inhibiteurs de PARP1/2, dans les cellules ERCC1-déficientes uniquement. Ces résultats suggèrent que les inhibiteurs de PARP1/2 pourraient représenter une nouvelle stratégie thérapeutique chez les patients dont la tumeur est déficiente en ERCC1 et un essai clinique va être mis en place pour évaluer cette hypothèse.

Afin d'explorer la présence de biomarqueurs de la fonctionnalité d'ERCC1, quatre approches ont été entreprises en parallèle dans le modèle isogénique de CBNPC déficient en ERCC1: (i) irradiation aux UV, afin d'évaluer la voie NER (Nucleotide Excision Repair); (ii) séquençage d'exome, dans le but de rechercher une signature génomique (ADN) ; (iii) analyse du transcriptome cellulaire, pour identifier des modifications d'expression d'ARN ; et (iv) SILAC (Stable Isotope Labeling by Amino acids in Cell culture) afin de comparer le protéome des cellules ERCC1-déficientes et ERCC1-proficientes. Ces approches ont permis d'identifier une potentielle signature génomique, ainsi que de biomarqueurs d'activité – guanine deaminase (GDA) et nicotinamide phosphoribosyltransferase (NAMPT). De plus amples validations et investigations mécanistiques de ces observations préliminaires sont actuellement requises.

Mots-clés : Cancer bronchique non-à-petites cellules (CBNPC) ; ERCC1 ; PARP1 ; létalité synthétiques ; réparation de l'ADN ; biomarqueur d'activité

Acknowledgements

I would like to thank Alan Ashworth and Chris Lord, who have welcomed me in their laboratory and have always been available for guidance, help and advice. In addition, I would like to thank them for the opportunities they have created for me, both during the time I was at The Institute of Cancer Research and for future collaborations.

I would also like to thank Jean-Charles Soria, who offered me the opportunity to go to The Institute of Cancer Research, and provided me essential support and guidance. I am also grateful to Fabrice André who accepted to be the Director of this PhD.

I would also like to acknowledge all people from the Institute of Cancer Research (especially Ilirjana, Rowan, Sara, Irene, Jess, Rachel, Yari and James) as well as from the U981 team (Ken, Luc, Nicolas and Céline) for their continuous assistance and the great times spent together. I guess the quantity of questions I have asked them during these two years probably reaches a volume of at least ten times the one of this manuscript…

I am also very grateful to Fabrice Barlesi and Olivier Delattre who accepted to review this thesis, with a special thanks to Olivier Delattre for the great start he has given to my scientific career and the way he has made me discover and love scientific research. I would also like to acknowledge Alexander Eggermont, who accepted to be the President of my PhD Viva jury.

And last but not least, I would like to thank my friends and family for their support over these two years, in particular Eléonore for her permanent happiness, cheeky sense of humour and extreme kindness towards her working Mum, and of course Romain, for his belief in me and unconditional love.

List of Figures and illustrations

List of Tables

List of Abbreviations

ALK	Anaplastic lymphoma kinase
BER	Base excision repair
BRCA	Breast cancer gene
CDK	Cyclin-dependent kinase
CHEK	Cell cycle checkpoint kinase
CIT	Citron Rho-interacting kinase
COX	Cyclo-oxygenase
CPD	Cyclobutane pyrimidine dimer
ddPCR	Droplet digital PCR
DDR2	Discoidin Domain Receptor 2
DGKA	Diacylglycerol kinase alpha
DR	DNA repair
DSB	Double-strand break
EC_{50}	Half maximal effective concentration
EGFR	Epidermal Growth Factor Receptor
ELISA	Enzyme-linked immunosorbent assay
ERCC1	Excision Repair Cross-Complementation group 1
FA	Fanconi Anemia
FDA	Food and drug administration
FGFR1	Fibroblast growth factor receptor 1
FISH	Fluorescence in situ hybridization
GDA	Guanine deaminase
HER	Human epidermal growth factor receptor
HIF1AN	hypoxia inducible factor 1, alpha subunit inhibitor
HR	Homologous recombination
HSP	Heat-shock protein
IC_{50}	Half maximal inhibitory concentration
ICL-R	Interstrand crosslink repair
ICR	The Institute of Cancer Research
IGFBP4	Insulin-growth factor binding-protein 4
IGR	Institut Gustave Roussy
IHC	Immunohistochemistry
IKKε	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase epsilon
KO	knock-out
KRAS	Kirsten rat sarcoma viral oncogene homolog
KTS	Kinome and tumour suppressor
LKB1	liver kinase B1
MAD	Median absolute deviation
MAPK	Mitogen-activated protein kinase
MEK	Mitogen-activated protein kinase kinase
MMR	Mismatch repair
MS	Mass spectrometry
MSH2	MutS protein homolog 2

MTH1	MutT homolog 1
mTOR	Mammalian target of rapamycin
NAD	Nicotinamide
NAMPT	Nicotinamide phosphoribosyltransferase
NAPRT	Nicotinic acid phosphoribosyltransferase
NER	Nucleotide excision repair
NF-κB	Nuclear factor of kappa light polypeptide gene enhancer
NHEJ	Non-homologous end joining
NNMT	nicotinamide N- methyltransferase
NPI	Normalised percentage inhibition
NSCLC	Non-small-cell lung cancer
NUDT1	nudix (nucleoside diphosphate linked moiety X)-type motif 1
OS	Overall survival
PAPPA	Pregnancy-associated plasma protein A
PARP	Poly(ADP-ribose) polymerase
PCA	Principal component analysis
PCR	Polymerase chain reaction
PDGFR	Platelet-derived growth factor β
PFS	Progression-Free Survival
PI3K	Phosphoinositide-3-kinase
PKM2	Pyruvate kinase M2
PLA	Phospho-ligation assay
PLK1	Polo-like kinase 1
PTEN	Phosphatase and tensin homolog
ROS	Reactive oxygen species
RRM1	Ribonucleoside-diphosphate reductase M1
SCLC	Small-cell lung cancer
SF	Surviving fraction
SHFM1	split hand-foot malformation type 1
SILAC	Stable isotope labeling by amino acids in cell culture
SIRT1	Sirtuin 1
SNV	Single nucleotide variant
SSA	Single-strand annealing
SSB	Single-strand break
TK1	Thymidine kinase 1
TKI	Tyrosine kinase inhibitor
TS	Thymidylate synthase
UBE2C	ubiquitin-conjugating enzyme E2C
UHRF1	Ubiquitin-like, containing PHD and RING finger domains, 1
UV	Ultra-violet
WB	Western blot
XP	Xeroderma Pigmentosum
XPF	Xeroderma Pigmentosum group F
γH2AX	phospho-histone 2AX

CHAPTER 1

Introduction

1.1. Lung cancer

1.1.1 Incidence, mortality and epidemiology

Lung cancer is the leading cause of cancer death worldwide. Every year, nearly 1.7 million new cases of lung cancer are diagnosed, with approximately 1.5 million deaths (representing almost 20% of all cancer deaths) (Jemal *et al.*, 2011; Lozano *et al.*, 2012). The median survival has recently improved from 9 to 12 months thanks to the introduction of novel targeted therapies (Sandler *et al.*, 2006; Reck *et al.*, 2013), but currently less than 15% of patients survive five years beyond their diagnosis (Jemal *et al.*, 2011). Two-thirds of patients present with locally advanced or metastatic disease (Morgensztern *et al.*, 2010), with half of them having distant metastases at presentation (Lozano *et al*, 2012). Although tobacco smoking accounts for the majority of lung cancer, approximately 10%-20% of patients with lung cancer are lifelong never smokers; lung cancer in the never smokers affects women disproportionately more often than men, and has completely different molecular characteristics (Subramanian *et al.*, 2007; Subramanian *et al.*, 2008; Yano *et al.*, 2011).

1.1.2. Histopathology and molecular classification

Historically, a binary classification was established that distinguished between small cell lung cancer (20% of cases) and non-small cell lung cancer (80% of cases). Subsequent histological advances have allowed the subdivision of non-small-cell lung cancers into adenocarcinoma, large-cell carcinoma and squamous cell carcinoma. In the last ten years, rapid progresses in cancer genome analysis have resulted in this classification being completely revisited in order to build a new molecular-based categorisation (Buettner *et al.*, 2013; Reck *et al.*, 2013) (Figure 1. 1).

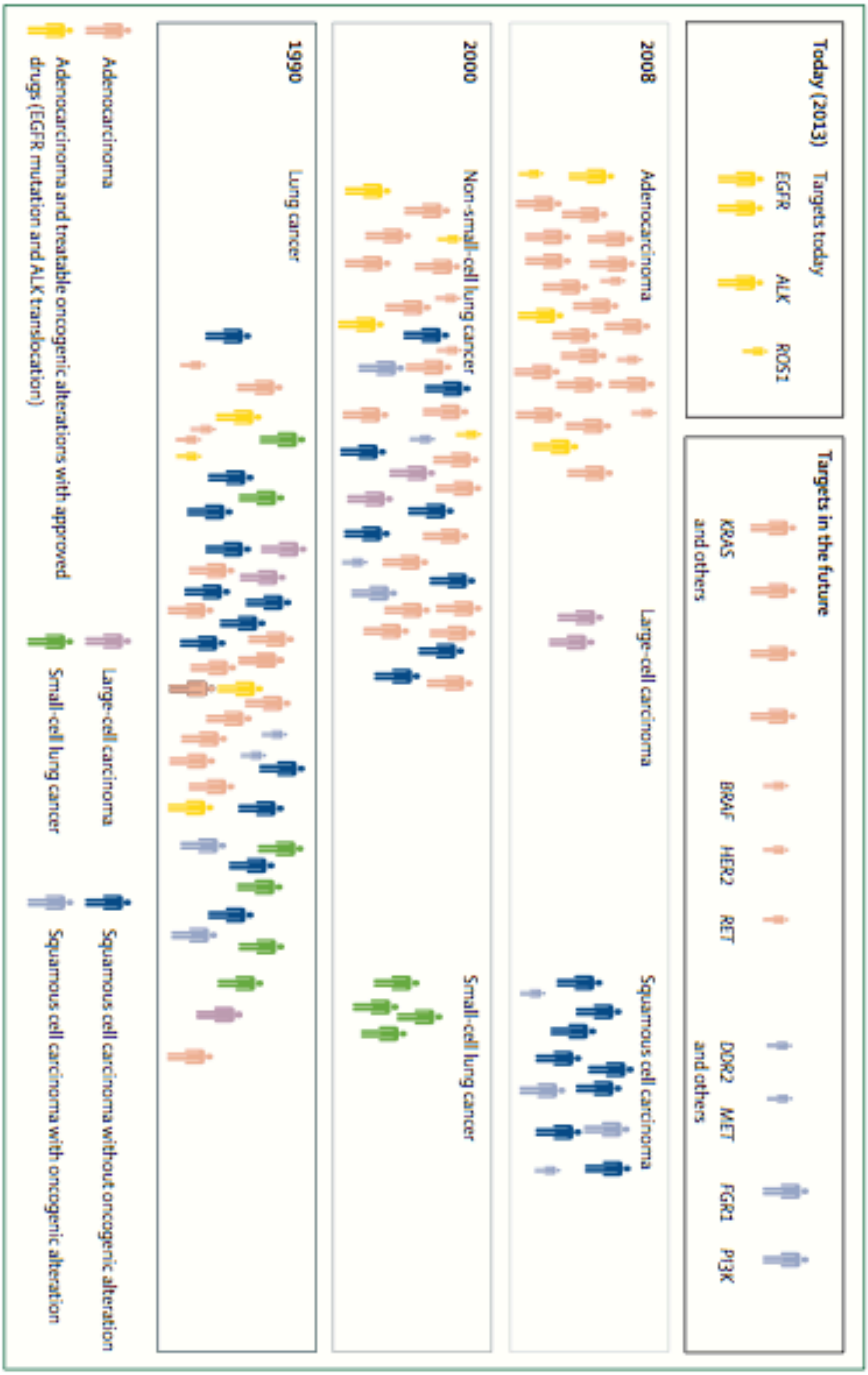


Figure 1. 1. Evolution of lung cancer histological and molecular subtypes over time (Figure and legends reproduced from Reck *et al.*, Lancet 2013)

Different colours denote different histological subtypes

Some of the molecular characteristics that define NSCLC have been identified. These alterations include mutations or translocations in oncogenes such as *KRAS* (mutated in 5–35% of patients) (Roberts *et al.*, 2013), *EGFR* (5–20%), *ALK* (5–10%), *HER2*, *BRAF*, *PIK3CA*, *MAPK2* and *MET* (all four mutated in <5% of patients) (Pao *et al.*, 2011) and tumour suppressor genes such as *TP53* (>50%) (Mogi *et al.*, 2011) and *LKB1* (50%) (Makowski *et al.*, 2008). Similar to most solid tumours, NSCLC is also characterized by genomic instability (Varella-Garcia, 2010), both in smoking and non-smoking associated NSCLC.

1.1.3. Treatment

Since the 1990s, the treatment of advanced NSCLC has consisted of platinum-based combination therapy, irrespective of the histological subtype. Platinum salts, either cisplatin or carboplatin, were combined with third-generation cytotoxic drugs such as gemcitabine, vinorelbine, docetaxel, and paclitaxel (Schiller *et al.*, 2002). The antifolate pemetrexed, approved in 2004 by the FDA, was the first antineoplastic agent that showed superior efficacy in a particular molecular subtype of NSCLC, i.e. non-squamous NSCLC – mainly adenocarcinoma (Scagliotti *et al.*, 2008).

In the last ten years, major progress has been made by the introduction of novel targeted therapies, associated with an appropriate patient selection based on the analysis of driver mutations. The three main molecular subsets that currently benefit from customised targeted therapies are the *EGFR*-mutated and *ALK*- and ROS-translocated populations. EGFR mutations are more frequent in Asian, women and non-smoker patients (Mok *et al.*, 2009; Rosell *et al.*, 2012). EGFR-targeting TKIs, such as erlotinib and gefitinib, have allowed a dramatic improvement in OS from 12 months to more than 2 years (Lynch *et al.*, 2004; Paez *et al.*, 2004; Kim *et al.*, 2008; Mok *et al.*, 2009; Rosell *et al.*, 2012). Similarly, the development of the ALK/ROS inhibitor crizotinib, has been paradigm-shifting for tumours bearing such driver alterations (Kwak *et al.*, 2010; Shaw *et al.*, 2011a; Shaw *et al.*, 2011b). Several additional mechanism-based targeted therapies are being developed for subsets of molecularly defined lung adenocarcinomas (Engelman *et al.*, 2008; Kohno *et al.*, 2012; Lipson *et al.*, 2012; Takeuchi *et al.*, 2012) or squamous cell carcinomas (such as *FGFR1* amplifications or *DDR2* mutations) (Weiss *et al.*, 2010; Hammerman *et al.*, 2011), which will hopefully soon transform the landscape of lung cancer treatment. Besides these therapies targeting driver oncogenic events, anti-angiogenic agents (bevacizumab) have demonstrated efficacy in combination with first-line platinum-based therapy with a median 2 months improvement in progression-free survival (Sandler, 2006; Reck *et al.*, 2010).

Finally, the use of maintenance therapy has been widely debated in NSCLC. Several studies have demonstrated that maintenance therapy prolongs overall survival in patients who have initially benefited from first-line platinum-based chemotherapy (Edelman *et al.*, 2012), and several drugs (including bevacizumab, pemetrexed and erlotinib), are currently approved in this indication (Sandler *et al.*, 2006; Reck *et al.*, 2010; Paz-Ares *et al.*, 2012).

1.2. DNA repair in NSCLC

Although most efforts have focused on targeting the "oncogene addiction" phenomenon through the identification of driver mutations, the targeting of DNA repair deficiencies also have the potential to become successful mechanism-based therapeutic approaches (Farmer *et al.*, 2005; Lord *et al.*, 2006; Martin *et al.*, 2006a; Lord *et al.*, 2012). Indeed, even if DNA damage and genomic instability are possible contributory factors to the aetiology of NSCLC, they also represent opportunities for therapeutic exploitation. Some of the favourable responses to DNA damaging chemotherapies or molecularly targeted therapies are influenced by tumour-specific DNA repair defects (Olaussen *et al.*, 2006; Jalal *et al.*, 2011; Lord *et al.*, 2012). Therefore, understanding the extent of DNA-repair defects in NSCLC is crucial if the standard therapies for this disease are to be used effectively.

1.2.1. DNA repair and carcinogenesis

The integrity of the DNA in each cell is continually challenged by hundreds of thousands of insults each day that can alter the sequence or chemical composition of the DNA. These lesions may come in the form of single-strand or double-strand DNA breaks, base damage, bulky adducts, intra and interstrand cross-links and breakdown of the replication fork. They may occur spontaneously (due to the inherent chemical instability of DNA) or be caused by agents such as UV light, ionizing radiation, environmental carcinogens, chemicals in cigarette smoke or chemotherapy (Jalal *et al.*, 2011).

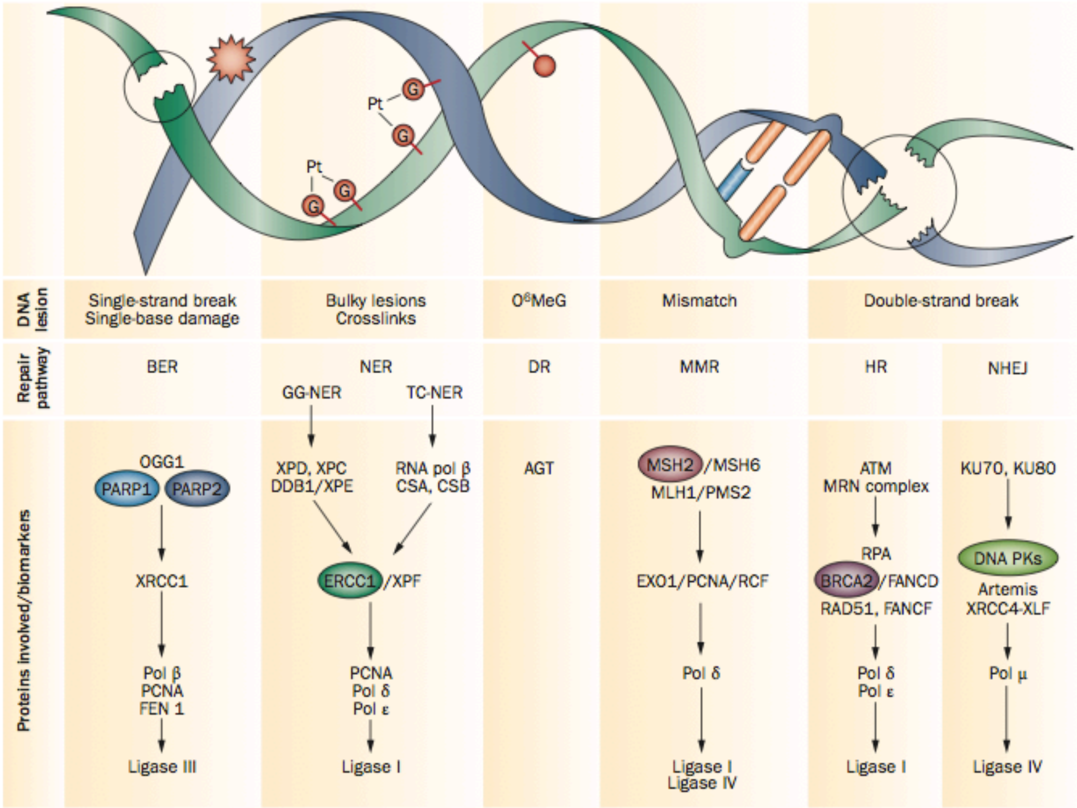


Figure 1. 2. The main DNA lesions and their corresponding DNA-damage-repair pathways. (Figure and legends reproduced from Postel-Vinay *et al.*, Nature Reviews Clinical Oncology 2012)

DNA lesions that affect a single strand without significantly disrupting the helical structure are generally repaired by BER, whereas DNA damage significantly distorting the DNA helix is repaired by NER. DR copes with small chemical changes affecting a single base, and MMR repairs mismatches in the pairing of DNA caused by replication errors. Finally, HR and NHEJ, although distinct pathways, are both involved in the repair of DNA double-strand breaks: HR allows 'error free' repair of the lesion whereas NHEJ is an 'error- prone' mechanism that repairs DNA but at the cost of introducing mutations into the genome. The selection of HR or NHEJ is primarily based on the phase of the cell cycle and the expression, availability and activation of DNA-repair proteins. Abbreviations: AGT, O6-alkylguanine-DNA alkyltransferase; ATM, ataxia telangiectasia mutated; BER, base excision repair; DR, direct repair; GG-NER, global genome NER; HR, homologous recombination; 0^6MeG, O^6-methylguanine; MMR, mismatch repair; NER, nucleotide excision repair; NHEJ, non-homologous end joining; TC-NER, transcription-coupled NER.

Normal cells exposed to DNA damage that threaten genomic integrity activate 'damage sensor' proteins, such as ataxia telangiectasia mutated (ATM), ataxia telangiectasia and Rad3-related (ATR), checkpoint kinase 1 and 2 (CHK1 and CHK2) or p53 (Kastan *et al.*, 2004). Once detected, each lesion can be repaired by at least one of the six major DNA repair pathways: BER (base excision repair); NER (nucleotide excision repair); DR (direct repair); MMR (mismatch repair); HR (homologous recombination) or NHEJ (non-homologous end joining) pathways. Other DNA repair pathways, such as Fanconi Anemia (FA), single-strand annealing or trans-lesion DNA synthesis, can also be activated to protect genomic integrity (Figure 1. 2) (Lord *et al.*, 2012; Postel-Vinay *et al.*, 2012).

Often described as a series of distinct pathways, DNA repair is more likely a complex and integrated network that coordinates the canonical pathways described above. This interwoven system enables cells to cope with the large burden of different types of DNA lesions and provides molecular redundancy when individual elements of the DNA-damage response are defective. As examples of pathway co-operation, both BER and NER can repair DNA lesions caused by alkylating agents or tobacco smoke (Pfeifer *et al.*, 2002), DNA interstrand cross-links caused by platinum salts are repaired by a sequential combination of NER and HR (Chen *et al*, 2009) and single-strand DNA breaks that fail to be repaired by BER often can ultimately cause double strand breaks (DSBs), a lesion repaired by HR.(Chen *et al.*, 2009) As these levels of functional interplay are starting to be unravelled, the potential for exploiting these effects therapeutically is also starting to be assessed.

In contrast to normal cells, cancer cells frequently fail to activate damage sensor proteins as DNA-repair pathways are often dysfunctional. This relative DNA repair deficiency stimulates mutagenesis and fosters tumourigenesis but, at the same time, may make tumour cells prone to the effects of DNA damaging chemotherapy (Helleday *et al.*, 2008; Lord *et al.*, 2012) From a clinical perspective, defects in DNA repair mechanisms are often associated with a bad prognosis as they likely enhance the progression of disease, but they may predict a better outcome after treatment as they may predispose cells to sensitivity to DNA damaging chemotherapy (Olaussen *et al.*, 2006). The key role of DNA repair in promoting tumourigenesis is highlighted by the numerous cancer predisposition syndromes and sporadic cancer cases associated with mutations in damage sensor or DNA repair genes (Olivier *et al.*; Peltomaki, 2003). NSCLC has mainly been linked with *p53* and *Rb* mutations (Birch *et al.*, 2001; Wikenheiser-Brokamp, 2006).

1.2.2. DNA repair as a therapeutic target in NSCLC: biomarkers of interest

In the future, the growing understanding of DNA-repair systems and DNA-repair defects in tumours may allow treatment refinement and hopefully deliver greater therapeutic windows (Lord *et al.*, 2012). In this context, a number of areas have received attention, such as defining molecular profiles in tumours that predict DNA repair defects and a favourable response to chemotherapy, and developing targeted agents that inhibit DNA repair enzymes (Figure 1. 3). These agents could be used either as radiosensitisers or chemosensitisers or as single agents in tumours with specific DNA repair defects. NSCLC displays frequent DNA repair deficiency (Table 1. 1), and several biomarkers have been evaluated in NSCLC clinical trials.

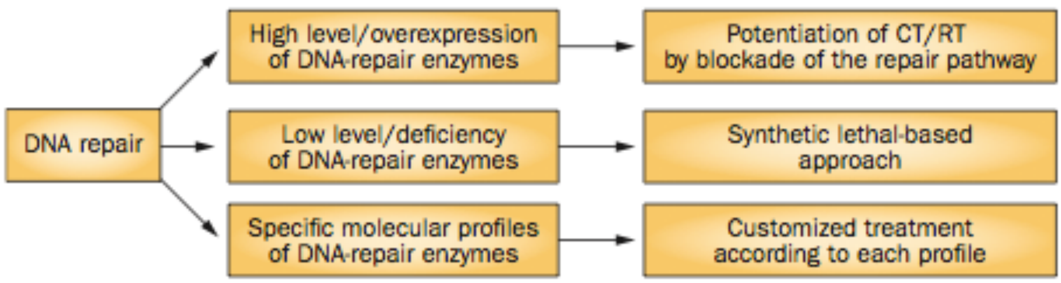


Figure 1. 3. DNA repair as a therapeutic target (Figure and legends reproduced from Postel-Vinay *et al.*, Nature Reviews Clinical Oncology 2012)

The therapeutic decision-making process can be guided by the DNA-repair profile. When the activity of DNA repair enzymes is high, chemotherapy is often inefficient as lesions caused by the DNA- damaging agent are adequately repaired. Therefore, specifically blocking the DNA-repair pathway can potentiate cytotoxicity. By contrast, when DNA-repair enzymes are deficient, synthetic lethality-based approaches can be used, in order to cause cell death following the inhibition of a distinct DNA-repair pathway from the one that is already deficient. Finally, when several isoforms are associated with different enzyme activity levels, molecular profiles should be analysed in order to customize treatment. Abbreviations: CT, chemotherapy; RT, radiotherapy.

1.2.2.1. ERCC1 and the NER pathway

ERCC1 is currently the most promising DNA-repair biomarker in the clinical treatment of NSCLC. ERCC1 is a structure-specific endonuclease that incises the damaged segment of the DNA and has a rate-limiting role in the NER-mediated repair of platinum adducts (Fagbemi *et al.*; Lindahl *et al.*, 1999). Consistent with this, low levels of ERCC1 are associated with better response to cisplatin (Altaha *et al.*, 2004; Olaussen *et al.*, 2006). As platinum-based therapy is the gold standard chemotherapy regimen in NSCLC (both in the adjuvant and in the advanced setting) evaluating NER functionality in this disease is of critical importance.

The prognostic effect of ERCC1 was first reported in 2005 by Simon and colleagues who reported a correlation of high *ERCC1* mRNA levels with better outcome in patients with resected NSCLC (median overall survival of 94.6 versus 35.5 months; $P = 0.01$) (Simon *et al.*, 2005). This finding was confirmed a year later by the International Adjuvant Lung Trial-bio (IALT-bio) study which assessed the protein levels of ERCC1 using immunochemistry (IHC) in 867 patients with NSCLC. Among patients who did not receive adjuvant chemotherapy, those with ERCC1-positive tumours survived longer than those with ERCC1-negative tumours (HR = 0.66 [adjusted for death]; 95% CI 0.49–0.90; $P = 0.009$) (Olaussen *et al.*, 2006; Olaussen *et al.*, 2007). The IALT-bio trial also evaluated the predictive value of ERCC1 expression for response to platinum-based therapy and revealed significantly prolonged survival among patients with ERCC1-negative tumours (HR = 0.65 [adjusted for death]; 95% CI 0.50–0.86; $P = 0.002$) compared with patients with ERCC1-positive tumours (HR = 1.14 [adjusted for death]; 95% CI 0.84–1.55; $P = 0.40$) (Bepler *et al.*, 2011) suggesting that patients with ERCC1-positive tumours did not benefit from adjuvant chemotherapy (Table 1. 2). Complementary analysis of the IALT also suggested that adjuvant chemotherapy was associated with an increased risk of brain metastasis only in patients with non-squamous cell carcinoma who were ERCC1-negative (HR = 1.4, 95% CI 0.90-2.1; $P = 0.04$) (Besse *et al.*, 2011).

In the advanced setting, the predictive value of ERCC1 for survival and sensitivity to platinum-based therapy has been widely studied although results have been variable and somewhat conflicting. A recent meta-analysis of 12 studies including a total of 836 patients that analysed ERCC1 status by IHC or real-time quantitative reverse transcription (qRT)-PCR, reported that median survival was significantly longer in patients with low levels of, or negative for, ERCC1 (median ratio MR = 0.77; 95% CI 0.47–1.07; *P*<0.00001) and that response to platinum-based therapy was significantly higher in this population of patients (odds ratio (OR) = 0.48; 95% CI 0.35–0.64, *P*<0.00001) (Chen *et al.*, 2010). Moreover, the results of the international phase III genotyping study–comparing a non-customized arm (patients assigned to receive docetaxel in combination with cisplatin regardless of their levels of *ERCC1*) with a customized arm (in which patients with low levels of *ERCC1* received docetaxel plus cisplatin and patients with high levels of *ERCC1* received docetaxel plus gemcitabine) reported a higher response rate in the customized arm (51% versus 39%, *P* = 0.02) and demonstrated feasibility of assessment of *ERCC1* mRNA levels in the clinical setting (Cobo *et al.*, 2007). However, this study did not confirm the predictive value for survival of low levels *ERCC1* (Lord *et al.*, 2002), as there was only a trend towards longer PFS in the customized arm without any improvement in overall survival (Cobo *et al.*, 2007).

Finally, although mutations in *ERCC1* in cancers are extremely rare, a synonymous polymorphism (a change in the base in the DNA sequence that does not alter the amino acid encoded) in exon 4 has been associated with changes in *ERCC1* mRNA levels and potentially correlates with poor prognosis in patients with advanced-stage colorectal cancer treated with platinum agents (Zhou *et al.*, 2004).

Several studies are currently ongoing, notably the phase III ET trial (ERCC1-Targeted Trial, NCT00801736), and the phase II CONTEST trial (Customized Neoadjuvant Versus Standard Chemotherapy in NSCL Patients With Resectable Stage IIIA (N2) Disease, NCT01784549), which evaluated customized *vs* standard chemotherapy according to ERCC1 status in advanced and resectable stage IIIA NSCLC, respectively. More importantly, results of the LACE (Lung Adjuvant Cisplatin Evaluation) Biology biomarker project, which aimed at validating the initial predictive value of ERCC1 reported in the IALT-bio trial, were reported during the course of this PhD. The LACE Biology project gathered tumour samples from three independent randomized phase III trials evaluating adjuvant cisplatin-based therapy, namely the IALT Biology trial, Cancer and Leukemia Group B (CALGB) 9633 study and the National Cancer Institute of Canada Clinical Trials Group JBR10 (Arriagada *et al.*, 2004; Winton *et al.*, 2005; Strauss *et al.*, 2008). The CALBG 9366 and JBR.10 trials constituted the validation series of the predictive value of ERCC1 detected in the IALT Biology samples in 2006 (Olaussen *et al.*, 2006). ERCC1 status was evaluated by immunohistochemistry using the 8F1 antibody. Surprisingly, results from the validation series failed to validate the previous results, which prompted to repeat the staining of formerly analysed samples obtained from the IALT Biology cohort. Several hypotheses were formulated to explain these results, including: (i) inadequate tools for evaluating ERCC1 expression; changes in the performance of the antibody used to perform the staining, which was initially reported as highly ERCC1-specific (Olaussen *et al.*, 2007), have been reported recently, notably with the recognition of a non-ERCC1 target (Ma *et al.*, 2012); (ii) underestimation of the level of biological complexity secondary to the presence of several isoforms, only one of which is functional (Friboulet *et al.*, 2013a). The strong sequence similarity between these isoforms precludes distinguishing between them, and does not allow the design of specific primers, antibodies or probes for qRT-PCR, ddPCR or FISH; and (iii) lack of predictive value for cisplatin-based chemotherapy of ERCC1.

Taken together, these data suggest that assessing ERCC1 status in NSCLC could provide essential information regarding prognosis and the likelihood of benefit from platinum therapy. Several hurdles still need to be overcome prior to using ERCC1 routinely as a predictive biomarker for cisplatin-based adjuvant chemotherapy. First, prospective validation is required, which had to be investigated within the phase II-III TASTE study (Tailored post-Surgical Therapy in Early stage NSCLC, NCT00775385), which involved more than 30 centres all over France. Results of the phase II part of the study were reported at ASCO 2013 (Soria et al. ASCO 2013), and demonstrated the feasibility of assessing at a national level a biomarker within 15 days after diagnosis. However, the phase III component of the study was cancelled due to the unexpected reliability of the ERCC1 IHC read-out, as explained above. Second, the optimal methodology to ascertain ERCC1 levels need to be determined: protein expression is a desirable end point for biologic significance, but fined-tuned measurements and functional assessments could not be carried out in the IALT-trial as specimens were stored and retrospectively collected (Bepler *et al.*, 2011). Moreover, the absence of validation of the initial IALT-Bio results is concerning, and technical biases that can interfere with sensibility or specificity of the results need to be resolved. Finally, further studies associating functional assessments of NER are warranted in order to develop mechanism-based therapeutic approaches.

Biomarker	Low protein levels of expression (% of cases)	Promotor methylation (% described; associated with low expression)	Other alterations	Correlations between biomarkers
ERCC1	51% (22–-66%) NSCLC	NA	NA	Correlation between ERCC1, RRM1 and BRCA1 levels (mRNA and protein)
RRM1	65% NSCLC	NA	NA	
BRCA1	37% NSCLC, 42% ADK, 30% of SCC	4%–30% (mainly in ADK and large cell carcinomas)	25% LOH, 14% promotor methylation of FANCF (conferring BRCAness phenotype)	BRCA1 mRNA levels correlated with RAP80 mRNA levels
BRCA2	34% NSCLC, 44% ADK and 24% SCC	0%–42% NSCLC	44% LOH	NA
MSH2	26% (18–38%) NSCLC, 9–35% SCC, and 29-30% of ADK	29–35% NSCL	MSI, 8%; LOH 54%	NA

Table 1. 1. Frequency of DNA repair alterations that can act as biomarkers in NSCLC

Abbreviations: ADK: adenocarcinoma; FANCF, fanconi anemia group f protein; LOH, loss of heterozygocity; MSI, micro-satellite instability; NSCLC, non-small-cell lung carcinoma; SCC, squamous cell carcinoma

Biomarker	Expres-sion level	Type of analysis	Pre-clinical	Pronostic and predictive value for OS and response [setting]	Prospective validation
ERCC1	Low	mRNA	Platinum sensitivity	Poor prognosis Longer OS for platinum-treated patients [advanced] Better response to platinum-based doublets [advanced]	Phase II (MADe IT) for feasibility Customized phase III
		protein	Platinum sensitivity	Poor prognosis Better response to platinum-based therapy [adjuvant; advanced]	Randomized Customized TASTE
		SNP	N/A	Longer OS for platinum-treated patients [advanced]	N/A
MSH2	Low	protein	Oxaliplatin sensitivity, cisplatin resistance	Poor prognosis Longer OS for cisplatin-treated patients [Adj] Better response to oxaliplatin + Gem [advanced]	N/A
		Promotor methylation	N/A	Poor prognosis in non-smoking women especially for early stages of NSCLC and ADK	N/A
		SNP	N/A	Good prognosis Better response to CDDP [advanced]	N/A
PARP	Low		Inhibition of PARP sensitizes cells to CDDP + RT		
BRCA1	High	mRNA	Modulator: Sensitivity to anti-microtubules; Resistance to DNA-DA and RT	Poor prognosis Longer OS for docetaxel-treated patients [advanced] Better response to docetaxel + gemcitabine [advanced]	Customized randomized Phase III BREC and SCAT studies (ongoing) N/A
	Low	mRNA	N/A	Longer OS for patients treated with gemcitabine plus cisplatin- [Neodajuvant and advanced]	N/A
RAP80	Low	mRNA	N/A	Longer OS for patients treated with gemcitabine plus cisplatin [advanced]	Customized randomized BREC study (ongoing)
RRM1	High	mRNA	Gemcitabine resistance	N/A	N/A
		Gene amplification	Gemcitabine resistance	N/A	N/A
	Normal	SNP	Gemcitabine sensitivity	N/A	N/A
	Low	mRNA	N/A	Longer OS for gemcitabine -treated patients [advanced] Better response to neo-adjuvant Gem Better response to Gem-based regimens [advanced]	Phase II (MADe IT) for feasibility
		Protein	N/A	Poor prognosis	N/A

Table 1. 2. DNA repair biomarkers in NSCLC and their predictive and prognostic value

Abbreviations: ADK, Adenocarcinoma; Adj, Adjuvant; Adv: Advanced; Nadj, Neoadjuvant; OS: Overall Survival; RT, radiotherapy; SCC, Squamous Cell Carcinoma; SNP: Single Nucleotide Polymorphism

1.2.2.2. BRCA1, PARP, RAP80 and the HR repair pathway

BRCA1 is a promising biomarker that could direct customized therapy in NSCLC. The tumour suppressor proteins BRCA1 and BRCA2 regulate the initial steps of HR by orchestrating the assembly of the DNA recombinase RAD51 onto broken DNA ends at the site of DSBs and stalled replication forks (Murphy *et al.*, 2010) This process can be visualised by IHC as nuclear RAD51 foci (Graeser *et al.*, 2010). Defects in *BRCA1* or *BRCA2* cause a profound HR defect that can be targeted by inhibiting PARP, a separate DNA repair enzyme, through synthetic lethality (Farmer *et al.*, 2005). Synthetic lethality relies on the fact that one DNA repair mechanism can compensate for deficiencies in another, and that simultaneous inhibition of both mechanisms causes cell death (Farmer *et al.*, 2005; Rehman *et al.*, 2010; Shaheen *et al.*, 2011). In some cases *BRCA*-mutant tumour cells are over 1,000 times more sensitive to potent PARP inhibitors compared with their *BRCA*-proficient counterparts (Bryant *et al.*, 2005; Farmer *et al.*, 2005). This level of genotype-specific selectivity and therapeutic potential prompted the clinical testing of these inhibitors as single agents rather than in combination with chemotherapy (Fong *et al.*, 2009).

Although germline or somatic *BRCA* mutations have only been described only rarely in NSCLC (Turner *et al.*, 2004), there seems to be some potential for exploiting PARP inhibitors in NSCLC if appropriate biomarkers can be developed. For example, a study of 98 samples of NSCLC reported low levels of BRCA1 or BRCA2 protein expression (<25% of tumour cells expressed these proteins) in up to 57% of NSCLC and 69% of adenocarcinomas (Lee *et al.*, 2007), events thought to occur as a consequence of epigenetic modulation of BRCA1 and BRCA2 (Marsit *et al.*, 2004). Another study on 126 samples of NSCLC reported that methylation of the promoter of the gene encoding Fanconi anemia group F protein (FANCF) occurred in 14% of NSCLC, potentially conferring a “BRCAness” phenotype (Marsit *et al.*, 2004) (i.e. a phenotype similar to the one observed in *BRCA*-mutated tumours, but which occurs in tumours without *BRCA* mutation). Interestingly, other synthetic lethal interactions have been described with PARP inhibition; notably, defects in PTEN or ATM can cause PARP inhibitor sensitivity (McCabe *et al.*, 2006; Mendes-Pereira *et al.*, 2009), and these genes are mutated in 5% and 6% of NSCLC, respectively. Furthermore, PTEN loss has been reported in 20–30% of NSCLC (Jin *et al.*, 2010; Rehman *et al.*, 2010). Consequently, therapeutic applications of PARP inhibitors may not be limited to the BRCA-deficient population, and evaluating these agents in patients with EGFR-mutant and PTEN-deficient NSCLC could be of interest, as PTEN loss contributes to erlotinib resistance in this population (Mendes-Pereira *et al.*, 2009). PARP inhibitors could also be combined with histone deacetylase (HDAC) inhibitors, as HDACs may be important enabling factors in HR. Finally, other synthetic lethal interactions, such as inhibition of CHK1 in FA-deficient tumours (Chen *et al.*, 2009) could also be exploited. However, as this approach has not been tested in NSCLC to date, more robust data are required before any clinical evaluation of this approach is carried out.

In addition to the synthetic lethality approach, inhibiting PARP could be used to potentiate chemotherapy as well as radiotherapy in NSCLC. In fact, *in vitro* studies report that PARP1 deficient cells are hypersensitive to DSBs (Farmer *et al.*, 2005) and that PARP inhibitors are strong radiation and cisplatin-sensitisers (Miknyoczki *et al.*, 2003; Albert *et al.*, 2007; Donawho *et al.*, 2007; Liu *et al.*, 2008; Powell *et al.*, 2010). No clinical data on PARP1 inhibitor efficacy in NSCLC is currently available, as the only phase III study that evaluated the combination of a putative PARP1 inhibitor to gemcitabine-carboplatin as first-line metastatic treatment in NSCLC used iniparib (ECLIPSE study, NCT01082549), which eventually turned out to display insufficient levels of PARP1 inhibition *in vitro*, and a ill-defined mechanism of action *in vivo* (Mendeleyev *et al.*, 1995; Liu *et al.*, 2012). The results of the phase II study evaluating the adjunction of veliparib to carboplatin-paclitaxel in advanced NSCLC (NCT01560104) are not available yet.

Although the clinical potential for using PARP inhibitors in NSCLC is yet to be established, there are fairly robust clinical data available describing the prognostic and predictive value of BRCA1 levels in NSCLC. High levels of *BRCA1* mRNA have been associated with poor prognosis in early stages of NSCLC in a study of 126 patients (Rosell *et al.*, 2007). These findings were replicated in two independent cohorts of 58 and 54 patients (Rosell *et al.*, 2007; Bartolucci *et al.*, 2009). This is noteworthy as high levels of DNA repair activity could be associated with better prognosis because they theoretically limit genomic instability and the progression of the disease. Although this finding could be related to the pleiotropic actions of BRCA1 — which are not limited to DNA repair — it highlights the importance of assessing both molecular levels of expression and functionality of biomarkers when studying DNA repair given that thresholds defined for molecular expression may not always reflect functional consequences.

Regarding the predictive value of BRCA1 expression, preclinical studies suggest that BRCA1 could modulate sensitivity to chemotherapy by enhancing apoptosis induced by antimicrotubule agents and conferring resistance to DNA damaging agents and radiotherapy (Quinn *et al.*, 2003; Chabalier *et al.*, 2006; Stordal *et al.*, 2009). In tumour cells isolated from NSCLC pleural effusions, low levels of *BRCA1* mRNA were correlated with sensitivity to cisplatin and resistance to docetaxel (Wang *et al.*, 2008). Similarly, low and high levels of expression of *BRCA1* mRNA have been associated with better outcome following neoadjuvant gemcitabine plus cisplatin therapy (Taron *et al.*, 2004) and gemcitabine/docetaxel treatment, respectively (Boukovinas *et al.*, 2008). The promise of BRCA1 status in the clinical setting was recently illustrated in two Spanish studies that customized treatment according to levels of *BRCA1* mRNA: cisplatin plus gemcitabine for patients with low levels of BRCA1; cisplatin plus docetaxel for patients with intermediate levels, and docetaxel alone for patients with high levels of *BRCA1*. The Spanish Customized Adjuvant Treatment (SCAT) pilot study, which was carried out in patients with completely resected stage II–IIIA NSCLC, suggested that there would be no detrimental effect on overall survival of docetaxel administered as a single agent in patients with high levels of expression BRCA1 (ASCO 2008, abstract 7533). As the number of patients was small, a prospective phase III validation study is currently ongoing (NCT00478699) to confirm the results of the pilot study. The second study, which was carried out in patients with metastatic disease with *EGFR* wild type tumours reported an excellent median 2-year survival of 41% in patients with the lowers BRCA1 expression (compared with 22% obtained in a recent randomized trial where patients received cisplatin-based regimens) (Rosell *et al.*, 2009). The randomized phase III BREC study (NCT00617656) is currently ongoing to confirm these results.

BRCA1 interacting proteins may also have predictive value for chemosensitivity response. For example, receptor associated protein 80 (RAP80) is a nuclear protein required for accumulation of BRCA1 and BRCA2 to sites of DNA damage. *In vitro* studies suggested that high RAP80 levels may compensate for BRCA1 deficiency and decrease platinum sensitivity in BRCA1-deficient cells (Yan *et al.*, 2007). The results of the Spanish study carried out in patients with metastatic NSCLC seemed to corroborate these findings; although mRNA levels of *RAP80* were correlated with mRNA levels of *BRCA1*, multivariate analysis revealed that RAP80 levels were an independent prognostic factor in patients treated according to levels of *BRCA1* (HR = 1.3, 95% CI 1–1.7; $P = 0.05$). More importantly, median overall survival was not reached in patients with low BRCA1 and low RAP80 levels, whereas it was 7 months in patients with high RAP80 levels and low BRCA1 levels (Rosell *et al.*, 2009).

1.2.2.3. MSH2 and the MMR pathway

MSH2 plays a key role in the MMR pathway through recognition of mispaired nucleotides resulting from replication errors as well as mismatched bases and DNA adducts induced by alkylating agents or antimetabolites (Lindahl *et al.*, 1999). Initial data regarding the prognostic and predictive value of MSH2 for response to chemotherapy in NSCLC were conflicting: two retrospective studies on 113 and 108 tumour samples, respectively, did not find any prognostic significance of MSH2 expression, (Cooper *et al.*, 2008; Kouso *et al.*, 2008) whereas Hsu *et al*. (Hsu *et al.*, 2005) reported that methylation of the promoter of *MSH2* was associated with poor prognosis in non-smoking women, especially for early stages NSCLC and adenocarcinomas. A clinical study of 93 patients with advanced-stage NCSLC reported that loss of expression of MSH2 was predictive of better response to oxaliplatin-based therapy and of resistance to cisplatin-based therapy (Scartozzi *et al.*, 2006). However, another study correlated the MSH2 gIV12-6T>C polymorphism — associated with low MSH2 expression — with a better response to cisplatin (Hsu *et al.*, 2007). The IALT-bio team recently studied the prognostic and predictive role of MSH2 by IHC on 673 tumour samples reporting that high MSH2 levels were a good-prognosis factor and there was a trend for chemotherapy to prolong overall survival in the presence of low levels of MSH2 (HR = 1.12; 95% CI 0.59–0.97; *P* = 0.03) (Kamal *et al.*, 2010) When combining MSH2 with ERCC1 into four subgroups (ERCC1-low and MSH2-low, ERCC1-low and HSH2-high, ERCC1-high and MSH2-high, ERCC1-high and MSH2-low), the benefit of chemotherapy decreased with the number of markers expressed at high levels. Looking at these results altogether, it is difficult to draw a clear picture of the role of MSH2 as a biomarker and its prognostic and predictive significance requires further investigation.

Interestingly, proof of concept that MSH2-deficient NSCLC could benefit from mechanism-based therapeutic approaches (e.g. synthetic lethality) was recently provided through the description of two synthetic lethal interactions between MSH2 deficiency and DNA polymerase-β inhibition,(Martin *et al.*, 2006b) and MSH2 and the antifolate methotrexate (Martin *et al.*, 2009). The interaction between MSH2 and methotrexate is of particular interest as low MSH2 expression has been reported in

18%–38% of NSCLC (Wang *et al.*, 2003; Hsu *et al.*, 2005; Kanellis *et al.*, 2006; Scartozzi *et al.*, 2006; Kamal *et al.*, 2010) (Table 1. 1) and methotrexate is a cousin of pemetrexed, a drug widely used in metastatic NSCC, NSCLC, SCLC and mesothelioma. Therefore, assessing whether pemetrexed benefit is higher in — or even limited to — MSH2-deficient patients could be incorporated into future NSCLC trials to further determine which population would benefit the most from antifolate-based therapy.

1.2.2.4. DNA protein kinase and the NHEJ pathway

NHEJ is an error-prone pathway that repairs DSBs by joining the ends of the broken DNA double-strands through the binding of a multi-protein complex containing DNA-dependent protein kinase (DNA-PK) (Lieber *et al.*, 2003). Reduced DNA-PK activity has been associated with an increased risk of NSCLC (Auckley *et al.*, 2001) and cytotoxicity of DNA-damaging agents can be enhanced by DNA-PK inhibitors in NSCLC cell lines (Eriksson *et al.*, 2001). One DNA-PK inhibitor, CC-115, is currently being evaluated in a phase I study (NCT01353625) and results are awaited. Moreover, DNA-PK belongs to the PI3K-related protein kinase family and PI3K inhibitors also inhibit DNA-PK, enhancing the cytotoxicity of radiation and topoisomerase inhibitors (Izzard *et al.*, 1999; Boulton *et al.*, 2000; Zhao *et al.*, 2006).

1.2.2.5. Nucleotide synthesis and DNA repair pathways

All DNA repair pathways require appropriate deoxyribonucleotides to synthesise new DNA at the site of damage. Ribonucleotide reductase 1 (RRM1) is the regulatory subunit of the ribonucleotide reductase enzyme that catalyses the reduction of ribonucleoside diphosphates to the corresponding deoxyribonucleotides and could also be a promising predictive biomarker in NSCLC. Interestingly, RRM1 is also the major molecular target of gemcitabine, which is widely used in metastatic NSCLC. *RRM1* is located on 11p15.5, a chromosome region with frequent loss of heterozygosity in NSCLC. The prognostic value of RRM1 was first reported in 2007 by Zheng and colleagues (Zheng *et al.*, 2007) in patients with stage I treatment-naive tumours who underwent complete surgical resection; levels of RRM1 were associated with a median overall survival of 60.2 and >120 months for low and high RRM1 expression, respectively.

RRM1 may also have predictive value; high RRM1 activity — subsequent to gene amplification (Tooker *et al.*, 2007), polymorphism (Kwon *et al.*, 2006), or mRNA overexpression (Davidson *et al.*, 2004) — has been associated with resistance to gemcitabine in NSCLC cell lines and animal models (Bergman *et al.*, 2005). Several large clinical studies have reported an association between low levels of *RRM1* mRNA and sensitivity to gemcitabine, in the neoadjuvant (Bepler *et al.*, 2008) and in the advanced setting (Rosell *et al.*, 2003; Rosell *et al.*, 2004a; Rosell *et al.*, 2004b; Ceppi *et al.*, 2006a; Boukovinas *et al.*, 2008; Souglakos *et al.*, 2008). Furthermore, RRM1 levels have also been reported to influence time to progression and overall survival in metastatic patients treated with gemcitabine plus cisplatin (Rosell *et al.*, 2003; Rosell *et al.*, 2004b). Based on these results, a customized phase II study was carried out, which included 85 patients assigned to received tailored chemotherapy according to *RRM1* and *ERCC1* mRNA levels, 53 of whom received treatment (Simon *et al.*, 2007). A partial response was observed in 44% of the patients and the median overall survival was 13.3 months. Although this phase II study had some limitations and comparison with historical data is not sufficient to drive robust conclusions, patient outcome seemed to be better than that reported in historical series (Chiappori *et al.*, 2005). Interestingly, as RRM1 levels have been reported to be closely related to ERCC1 and BRCA1 levels (Rosell *et al.*, 2004a; Rosell *et al.*, 2007; Simon *et al.*, 2007; Zheng *et al.*, 2007), the best chemotherapy regimen for patients expressing low levels of ERCC1 may be a combination of cisplatin and gemcitabine (Rosell *et al.*, 2003; Rosell *et al.*, 2004a; Ceppi *et al.*, 2006a).

Thymidylate synthase (TS) TS is a folate-dependent enzyme that catalyzes the methylation of deoxyuridine-5'-monophosphate (dUMP) using 5,10-methylene-tetrahydrofolate (5,10-CH2-THF) as the methyl donor to form deoxythymidine-5'-monophosphate (dTMP) (Carreras *et al.*, 1995; Rahman *et al.*, 2004). This function maintains the dTMP pool that is critical for DNA replication and repair, which in turn is essential for cell proliferation. TS is also the primary target of the multifolate inhibitor pemetrexed (Galvani *et al.*, 2011). Increased baseline expression of TS has initially been reported in squamous cell carcinoma tissues from chemonaïve patients, as compared to adenocarcinomas ($p<0.0001$) (Ceppi *et al.*, 2006b). Increased levels of TS have subsequently been correlated with reduced sensitivity to pemetrexed, both in NSCLC cell lines (Giovannetti *et al.*, 2005; Giovannetti *et al.*, 2008), and xenografts obtained with TS-overexpressing cells (Takezawa, BJC 2011). In the clinical setting, several studies have reported that lower TS expression levels are associated with better clinical outcome (Chen *et al.*, 2011; Sun *et al.*, 2011).

Pemetrexed is currently approved for non-squamous NSCLC patients as first line therapy in combination with cisplatin as well as for maintenance therapy. A prospective validation of the role of TS as a predictive biomarker for response to pemetrexed-based chemotherapy is currently ongoing through the international phase III ITACA trial (International Tailored Chemotherapy Adjuvant; EudraCT #: 2008-001764-36).

1.2.3. Challenges of developing biomarkers for DNA repair

If indeed the activity of particular DNA repair pathways affects therapy it would be helpful to have methods of directly analysing the activity of these pathways in patient samples. Some of the problems associated with DNA repair biomarkers are common to all biomarkers. These include consistency, threshold assignment and prospective validation. In the context of DNA repair, relevant biomarkers should ideally reflect the functionality of DNA repair pathways rather than only providing information limited to the level of expression or mutation status of the protein of interest. For example, studies assessing *BRCA1* status have almost exclusively used qPCR. Consequently, in the majority of these studies, patients were classified according to their gene expression levels by terciles (Boukovinas *et al.*, 2008; Rosell *et al.*, 2009). As cut-offs for defining the "low" and "high" level of expression were respectively defined as the lowest tercile and the highest tercile, and variable thresholds have been reported.

More interesting than the evaluation of the expression level of a single biomarker is the functional evaluation of a pathway. An example of functional assay to assess the activity of BRCA1 in the HR pathway is the formation of RAD51 foci after DNA damage (Graeser *et al.*, 2010). Lack of RAD51 focus formation (reflecting BRCA1 deficiency) might allow selection for those patients that might benefit from PARP inhibitors. Similarly, ERCC1 activity (and consequently NER functionality) could be indirectly assessed by IHC using the R-C18 antibody, which detects platinum adducts on tumour cells after treatment with cisplatin *in vitro*.(Dzagnidze *et al.*, 2007; Nel *et al.*, 2013)(Nel, BJC 2013) Such an assay could identify important applications in the clinical setting to detect patients who would (or would not) benefit from a platinum-based therapy. This is very important, for example, in the adjuvant setting, in which all patients with stage Ib-IIIa NSCLC indiscriminately receive platinum-based therapy for an absolute benefit of only 5.3% after 5 years. As highlighted by the initial results of the IALT-bio trial — ERCC1-positive population do not benefit from adjuvant platinum-based therapy — appropriate patient selection is crucial to avoid ineffective, or even deleterious treatments. This may be of most importance in stage I, in which the benefits of adjuvant chemotherapy have not been clearly demonstrated. In addition to IHC assays, high-throughput DNA sequencing platforms, multiplexed assays designed for NSCLC (such as the SNaPshot/PCR-based platform described by Su *et al.* (Su *et al.*, 2011)) or comparative genomic hybridization could provide key information about genotypes and genomic instability, which reflects the DNA-repair capacity of cancer cells. Other assays, such as host-cell reactivation, COMET, γH2AX foci formation and mutagen sensitivity assays, could also be applied although they have mainly been used in the context of epidemiological, screening or cancer prevention studies (Gorlova *et al.*, 2008; Orlow *et al.*, 2008). The report that the DNA repair capacity of peripheral lymphocytes evaluated by host-cell reactivation assay predicts survival of patients with NSCLC treated with platinum-based therapy, also opens new perspectives (Wang *et al.*, 2011b).

Finally, although mutational signatures (genomic scars) have mainly been used so far to correlate a genetic profile to tumour aetiology or mechanisms of tumourigenesis (i.e. exposure to mutagens etc.) (Pfeifer *et al.*, 2005; Stratton *et al.*, 2009; Pfeifer, 2010; Nik-Zainal *et al.*, 2012a; Nik-Zainal *et al.*, 2012b; Pena-Diaz *et al.*, 2012; Burns *et al.*, 2013; Korbel *et al.*, 2013), they could also represent promising selection biomarkers, if associated with specific defects in DNA repair pathways. Such genomic scars have recently been described in tumours with defective HR or microsatellite instability (Jiricny, 2006; Stephens *et al.*, 2009; Barber *et al.*, 2011).

Another key issue is the material on which the biomarker should be assessed: primary or secondary tumours, circulating tumour cells, or host. Regarding tumour analysis, core biopsies and fine-needle aspiration are usually sufficient to allow histological and IHC characterization. However, tissue samples are often limited in quantity, and the difficulty of obtaining serial lung tumour specimens, which are imperative in assessing the pharmacodynamic activity of a drug and exploring predictive biomarkers to personalize therapy, is obvious. Moreover, analysis based on a limited biopsy or cytology specimen is potentially confounded by the issue of cancer heterogeneity.

Beyond the choice of the material and method to analyse DNA repair, remains the question of the timing: when should the tumour DNA repair capacity be analysed?

Pre-treatment biopsies are useful in the case of germline mutations, but of limited interest for functional tests if the tumour has not been exposed to any DNA damaging agent. As exposing a patient to a toxic drug, followed by re-biopsy only in order to assess a molecular biomarker of response is obviously not acceptable and unethical, *in vitro* assays have to be developed. Evaluation of HR functionality by assessing RAD51 foci formation on tumour biopsy after *ex-vivo* tumour irradiation is one approach. Other techniques – e.g. exposing cells to cisplatin and evaluating the clearance of platinum adducts – are currently being investigated (Nel *et al.*, 2013).

Finally, the germline characteristics of patients need to be considered as these are at least partially maintained in the tumour. Several polymorphisms in DNA repair genes have been correlated with treatment outcome in NSCLC (Gurubhagavatula *et al.*, 2004; Zhou *et al.*, 2004; de las Penas *et al.*, 2006; Camps *et al.*, 2007; Matakidou *et al.*, 2007). For example, a study of Japanese patients revealed that germline polymorphisms in *TP53* and *PARP1* were correlated with sensitivity to platinum-based doublets in NSCLC patients (Shiraishi *et al.*). This may be an important observation, if it were to be widely applied to other populations. As single nucleotide polymorphisms (SNPs) can easily be examined using blood cells or buccal swab at low cost, they also represent very promising biomarkers for treatment guidance. However, using germline data as a predictive biomarker in the decision-making process of chemotherapy treatment should be validated extremely carefully prior to any translation in the clinical setting, especially if it potentially results in treatment reduction. If combined, all techniques described above could allow the formulation of a comprehensive “DNA repair molecular portrait” of the tumour, which could eventually guide the therapeutic strategy and allow optimization of the treatment choice (Figure 1. 4).

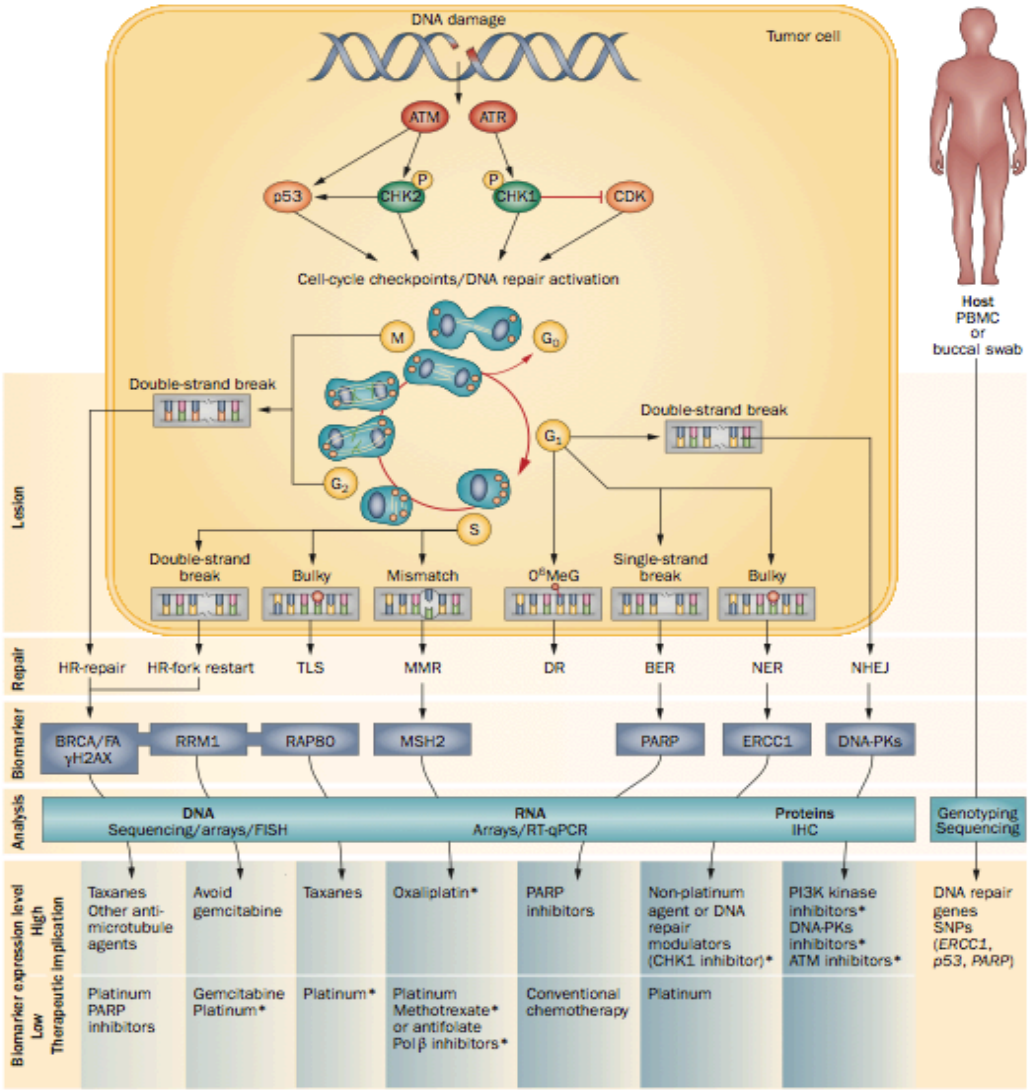


Figure 1. 4. DNA-repair biomarkers and therapeutic implication (Figure and legends reproduced from Postel-Vinay *et al.*, Nature Reviews Clinical Oncology 2012)

In cancer cells, DNA lesions activate damage-sensor proteins such as ATM–ATR and downstream proteins such as CHK1 and CHK2. ATM–CHK2 and ATR–CHK1 complexes orchestrate cell-cycle arrest through inhibition of CDK and phosphorylation of p53. Depending on the cellular context, p53 triggers a transcriptional program causing DNA repair and survival, apoptosis or senescence. DNA lesions can be repaired by several repair pathways depending on the cell-cycle phase and the type of the lesion. Analysis of the expression level of each corresponding biomarker has potential therapeutic implications. Similarly, the host DNA-repair profile analysis can bring significant information to guide therapeutic choice. *Preclinical or not yet validated data. Abbreviations: ATM, ataxia telangectasia mutated; ATR, ATM and Rad3-related; BER, base excision repair; CDK, cyclin-dependant kinase; CHK1 and CHK2, checkpoint kinase 1 and 2; DR, direct repair; FISH, fluorescence *in situ* hybridization; HR, homologous recombination; IHC, immunohistochemistry; 06MeG, O6-methylguanine; PBMC, peripheral blood mononuclear cell, Abbreviations: MMR, mismatch repair; NER, nucleotide excision repair; NHEJ, non-homologous end joining; RT-qPCR, quantitative reverse-transcription PCR; SNP, single nucleotide polymorphism; TLS, translesion DNA synthesis.

1.2.4. DNA repair as a therapeutic target: synthetic lethality

The therapeutic decision making process can be guided by the tumour DNA repair profile, as illustrated above. Among the mechanism-based approaches exploiting DNA repair as a therapeutic target that have been described (Lord *et al*., 2012), synthetic lethality is currently the most promising.

Synthetic lethality describes a relationship between two genes whereby loss of either gene is compatible with cell viability, but concomitant inhibition or loss of both genes causes cell death (Figure 1. 5) (Farmer *et al.*, 2005; Bryant *et al.*, 2005). An extension of this concept is synthetic sickness, which is the situation where the combination of two defects impairs cell fitness without overtly killing them (Kaelin, 2005). These concepts easily find clinical applicability when a synthetic sickness / lethality (SSL) relationship can be found between a tumour suppressor or deficient DNA repair gene, and a second gene. This latter gene then becomes a candidate therapeutic target, which allows selective cell kill (or impact of cell fitness) in cancer cells deficient for the tumour suppressor or DNA repair gene only, while normal cells are left intact.

The clinical applications of synthetic lethality are currently srestricted to the use of PARP inhibitors in *BRCA*-deficient tumours, but several *in vitro* and preclinical results describing other synthetic lethal relationship could be clinically exploited in the new future (Lord *et al.*, 2012). PARP1 and PARP2 are the best characterised members of the PARP family of proteins. These enzymes modify conformation, activity or stability of target proteins by adding polymers of ADP-ribose (PARsylation) using βNAD+ as a substrate (Rouleau *et al.*, 2010). PARP1 plays a key role in the initiation of BER, where it detects and binds SSB. This triggers the activation of the catalytic activity of the protein, which results in the PARsylation of PARP1 itself (auto-PARSylation) as well as a number of other proteins directly or indirectly involved in the repair process (such as XRCC1 or histone H2B) (Rouleau *et al.*, 2010). When PARP1 activity is inhibited, single strand DNA breaks are not effectively repaired. These persistent single strand DNA breaks have the capacity to stall and subsequently cause the collapse of replication forks when cells attempt to replicate DNA, resulting in potentially lethal breaks in both strands of the double helix (Tutt *et al.*, 2005).

In normal cells where the DNA repair machinery is intact, these DSB can be efficiently repaired by HR, a high-fidelity mechanism of DSB repair. In cancer cells where HR is defective, such as *BRCA1* or *BRCA2* mutated tumours, these DSB cannot be repaired and cause cell death. In some *in vitro* models, BRCA-mutant cells are over 1000 times more sensitive to PARP inhibitors than the BRCA-wild type cells (Farmer *et al.*, 2005). Importantly, the profound sensitivity of BRCA-deficient cells is not limited to a specific histology, as similar results have bee observed in breast, pancreatic and cervical tumour cell lines with pre-existing or experimentally-imposed *BRCA* gene defects (Farmer *et al.*, 2005; Edwards *et al.*, 2008). Moreover, rather than being specific of a peculiar genotype, this exquisite PARP inhibitor sensitivity is rather defined by a phenotype of HR deficiency, also called the “BRCAness phenotype” (Turner *et al.*, 2004)(Turner, NRC 2004), which includes other genetic defects such as *RAD51*, *ATM*, *ATR*, *CHEK1*, *CHEK2*, *CDK1*, *SHFM1* and the *FANC* family of genes (McCabe *et al.*, 2006; Akamatsu *et al.*, 2010; Johnson *et al.*, 2011; Moskwa *et al.*, 2011; Peasland *et al.*, 2011).

The proof-of-concept phase I study that evaluated olaparib (AstraZeneca, AZD2281) enrolled 60 patients. Although the initial inclusion criteria of the dose-escalation phase did not mandate a germline BRCA1 or BRCA2 mutation, only BRCA mutation carriers were eligible for being enrolled in the expansion cohort. This appropriate patient selection allowed demonstrating, in the clinical setting, the dramatic activity of PARP inhibitors in BRCA mutation carriers (Fong *et al.*, 2009). These results were subsequently confirmed in two phase II studies in breast and ovarian cancer (Audeh *et al.*, 2010; Tutt *et al.*, 2010), which both limited patient inclusion criteria to *BRCA* mutation carriers. As traditional histopathological methods and gene-expression profiling have suggested some overlaps between triple-negative breast cancer, basal-like breast cancer and BRCA1-mutant familial breast cancer, it was hypothesised that this subset of breast cancer could also respond to PARP inhibitors. Several clinical trials that evaluate PARP inhibitors in patients with triple-negative are currently ongoing (www.clinicaltrial.gov), but it is at this time too early to confirm in the clinical setting the putative PARP inhibitor sensitivity of this patient population (Balmana *et al.*, 2011).

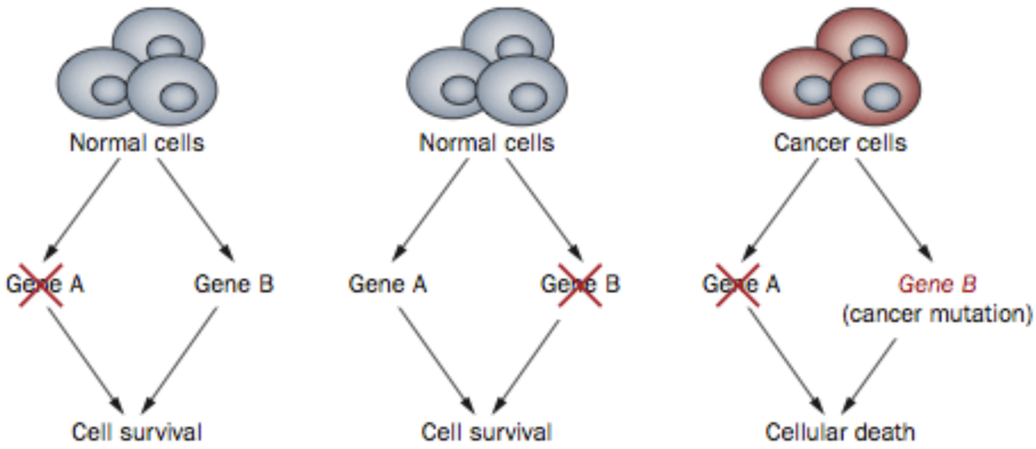


Figure 1. 5. Synthetic lethality (Figure and legends reproduced from Rehman *et al*., Nature Reviews Clinical Oncology 2010)

Loss of either gene 1 or gene B in normal cells is compensated by the action of the remaining gene. In tumour cells, a mutation in one of these genes leaves the cell vulnerable to loss of the other gene by drug inhibition. This approach is the basis of drugs that target synthetic lethal relationships. By contrast, normal tissues are spared any toxic effects

1.3. Excision Repair Cross-Complementation Group 1 (ERCC1)

As described above, ERCC1 is currently the most promising DNA repair biomarker for the clinical treatment of NSCLC, and has been the centre of much interest in this disease.

1.3.1. DNA repair functions of ERCC1

ERCC1 (Excision Repair Cross-Complementation Group 1) is a 297 amino acid protein that plays a central role in the NER pathway, and has additional roles in ICL-R, and DSB repair mechanisms (Kirschner *et al.*, 2010; Gregg *et al.*, 2011; McNeil *et al.*, 2012).

The *ERCC1* gene, located on 19q13.2-q13.3, has 10 exons and codes for at least four different isoforms by alternative splicing, namely isoform 201, 202, 203 and 204, which were largely uncharacterised at the beginning of this PhD. Work performed by the Institut Gustave Roussy U981 INSERM Unit during the PhD demonstrated that isoform 202 was the only functional isoform, at least with regards to all DNA repair functions of ERCC1 (Friboulet *et al.*, 2013a; Friboulet *et al.*, 2013b). Deletions in the interaction domain with XPF have been described in ovarian cell lines but not in lung cancer cell lines (Sun *et al.*, 2009). The ERCC1 main protein partner, XPF (also known as ERCC4) possesses the catalytic activity within a nuclease domain, and a also harbours a double helix-hairpin-helix $(HhH)_2$ domain that ensures the interaction with ERCC1 (Tripsianes *et al.*, 2005; Su *et al.*, 2012). Although only XPF carries the nuclease domain, it requires ERCC1 for subsequent nuclease activity (Tripsianes *et al.*, 2005). Interestingly, ERCC1 central domain resembles the nuclease domain of XPF structurally, but the active endonuclease site is replaced by basic and aromatic residues for ssDNA (single-strand DNA) binding (Su *et al.*, 2012) (Figure 1. 6). The ERCC1/XPF heterodimer forms a structure specific endonuclease, which binds the DNA on the 5' side of the lesion: indeed, the central domain of ERCC1 binds ssDNA/dsDNA (single-strand DNA / double-strand DNA) junctions with a defined polarity, preferring a 5' single-stranded overhang (Tsodikov *et al.*, 2005).

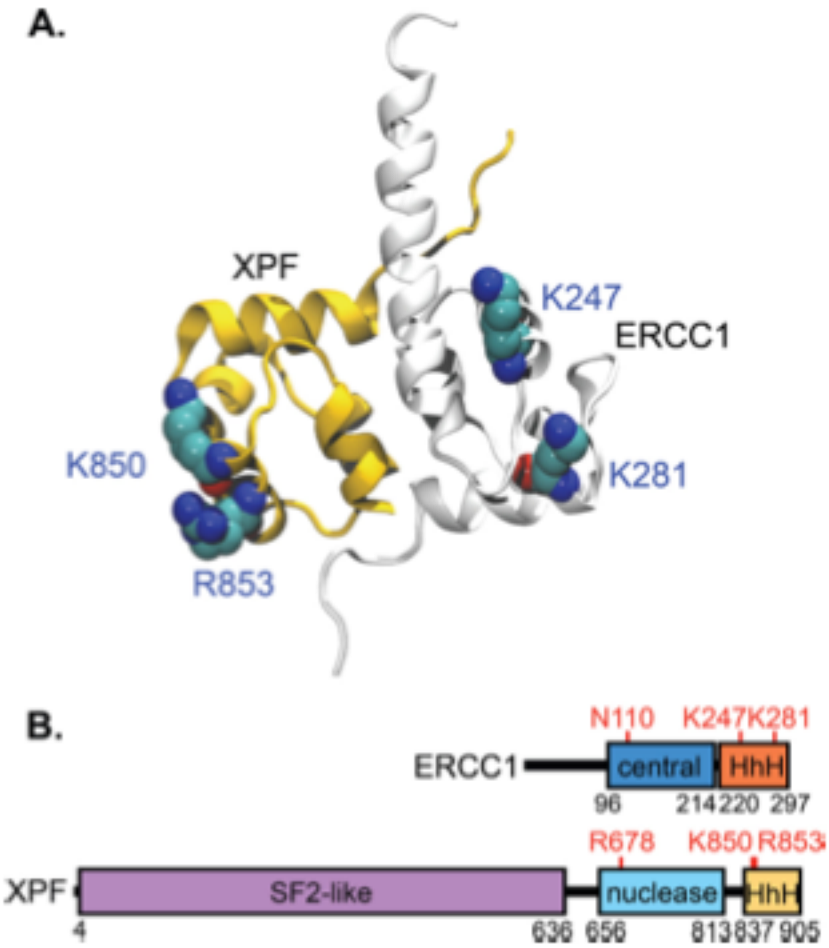


Figure 1. 6. Structure of the ERCC1/XPF HhH_2 domains and scheme of the DNA binding domains in ERCC1/XPF (Figure and legends reproduced from Su *et al*., Journal of Biological Chemistry 2012)

A) HhH2 domains of ERCC1 (grey) and XPF (yellow) contribute to dimerization. The residues shown in atom colour are believed to contribute to DNA binding. B) Five domains of ERCC1/XPF that may contribute to DNA binding are shown. Residues in the central and HhH2 domains of ERCC1 and in the nuclease and HhH2 domains of XPF that are believed to play an important function are highlighted in red

Other protein partners of ERCC1 include XPA and MSH2 (Tripsianes *et al.*, 2005; Tripsianes *et al.*, 2007; Tsodikov *et al.*, 2007; Friboulet *et al.*, 2013a). XPA binding has been reported to be exclusively important in NER, whereas XPF is involved in all DNA repair functions of ERCC1. It has been suggested that ERCC1 and XPF were unstable in the absence of each other (Gaillard *et al.*, 2001).

1.3.1.1. Nucleotide Excision Repair (NER)

ERCC1 is a key component of both GG-NER (Global Genome NER) and TC-NER (Transcription-Coupled NER) where it plays a rate-limiting role (Fagbemi *et al.*, 2011). NER is involved in the repair of all “bulky” helix-distorting lesions blocking DNA or RNA polymerases. These include mostly exogenous lesions, such as intra- or inter-strand adducts (such as platinum-, polycyclic aromatic hydrocarbons- or tobacco-induced lesions) and pyrimidine dimers or 6-4 pyrimidine-pyrimidinone photoproducts induced by UV light (Lindahl *et al.*, 1999). Mutations in genes encoding proteins involved in NER are responsible of the xeroderma pigmentosum or trichothiodystrophy (XPA-XPG) and Cockayne’s (CSA and CSB) syndromes, all three characterized by a profound sun sensitivity. Interestingly, only xeroderma pigmentosum is associated with an increased skin cancer risk (Lehmann, 2001), whereas CSA/B deficiency is rather associated with increased cell sensitivity to apoptosis or impairment of transcription (Lu *et al.*, 2001).

Two separate NER pathways have been described: the global genome NER (GG-NER, acting anywhere in the genome at any time) and the transcription-coupled NER (TC-NER, only detecting lesions that interfere with elongation of the RNA polymerases) (Hoeijmakers, 2001). GG-NER is initiated by the recognition complex XPC-HR23B whereas TC-NER initially involves the CSA and CSB. The following steps are common to both pathways: DNA is locally unwound around the damaged site by a TFIIH complex; XPB (ERCC3) and XPD (ERCC2) helicases maintain, with opposite polarity, the DNA-helix in an open conformation, allowing the recruitment of the repair enzymes (XPA-RPA complex and XPG) (Figure 1. 7).

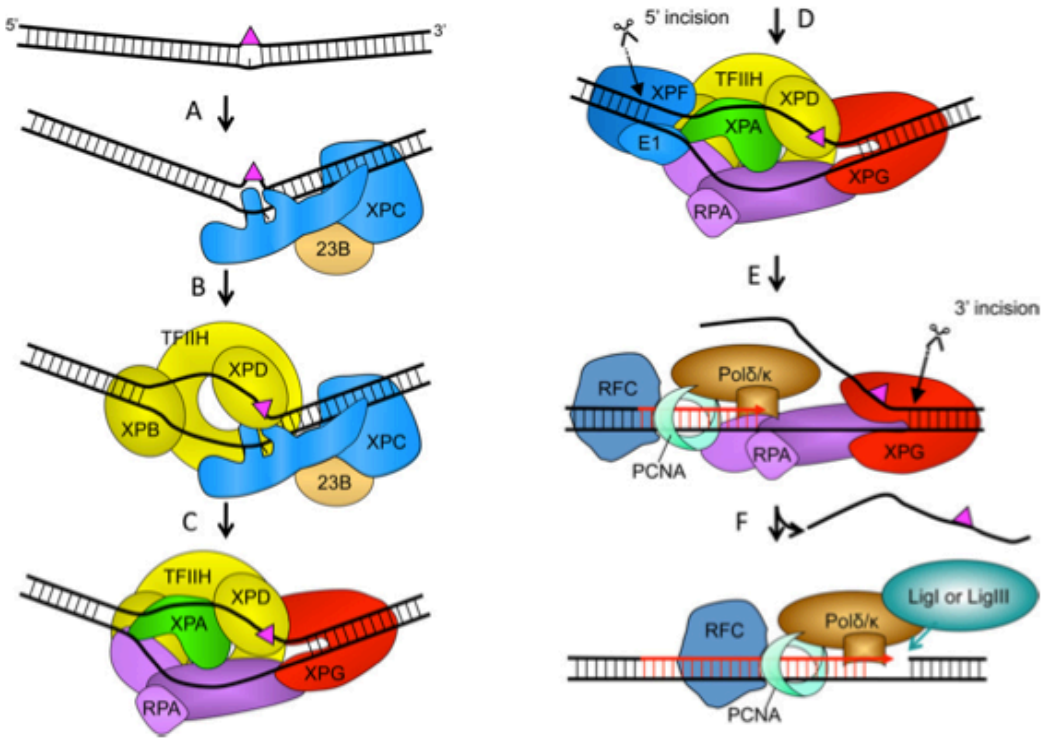


Figure 1. 7. Model for the nucleotide excision repair pathway (NER) (Figure and legends reproduced from Fabgemi *et al*., DNA repair 2011)

A) A bulky DNA lesion is recognized by XPC-RAD23B, which binds to the undamaged strand of DNA, allowing for the recruitment of TFIIH. B) The lesion is verified by XPD, resulting in the recruitment of XPA, RPA and XPG leading to the formation of the preincision complex. C) ERCC1/XPF is recruited to the preincision complex through interaction with XPA leading to DNA incision 5' to the damage, which produces (D) a free 3'-OH group available for initiation of repair synthesis by the replication machinery, and in turn triggers 3' incision by XPG. E) Repair synthesis is completed and DNA ligase IIIα or ligase I seals the nick to complete the process (F)

XPA recognises and verifies the lesion while RPA binds the single-stranded DNA. ERCC1 first interacts with XPA at the pre-incision complex (Li *et al.*, 1994; de Laat *et al.*, 1999). Interaction of both XPF and ERCC1 with RPA (replication protein A) is then triggered and enables the optimal positioning of the complex (Matsunaga *et al.*, 1996; Bessho *et al.*, 1997). ERCC1/XPF heterodimeric endonuclease then cuts the DNA on the 5' side of the lesion at the double-strand / single-strand junction, followed by XPG (ERCC5) that excises the damaged strand on the 3' side of the lesion, liberating a 24-32 nucleotides fragment. The polymerisation and ligation of the newly synthesized DNA strand can then occur, involving several DNA polymerases (notably Polδ or Polε) (Fagbemi *et al.*, 2011).

1.3.1.2. Interstrand crosslink repair (ICL-R)

Fewer molecular and mechanistic details are known regarding the role of ERCC1/XPF in ICL-R and several models have been proposed (as exemplified in Figure 1. 8 and Figure 1. 9, which illustrate two of the mechanisms that have been proposed). Interstrand crosslink repair is an interesting example of the cooperation between the canonical DNA repair pathways, as it involves sequentially four different mechanisms. ICLs result mainly from cisplatin derivatives or mitomycin C treatment; their presence causes important distortions from the DNA helix as well as prevents strand separation during DNA replication. The repair of the lesion occurs generally during the S phase: schematically, the FANCM protein (FA pathway) initially recognises the lesion and the FA core complex as well as endonucleases (notably ERCC1/XPF) that cleave DNA on both sides of the lesion allowing the extrusion of the covalent binding (Niedernhofer *et al.*, 2004; Zhang *et al.*, 2007; Al-Minawi *et al.*, 2009) (Figure 1. 8). Trans-lesion Synthesis (TLS) then allows the bypass of the cross-link, followed by NER, which excises the damaged DNA. After filling of the gap, HR mechanism can take place: XRCC1 and XRCC2 (paralogs of RAD51) are recruited to ensure “invasion” of the damaged DNA in the normal sequence of the homolog sister where the correction of the lesion can be achieved. ICL-R thus allows an error-free repair at the replication fork (Deans *et al.*, 2011).

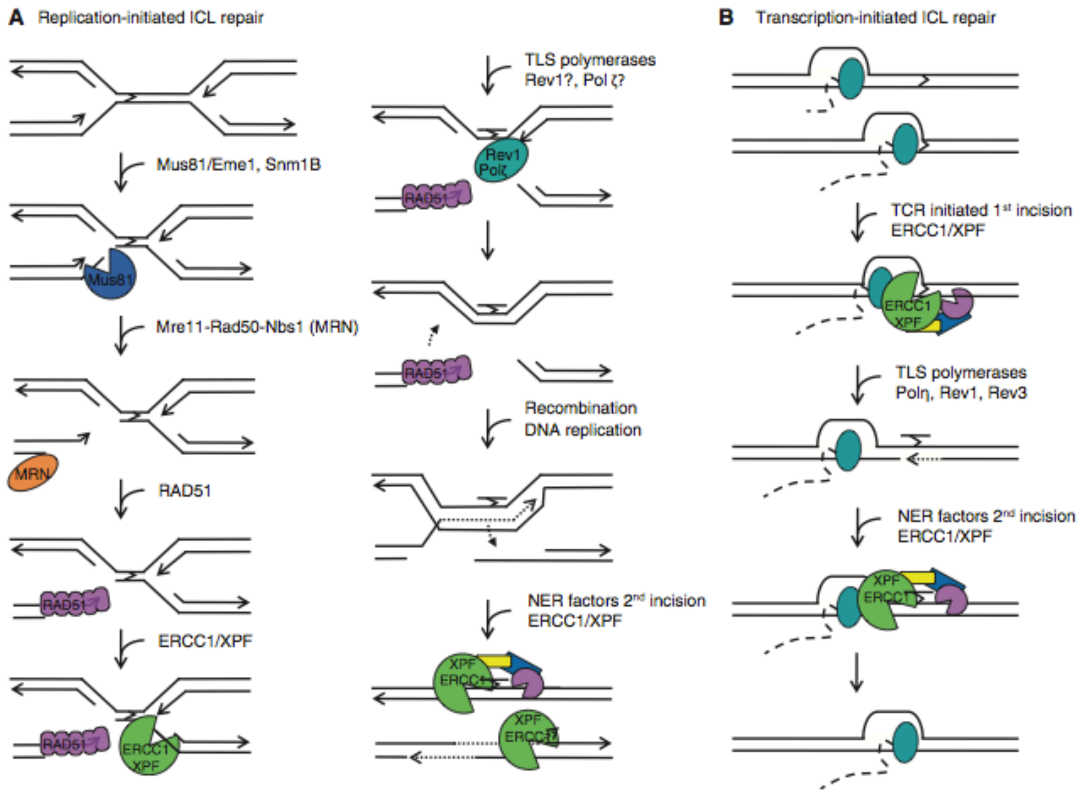


Figure 1. 8. A first model for the role of ERCC1 in ICL-R in mammalian cells (Figure and legends reproduced from Al-Minawi *et al.*, Nucleic Acids Research 2009)

ICL repair can be initiated either at a replication fork (A) or at a stalled RNA polymerase (B). A) ICLs in DNA initially stall replication forks that collapse into a one-sided DSB by Mus81 activity. The released one-sided DSB Is likely resected by the MRE11-RAD51-Nsb1 complex, and subsequently coated with the RAD51 protein to promote HR at a later stage. An opposing second stalled replication fork is possibly processed by the ERC1/XPF complex to unhook the crosslink and to allow TLS. The DNA molecule would be invaded by RAD51 during HR to initiate synthesis-dependent strand annealing repair and the final lesion removed by NER. B) ICLs in DNA will stall RNA polymerase during transcription that will initiate TCR. The RNA polymerase with either backtrack or be degraded during subsequent repair, which proceeds through TLS. NER factors will be attracted for second incision that will remove the ICL to allow resumption of transcription.

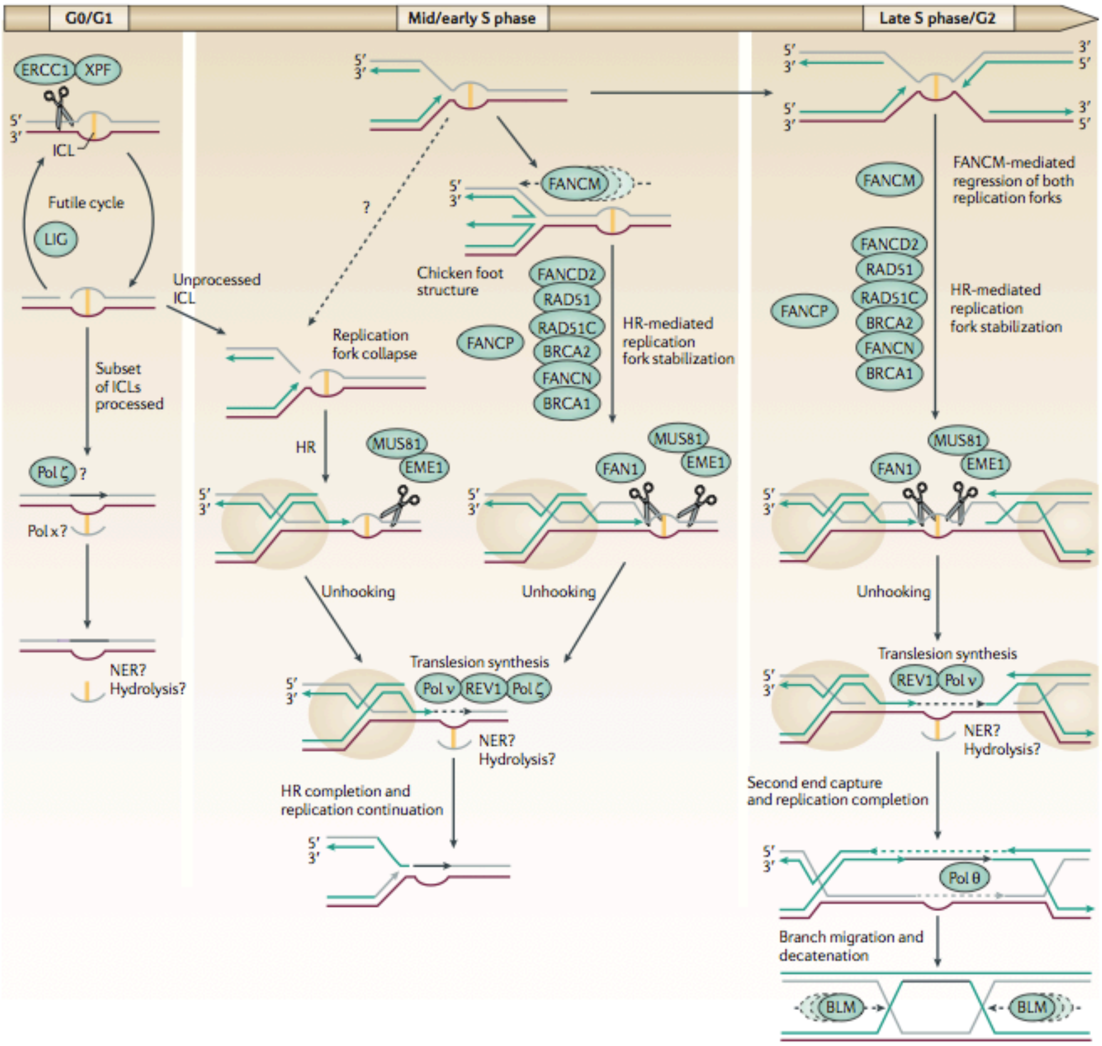


Figure 1. 9. A different model example for the role of ERCC1/XPF in ICL-R (Figure and legends reproduced from Deans & West, Nature Reviews Cancer 2011)

In cells in G1 phase of the cell cycle, NER can remove a subset of ICLs. However, some lesions cannot be bypassed in this manner and a futile cycle of repair involving excision and ligation occurs. During S phase, a partially processed intermediate may be encountered by a replication fork, which leads to the collapse of the replication fork and formation of a one-ended DSB. FANCM-mediated fork regression and stabilization allows the nascent leading strand of DNA synthesis to be extended by TLS. Unhooking of the ICL by coordinated incision by the 5’ and 3’-flap endonucleases allows extension past the lesion. HR is then completed and replication can be re-established.

In the absence of ERCC1/XPF, replication-dependent crosslink-induced DSBs occur, indicating that ERCC1/XPF is required for the extrusion and/or the removal of the lesion (Niedernhofer *et al.*, 2004). More recently, it was reported that the unhooking of an ICL by XPF-ERCC1 was necessary for the stable localization of FANCD2 to the chromatin and subsequent homologous recombination-mediated DSB repair, suggesting a role for ERCC1 in the Fanconi Anemia pathway of crosslink repair (Bhagwat *et al.*, 2009a). The cell cycle phase at which ERCC1/XPF takes part and whether this would be part of the NER or FA pathway is still unknown (Bhagwat *et al.* 2009a, Deans *et al.*, 2011) (Figure 1. 9). Finally, a role for ERCC1 in the completion of HR in at DNA replication forks stalled by interstrand crosslinks has also been reported (Al-Minawi *et al.*, 2009).

1.3.1.3. Double-strand break repair

DSBs are considered to be the most lethal type of DNA lesion (Ward *et al.*, 2012): failure to repair DSB can cause cell death or significantly threat cell survival by generating major genetic instability through a variety of alterations such as deletions, loss of heterozygosity, translocations and chromosome loss (Hoeijmakers *et al.*, 2001). Two main mechanisms are involved in the repair of these lesions: homologous recombination (HR) and non-homologous end-joining (NHEJ). The choice of the repair pathway is mainly based on the cell cycle phase during which the lesion is occurring and the expression, availability and activation of repair proteins (Shrivastav *et al.*, 2008).

HR mainly occurs during the late S/G2 to M phase, following DNA replication, and allows a reliable error-free repair of DSBs and covalent interstrand DNA cross-links. The DNA DSB is initially recognised by the MRN complex associating MRE11 (meiotic recombination 11) to RAD50 and NBS1 (Nimjen Breakage Syndrom 1) (Shrivastav *et al.*, 2008). The phosphorylated form of H2AX then binds to chromatin areas flanking the break while the phosphorylation of MRN by ATM stimulates its 5’-3’ exonuclease activity allowing the formation of two protrusive 3’ single strand extremities. These latter proteins are immediately coated and protected by RPA (Replication Protein A), subsequently replaced by RAD51. These initial steps are thoroughly regulated by BRCA1 and BRCA2, which play a crucial role and notably stimulate (in association with RAD52) the assembly of the RAD51 complex by displacing RPA. The RAD51 complex guides homology search and allows the “DNA invasion" of the single-strand coated DNA into the normal sequence of the undamaged sister chromatid or homologous chromosome. Subsequent DNA synthesis and ligation from the 3' end of the broken helix leads to the formation of "double Hollidays junctions" resolved notably by the ReQ helicase BLM in association with topoisomerase IIIα (Sharma *et al.*, 2007).

NHEJ is an error-prone pathway of repairing DSBs, which can result in deletion or insertion of base pairs and occurs during the G1 phase of the cell cycle (Lieber *et al.*, 2003). Instead of using a single-strand template for repair, this mechanism consists through end-joining of the broken double-strands DNA trough the binding of a multi-protein complex to the broken DNA extremities. This complex includes the heterodimer Ku70/80, which exhibits sequence-independent affinity for double-strand extremities and recognises the lesion. It subsequently recruits the catalytic subunit of DNA-dependant protein kinase (DNA-PKcs) and the Artemis nuclease (Lieber *et al.,* 2003), which processes DNA ends. Ligation is then completed by another complex including XRCC4 and DNA ligase IV.

Single-strand annealing (SSA) and microhomology-mediated end-joining (MMEJ) are a particular form of homology-directed repair (HDR) occurring when homologous DNA sequences (usually tandem repeat sequences) are flanking the lesion (Jackson *et al.*, 2002; Zheng *et al.*, 2006). The matching of these two lateral sequences requires RAD52, RAD59, as well as ERCC1/XPF, and usually allows "illegitimate" error-prone recombination. This mechanism is however rarer and less reliable than the classical HR using single-strand DNA invasion: SSA is usually used when HR or NHEJ are impaired (Ahmad *et al.*, 2008) and for interstrand crosslinks (Zheng *et al.*, 2006).

The ERCC1/XPF endonuclease is required for efficient SSA and gene conversion in mammalian cells, but the role of ERCC1 in these mechanisms is still ill-defined. Several models have been proposed (Sargent *et al.*, 1997; Niedernhofer *et al.*, 2001; Ahmad *et al.*, 2008; Al-Minawi *et al.*, 2008). Overall, it seems that the heterodimer is mainly involved when the broken DNA ends contain 3'-overhanging unmatched sequences or ends that cannot be used to prime DNA synthesis (Ahmad *et al.*, 2008; Gregg *et al.*, 2011). Finally, the interaction reported between ERCC1/XPF and SLX4 (aka FANCP) may mediate its role in the resolution of specific DNA structures formed during the HDR or DSB, ICR or collapsed replication forks (Al-Minawi *et al.*, 2009; Bhagwat *et al.*, 2009a; Munoz *et al.*, 2009; Svendsen *et al.*, 2009)

1.3.1.4. Telomere maintenance

A potential role for ERCC1/XPF has been reported in telomere maintenance, where it interacts with the telomere binding protein 2 (TRF2). When ERCC1 is defective in a TRF2-deficeint background, ERCC1-deficient cells accumulate double-minutes (Zhu *et al.*, 2003; Munoz *et al.*, 2005; Munoz *et al.*, 2009). This suggests that ERCC1/XPF has a role in the cleavage and degradation of G-rich 3'-overhangs rendering chromosomes vulnerable to end-to-end fusions.

1.3.1.5. ERCC1 at fragile sites

Very recent work published during the course of this PhD reported that ERCC1 co-localised with FANCD2 on mitotic chromosomes, and that depletion of ERCC1 led to an increase in the frequency of chromosome bridges during anaphase, as well as fragile site damage at the following G1 phase (Naim *et al.*, 2013).

1.3.2. Non-DNA repair functions of ERCC1

Two main non-DNA repair functions have been described for ERCC1. First, ERCC1 has been be involved in mitotic progression. Hepatocytes of ERCC1-deficient mice display a high polyploidy rate (Nunez *et al.*, 2000), and ERCC1-KO cells present an increased multinucleation associated with cytokinesis defects (Rageul *et al.*, 2011). Interestingly, multinucleation was not observed in XPF-KO cells, suggesting a role for ERCC1 in mitotic progression that would be independent from XPF. Second, very recent work published during the course of the PhD reported a role for ERCC1/XPF in transcription initiation: using promotor occupancy studies combined with expression profiling of livers of ERCC1-/- mice, Kamileri and colleagues showed that ERCC1/XPF assembled on active promoters *in vivo* and facilitates chromatin modifications for transcription during mammalian development. This role in initiating transcription of genes associated with growth was suggested as a major causal contribution to NER developmental disorders (Kamileri *et al.*, 2012).

1.3.3. ERCC1 deficiency in mice

ERCC1 knockout mice display severe postnatal growth failure and die before weaning, at approximately 3 weeks of age, with a body weight of only about 20%

compared to their normal littermates and an aging phenotype (McWhir *et al.*, 1993; Weeda *et al.*, 1997). ERCC1-/- mice spontaneously develop symptoms characteristic of progressive neurodegeneration, such as dystonia and ataxia (Niedernhofer *et al.*, 2006).

1.3.4. ERCC1 deficiency in humans

ERCC1-deficiency is rare in humans, and only three patients have been described so far. Overall, ERCC1-deficiency in humans is associated with much more severe symptoms and premature aging (Gregg *et al.*, 2011) than other NER-associated genetic disorders, namely xeroderma pigmentosym (XP), trichothiodystrophy (TTD) and Cockayne syndrome (CS) (reviewed in (Kraemer *et al.*, 2007)).

The first patient identified with *ERCC1* mutation presented severe pre- and postnatal developmental defects (Jaspers *et al.*, 2007), associated with cerebro-oculo-facia-skeletal syndrome (COFS). Main disabilities at birth included skeletal defects such as microcephaly, arthrogryposis and rocker-bottom feed, together with neurological alterations and cerebral hypoplasia. Other COFS have been reported in patients with mutations in *ERCC6*/*CSB*, *ERCC5*/*XPG* and *ERCC2*/*XPD* (Nouspikel *et al.*, 1997; Meira *et al.*, 2000). One truncating mutation was found in the maternal allele, resulting in the lack of the whole XPF binding domain, and a missense mutation was found on the paternal allele, in the conserved HhH_2 domain required for XPF binding. Fibroblasts from the patient had 15% of the normal NER, suggesting that the mutations might affect either the stability of the protein or its nuclear localisation, rather than its activity (Gregg *et al.*, 2011). A second patient with mutations in *ERCC1* was described in 2007, with a less severe phenotype: the patient displayed progressive neurodegeneration beginning at age 15 and died at 37 (Imoto *et al.*, 2007). This patient had a splicing mutation, and a non-sense mutation lying in the HhH_2 domain. During the course of this PhD, a third patient, initially thought to suffer from Cockayne Syndrome, was reported (Kashiyama *et al.*, 2013). *ERCC1* sequencing revealed a homozygous missense mutation in exon 7, associated with undetectable levels of ERCC1. The patient also displayed severe COFS and skeletal abnormalities at birth, and died at the age of 2.5 years.

1.4. Objectives of this PhD

In the era of personalized medicine, DNA repair biomarkers will likely play a crucial role in determining the optimal use of chemotherapy. Therapeutic successes obtained with synthetic lethality demonstrate that DNA repair can be considered as a therapeutic target, and the promising results obtained in HR-deficient breast and ovarian cancer might be translated into NSCLC.

ERCC1 is currently the most promising DNA repair-related biomarker in NSCLC, and low levels have been correlated to platinum salts sensitivity (Olaussen *et al.*, 2006; Postel-Vinay *et al.*, 2012). Unfortunately, platinum administration has to be halted after 4 to 6 cycles, because of cumulative haematological toxicities or irreversible neurotoxicity, even if tumour shrinkage is still observed. Also, not all patients are eligible for platinum therapy (because of comorbidities or impaired general fitness), and resistance ultimately always develops. Therefore, the first aim of this PhD was to investigate DNA repair mechanism-based approaches for ERCC1-deficient NSCLC, and notably synthetic lethal relationships, through high-throughput approaches using RNA interference or drugs chemicals (Chapters 3 and 4).

The feasibility of using biomarkers such as ERCC1, BRCA1 or RRM1 routinely is currently being investigated in NSCLC. Beyond the assessment of a single biomarker, a functional assessment of DNA repair pathways is essential and clinical trials that evaluate the assays currently under development are warranted. Such functional evaluation would allow to eventually adding a functional DNA repair classification in addition to the classical histological and current molecular ones, ultimately leading to the design of new therapeutic approaches and clinical benefit. The second aim of this PhD was accordingly to investigate surrogate biomarkers of ERCC1 / NER activity in NSCLC, which would allow customising treatment (Chapter 5).

Eventually, the concept of synthetic lethality may be extended to interactions between DNA repair deficiencies and other cell signalling abnormalities, such as activated oncogenes or growth factors (Lord *et al.*, 2012). Therefore, the third aim of this PhD was to explore the therapeutic potential of mechanism-based approached in different molecular or functional subsets of NSCLC through a comprehensive profiling of NSCLC cell lines, using high-throughput screening (Chapter 6).

CHAPTER 2

Material and Methods

2.1. Reagents

2.1.1. General chemicals and solutions

All chemicals were purchased from Sigma unless otherwise stated.

PBS: 137 nM NaCl, 2 mM KCL, 8 mM Na_2HPO_4, 1.5mM KH_2PO_4 in H_2O, pH adjusted to 7.4 with HCl.
IFF: immunofluorescence buffer (PBS plus 1% bovine serum albumin (Sigma, UK) and 2% foetal calf serum (Invitrogen, UK), filtered through a 0.2 µm filter).
NP250 lysis buffer: 20 mM TrisCl pH 7.6, 250 mM NaCl, 1mM EDTA pH8, 0.5% NP40, water, phosphatase and protease inhibitors (complete protease inhibitor, Roche).
PFA: 4% (w/v) paraformaldehyde in PBS.
PI staining solution: propidium iodide 20 µg/ml, RNase A 100 µg/ml in PBS.
Triton solution: 0.25% Triton-X 100 in PBS.
Transfer buffer: 14.4 g gylcine, 3.03 g Tris, 200ml methanol made up to 1 litre with H_2O.
10x TBS-T: 200ml 1 M Tris pH 7.5, 300 ml 5M NaCl, 10 ml Tween in 1 L of H_2O.
Sulforhodamine B: 0.057% sulforhodamine B (w/v) in 1% acetic acid (v/v) in H_2O.
TCA: 10% trichloroacetic acid in H_20.

2.1.2. Drugs and chemotherapeutics

The in-house drug screen library drugs are listed in Supplementary table 1. Each compound was dissolved in 100% DMSO to give 5mM stocks and then diluted to 0.5, 0.05, 0.005, 0.0005mM stocks in 96-well 2D-matrix plates. Daughter plates in 384-well format were prepared from these 96-well 2D-matrix racks using the Hamilton Microlab Star robotic platform. Compounds were stored under a nitrogen atmosphere using a StoragePod® (Roylan Developments).

Other drugs used in this PhD are listed below:

- Olaparib: from Selleck, dissolved in DMSO
- Niraparib: from Selleck dissolved in DMSO
- Cisplatin: from Sigma, dissolved in 0.9% NaCl
- BMN 673: from BioMarin Pharmaceuticals, dissolved in DMSO
- FK866: from Sigma, dissolved in H_2O
- MK-1175 (Wee1 inhibitor): from ChemieTek, dissolved in DMSO

2.1.3. Antibodies

Antibody	Application	Dilution	Incubation	Manufacturer	Product #
β actin HRP	WB	1:1000	1h RT	Santa Cruz	sc-1616
β tubulin	WB	1:5000	o/n 4° C	Sigma	T8328
BRCA2	WB	1:1000	o/n 4° C	Calbiochem	OP-95
CIT	WB	1:1000	o/n 4° C	Abcam	ab86782
CPD	IF	1:1000	1h RT	Cosmo Bio	CAC-NM-DND-001
CKS1	WB	1:1000	o/n 4° C	Invitrogen	370200
DGKA	WB	1:1000	o/n 4° C	Abcam	ab88672
ERCC1	WB	1:1000	o/n 4° C	Santa Cruz	sc-53281
ERCC1	WB	1:1000	o/n 4° C	Santa Cruz	sc-10785
GDA	WB	1:1000	o/n 4° C	Abcam	ab155773
HSPA2	WB	1:1000	o/n 4° C	Novus Biologicals	NBP1-33501
IGFBP4	WB	1:1000	o/n 4° C	Millipore	06-109
NAMPT	WB	1:1000	o/n 4° C	Novus Biologicals	NB100-594
NAPRT	WB	1:1000	o/n 4° C	Santa Cruz	sc-87326
NNMT	WB	1:1000	o/n 4° C	Santa Cruz	sc-376048
PAPPA	WB	1:1000	o/n 4° C	Pierce	MA1-46425
PARP1	WB	1:1000	o/n 4° C	Santa Cruz	sc-8007
PARP14	WB	1:1000	o/n 4° C	Santa Cruz	sc-377150
Phospho (Ser 139) H2AX	IF	1:1000	1h RT	Millipore	05-636
PKM2	WB	1:1000	o/n 4° C	Cell Signaling Technology	3198S
RAD51	IF	1:1000	1h RT	Santa Cruz	sc-8349
SIRT1	WB	1:1000	o/n 4° C	Santa Cruz	sc-15404

TK1	WB	1:1000	o/n 4° C	Abcam	ab76495
UBE2C	WB	1:1000	o/n 4° C	Abcam	ab56861
UHRF1	WB	1:1000	o/n 4° C	Abcam	ab57083
XPF	WB	1:1000	o/n 4° C	Abcam	ab85140

Table 2. 1. List of antibodies

IF=Immunofluorescence; WB=Western blotting; o/n=overnight

2.1.4. siRNA library targeting the human genome

The Dharmacon® SMARTpool® kinase siRNA library and DNA Repair Custom library were obtained from Dharmacon. The libraries target 1763 phylogenetically related protein kinase family genes and DNA repair genes. The libraries were supplied in 96 well plate format with each gene represented by one well, with 80 genes per plate. Each well contained a siRNA SMARTpool® with four individual siRNAs targeting the same gene but with different target sequences. Target sequence information was not supplied. The siRNA libraries were aliquotted using the Hamilton Robot into 384 well plates for the screens.

2.1.5. siRNA oligonucleotides

siRNA oligonucllеotides were purchased from Dharmacon or Qiagen and were supplied 2'ACE protected and lyophilized. These were reconstituted to 20 μM in DEPC treated water (Ambion) and stored in aliquots at -20°C. The following non-targeting siRNA negative controls are used in this thesis and target no known human genes: siCON1 or siCON2 (Dharmacon) and AllStar negative control siRNA (Qiagen). The final siRNA concentration per well in the 384 screening plates was 20 nM.

2.2. Protocols

2.2.1. Tissue culture

2.2.1.1. Cell lines

All cell lines were obtained from ATCC (USA) unless otherwise stated and maintained according to the supplier's instructions.

Cell line	Tissue of origin	Source	Media
A427	Lung	ATCC	EMEM (Gibco)
A549	Lung	ATCC (Institut Gustave Roussy)	DMEM (Gibco)
A549	Lung	ATCC (The Institute of Cancer Research)	RPMI (Gibco)
BEN	Lung	Institut Gustave Roussy	DMEM (Gibco)
DLD1 BRCA2 -/- (null/deficient)	Colorectal	Horizon	DMEM (Gibco)
DLD1 BRCA2 +/+ (wild type)	Colorectal	Horizon	DMEM (Gibco)
HOP-62	Lung	Institut Gustave Roussy	DMEM (Gibco)
NCI-H23	Lung	ATCC	RPMI 1640 (Gibco)
NCI-H292	Lung	ATCC	RPMI 1640 (Gibco)
NCI-H358	Lung	ATCC	RPMI 1640 (Gibco)
NCI-H460	Lung	ATCC	RPMI 1640 (Gibco)
NCI-H727	Lung	ATCC	RPMI 1640 (Gibco)
NCI-H1299	Lung	ATCC	RPMI 1640 (Gibco)
NCI-H1650	Lung	ATCC	RPMI 1640 (Gibco)
NCI-H1793	Lung	ATCC	DMEM/Ham's F12 (Gibco)
NCI-H1838	Lung	ATCC	RPMI 1640 (Gibco)
NCI-H1975	Lung	ATCC	RPMI 1640 (Gibco)
NCI-H2228	Lung	ATCC	RPMI 1640 (Gibco)
NCI-H2342	Lung	ATCC	DMEM/Ham's F12 (Gibco)
U2OS	Osteosarcoma	ATCC	Mc Coy's 5A (Gibco)

Table 2. 2. List of cell lines

2.2.1.2. Derivative cell lines

The generation of the ERCC1-deficient A549 cell lines using zinc finger nuclease gene-targeting has been perforemed at Institut Gustave Roussy (U981) and described previously, along with methods for re-expressing different ERCC1 isoforms (Friboulet *et al.*, 2013a). ERCC1 gene in A549 lung cancer cell lines was knocked-out with Zinc Finger Nucleases (ZFN®) technology (Sigma-Aldrich) to establish cell lines without basal expression of ERCC1 isoforms.1 To create ERCC1-deficient

A549 cells deficient, A549 NSCLC cells were transfected with plasmids encoding Zinc Finger Nucleases (ZFNs) targeting *ERCC1* with Lipofectamine®. To increase ZFN efficiency, cold-shock treatments was used by setting the incubator at 30°C and 5% CO2. Surveyor nuclease (Cel-1) assay (706025; Transgenomic® Ltd), a nucleotide mismatch assay, was used to quantify the effectiveness of ZFN on extracted genomic DNA from transfected cells. Isolated clones were screened by western blot for ERCC1 protein detection. All four *ERCC1* alleles (determined by FISH, data not shown) from these clones were sequenced after TOPO®-cloning to ensure the presence of a ZFN double strand break-induced coding mutation in all genomic copies of *ERCC1* present in A549 cells (Figure S5B). To assess the role of each *ERCC1* isoform, stable A549 cell lines re-expressing a single isoform were generated. *ERCC1* isoform cDNA purchased from Genscript® (Piscataway) were cloned into lentiviral vectors using ViraPowerTM HiPerformTM Lentiviral FastTiterTM TOPO® Expression Kit (k532000; Gibco-Invitrogen) and transduced in A549 ERCC1-deficient clones (Ac375), previous to cell sorting by positive GFP selection. An empty lentivirus (Control vector) was used as control.

2.2.1.3. Media

Dulbecco's Modified Eagle Medium (DMEM), Eagle's Minimum Essential Medium (EMEM), RPMI 1640 medium, DMEM/Ham's F12 medium and McCoy's 5A medium were all supplied by Gibco and supplemented with 10% Fetal Bovine Serum (FBS) (Gibco). Antibiotics were not routinely used in cell culture.

2.2.1.4. General culture conditions

Cell lines were grown in 5% CO_2 at 37°C in their respective media. When passaging cells, growth medium was removed, cells washed in PBS (Sigma) and the cells incubated with a covering volume of trypsin/EDTA (Sigma). After cells had detached the trypsin was neutralized by adding media to the cells and the cells seeded into new flasks or plates. To count cell numbers, re-suspended cells were run through a pipette multiple times to ensure a single cell suspension before staining a 10 μl aliquot with trypan blue and counting using the Countess automated cell counter (Invitrogen) which also estimates viable cell numbers. Frozen cell stocks were maintained in liquid nitrogen in a solution of 90% FBS with 10% DMSO. Cell line stocks were tested monthly for mycoplasma infection (as per laboratory protocol) and identities were confirmed by STR typing.

2.2.1.5. siRNA transfection

For reverse transfections, a transfection mix of transfection reagent (lipofectamine 2000, RNAimax, or Dharmafect 3), and serum free media (Opti-MEM®, Gibco) was prepared and incubated for 5 minutes according to the manufacturer's instructions. The siRNA was added the transfection mix and incubated for 30 minutes as per the suggested protocol. Cell suspensions were prepared by trypsinising, counting and diluting cells in antibiotic-free media according to cell numbers required for the experiment. These suspensions were added to the transfection reagents, mixed and plated into the required plate.

If siRNA had to be followed by drug exposure (combination experiment assessing the effect of the siRNA on drug sensitivity), drug was added 48h after transfection and transfection efficacy concomitantly checked by western blot.

2.2.1.6. Individual siRNA silencing revalidation

Validation of RNAi gene silencing was determined by western blotting and by viability assays of silencing effects with individual siRNA oigonucleotides. Cells were reverse transfected with individual siRNA oligos (Dharmacon). After 24 hours half the cells were plated into 96 well or 6 well plates for dose response assays as already described and the rest of the cells were collected 48 hours following transfection for preparation of protein lysates and western blot analysis.

2.1.1.7. Clonogenic / colony formation survival assay (CFA)

Cells were plated in six well plates at a density of 500-1000 cells per well in 2ml of media and left overnight to settle. The following day, cells were treated with serial dilutions of the investigative drug in media. Solutions of vehicle (DMSO) in media were added to control wells. Media and drug or vehicle was replenished three times a week with freshly made dilutions. The assays were performed over 10-14 days at which point were fixed with 10% cold trichloroacetic acid (TCA) and stained with 0.057% sulforhodamine B before rinsing four times in 1% (v/v) acetic acid and leaving to dry. Cell viability was assessed either by colony counting either manually or using a colony counting machine (ColCount®, Oxford Optronix). Surviving fractions were calculated as SF = mean in treated sample/mean of untreated samples (vehicle). Values were plotted using GraphPad Prism to generate dose response curves for each cell line. SF_{50} values were calculated using Graphpad Prism software.

2.2.1.8. Short-term assays (96-well plate)

Cells were trypsinised and counted and plated in 96-well (250-1000 cells/well) plates in 80 μl. After 24 hours drug or vehicle (DMSO) dilutions in media were added to the cells to make a total volume of 100 μl. Cells were left in constant exposure to drug for five to six days depending on cell confluence as judged by light microscopy. Cell viability was assessed using the luminescent CellTiter-Glo® reagent (Promega) whereby 50 μl of reagent diluted 1 in 4 in PBS was added to each well, which was shaken for 10 minutes according to the manufacturer's protocol. After a 10-minute incubation at room temperature luminescence was measured using the Victor X5 Multilabel plate reader (Perkin Elmer). Cell viability (surviving fraction) was calculated as a fraction of luminescence in vehicle treated wells and dose response curves plotted in GraphPad Prism.

2.2.1.9. High throughput drug screen

An in-house drug library encompassing 80 drugs either used in clinical practice or in late-stage development was used. Each compound was dissolved in 100% DMSO to give 5mM stocks and then diluted to 0.5, 0.05, 0.005, 0.0005mM stocks in 96-well 2D-matrix plates. Daughter plates in 384-well format were prepared from these 96-well 2D-matrix racks using the Hamilton Microlab Star robotic platform. Compounds were stored under a nitrogen atmosphere using a StoragePod® (Roylan Developments).

Cell lines were seeded (500 cells/well) into 384-well plates using a MultiDrop Combi Dispenser (Thermo) and incubated overnight at 37 ^{o}C, 5% CO_2. Replicate cell plates were then loaded onto Microlab Star screening platform and drug plates were serially diluted in appropriate media before being added to the cell plates. The final drug concentrations used for each drug were 1000, 100, 10 and 1 nM. The final DMSO concentration in all wells was 0.2% (v/v). Controls included 0.2% (v/v) DMSO and 10μM staurosporine (Sigma Aldrich) or 20 μM CPT11 (Sigma Aldrich). After incubation in drug-containing media for four days, cell viability was quantified with CellTiter-Glo (Promega) using a Victor X5 Multilabel plate reader luminescence protocol (Perkin Elmer). Luminescence data from each well was normalised to the median signal from DMSO-containing wells to calculate the survival fraction.

2.2.1.10. High throughput siRNA screen

All siRNA libraries were purchased from Dharmacon. siRNA were first reconstituted in RNAse-free DEPC-treated water (Ambion) in 96-well mother matrix plates, diluted to reach the final concentration of 20μM and aliquoted in 384-well plates of the daughter plates. Each well contained a SMARTpool® of four distinct siRNA species targeting different sequences of the target transcript. Each plate was supplemented with negative siCONTROL (12 wells; Dharmacon) and positive control (4 wells, siPLK1, Dharmacon). Cell lines were reverse transfected (according to optimal transfection conditions) into 384-well plates using a MultiDrop Combi Dispenser (Thermo) and incubated at 37 ^{o}C, 5% CO_2. After incubation in siRNA-containing media for seven days, cell viability was quantified with CellTiter-Glo (Promega) using a Victor X5 Multilabel plate reader luminescence protocol (Perkin Elmer). Luminescence data from each well was normalised to the median signal from DMSO-containing wells to calculate the survival fraction. For siRNA screens combined with olaparib, drug or vehicle (DMSO) was added 48h after transfection at 1μM concentration in media and cells were exposed to olaparib for five days.

2.2.1.11. High throughput screen siRNA revalidation

Revalidation of the siRNA screen hits was performed by deconvolution of the SMARTpool® into four distinct oligonucleotide species targeting four different sequences of the gene. Reverse transfection and viability analysis was performed as described above.

2.2.2. Confocal microscopy

Cells were grown on coverslips and exposed to olaparib (10 μM) or treated with increasing UVB doses (Stratalinker1800®). 24 hours after drug treatment or exposure to UVBs, cells were fixed with 4% paraformaldehyde (PFA) for 30 to 60 minutes, after which time coverslips were kept at 4ºC in PBS until stained. Cells were permeabilised in 0.2% (v/v) Triton X-100 in PBS for 10 minutes, washed 3x in PBS, and blocked in IFF for 30 minutes. The cells were then incubated for an hour at room temperature with the primary antibody diluted in IFF (anti-γH2AX, anti-RAD51 or anti-CPD). After 3 x 5 minute washes in PBS, the cells were incubated with a secondary antibody conjugated with ALEXA 488 nm (Molecular Probes) or ALEXA 555 nm diluted 1:1000 in IFF for 60 minutes at room temperature. The cells were subsequently washed 3 x 5 minutes with DAPI nuclear stain diluted 1:10,000 in PBS to visualise cell nuclei. Coverslips were rinsed in H_2O, drained and placed cell-side down on 4 µl Vectashield mounting medium on a microscope slide. Coverslips were secured to the slide by sealing with nail varnish at the edge and viewed on a LEICA TCS-SP2 confocal microscope. To ensure comparability between the images, acquisition settings were set up on the CPD / γH2AX / RAD51 staining producing the highest fluorescence intensity, and were maintained identical throughout the acquisition of all pictures of a given experiment.

2.2.3. Flow cytometry and analysis of cell cycle distribution by FACS

For cell cycle analysis by DNA content, cells were plated in 10cm dishes and exposed to olaparib 10µM 24h after plating. After one day of drug exposure, cells were harvested, washed with PBS, and fixed in 70% ethanol. Cells were then washed twice in PBS and resuspended in 1 ml of PBS containing 100 µl RNase A (Sigma) and 20 µl propidium iodide (PI) (Sigma). Total DNA content was quantified and analyzed by flow cytometry on a Becton Dickinson fluorescence-activated cell scan cytometer and data was analysed using CellQuest pro® (Becton Dickinson, USA).

2.2.4. Protein manipulation

2.2.4.1. Western blotting

Whole-cell protein extracts were prepared from cells lysed in NP250 buffer (20mM Tris pH .7.6, 1mM EDTA, 0.5% NP40, 250mM NaCl) supplemented with protease inhibitor cocktail tablets (Roche, UK). Protein concentrations were measured using the Biorad Protein Assay Reagent (BioRad, Hemel Hempstead, UK). For Western blot analysis, 20-50 µg of whole cell lysates were electrophoresed on Novex 4–12% gradient *bis–tris* pre-cast gels (Invitrogen) with full range rainbow molecular weight marker (GE Healthcare, UK) as a size reference and proteins were transferred to nitrocellulose membrane (Bio-Rad, USA), blocked and probed with primary antibody diluted 1 in 1000 in 5% milk overnight at 4°C. Secondary antibodies were diluted 1 in

5000 in 5% milk and incubated for one hour at room temperature. Protein bands were visualised using ECL (GE Healthcare, UK) and Kodak BioMAX XAR film (Kodak).

2.2.4.2. Stable isotope labeling by amino acids in cell culture (SILAC)

Specific Material

- SDT-lysis buffer: 2%(w/v) SDS, 100mM Tris/HCl pH 7.6, 0.1M DTT
- UA: 8 M urea (Sigma, U5128) in 0.1 M Tris/HCl pH 8.5 (60 ml per sample)
- IAA solution: 0.05 M iodoacetamide in UA (2ml per sample)
- Trypsin Gold (Promega): Stock 0.4 µg/µl
- ABC: 0.04M NH4HCO3 in water (30ml per sample).
- Amicon Ultra 30K 30,000MWCO (Millipore)
- Sep Pak C18 Plus Light Cartridge (Waters)
- Labelled aminoacids: $^{13}C_6$ $^{15}N_2$ L-lysine and $^{13}C_6$ $^{15}N_4$ L-arginine (Pierce)
- Unlabelled aminoacids: L-lysine and L-arginine (Sigma)
- SILAC DMEM (Pierce)

Protocol

Cells were grown in SILAC DMEM supplemented with either labelled or unlabelled amino acids for eight doubling times. Cell were lysed by adding 1-2 ml of lysis buffer per 10cm dish of confluent cells and boiled for 10 mins. Protein concentration was measured using a BSA scale. Equal concentration of heavy- and light-labelled cells were mixed (~1ml of lysate, equivalent ~1-15 mg total protein) with 8 mL of UA buffer and load into an Amicon Ultra 15 Ultracel 30k device, with a 1:8 lysis buffer to UA ratio. After concentrating at 4000*g* for 15-30 mins (at least 10 fold concentration), 10 mL of UA was added to the device, which was concentrated again again (4000*g* for 15-30 mins). 2mL of IAA was added and samples were incubated in darkness at RT for 30 min, prior to 15 min centrifugation (4000*g*). 10 mL of UA solution were added and the sample was concentrated again (this step was repeated three times in total) prior to washing twice 10 mL of ABC. The filter was put onto a new tube and trypsin was added (1:100 of the lysate amount) in 1 mL of ABC. Digestion was conducted overnight at 37°C in a wet chamber. Peptides were collected by centrifugation (15 min), prior to adding 0.5 mL of ABC and spinning for 15 min, followed by adding 0.5ml of 20% Acetonitrile in ABC. Extracts were acidified and desalted on a C_{18} cartridge (Sep-Pak), prior to off-gel isoelectric focusing using a 3100 Off-Gel Fractionator® (Agilent Technologies).

Mass spectrometry

Reversed phase chromatography was performed using an HP1200 platform (Agilent, Wokingham, UK). Twenty percent of each sample was analysed as a 6 µl injection, including a known spiked concentration of a three- peptide mixture QC sample spike. Peptides were resolved on a 75 µm I.D. 15 cm C18 packed emitter column (3 µm particle size; NIKKYO TECHNOS CO., LTD, Tokyo, Japan) over 120 mins (SET 1 (runs 1 and 2) only) or 90 mins using a non-linear gradient of 96:4 to 50:50 buffer A:B (buffer A: 2% acetonitrile / 0.1% formic acid; buffer B: 80% acetonitrile / 0.1% formic acid) at 250nl/min. Peptides were ionised by electrospray ionisation using 1.8kV applied immediately pre-column via a microtee built into the nanospray source. Sample was infused into an LTQ Velos Orbitrap mass spectrometer® (Thermo Fisher Scientific, Hemel Hempstead, UK) directly from the end of the tapered tip silica column (6-8 µm exit bore). The ion transfer tube was heated to 200°C and the S-lens set to 60%. MS/MS were acquired using data dependent acquisition based on a full 30,000 resolution FT-MS scan with preview mode disabled and internal lock mass calibration against the polysiloxane ion at 445.120025 *m/z*. The top 20 most intense ions were fragmented by collision-induced dissociation and analysed using normal ion trap scans. For sample set 2, the top 10 ions were also fragmented and analysed using enhanced ion trap scans. Automatic gain control was set to 1,000,000 for FT-MS and 30,000 for IT-MS/MS, full FT-MS maximum inject time was 500 ms and normalised collision energy was set to 35% with an activation time of 10 ms. Wideband activation was used to co-fragment precursor ions undergoing neutral loss of up to -20 m/z from the parent ion, including loss of water/ammonia. MS/MS was acquired for selected precursor ions with a single repeat count followed by dynamic exclusion with a 10ppm mass window for 60s (120 min gradient) (SET 1 only) or 45s (90 min gradient) based on a maximal exclusion list of 500 entries.

2.2.5. DNA manipulation

Genomic DNA was isolated from the cell lines using the Puregene - blood, cell and tissue kit® (Qiagen) according to manufacturer's instructions, eluted in 50 µl H_2O and stored at -20°C. DNA concentration was measured using either a spectrophotometer measuring the UV absorbance at 260nm (samples for exome sequencing) or using a Qubit® 2.0 fluorometer (samples for 8-oxoguanine dosage).

For exome sequencing, three micrograms of genomic DNA was fragmented to 200 bp using a Covaris E Series instrument (Covaris Inc.) and the resultant library subjected to DNA capture using either the 38 Mb SureSelect Human All Exon kit® (Agilent). Illumina paired-end libraries were prepared from the captured target regions and quantified using a Bioanalyzer DNA chip® (Agilent), followed by sequencing on a HiSeq2000® platform (Illumina), acquiring 2 x 76 bp reads. Casava software (v1.8, Illumina) was used to make base calls. Sequences were output in fastq format. Reads failing the Illumina chastity filter were removed before further analysis.

2.2.6. RNA manipulation

RNA was extracted from cell lines using the RNAeasy kit® (Qiagen) according to the manufacturer's instructions, eluted in 50μl H_2O and stored at -80°C. RNA concentration and quality was measured using a 2100 Bioanalyzed® platform (Agilent® technologies) and a spectrophotometer. Whole transcriptomic analysis was performed on an Illumina BeadArray® HumanHT-12 v4 device at the Wellcome Trust Centre for Human Genetics, University of Oxford.

2.2.7. Metaphase spreads preparation

Cells were plated (400000 cells / T75) and treated with olaparib 10μM or DMSO. After 24h, 10μM BrdU was added for 2 replicating cycles prior to collecting metaphases by adding 0.2μg/ml of media of colcemid for 2 hours. Flasks were banged strongly to detach metaphasic cells, which were subsequently collected and washed three times with PBS. 4ml of 75mM KCL was subsequently added for 8 minutes, prior to adding 4ml of fixative (3:1 methanol : glacial acetic acid) to the KCL solution, in order to fix metaphases. Cells were pelleted and washed once more with 4ml of fixative solution, prior to final resuspension of the pellet in 60μl of fixative. Metaphases were stored at -20° prior to spreading on a slide (30μl of cells added drop by drop on a wet slide). Sides were dried overnight before DNA staining with DAPI 1:10000, and mounting with 4 μl Vectashield medium on a microscope slide. Coverslide was secured to the slide by sealing with nail varnish at the edge and viewed on a Zeiss confocal microscope.

2.2.8. UVB irradiation

Cells were irradiated with UVBs in 96-, 24- or 6-well plates at increasing doses on a Stratalinker 1800 according to the manufacturer's instructions.

2.2.9. NAD/NADH quantification

Exponentially growing cells were trypsinised and NAD/NADH was subsequently quantified using the NAD/NADH Assay Kit® (ab65348, Abcam) according to the manufacturer's instructions.

2.2.10. 8-oxoguanine quantification

Genomic DNA was extracted as previously described, quantified using a Qubit® 2.0 fluorometer, and 8-oxoguanine was subsequently quantified using the OxiSelect™ Oxidative DNA Damage ELISA Kit® (Cell Biolabs) according to the manufacturer's instructions.

2.2.11. Pyrimidine dimers quantification

Pyrimidine dimers quantification was performed at different time points after UVB irradiation using a Stratalinker 1800® at a range of increasing doses, using the CycLex Cellular UV DNA-Damage Detection Kit® MBL international.

2.3. Statistical analyses

2.3.1. General statistical analysis

Statistical tests were performed using Microsoft Excel or GraphPad Prism. All tests were two-sided unless otherwise stated. Mann-Witney tests were used to compare non-parametric datasets and Student's t-tests used for parametric datasets. Statistical significance was set at p-value < 0.05.

2.3.2. Dose-response curves

Drug treated wells were compared to vehicle treated wells to generate surviving fractions where:

"Surviving fraction = (luminescence in drug treated well)/(luminescence in vehicle treated wells)"

Surviving fractions were expressed as a percentage and plotted in GraphPad Prism where dose response curves were drawn using a five-parameter logistic equation with or without preliminary Log-transformation of the data. SF_{50} or SF_{80} values were calculated from these curves using GraphPad Prism software. Comparisons of dose response curves were performed using two-way ANOVA testing.

2.3.3. Viability assessment post-siRNA transfection

Wells treated with the siRNA of interest were compared to Mock (no siRNA) and negative control (AllStar®, Qiagen) siRNA. Surviving fraction was calculated following:

"Surviving fraction = (luminescence in siRNA of interest treated well) / (luminescence in siCON treated wells)"

2.3.4. High-throughput screens

2.3.4.1. High throughput drug screen analysis

Raw luminescence values were initially analysed using the CellHTS2 package of the R v3.0.1 software (www.r-project.org). This outputs a range of quality control statistics such as plate-to-plate raw luminescence values and post-normalisation plots. The raw luminescent cell viability values for each well in the screen were normalized to the median viability of the negative controls (DMSO) of each plate, in order to calculate a survival fraction (SF). The median survival fraction of triplicates at each drug concentration was then used for the final analysis.

2.3.4.2. High-throughput siRNA screen analysis

Raw luminescence values were initially analysed using the CellHTS2 package of the R v3.0.1 software (www.r-project.org). This generated a number of quality control metrics including the Z′ factor which was calculated using the positive (siPLK1) and negative (siCON1 and Allstar) controls to assess the discriminatory power of the screen. The Z′ factor was calculated using: $Z' = 1- [(3X(A+B))/(C-D)]$, where A equals the mean of luminescence measurements from the positive controls (siPLK1), B the mean of luminescence measurements from the negative controls (siCON and Allstar) and C and D, standard deviations of luminescence measurements from positive and negative controls, respectively. The Z' factor is a measure of the ability of the screen to distinguish between positive and negative controls. For population screens a Z' factor of > 0.5 is considered excellent and > 0 borderline. For cell based screens such as this, this level of reproducibility is considered excessive and a Z' factor of >0 was considered adequate. Pearson correlation coefficients (r) between plates were squared to calculate r^2 values. The raw luminescence values were $\log_2$ transformed and then centred by the plate median.

Each screen was performed in triplicate with results of all three screens being used in the final analysis. The effect of individual siRNAs on cell viability was assessed by calculating the Z-score for each oligo, where the Z-score represents the number of standard deviations that separates the effect of this particular siRNA from the mean of the dataset. For example, for siRNA "X":

"Z-score of siRNA X = (cell viability effect of siRNA X – median cell viability effect of all SMARTpool® siRNAs of the library) / standard deviation of all SMARTpool® siRNAs of the library"

The assumption that the vast majority of siRNAs has no effect on cell viability was made to ensure the relevance of this calculation. Consequently, a Z-score of 0 would represent no effect on cell viability, whereas a negative Z-score of < -2 would reflect a significant loss of cell viability. Raw luminescence values from each plate were first log2-transformed and normalised to the median luminescence score for all experimental wells on one plate, in order to take into account plate-to-plate variation in transfection and when acquiring luminescence. The three replicates were combined in the final analysis, where the median Z-score for each individual siRNA was calculated using the median normalized value of each siRNA and the median absolute deviation (MAD), in order to account for the variability within the screen.

2.3.4.3. High-throughput siRNA screen with olaparib

The effect of each siRNA on PARP inhibitor sensitivity was calculated by calculating a drug effect (DE) score for each siRNA. Drug effect (DE) scores were calculated as the difference between the median of replicate wells with drug and median of replicate corresponding wells with no drug for each siRNA. Calculation of the Median Absolute Deviation (MAD) was used to estimate the variance of the DE data and DE Z-scores calculated as:

"DE Z-score of siRNA = (drug effect of each siRNA – median drug effect of all siRNA) / standard deviation of DE values for all siRNA SMARTpools®"

Negative DE Z-scores suggest the siRNA increases sensitivity to the drug and positive DE Z-scores suggest the siRNA causes resistance. DE Z-scores of < -2 and > 2 were selected as significant sensitising and resistance causing effects.

2.3.4.4. High-throughput revalidation siRNA screen

As silencing of PLK1 produced different loss of viability among cell lines, raw luminescence values were analysed using the NPI formula (Normalised Percentage Inhibition), where
"NPI = 1-((Raw sample - Raw Negative Mean) / (Raw Positive Mean - Raw Negative Mean))."

2.3.5. Exome sequencing

Exome sequencing analysis was performed according to the following steps by Dr James Campbell, ICR.

<u>Step 1. FASTQ file creation, demultiplexing and filtering failed reads</u>
The configure_bcl2fastq.pl script (part of CASAVA 1.8, Illumina) was used to configure the demultiplexing of samples and creation of fastq files. The resulting fastq files contained reads that fulfilled and failed the Illumina chastity filter, these latter being subsequently removed.

<u>Step 2. Alignment to reference sequence</u>
BWA (Burrows-Wheeler Aligner, bio-bwa.sourceforge.net) was used to align short reads to a reference sequence (GRCh37). For paired end reads, this involved alignment of reads 1 and 2 separately and then conversion of the resulting sai files into sam format. The sam file was then compressed to bam format, sorted and indexed using samtools. The bam files were next processed using Picard tools to add read group tags indicating the library ID and to remove duplicates. Quality Control information based on the alignment was generated using bedtools to determine coverage of the targeted regions and samtools to extract statistics describing the proportion of reads properly aligned. Bedtools was also used to remove reads not mapping to the target regions.

<u>Step 3. Realignment, recalibration and variant calling</u>

The bam file containing aligned reads from the “tumour” (ERCC1-deficient clones or non-isogenic NSCLC cell lines) and “normal” (ERCC1-proficient A549 cell line or reference genome, respectively) matched samples were next processed using the Genome Analysis Tool Kit (GATK) (McKenna *et al.*, 2010). The analysis was not restricted to simple pairs of samples but can practically be performed on larger sets of matched samples. Contrastive analyses should only be performed on samples that have been processed as a matched set using the following GATK Broad Best pipeline V2 with standard settings (www.broadinstitute.org/gatk/guide/best-practices).

Step 4. Merging and annotation of variant calls

The aligned reads (in bam format) for each chromosome were merged using samtools merge, sort and index. The variant calls (as vcf files) for each chromosome were merged by concatenating the vcf files together. Variant calls were filtered using the following heuristic criteria: (i) variants called in regions not covered by the exome capture probes were excluded; (ii) variants marked as low quality (QUAL below 20) were excluded; (iii) variants with fewer than 10 reads covering the locus in all samples were excluded. Common SNPs (those reported to have a global minor allele frequency of greater than 5% in any of the 1000 genomes project data sets) were also removed from the main analysis. The remaining variants were annotated using the Ensembl variant effect predictor script (Ensembl v61).

2.3.6. Transcriptomic analysis

Transcriptomic analysis was performed by Dr Céline Lefebvre (IGR). Illumina raw data were acquired and normalized with the lumi package (Du *et al.*, 2008) in R 3.0.1. Specifically, raw data was background corrected and normalized with variance stabilization and robust spline normalisation (vst and rsn methods). Quality of the data was checked before and after normalisation by checking distributions, sample clustering with hierarchical clustering and principal component analysis. Normalized expression profiles contain 47,231 probes for 27 samples. The expression profiles of probes mapping to validated genes contained 24,845 probes and 27 samples.

Differential expression was computed with the package limma in R 3.0.1. This package uses linear models to assess differential expression in any context including multifactor designed experiments. Multiple hypothesis testing corrections were done using either False Discovery Rate (FDR with Benjamini Hochberg’s method (Benjamini *et al.*, 1995) or Bonferroni correction.

2.3.7. SILAC analysis

SILAC samples were run on Orbitrap Velos® (Thermo Scientific) by Dr Andrew Thompson at the ICR proteomics core facility. Data analysis was performed by Dr Faraz Mardakheh using MaxQuant according to Cox et al (Cox *et al.*, 2009). Briefly, Raw MS/MS data we're searched and quantified on MaxQuant (www.maxquant.org) against Human IPI database (v3.68). A False Detection Rate of 1% was used for peptide and protein identifications, and hits were filtered for having at least two ratio counts. All statistical analyses were performed by Perseus software from the MaxQuant package.

CHAPTER 3

An integrated approach for the identification of synthetic lethal interactions with ERCC1-deficiency

3.1. Introduction

Approximately 30% of NSCLC patients will respond to platinum therapy, but the administration of this cytotoxic agent is unfortunately limited in duration and has to be halted after a maximum of 6 cycles for toxicity reasons, even if some degree of tumour shrinkage continues to be observed (Soria *et al.*, 2013). In the advanced and metastatic setting, all these patients will ultimately relapse after a few months and usually die within one year after diagnosis (Jemal *et al.*, 2011). Therefore, there is an urgent need to find alternative non-toxic therapies that would selectively target genetic deficiencies underlying platinum-sensitivity (Postel-Vinay *et al.*, 2012). ERCC1 is the most promising biomarker of cisplatin response in NSCLC, and a significant proportion of tumours display low levels of ERCC1 (Besse *et al.*, 2013).

Several experimental approaches may be utilised to investigate synthetic lethal interactions and genetic dependencies. Beyond candidate-based approaches, where the role of specifically chosen proteins is studied, high throughput approaches offer the advantage of rapidly examining the effect of many drugs or the role of multiple genes in parallel (Lord *et al.*, 2009; Brough *et al.*, 2011). In this thesis, we have exploited the potential of several parallel high-throughput approaches - namely drug screens and siRNA screens, the results of which will be presented in this Chapter.

Beyond the choice of an appropriate approach, the choice of a relevant model is crucial for studying meaningfully the deficiency in a given gene and identifying synthetic lethal interactions. Several cell-lines based models can be used, which are most frequently divided into “isogenic” or “non isogenic” systems. Isogenic cell lines all derive from a unique progenitor, in which genetic manipulation (e.g. by gene targeting or RNA interference) has been used to disable the gene of interest in the resulting daughter cell lines (Rehman *et al.*, 2010). Therefore, isogenic systems provide a simplified system for identifying the consequences of a single genetic modification; as the number of differences between the parental and daughter cells is minimized, the causality of the observed changes can largely be attributed to the alteration of the gene of interest. Non-isogenic models constitute a large panel of cell lines which arise from different progenitors, and as such harbour several differences in genetic background, but can be categorized according to the alteration of interest (e.g. ERCC1-high expression *vs* ERCC1-low expression, *KRAS*-wild type *vs* *KRAS*-mutant). Considering the diversity of the genetic background of non-isogenic models systems, the causality of an effect is usually more difficult to assess in such systems, and large panels of cell lines are required to limit confounding factors. However, non-isogenic models better represent tumour heterogeneity, and are as such a better reflection of the clinical reality. Ideally, the use of both models should be considered, for example combining a “discovery and mechanistic” step in an pure isogenic model, and a “validation step” to demonstrate the potential clinical relevance in non-isogenic models.

Although ERCC1 was the first mammalian repair gene to be cloned (Westerveld *et al.*, 1984) and targeted in mice (McWhir *et al.*, 1993), only three cases of patients presenting with ERCC1-deficiency have been reported to date. Strikingly, these patients all displayed a considerably more severe phenotype than XP-deficient patients, notably suffering from pre-and postnatal developmental failure followed by death in early infancy (Jaspers *et al.*, 2007; Kashiyama *et al.,* 2013). Complete ERCC1-deficiency is reported to be lethal for human cells, and the only available isogenic models of ERCC1-deficiency at the beginning of the PhD were the Chinese Hamster Ovary (CHO) derived AA8 (ERCC1-wild type) and UV20 (ERCC1-deficient) cell lines. Given the important differences between hamster and human cells and the limited translatability of the findings from CHO to human NSCLC cells, considerable efforts were made at the Institut Gustave Roussy to generate an NSCLC isogenic model of ERCC1-deficiency (showing undetectable ERCC1 protein expression by western blot) (Friboulet *et al.*, 2013a). As described in this Chapter, this latter model provided the ideal tool to investigate ERCC1-deficiency in NSCLC and has been the cornerstone of the main findings of this PhD.

3.2. Results

3.2.1. Choice of model of ERCC1-deficiency

3.2.1.1. Isogenic NSCLC model of ERCC1-deficiency

When generating isogenic models, cell line selection is critical both in terms of phenotypic characteristics and relevance of the model. The A549 cell line expresses relatively high levels of ERCC1 and is one of the most commonly used NSCLC cell lines, as reported in the literature. This cell line is relatively resistant to cisplatin (IC_{50} ≈ 15 μM), highly transfectable, with robust and rapid growth following transfection, making it ideal for genetic manipulation. Therefore, A549 cells were used to generate a series of isogenic model of ERCC1-deficiency. *ERCC1* gene targeting was performed using a Zinc finger nuclease technology (ZFN®, Sigma-Aldrich) (Gaj *et al.*, 2013), which offers the advantage of cutting DNA at highly specific sites and with satisfactory efficacy.

3.2.1.1.1. Generation of ERCC1-deficient isogenic model

The overall procedure for generation of ERCC1-deficient cells is described in Figure 3. 1. and was performed at the Institut Gustave Roussy by Dr Luc Friboulet and colleagues (U981 INSERM). ERCC1 exists as four distinct isoforms that share exons 4, 5, 6 and 7. A short sequence in exon 7 was consequently chosen as target for the Zinc finger nucleases (ZFN), in order to be able to simultaneously inactivate all four isoforms, with the potential for reintroducing each isoform by cDNA expression as a mean to study isoform-specific effects (a complete description of *ERCC1* sequence, sequence targeted by the ZFN together with its corresponding exon location, cDNA and protein sequence of the four isoforms, is detailed in Supplementary Figure 1).

A549 cell lines are aneuploid and *ERCC1* is present as four copies in the parental cell line. The creation of ERCC1-deficient cells therefore required altering each of these copies. After screening by western blot more than 500 colonies at the Institut Gustave Roussy, one ERCC1-heterozygous cell line with three altered copies was identified; further treatment with the Zinc finger nucleases allowed the generation of three ERCC1-deficient clones, harbouring four altered copies of *ERCC1* and undetectable levels of ERCC1 by Western blot (Figure 3. 2).

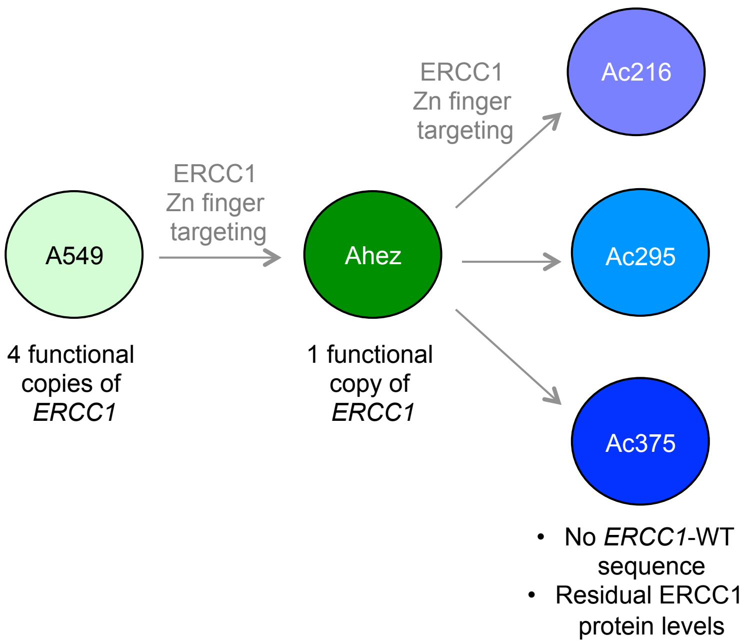


Figure 3. 1. Generation of the ERCC1-deficient clones.

To create ERCC1 deficient A549 cells, this NSCLC model was transfected with plasmids encoding Zinc Finger Nucleases (ZFNs) targeting *ERCC1*. Isolated clones were screened by western blot for ERCC1 protein detection. All four *ERCC1* alleles (determined by FISH, data not shown) from these clones were sequenced after TOPO®-cloning to ensure the presence of a ZFN double strand break-induced coding mutation in all genomic copies of the *ERCC1* gene present in A549 cells. Three ERCC1-deficient clones, harbouring mutations in four *ERCC1* alleles, were eventually derived from a common heterozygous ancestor, after screening of more than 500 colonies.

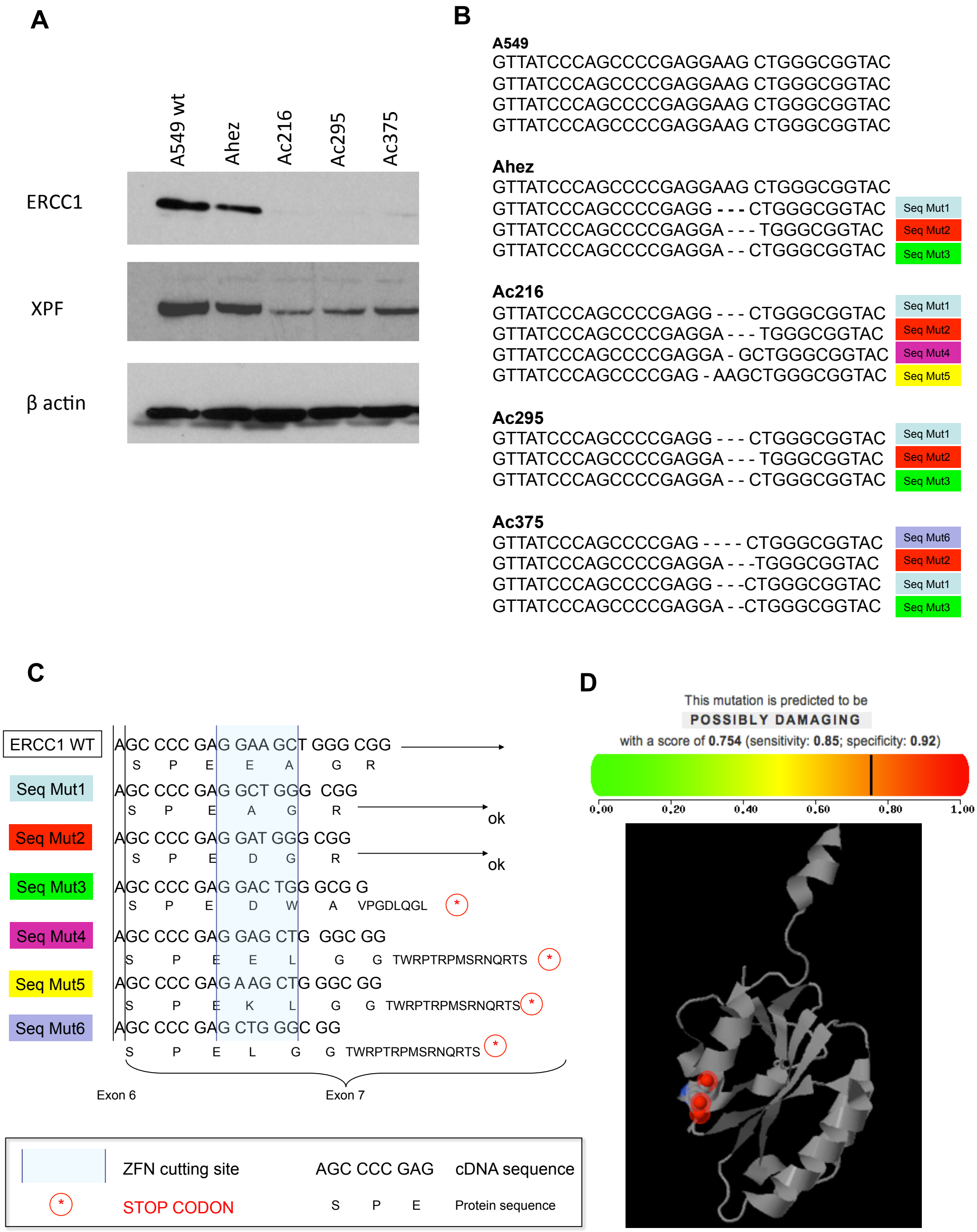


Figure 3. 2. Characterisation of ERCC1-deficient cells (see legend next page)

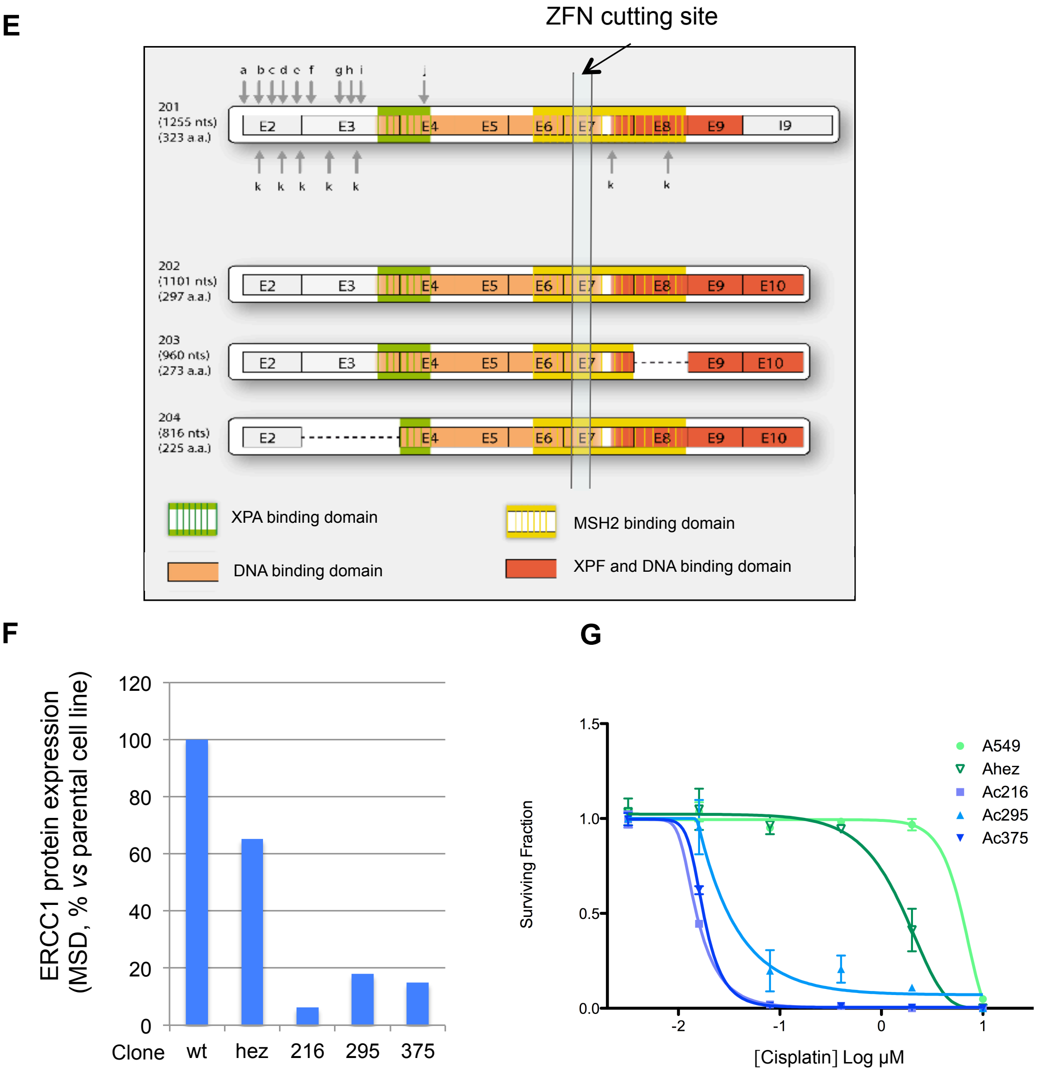


Figure 3.2. Characterisation of ERCC1-deficient cells (continued from previous page)

A) Western Blot detection of ERCC1 and XPF in A549 Wild Type (WT), heterozygous mutant (Ahez) or homozygous mutated cells (Ac216, Ac295, Ac375). B) DNA sequence (obtained by TOPO Cloning®) of the 4 copies of *ERCC1* in *ERCC1* mutant, heterozygous and homozygous clones. The mutation of only 3 sequences indicates that the same deletion was detected in 2 copies of the gene. C) Consequence of the DNA mutation on the protein sequence. Results are presented by mutation type (Mut1 to Mut 6), with colour boxes corresponding to the colour boxes represented on each clone sequence on Figure 3.2.B, to ease visualisation. As depicted, four mutant sequences result in a frameshift mutation with creation of a stop codon, whereas two of them create a deletion +/- a substitution of one amino acid with no frameshift mutation. D) Prediction of a possibly damaging functional consequence of a single nucleotide variation at the localisation of the mutations (amino acid 204 of the protein, within the HhH DNA binding domain) according to Polyphen-2 (genetics.bwh.harvard.edu/pph2) and Pfam (pfam.sanger.ac.uk). E) Schematic representation of each ERCC1 isoform exon composition together with the functional domains composition and indication of the Zn finger nuclease cutting site at the beginning of exon 7. F) MSD quantification of ERCC1 expression, relative to the parental cell line. G) Short-term assay (five days of drug exposure) evaluating cisplatin sensitivity of the *ERCC1* isogenic model.

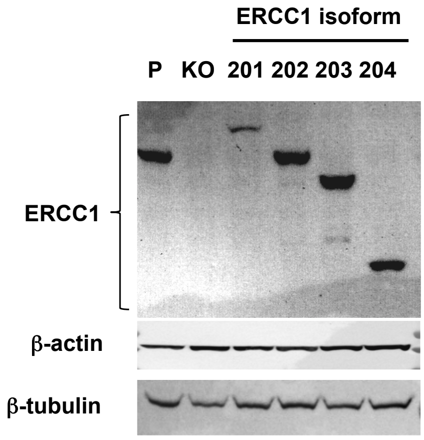
Alterations resulted either in a frameshift mutation (for four out of six mutations), in the loss of one amino acid (one mutation), or in a coding single nucleotide polymorphism associated with the loss of one amino acid (one mutation) (Figure 3. 2. B and C). The study of functional domains by Pfam (pfam.sanger.ac.uk) revealed that this location corresponded to the beginning of the DNA binding Helix-hairpin-Helix (HhH) domain of ERCC1, which has been described as playing a key role in ERCC1/XPF the heterodimer function (Tripsianes *et al.*, 2005; Das *et al.*, 2008; Das *et al.*, 2012). The study by Polyphen-2 of the functional consequences of an E (glutamic acid) to A (alanine) or E to D (aspartic acid) substitution (corresponding to mutant 1 and mutant 2 sequences without any amino acid deletion, respectively) at position 204, revealed a high probability of functionally damaging consequence (score = 0.75; sensitivity = 0.85, specificity = 0.92; Figure 3. 2. D). A more accurate protein dosage analysis by MSD (Meso Scale Discovery®, Meso Scale) revealed that the heterozygous cell line (Ahez) and the three ERCC1-deficient clones (Ac216, Ac295 and Ac375), expressed respectively 65%, 6%, 18% and 15% of the original ERCC1 protein amount (Figure 3. 2. E). Importantly, this biological model of ERCC1-deficiency was also a relevant reflection of the clinical reality, as illustrated by the extreme sensitivity to cisplatin of the ERCC1-deficient clones (Figure 3. 2. F; Supplementary Table 2). Interestingly, XPF – which forms a heterodimer with ERCC1 for the repair of DNA lesions and possesses the endonuclease activity) – was also significantly decreased in the ERCC1-deficient models.

3.2.1.1.2. Investigation of the DNA repair-related role of each ERCC1 isoform

This isogenic model also offered the opportunity to further investigate the role of each ERCC1 isoform. ERCC1-deficient cells were transduced with lentiviral vectors driving the ectopic re-expression of a unique ERCC1 isoform (Figure 3. 3. A). During this PhD, isoform 202 was identified as the only functional isoform with regards to response to cisplatin (Friboulet *et al.*, 2013a), but the role and functionality of each remaining isoform remained largely unknown.

*Note: Although I have contributed to the design of the investigations and writing of the manuscript regarding the role of each ERCC1 isoform, experiments described in this subsection (*Figure 3. 3*,* Figure 3. 4*,* Figure 3. 5*) were performed at the Institut Gustave Roussy within the U981 team.*

A



B

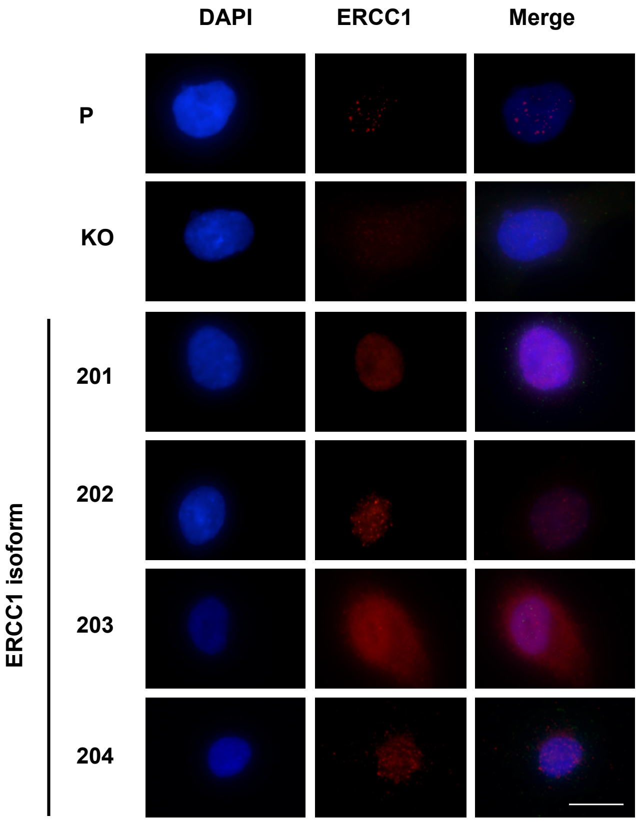


Figure 3. 3. Subcellular localisation of ERCC1 isoforms (Figure and legends reproduced from Friboulet *et al.*, Cell Cycle 2013)

A) ERCC1 expression was assessed by immunoblotting in parental A549 (P), A549 depleted for ERCC1 (KO) and A549 expressing individually each of the four ERCC1 isoform (201, 202, 203 and 204). β-actin and β-tubulin were used as loading controls. B) Immunofluorescence staining of parental A549 (P), A549 knocked-down for ERCC1 (KO) and A549 expressing individually each of the four ERCC1 isoform (201, 202, 203 and 204) for ERCC1 (FL297 antibody, Red). Nuclear DNA was counterstained with DAPI (Blue). Scale bar, 10 µm.

First, using confocal microscopy to study the subcellular localization of each ERCC1 isoform revealed that all isoforms were localized in the nucleus, with the exception of isoform 203 which also generated a significant cytoplasmic signal (Figure 3. 3. B). This suggested a potential non-DNA repair activity of ERCC1 isoform 203, or a potential re-localisation of isoform 203 to the nucleus in response to DNA damage. The best-characterised partner of ERCC1 is XPF, which forms a heterodimer with ERCC1 and possesses the endonuclease activity (Tripsianes et al., 2005). The ability of each isoform to interact with XPF was examined using proximity ligation assays (PLA®, Duolink®; Sigma-Aldrich) technology. This revealed that only ERCC1-isoform 202 was able to physically interact with XPF and produce detectable nuclear foci (Figure 3. 4). Of note, this was also true for two other ERCC1-interactors that contribute to the proper localization and activity of ERCC1/XPF during DNA repair (Li *et al.*, 1994; Matsunaga *et al.*, 1996; Bessho *et al.*, 1997; de Laat *et al.*, 1999), namely RPA and XPA, which could only be detected at similar levels as in the parental cell line after reintroduction of ERCC1-isoform 202 (Figure 3. 4). As previous studies have suggested that XPF is unstable in the absence of interaction with ERCC1 (Arora *et al.*, 2010), XPF, RPA and XPA protein levels were evaluated by western blot after re-expression of each distinct ERCC1-isoform. This analysis suggested a clear dependency of XPF protein levels on the presence of ERCC1 isoform 202 could be observed (Figure 3. 5).

Taken together, these results strongly supported the use of this NSCLC isogenic model of ERCC1-deficiency, which offered the following advantages: (i) being constituted of several clones harbouring different *ERCC1* mutations; (ii) being clinically relevant, as illustrated by the extreme sensitivity of the clones to cisplatin and (iii) facilitating the study the role of each known ERCC1 isoform.

3.2.1.2. Non-isogenic NSCLC model of ERCC1-deficiency

Non-small cell lung cancer cell lines harbour different levels of ERCC1 expression, and can as such be classified as ERCC1-high and ERCC1-low (cell lines expressing high and low levels of ERCC1, respectively). In order to investigate whether such models could be useful for the study of ERCC1-deficiency synthetic lethal relationships, or the validation of hits identified using the isogenic model, we evaluated ERCC1 protein level in a panel of 15 NSCLC cell lines.

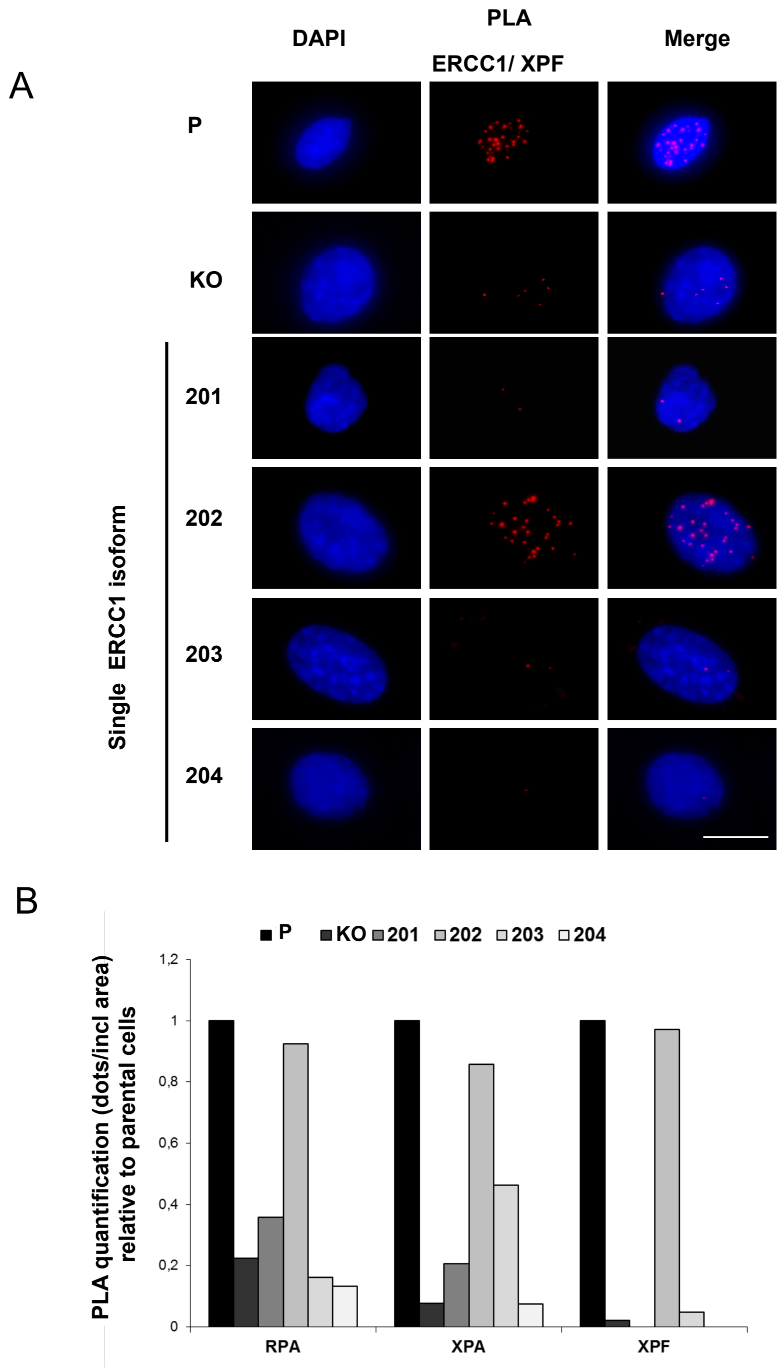


Figure 3. 4. ERCC1 interaction with NER factors (Figure and legends reproduced from Friboulet *et al.,* Cell Cycle 2013)

A) Representative images of proximity ligation assay (PLA®, Duolink®) detection of ERCC1/XPF heterodimers (Red) in parental A549 (P), A549 knocked-down for ERCC1 (KO) and A549 expressing individually each of the four ERCC1 isoform (201, 202, 203 and 204). Nuclear DNA was counterstained with DAPI (Blue). Scale bar, 10 µm. B) Quantification of PLA assay detection of the interaction between ERCC1 and RPA, XPA and XPF performed as in A). The nuclear dots per nucleus were counted using ImageJ software. The ratio of the signal in KO and single isoform-expressing A549 over WT cells was plotted.

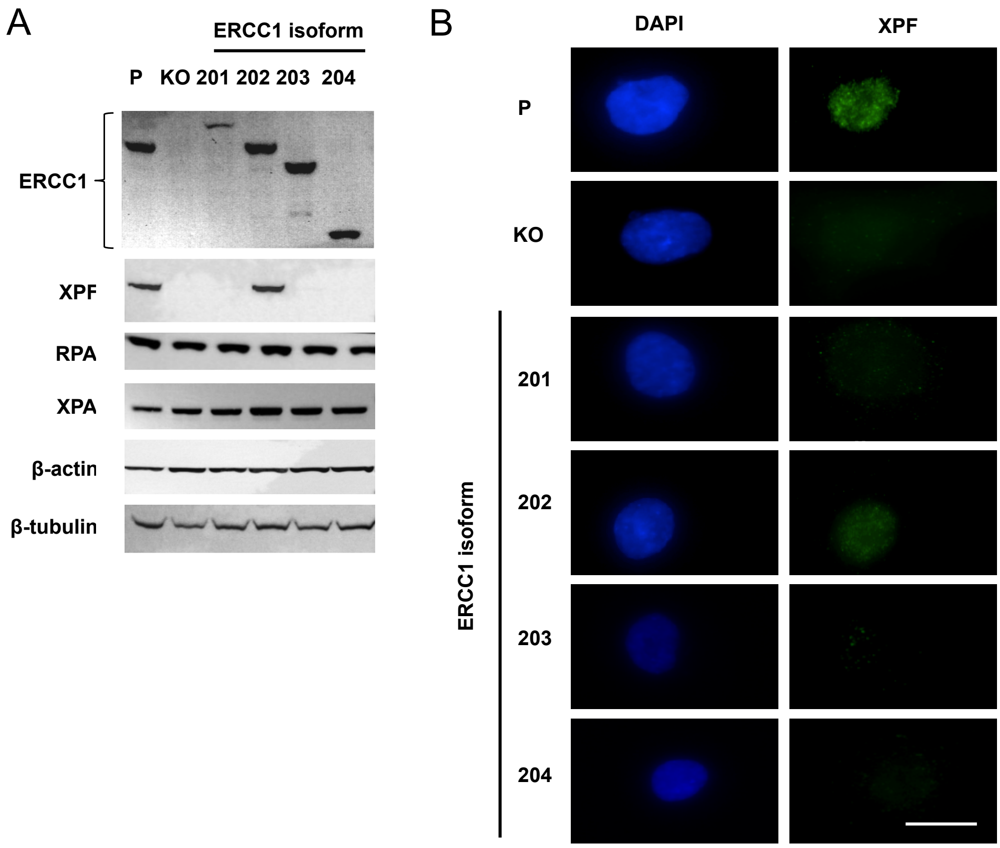


Figure 3. 5. ERCC1 isoform 202 is required for XPF expression (Figure and legends reproduced from Friboulet *et al.,* Cell Cycle2013)

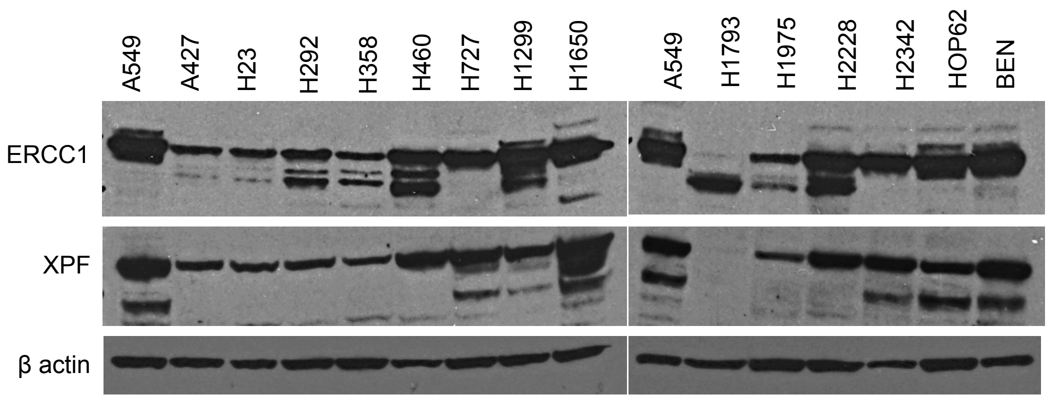
A, Immunoblot analysis of ERCC1, XPF, RPA, and XPA in parental A549 (P), A549 depleted for ERCC1 (KO) and A549 expressing individually each of the four ERCC1 isoform (201, 202, 203 and 204). β-actin and β-tubulin were used as loading controls. B, Immunofluorescence staining of parental A549 (P), A549 depleted for ERCC1 (KO) and A549 expressing individually each of the four ERCC1 isoform (201, 202, 203 and 204) for XPF (ab85140 antibody, red) as on figure 1B. Nuclear DNA was counterstained with DAPI (Blue). Scale bar, 10 μm.

This panel encapsulated models from both male and female patients, smokers and non-smokers, adenocarcinoma and large cell carcinomas. Also, this panel presented a variety of mutational status for the main NSCLC oncogene drivers or tumour suppressor genes (notably for *KRAS*, *EGFR* and *LKB1*), in order to minimise the likelihood of confounding factors in the interpretation of the results (Figure 3. 6). As shown in Figure 3. 6, these cell lines harboured different levels of ERCC1 expression, and could be classified as ERCC1-high (9 cell lines) or ERCC1-low (6 cell lines) following the quantification of the relative ERCC1 expression by western blot using the Image J software (quantification of the number of pixels on the ERCC1 band, normalized over the number of pixels of the β actin band of the corresponding cell line) (Figure 3. 6. B). These cell lines did not only harbour different degrees of ERCC1 isoform 202 expression (the predominant isoform), but also displayed variable levels of other ERCC1 isoforms, as shown in Figure 3. 7, where a longer western blot exposure was required to detect all relevant isoforms.

In order to assess the clinical relevance of this isogenic model, the cisplatin sensitivity of each cell line was assessed and correlated to ERCC1 relative expression. No correlation could be detected between cisplatin sensitivity and ERCC1 status, both when classifying the cell lines in 2 groups (ERCC1-high *vs* ERCC1-low), or when considering ERCC1 as a continuous variable (Figure 3. 8). A potential explanation for this absence of correlation could have been the induction of *ERCC1* expression under platinum therapy. Controversial results have been obtained so far regarding this hypothesis and further investigation is required (Dr Ken Olaussen, personal communication).

As a result, considering the absence of correlation between ERCC1-isoform 202 expression and cisplatin sensitivity - as assessed by cisplatin IC_{50} – it was deemed that the use of the non-isogenic model had little clinical relevance at this stage. The vast majority of the experiments investigating ERCC1-deficiency synthetic lethal relationship (Chapter 4) or ERCC1 functionality (Chapter 5) have consequently been performed by using the isogenic model of ERCC1-deficiency, as this represented a better reflection of the clinical reality with regards to the use of ERCC1 as a predictive biomarker for cisplatin sensitivity. Chapter 6 will illustrate how investigations performed on the non-isogenic model could be exploited, as exemplified by the study of *KRAS* mutated NSCLC.

A



	A549	A427	H23	H292	H358	H460	H727	H1299	H1650	A549	H1793	H1975	H2228	H2342	HOP62	BEN
ERCC1 expression																
EGFR									S			R				
KRAS	X	X	X	X (?)	X	X	X			X					X	
NRAS								X								
MYC						GA-MYC										
PIK3CA						X						X				
PTEN									X							
RB1													X			
TP53			X		X		X	X	X		X	X	X	X	X	X
STK11	X	X	X			X				X						
SMARCA4	X	X	X					X		X						
CDKN2A	X	X	X	X	X	X		X	X	X	X	X	X		X	
other		CTNNB1									TSC1	APC		MAP2K4		
Smoking habit			yes		no			yes	yes							
Histology	AC		AC		AC (BAlv)	LCC		LCC	AC	AC	AC	AC	AC	AC	AC	C

B

Cell line	ERCC1 relative expression	ERCC1 status
H1793	0.1667	low
H1975	0.1862	low
H358	0.2200	low
H23	0.3066	low
A427	0.3149	low
H292	0.3902	low
HOP62	0.8063	high
H727	0.8391	high
H2228	0.8612	high
A549_2	0.9937	high
H1299	1.0006	high
A549_1	1.0238	high
H460	1.1443	high
H1650	1.2885	high
BEN	1.3794	high
H2342	1.4523	high

Figure 3. 6. ERCC1 expression and mutational characteristics of the non-isogenic panel of 15 NSCLC cell lines

ERCC1 and XPF protein expression quantified by Western blot in the non-isogenic panel of 15 NSCLC cell lines. ERCC1 relative expression (high: red; blue: low) are displayed, as well as the main described mutations for each cell line (data pooling from COSMIC (Catalogue of Somatic Mutation in Cancer; cancer.sanger.ac.uk) and CCLE (Cancer Cell Line Encyclopedia; www.broadinstitute.org) databases, as well as in-house exome sequencing). B) ERCC1 relative expression was determined using ImageJ software, by calculating the ratio of the number of pixels of the ERCC1 signal over the corresponding β actin signal; this allowed classifying the cell lines in two main groups: ERCC1-low (ratio ERCC1/β actin < 0.4) and ERCC1-high (ratio ERCC1/β actin > 0.8).

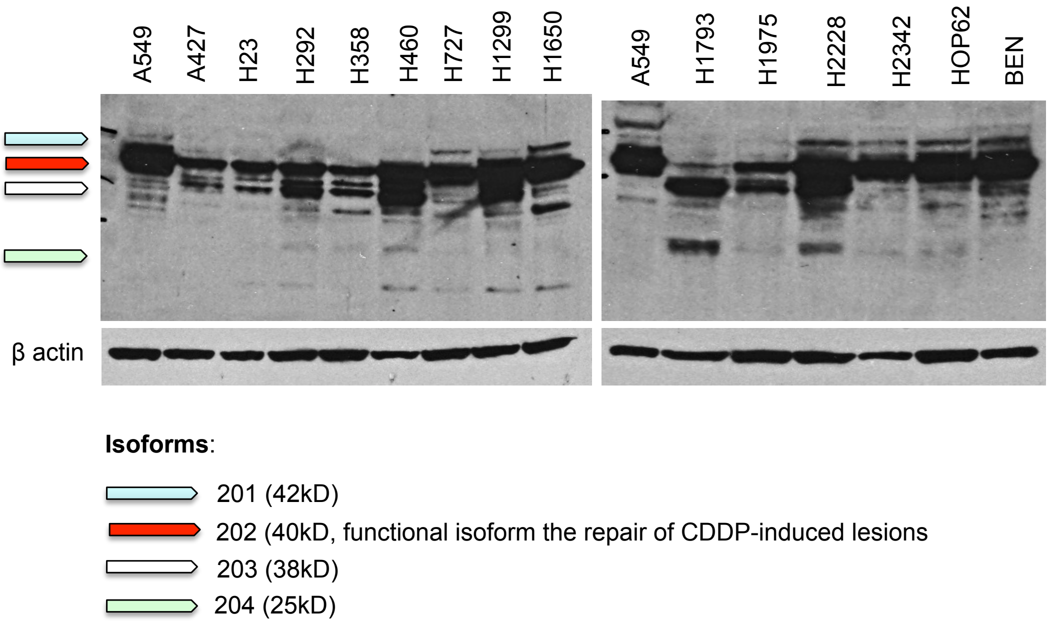


Figure 3. 7. Non-isogenic NSCLC cell lines display variable expression of ERCC1 isoforms

Expression of each ERCC1 isoform assessed by western blot (FL297, Santa Cruz) in the non-isogenic panel of NSCLC cells. Prolonged blot exposure was required to allow detection of all four distinct isoforms.

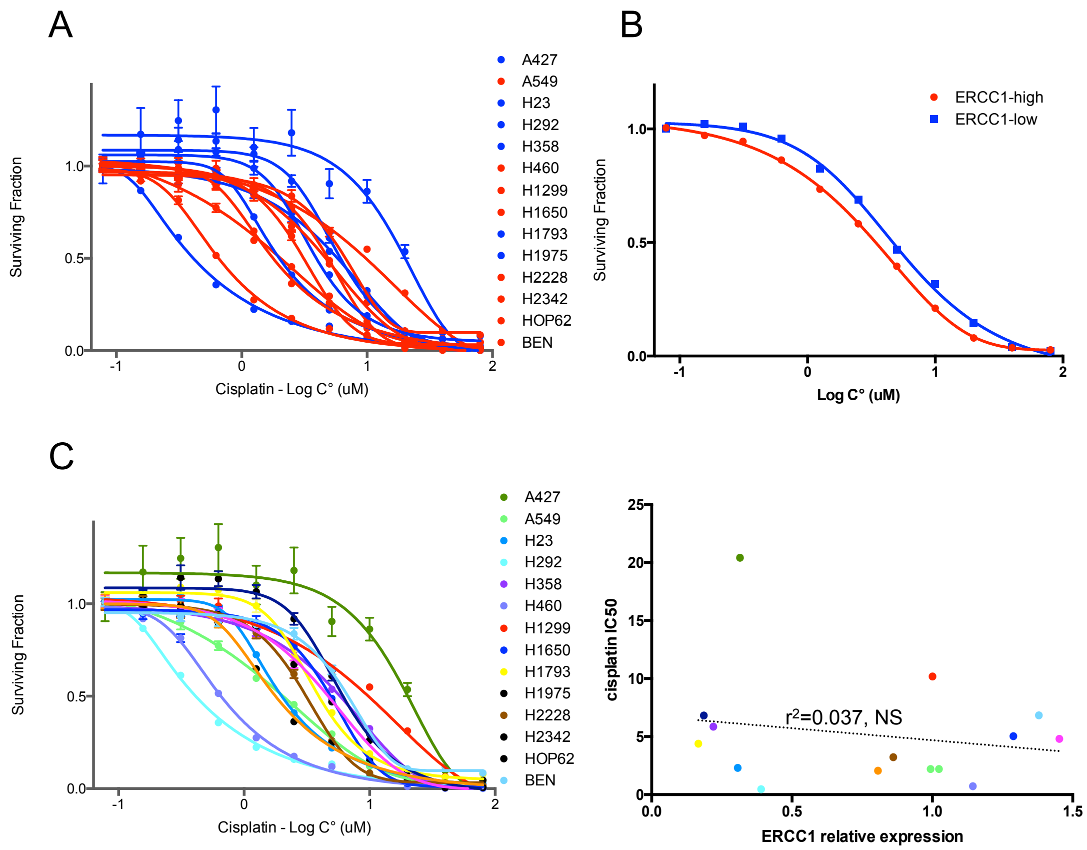


Figure 3. 8. Absence of correlation between cisplatin sensitivity and ERCC1 expression in the non-isogenic NSCLC model

A) Short-term assay (5 days drug exposure) assessing cisplatin sensitivity in the 15 non-isogenic NSCLC cell lines. Dose-response curves of ERCC1-high and ERCC1-low cell lines are depicted in red and blue colours, respectively. B) Cisplatin sensitivity of ERCC1-high and ERCC1-low groups; represented is the median of each group of cells. C) Correlation between cisplatin IC_{50} and ERCC1 relative expression as assessed in Figure 6. Linear regression showed a non-significant correlation coefficient (r^2=0.037).

3.2.2. High throughput screens

3.2.2.1. High throughput drug screen

3.2.2.1.1. Performance of the screen

In order to identify ERCC1-selective agents, we performed a drug sensitivity screen using a library of 80 drugs either already used in oncology or in late stage development (Figure 3. 9). Each cell line of the isogenic model was plated in 384-well plates and exposed to the drug library for five days, with each drug represented at four concentrations (see Chapter 2, Material and Methods). In total, each cell line was screened in triplicate, and replica data were combined in the final analysis. CellTitre Glo raw luminescence values were analysed using the CellHTS2 package of the R software (BioConductor) (Boutros 2006).

3.2.2.1.2. Screen Quality Control

In order to be eligible for further analysis, screens had to meet pre-defined quality criteria: (i) a Spearman rank correlation > 80 and (ii) a Z prime (Z') factor above 0.5. The Spearman rank correlation coefficient is a measure of correlation between two replicates; a result close to zero means there is no correlation between the two replicates, whereas a result close to 1 means there is a strong correlation between them. The Z' factor reflects the ability to separate between positive and negative controls, by taking into account four parameters (namely the means and standard deviation of both positive and negative controls); a Z' factor > 0.5 is usually considered as excellent, a Z' factor > 0 is acceptable but borderline, and a Z' factor < 0 precludes from further analysis as it reflects too much overlap between positive and negative controls (Boutros *et al.*, 2006) (Figure 3. 10. A-B). The reproducibility of each screen was also assessed in order to look for abnormal spatial pattern or high values indicating a problem with one of the replicates (e.g. row or edge effects, as represented in Figure 3. 10 on the first row). A plate plot of normalized intensities was also created in order to help identifying spatial abnormalities (Figure 3. 10. C-D). Quality controls of all drug screens included in the final dataset are summarized in Table 3. 1.

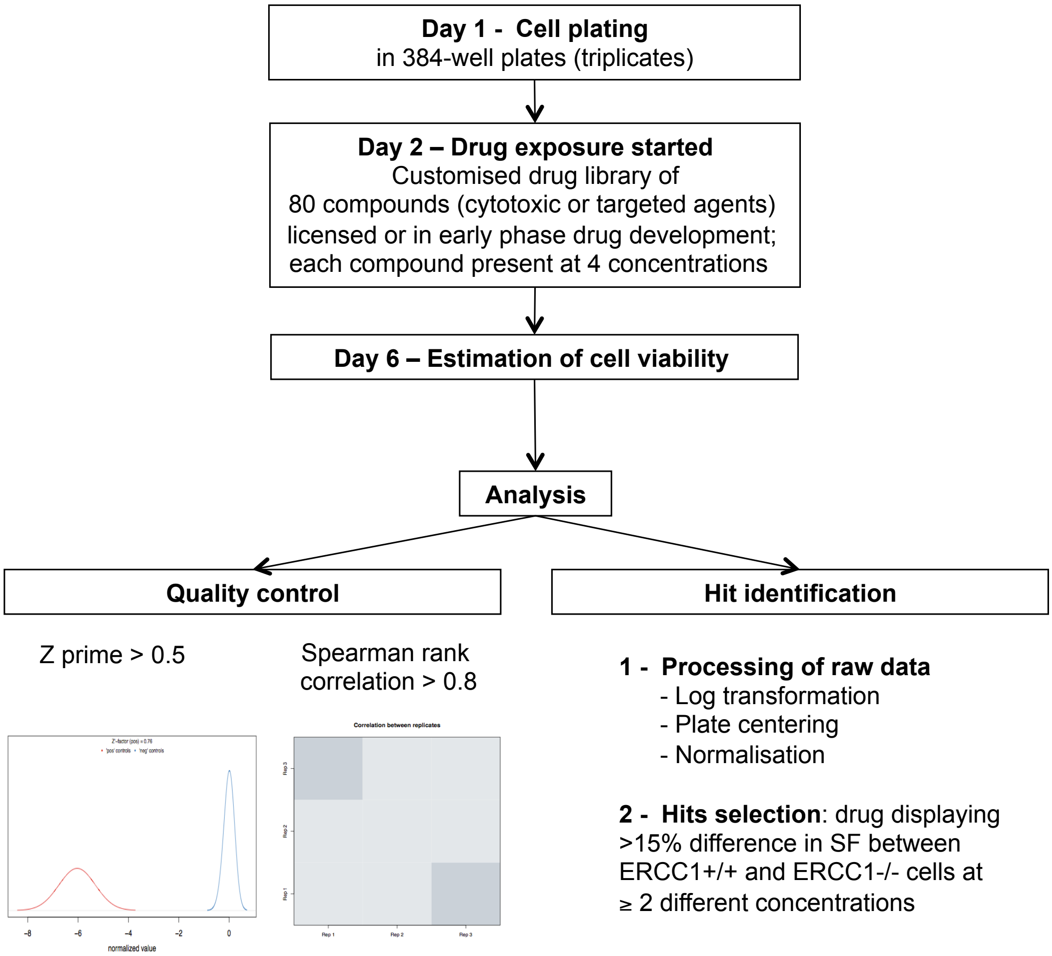


Figure 3. 9. High-throughput drug screen flowchart

High-throughput drug screen flowchart. Cells were plated at day 1 and drug was added 24h after plating. After five days of drug exposure, cell viability was assessed by CellTitre-Glo® (Promega) luminescence reading. Only screens fulfilling the pre-established quality criteria (Z prime > 0.2 and Spearman rank correlation between replicates > 0.8) were further considered. After processing of the data, potentially interesting compounds were selected based on their ability to show > 15% difference in survival fraction between the ERCC1-proficient and ERCC1-deficient populations at ≥ 2 different concentrations

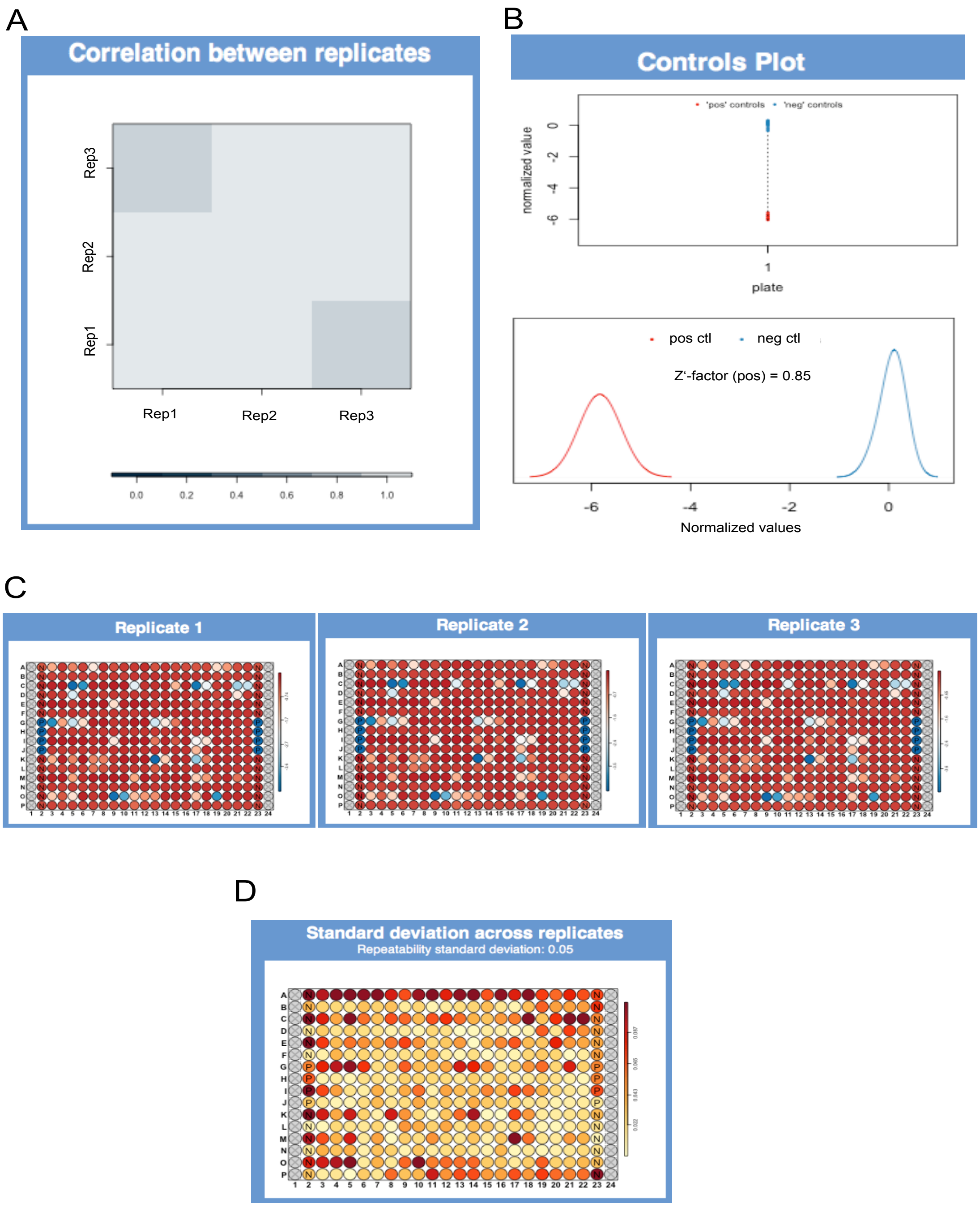


Figure 3. 10. High-throughput drug screen quality control criteria - examples

A) Spearman rank correlation between replicates of a given screen plate; depicted in this example are an excellent correlation (r^2=1) between replicates 1&2 and replicate 2&3, and an very good correlation (r^2=0.8) between replicates 1&2. B) Visual representation of the Z prime factor, with an excellent separation between positive and negative controls (Z' = 0.85). C) Plate plot of luminescence intensities, to help identifying spatial abnormalities for example due to cell plating; red colours are indicative of high cellular viability, whereas blue colours reflect the absence of viable cells. D) Plate plot of the reproducibility and standard deviation across replicates; abnormally spatial pattern or high values indicate problems with one of the replicates, such as edge effects (row A, P, and columns 2 and 23 of the example plate), or row effects (wells 2-5 of rows C, E, G, I, K, M, O) with typical pattern of cell plating issue in the displayed example. Abbreviations: rep: replicates; pos ctl: positive controls; neg ctl: negative controls

	Spearman rank correlation	Z prime factor		
		replicate 1	replicate 2	replicate 3
A549	0.88-0.95	0.76	0.74	0.79
Ahez	0.93-0.94	0.71	0.77	0.8
Ac216	0.85-0.88	0.83	0.75	0.78
Ac295	0.82-0.86	0.8	0.85	0.83
Ac375	0.82-0.87	0.79	0.8	0.47

Table 3. 1 - Quality criteria of the isogenic drug screen

Drug screens performed in the isogenic model, which were included in further analysis, all met pre-defined excellent quality criteria (Spearman rank correlation > 0.8 and Z' > 0.5).

3.2.2.1.3. Raw data analysis

Among the three main possible ways for analysing high throughput screens - namely calculation of the Surviving Fraction (SF), Z-scores, or NPI analysis - we chose to analyse these drug screen data based on the effect on cell Survival Fraction. The Z score is a statistical tool used in high-throughput screens which assumes that most of the tested drugs / siRNAs have no effect on cell survival; this was not relevant for our analysis where most drugs were expected to have an effect on cell viability. NPI analysis is preferentially used when important variations between cell lines are observed in the raw values range, i.e. the interval between positive (camptothecin, CPT11 20µM) and negative (DMSO) control; NPI (Normalized Percentage Inhibition) allows an inter-cell lines normalisation of the distribution and consequently limits the effect of cell line to cell line variability. This was nevertheless not relevant for our isogenic model, where DMSO had no effect on cell viability and CPT11 induced almost 100% cell kill in all models. We therefore investigated the presence of ERCC1-selective effects by calculating the median SF of the three replicates for each drug concentration.

3.2.2.1.4. Drug screen results

To focus our analysis on ERCC1-selective effects, we identified those drugs where there was a > 15% difference in SF between parental ERCC1-proficient and ERCC1-deficient clones at two or more drug concentrations. This approach identified 26 drugs for subsequent validation, including six different PARP1/2 inhibitors (Figure 3. 11. A; Supplementary table 1) that delivered ERCC1-selective effects among all ERCC1-deficient clones. Interestingly, several drugs targeting the EGFR or PI3K-AKT-mTOR pathways also showed significance based on our selection criteria (Figure 3. 11. B; Supplementary table 1).

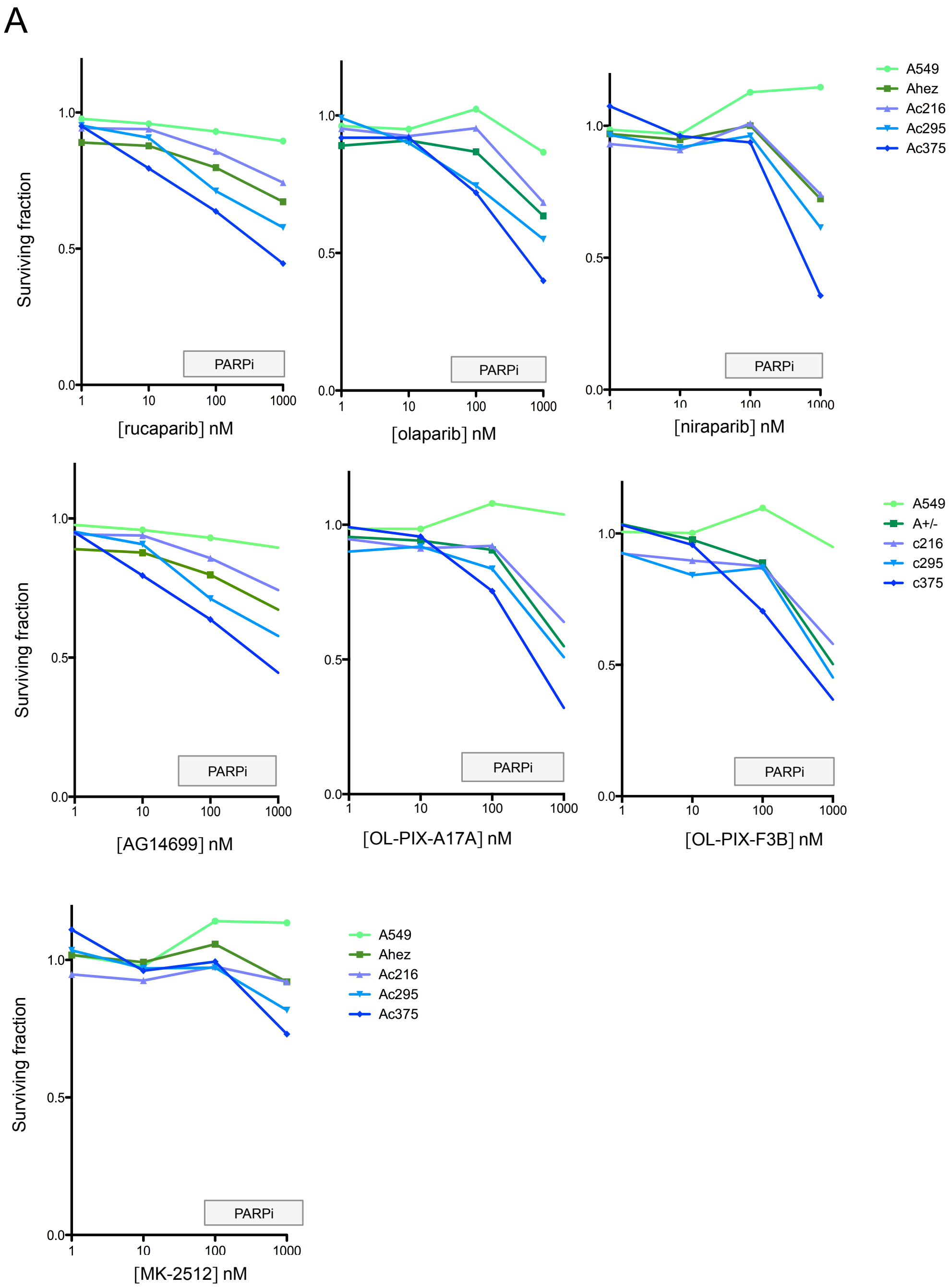


Figure 3.11. Results of the high-throughput drug screen – Hits meeting pre-defined ERCC1 selectivity criteria (see legend below).

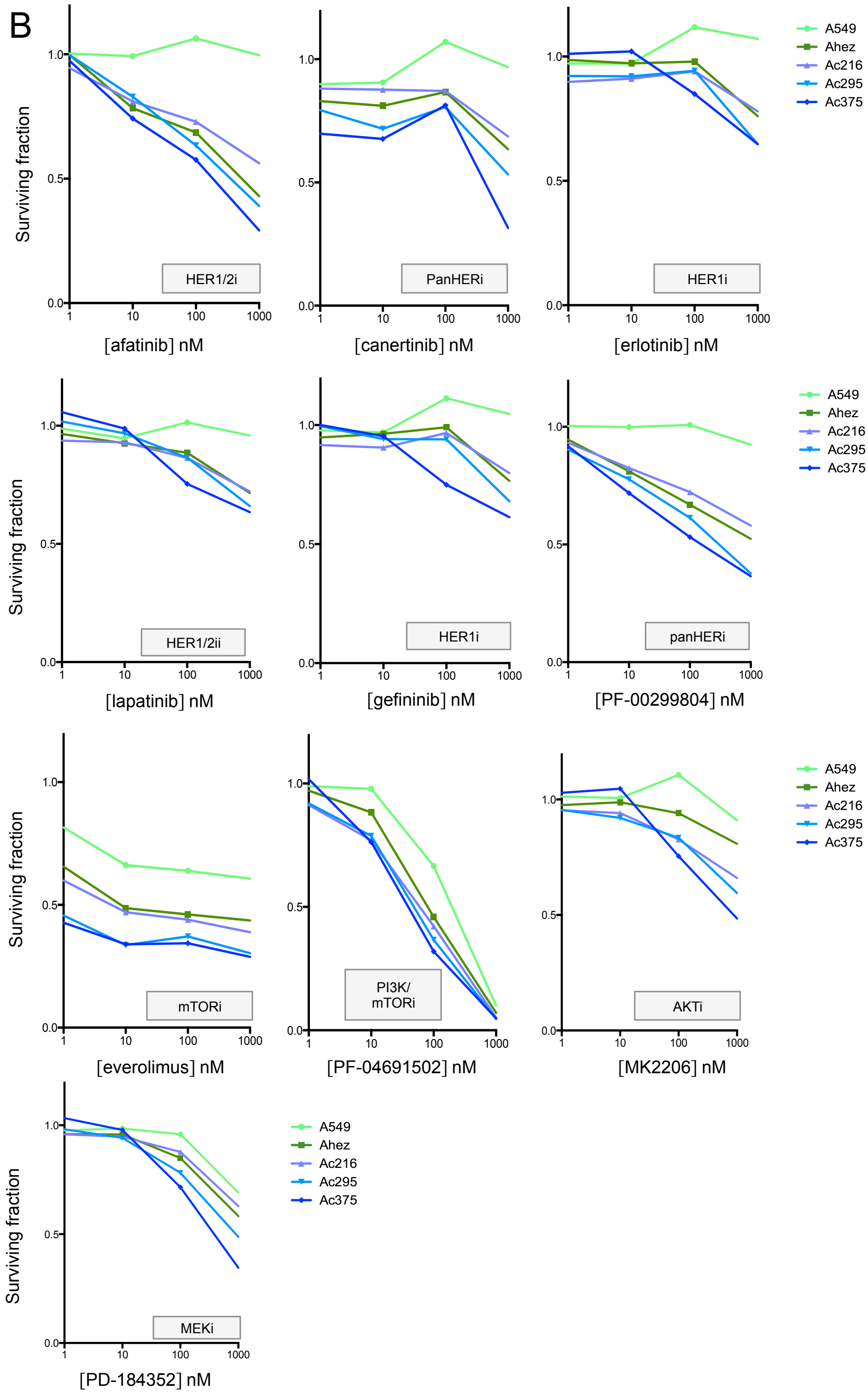


Figure 3. 11. Results of the high-throughput drug screen – Hits meeting pre-defined ERCC1 selectivity criteria (see legend below).

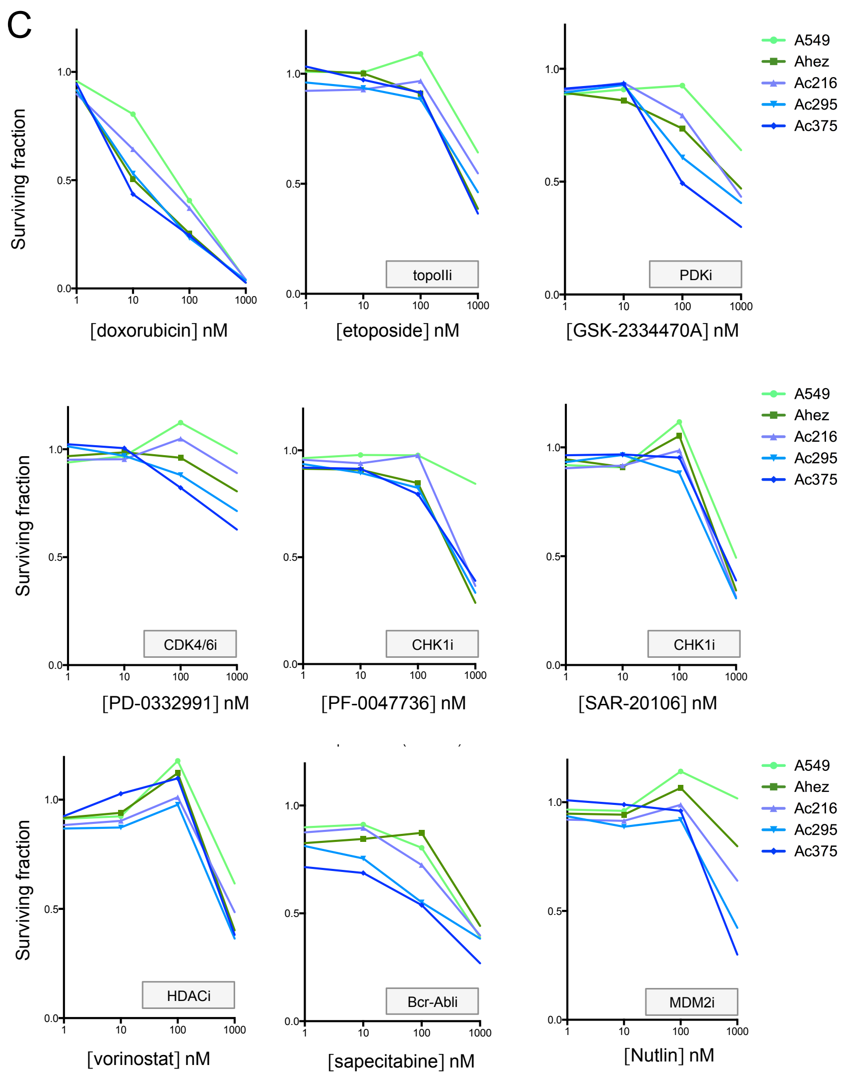


Figure 3. 11 - Results of the high-throughput drug screen – Hits meeting pre-defined ERCC1 selectivity criteria (continued from previous two pages)**.**

ERCC1-seiectivity criteria was defined as drugs displaying >15% difference in SF between ERCC1-proficient and ERCC1-deficient cells at two or more different concentrations. Names in the grey box on the bottom right corner of each plot represent display the targets of each drug. A) Seven distinct PARP1/2 inhibitors are identified as being ERCC1-selective. B) Six distinct HER1/2/3 inhibitors as well four inhibitors of downstream targets (namely MEK, mTOR, PI3K and AKT) are identified as being ERCC1-selective. C) Nine other targeted therapies or conventional chemotherapies were identified as being ERCC1-selective; among these, three targeted cell cycle regulators (CHK1, two molecules, and CDK4/6, one molecule).

3.2.2.1.5. Drug screen hit revalidation

As high throughput screens often deliver a number of false-positive results, we then aimed at validating these findings in low-throughput assays. The effect of each of the identified drug was evaluated in 96-well plate format where cells were exposed to the compound for five consecutive days (Figure 3. 12; Supplementary Table 2), and cell viability was assessed by luminescence using CellTitre-Glo® (Promega). The most promising hits were subsequently revalidated in colony formation assay (CFA), in order to increase the sensitivity of the method and be able to distinguish between cell proliferation arrest and cell death. Indeed, CellTitre-Glo® luminescence values are a reflection of the cellular ATP metabolism, and as such are only an indirect reflect of cellular viability (Figure 3. 13; Supplementary Table 2).

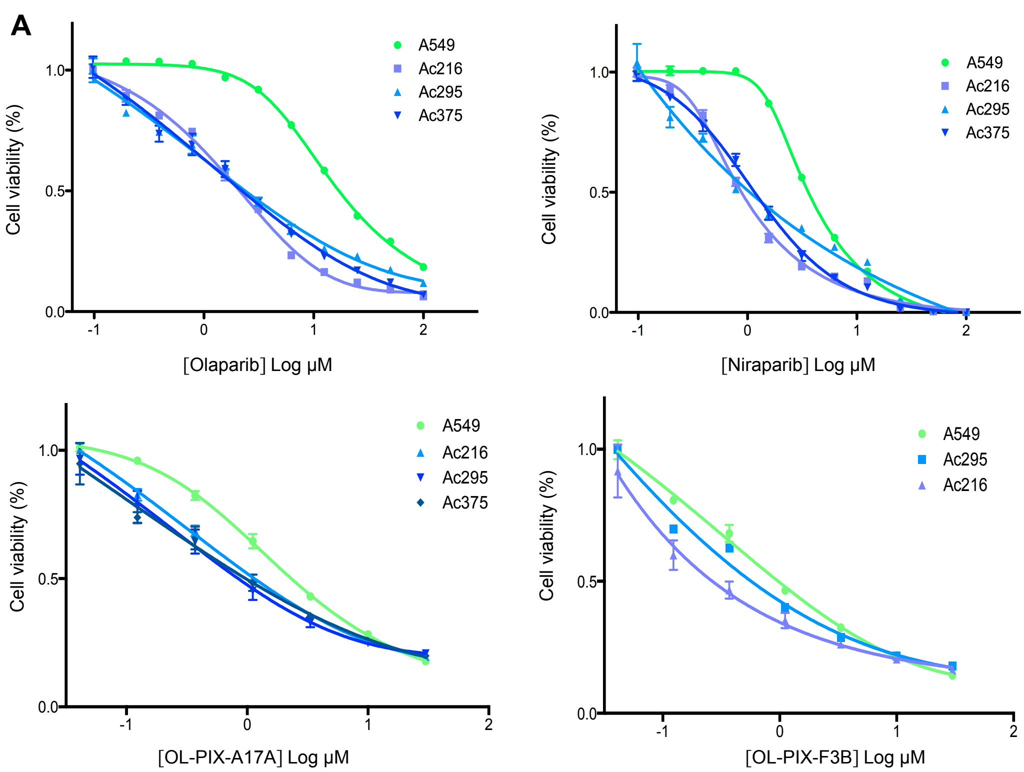


Figure 3. 12. Revalidation of the high-throughput drug screen hits in low throughput format – short term assay (legend next page).

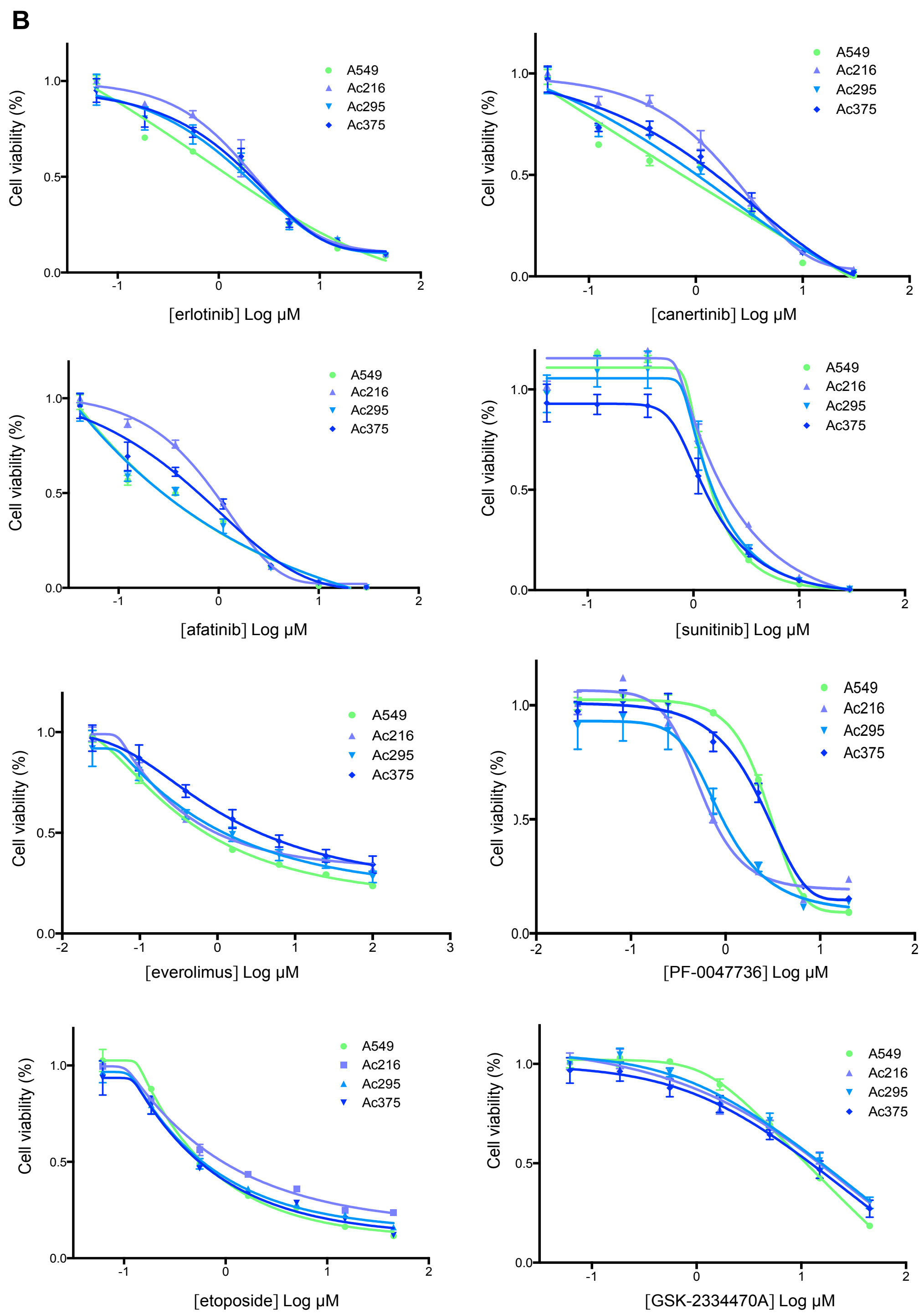


Figure 3. 12. Revalidation of the high-throughput drug screen hits in low throughput format – short term assay (continued from previous page)

Short-term assay evaluating the sensitivity of ERCC1-deficient clones compared to parental A549 cell lines to 11 of the 26 hits identified with the high throughput drug screen. Cells were exposed 5 days to the drug; error bars represent the standard deviations from the mean of three independent experiments. A) ERCC1-selective effects observed with PARP1/2 inhibitors revalidated consistently; B) Selective effects of molecules other than PARP inhibitors revalidated more inconsistently, and most of them did not show anymore selectivity towards the ERCC1-deficient population in short-term assay.

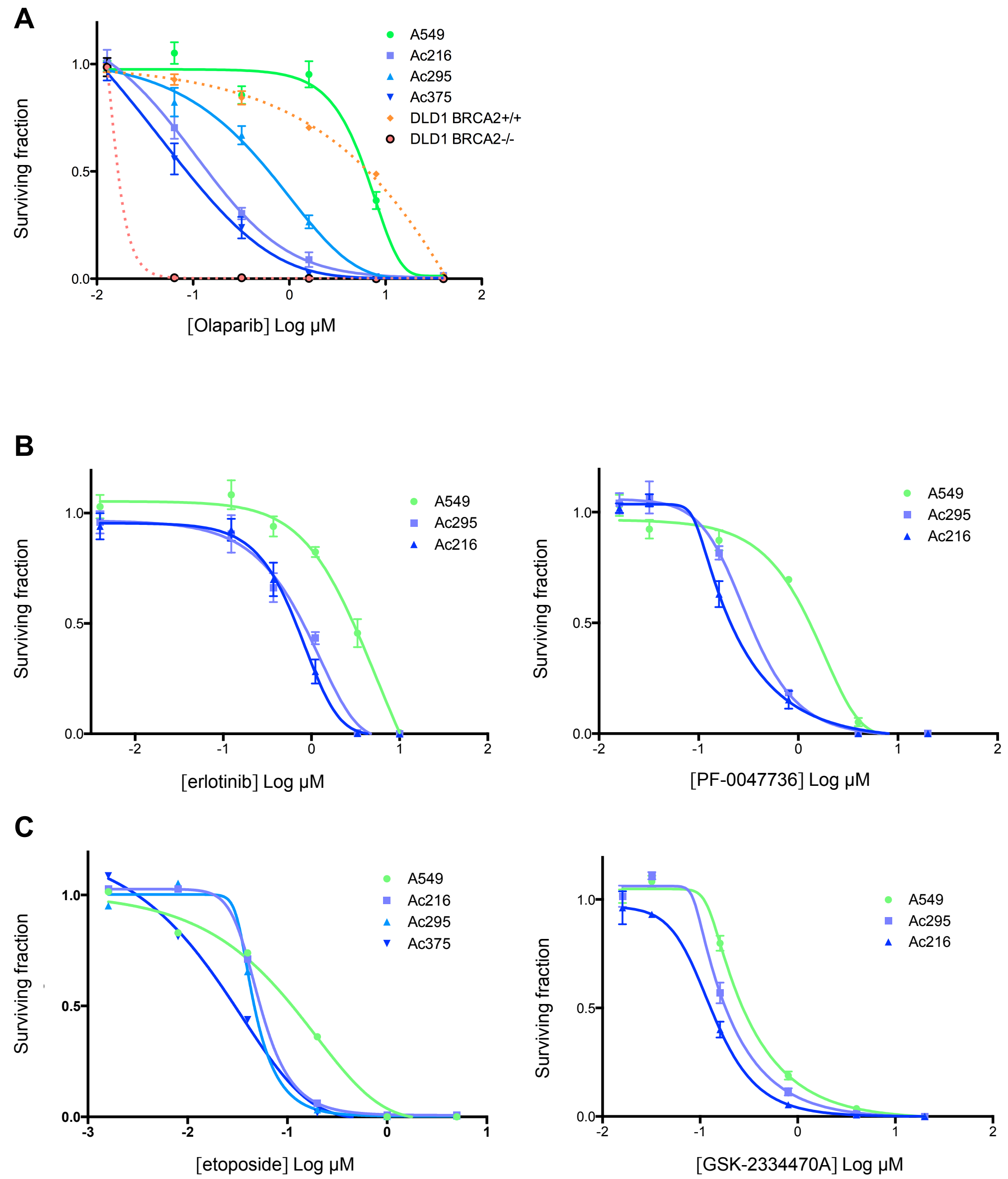


Figure 3. 13. Revalidation of the high-throughput drug screen hits in low throughput format – long term assay

Long-term assay (colony-formation assay, CFA) evaluating the sensitivity of ERCC1-deficient clones compared to wild-type A549 cell lines to five of the 26 hits identified with the high throughput drug screen. Cells were exposed to the drug for at least 15 days; error bars represent the standard deviations from the mean of three independent experiments. A) ERCC1-selective effects observed with PARP1/2 inhibitors revalidated consistently in CFA. DLD1 *BRCA2*+/+ and *BRCA2*-/- cell lines are displayed as negative and positive controls, respectively. B) ERCC1-selective effects observed with erlotinib (EGFR inhibitor) and PF-0047736 (CHK1 inhibitor) were revalidated in CFA. C) Etoposide and GSK-2334470A (PDK1 inhibitor) did not show any selective effect towards ERCC1-deficient cells when evaluated in CFA.

A major weakness of the isogenic model was the use of a single parental cell line as reference for the ERCC1-proficient population, as compared to three ERCC1-deficient clones. Therefore, any variation in that cell line could be falsely attributed to ERCC1-proficiency, as no other model of ERCC1-proficiency was used as control. In order to minimise the potential for these ERCC1-selective effects being due to the genetic drift often observed in cultured tumour cell lines, we also assessed the sensitivity to the 26 identified drug screen hits of two *ERCC1* wild-type A549 cultures maintained at two different sites - The Institute of Cancer Research (ICR) and Institut Gustave Roussy (IGR). These latter differed notably by their number of passages and cultured media (RPMI *vs* DMEM). Results obtained with *ERCC1* wild-type A549 cell lines from ICR and IGR were comparable for PARP inhibitors (Figure 3. 14).

Taken together, these subsequent validation experiments, using the same experimental procedure as for the high-throughput screen, suggested that of the 26 drugs identified in the initial analysis, only the PARP1/2 inhibitors showed a reproducible ERCC1-selective effect as described in Chapter 4. Importantly, these selective effects were consistent among all ERCC1-deficient clones and were observed using several different PARP1/2 inhibitors.

3.2.2.2. High throughput siRNA screen

RNA interference (RNAi) screens, using either short hairpin RNA (shRNA) or short interfering RNA (siRNAs), have proven highly effective in identifying novel genes involved in biological processes (Willingham *et al.*, 2004; Lord *et al.*, 2009; Brough *et al.*, 2011). In particular, siRNA (short interfering RNA) screens represent a rapid and relatively unbiased way of identifying genetic determinants that modulate cellular phenotypes and are involved in synthetic lethal relationships. This tool exploits intrinsic intracellular mechanisms that tightly regulate messenger RNA (mRNA) levels and artificially control gene expression by causing the degradation of specific mRNAs. siRNAs are short 21-28 nucleotide sequences that are then incorporated into the RISC multiprotein endonuclease complex (RNA-induced silencing complex), where a helicase unwinds duplex siRNA and allows its antisense strand to bind to mRNA. The binding of the anti-sense strand to the protein coding mRNA strand results into the cleavage and subsequent destruction of the targeted mRNA by a RNase contained within the RISC complex, ultimately resulting in reduced protein production and gene silencing or “knock-down” (Iorns *et al.*, 2007).

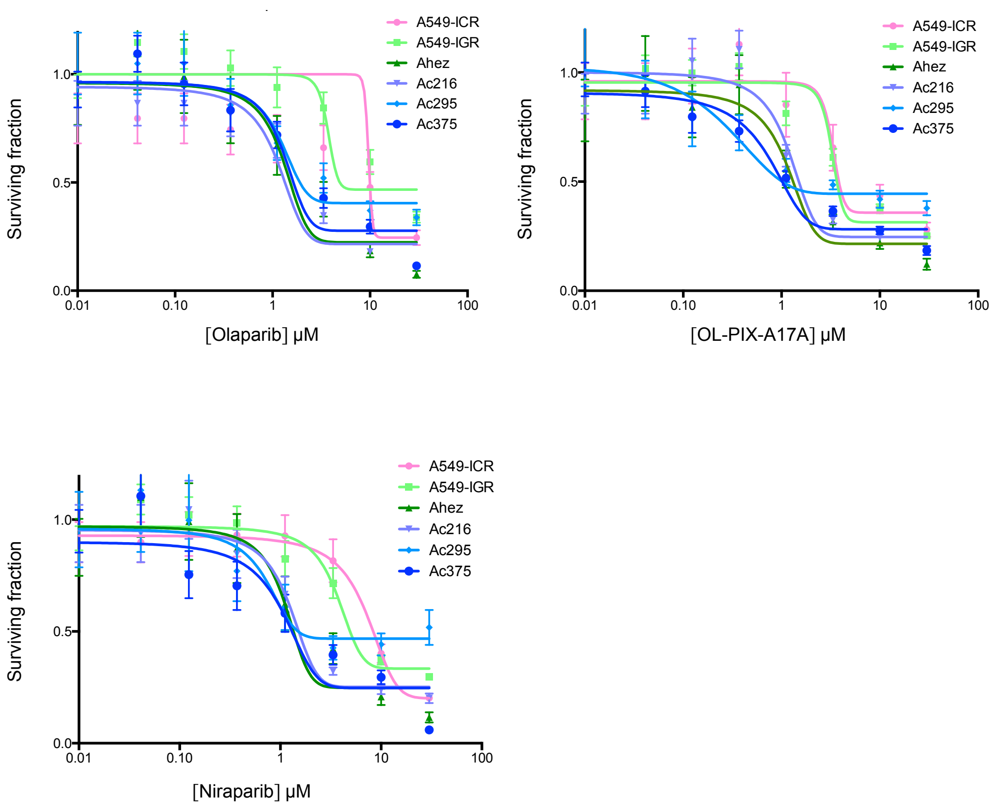


Figure 3. 14. Representative examples of the sensitivity to PARP1/2 inhibitors of two A549 wild-type cell lines from different suppliers.

Short-term assay evaluating olaparib, OL-PIX-A17A and niraparib sensitivity of ERCC1-deficient clones compared to two wild-type A549 cell lines from different suppliers (Institut Gustave Roussy and Institute of Cancer Research - originally ATCC); sensitivity of the two ERCC1 wild-type cell lines to PARP1/2 inhibitors is similar. Results obtained with other PARP1/2 inhibitors were similar, with comparable sensitivity of the two ERCC1 wild-type A549 cell lines to PARP inhibition (data not shown). Error bars represent the standard deviations from the mean of 8 independent experiments.

As siRNAs target specific sequences of particular mRNAs of interest, (Meister 2004), they allow to evaluate simultaneously the phenotypic effect of the loss of function of multiple genes when used in high-throughput arrayed libraries (Iorns *et al.*, 2007). The main key steps in performing a siRNA screen are described in Figure 3. 15.

3.2.2.2.1. Optimisation of siRNA screen conditions

The initial step in siRNA screen assay is the identification of optimal transfection conditions, or optimisation. Optimal transfection conditions have to fulfil the following criteria: (i) selection of an appropriate transfection reagent and appropriate concentration of this latter, in order to allow entry of the siRNA into the cell following membrane permeabilisation; (ii) absence of intrinsic toxicity (i.e. in absence of siRNA) of the selected transfection reagent to the cells at the selected concentration; this ensures that the cells can be transfected without excessive non-specific toxicity, which could reduce the sensitivity of the screen; (iii) appropriate cell concentration, which allows both ensuring efficient transfection conditions and detecting significant separation between negative and positive controls at the time of reading, i.e. day 7 after plating. Three siRNA were used as negative controls (All star, siCON1, siCON2) and one for positive control (siRNA targeting *PLK1*, a gene that is essential for cancer cell viability (Liu *et al.*, 2003)). A cell number of 2-4M cells in the negative control wells (i.e. 90-100% confluence with >80% viability), and < 20% viability in the positive control wells at the time of reading, were targeted as optimal conditions (Figure 3. 15). The efficacy of three transfection reagents (Lipofectamine 2000®, RNAimax® and Dharmafect 3®) was assessed at various cell densities and optimal transfection conditions were selected based on the above criteria (Figure 3. 16).

3.2.2.2.2. Performance of the screen

Six in-house customized short interfering RNA (siRNA) libraries targeting 2060 distinct genes were purchased from Dharmacon, with targets of the selected siRNAs being distributed as follows: kinases (779 genes), phosphatases (256 genes), tumour suppressors (17 genes), DNA repair related genes (242 genes), PARP-related genes (44 genes), and metabolic targets (722 genes). Each gene was represented in one well of a 96 well mother matrix plate, with each well containing a pool (SMARTpool®) of four different siRNA targeting the same gene. A maximum of eighty SMARTpools® were arrayed on each plate, and each plate was in addition supplemented with twelve wells of negative siControl (siCON1, siCON2, AllStar; 4 wells each), and four wells of positive siControl (siPLK1).

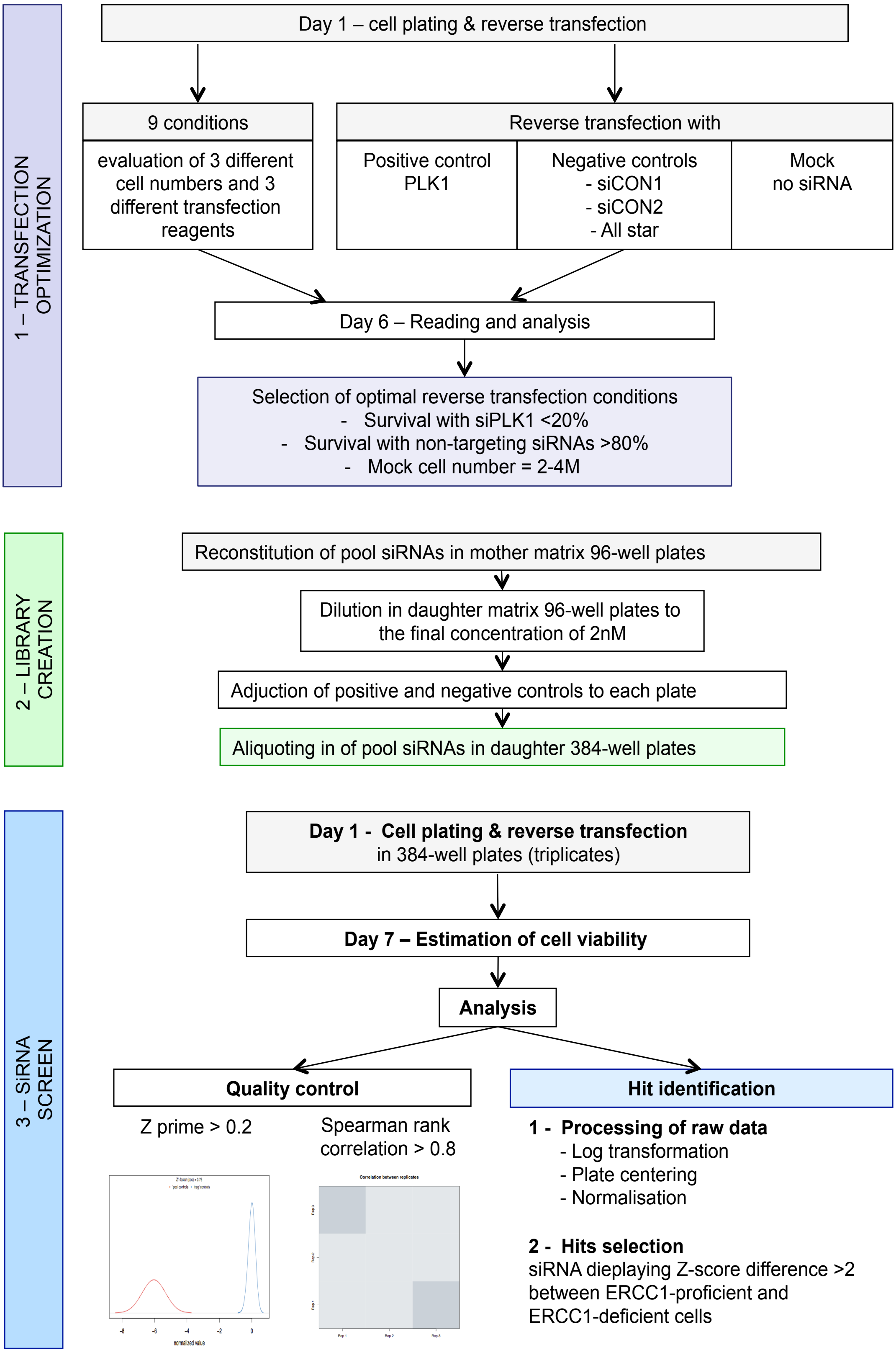


Figure 3. 15. High-throughput siRNA screen flowchart (legend next page)

Figure 3.15. High-throughput siRNA screen flowchart (figure previous page)

The completion of the high throughput siRNA screen required three main steps.

1 – Transfection optimization: cells were plated at day 1 and reverse transfected in 384-well plates under variable conditions in terms of cell density and transfection reagent. Reading occurred at day 7 in order to mimic the conditions in which the siRNA screen would further be realized. Optimal reverse transfection conditions were determined to allow best separation between positive and negative controls, and avoid excessive cell confluence, which would reduce the sensitivity of the screen.

2 – Customised library creation. The selected library containing the siRNAs of interest was reconstituted in 96-well matrix plates and further aliquoted in 384-well plates, with each well containing 5µL of the siRNA of interest at a 200nM concentration.

3 – Screen completion: Cells were plated and reverse transfected at day 1 in triplicates. Cell viability was assessed at day 7 by CellTitre-Glo® (Promega) luminescence reading. Only screens fulfilling the pre-established quality criteria (Z prime > 0.2 and Spearman rank correlation between replicates > 0.8) were further considered. After processing of the data, potentially interesting siRNAs were selected based on their ability to show a > 2 difference in median Z-scores between the ERCC1-proficient and ERCC1-deficient populations.

Figure 3.16 - Optimisation of siRNA screen transfection conditions (Figure next page)

A) ERCC1-isogenic cell lines were transfected either with one the negative siRNA controls (siCON1, siCON2 or Allstar), with the positive control (PLK1), or were plated in wells where no siRNA was added (Mock). For each cell line, three different transfection reagents and three different cell densities at plating were assessed. Cell viability was assessed after 7 days using CellTitre-Glo and luminescence reading on a Victor X5 plate reader, and mean raw luminescence values were normalised to the Mock. Optimal conditions were further selected as described in Figure 3.16. B) Bar plots of the selected optimal transfection conditions, with the mean raw luminescence values, with error bars representing the standard deviation from the mean of 4 replicates. When negative control siRNAs produced excessive toxicity (e.g. siCON#1 and siCON#2 in the Ac375 cell line), these were further removed from further screen QC analysis, with AllStar being kept as the only negative control only.

A

		Transfection reagent (0.05µL/well)		
Cell line	**Cell number**	lipofectamin	RNAimax	Dharmafect 3
A549	250	✓✓✓	✓✓✓	✗
	500	✓	✓✓	✗
	750	✗	✓	✗
Ahez	250	✗	✓✓✓	✗
	500	✗	✓✓	✗
	750	✗	✓✓	✗
Ac216	250	✓✓	✓	✗
	500	✗	✓✓✓	✗
	750	✗	✓✓✓	✗
Ac295	250	✓	✓	✗
	500	✗	✓✓	✗
	750	✗	✓✓✓	✗
Ac375	250	✓✓	✗	✗
	500	✓	✗	✗
	750	✓	✓✓	✗

✓✓✓ selected optimal transfection condition

B

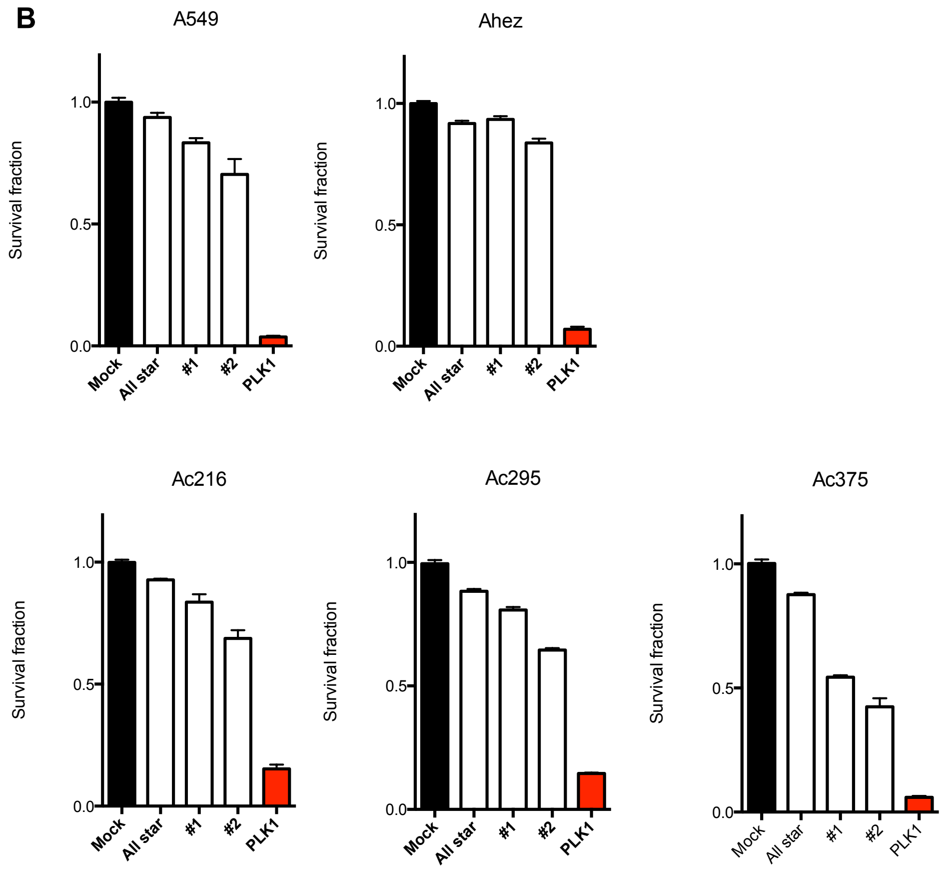


Figure 3. 16. Optimisation of siRNA screen transfection conditions (legend previous page)

The screen was performed in triplicates and carried out as described in Figure 3. 15. Briefly, cells were plated and reverse transfected following optimal conditions at day 1; raw luminescence values were read at day 7 after media removal followed by addition of CellTitre-Glo® to each well.

3.2.2.2.3. Screen Quality Control

Similar to what has been described above regarding drug screens, siRNA screens had to meet the following pre-defined quality criteria in order to be included in the final dataset: (i) a Spearman rank correlation > 80 and (ii) a Z prime (Z') factor above 0.5. The reproducibility of each screen was also assessed in order to look for abnormal spatial pattern or high values indicating a problem with one of the replicates, together with the generation of a plate plot of normalized intensities. Quality controls of all drug screens included in the final dataset are summarized in Figure 3. 17.

3.2.2.2.4. siRNA screen analysis

Each screen was performed in triplicate with results of all three screens being used in the final analysis. The effect of individual siRNAs on cell viability was assessed by calculating the Z-score for each oligonucleotide: the Z-score represents the number of standard deviations that separates the effect of this particular siRNA from the mean of the dataset. This allows to convert data from different data sets into scores that can be more accurately compared to each other, which is essential when several screens are performed on multiple cell lines which display variable transfection efficacy. The assumption that the vast majority of siRNAs has no effect on cell viability was made to ensure the relevance of this calculation. Consequently, a Z-score of 0 would represent no effect on cell viability, whereas a negative Z-score of < -2 would reflect a significant loss of cell viability.

Raw luminescence values from each plate were first $\log_2$-transformed and normalised to the median luminescence score for all experimental wells on one plate, in order to take into account plate-to-plate variation in transfection and when acquiring luminescence. The three replicates were combined in the final analysis, where the median Z-score for each individual siRNA was calculated using the median normalized value of each siRNA and the median absolute deviation (MAD), in order to account for the variability within the screen.

A

		Spearman's rank correlation				Z prime		
		plate 1	plate 2	plate 3	plate 4	replicate 1	replicate 2	replicate 3
Kinome Tum Suppr DNA Repair	A549	0.9-0.94	0.82-0.88	0.86-0.94	NA	0.58	0.63	0.45
	Ahez	0.96-0.97	0.93-0.95	0.94-0.96	NA	0.76	0.78	0.71
	Ac216	0.89-0.9	0.87-0.89	0.86-0.89	NA	0.52	0.5	0.62
	Ac295	0.95-0.96	0.96-0.96	0.87-0.93	NA	0.52	0.31	0.54
	Ac375	0.88-0.97	0.92-0.97	0.9-0.98	NA	0.39	0.51	0.19
Mbolic Ptase PARP	A549	0.86-0.9	0.94-0.94	0.85-0.9	0.91-0.93	0.58	0.66	0.69
	Ac216	0.9-0.94	0.93-0.94	0.93-0.94	0.94-0.95	0.64	0.61	0.59
	Ac295	0.19-0.93	0.73-0.84	0.8-0.85	0.63-0.77	0.45	0.1	0.34
	Ac375	0.94-0.97	0.95-0.96	0.94-0.96	0.96-0.97	0.66	0.73	0.72

B

C

D

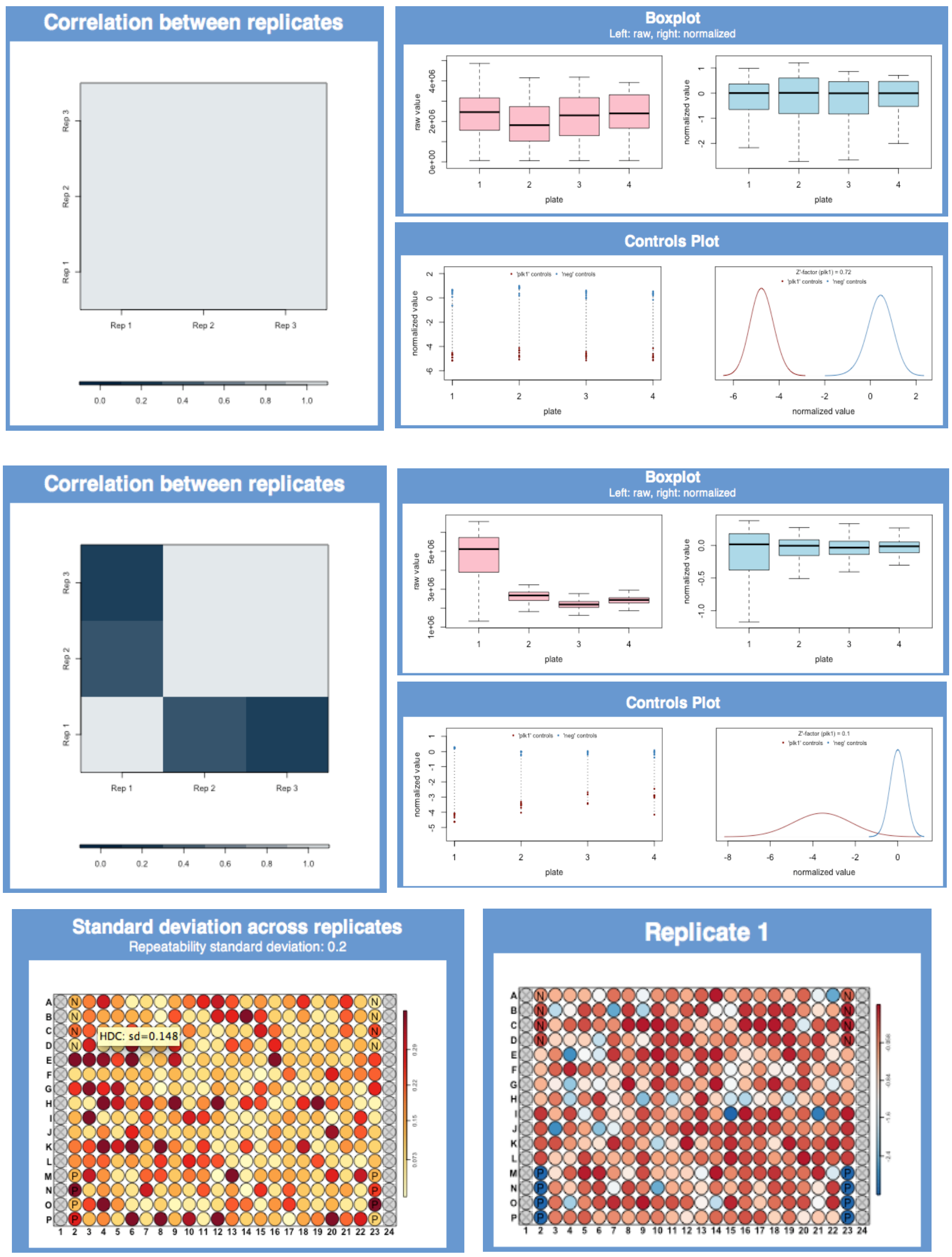


Figure 3. 17. High-throughput isogenic siRNA screen quality control (legend next page)

Figure 3.17. High-throughput isogenic siRNA screen quality control (figure previous page)

A) Quality Control of the siRNA screens performed in the isogenic model, which were included in further analysis. All screens met pre-defined excellent quality criteria, including Spearman rank correlation > 0.8 and Z' > 0.5 (20/27 plates) or Z' > 0 (7/27 plates). B) Spearman rank correlation between replicates of a given screen plate; depicted in this example are an excellent correlation (r^2=1) between all replicates. B) Visual representation of the Z' factor, with an excellent separation between positive and negative controls (Z' = 0.72). C) Illustration of problem encountered with one of the plates of the metabolic-phosphatase screen of the Ac295 cell line, impacting on all QC criteria. Such isolated issue could be easily identified thanks to the use of CellHTS2 package during R analysis of the screen results. D) Examples of plate plots of reproducibility and luminescence intensities, to help identifying spatial abnormalities due to cell plating.

3.2.2.2.5. siRNA screen results - hits for ERCC1-deficient cells

Hits were selected based on (i) a Z-score difference >2 between the parental cell line and the median of the three ERCC1-deficient clones, (ii) a consistent effect among all clones and (iii) a Z-score comprised between -1 and 1 in the parental cell line (i.e. the absence of effect of the oligonucleotides of interest in the reference population). This allowed the identification of 69 hits, 33 of which were selected for subsequent revalidation. The siRNAs that induced a Z-score difference > 2 between the parental cell line and the median of the clones are listed in Figure 3. 18 and Figure 3. 19, where each table corresponds to a different library according to the time at which the screen was performed. In order to look for enriched pathways in the results of each screen, interactions between hits were visualized using the String server and database (Figure 3. 18, Figure 3. 19).

A

Gene name	A549 Z-score	Ahez Z-score	Ac216 Z-score	Ac295 Z-score	Ac375 Z-score	Z-score A549 minus median Z-score clones
GALK1	-0.27764	-1.06808	0.92186	-6.08560	-5.55174	**5.27410**
DGKD	0.28802	0.61311	1.26965	-4.89673	-4.83719	**5.12521**
MAPK10	1.14688	0.82332	1.08640	-4.92333	-3.79533	**4.94222**
ERCC2	0.68272	0.22180	0.90639	-3.91078	-5.39703	**4.59350**
EXT1	1.03237	0.31214	0.68396	-3.06825	-3.41680	**4.10062**
MAPK1	-2.45042	-4.43714	-4.85757	-7.47226	-6.40180	**3.95138**
XAB2	-11.49758	-9.08189	-15.43453	-18.62313	-11.74848	**3.93694**
CHEK1	-2.66631	-4.44969	-6.87620	-6.60262	-2.83883	**3.93631**
WEE1	0.27406	-1.04245	-5.06474	-3.64090	0.35354	**3.91497**
RPS6KA2	-4.11472	-7.11541	-4.28403	-10.27430	-7.88840	**3.77368**
CALM3	-0.30660	-2.80466	-1.87663	-3.97349	-3.97042	**3.66382**
PRKAG3	-3.67853	-6.55832	-7.04329	-4.44861	-7.91605	**3.36476**
PCTK3	-0.96730	-6.13710	-3.73893	-4.31528	-5.16784	**3.34798**
PRKCZ	-1.69565	-5.24193	-1.15767	-8.31617	-4.92842	**3.23277**
PCNA	-1.08178	-4.30064	-4.26863	-2.98820	-4.59978	**3.18685**
TAF1	-3.62892	-3.75595	-3.28794	-10.02320	-6.81550	**3.18657**
STK32A	0.18178	0.00248	-1.29906	-2.98850	-3.09569	**3.17028**
MLH1	-0.36393	0.39514	0.85747	-3.97378	-3.43882	**3.07489**
CSNK1E	-0.86892	-4.87900	-3.85806	-3.23420	-4.64831	**2.98913**
MAP2K1	1.03928	0.93933	0.97099	-2.93374	-1.93090	**2.97018**
CDC2L2	-11.46979	-8.51128	-14.54891	-14.28408	-11.01549	**2.81430**
CDC2L1	-2.77640	-4.38654	-5.54715	-5.01555	-6.44498	**2.77075**
AKT3	-0.73464	-2.57375	-3.77250	-2.08549	-3.49718	**2.76253**
PRKCL2	-0.70275	-2.38392	-4.22641	-3.43049	-1.84330	**2.72774**
PHOX2B	0.46327	-0.12097	0.28444	-2.24783	-2.47732	**2.71110**
CSNK1G1	-1.03827	-6.96262	-4.86595	-3.60413	-3.68371	**2.64544**
NME6	-0.38647	-3.39259	-2.59279	-3.17984	-2.93260	**2.54612**
MASTL	-0.45178	-3.46360	-2.96321	-2.49527	-3.26626	**2.51142**
PAK3	0.24394	-0.17849	-2.25137	-0.92675	-2.25748	**2.49531**
PIK3C2B	-0.65388	-1.11714	-3.13402	-2.94903	-3.77244	**2.48014**
TTBK1	-3.94444	-4.49889	-7.73744	-6.38141	-3.22016	**2.43697**
PIK3C2A	-0.36193	-0.85311	-2.25397	-2.73021	-4.05845	**2.36829**
PMS2	-0.31701	0.02151	-0.03656	-3.76595	-2.66346	**2.34645**
XPA	0.79120	0.85738	0.68558	-2.81891	-1.55190	**2.34310**
DDB2	1.06824	0.43240	1.06496	-1.22007	-1.35261	**2.28831**
STK35	-0.17073	-0.84735	-3.76630	-0.92414	-2.40833	**2.23760**
PRKCD	-1.69005	-3.08754	-0.00538	-4.61460	-3.92147	**2.23141**
CDKL2	1.11499	-0.17627	1.13087	-2.31868	-1.07031	**2.18530**
PNCK	-0.45097	-2.52258	-2.63445	-2.31043	-3.50302	**2.18347**
PKM2	-0.10433	-2.21861	-2.45452	-1.18171	-2.26729	**2.16296**
PRPS1	-0.64755	-0.77146	0.25094	-2.80008	-2.87718	**2.15252**
TTK	-3.31535	-4.69968	-8.30702	-5.44491	-4.53876	**2.12956**
CIT	-0.77379	-1.69481	0.00679	-2.86784	-4.70203	**2.09405**
BCKDK	-1.04296	-2.83878	-4.39073	-1.94434	-3.13125	**2.08829**
PIP5K1A	0.50859	0.44305	-1.84498	-0.39183	-1.57415	**2.08274**
MAP2K3	-0.86598	-1.90123	-0.27857	-3.41559	-2.94635	**2.08037**
PER1	0.64624	0.35273	0.89600	-1.57238	-1.42863	**2.07487**
CAMKIINALPH	-0.19199	-4.42688	-2.22126	-1.11672	-2.99053	**2.02927**
CHEK1	-4.17510	-3.56813	-7.33565	-6.18986	-2.63891	**2.01476**
AURKC	0.28191	-4.42837	-0.73985	-1.72578	-2.78844	**2.00769**
CDH1	0.56362	0.44215	1.06919	-1.44302	-1.66878	**2.00665**
AAK1	-0.85242	-0.04963	-3.38190	-2.11607	-2.85540	**2.00298**

Figure 3.18. Hits from the Kinome - Tumour Suppressors – DNA repair plates showing selectivity for the ERCC1-deficient population (legend next page)

B

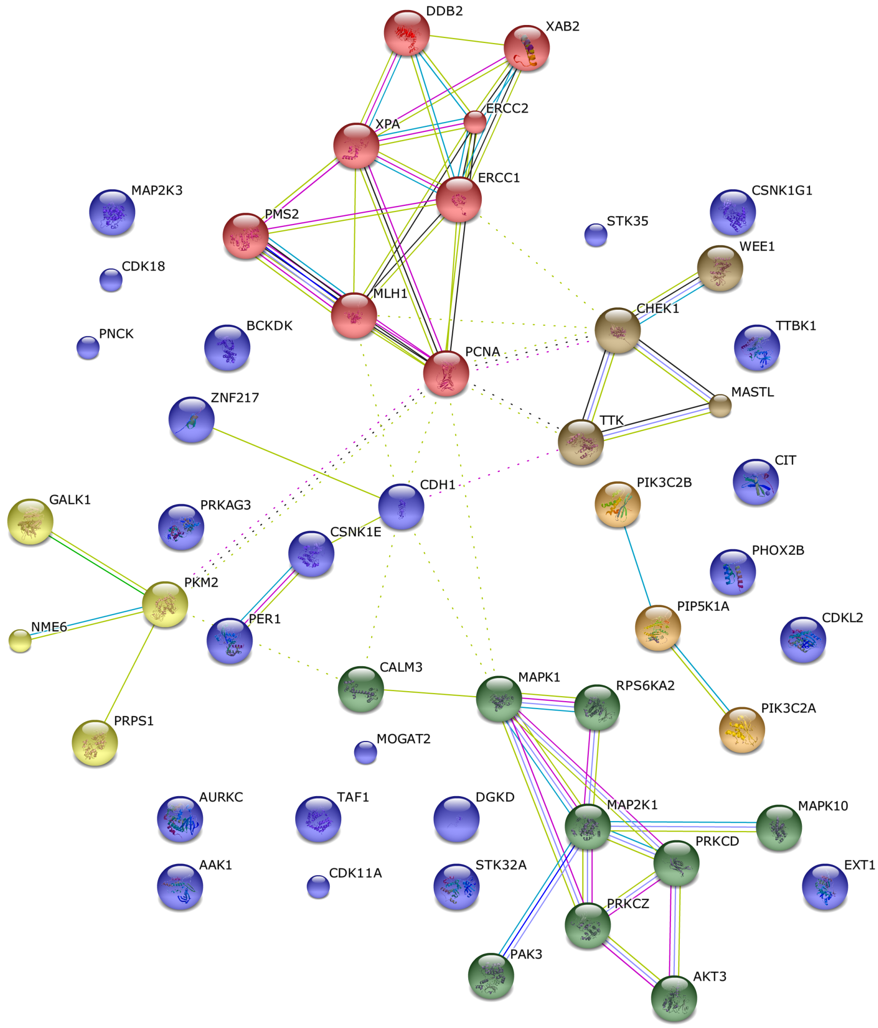


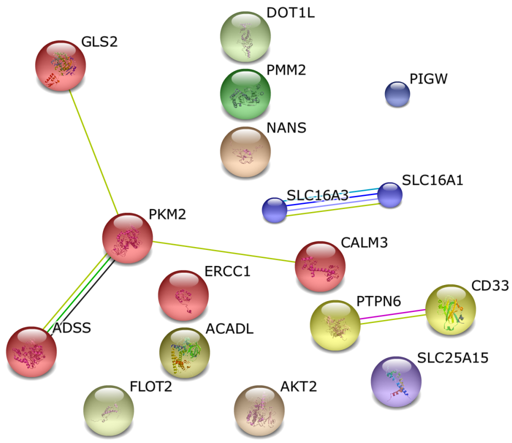
Figure 3. 18. Hits from the Kinome - Tumour Suppressors – DNA repair plates showing selectivity for the ERCC1-deficient population (continued from previous page)

A) List of hits from siRNA screen in difference Z score order, with siRNA target gene, sensitivity Z score for each cell line, and the difference in median Z-scores between each population. B) Visual representation of the hits from Kinome – Tumour Suppressors – DNA repair plates of the siRNA screen using STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) v9.1 database, to help identifying functional networks within the hit list. Hits were further separated into clusters following a Markov Cluster Algorithm of 2.

A

Gene name	A549-FR Z-score	Ac216 Z-score	Ac295 Z-scores	Ac375 Z-score	median Z-score A549 minus median Z-score clones
NANS	-0.7077033	-4.1026845	-7.1713833	-4.0351237	**3.394981175**
PIGW	0.3825628	-2.4057106	-3.5070456	-2.2118605	**2.788273398**
AKT2	-0.2725122	-1.9568805	-17.685096	-2.8427813	**2.570269119**
DOT1L	-0.260446	-2.7424996	-2.9648202	-1.7635262	**2.482053589**
PMM2	0.9319885	-1.7381581	-0.7661784	-1.5356562	**2.467644664**
CALM3	0.85316684	-1.881007	0.19332365	-1.5983761	**2.451542945**
GLS2	0.49438269	-2.2276897	-1.2779788	-1.8074496	**2.301832266**
SLC25A15	-0.0778927	-2.6429957	-2.3637613	-2.1901652	**2.285868557**
CD33	0.79251543	-1.7265478	0.94592982	-1.4508601	**2.243375492**
PTPN6	-0.10548	-1.7657369	-2.3280629	-2.7799449	**2.222582901**
FLOT2	1.07077582	-1.0899442	0.28538603	-1.3389539	**2.160719988**
SLC16A3	-0.2282745	-2.2308397	-2.4957048	-2.3802032	**2.151928682**
ADSS	1.32617733	-0.7273055	1.33601007	-1.0004559	**2.05348285**
SLC16A1	0.65018241	-1.3709844	-0.6856968	-1.5655597	**2.021166815**
ACADL	-0.1938044	-2.2083639	-3.0960234	-1.7013334	**2.014559488**
PKM2	0.04727656	-1.9574963	-2.611963	-1.7258831	**2.004772864**

B



C

Gene name	A549-FR Z-score	Ac216 Z-score	Ac375 Z-score	median Z-score A549 minus median Z-score clones
ADPRHL1/ARH2	0.96956498	-5.5158821	-5.0587913	**6.256901649**
ADPRHL1/ARH2	0.8375023	-5.6764915	-4.7065228	**6.029009465**

Figure 3. 19. Hits from the Metabolic – Phosphatase, and PARP plates showing selectivity for the ERCC1-deficient population

A) List of hits from siRNA screen in difference Z score order, with siRNA target gene, sensitivity Z score for each cell line, and the difference in median Z-scores between each population. B) Visual representation of the hits from Metabolic - Phosphatase plates of the siRNA screen using STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) v9.1 database, to help identifying functional networks within the hit list. Hits were further separated into clusters following a Markov Cluster Algorithm of 1. C) Single hit from the PARP plate siRNA screen (each siRNA was aliquoted twice in this plate, and both replicates came up consistently).

3.2.2.2.6. siRNA screen revalidation – hits for ERCC1-deficient cells

It is now well recognised that siRNAs, as well as silencing expression of the target gene, also suppress the expression of other genes without perfect sequence complementarity. The mechanism of these off-target effects is thought to be due to off-target processing of the siRNA in a similar manner to microRNAs (Jackson *et al.*, 2003; Fedorov *et al.*, 2006). To identify the siRNAs that modified response due to a specific on-target effect, and exclude those that modified response due to off-target effects, 33 of the 69 hits identified during the high throughput assay were selected for revalidation, where the effect of each individual deconvoluted oligo specie that had initially been evaluated within the SMARTpool was evaluated separately. It is generally considered that observation of a phenotype caused by two distinct siRNA species indicates that it is unlikely to be the result of an off-target effect (Echeverri *et al.*, 2006). Therefore, effects were taken as being potentially on-target when two or more siRNA targeting the same gene decreased cell survival in ERCC1-deficient clones by more than 20% as compared to the parental cell line, and effects were deemed as being highly likely to be on-target when a similar effect was observed with three or more individual siRNAs (validation by redundancy). Genes that only modified ERCC1-deficient cell survival with one siRNA were deemed as likely to be off-target effects.

Revalidation was conducted in 384-well plates, after reconstituting and aliquoting the customised revalidation library as described above (Figure 3. 15). Two independent screens were performed in triplicates for each cell line, and an extra-screen was performed in triplicates using the A549 cell line maintained at the Institute of Cancer Research to account for a potential genetic drift of the parental Institut Gustave Roussy A549 cell line, as well as increase the robustness of data arising from the reference cell lines. Raw data analysis was performed using the NPI analysis (Normalised Percentage Inhibition); indeed, Z-scores could not be used in this setting given that the assumption that most siRNAs had no effect was not relevant anymore, and that discrepancies in the cell killing resulting from the positive control (siPLK1) precluded the use of survival fraction analysis.

Among the 33 selected siRNAs included in the revalidation plate, four of them showed effects that were deemed to be at least potentially on-target following the observation of redundancy among deconvoluted siRNAs (Figure 3. 20): AAKT, MAP2K1, MAPK10, and NUTD1 (caution was exercised for MAP2K1 and MAPK10 as discrepancies were observed between the independent screen replicates for the clones). Of note, this proportion of revalidation is within what is usually observed when revalidating high-throughput screens (in A. Ashworth's lab and other labs).

3.2.2.2.7. ERCC1-high hits of the siRNA screen

The main purpose of the thesis was to study ERCC1-deficiency. As ERCC1-high tumours are usually resistant to platinum therapy, there is also a need for novel therapeutic approaches in this population. However, ERCC1-high hits of mechanistic interest were identified following the high-throughput siRNA screens (by performing the reverse analysis than the one described above for the identification of ERCC1-low hits), and further revalidated (Figure 3. 21, Figure 3. 22). This allowed the identification of 162 hits (37 from the Kinome - Tumour Suppressor – DNA repair plates; 115 from the Metabolic -Phosphatase plates; 10 from the PARP plates), 73 of which were selected for subsequent revalidation. Among those, 10 of them showed effects that were deemed to be potentially on target following the observance of redundancy among deconvoluted siRNAs: RPS6KB1, EIF4E2, EHMT1, HIF1AN, SOS1, TGFBR2, BRCA2, XPA, PMS1 and PMS2.

3.3. Discussion

Presented in this chapter are the reasons that motivated the choice of the isogenic model, which drives most of the following work on ERCC1-deficiency in NSCLC, as well as the results of high-throughput drug screens and siRNA screens.

Although tumour heterogeneity may be better represented by non-isogenic models, isogenic models are the gold-standard for studying in a "pure" fashion genetic dependencies and investigating synthetic lethal relationships (Rehman *et al.*, 2010). Only three patients have been described with ERCC1-deficiency so far (Jaspers *et al.* 2007, Kashiyama *et al.,* 2013), and fibroblasts derived from these patients were not available for our study. Moreover, patient-derived fibroblasts, even when immortalized by SV40, are often difficult to maintain in culture or manipulate genetically (e.g. by siRNA transfection), and results obtained with such model often diverge from those obtained with cancer cells.

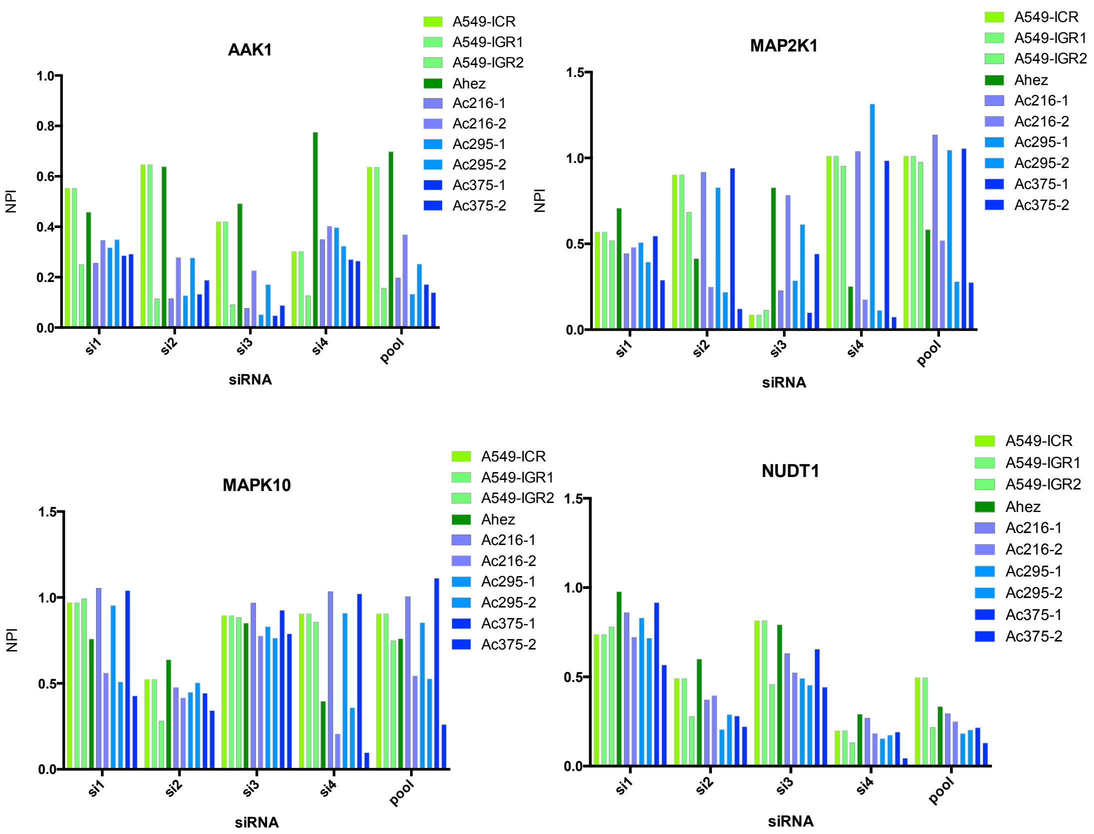


Figure 3. 20. Validation of ERCC1-low siRNA screen hits from the siRNA screen.

siRNA sensitivity assay repeated in two indendent experiments, performed in triplicate each, with four distinct siRNA targeting the same gene. Genes were selected based on their >2 difference in median Z-score between the ERCC1-proficient and ERCC1-deficient populations, the absence of toxicity in the ERCC1-proficient cells and the consistency of the results among ERCC1-deficient clones. Genes were considered as validating when > 20% decrease in cell viability was observed with the SMARTpool® as well as at least two of the four individual siRNAs. Displayed are results of the NPI analysis with the four individual siRNAs per gene and the SMARTpool®. One extra screen was performed with the ICR A549 cell line to minimize the risk of artefacts resulting from the use of a single control cell line in the ERCC1-proficient group.

A

Gene name	A549 Z-score	Ahez Z-score	Ac216 Z-score	Ac295 Z-score	Ac375 Z-score	median Z-score clones minus Z-score A549
HRPT2	-7.36162	-3.74919	-2.59840	0.82480	0.64720	**8.00883**
GUK1	-7.68825	-4.58372	-7.50730	0.45662	0.03640	**7.72465**
PRKAR2B	-6.44562	-2.72682	-2.71781	0.80372	0.57099	**7.01661**
EFNB3	-5.14324	-3.50975	0.63494	0.79482	0.69935	**5.84258**
TOPBP1	-4.82343	-3.82246	-7.31673	0.21551	0.24153	**5.03894**
MAP2K4	-4.91613	-0.37521	0.04858	0.25419	-0.42606	**4.96471**
MGC45428	-4.88170	-0.62305	-0.60288	-0.03169	0.08064	**4.85002**
PCK2	-3.05940	-1.80977	1.08317	0.77101	0.43200	**3.83041**
RAD50	-5.03881	-2.34992	-1.82576	-1.13082	-1.35998	**3.67883**
BUB1B	-3.03049	-2.44093	-5.56340	0.93137	0.64552	**3.67601**
MYLK2	-5.25582	-4.36203	-1.66336	-1.51989	-3.63973	**3.59246**
EXOSC10	-2.98136	-5.51819	-4.80671	0.36740	0.13439	**3.11574**
MEN1	-2.62013	-1.64355	-2.57760	0.58999	0.41715	**3.03728**
PRKCB1	-2.71865	-0.29118	-1.09231	0.57098	0.27556	**2.99421**
ERCC3	-2.84774	-0.51468	0.40952	0.01258	-0.15734	**2.86032**
PANK4	-1.87555	-2.02772	0.37308	0.98378	0.81123	**2.68678**
PRKAR1B	-2.88097	-1.51259	-1.21285	-0.09916	-0.20555	**2.67541**
ILK	-2.67687	-1.71133	-6.23262	0.49501	-0.03761	**2.63926**
PRPF4B	-2.25685	-3.44093	-5.32513	0.36128	0.37156	**2.61813**
CYLD	-1.93981	-2.75080	-3.26768	0.77378	0.66562	**2.60543**
FLJ21816	-2.23143	-1.77369	-1.39952	0.34642	0.34321	**2.57464**
FLJ35220	-2.19608	0.36103	0.62191	0.22812	0.32312	**2.51919**
MDC1	-2.93140	-0.63778	-1.95221	-0.45899	0.04216	**2.47242**
SHFM1	-4.27716	-1.82702	-5.26911	-1.62581	-1.81840	**2.45876**
FANCF	-1.85832	0.16922	0.80101	0.52020	-0.20850	**2.37852**
BRCA2	-2.26594	-0.68412	-3.47086	0.36455	0.06782	**2.33377**
HUNK	-2.63620	-3.54180	-0.39860	-3.66688	-0.13846	**2.23760**
BUB1B	-4.94025	-2.70497	-6.17783	-2.80195	-0.77511	**2.13831**
PGK2	-0.97785	0.20460	0.35131	1.44446	1.15911	**2.13696**
MAP2K4	-2.01127	-0.08527	-0.74820	0.81931	0.04942	**2.06069**
RPA1	-8.80079	-6.24910	-9.93405	-6.75084	-5.90718	**2.04994**
PMS1	-2.08219	0.32203	0.00005	-0.26562	-0.04055	**2.04164**
MAP3K5	-1.40653	1.13306	0.07738	0.82439	0.63138	**2.03792**
MAP2K5	-1.61468	0.55852	-0.30367	0.66473	0.41967	**2.03435**
GNE	-1.09239	0.87959	0.25699	1.23049	0.92439	**2.01678**
FER	-1.59155	0.72594	-0.46793	0.43276	0.42390	**2.01545**
TREX2	-1.85008	-2.08733	-0.69738	0.23558	0.15450	**2.00458**

B

Gene name	A549-FR Z-score	Ac216 Z-score	Ac375 Z-score	median Z-score clones minus median Z-score A549
TNKS2	-2.6222142	-0.616672	-0.3607324	**2.133511971**
SIRT1	-3.9452985	-1.6356454	-1.6407744	**2.307088607**
PARP10	-2.1553634	-0.0228371	0.40227331	**2.345081483**
SIRT1	-3.6560135	-1.3717994	-1.1140661	**2.413080755**
TNKS2	-2.8258575	-0.477241	-0.2490914	**2.46269132**
PARP9	-3.4011722	-1.0170329	-0.8254437	**2.479933864**
PARP16	-5.7505024	-1.9345794	-1.7579876	**3.904218959**
PARP16	-7.624049	-2.5618869	-2.1377016	**5.274254693**
PARP1	-8.4499518	-0.5528146	-0.0269272	**8.160080901**
PARP1	-9.0346646	-1.0204973	0.1414234	**8.595127639**

Figure 3. 21. Hits from the Kinome – Tumour Suppressor – DNA repair and PARP plates showing selectivity for the ERCC1-proficient population (legend below)

C

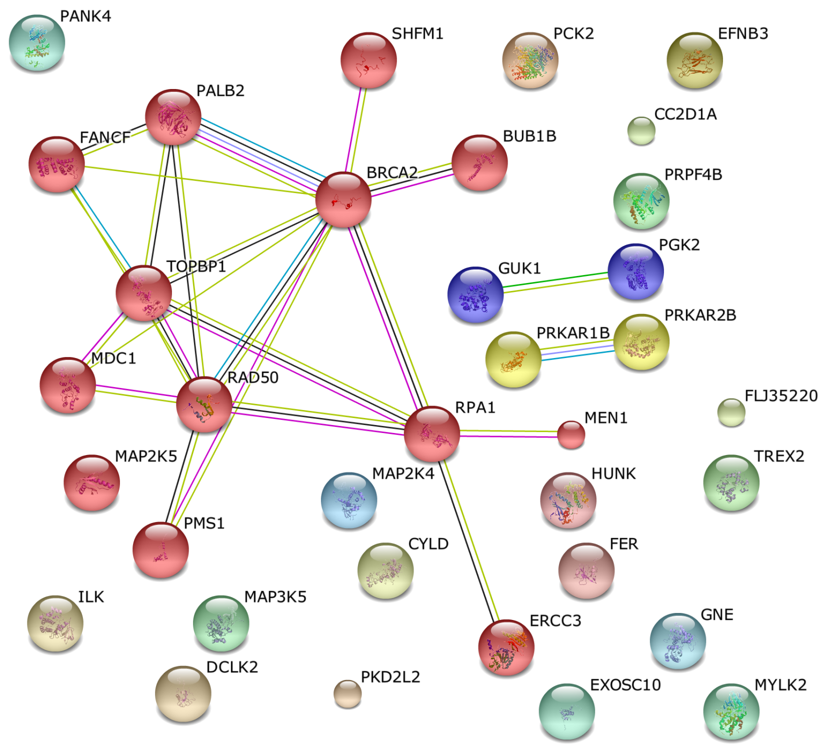


Figure 3. 21. Hits from the Kinome – Tumour Suppressor – DNA repair and PARP plates showing selectivity for the ERCC1-proficient population

A) List of hits from the Kinome – Tumour Suppressor – DNA repair plates siRNA screen in difference Z score order, with siRNA target gene, sensitivity Z score for each cell line, and the difference in median Z-scores between each population. B) List of hits from the Kinome – Tumour Suppressor – DNA repair plates siRNA screen in difference Z score order, with siRNA target gene, sensitivity Z score for each cell line, and the difference in median Z-scores between each population (each siRNA was aliquoted twice in this plate, which explains that some siRNA replicates came up consistently). C) Visual representation of the hits from Metabolic - Phosphatase plates of the siRNA screen using STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) v9.1 database, to help identifying functional networks within the hit list. Hits were further separated into clusters following a Markov Cluster Algorithm of 1.

A

Gene name	A549-FR Z-score	Ac216 Z-score	Ac295 Z-scores	Ac375 Z-score	median Z-score clones minus median Z-score A549
PPP2R2A	-16.809543	-2.9391265	-2.2536868	-2.1822251	**14.55585616**
PRKAG3	-21.028331	-6.5344297	-19.525966	-5.077673	**14.49390119**
CFLAR	-17.373858	-4.1501549	-3.9927378	-2.8733448	**13.38112062**
AKAP10	-14.482688	-1.6810414	-1.3259827	-2.966428	**12.80164666**
SLC7A5	-12.567161	-0.0983801	0.87006481	-1.2526378	**12.46878118**
RRM2	-17.891636	-5.7174712	-21.297139	-4.7842281	**12.17416466**
TYMS	-17.707813	-5.6395454	-21.158629	-4.6525379	**12.06826792**
EIF4E2	-12.838056	-1.38034	0.55534456	-1.0382227	**11.79983287**
RPS6	-15.46595	-3.792872	-14.683782	-3.025768	**11.67307837**
AMD1	-15.158557	-4.401888	-2.8755656	-3.6714511	**11.48710545**
NAGK	-14.841292	-3.5201464	-6.8335701	-3.2943993	**11.32114538**
PYCS	-12.759069	-2.3060823	-3.1152358	-0.9136954	**10.45298683**
CMAS	-10.997329	-1.1239672	-2.1080939	-0.7899972	**9.873361467**
CREB3L1	-12.984772	-3.3544567	-5.1176843	-3.2683651	**9.630315514**
OGDH	-10.08125	-0.5011254	-8.5584038	-1.2012597	**8.879990571**
ODC1	-12.414465	-3.8864337	-11.297799	-3.0316936	**8.528030828**
EIF4G1	-12.613434	-4.2090013	-7.8398538	-3.3155377	**8.404432589**
SOS1	-10.055812	-1.7062554	-2.4029065	-1.6121638	**8.349556922**
FLJ20604	-10.343048	-2.0085536	-5.9428973	-1.7741414	**8.334494496**
ERN1	-9.0720315	-1.6967575	0.25664562	-0.7976572	**8.274374244**
PMAIP1	-9.3614058	-1.1183118	-0.3096717	-1.4187277	**8.243094003**
UMPS	-11.018033	-2.8818527	-9.2667731	-2.827125	**8.136180231**
DYNC1H1	-11.01096	-2.9296156	-9.8992453	-2.3382975	**8.081343978**
PR48	-11.766317	-3.7100386	-8.7428236	-3.3013116	**8.056278244**
HIF1AN	-9.5731924	-1.2388303	-2.1387893	-1.5417688	**8.03142363**
GPI	-9.3457788	-1.5691235	-0.2083011	-1.3639759	**7.981802925**
NOS3	-7.6847323	0.26388655	0.28936406	0.15324832	**7.948618825**
PRKCZ	-11.935417	-3.0965853	-7.336737	-4.0140172	**7.921399519**
CDC2	-10.50861	-2.8950363	-7.3801376	-2.7740697	**7.613573651**
EXOC7	-9.7381847	-2.2121115	-3.5616536	-1.3593082	**7.526073223**
PPAT	-9.3778466	-1.8893722	-2.8781511	-1.7071047	**7.48847438**
EIF4E	-10.367816	-2.9363131	-8.2530408	-2.1825722	**7.43150319**
PRKAR1B	-7.9632656	-0.6687733	-0.6567363	-0.3408613	**7.306529293**
PRKAR2B	-10.176744	-2.9164781	-5.2802355	-2.9559882	**7.220756134**
SLC22A5	-6.602092	0.43164328	0.78431403	0.60723005	**7.209322083**
SNX9	-11.463418	-4.3149846	-2.2533186	-4.3795458	**7.1484338**
CAD	-9.5478439	-2.4346433	-7.1179765	-2.1157869	**7.113200607**
EPHA3	-7.8875201	-0.8637689	-0.0175062	-0.8510092	**7.036510905**
COX5A	-7.2788333	-0.4008746	-0.1303991	-0.5369828	**6.877958762**
DHFR	-8.697589	-1.8815472	-3.0710632	-2.2833532	**6.414235891**
GNPDA1	-11.368908	-5.2474637	-13.431262	-4.2980197	**6.121444596**
PPAP2A	-8.5811309	-2.5901942	-2.9794246	-2.514426	**5.990936727**
PISD	-11.377379	-5.4503641	-6.6011015	-5.016823	**5.927015077**
FLJ23751	-8.0709933	-2.4025784	-0.7714713	-2.1932678	**5.877725505**
ACO1	-6.6880931	-1.1489865	-4.5371013	-0.7927451	**5.539106577**
KIAA0377	-7.0435851	-2.7434291	-1.600524	-1.4024633	**5.443061087**
EIF4G2	-8.3499837	-3.1044194	-6.6342148	-1.3593342	**5.245564337**
MTMR4	-6.0778847	-1.0280398	0.33758919	-0.8423619	**5.235522769**
PIGK	-8.2745539	-3.0553841	-3.5784139	-2.9564861	**5.219169779**
HDC	-5.8105616	-1.3929132	0.68561256	-0.6962718	**5.114289737**
NDUFAB1	-5.3753705	-0.2857931	-27.648878	-0.084263	**5.089577423**
PSPH	-5.3540495	0.09576601	-1.6852189	-0.3468357	**5.007213772**
ENO2	-4.039685	0.85761789	1.44984195	0.61292902	**4.897302884**
WBSCR22	-7.2687756	-2.4383504	-8.2332301	-2.0612473	**4.830425126**
CD81	-6.421161	-1.6552385	-0.74077	-1.6541562	**4.767004821**
ELF3	-5.2130646	-0.7636757	0.09183566	-0.4794186	**4.733646003**
PPOX	-8.2923987	-3.86723	-3.5809914	-3.5788512	**4.71140738**

Figure 3. 22. Hits from the Metabolic – Phosphatase, and PARP plates showing selectivity for the ERCC1-deficient population (legend below)

A (continued)

Gene name	A549-FR Z-score	Ac216 Z-score	Ac295 Z-scores	Ac375 Z-score	median Z-score clones minus median Z-score A549
UGDH	-4.4140256	0.33177665	0.29209949	-0.04899	**4.706125115**
FRAP1	-6.6541382	-1.9988412	-2.486667	-1.6942812	**4.655297046**
SLC25A14	-10.085782	-5.4996318	-12.207495	-5.2820416	**4.586150169**
MAT2A	-8.0703504	-3.5142427	-16.690364	-2.4304843	**4.556107699**
PPAP2B	-9.6591899	-5.130459	-7.4992323	-4.381096	**4.52873091**
FLJ20522	-7.1279635	-2.4047982	-6.5374986	-2.7324624	**4.395501105**
SUCLG2	-4.4205289	-0.0379561	0.43892634	-0.178449	**4.38257283**
MTMR2	-6.3367775	-1.9035836	-3.0148864	-2.2780512	**4.05872634**
MTMR6	-3.8145228	0.06901829	0.54163233	0.24402946	**4.05855223**
GART	-6.5090141	-2.4809827	-11.128955	-1.3210325	**4.028031442**
PSPH	-4.0311237	0.08096939	-1.6852189	-0.0205358	**4.010587914**
GARS	-5.9689685	-1.8502397	-3.5790392	-1.9951859	**3.973782576**
BCKDHA	-7.7389016	-3.7747138	-10.856229	-2.5477855	**3.964187723**
MAOA	-4.4424818	-0.4130494	-0.4985609	-1.013743	**3.943920865**
CA12	-3.4504684	0.52813056	-1.4034854	0.42890781	**3.879376251**
UGT2B17	-3.2398133	0.53867723	1.22403947	0.53953521	**3.779348557**
SREBF1	-5.0466387	-1.2931688	-1.3674098	-0.5755177	**3.75346982**
PRPS1L1	-3.6489702	0.23186024	-0.3493331	0.02518687	**3.674157068**
ALDRL6	-6.1798083	-2.5286199	-6.5983634	-1.783607	**3.651188377**
NT5M	-4.9624232	-1.8178805	-1.403968	-1.2194926	**3.558455215**
CKMT1B	-5.8451219	-2.3551004	-3.495396	-1.4758935	**3.490021447**
SLC25A3	-3.5394779	0.45539732	-0.7678601	-0.1401663	**3.399311631**
MTCH2	-3.9012531	-0.9399365	0.71315101	-0.5739397	**3.32731341**
MUT	-4.1855352	-1.0889347	1.17724376	-0.868163	**3.317372196**
ETFB	-3.8626193	-0.5667479	0.18645331	-0.6778265	**3.295871441**
PRKY	-5.1975993	-1.9435191	-3.1175864	-1.5301467	**3.254080243**
ALDH1A3	-5.6659154	-2.3342168	-6.4284697	-2.4383533	**3.227562106**
EHMT1	-4.0138952	-1.0588925	1.19154388	-0.8446761	**3.169219168**
MGC5601	-6.026427	-2.9054032	-6.1428103	-2.7386866	**3.121023787**
ACO2	-4.1287818	-1.0099709	-2.5265209	-0.4645609	**3.118810832**
PPP1R12C	-4.1157274	-1.1733994	-0.1498353	-1.1371709	**2.978556461**
CBLC	-3.2423871	-0.3401309	0.33827527	-0.3325793	**2.909807734**
ARG2	-4.9436347	-1.7350601	-4.385606	-2.1531678	**2.790466818**
PYCR1	-4.3570564	-1.6371621	-2.4480025	-1.3312861	**2.719894362**
ACLY	-2.0381091	0.77459982	-0.3222216	0.64340948	**2.68151857**
PCCB	-2.864106	-0.3937053	0.66089016	-0.1967081	**2.667397926**
COQ3	-6.0105571	-3.3911275	-9.4056726	-2.5268155	**2.619429639**
GRB2	-3.821467	-1.2177501	1.27232168	-1.478863	**2.603716843**
PAICS	-3.1733903	-0.5803289	-1.3739604	-0.5417227	**2.593061395**
ARG1	-3.3626173	-0.8329036	-7.5224471	-0.6996483	**2.529713654**
RPS6KB1	-2.890913	-0.5562391	1.02377102	-0.4850993	**2.405813725**
HRAS	-2.6865364	-0.7669139	-0.2846962	-0.0503191	**2.401840239**
CLDN5	-6.065109	-3.6675238	-4.5304217	-1.8623766	**2.397585161**
PFKP	-2.9276805	-1.5439158	-0.4585271	-0.5520275	**2.375653065**
PRPS1	-3.1308234	-0.2183932	-0.756942	-0.9098288	**2.37388134**
SLC25A12	-3.4907286	-2.2622298	-1.1347042	-1.0329021	**2.356024386**
IDH2	-1.8602296	-0.1656309	0.74739932	0.47803011	**2.338259689**
ATP5F1	-4.4137898	-1.9738487	-12.0826	-2.1020218	**2.31176798**
PPP1CB	-3.2474141	-1.9469801	-0.7825484	-0.9820344	**2.265379675**
INSR	-3.3225441	-1.0667686	-5.4460596	-1.0416286	**2.255775496**
PYCR2	-3.621368	-1.9244449	0.0051972	-1.3964663	**2.224901755**
NME7	-1.6851613	0.33483122	1.40792427	0.5063486	**2.19150988**
SLC25A22	-1.7521201	-0.0357766	0.58751392	0.39346967	**2.145589799**
C9ORF12	-4.6238542	-2.2776174	-4.4502951	-2.4802939	**2.143560269**
CHKA	-3.033511	-0.7424556	-4.6386216	-0.9046701	**2.128840877**
HPRT1	-4.4691113	-2.3951299	-13.317132	-2.2680746	**2.073981373**
PTPDC1	-2.7878086	-0.7203634	0.46561473	-1.2018212	**2.067445243**
SLC27A3	-1.627689	0.4573962	0.15822337	0.4079634	**2.035652358**

Figure 3. 22. Hits from the Metabolic – Phosphatase, and PARP plates showing selectivity for the ERCC1-deficient population (legend below)

B

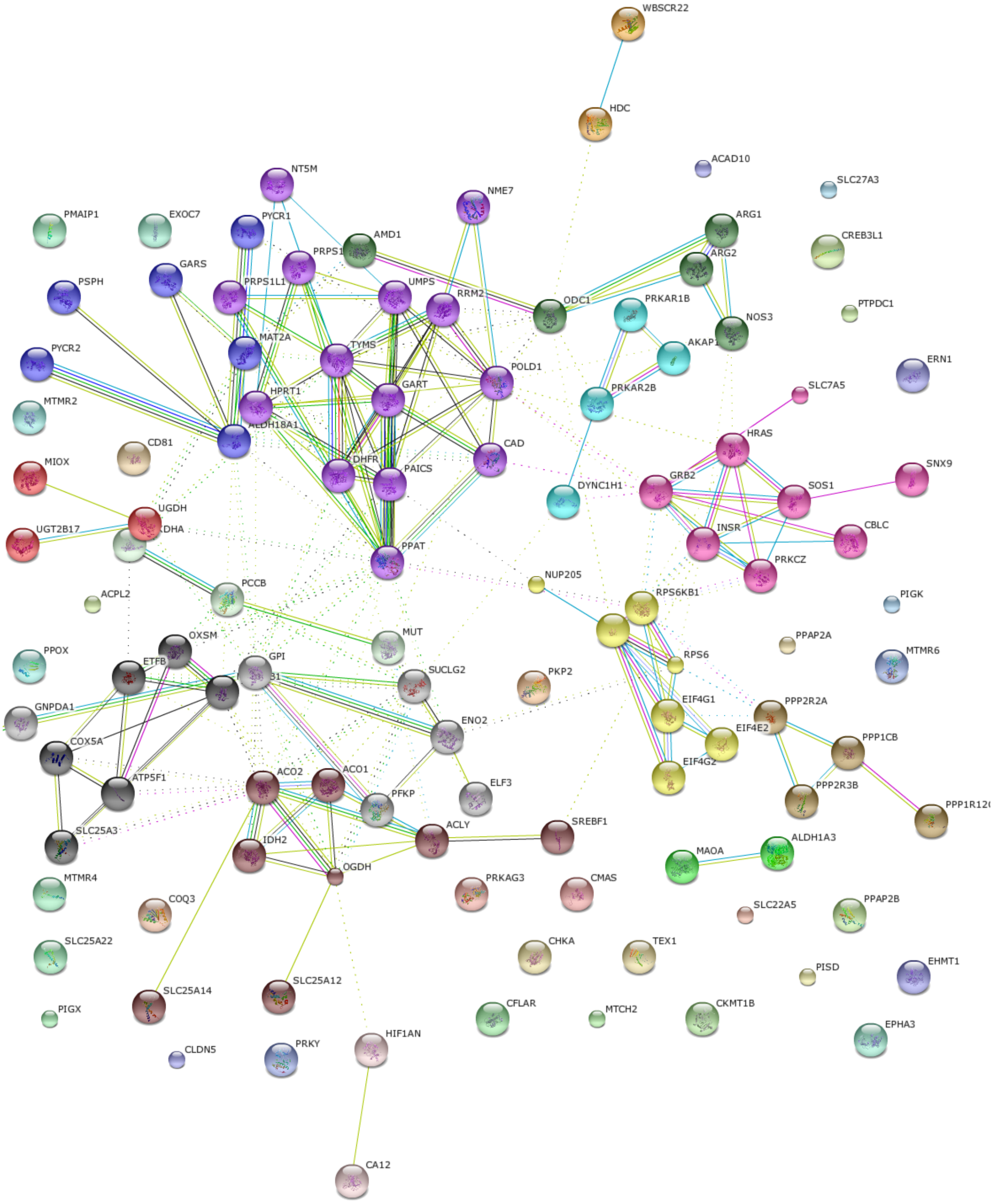


Figure 3. 22. Hits from the Metabolic – Phosphatase, and PARP plates showing selectivity for the ERCC1-deficient population

A) List of hits from siRNA screen in difference Z score order, with siRNA target gene, sensitivity Z score for each cell line, and the difference in median Z-scores between each population. B) Visual representation of the hits from Metabolic - Phosphatase plates of the siRNA screen using STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) v9.1 database, to help identifying functional networks within the hit list. Hits were further separated into clusters following a Markov Cluster Algorithm of 2.

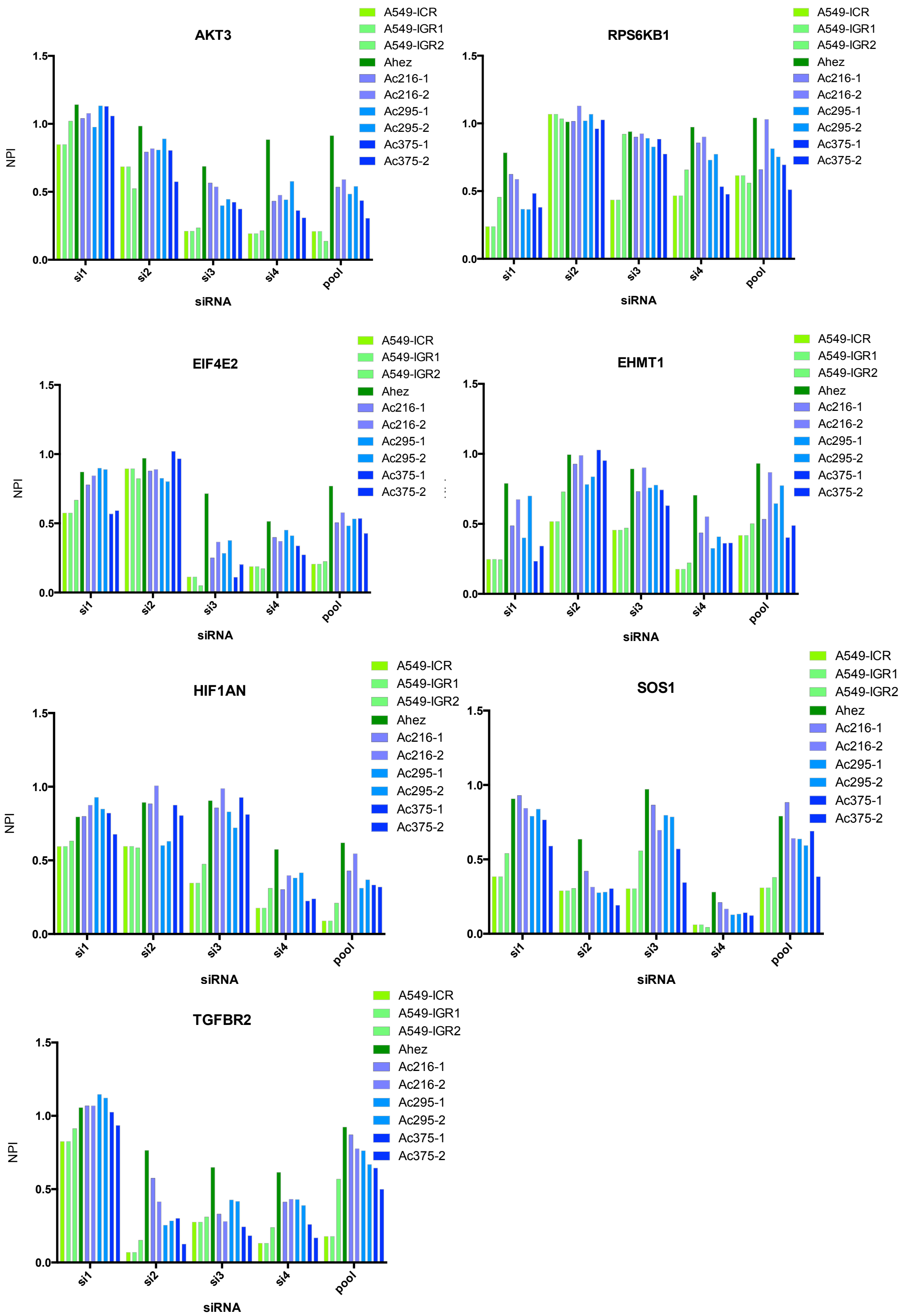


Figure 3. 23. Revalidation of ERCC1-high siRNA screen hits from the siRNA screen (legend below)

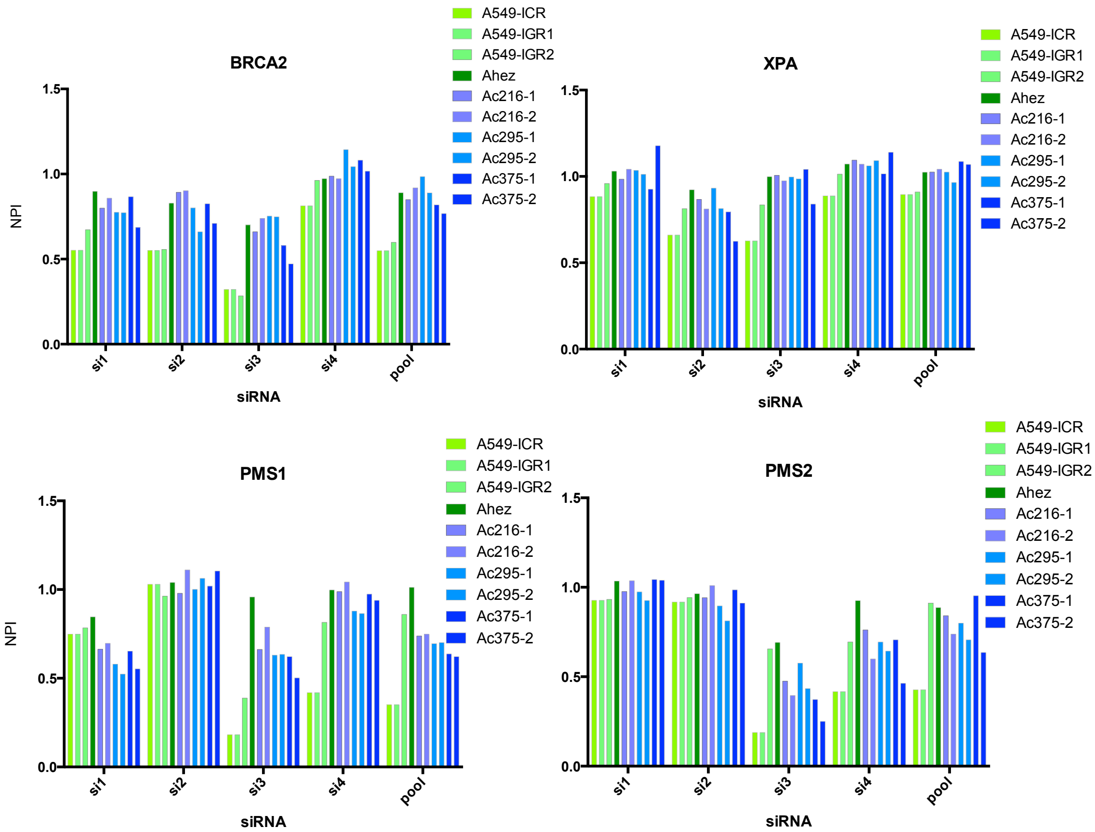


Figure 3. 23. Revalidation of ERCC1-high siRNA screen hits from the siRNA screen.

siRNA sensitivity assay repeated in two independent experiments, performed in triplicate each, with four distinct siRNA targeting the same gene. Genes were selected based on their > 2 difference in median Z-score between the ERCC1-proficient and ERCC1-deficient populations, the absence of toxicity in the ERCC1-proficient cells and the consistency of the results among ERCC1-deficient clones. Genes were considered as validating when > 20% decrease in cell viability was observed with the SMARTpool® as well as at least two of the four individual siRNAs. Displayed are results of the NPI analysis with the four individual siRNAs per gene and the SMARTpool®. One extra screen was performed with the ICR A549 cell line to minimize the risk of artefacts resulting from the use of a single control cell line in the ERCC1-proficient group.

The only available cancer cell model of total ERCC1-deficiency at the beginning of this PhD was the CHO-derived match-pair of AA8 and UV20 cells, but these represented only a single model and had limited translatability to human lung cancer cells. It was therefore crucial to create an appropriate NSCLC model of ERCC1 deficiency, and the generation of the ERCC1-isogenic NSCLC model represented a fundamental step of this work. However, as an intrinsic weakness of this model was the comparison of several ERCC1-deficient clones to a single control ERCC1-proficient parental cell line, the use of an independent ERCC1-proficient A549 cell line, maintained in a different institution, growing in a different media and presenting a different number of passages, allowed us to limit this pitfall by the addition a second isogenic control.

In isogenic models, several clones deficient for the alteration of interest are usually required, to account for clone-to-clone variability. Although we used three clones, which were generated independently for the loss of their last *ERCC1* copy, they all derived from the same *ERCC1*-heterozygous cell line; this point is important to keep in mind as the three clones consequently share a common ancestor and cannot be considered as completely independent. Therefore, effects that were observed in the three ERCC1-deficient clones but not in the heterozygous cell line were deemed to be almost certainly ERCC1-related, whereas effects observed both in the ERCC1-deficient clones and in the heterozygous cell line were deemed to be potentially ERCC1-related; indeed, these latter effects could either be ERCC1-related but observed under a lower threshold of ERCC1-deficiency, or be ERCC1-independent and related to another mutation that had occurred during the generation process of the heterozygous cell line.

Unfortunately, the non-isogenic model constituted of 15 NSCLC cell lines could not meaningfully be exploited for the purpose of this work as: (i) ERCC1 is present as four distinct isoforms, which show a high degree of homology and cannot easily be distinguished with currently available techniques (western blot, RT-PCR, ddPCR, FISH); only isoform 202 has proven to be functional with regards to ERCC1 DNA repair activity, and overexpression of another isoform could therefore lead to misclassification; and (ii) ERCC1-isoform 202 relative expression, measured by western blot, could not be correlated to platinum sensitivity in this model. The non-isogenic model was therefore considered as not being a sufficiently relevant reflection of the clinical reality and as having limited translatability at this stage.

Further experiments would include the evaluation of a putative induction of *ERCC1* expression after cisplatin exposure, and the correlation between platinum sensitivity and the ERCC1/XPF heterodimer expression, assessed by Duolink® (PLA®, Sigma-Aldrich).

Analysis of the drug screen revealed three interesting trends. First, out of 24 molecules that showed differential effects between ERCC1-proficient and ERCC1-deficient cells, six of them were PARP inhibitors of different structures, chemical families and potencies (Figure 3. 24). This differential effect was consistently revalidated in all assay formats as well as with several molecules, reinforcing the robustness of these findings. No previous synthetic lethal relationship had been reported between NER and BER deficiency. However, recent studies have provided insight into the mechanism of action of PARP inhibitors and show that PARP inhibitor selectivity may not be limited to HR-deficient cells (Zhang *et al.*, 2011; Kedar *et al.*, 2012). An independent study performed at the same time as this thesis reported in parallel that PARP inhibitors could synergize with cisplatin in ERCC1-low NSCLC cells and not in ERCC1-high cells, suggesting that ERCC1-deficiency could at least partly drive both platinum and PARP inhibitor sensitivity (Cheng *et al.*, 2013). Further analysis and mechanistic insight regarding this ERCC1-deficient cells - PARP inhibitors synthetic lethal relationship will be provided in Chapter 4.

A second major group of hits arising from the drug screen was constituted of several molecules inhibiting the PI3K-AKT-mTOR pathway, which also showed consistent differential effects towards ERCC1-deficient cells. These ERCC1-selective effects could not be consistently revalidated in short-term assays, but some degree of revalidation was obtained in long-term assays. Interestingly, two of the siRNA screen hits that could be revalidated with at least two of the deconvoluted oligos involved MAP Kinases (MAP2K1, MAPK10), which are also part of pivotal cell proliferation pathways and are involved in cross-talk with the PI3K-AKT-mTOR pathway. Interactions between DNA repair and other hallmarks of cancer, such as uncontrolled proliferation and loss of dependency towards growth factors (Hanahan *et al.*, 2011) are just starting to be investigated and are a matter of growing interest (Bassi *et al.*, 2013; Qu *et al.*, 2013). Whether the activation of a pro-proliferating pathway could counterbalance enhanced cell death resulting from a DNA repair deficiency, and whether ERCC1-deficient cells could become addicted to such pathway is a potential hypothesis.

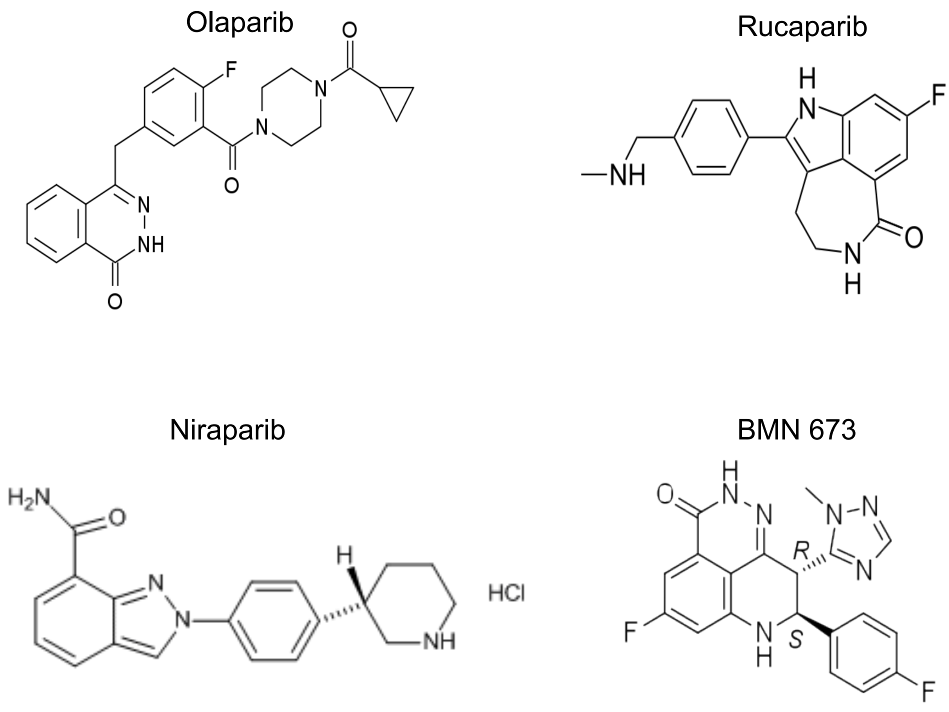


Figure 3. 24. Chemical structure of the four PARP inhibitors used in the experiments and currently in clinical development

Third, ERCC1-deficient cells appear to be consistently more sensitive to CHK1 inhibitors, an effect which could be revalidated in multiple short- and long-term assays (Figure 3. 12, Figure 3. 13). CHK1 is a serine-threonine kinase that phosphorylates the CDC25 phosphatase, thereby preventing cells from entering mitosis (Bartek *et al.*, 2003; Niida *et al.*, 2005). Consequently, the increased sensitivity of ERCC1-deficient cells to CHK1 inhibitors is in line with the mitotic defects observed in this population (Friboulet *et al.,* 2013b*)*: one hypothesis is that CHK1 inhibition may accelerate cell death in ERCC1-deficient cells either by increasing genomic instability (secondary to unrepaired lesions over multiple cell cycles), or by causing mitotic catastrophe (Niida *et al.*, 2005). Interestingly, results of the high-throughput siRNA screen showed that inhibition of *CHEK1* by siRNA also caused increased cell death in the ERCC1-deficient population. However, this effect was inconsistent in revalidation, with paradoxical and unpredictable effects of *CHEK1* silencing in asynchronous cells. ERCC1-deficient cells also showed a trend to decreased CHK1 expression as measured by western blot (data not shown), but – similarly to the effects observed with siRNAs - variability among experiments in asynchronous cells were observed. However, these observations should be kept in mind, notably considering the current development of several potent and selective CHK1 inhibitors (such as LY2603618, LY2606368 or SCH 900776; source: clinicaltrial.gov) and the actual absence of predictive selection biomarker for these drugs. The possibility of using CHK1 inhibitors in monotherapy or combination for patients with ERCC1-deficient tumours might therefore deserve further investigation.

Several hits were also identified through high-throughput siRNA screens that evaluated the effect of 2060 distinct oligos. A key step in an RNAi screen is the validation of the results to confirm the phenotype observed is due to specific inhibition of the target gene and not due to off-target effects of the siRNA. Although siRNA were initially thought to be highly specific, it is now largely accepted that siRNA also silence genes other than the intended target through off-target effects in a less sequence-specific manner, in a similar fashion to the microRNA machinery (Jackson *et al.*, 2003; Birmingham *et al.*, 2006). Minor chemical modifications which limit the incidence of off-target effects have been introduced (Jackson *et al.*, 2006), but the demonstration that the suppression of the targeted gene is responsible for the observed phenotype remains crucial and mandatory. Two major ways can be used to perform such demonstration: “rescue” or “redundancy” (Echeverri *et al.*, 2006). Rescue involves reversing the phenotype induced by the siRNA by re-expressing an siRNA-resistant cDNA (i.e. one that has mismatches with the siRNA and cannot be targeted for degradation). With the large number of hits that result from RNAi screens, this method is obviously not feasible in routine laboratory practice; moreover, over-expression of a cDNA in ectopic localizations leads to aberrant effects. Redundancy was therefore the method of choice to confirm specificity in the context of this PhD. This involves demonstrating that multiple siRNA species targeting different regions of the same target mRNA (i.e. deconvoluted siRNAs) cause the same phenotype. It is generally agreed that if two or more different siRNA species targeting the same gene generate the observed phenotype, the effect is likely to be on-target (Echeverri *et al*., 2006).

The proportion of hits that were revalidated in our series is in line with what has previously been obtained in similar experiments in the Ashworth lab and elsewhere. Among four hits that consistently revalidated, MAP2K1 and MAPK10 were of particular interest given the consistency with the sensitivity notably to MEK and AKT inhibitors observed in the drug screen, as discussed above. Moreover, NUDT1 represented another particularly interesting hit: *NUDT1* encodes for the 7,8-dihydro-8-oxoguanine triphosphatase (also called *MTH1*, MutT Homolog 1), an enzyme that prevents incorporation of oxidized nucleotides into DNA/RNA during replication and transcription by hydrolysing oxidized purine nucleoside triphosphates - such as 8-oxo-dGTP, 8-oxo-dATP, 2-hydroxy-dATP, and 2-hydroxy rATP - to monophosphates (Sakumi *et al.*, 1993; Furuichi *et al.*, 1994). Contrary to some other oxidized bases, 8-oxo-Guanine can be incorporated into nucleic acids without blocking their synthesis, where it induces base mispairing that mostly result in A:T>C:G or G:C>T:A transversions (Takagi *et al.*, 2012). Therefore, MTH1 is responsible for ensuring accurate DNA replication as well as preventing erroneous protein synthesis under oxidative stress. Whether ERCC1 would also protect from oxidative stress, for example by playing a role either in the removal of incorporated oxidized nucleotides, or whether ERCC1-deficient cells produce more reactive oxygen species than their ERCC1-proficient counterpart, remains to be studied. Such role for ERCC1 would render ERCC1-deficient cells more reliant on MTH1 for preventing oxidative stress, and perhaps explain the selective sensitivity of ERCC1-deficient cells to *NUDT1* silencing.

Among the ERCC1-high siRNA screen hits that could be revalidated, three main groups could be identified and were of particular interest. First, a DNA repair group constituted of the BRCA2, XPA, PMS1 and PMS2 proteins. Although the increased sensitivity of the ERCC1-high population to XPA silencing could be considered as a positive control (as XPA interacts with ERCC1 as a heterodimer in the repair of DNA lesions through NER), the increased sensitivity of ERCC1-proficient cells to BRCA2 silencing was unexpected. BRCA2 plays a crucial role in DSB repair (and notably HR) but several additional DNA repair roles have been described for this protein (Jasin, 2002; Tutt *et al.*, 2005). ERCC1 has also been involved in some forms of DSB repair, and the decreased sensitivity of ERCC1-low cells to BRCA2 silencing might suggest that ERCC1-deficient cells do not rely as much as ERCC1-proficient cells on HR for their survival, and that they can compensate their inherent DNA repair defect by other mechanisms to ensure survival (such as dependency on alternative pathways or metabolic changes). Also, ERCC1-proficient cells were more sensitive to the silencing of PMS1 and PMS2, which are two key proteins of the MMR pathway; this differs with preliminary results arising from high-throughput synthetic lethal screens performed in yeast (BioGRID, thebiogrid.org), where some MMR proteins showed synthetic lethal interactions with RAD10, the yeast ERCC1 equivalent, and highlights the limited translatability of results obtained in yeast to human cancer cells.

The second main group identified among ERCC1-high hits were several proteins involved in the PI3K-AKT-mTOR pathway, namely AKT3, RPS6KB1 and EIF4E2, suggesting that ERCC1-proficient cells might be more dependent on this pathway for their survival. Paradoxically, AKT3 was initially identified as an ERCC1-low hit, and results of the drug screen showed increased sensitivity of ERCC1-deficient cells to several PI3K or mTOR inhibitors. These contradictory results may reflect the complexity of this pathway, notably in terms of regulatory feedback loops, cross-talks, multiplicity of protein isoforms or subunits harbouring distinct roles (such as PI3Kα, β, γ, δ, or ε), redundancy and salvage pathways (Rodon *et al.*, 2013). Whether compounds targeting specific elements of this pathway could have any clinical applicability, either for the ERCC1-low or the ERCC1-high population remains to be determined, but a deeper and thorough understanding of the role of each protein in the pathway itself as well as of the potential interactions with DNA repair mechanisms will be required.

The last group of ERCC1-high hits constituted TGFBR2 and HIF1AN, which were grouped because of their potential for interacting with the extra-cellular micro-environment (Kaklamani *et al.*, 2004; Lisy *et al.*, 2008). Considering the potential for interactions between several hallmarks of cancer, such as DNA repair defects and microenvironment changes, these actionable hits should be further investigated.

Specific drugs targeting either TGFBR2 or the pro-angiogenic pathways were not evaluated (with the exception of sunitinib that was present in the drug screen library but showed no significant difference between both populations), but may constitute a first step for investigating the relevance of these hits (Hu *et al.*, 2013).

Finally, it did not escape our attention that the siRNA targeting *PARP1* appeared twice as top hit in the ERCC1-high population. At first sight, this might seem to be contradictory with the PARP1/2 inhibitor sensitivity observed in the ERCC1-deficient population observed in the drug screen. However, inhibiting *PARP1* expression by siRNA might have very different consequences from inhibiting its catalytic activity via a small molecule inhibitor. It is possible that different modes of PARP1 inhibition elicit very different phenotypes.

In summary, several potential synthetic lethal interactions with ERCC1-deficiency were identified through high-throughput techniques (siRNA or drug screens). In order to maximize the potential for getting a rapid clinical applicability and potential benefit for patients from the above described screen results, we decided to focus on (i) reproducible screen results that could be consistently and robustly revalidated; (ii) hits representing clinically actionable targets; and (iii) hits for which drugs were licensed or in clinical development in humans. PARP inhibitors were the best hits meeting all these selection criteria, and investigations that have been performed on the PARP – ERCC1 synthetic lethal interaction, as well as potential clinical applications, will be described in Chapter 4.

CHAPTER 4

PARP1/2 inhibitors as potential therapy for ERCC1-deficient NSCLC

4.1. Introduction

The poly (ADP-ribose) polymerase (PARP) superfamily of proteins comprises 18 different members, the roles of most of which are still being elucidated (Ame *et al.*, 2004; Vyas *et al.*, 2013). All PARPs use βNAD+ as their substrate to modify acceptor proteins with ADP-ribosylation (Pears *et al.*, 2012), where units of ADP-ribose are added onto acceptor proteins. ADP-ribosylation regulates several key cellular processes involved in stress response (including DNA damage (Lord *et al.*, 2008), apoptosis (Koh *et al.*, 2005), heat shock (Petesch *et al.*, 2008), cytoplasmic stress (Leung *et al.*, 2011), or in normal cell physiology (such as cell division (Chang *et al.*, 2004), chromatin structure regulation or transcription (Schreiber *et al.*, 2006). The most studied of the family is PARP1, an abundant nuclear protein that comprises three function domains: a DNA binding domain, an auto-modification domain and a catalytic domain (Slade *et al.*, 2011). PARP1 catalyses poly(ADP-ribosyl)ation of itself (auto-parsylation) and of other proteins, by forming poly (ADP-ribose) chains (pADPr) (Hassler *et al.*, 2012). This post-translational modification occurs within a few seconds in response to single strand breaks and is part of the SSB repair pathway, as a part of base excision repair (BER) (Haince *et al.*, 2008). Aside from its role in BER repair, alternative roles for PARP1 have been proposed (Rouleau *et al.*, 2010). PARP1 has for example been implicated in double-strand break repair, notably homologous recombination – in which the localisation of MRE11 and ATM are dependent on pADPr formation (Haince *et al.,* 2008) - and in non-homologous end joining (Wang *et al.*, 2006). PARP1 also has roles in regulating transcription that are related to its ability to alter chromatin structure or methylation patterns and to interact with multiple transcription factors (Kraus, 2008; Caiafa *et al.*, 2009) including members of the ETS family (Brenner *et al.*, 2012), and androgen receptor (Schiewer *et al.*, 2012).

The high-throughput drug screen described in Chapter 3 allowed the identification of multiple clinical PARP1/2 inhibitors such as olaparib (AZD-2281), niraparib (MK4827), or rucaparib (AG-014699 or PF-01367338) as being selective for ERCC1-

deficiency. These results were revalidated in multiple formats (short and long-term assays) and using several PARP inhibitors (Chapter 3). Given the potential for using PARP1/2 inhibitors in the clinical setting, further investigations were performed using two clinically relevant PARP1/2 inhibitors, namely olaparib (AZD-2281, Astra Zeneca) and niraparib (MK-4827, TesaroBio). Both compounds displayed significant selectivity towards the ERCC1-deficient clones, which were 10 to a 100 times more sensitive to the PARP1/2 inhibitors than their ERCC1-proficient counterpart (Supplementary table 2). Results of these investigations, including further revalidation in different models and mechanistic dissection of the observed synthetic lethality, are presented in this chapter.

4.2. Results

4.2.1. Revalidation of the PARP1/2 inhibitors selective effects using multiple cellular models

In addition to the original ERCC1-isogenic model, we also aimed at assessing the generality of our findings by silencing *ERCC1* by RNA interference. Although it had been feasible to generate ERCC1-deficient A549 clones by gene targeting, siRNA-mediated silencing of *ERCC1* in NSCLC models caused acute cytotoxicity (Figure 4. 1. A), precluding their use in siRNA experiments. However, we noted that Zhang and colleagues had previously silenced *ERCC1* in U2OS osteosarcoma cells with minimal cytotoxic effects (Zhang *et al.*, 2011). U2OS cells were reverse transfected in 6-well plates at a concentration of 200,000 cells / well. Twenty-four hours after transfection, cells were transferred to new 6-well plates at a minimal cell concentration (500 cells / well) to allow formation of individual colonies. Drug was added 48h after the initial transfection (i.e. 24h after the plate transfer), which corresponded to the time where effects of the siRNA silencing were deemed to be maximal. Cells were subsequently exposed to olaparib for 15 days in order to allow formation of visible colonies, with drug containing media being replaced three times a week. Using this system, we found that ERCC1 siRNA caused olaparib sensitivity compared to control transfected cells (Figure 4. 1. B, Supplementary Table 2).

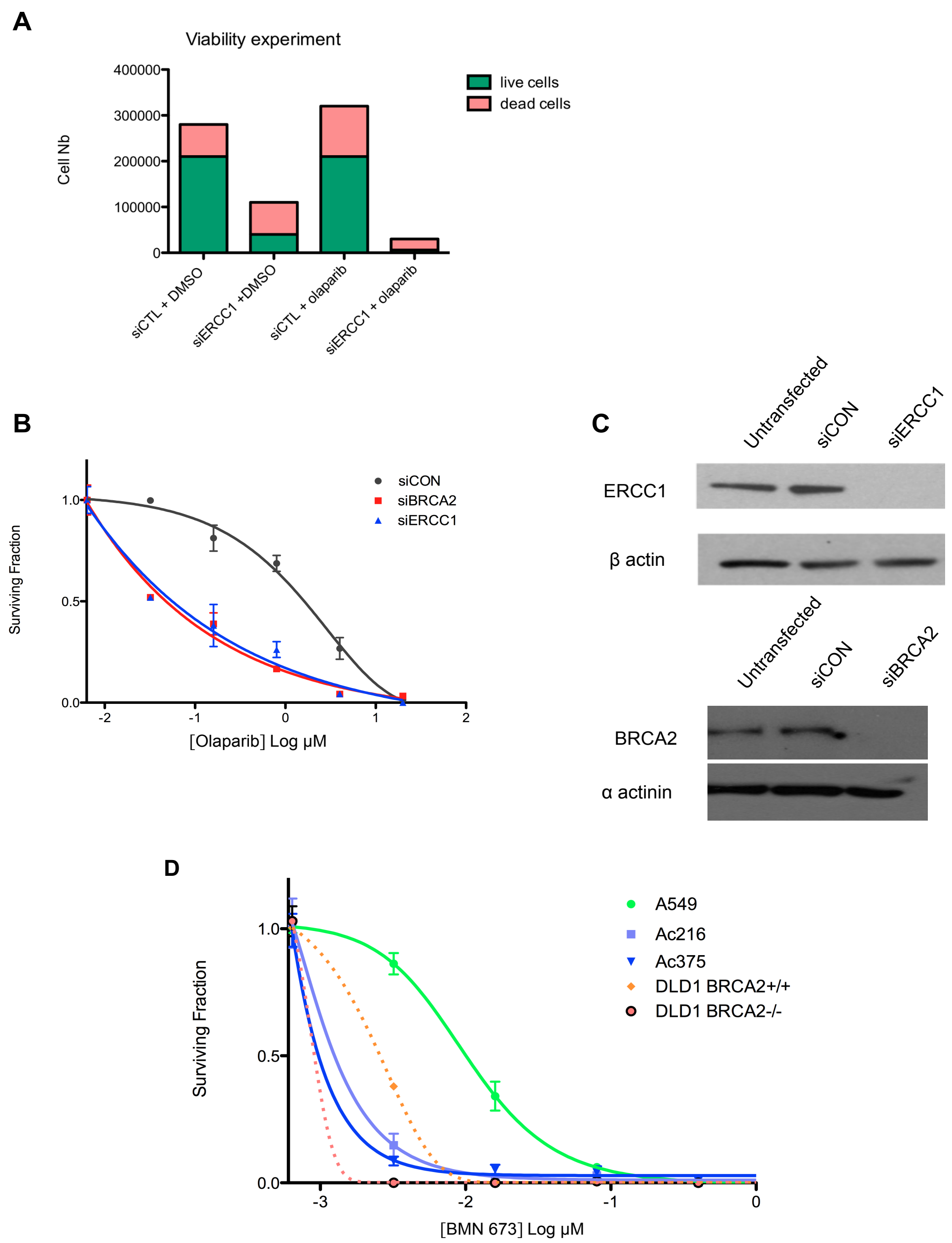


Figure 4. 1 - Revalidation of the drug screen hits using multiple cellular models and extension to a hyperpotent and highly selective PARP inhibitor

A) Viability experiment assessing cell toxicity of silencing *ERCC1* by siRNA in NSCLC cell lines (results of A549 cells are displayed as representative example). Viability was assessed by counting live cells at 1 week after reverse transfection. Acute toxicity of *ERCC1* silencing by siRNA in NSCLC cell lines precluded from using these models for further experiments and revalidation of the high throughput screens. B) Revalidation of olaparib sensitivity using siRNA silencing of *ERCC1* in U2OS cells, which had been used previously in similar experiments and better tolerated *ERCC1* silencing. Olaparib was added 48h after reverse-transfection and cells were exposed to the drug for five days. Error bars represent the standard deviation from the mean of three independent experiments. C) Western blot showing ERCC1 and BRCA2 silencing after siRNA transfection in U2OS cells. D) Clonogenic survival experiment evaluating BMN 673 sensitivity of ERCC1-deficient clones. Error bars represent the standard deviation from the mean of three independent experiments.

Strikingly, the effect of ERCC1 siRNA was comparable to the effect of BRCA2 siRNA that was transfected concomitantly as positive control. Validation of appropriate silencing by ERCC1 and BRCA2 siRNAs at 48h after transfection was validated by western blot (Figure 4. 1. C). This suggested that ERCC1-selective effects of PARP1/2 inhibitors were not limited to NSCLC models and could be consistently revalidated using different gene targeting approaches.

4.2.2. ERCC1-deficiency in NSCLC sensitizes cells to BMN 673, a novel hyperpotent PARP1/2 inhibitor

The majority of clinical PARP1/2 inhibitors have biochemical IC_{50}s in the nM to mM range. BMN 673 is a highly potent PARP1/2 inhibitor that selectively inhibits PARP1 at sub-nanomolar concentrations (Wang *et al.*, 2011a) and is currently being assessed in Phase 1 clinical studies. In order to assess whether ERCC1-deficient cells were also selectively more sensitive to BMN 673 than their ERCC1-proficient counterpart, the effect of BMN 673 in the ERCC1-isogenic system were evaluated in colony formation assay. Cells were plated at a 500 cells / well concentration for ERCC1-deficient clones and 300 cells / well for ERCC1-proficient cells, in order to adjust for differences in growth rates. Drug-containing media was replenished three times a week. DLD1 *BRCA2*+/+ and DLD1 *BRCA2*-/- breast cancer cells were evaluated in parallel as negative and positive controls, respectively, and plated at a 1000 cells / well concentration. As shown in Figure 4. 1. D, ERCC1-deficient clones were significantly more sensitive to BMN 673 than their ERCC1-proficient counterparts (Supplementary Table 2). Consistent with the enhanced potency of this compound, the ERCC1-selective effect of BMN 673 was achieved at considerably lower concentrations of PARP1/2 inhibitor than for the other clinical inhibitors (compare with Figure 3.13; Supplementary Table 2).

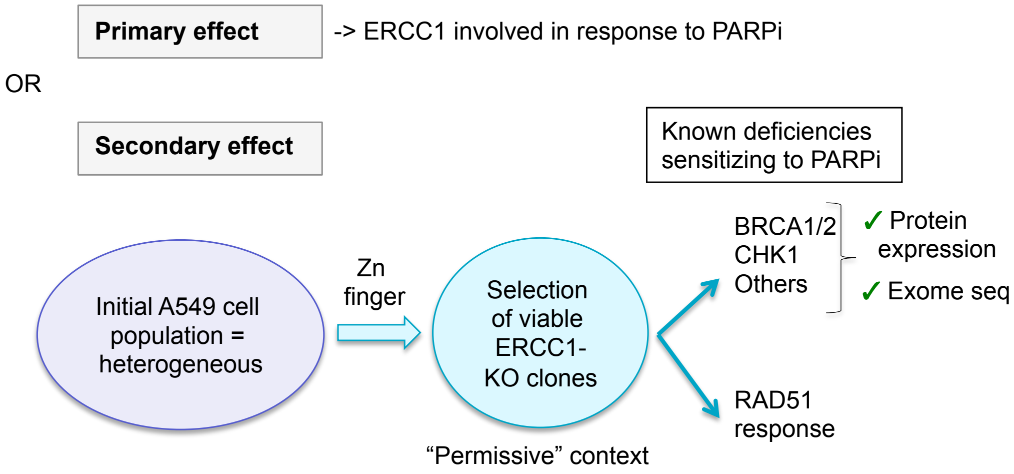


Figure 4. 2. Diagram of potential hypotheses supporting the ERCC1-low selective effect of PARP inhibitors

Top panel: PARP inhibitor sensitivity is a primary effect of ERCC1 deficiency: as such, ERCC1 would be involved in response to PARP inhibitors. Lower panel: PARP inhibitor sensitivity would not directly be related to ERCC1-deficiency, but to another characteristic of the ERCC1-deficient clones that would act as confounding factor. Indeed, the process of generating the clones by Zinc finger targeting could have introduced artefact mutations in genes involved in response to PARP inhibitors; ultimately, the long generation process of the clones under selective pressure of knocking-down *ERCC1* could have selected a subcellular population of A549 cells harbouring a "permissive" genetic context, which would itself sensitize to PARP inhibitors. The main known deficiencies sensitizing to PARP inhibitors were consequently assessed, in order to rule out this possibility.

4.2.3. PARP 1/2 inhibitor sensitivity is a primary effect of ERCC1-deficiency

Beyond *BRCA1/2* deficiency, several DNA repair defects have been shown to modify the cellular response to PARP1/2 inhibitors, including *PTEN, ATM, ATR, CDK1, CHEK1, CHEK2,* and the *FANC* family of genes (Johnson *et al.*, 2011; Murai *et al.,* 2012; Rehman *et al.,* 2010; Turner *et al.*, 2008). Therefore, two hypotheses could be formulated at this stage (Figure 4. 2): (i) PARP inhibitor selective effect was a direct consequence of ERCC1 deficiency, or (ii) ERCC1-deficent clones carried another occult DNA repair defect underlying the PARP inhibitor sensitivity. This was possible as absence of ERCC1 function is theoretically lethal; since the process of isolating three ERCC1-deficiant clones had required the screening of several hundreds of colonies, it was conceivable that ERCC1-deficient clones might have arisen from a selected sub-population of A549 cells having a selective advantage under ERCC1-negative selection pressure. It is indeed well established that cell lines maintained in culture undergo genetic drift and are as such not perfectly genetically identical or homogeneous. Also, although zinc finger nucleases were theoretically designed to recognize a unique *ERCC1*-specific DNA sequence, all gene-targeting systems have off-target effects. As such, small deletions could have disabled another DNA repair gene involved in the PARP inhibitor response.

4.2.3.1. ERCC1-deficient cells do not display overt HR deficiency

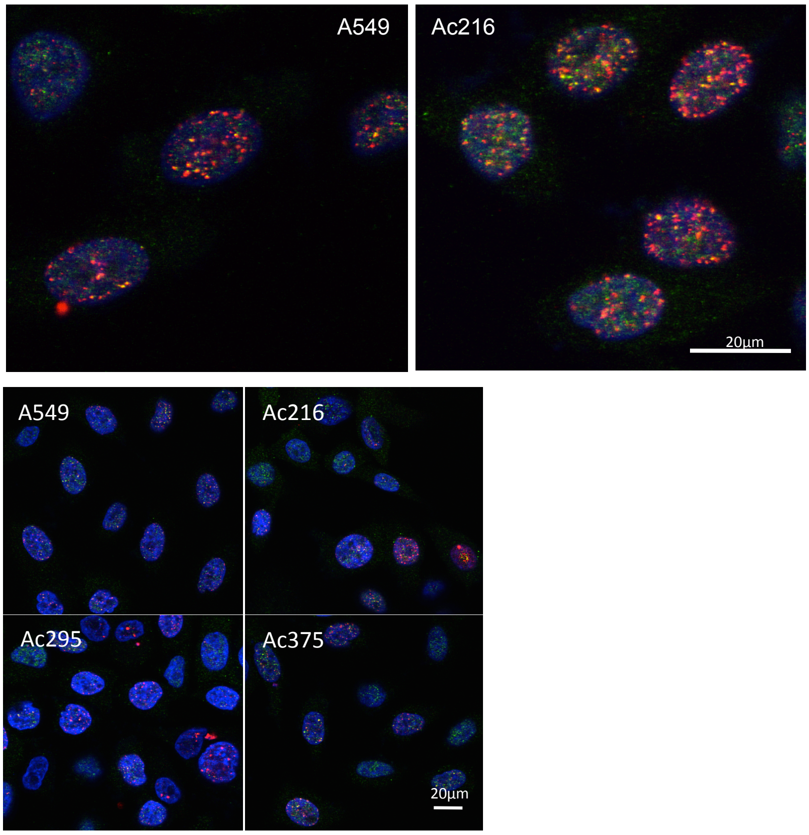
In order to eliminate the hypothesis of an ERCC1-independent PARP1/2 inhibitor sensitivity, several investigations were performed. First, whole exome sequencing of the ERCC1-deficient clones was examined and sequences of ERCC1-deficient models were compared to that of their parental ERCC1-proficient counterpart (see further details on exome sequencing procedure in Chapter 5). All protein-altering mutations were screened individually, and no alteration could be found in a gene coding for a known determinant of PARP inhibitor response. Secondly, the appropriate expression of BRCA1 and BRCA2 at the protein level was assessed, and no decrease in BRCA1 or BRCA2 expression could be detected by western blot in ERCC1-deficient clones (data not shown). Finally, the ability of cells to form nuclear RAD51 foci after olaparib exposure was assessed to investigate a functional HR deficiency. Indeed, the profound sensitivity of *BRCA1* or *BRCA2* mutant cells to PARP1/2 inhibitors is most likely caused by a defect in the recruitment of the recombinase RAD51 to sites of DNA damage. In normal dividing cells, RAD51 recruitment (which can be monitored by visualising nuclear RAD51 foci using immunocytochemistry) precedes DNA strand invasion as part of the process of HR. In order to investigate whether ERCC1-deficient cells displayed such defects, cells were plated on coverslips, exposed to olaparib at a 10μM concentration for 24h and subsequently fixed prior to RAD51 and γH2AX immunostaining. Olaparib exposure elicited the formation of nuclear RAD51 and γH2AX foci in both ERCC1-proficient and deficient models, and that ERCC1-deficient cell lines did not show the overt RAD51 defect found in BRCA-deficient models (Figure 4. 3). Of note, siRNA silencing of *BRCA2* in A549 NSCLC –that was used as positive control - was able to cause a profound defect in olaparib-induced RAD51 response. This suggested that ERCC1 deficiency in NSCLC cells did not abrogate RAD51 function as a mechanism of PARP1/2 inhibitor sensitivity.

4.2.3.2. ERCC1 isoform 202 rescues PARP1/2 inhibitor sensitivity in ERCC1-deficient NSCLC

ERCC1 is expressed as four distinct isoforms, 201, 202, 203 and 204 (Friboulet, 2013). Isoforms 201, 203 and 204 lack amino acids encoded by exons 10, 8 and 3 respectively, whereas ERCC1 isoform 202 is the only isoform to encompass the full XPA, XPF, MSH2, single-strand DNA and double-strand DNA binding domains (Tripsianes *et al.*, 2005; Friboulet *et al.*, 2013a). Very recent work performed by the IGR team during the course of this PhD has demonstrated that isoform 202 is a major determinant of platinum sensitivity in NSCLC, when compared to the other isoforms (Friboulet *et al.*, 2013a) (Figure 4. 4).

To test whether the four distinct ERCC1 isoforms had differential effects on the PARP inhibitor response, previously validated ERCC1 isoform cDNA expression constructs (Chapter 3) were transfected and stably re-expressed into ERCC1-deficient A549 cells. Cells were plated in 96-well plates and cell viability was assessed after 6 days of drug exposure by CellTitre-Glo® luminescent reading. Only the construct encoding isoform 202 restored PARP1/2 inhibitor resistance in ERCC1-deficient clones whereas the other isoforms had no effect (Figure 4. 5; Supplementary Table 2), suggesting that like the response to cisplatin, the response to PARP inhibitors was also determined by ERCC1 isoform 202.

A



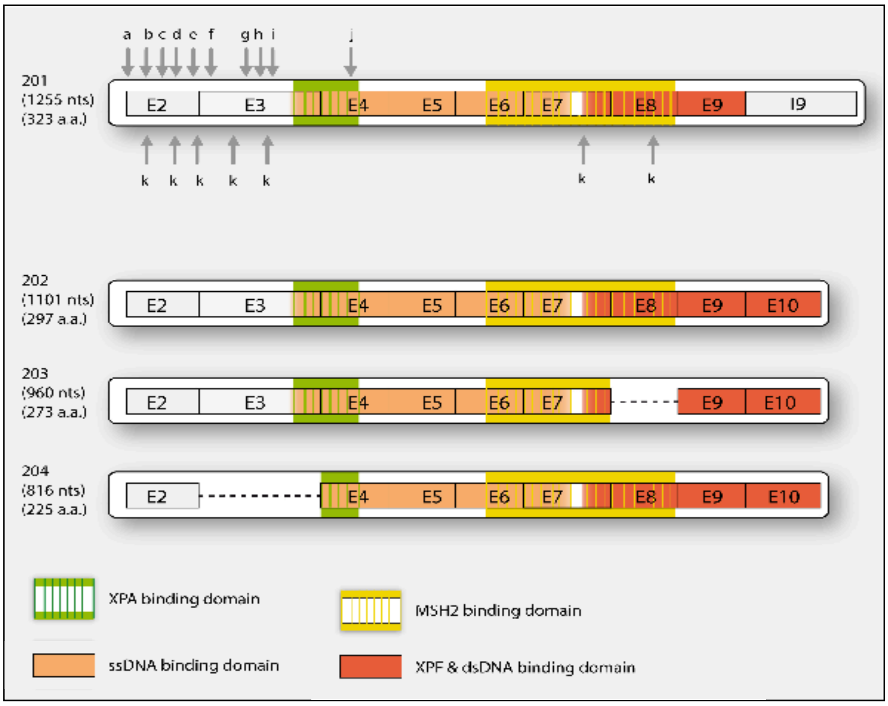
B

Cell line	Treatment	Cells with >5 RAD51 foci (%)
A549	DMSO	5
Ac216	DMSO	10
Ac295	DMSO	7
Ac375	DMSO	6
A549	olaparib	47
Ac216	olaparib	55
Ac295	olaparib	46
Ac375	olaparib	29
A549 + siBRCA2	olaparib	1.5

Figure 4. 3. Selective PARP1/2 inhibitor sensitivity of ERCC1-deficient cells is not related to overt HR deficiency

A) Representative examples of immunofluorescence staining of RAD51 (488 wavelength, green) and γH2AX (555 wavelength, red) foci in ERCC1-isogenic cell lines after olaparib exposure for 24h at 10μM concentration. B) Quantification by confocal microscopy of RAD51 foci formation in ERCC1-isogenic cell lines after DNA damage by olaparib (24h exposure at 10μM concentration). The percentage of RAD51 foci in untreated cells and A549 cells transfected with BRCA2 siRNA are displayed as negative and positive controls, respectively.

A



B

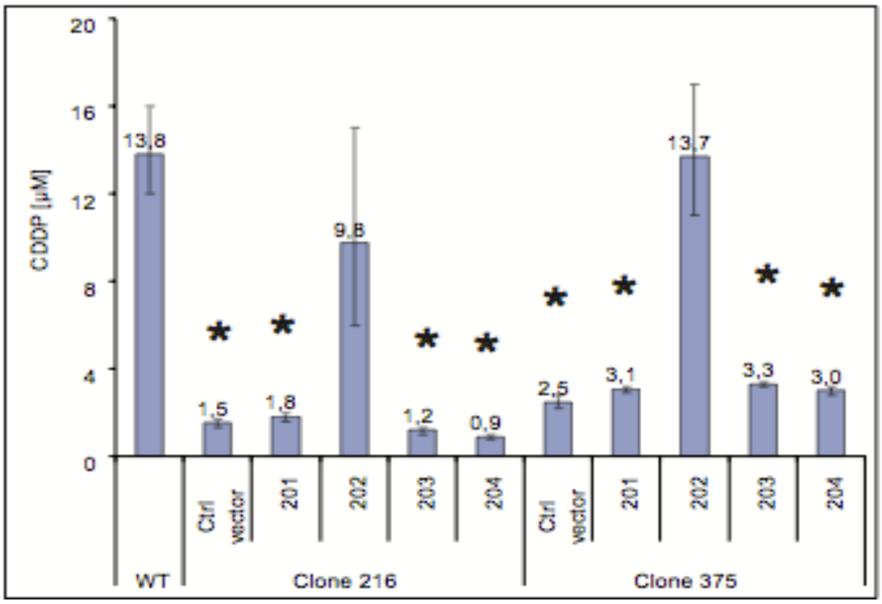
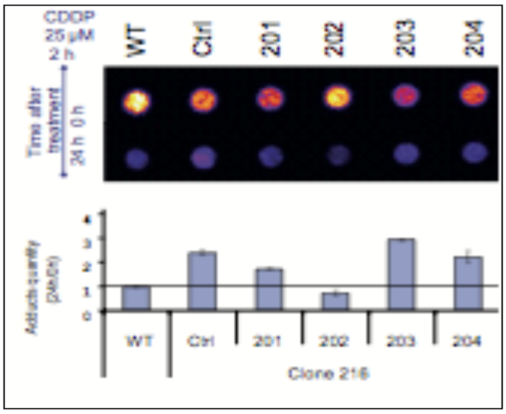


Figure 4. 4. ERCC1 isoform 202 is the only functional isoform with regards to NER and platinum-resistance in vitro and in vivo (Figure and legends reproduced from Friboulet *et al.*, NEJM 2013) (continued next page)

C



D

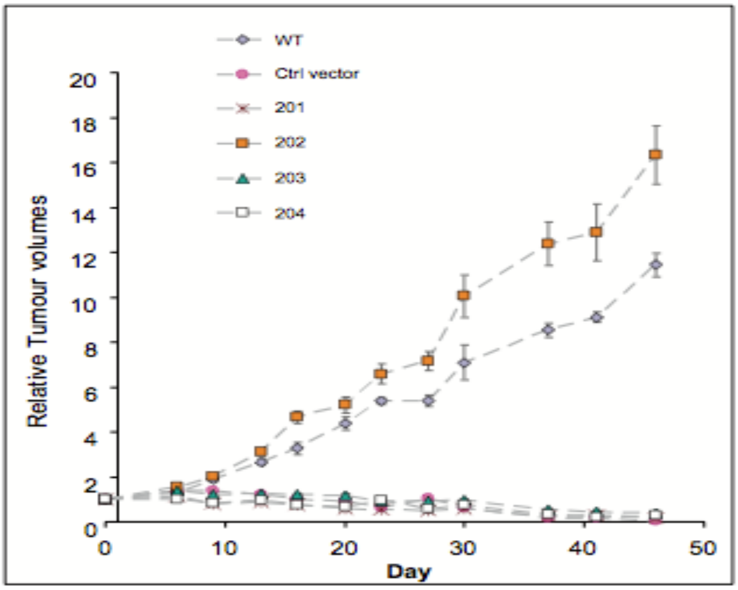


Figure 4. 4. ERCC1 isoform 202 is the only functional isoform with regards to NER and platinum-resistance in vitro and in vivo (Figure and legends reproduced from Friboulet *et al.*, NEJM 2013)

A) Composition of the four distinct ERCC1 isoforms, and differences between these latter in terms of sequence and protein binding domains. Black arrows show epitope binding sites of the most commonly used ERCC1 antibodies, which recognize all ERCC1 isoforms. B) Short-term growth assay for A549 wild-type, ERCC1-deficient, and isoform-expressing cells treated for 48 hours with increasing doses of cisplatin. Numbers above the bars represent the 50% inhibitory concentration for cisplatin. Asterisks indicate significant differences from wild-type cells ($P<0.05$). The I bars represent 95% confidence intervals. C) Dot–blot analysis and quantification of removal of cisplatin–DNA adducts 24 hours after a 2-hour cisplatin treatment at a concentration of 25 µmol/L in A549 wild-type, ERCC1-deficient clones 216 and 375 (the control vector) and cells expressing single ERCC1 isoforms (201, 202, 203, or 204). The adduct quantity ratio (the ratio of the number of adducts at 24 hours to the number at 0 hours) in A549 wild-type cells was set at 1. D) Relative tumour volumes over the course of treatment after 10^6 ERCC1-deficient cells, with single isoform expression, were injected subcutaneously into nude mice. Twice weekly, tumours were measured and cisplatin was injected intraperitoneally at a dose of 1 mg per kilogram of body weight. The I bars represent 95% confidence intervals.

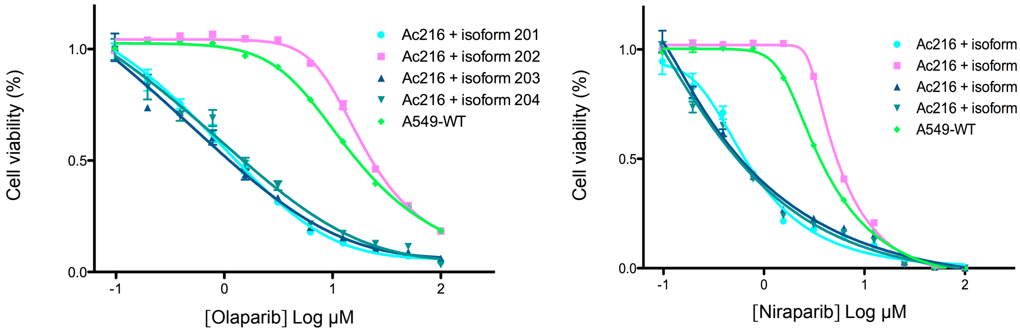


Figure 4. 5. PARP1/2 inhibitor sensitivity is rescued by the functional ERCC1 isoform

Short-term assay evaluating PARP1/2 inhibitor sensitivity in A549-WT cell lines and in ERCC1-deficient clone Ac216 in which each individual ERCC1 isoform had been stably reintroduced (see Chapter 3, Figure 3.3). Only reintroduction of ERCC1 isoform 202 allows rescuing PARP1/2 inhibitor sensitivity with two clinically relevant PARP inhibitors and restore a degree of resistance comparable to that of the parental cell line. Error bars represent the standard deviation from the mean of three independent experiments.

4.2.4. Mechanistic dissection of NSCLC cell sensitivity to PARP1/2 inhibitors

To understand the mechanism of PARP1/2 inhibitor sensitivity in ERCC1-deficient NSCLC cells, a number of molecular genetic modifications or phenotypes associated with the response to PARP1/2 inhibitors were investigated, namely (i) the effect of silencing DNA repair and kinome genes on olaparib sensitivity; (ii) the formation and resolution of nuclear γH2AX foci following olaparib exposure; (iii) the effect on the cell cycle of PARP1/2 inhibitors, and (iv) the effect of PARP1 ablation on PARP1/2 inhibitor sensitivity.

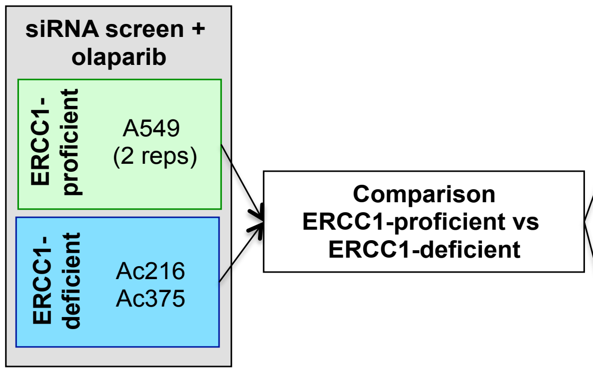
4.2.4.1. The effect of ERCC1 deficiency on PARP1/2 inhibitor sensitivity is epistatic with defects in genes that control the nuclear localisation of RAD51

To further investigate the mechanism by which ERCC1 deficiency led to PARP1/2 inhibitor sensitization, an olaparib siRNA sensitisation screen was performed, which simultaneously evaluated the effect of 911 different genes on the extent of PARP1/2 inhibitor sensitivity in both ERCC1-proficient and deficient clones. For this screen, the siRNA library targeting kinase, tumour suppressor and DNA repair genes, was used. This approach had been previously successfully used in the Gene Function Laboratory to identify determinants of PARP inhibitor sensitivity (Turner, 2008). Screens were performed following conditions described in Chapter 3 for siRNA transfection; olaparib or vehicle (DMSO) was added 48h after siRNA transfection (at the time where effects of the silencing were deemed to be maximal) at a non-toxic concentration corresponding to 80% SF in a 384-well plate assay. Cells were exposed to the drug for five days and results were read using CellTitre-Glo® luminescent viability assay (Promega). Screens were performed in triplicate in the parental ERCC1-proficient A549 cell line (two independent triplicates to increase robustness of the results obtained in the control group) and in the two ERCC1-deficient clones that displayed the lowest levels of ERCC1 expression (Figure 4. 6). All screens were used in the final analysis. The effect of each siRNA on olaparib sensitivity was quantified by calculating Drug Effect (DE) Z-scores. Briefly, the raw luminescence values from each plate were first log2-transformed and normalised to the median luminescence score for all experimental wells of the plate. This operation aimed at taking into account plate-to-plate variation in transfection efficacy and variability in acquisition of luminescence values. The effect of each siRNA

SMARTpool® was then evaluated, both in presence of vehicle (DMSO) and drug (olaparib), and a DE Z-score was generated for each siRNA which allowed distinguishing between the effect of the siRNA on its own (DMSO) and its effect when combined with olaparib (treated wells). First, the DE score was calculated by taking the difference between the median values of the log2-transformed and plate-centred data with and without drug. This DE score was subsequently standardised into a DE Z-score adjusted to the MAD (i.e. the variance of the screen) (see Chapter 3) (Lord *et al.*, 2009; Boutros *et al.*, 2006). A Z-score of ≤ -2 was used to define statistically significant olaparib sensitising effects.

The comparison of sensitisation effects in ERCC1-proficient and deficient clones indicated that siRNAs targeting well-established HR genes that control the localisation of RAD51 to the site of DNA damage, such as *BRCA1*, *BRCA2*, *ATR* and *SHFM1* (aka *DSS1*) enhanced the olaparib sensitivity in ERCC1-proficient NSCLC cells but not in ERCC1-deficient cells (Figure 4. 6). As a sign of the quality of the screens, BRCA2 siRNA were plated in duplicate within the library and both BRCA2 siRNA pools returned DE Z-scores of < -2 in the ERCC1-proficient cells but not the ERCC1-deficient clones. As high-throughput screens generate a number of false positive results, the effect of BRCA2 silencing on olaparib sensitivity in ERCC1-deficient and ERCC1-proficient cells was independently validated. ERCC1-proficient and ERCC1-deficient cells were reverse transfected with BRCA2 siRNA, and olaparib was added 48h after transfection. Cell viability was assessed after five days of drug exposure using CellTite-Glo® luminescent viability assay (Promega). As displayed in Figure 4. 7, BRCA2 silencing had no effect on olaparib sensitivity in ERCC1-deficient cells, contrary to ERCC1-proficient cells (Supplementary Table 2). Moreover, this effect was also observed with another PARP inhibitor, niraparib, to a similar degree (Figure 4. 7. B). These observations suggested that ERCC1 deficiency and HR gene deficiency were in fact *epistatic*, such that the effect of modulating ERCC1 masked the phenotypic effect of modulating well-known HR genes. As observation of epistasis between genes is usually indicative of involvement in a shared process, these results also suggested that, although ERCC1 deficiency had no effect on the RAD51 response, ERCC1 function might be linked to HR gene function in the response to PARP1/2 inhibitors.

A



Hits sensitizing only ERCC1-proficient cells to olaparib:

	median Z-score	
siRNA	**ERCC1-prof.**	**ERCC1-def.**
BRCA2	**-6.9735**	-0.6682
BRCA2	**-7.3972**	0.5153
ATR	**-3.3420**	-0.5668
SHFM1	**-3.6564**	0.8222
BRCA1	**-2.6386**	-0.9221

Hits sensitizing only ERCC1-deficient cells to olaparib:

	median Z-score	
siRNA	**ERCC1-prof.**	**ERCC1-def.**
GAK	0.46941	**-3.98736**
SMG1	0.15780	**-3.58153**
TGFBR2	1.57261	**-3.21144**
PRPS1	0.49857	**-3.09022**
KSR	0.54998	**-3.05093**
GTF2H3	-0.11419	**-2.60892**
GRK7	-0.57582	**-2.50933**
ALS2CR7	-0.05633	**-2.41465**
NEK1	0.11241	**-2.18185**

B

siRNA	**A549 rep 1 Z-score**	**A549 rep 2 Z-score**	**Ac216 Z-score**	**Ac295 Z-score**
BRCA2	**-4.001269**	**-9.945722**	-0.7319837	-0.6043214
BRCA2	**-3.160561**	**-11.63386**	1.33327708	-0.3027674
ATR	**-2.088426**	**-4.595545**	-0.5451774	-0.5884155
SHFM1	**-2.631897**	**-4.680863**	0.81244135	0.8319544
BRCA1	**-2.20534**	**-3.071778**	-0.0695399	-1.7745903

Figure 4. 6. Flowchart and results of a siRNA screen in combination with olaparib in the ERCC1-isogenic model

A) Olaparib was added at a non-toxic (80% survival fraction) concentration to a siRNA screen of 784 kinases and tumour suppressor genes as well as 127 DNA repair genes 48h after reverse-transfection. The results were obtained by comparing the ERCC1-proficient *vs* ERCC1-deficient cells. Hits displayed followed significance criteria in both ERCC1-deficient clones when compared to the ERCC1-proficient isogenic cell line. Abbreviations: reps: replicates; ERCC1-prof.: ERCC1-proficient; ERCC1-def.: ERCC1-deficient. B) Drug effect values of a siRNA screen evaluating 911 kinase, tumour suppressor and DNA repair genes, combined with olaparib treatment at non-toxic concentration (80% surviving fraction). The screen was performed in triplicate. The sensitizing effect of each siRNA (Drug Effect, or DE) was estimated using a DE Z-score of ≤ -2 to define statistically significant olaparib sensitising effects. BRCA1, BRCA2, ATR and SHFM1 were the only siRNA sensitising the ERCC1-proficient cells only to olaparib, as assessed by a DE Z-score ≤ -2. A549a and A549b represent two individual screens independently performed in triplicate each in A549 cell line; Ac216: ERCC1-deficient clone 216; Ac375: ERCC1-deficient clone 375.

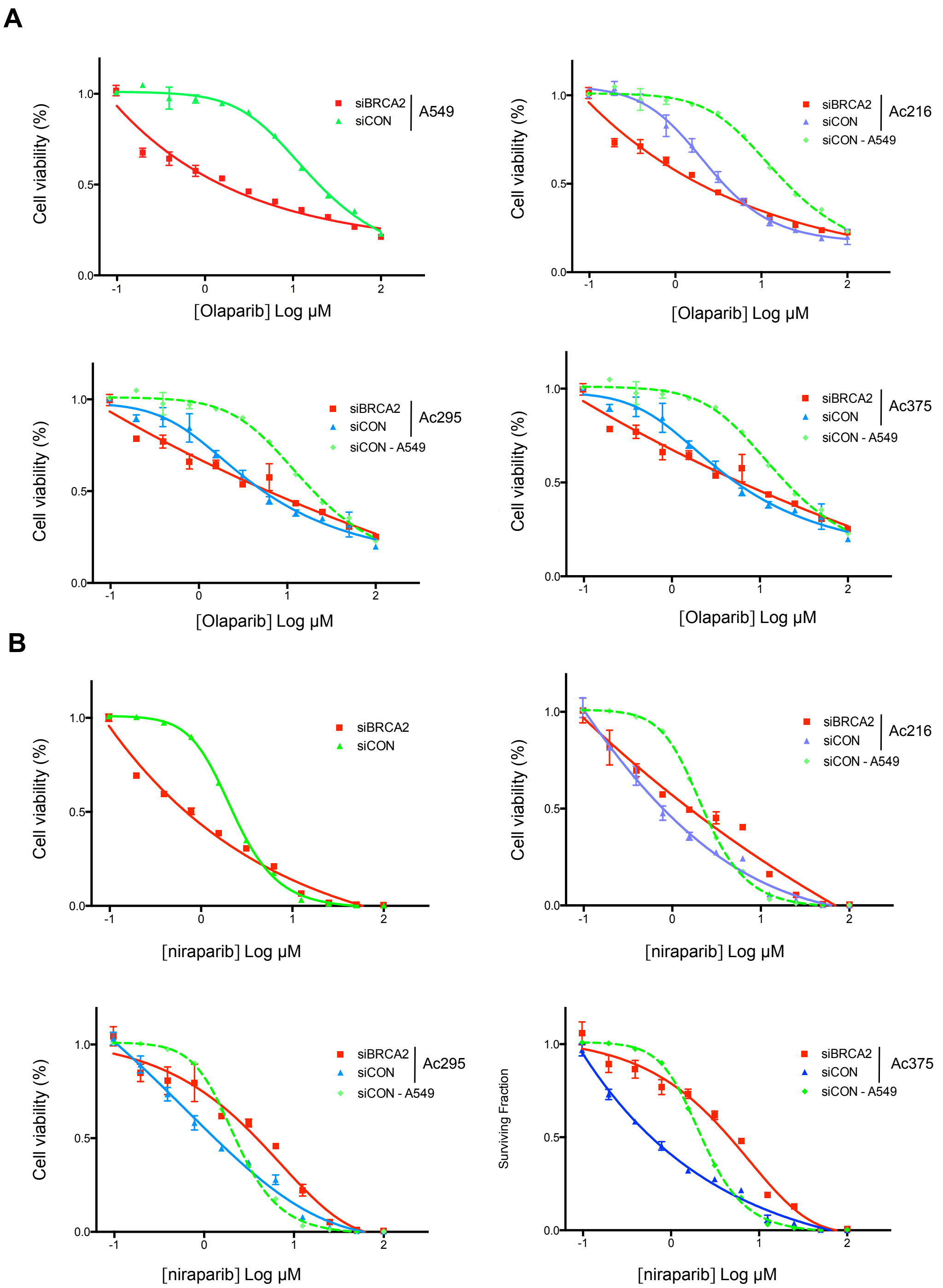


Figure 4. 7. BRCA2-silencing is epistatic with ERCC1-deficiency in mediating PARP1/2 inhibitor sensitivity

Effect of *BRCA2* knock-down by siRNA on sensitivity of ERCC1-isogenic cell lines to olaparib (panel A) and niraparib (panel B). Cells were reverse transfected with BRCA2 siRNA and drug was added 48h after transfection. Cells were exposed to the drug for five days. Error bars represent the standard deviation from the mean of three independent experiments.

4.2.4.2. ERCC1-deficient cells display a delay in the repair of DNA damage following PARP1/2 inhibitor exposure

The ability of ERCC1-deficient cells to resolve DNA damage following olaparib treatment was also assessed. In addition to the formation of RAD51 nuclear foci, one of the other characteristics of exposure to PARP1/2 inhibitors is the formation of nuclear γH2AX foci, a marker of the phosphorylation of histone H2AX at the site of DNA double-strand breaks and stalled replication forks. ERCC1-proficient and -deficient NSCLC cells were exposed to olaparib (10μM) for 24h, and γH2AX foci formation after drug removal, was monitored using immunocytochemistry. Prior to olaparib exposure, the frequency of cells with γH2AX foci in untreated ERCC1-proficient and -deficient clones was not significantly different, with all clones exhibiting approximately 10% of cells with more than ten γH2AX foci (data not shown).

After 24 hours of olaparib exposure, the frequency of cells with γH2AX foci increased, with 55-80% of cells exhibiting more than ten γH2AX foci, regardless of ERCC1 genotype (time point T=0, Figure 4. 8. A, frequency of cells with more than ten γH2AX foci A549 *vs* Ac216: p=0.24; A549 *vs* Ac295: p=0.06; A549 *vs* Ac375: p=0.06, Student's t test). By contrast, the resolution of γH2AX foci after PARP1/2 exposure was significantly delayed in ERCC1-deficient clones when compared to the ERCC1-proficient parental NSCLC cells, with 25-40% of ERCC1-deficient cells exhibiting more than ten γH2AX foci, compared to only 8% in the ERCC1 proficient parental clone at 76h after drug removal (A549 *vs* Ac216: p=0.002; A549 *vs* Ac295: p=0.002; A549 *vs* Ac375: p=0.004, Student's t test, Figure 4. 8). These observations were consistent with the hypothesis that ERCC1-deficient cells displayed a defect in the resolution of DNA damage caused by PARP1/2 inhibitors.

4.2.4.3. The delay in repair of DNA lesions following PARP1/2 inhibitor exposure is associated with G2/M cell cycle arrest

Interestingly, the maximum proportion of cells displaying more than ten γH2AX foci was not observed at drug removal, but 6-12 hours after drug removal (Figure 4. 8). This suggested a cell cycle effect in the formation of γH2AX foci, rather than a direct induction of DSB by olaparib itself (by contrast with ionising radiation (IR), for which the maximum of γH2AX foci would have been observed at T0 following irradiation). In order to investigate this hypothesis, the cell cycle response to olaparib exposure in

ERCC1-deficient NSCLC cells was assessed. As in the previous experiment, cells were exposed to olaparib 10µM for 24h and changes in the cell cycle after drug removal were monitored, using propidium iodide flow cytometry. Although both ERCC1-deficient and ERCC1-proficient models exhibited a G_2/M arrest in response to olaparib exposure, this arrest was much more profound and prolonged in ERCC1-deficient cells (Figure 4. 9 and Table 4. 1, % cells in G_2 at drug removal for A549 = 26.7, Ac216 = 51.8, Ac295= 51.8, Ac375= 54.9). This difference in G_2/M arrest was most pronounced 6h after drug removal (% cells in G_2 at 6 hours after drug removal for A549 = 31.1, Ac216 = 64.3, Ac295 = 59.7, Ac375 = 63.3), coinciding with the maximal formation of γH2AX foci (Figure 4. 8), consistent with the hypothesis that the resolution of DNA damage in ERCC1-deficient clones was delayed in response to a PARP1/2 inhibitor, when compared to ERCC1-proficient cells.

4.2.4.4. The combination with a G_2/M inhibitor accelerates PARP1/2 inhibitor-induced cell death in ERCC1-deficient cells

As PARP1/2 inhibitors induced a profound and prolonged G_2/M blockage in ERCC1-deficient cells only, it was hypothesised that the addition of a G_2/M inhibitor to the PARP1/2 inhibitor treatment could precipitate cell death in ERCC1-deficient cells only – for example through mitotic catastrophe - and as such increase the therapeutic window between ERCC1-proficient and ERCC1-deficient cells. Wee 1 is serine-threonine kinase which plays a key role in the regulation of the G_2/M checkpoint and inhibits entry into mitosis through inhibiting Cdk1. Notably, Wee1 inhibitors, such as MK1175, can force cells to enter into mitosis prematurely by abrogating the G_2/M checkpoint. The effect of combining MK1175 at non-toxic concentrations with olaparib was consequently assessed in parallel in two different formats: a short-term assay (3 days of drug exposure) and a mid-term assay (6 days of drug exposure). The objective of the 3-day exposure assay was to evaluate the effect of inhibiting the G_2/M checkpoint at a time where ERCC1-deficient cells would still be blocked into G_2 following PARP1/2 inhibitor treatment, whereas ERCC1-proficient cells would already have recovered a baseline cell cycle profile. By contrast, the objective of the 6-day assay was to evaluate the effect of combining the Wee1 inhibitor at a time where ERCC1-proficient and ERCC1-deficient cell lines had both recovered their baseline cell cycle profile.

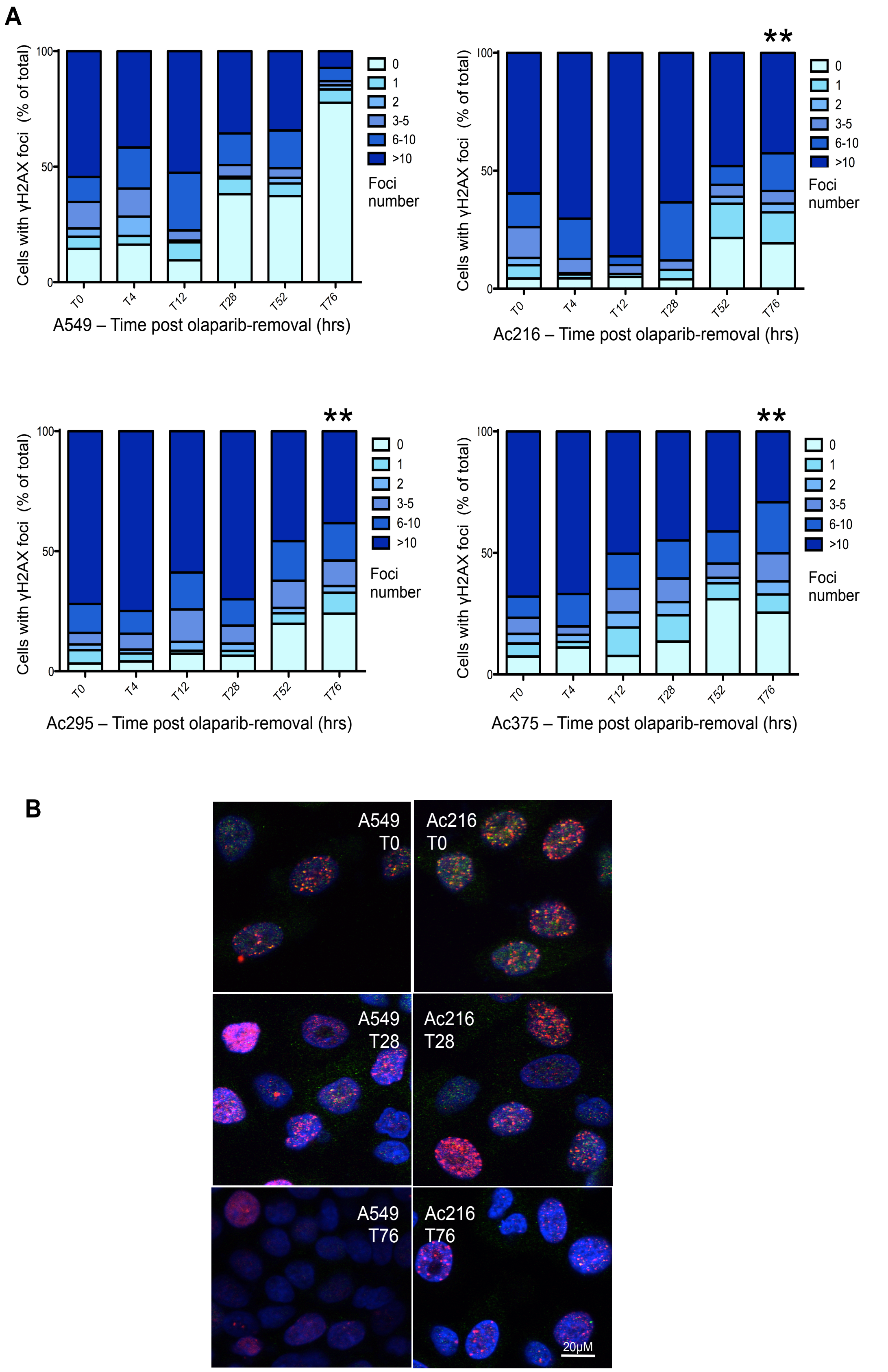


Figure 4. 8. Kinetics of γH2AX foci formation following olaparib treatment
(legend next page)

Figure 4.8. Kinetics of γH2AX foci formation following olaparib treatment (figure previous page)

A) Quantification of γH2AX foci per cell following olaparib treatment in isogenic ERCC1-proficient and ERCC1-deficient cell lines. Cells were continuously exposed to 10μM olaparib for 24h, prior to drug removal (T0). Foci were counted at different time points following olaparib removal. The proportion of cells presenting more than ten γH2AX foci was not significantly different between ERCC1-proficient and ERCC1-deficient cell lines at T0 (A549 *vs* Ac216: p=0.24; A549 *vs* Ac295: p=0.06; A549 *vs* Ac375: p=0.06, Student's t test); by contrast, this proportion was significantly higher at 76h after drug removal in ERCC1-deficient clones as compared to the parental cell line (A549 *vs* Ac216: p=0.002; A549 *vs* Ac295: p=0.002; A549 *vs* Ac375: p=0.004) B) Representative images of kinetics of RAD51 and γH2AX foci formation after olaparib treatment. Cells were exposed to 10μM olaparib for 24h prior to drug removal. Kinetics of RAD51 (488 wavelength, green) and γH2AX foci (555 wavelength, red) formation was then assessed by confocal microscopy at different time points. Examples of pictures obtained with the A549 ERCC1-proficient cell line and one ERCC1-deficient clone at olaparib removal, 28h and 76h after olaparib removal (T0, T28 and T76, respectively).

	A549			Ac216			Ac295			Ac375		
	G1	S	G2	G1	S	G2	G1	S	G2	G1	S	G2
DMSO	66.9	18.4	13.5	66.6	15.3	17	66.4	15.9	15.9	68.7	13.4	16.9
T0	57.8	13.9	26.7	37.1	9.6	51.8	39.9	7.3	51.8	36.7	7.7	54.9
H6	53.5	14.1	31.1	28.1	3.9	64.3	32	6.8	59.7	29.1	6.4	63.3
H24	58	17.4	22.2	43.8	5.4	48.5	50.5	5.3	42.2	46.4	3.7	48.7
H30	61.3	13.9	23.6	41.5	11.6	44.3	48.5	9.9	39.6	41.6	9.5	47.5
D3	68.8	12.5	17.6	55.5	9.6	32.8	56.4	8.5	31.9	59.5	8	30.8
D4	75.7	9.6	13	68.5	8.8	21.1	72.7	5.7	18.3	68.5	9.7	20.1
D5	82.6	7.2	9.4	76.1	7.6	14.4	80.2	4.2	13.4	74.7	6.4	16.6

>40% cells   20-40% cells   <20% cells

Table 4. 1. Cell cycle analysis at different time points following olaparib removal.

Numbers represent the percentage of cells in each phase of the cell cycle at different time points following olaparib removal, analysed by propidium iodide FACS (see Figure 4.9).

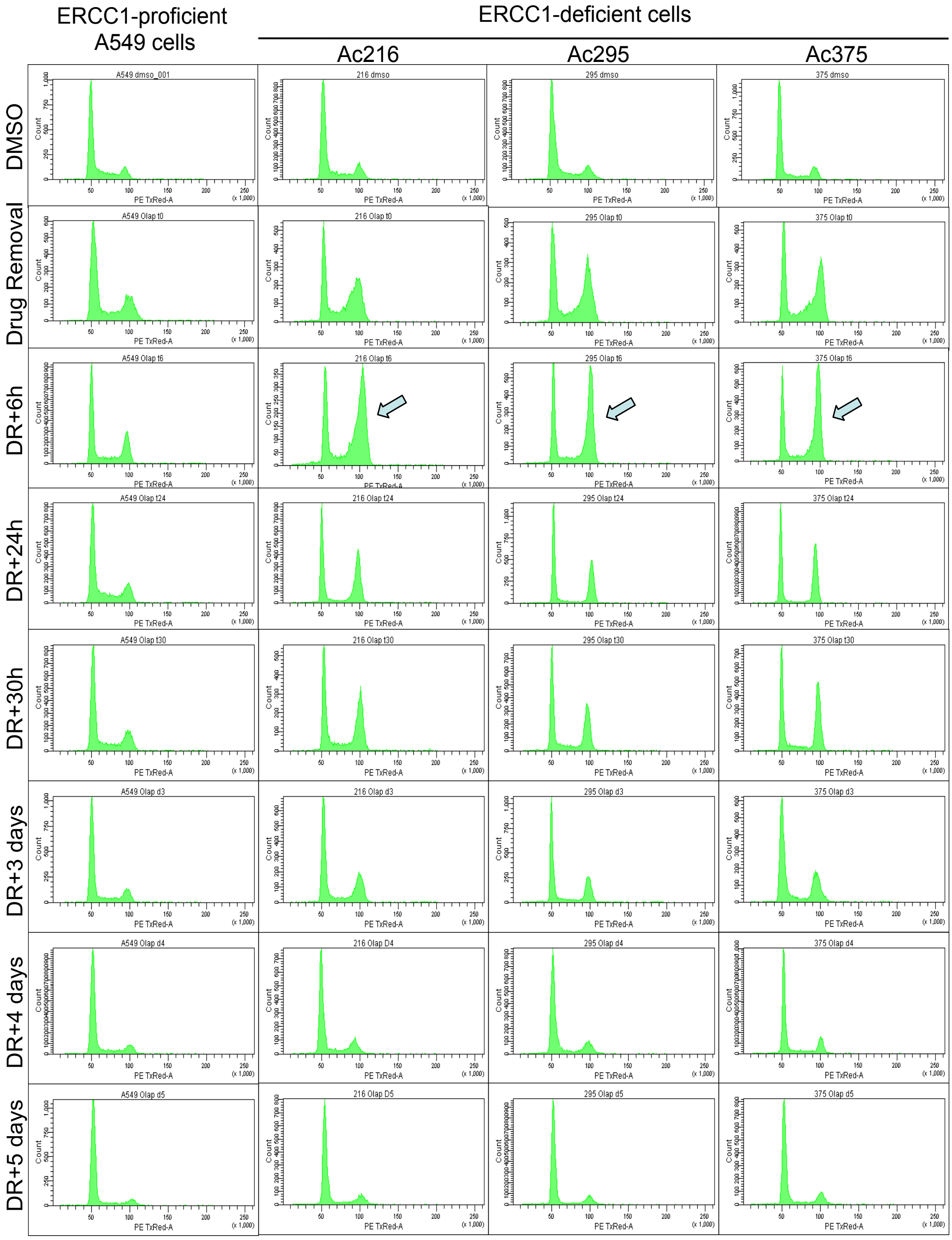


Figure 4. 9. Cell cycle analysis following olaparib treatment

FACS profile of ERCC1-proficient and ERCC1-deficient cells before and after olaparib treatment at different time points. Cells were exposed to 10μM olaparib for 24h, prior to drug removal. Cells were stained with propidium iodide at several time points after drug removal (DR) for analysis of the DNA content and cell cycle phase. Arrows indicate the G_2/M blockade observed at 6h after drug removal in the ERCC1-deficient clones only.

For both formats, the Wee1 inhibitor was added concomitantly to olaparib (i.e. at day 1 after cell plating) given the use of asynchronous cells and the use of an inhibitor that could remain stable and active for several days in culture.

Although the 3-day experiment format increased the therapeutic window between ERCC1-proficient and ERCC1-deficient cells, the addition of a Wee1 inhibitor for 6 days had no effect on olaparib selectivity (Figure 4. 10). The absence of selective effect of olaparib after three days only of drugging was expected, as longer exposure is usually required to observe PARP1/2 inhibitors-mediated cytotoxicity. However, the increase in therapeutic window following the adjunction of a Wee1 inhibitor suggested that abrogating the G_2/M checkpoint may precipitate cell death in the ERCC1-deficient cells only, potentially through mitotic catastrophe; the disappearance of the beneficial effect of adding a Wee1 inhibitor for six days of drug exposure suggested that, at later time points, PARP1/2 inhibitor-induced cell death may occur through an alternative mechanism, such as apoptosis.

Figure 4.10. Increase in therapeutic window when combining short PARP1/2 inhibitor exposure with a Wee1 inhibitor (figure next page)

A) Combination of olaparib with non-toxic concentrations of a Wee1 inhibitor (MK1175) for three days of drug exposure. No ERCC1-selective effect is observed after three days of drug exposure with any of the agents as monotherapy, whereas an ERCC1-selective effect (increase in therapeutic window) of olaparib can be observed when combining both agents. B) Combination of olaparib with non-toxic concentrations of a Wee1 inhibitor (MK1175) for six days of drug exposure. No ERCC1-selective effect is observed with MK1175 as monotherapy, contrary to olaparib as previously described. The therapeutic window observed after longer exposure to olaparib between ERCC1-proficient and ERCC1-deficient cells is not further increased after six days of exposure to the combination. Error bars represent the standard deviation from the mean of three independent experiments.

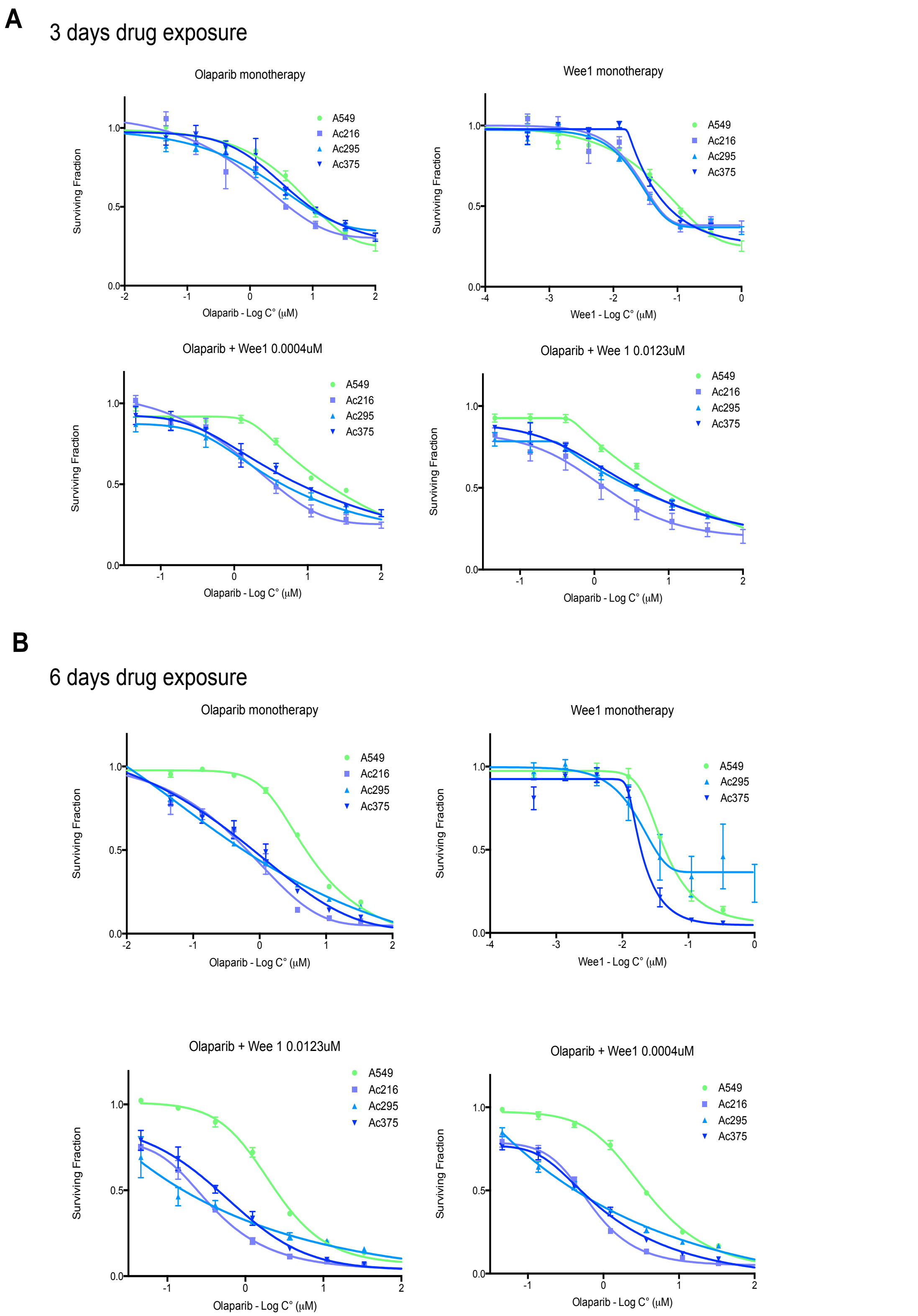


Figure 4. 10. Increase in therapeutic window when combining short PARP1/2 inhibitor exposure with a Wee1 inhibitor (legend previous page)

4.2.4.5. PARP1 silencing causes PARP1/2 inhibitor resistance in ERCC1 deficient NSCLC cells

Several overlapping mechanisms have been suggested to explain the cytotoxicity of PARP1/2 inhibitors, including the formation of DNA double-strand breaks subsequent to the failure of single-strand break repair caused by PARP1 inhibition (Farmer *et al.*, 2005). More recently, the observation that the cytotoxic response to small molecule PARP1/2 inhibitors can be abrogated by the genetic suppression of PARP1 levels has led to the hypothesis that PARP1 trapped onto DNA as a result of its catalytic inhibition might be a key cytotoxic DNA lesion (Kedar *et al.*, 2012; Murai *et al.*, 2012; Pettitt *et al.*, 2013). This observation is consistent with the idea that auto-PARylation of PARP1 is required for the dissociation of this enzyme from damaged DNA and that, in the absence of a PARP1 substrate, the PARP1/DNA lesion is not formed, resulting in a minimisation of the effects of PARP1/2 inhibitors in certain contexts. In order to investigate this hypothesis, the effects of silencing PARP1 were assessed under PARP1/2 inhibitors exposure. ERCC1-deficient models were reverse-transfected with PARP1 siRNA and PARP1/2 inhibitors were added 48h after transfection. Cells were exposed to the drug for five consecutive days prior viability assessment by CellTitre-Glo® luminescent reading. As shown in Figure 4. 11, PARP1 siRNA transfection rescued PARP1/2 inhibitor sensitivity in ERCC1-deficient clones, but PARP1 depletion did not affect the sensitivity of ERCC1-proficient cells to PARP inhibition (Supplementary Table 2). This effect was more pronounced following niraparib than following olaparib exposure, which is consistent with the reported greater ability of niraparib to bind PARP1 to the DNA (Murai *et al.*, 2012). This suggested that the selective cytotoxicity of PARP1/2 inhibitors towards ERCC1-deficient cells may be primarily mediated by the trapping of PARP1 on the DNA.

4.2.4.6. Metaphase spreads of ERCC1-deficient cells after olaparib exposure reveals the appearance of radial structures

In order to investigate the consequences of PARP1 catalytic activity inhibition at the chromosome level and examine whether this effect was different in ERCC1-proficient and ERCC1-deficient cells, proliferating ERCC1-proficient and -deficient cells were exposed to olaparib (10μM) or vehicle (DMSO) for 24h. Cells were collected and metaphase spreads prepared in order to study chromosomal abnormalities.

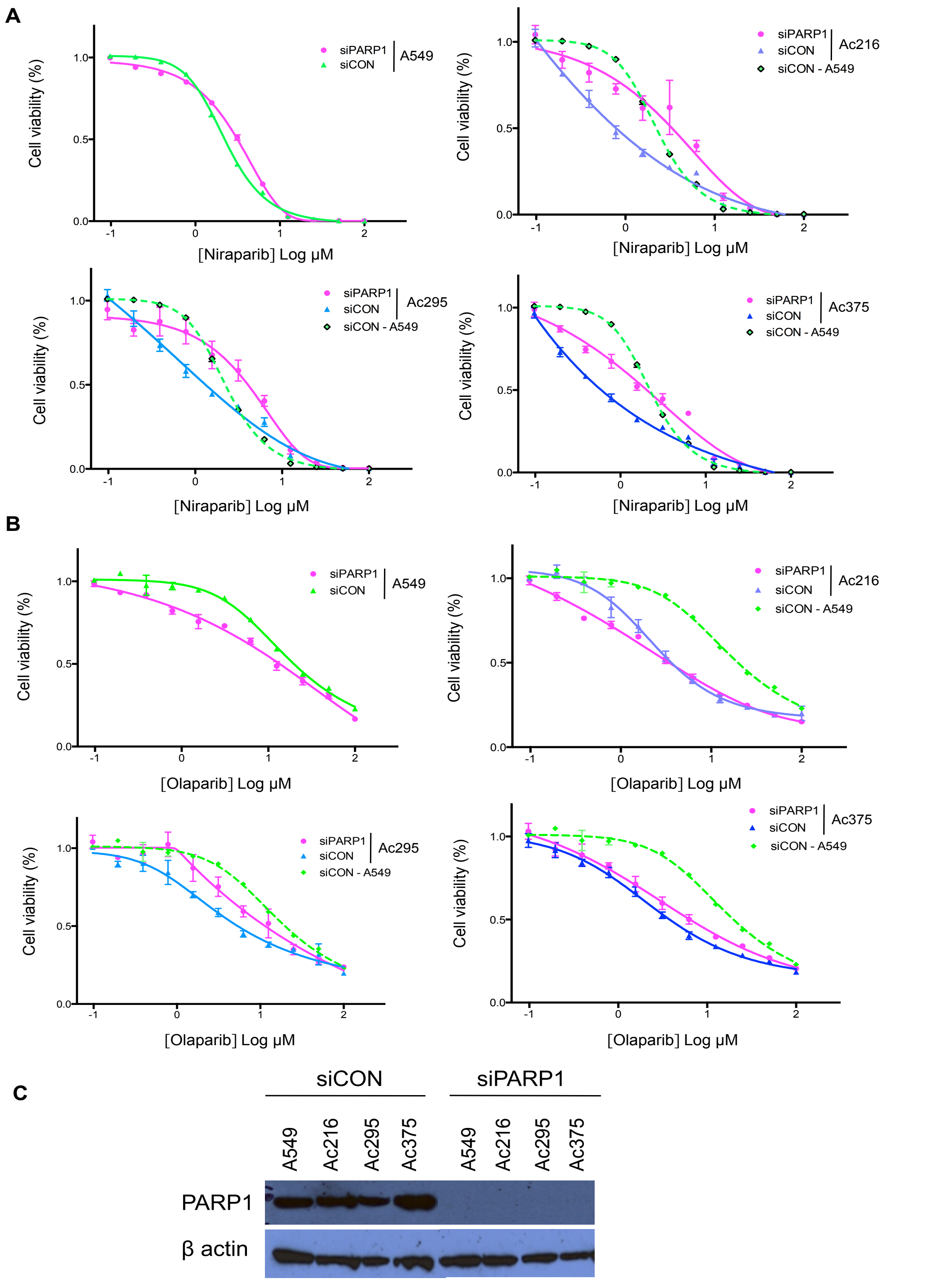


Figure 4. 11. PARP1 silencing effect on PARP1/2 inhibitor sensitivity in ERCC1-isogenic cell lines

Effect of *PARP1* knock-down by siRNA on sensitivity of ERCC1-isogenic cell lines to niraparib (panel A) and olaparib (panel B). Cells were reverse transfected with PARP1 siRNA and drug was added 48h after transfection. Cells were exposed to the drug for five days. Error bars represent the standard deviation from the mean of three independent experiments. C) Western blot showing appropriate PARP1 silencing after siRNA transfection.

Twenty to fifty metaphase images were collected using a Zeiss LSM 710 confocal microscope and subsequently analysed by a cytogeneticist (Dr Géraldine Pottier, Comissariat à l'Energie Atomique, CEA). All cell lines displayed a high degree of aneuploidy (with a higher proportion of polyploidy cells in ERCC1-deficient models) and chromosomal abnormalities were overall rare. DMSO-treated ERCC1-proficient and ERCC1-deficient cells displayed the same proportion of abnormalities, suggesting that the absence of ERCC1 is not associated with increased genomic instability in untreated cells (Figure 4. 12). In spreads from cells treated with olaparib, aberrations also seemed to be as common in ERCC1-proficient as in ERCC1-deficient cells, except for one ERCC1-deficient clone (Ac216, which expressed the lowest level of ERCC1) in which the number of abnormalities was increased. Importantly, Ac216 cells treated with olaparib displayed the appearance of radial structures (circles in red in Figure 4. 12. B), a structure thought to result from the fusion of the broken arms of non-homologous chromosomes that cannot be properly segregated in most cells, resulting in either chromosome breakage or a failure in cell division (Deans *et al.*, 2011). This chromosome structure is also characteristic of failure of interstrand crosslink repair (Deans *et al.,* 2011), and similar observations have already been reported in ERCC1-/- ES cells exposed to the crosslinking agent mitomycin C (Niedernhofer *et al.,* 2004). This suggested that the mechanism underlying the selective toxicity of PARP1/2 inhibitor on ERCC1-deficient cells was more likely the creation of a DNA lesion – perhaps PARP1 itself, Cf Figure 4. 11 – rather than the inhibition of PARP1 activity in single-strand break repair. Of note, parallel experiments performed at Institut Gustave Roussy by the U981 team revealed that ERCC1-induced polyploidy was rescued by the reintroduction of isoform 202 (Figure 4. 13), and that this latter isoform was also responsible for the repair of interstrand crosslinks repair (Figure 4. 14).

A

Cell line	Treatment	Chromosome number	Centromere number	Dicentric	Acentric	Ring	Chromatide break	Chromosome break	Tricentric	Double-minute	Radial
A549	DMSO	2990	3017	1.07	0.26	0.03	0.47	0.03	0.03	0.13	0
	olaparib	1170	1177	1.11	0.43	0	0.43	0	0	0.08	0
Ac216	DMSO	1898	1941	0.21	0.16	0	0.1	0	0	0	0
	olaparib	985	1001	1.62	0.2	0	0.41	0	0	0.51	0.3
Ac295	DMSO	772	714	0.77	0	0	0.52	0	0	0	0
	olaparib	1802	1825	0.39	0	0	0	0	0	0.11	0
Ac375	DMSO	2782	2804	0.82	0.07	0	0.07	0	0	0	0
	olaparib	731	738	0.96	0	0	0	0	0	0	0

B

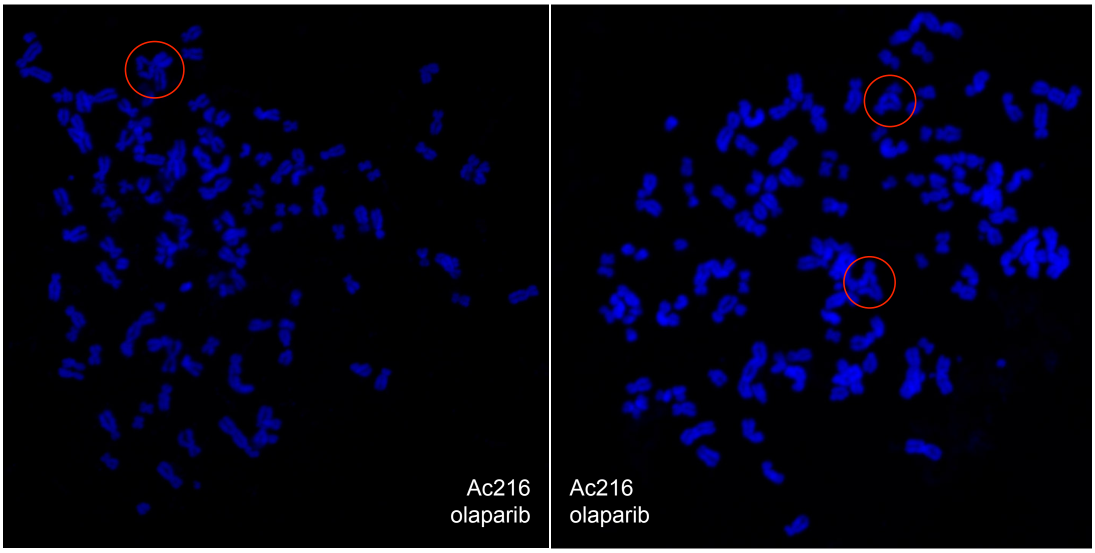


Figure 4. 12. Metaphase spreads of ERCC1-WT and ERCC1-deficient cells after olaparib exposure reveals appearance of radial structures

Cells were exposed to 10μM olaparib for 24h prior to drug removal and collection of metaphase cells. A) Quantitative assessment and classification of chromosomal abnormalities observed following olaparib exposure or vehicle (DMSO). An increase in the number of decentric or double-minute chromosomes, as well as chromatide break and radial structure was observed in the Ac216 clone, which expresses the lowest levels of ERCC1. B) Images of the Ac216 cell line where radial structures were observed (as indicated by red circles. That data was analysed by Geraldine Pottier (CEA).

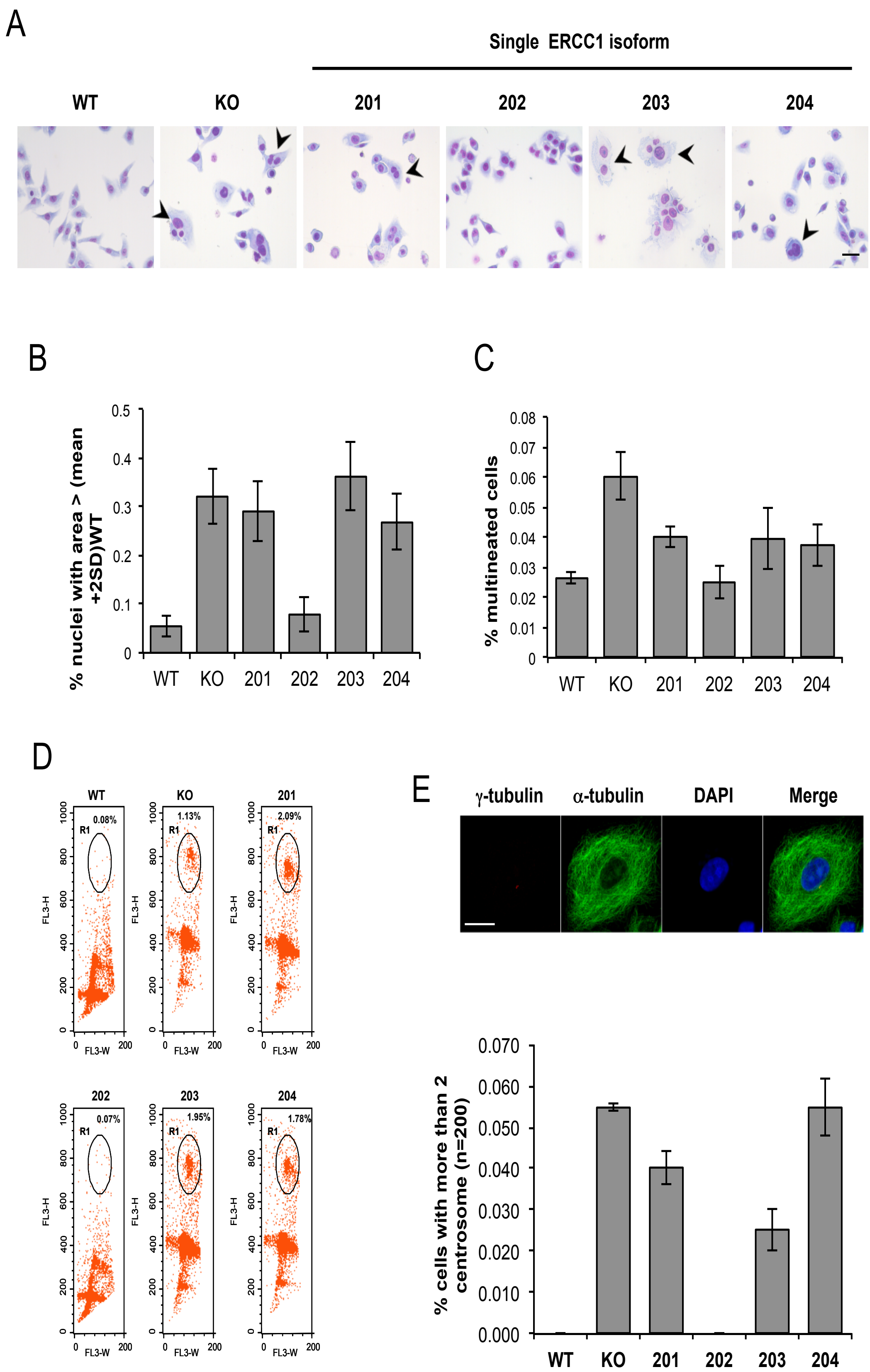


Figure 4. 13. ERCC1 deficiency-induced polyploidy is rescued by ERCC1 isoform 202 (Figure and legends reproduced from Friboulet et al, Cell Cycle 2013) (legend next page)

Figure 4. 13. ERCC1 deficiency-induced polyploidy is rescued by ERCC1 isoform 202 (Figure and legends reproduced from Friboulet et al, Cell Cycle 2013) (figure previous page)

A) Diff-Quick^{TM} staining of wild-type A549 (WT), A549 knocked-down for ERCC1 (KO) and A549 expressing individually each of the four ERCC1 isoform (201, 202, 203 and 204). Arrowheads point to abnormally sized nuclei and multinucleated cells. Scale bar, 20 µm; B) Nuclear area were determined using ImageJ software in wild-type A549 (WT), A549 knocked-down for ERCC1 (KO) and A549 expressing individually each of the four ERCC1 isoform (201, 202, 203 and 204) stained with Diff-Quick^{TM} as in A. The percentage of cells with nuclear area superior to the average nuclear area of WT cells was plotted. Error bars indicate SEM; C) The percentage of multinucleated cells was scored manually on Diff Quick^{TM} stain cells (n = 200). Error bars indicate SEM; D) wild-type A549 (WT), A549 knocked-down for ERCC1 (KO) and A549 expressing individually each of the four ERCC1 isoform (201, 202, 203 and 204) were blocked in G2/M by a 6h colcemid treatment (Karyomax) and then processed for PI DNA profiling. Dot plot representation of the flow cytometry analysis are shown. Circled population corresponds to aneuploidy single cells containing more than 4N DNA E) Images show representative γ- and α-tubulin immunofluorescence staining of A549 cells used to score for centrosome number per cell (Scale bar, 10 µm). The percentage of cells with more than 2 centrosomes was plotted. At least 200 wild-type A549 (WT), A549 knocked-down for ERCC1 (KO) or A549 expressing individually each of the four ERCC1 isoform (201, 202, 203 and 204) were counted. Error bars indicate the SEM of a representative experiment.

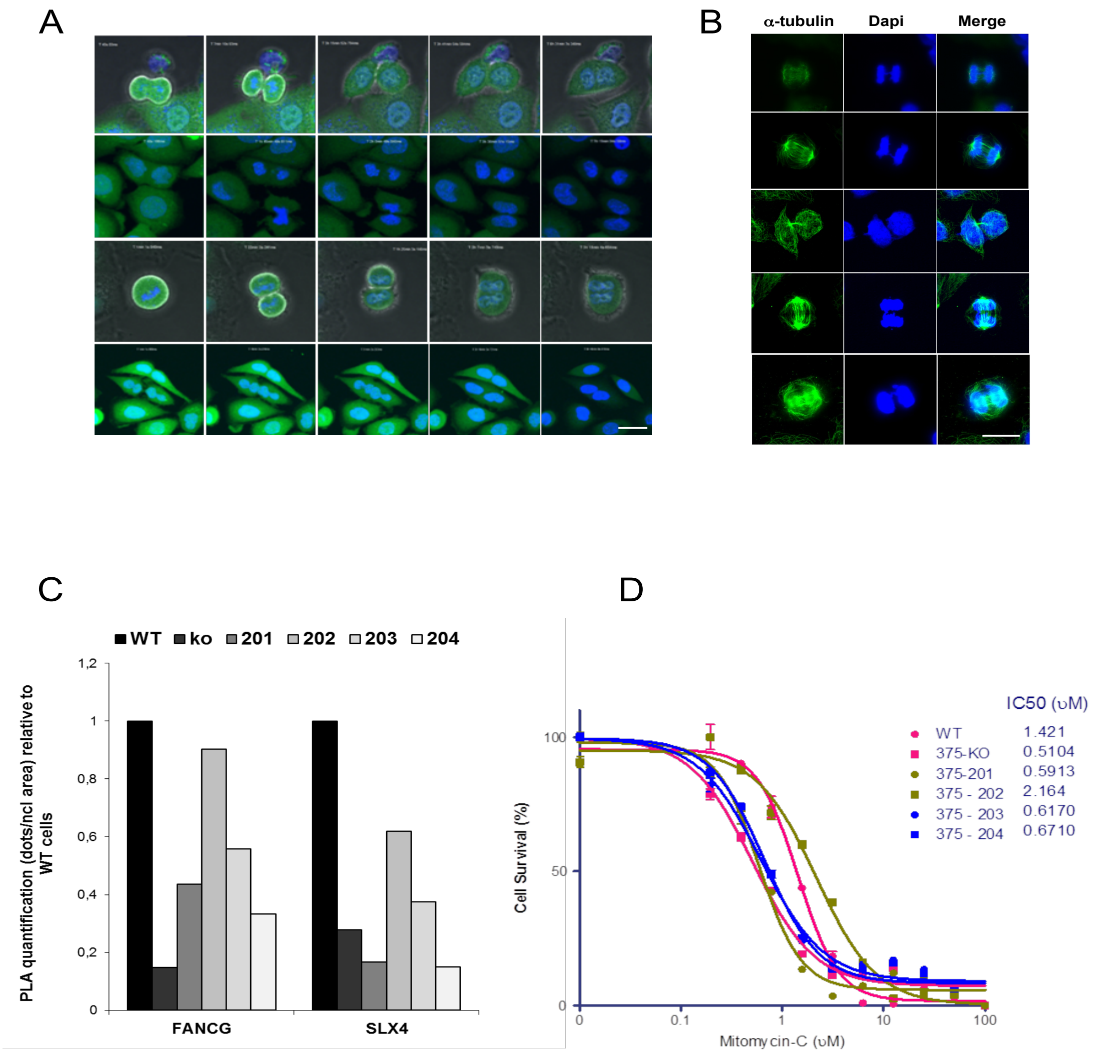


Figure 4. 14. The role of ERCC1 in interstrand cross-link repair is fulfilled by ERCC1 isoform 202 (Figure and legends reproduced from Friboulet et al, Cell Cycle 2013)

A) Time-lapse videomicroscopy was performed on ERCC1-deficient A549 cells maintained in growing medium containing tracker green and Hoechst 33342 used for cytoplasmic and nuclear staining, respectively. Representative snapshots of four failed mitosis are shown. Scale bar, 20 μm; B) Immunofluorescence images of ERCC1-deficient A549 cells stained for a-tubulin for visualisation of mitotic spindles and counterstained with DAPI. DNA bridges are visible in anaphase as well as during cytokinesis. Scale bar, 10 μm; C) ERCC1 interaction with the FA gene products FANCG and SLX4 was assessed in PLA. The PLA signal detected in wild-type A549 (WT), A549 knocked-down for ERCC1 (KO) and A549 expressing individually each of the four ERCC1 isoform (201, 202, 203 and 204) was quantified using the ImageJ software. The ratio of the signal in KO and single isoform-expressing A549 over WT cells was plotted; D) Wild-type A549 (WT), A549 knocked-down for ERCC1 (KO) and A549 expressing individually each of the four ERCC1 isoform (201, 202, 203 and 204) were treated for 48h with Mitomycin C or with vehicle as indicated. Cell viability was then assessed by WST-1 assay. Percentage of surviving cells was plotted and IC50 determined. Error bars indicate the SEM of triplicate measurements of a representative experiment.

4.2.4.7. Correlation between cisplatin sensitivity and olaparib sensitivity in non-isogenic NSCLC models

Since the use of isolated isogenic models does not recapitulate the impact of genetic heterogeneity on drug response, olaparib sensitivity was examined in the non-isogenic panel of 14 NSCLC cell lines (Chapter 3). Cisplatin sensitivity was assessed in parallel, as some determinants of cisplatin sensitivity (e.g. *BRCA1*/2, the *FANC* family, or *ATR*) have also been associated with olaparib sensitivity. Cells were plated at densities ranging from 300 to 1500 cells / well in a 96-well plate (to account for differences in growth rate) and exposed to olaparib or cisplatin for five days. Cell viability was then assessed using the CellTitre-Glo® luminescent viability assay (Promega). When comparing the olaparib sensitivity of NSCLC models to the expression of ERCC1 (as detected by western blotting), no correlation between reduced ERCC1 expression and increased olaparib sensitivity could be detected (data not shown). However, the examination of cisplatin sensitivity in the same NSCLC cell line panel revealed that cisplatin sensitivity was significantly correlated to olaparib sensitivity (r^2=0.5409, $p<0.05$, Pearson r correlation; Figure 4. 15), despite the absence of correlation to the level of ERCC1 protein (Chapter 3). This added to the list of previously described common determinants of PARP inhibitors and platinum sensitivity, and suggested that the mechanism of action of PARP1/2 inhibitors might be closer than previously thought to the one of platinum salts.

4.2.4.8. A model for ERCC1-deficient NSCLC sensitivity to PARP1/2 inhibitors

The mechanistic dissection of PARP1/2 inhibitor sensitivity in ERCC1-deficient NSCLC suggested the following: (i) ERCC1-deficient NSCLC cells are not profoundly deficient in terms of RAD51 foci response or BRCA1/BRCA2 expression; (ii) ERCC1-deficiency is epistatic with HR gene silencing in terms of PARP1/2 inhibitor sensitivity; (iii) γH2AX foci resolution in response to PARP1/2 inhibitors is delayed in ERCC1-deficient cells; (iv) persisting DNA damage is observed in ERCC1-deficient cells compared to wild-type counterparts following PARP1/2 inhibitor exposure, which results in a delay in cell cycle progression; and (v) silencing of PARP1 prior to PARP1/2 inhibitor treatment is able to minimise the ERCC1-selective effects of PARP1/2 inhibitors.

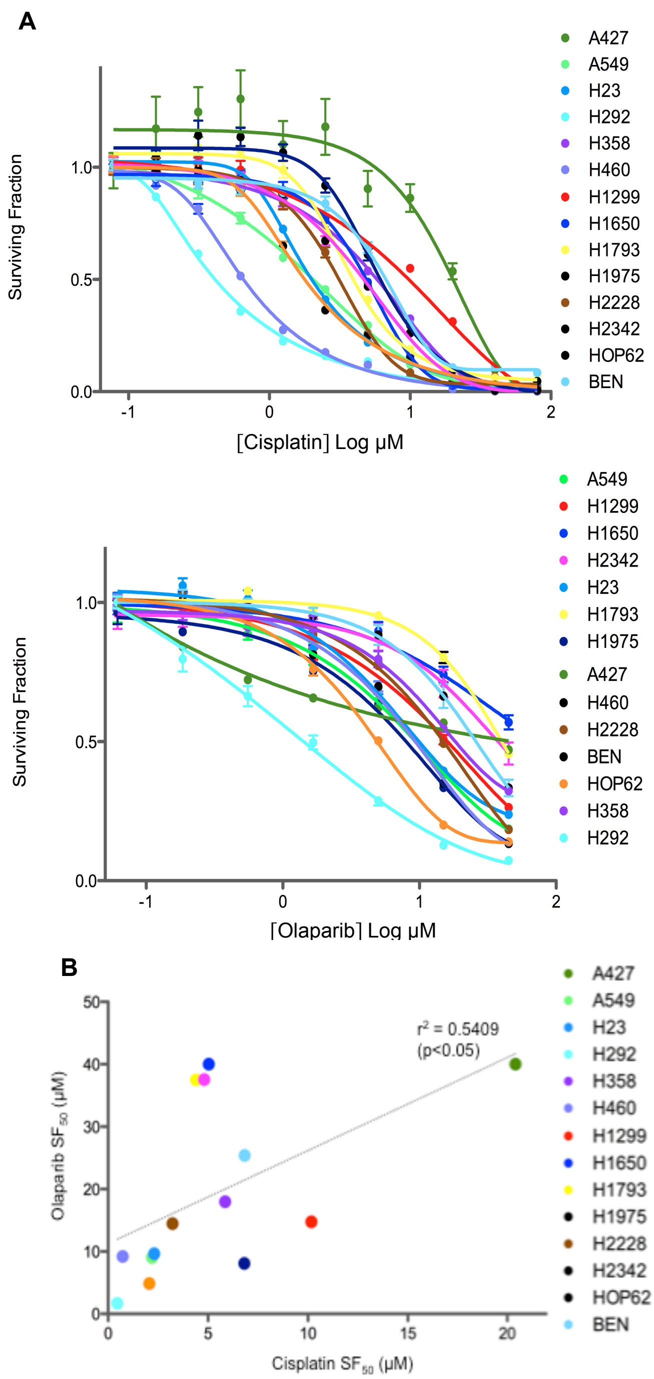


Figure 4. 15. Correlation between cisplatin sensitivity and olaparib sensitivity in a panel of 14 non-isogenic NSCLC cell lines.

A) Cisplatin and olaparib sensitivity of the panel of 14 non-isogenic NSCLC cell lines. Cells were exposed to the drug for five days before assessing cell viability. Error bars represent the standard deviation from the mean of three independent experiments. B) Significant correlation (Pearson r^2=0.5409, $p<0.05$) between cisplatin SF_{50} and olaparib SF_{50} (evaluated in a short-term assay) in a non-isogenic panel of 14 NSCLC cell lines.

Although a number of possible scenarios might explain these observations, the following proposed working model may be most consistent with the data (Figure 4. 16): (i) PARP1 binds DNA in response to a commonly occurring DNA insult but in the presence of a catalytic inhibitor, is trapped onto DNA (Figure 4. 16. A). This is consistent with recent data (Murai *et al.*, 2012) and the observation that silencing PARP1 by siRNA causes PARP1/2 resistance in ERCC1-deficient NSCLC cells (Figure 4. 11). (ii) When cells are in S phase, DNA trapped PARP1 stalls the oncoming replication fork (Figure 4. 16. B), and causes a γH2AX response (as demonstrated in Figure 4. 8). In some cases, fork arrest leads to replication fork collapse and formation of a DNA double-strand break (DSB) (Figure 4. 16. C, D), consistent with the formation of RAD51 foci in both ERCC1-deficient and proficient NSCLC cells (Figure 4. 3). As stalling occurs upstream of the DNA lesion, the creation of the DSB does not allow the removal of trapped PARP1. (iii) The DSB creation results in the formation of a branched structure on the 5' side of the DNA, thereby creating a substrate for the ERCC1/XPF DNA endonuclease, which excises and removes trapped PARP1 (Figure 4. 16. E-F). In the absence of ERCC1, the DNA lesion is presumably not processed past point E; cells remain trapped in S phase and display G_2/M arrest, the γH2AX response is still activated (Figure 4. 8, Figure 4. 9) and as DNA DSBs are particularly lethal, cells either die at this point or use alternative forms of repair that are presumably sub-optimal, thus impairing their overall fitness. (iv) In ERCC1-proficient cells, gap filling is performed after PARP1 excision by ERCC1/XPF via conventional DNA polymerases (Figure 4. 16. G), which generates a final substrate for HR (Figure 4. 16. H), eventually followed by replication fork restart and cell cycle progression. The necessity for ERCC1 activity on the PARP1/DNA lesion before HR can restore the replication fork is consistent with the epistasis observed between HR genes and ERCC1-deficiency in terms of olaparib sensitivity.

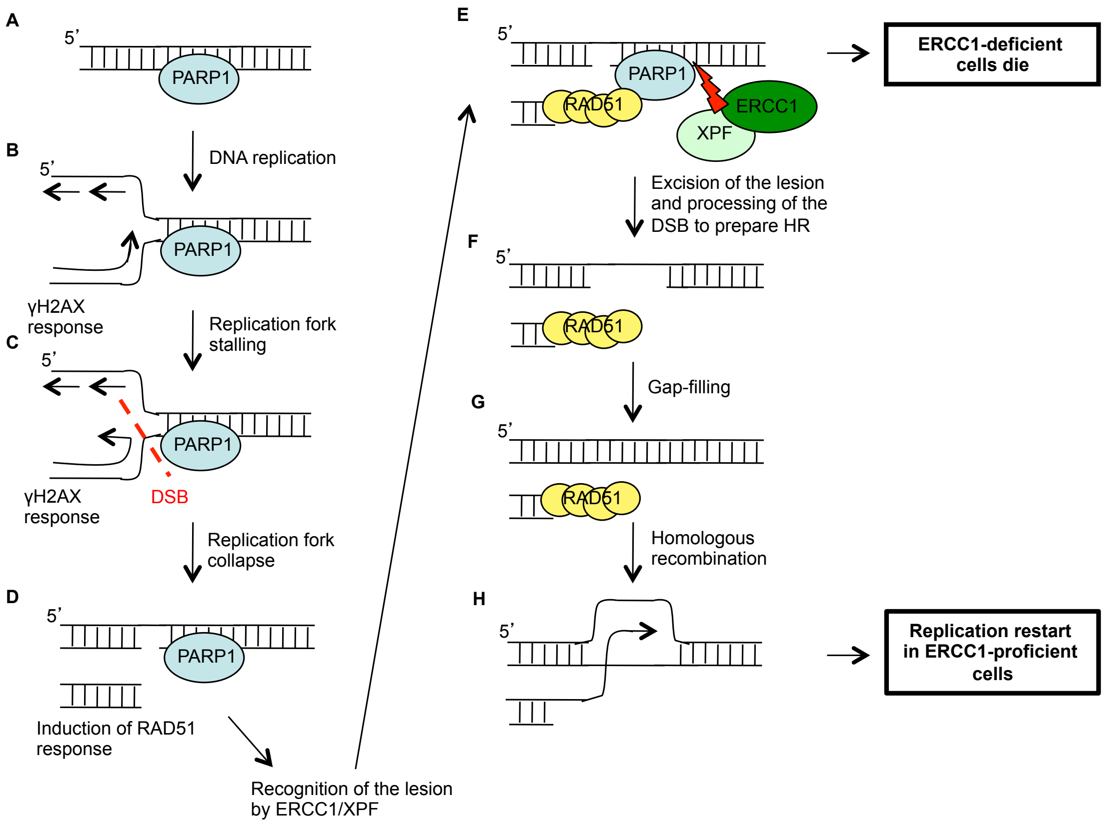


Figure 4. 16. Proposed model for explaining PARP1/2 inhibitors selectivity in ERCC1-deficient cells

A: PARP1 binds DNA in response to a commonly occurring DNA damage, for example following the formation of spontaneous single-strand break, but is trapped onto the DNA by the PARP1/2 inhibitor. B: During DNA replication, PARP1 bound to DNA causes stalling of the replication fork. C: This leads to fork regression and formation of a double-strand break, which is ERCC1-independent. The corresponding γH2AX response can be detected by the formation of γH2AX foci. D: The resulting structure creates a substrate for the ERCC1/XPF endonuclease. E: ERCC1/XPF removes the lesion while the DSB is processed to prepare for homologous recombination. In the absence of ERCC1, cells either die due to the toxicity of unresolved DSBs, or these DNA lesions are repaired by processes that ultimately impair cellular fitness. F: In ERCC1-proficient cells, the PARP1 lesion is removed, gap-filling is ensured by conventional DNA polymerase (G) and homologous recombination can then occur (H) allowing the restart the replication fork.

4.2.5. *In vivo* evaluation of olaparib sensitivity on ERCC1-deficient cells xenografts

On the basis of *in vitro* selective effects of PARP1/2 inhibitors on ERCC1-deficient cells, *in vivo* experiments were performed. Nude mice were injected with 10^6 A549 ERCC1-proficient or Ac216 ERCC1-deficient cells and effects of olaparib (50mg/kg) on tumour size and weight were examined. Briefly, three settings were evaluated to mimic different clinical settings: 1) a so-called “adjuvant” setting, where treatment by olaparib was started the day following the injection of the cells (Figure 4. 17. A); 2) a so-called “advanced” setting, where olaparib treatment was started when tumours reached a minimal volume of $80mm^3$ (Figure 4. 17. B); 3) a so-called “maintenance” setting, in which cisplatin (1mg/kg) was administered first followed by olaparib maintenance. Two maintenance experiments were performed: for the first one, four injections of cisplatin were arbitrarily administered to mimic the four to six cycles of platinum-based therapy administered to patients with advanced or metastatic NSCLC. As displayed in Figure 4. 17. C, the cisplatin sensitivity of ERCC1-deficient clones was so profound that tumours continued to shrink in a similar proportion in the control arm and in the treatment arm after platinum cessation. A second experiment, where a single injection of cisplatin was administered, was therefore set up (Figure 4. 17. D).

As shown in Figure 4. 17, none of these experiments provided promising results, and in vivo experiments were stopped at this point. Intriguingly, olaparib treatment even seemed to have a detrimental effect on ERCC1-proficient cells-derived xenografts (Figure 4. 17. A-B), which was not observed on tumours derived from ERCC1-deficient cells.

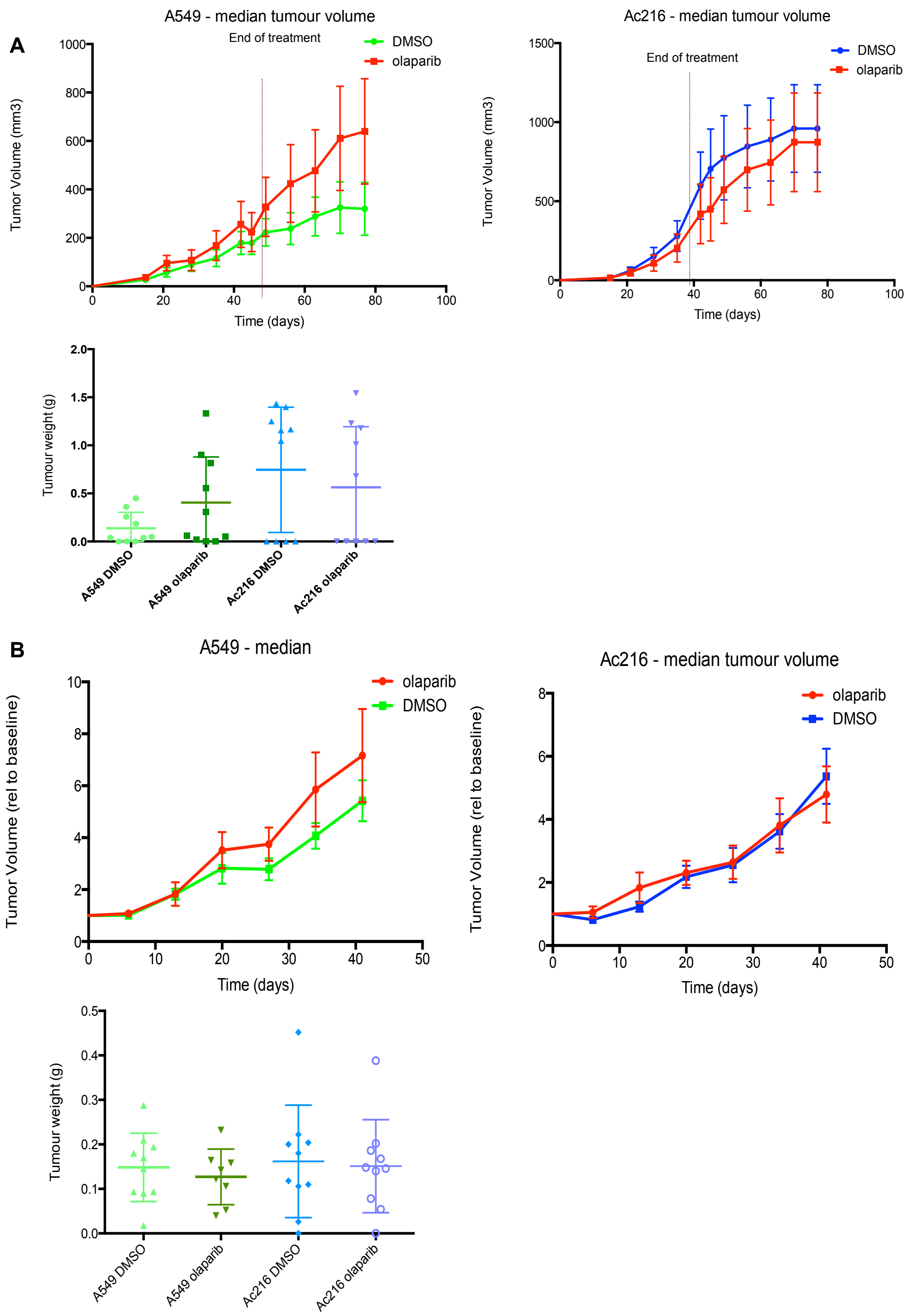


Figure 4. 17. In vivo experiments evaluating olaparib monotherapy in different settings (legend next page)

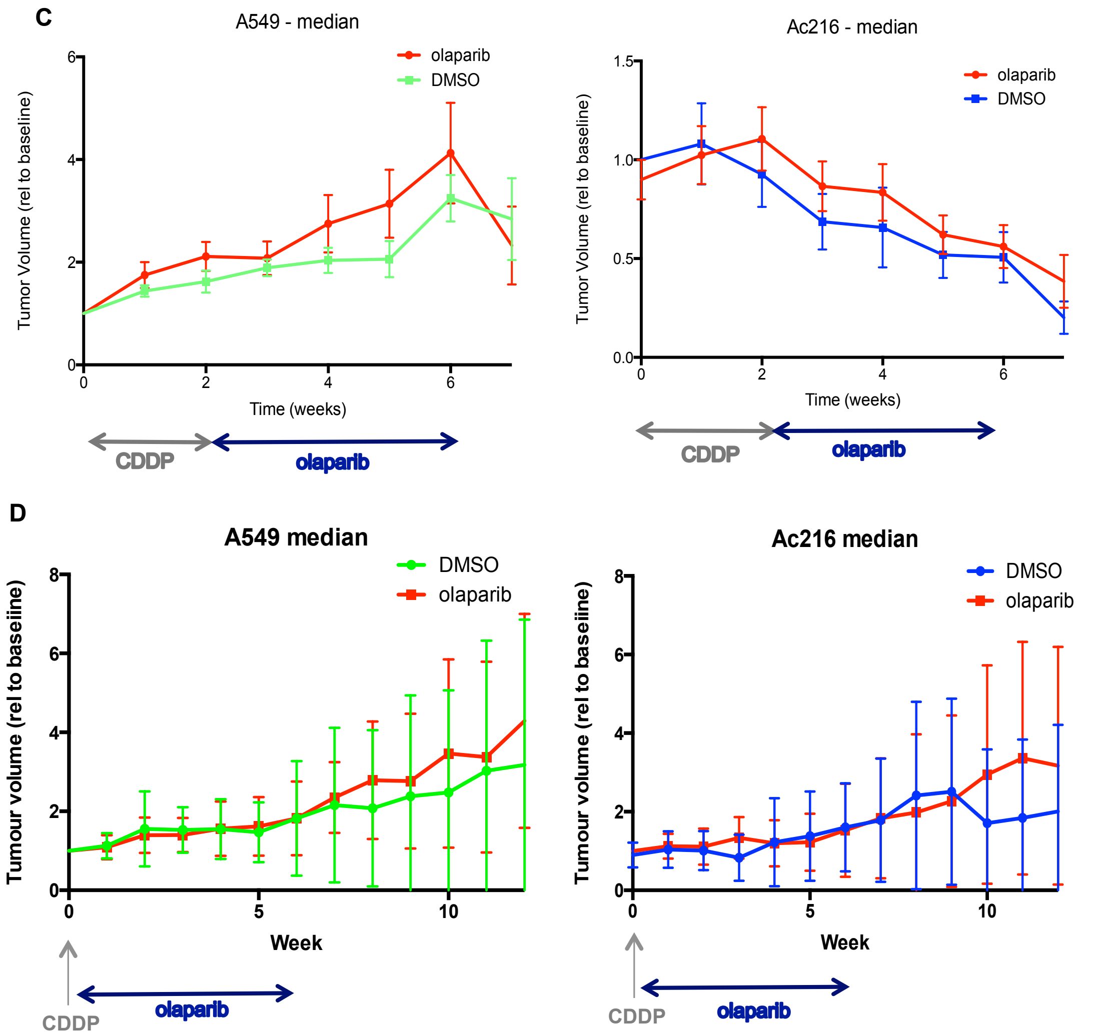


Figure 4. 17. In vivo experiments evaluating olaparib monotherapy in different settings (continued from previous page)

Tumor volumes over the course of treatment after 10^6 ERCC1-proficient or ERCC1-deficient cells were injected subcutaneously into nude mice. Total treatment duration was 6 weeks for all experiments. Tumors were measured weekly and olaparib or vehicle (DMSO) was injected intraperitoneally daily. The I bars represent 95% confidence intervals. A) Real tumour volumes (left and middle panels) and tumour weight (right panel) over the course of olaparib as adjuvant treatment. Drug was started before tumour establishment (pseudo “advanced” setting) and was administered at a dose of 50mg per kilogram of body weight. B) Relative tumour volumes (left and middle panels) and tumour weight (right panel) over the course of olaparib as palliative treatment. Drug was started when tumour reached a > $80mm^3$ volume (pseudo “metastatic” setting) and was administered at a dose of 50mg per kilogram of body weight. C) Relative tumour volumes over the course of olaparib as maintenance treatment. Drug was started when tumour reached a > $80mm^3$ volume. Treatment consisted of cisplatin 1mg/kg (twice weekly for a total of 4 injections) followed by olaparib 100mg/kg daily (pseudo “maintenance” setting). D) Relative tumour volumes over the course of olaparib maintenance treatment. Drug was started when tumour reached a > $80mm^3$ volume. Treatment consisted of cisplatin 1mg/kg (one single injection) followed by olaparib 100mg/kg daily (pseudo “maintenance” setting).

4.3 - Discussion

One major challenge in the era of personalised medicine is the identification of predictive biomarkers for drug response. ERCC1 expression has previously been correlated with cisplatin response in non-small cell lung cancer (NSCLC) and other tumour types (Lord *et al.*, 2002; Olaussen *et al.*, 2006; Steffensen *et al.*, 2009; Metzger *et al.*, 2010; Vilmar *et al.*, 2011; Yin *et al.*, 2011; Langer, 2012; Postel-Vinay *et al.*, 2012). The drug screen described in Chapter 3 identified PARP1/2 inhibitors as a potential novel therapeutic strategy for ERCC1-deficient NSCLC cells. In the present chapter, deeper investigation of these ERCC1-selective effects showed that ERCC1-deficient NSCLC cell line models were not only sensitive to a range of clinical PARP1/2 inhibitors, but also that ERCC1 isoform 202, the isoform that modulates cisplatin response, also caused PARP1/2 inhibitor resistance in NSCLC models. The epistasis between ERCC1 dysfunction and HR gene silencing in terms of PARP1/2 inhibitor sensitivity, together with the lack of a profound RAD51 dysfunction in ERCC1-deficient cells, suggested that the role of ERCC1 in the processing of PARP1/2 inhibitor-related DNA lesion might not be in HR itself but rather in the processing of the DNA lesion as a precursor to its final repair by RAD51 mediated-HR. Together with previous data suggesting the nature of DNA lesions caused by PARP1/2 inhibitors (Murai *et al.*, 2012), this allowed building a model in which PARP1 itself trapped on the DNA by PARP1/2 inhibitor might constitute a substrate lesion for ERCC1/XPF - prior to HR – which would cause the selectivity observed.

PARP1/2 inhibitors have shown remarkable activity in BRCA-deficient breast and ovarian cancers (Farmer *et al.*, 2005; Bryant *et al., 2005*). The present study provides evidence that PARP1/2 inhibitor selective sensitivity may not be limited to this population. Interestingly, ERCC1 also emerged as a determinant of PARP1/2 inhibitor sensitivity in a wide siRNA screen designed to identify modifiers or olaparib response, with a DE Z-score of -2.248 (data not shown). Findings presented in this chapter add *ERCC1* to the panel of clinically relevant DNA repair genes that modulate the cellular response to these agents, including *PTEN, ATM, ATR, CDK1, CHEK1, CHEK2,* and the *FANC* family of genes (Johnson *et al.*, 2011; Murai *et al.*, 2012; Rehman *et al.*, 2010; Turner *et al.*, 2008). The observation that only *ERCC1* isoform 202 was able to rescue the PARP1/2 inhibitor selective effect also suggests that the processing of PARP1/2 inhibitor-generated DNA lesions might be more similar to the molecular response to platinum adducts that previously thought (Dabholkar *et al.*, 1995; Sijbers *et al.*, 1996; Sun *et al.*, 2009; Friboulet *et al.*, 2013a). Moreover, the relative correlation observed in the non-isogenic panel of 14 NSCLC cell lines between cisplatin sensitivity and olaparib sensitivity also supports this hypothesis. Taken together, these observations support the proposition that platinum sensitivity could be a surrogate biomarker of PARP1/2 inhibitor sensitivity. Platinum administration has to be halted after a few cycles, and platinum-sensitive patients could benefit from "switch maintenance therapy" (i.e. introduction of a new agent following platinum-based therapy) in order to prolong tumour shrinkage. Given the excellent tolerability profile of PARP1/2 inhibitors as monotherapy, these agents could be evaluated as switch maintenance therapy in platinum-sensitive NSCLC patients. A clinical phase II protocol evaluating this hypothesis has been designed, and its realisation has been approved by AstraZeneca, the owner of olaparib (AZD2281), which is the most clinically advanced PARP1/2 inhibitor. The design of this protocol, which will be a collaborative study between Institut Gustave Roussy, the Institute of Cancer Research, and multiple other French and British centres, is summarized in Figure 4. 18. Importantly, the clinical protocol will be supported by a comprehensive translational research programme aiming at understanding the determinants of platinum and PARP1/2 inhibitors sensitivity in NSCLC, including Next Generation Sequencing (NGS), *in vitro* and *in vivo* work (Chapter 7). Besides the maintenance setting, PARP1/2 inhibitors as monotherapy could also be used as first line treatment (as an alternative to platinum) for NSCLC patients with ERCC1-deficient tumours who are not eligible for platinum-based treatments for reasons such as poor performance status or co-morbidities.

In many cases, demonstrating *in vivo* efficacy in a mouse model of cancer is a pre-requisite for clinical implementation. However, in this particular case, olaparib did not elicit an over anti-tumour response in ERCC1-deficient xenografted human tumour cells (Figure 4. 17). There might be a number of reasons explaining this. For example, it is possible that a more prolonged exposure *in vivo* is required to elicit ERCC1 selectivity than that used in the experiment described in Figure 4. 17. Furthermore, as only tumour volume and weight were assessed, it is possible that anti-tumour response did occur but might only have been detected at a cellular level, (i.e. tumour cell apoptosis, necrosis, etc.) and might not have been observable in terms of tumour volume and weight.

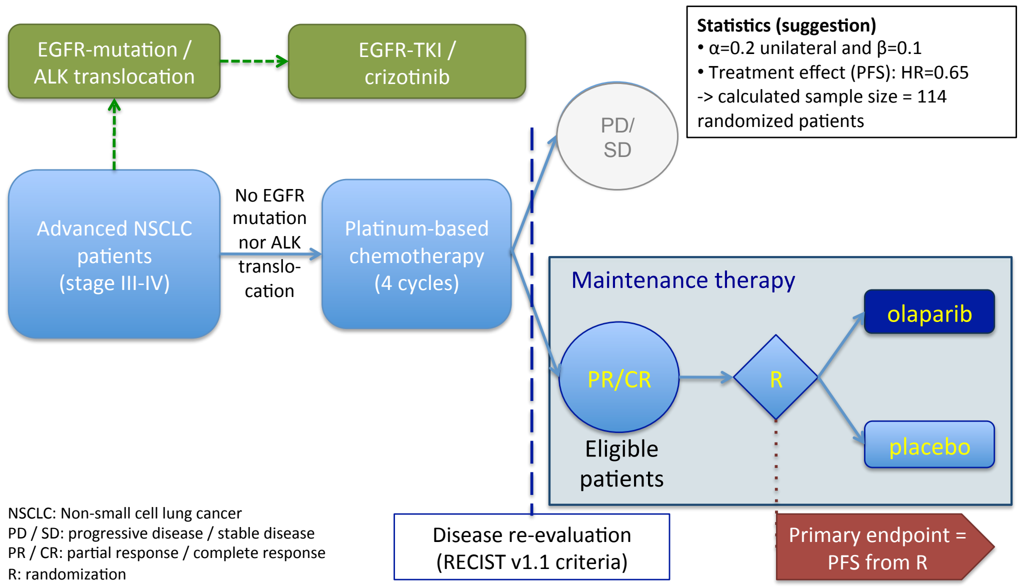


Figure 4. 18. Olaparib maintenance clinical trial - PIPSeN study design

Advanced NSCLC patients whose tumour do not harbor any EGFR-mutation nor ALK translocation will receive platinum-based therapy for 4-6 cycles as per institution's procedures. After tumour evaluation by RECIST v1.1 criteria, patients whose tumour shows partial or complete response to platinum therapy will be randomized between olaparib or placebo. Patients with progressive or stable disease following platinum therapy will be treated as per institution's procedures. The primary endpoint will be the true Progression Free Survival (PFS) from the time of randomisation. With a risk of first species (alpha) equal to 0.2, a risk of second species (beta) equal to 0.1 and an expected improvement in PFS of 35% in the treatment arm (hasard ratio of PFS equal to 0.65), the randomisation of 114 patients is required to complete the trial. Of note, the selection criteria for including patients into the randomised arm is not ERCC1 status, but platinum sensitivity, which is a surrogate biomarker of ERCC1 activity. Platinum sensitivity was chosen as selection criteria as (i) there is currently no reliable assay evaluating ERCC1 isoform 202 only or ERCC1 functionality, (ii) tumours that respond to platinum salt often present known or occult DNA repair defects that also account for PARP inhibitor sensitivity.

At present, such a model does not exist, but the limitations of using human tumour cell subcutaneous xenografts, which in this case does not model the site of tumour nor the impact of the previce tumour microenvironment, highlight the potential utility of more refined animal model systems.

The use of a relatively novel isogenic model of ERCC1 deficiency exemplifies the utility that such genetically controlled systems can have in the identification of synthetic lethalities. During the time of this PhD, Cheng *et al* independently reported the potential for using PARP1/2 inhibitors combined with platinum in ERCC1-low cells (Cheng *et al.*, 2012), using a non-isogenic panel of four NSCLC cell lines and two different PARP1/2 inhibitors – namely veliparib (ABT888; Abbott, USA) and olaparib (AZD2281; Astra-Zeneca, USA). This represents a different but complementary approach to the approach that has been taken here; isogenic models have the advantage of limiting the number of genetic changes between wild-type and mutant cells, so that differences observed can largely be explained by changes in the gene of interest. Although non-isogenic panels may better represent the impact of tumour genetic and epigenetic heterogeneity, the results from non-isogenic analyses are often more difficult to interpret, given the number of genetic variables in a non-isogenic cell line panel (Rehman *et al.*, 2010). Furthermore, a major pitfall - and challenge - in classifying cell lines according to ERCC1 status is the absence of reliable assay, as the strong similarity among all isoforms precludes the distinction between ERCC1-isoform 202 (the unique functional isoform) and other nonfunctional isoforms (Friboulet *et al.*, 2013a). Expression of nonfunctional isoforms can therefore result in misclassification, and thus the development of functional assays, such as the Duolink® technology that detects the ERCC1/XPF heterodimer, will be crucial to overcome this hurdle and create a meaningful classification of ERCC1 functionality. Further, other approaches looking for independent surrogate biomarkers of ERCC1 activity could be considered, and those that have been investigated during the course of the PhD will be presented in Chapter 5. With regards to the isogenic model, the experiment where PARP1/2 inhibitor sensitivity rescue is observed when re-expressing ERCC1 functional isoform 202 (Figure 4. 5) provides evidence that PARP1/2 inhibitor sensitivity is very likely a direct consequence of ERCC1 deficiency.

The mechanistic dissection of the sensitivity of ERCC1-deficient cells to PARP1/2 inhibitors revealed that ERCC1 deficiency was epistatic with HR deficiency towards PARP1/2 inhibitor sensitivity. Also, it was shown that ERCC1-deficient cells displayed a significant delay in DNA damage repair associated with a G_2/M cell cycle arrest following PARP1/2 exposure and appearance of radial chromosomal structures (Figure 4. 12) (in the most ERCC1-deficient cell line only, Ac216, for this last point). A similar observation was previously described in ERCC1-null myoepithelial fibroblasts and embryonic stem cells following Mitomycin C exposure in a study investigating the role of ERCC1/XPF in the removal of DNA interstrand-crosslinks (Niedernhofer *et al.*, 2001; Niedernhofer *et al.*, 2004). Together with the observation that *PARP1* silencing could rescue PARP1/2 inhibitor sensitivity (Figure 4. 11), as well as the ability of ERCC1 isoform 202 to repair interstrand crosslinks (Figure 4. 14) and restore PARP1/2 inhibitor resistance (Figure 4. 5), this suggests that ERCC1/XPF may be involved in the removal of a lesion constituted of PARP1 trapped onto the DNA by the PARP1/2 inhibitor (Kirschner *et al.*, 2010; Murai *et al.*, 2012). This working model is consistent with the recent description of the crystal structure of PARP1 bound to a DNA break (Langelier *et al.*, 2011; Langelier *et al.*, 2012): the major bulk created by trapped PARP1 may support the hypothesis that removing PARP1 from the damaged DNA strand is required for the DNA repair machinery to have access to the intact DNA strand. Furthermore, recent observations by Pommier and colleagues provide strong evidence for PARP1 "trapping" by PARP1/2 inhibitors (Murai *et al.*, 2012). Also, the limited double-helix distortion in the latter working model favours that PARP1/2 sensitivity is related to the role of ERCC1 in DSB repair rather than in NER.

In conclusion, high-throughput drug screens performed in an isogenic model of ERCC1-deficient NSCLC cell lines identified PARP1/2 inhibitors as being selectively toxic to ERCC1-deficient cells. ERCC1-selective effects of PARP1/2 inhibitors were consistently revalidated with several *in vitro* models using multiple PARP1/2 inhibitors, and a mechanistic explanation was proposed. Clinical trials in appropriately selected patients, associated with translational studies to further examine the determinants of PARP1/2 sensitivity in this context, are warranted.

CHAPTER 5

Biomarkers of ERCC1 functionality

5.1. Introduction

ERCC1 as a prognostic or predictive biomarker of patient survival or platinum sensitivity, respectively, has been evaluated at the constitutional DNA level (single-nucleotide polymorphisms), somatic tumour DNA level (mutations or epigenetic alterations), transcriptional level (RT-PCR) and protein level (immunohistochemistry) (reviewed in Postel-Vinay *et al.*, 2012; Besse *et al.*, 2013). Although the prognostic significance of ERCC1 has been consistently observed, study results regarding the predictive value of ERCC1 for survival and sensitivity to platinum-based therapy have been variable and sometimes conflicting. This led to a recent meta-analysis of 12 published studies in NSCLC including a total of 836 patients (Chen *et al.*, 2010), and in which ERCC1 status was analysed by IHC or qRT-PCR. This study reported that median survival was significantly longer in patients with low levels of, or negative for, ERCC1 expression (median ratio = 0.77; 95% CI 0.47–1.07; P <0.00001) and that response to platinum-based therapy was significantly higher in this population of patients (odds ratio = 0.48; 95% CI 0.35–0.64, P <0.00001). Several hypothesis have been formulated to support the observed discrepancies between studies, including the specificity of antibody used for IHC (Niedernhofer *et al.*, 2007; Olaussen *et al.*, 2007; Bhagwat *et al.*, 2009b; Olaussen *et al.*, 2010; Bepler *et al.*, 2011), the choice of the threshold for defining the ERCC1-low or ERCC1-high population, and the absence of correlation between the RNA and protein levels. As an illustration, the LACE study and repeated staining of samples from the IALT-Bio trial – which initially strongly established the predictive value of ERCC1 for response to platinum therapy in the adjuvant setting - did not manage to confirm the initial results, possibly due to a change in the performance of the antibody batch (Friboulet *et al.*, 2013a). More importantly, the biological complexity of ERCC1 function had been initially underestimated. This almost certainly contributed to these contradictory results. Indeed, as described in the above chapters, *ERCC1* generates four isoforms by alternative splicing, only one of which – namely isoform 202 - has proven functionality in NER, DNA repair and cell cycle progression (see Chapters 3 and 4) (Dabholkar *et al.*, 1995; Friboulet *et al.*, 2013a; Friboulet *et al.*, 2013b; Sjibers *et al.*, 1996; Sun *et al.*, 2009). The high degree of sequence similarity between isoforms precludes from distinguishing between each other with the currently available tools -

RT-PCT, ddPCR, or immunohistochemistry – and does not allow the design of isoform-specific probes. Consequently, abundant expression of one of the non-functional isoforms could lead to misclassification of the tumour as ERCC1-positive. The absence of correlation between ERCC1 status and response to platinum therapy in the isogenic model of 15 NSCLC cell lines (Chapter 3) illustrates the difficulty of assessing ERCC1 status, even *in vitro* on cell lines.

Therefore, several hurdles were faced, which needed to be overcome before using ERCC1 routinely as a predictive biomarker for cisplatin-based therapy. First, the optimal methodology to ascertain ERCC1 levels had to be determined: although protein expression is a desirable end point for biologic significance, this is only relevant when protein levels reliably reflect protein functionality. Second, the threshold of ERCC1 deficiency that correlates with pathway dysfunctionality remained to be defined. Third, a customized assay able to measure isoform 202 specifically had to be designed. Based on this observation and the difficulty in fulfilling all these criteria, we sought to investigate whether a biomarker of ERCC1 functionality, i.e. an indirect measurement of ERCC1 activity, which would be routinely used in the clinical practice, might be found. Four parallel approaches were undertaken for this purpose (Figure 5. 1): (i) UV irradiation, to evaluate the NER pathway; (ii) whole exome sequencing, to look for an ERCC1-associated genomic scar at the DNA level; (iii) transcriptomic analysis, to investigate changes at the RNA expression level; and (iv) SILAC analysis, to compare proteomic profiles between ERCC1-proficient and ERCC1-deficient cells.

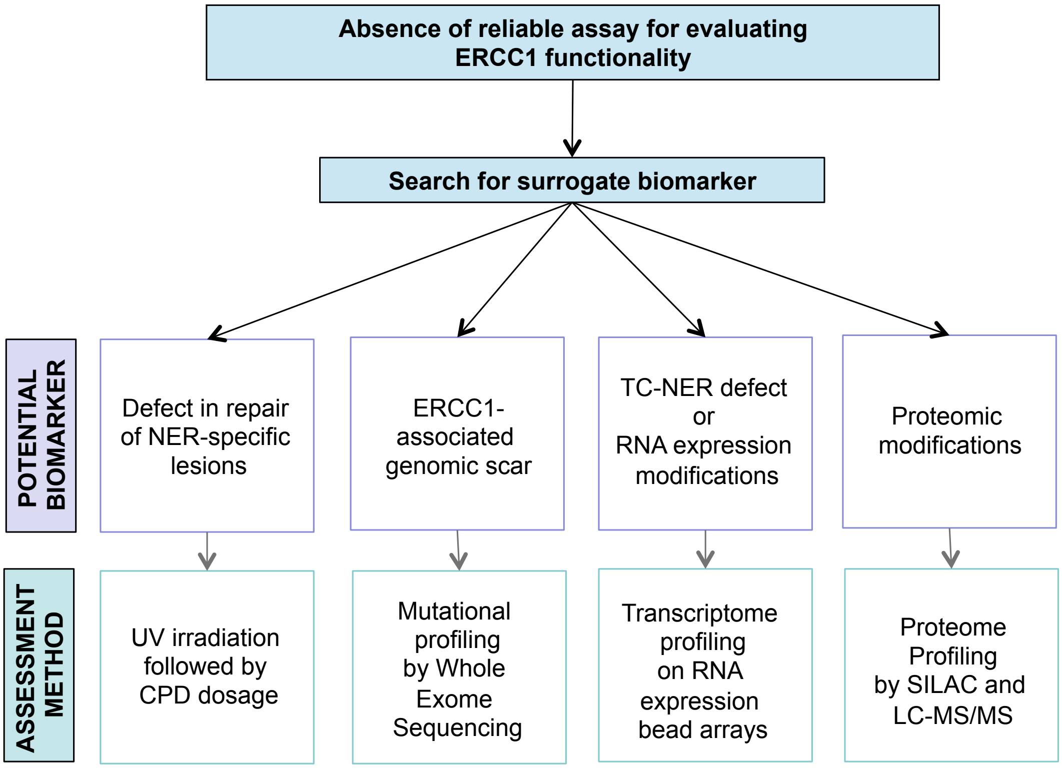


Figure 5. 1. Flowchart summarising the four different approaches to look for surrogate biomarker of ERCC1 functionality

Schematic representation of the four approaches undertaken in parallel to investigate the presence of a surrogate biomarker of ERCC1 functionality

5.2. Results

5.2.1. UV irradiation and cyclobutane pyrimidine dimers evaluation

ERCC1/XPF endonuclease activity is the rate-limiting step of the NER pathway. As such, an evaluation of NER functionality would reflect ERCC1/XPF activity and may represent a surrogate biomarker of the heterodimer ability to remove platinum-induced DNA adducts. An approach that has been used for several years to evaluate NER functionality is the quantification of the ability of the cell to remove DNA adducts after platinum exposure, by using specific anti-DNA adducts antibodies and quantifying the disappearance of the signal over time. For example, the monoclonal R-C18 antibody specifically recognises guanine-guanine intrastrand crosslinks induced by cisplatin exposure and can be used to quantify these lesions by dot blot on cultured cells or by immunofluorescence staining on tissues (Dzagnidze *et al.*, 2007; Friboulet *et al.*, 2013a; Nel *et al.*, 2013). This approach offers the advantage of directly studying the effects of cisplatin itself. However, it is also limited by several bias-introducing factors, such as the narrow therapeutic window and high variance of cisplatin IC_{50}, the variability of penetration of the drug into the cell, the presence of efflux or detoxification systems or other mechanisms of primary platinum resistance (Galluzzi *et al.*, 2012), or simply the inability of the antibody to access the cellular DNA. Other types of DNA damage repaired by NER are UV-induced lesions, including cyclobutane pyrimidine dimers (CPDs, the most abundant lesions), 4-6 photoproducts (4-6PPs) and Dewar valence isomers (Rastogi *et al.*, 2010). Contrary to 4-6PPs, the complete removal of CPDs by the NER machinery takes several hours (Pathania *et al.*, 2011), during which the differential cellular level of CPDs allows a distinction between NER-proficient and NER-deficient cells. Previous similar experiments have used X-Rays irradiation followed by the quantification of RAD51 foci in order to evaluate the functionality of the HR pathway (Willers *et al.*, 2009; Birkelbach *et al.*, 2013). By analogy, it was hypothesised that the quantification of CPDs after UV-irradiation may allow the assessment of NER functionality, and may not be subject to the bias and variability encountered with cisplatin treatment. Among all groups of UV radiation, UVB (280-315nm) produce the most adverse effects (Rastogi *et al.*, 2010) and have the highest tissue penetration capability. We therefore evaluated whether the ability of NSCLC cells to remove CPDs following UVB irradiation may be used as surrogate biomarker of ERCC1/XPF activity.

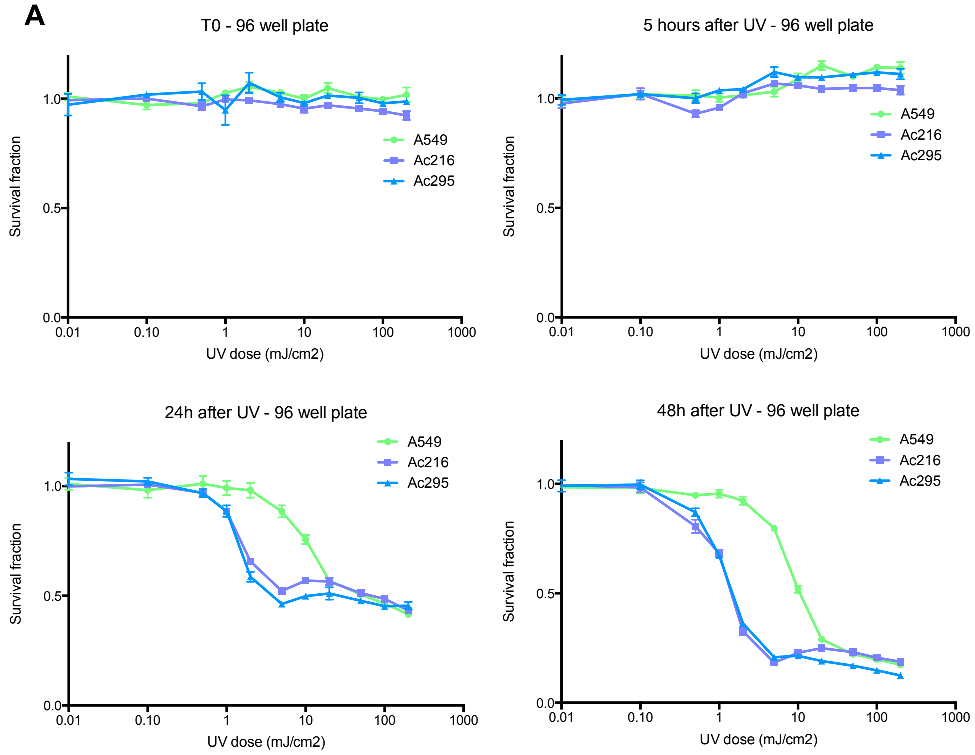
5.2.1.1. Selective sensitivity of ERCC1-deficient cells to UVB irradiation

To investigate whether ERCC1-deficient cells were sensitive to UV irradiation, i.e. a phenotype of NER deficiency, the ERCC1-isogenic panel of cell lines was exposed to UVB irradiation and cell viability was assessed at several time points after exposure. Contrary to X-rays, UVs are not able to penetrate tissue or material more than a few hundred nanometres (Chadwick *et al.*, 1995) and are stopped by plastic edges of culture plates. The pilot experiment was therefore initially performed in two formats: 96-well plate and 24-well plate in order to assess the possibility of decreased sensitivity of the assay due to edge effects limiting penetration of the UVs. Cells were plated at appropriate density to reach 75% confluence at the time of irradiation, which was performed using a Stratalinker® 1800 at doses ranging from 0.01 to 200mJ/cm^2. Cell viability was assessed CellTitre-Glo® luminescent viability assay at different time points after irradiation (Figure 5. 2). Dose-response curves analysed at 24 and 48h after irradiation revealed a significant decrease in ERCC1-deficient cells viability as compared to their ERCC1-proficient counterparts for doses ranging from 0.1 to 50mJ/cm^2. This difference was observed in both plate formats.

5.2.1.2. ERCC1-deficient cells accumulate CPDs after UVB irradiation

In order to investigate whether ERCC1-deficient cells were not only more sensitive to UVB irradiation but also developed more UV-induced lesions, a quantification of the number of remaining CPDs at different time points after irradiation was performed. Cells were plated in 96-well plate at appropriate density to reach 75% confluence at the time of irradiation, which was performed using a Stratalinker® 1800 at doses ranging from 0.01 to 20mJ/cm^2. Higher doses were not used as ERCC1-deficient cells displayed an approximate 100% cell death rate at 48h after irradiation with doses above 5mJ/cm^2. Quantitative evaluation of CPDs was performed by ELISA (CycLex Cellular UV DNA-Damage Detection Kit®) at 4h and 12h after irradiation and cell viability was concomitantly assessed using CellTitre-Glo® luminescent viability assay. The total number of CPDs detected at both time points at any dose was comparable between each cell line (Figure 5. 3). CPD raw values were subsequently normalised to cell viability to account for the excess of cell death of the ERCC1-deficient group at doses above 0.1mJ/cm^2 (Figure 5. 3). After adjustment for cell viability, ERCC1-deficient cells displayed a significantly higher number of CPDs per cell than their ERCC1-proficient isogenic counterparts.

A



B

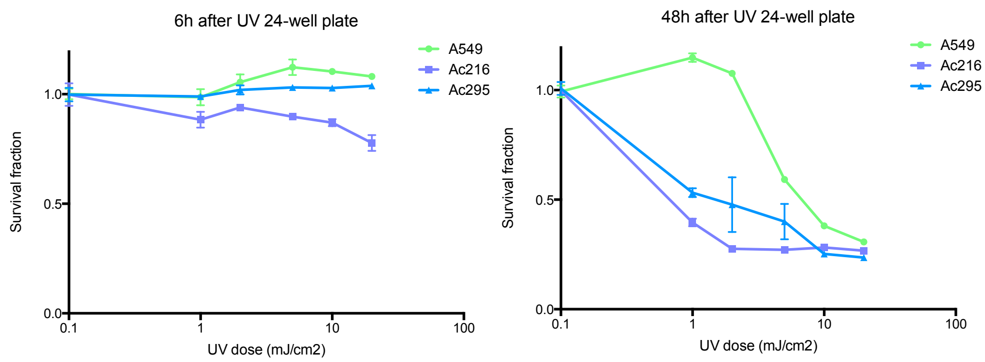
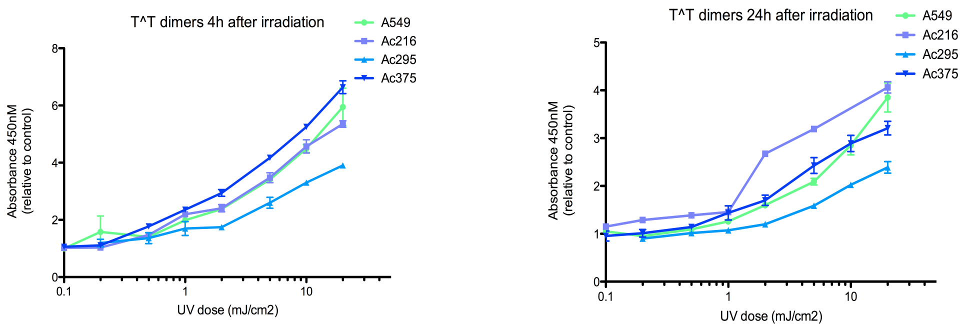


Figure 5. 2. Selective sensitivity of ERCC1-deficient cells to UV irradiation

ERCC1-proficient (A549) and ERCC1-deficient (Ac216 & Ac295) cells were plated at 1000 cells/well or 5000 cells/well in 96- and 24-well plates, respectively, prior to UVB irradiation at several doses (Stratalinker® 1800). Cell viability was assessed using CellTitre-Glo® luminescent reading at different time points after irradiation. Both assay formats were run in parallel in order to detect potential edge effects and UV penetration issues due to size and depth of the wells of the 96-well plates. A) Survival curves after UVB irradiation of ERCC1-proficient and ERCC1-deficient cells plated in 96-well plates. Time points: Top left graph = time of irradiation; top right graph: five hours after irradiation; bottom left graph = 24 hours after irradiation; bottom right graph = 48 hours after irradiation. B) Survival curves after UVB irradiation of ERCC1-proficient and ERCC1-deficient cells in 24-well plates. Time points: left graph = six hours after irradiation; right graph = 48 hours after irradiation.

A



B

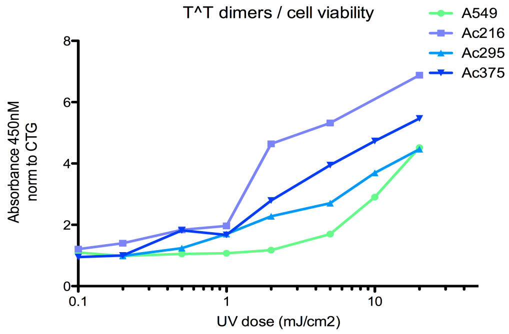


Figure 5. 3. ERCC1-deficient cells accumulate cyclobutane pyrimidine dimers after UVB irradiation

ERCC1-proficient (A549) and ERCC1-deficient cells (Ac216, Ac295 and Ac375) were exposed to UVB irradiation at several doses. Quantitative evaluation of pyrimidine dimers was performed by ELISA (CycLex Cellular® UV DNA-Damage Detection Kit) at various time points after irradiation. A) Quantification of the total number of pyrimidine dimers at four hours (left panel) and 24 hours after irradiation (right panel). Relative quantification was performed by calculating the ratio over baseline values (no UVB irradiation). No difference was observed between ERCC1-proficient and ERCC1-deficient cells. B) Quantification of the total number of pyrimidine dimers adjusted for cell viability (assessed by CellTitre-Glo® luminscent reading). Relative quantification was performed by calculating the ratio of the ELISA absorbance value (pyrimidine dimer quantification) over the CellTitre-Glo® luminescence value (viability measurement) at each UVB irradiation dose. After adjustment on cell viability, a higher number of pyrimidine dimers per live cell was observed in the ERCC1-deficient clones as compared to the parental cell line.

5.2.1.3. Detection of CPDs and ERCC1 by immunofluorescence

To be able to use UV-induced lesions on tumour cells as a surrogate biomarker of NER functionality, the detection of CPDs on UV-irradiated cells needed to be achieved by immunofluorescence, by analogy with the methods used to detect RAD51 foci on X-irradiated cells. In order to investigate the feasibility of this detection, cells were grown on coverslips and UV-irradiated at doses ranging from 0 to 20mJ/cm^2; cells were fixed at 24h after irradiation and the detection of CPDs and ERCC1 was performed by confocal microscopy after immunostaining. CPDs could be detected as a nuclear continuous staining (Figure 5. 4, green), which increased in intensity with increments in irradiation dose. ERCC1 was detected on ERCC1-proficient cells as nuclear punctiform staining, which did not vary in intensity with the irradiation dose (Figure 5. 4, red).

When considering only CPDs detection, an enhancement in signal intensity could be observed for ERCC1-deficient clones at doses above 5mJ/cm^2, when compared to their ERCC1-proficient isogenic counterpart (Figure 5. 5). No background CPD staining was observed in absence of UVB irradiation. As UVBs do not penetrate tissues over a few hundred nanometres and considering that the penetration depth varies according to the wavelength (Chadwick *et al.*, 1995), we aimed at performing a pilot experiment on ERCC1-isogenic NSCLC-derived xenografts. Mice tumours arising from A549 ERCC1-proficient and Ac216 ERCC1-deficient cells were harvested, irradiated at various doses ranging from 5 to 200mJ/cm^2 and oriented so that the irradiated surface could be easily identified. After 4h incubation in 10%FBS supplemented DMEM at 37°C in a 5%CO_2 atmosphere, tumours were snap-frozen in liquid nitrogen. Cryosections were performed, with notification of the orientation of the section, and fixed sections were subsequently stained for CPDs. Although the antibody used for detecting CPDs was a mouse primary antibody, it was hoped that the nuclear signal induced by CPDs could be differentially detected from the stromal background after addition of the secondary anti-mouse antibody. Unfortunately, the strong background staining of the tumour stroma did not allow a reliable detection of the CPD-induced signal, thereby precluding from further experiments using mice xenografts (Figure 5. 6).

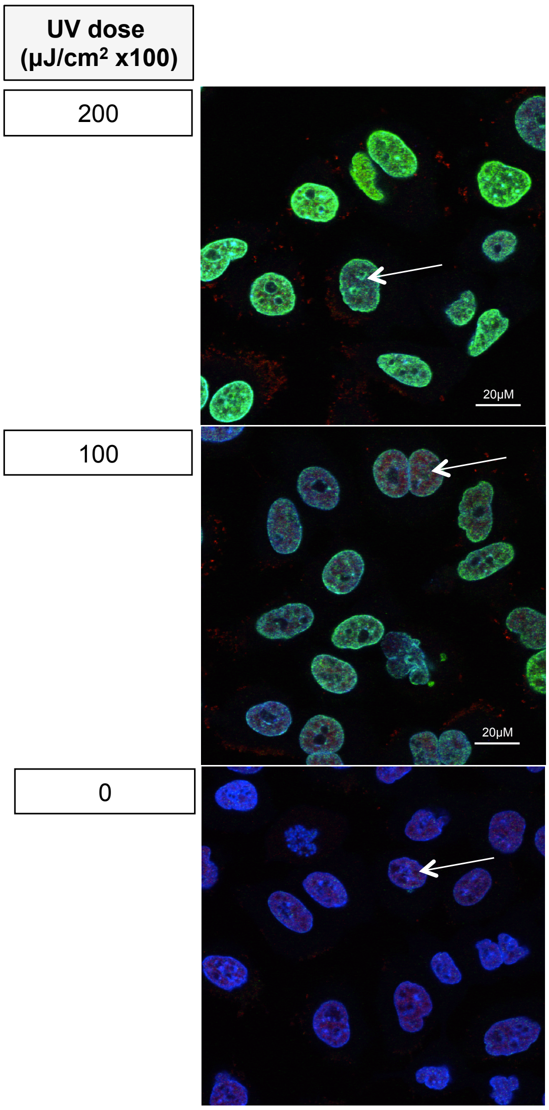


Figure 5. 4. Absence of increase in ERCC1 expression with UVB dose irradiation doses, as opposed to cyclobutane pyrimidine dimers

Representative confocal microscopy images of immunofluorescent detection of ERCC1 (555 wavelength, red staining) and cyclobutane pyrimidine dimers (CPD; 488 wavelength, green staining) in wild-type A549 cells. Nuclear DNA was counterstained with DAPI (Blue). No increase in ERCC1 expression (red staining; arrow pointing at representative examples at each dose level) was detected in UVB irradiated cells (100 and 200μJ/cm^{2}x100) as compared to baseline (no UVB irradiation, bottom picture). On the contrary, CPDs staining (green) increased with the dose. No specific pattern of foci or co-localization of ERCC1 and CPD stainings was observed.

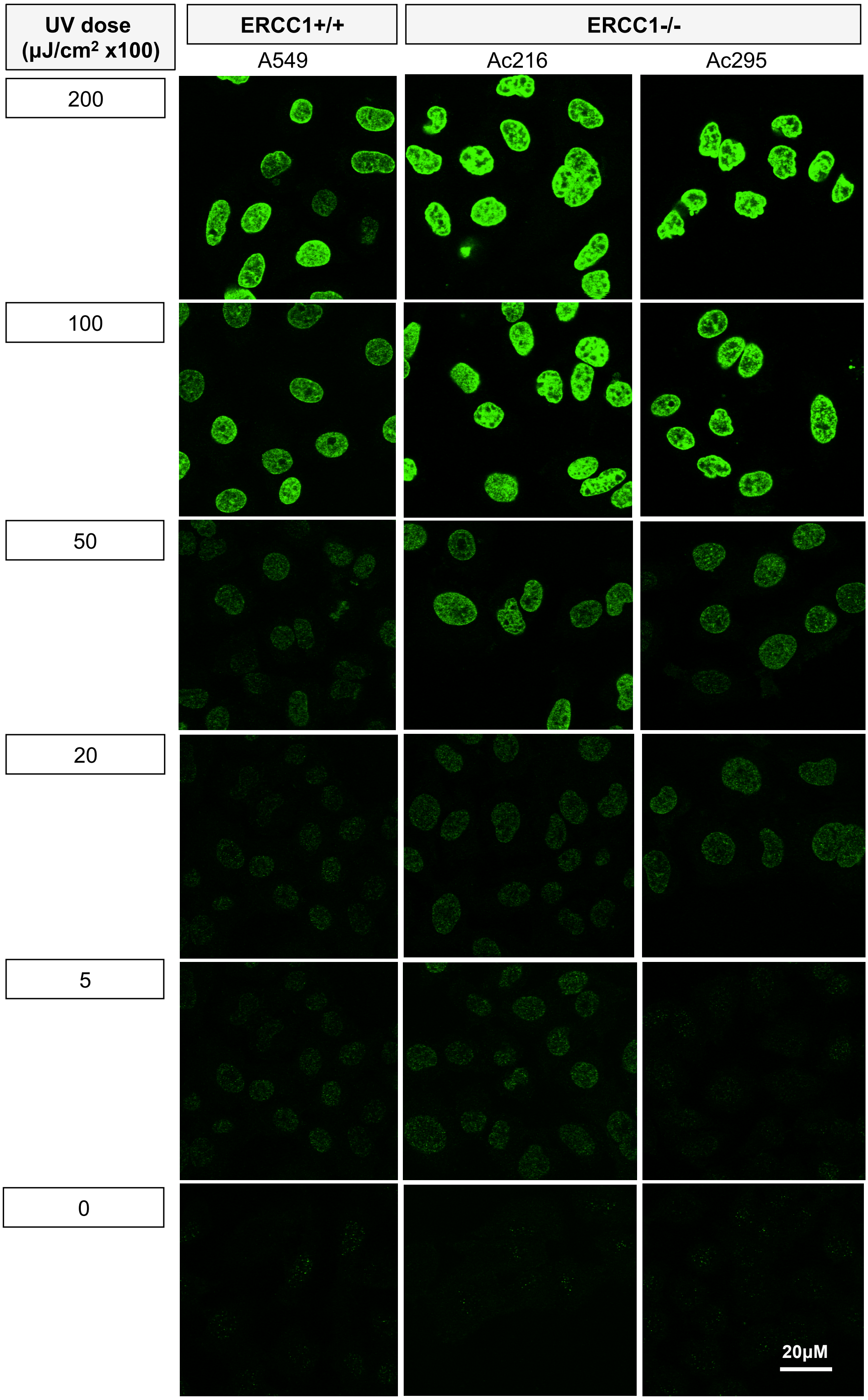


Figure 5. 5. Relative increase of cyclobutane pyrimidine dimers (CPDs) in ERCC1-deficient clones with UV dose irradiation (legend next page).

Figure 5.5. Relative increase of cyclobutane pyrimidine dimers (CPDs) in ERCC1-deficient clones with UV dose irradiation (figure previous page)

ERCC1-proficient (A549) and ERCC1-deficient (A216 and Ac295) cells were irradiated on coverslips in 6-well plates with increasing doses of UVB. Cells on coverslips were fixed 24h after irradiation and immunostaining of CPDs (488 wavelength, green staining) was then performed. Settings of the confocal microscope were first set at the maximum fluorescence (Ac216 at a 200µJ/cm^{2}x100 UV dose irradiation) and then maintained identical over all images to allow comparability of the signal acquisition and fluorescence intensity between pictures. ERCC1-deficient cells showed an increase in fluorescence (i.e. CPD staining) at doses of 100 and 50µJ/cm^{2}x100, when compared to their ERCC1-proficient counterpart.

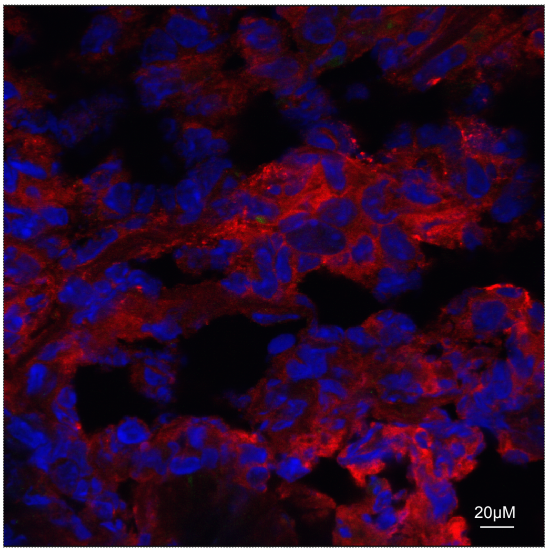


Figure 5. 6. Staining of A549 mice xenografts with CPDs after irradiation: failure of the pilot experiment

Prior to performing any experiment on human tissues, a pilot experiment on mice xenografts (A549 ERCC1-proficient and Ac216 cells) was attempted in order to evaluate the potential for UVBs to penetrate tumour tissues. Tumours were harvested, cut into two equal parts and irradiated with increasing UVB doses after registration of the irradiation orientation. Irradiated tumours were put in culture media (DMEM) and left incubating 37°C, 5% CO2) for various times. Tumours were then snap-frozen in liquid nitrogen, cut using a cryostat and stained for cyclobutane pyrimidine dimers. The use of a mice antibody for detecting CPDs precluded from further experiments using mice xenografts, because of the overcoming background staining of the stroma.

5.2.2. Exome sequencing: search for an ERCC1-associated genomic scar

DNA repair plays a crucial role in tumourogenesis and many tumours are associated with a long process of genomic instability causing abnormally high number of mutations or epigenetic changes. Specific patterns of somatic mutations in tumours, or genomic scars, have been previously associated with unique DNA repair defects (Pfeifer *et al.*, 2003; Pfeifer *et al.*, 2005; Stratton *et al.*, 2009; Pfeifer, 2010; Nik-Zainal *et al.*, 2012a; Alexandrov *et al.*, 2013; Lawrence *et al.*, 2013; Tomasetti *et al.*, 2013). While biomarkers currently used for treatment decisions almost exclusively are pharmacologically targetable drivers of the disease, genomic DNA damage caused by DNA repair deficiency could also be used as a biomarker to inform therapy. For example, microsatellite instability associated with mismatch-repair deficiency is a predictive marker of sensitivity to 5-FU (5-fluorouracil) and metothrexate (Hewish *et al.*, 2010). By analogy, a specific genomic scar resulting from homologous-recombination deficiency might be used to select patients that would benefit from PARP inhibitors (Lord *et al.*, 2012). It is now reasonable to envision the systematic characterization of tumour DNA repair defects and the use of this information to personalize treatments. As ERCC1 is the rate-limiting factor of the NER pathway, it was hypothesized that ERCC1-deficiency could be associated with specific genomic DNA fingerprints that would result from NER deficiency. Whether such an ERCC1-deficiency associated signature could be used as a predictive biomarker of platinum sensitivity was studied in this chapter.

5.2.2.1. Exome sequening of ERCC1-isogenic model: overall design

In order to investigate the possibility for an ERCC1-deficiency associated genomic scar, DNA was extracted from all ERCC1-isogenic cell line models and whole exome was sequenced using an Illumina® HiSeq2000 after targeted exome capture. Following alignment to the reference genome (GRCh37) and removal of PCR duplicates, variant calling was performed using the GATK Broad Best pipeline v2 with standard settings (Figure 5. 7) (www.broadinstitute.org/gatk/guide/best-practices). Annotation of the predicted mutation consequences was performed using the Ensembl variant effect predictor (Ensembl v61).

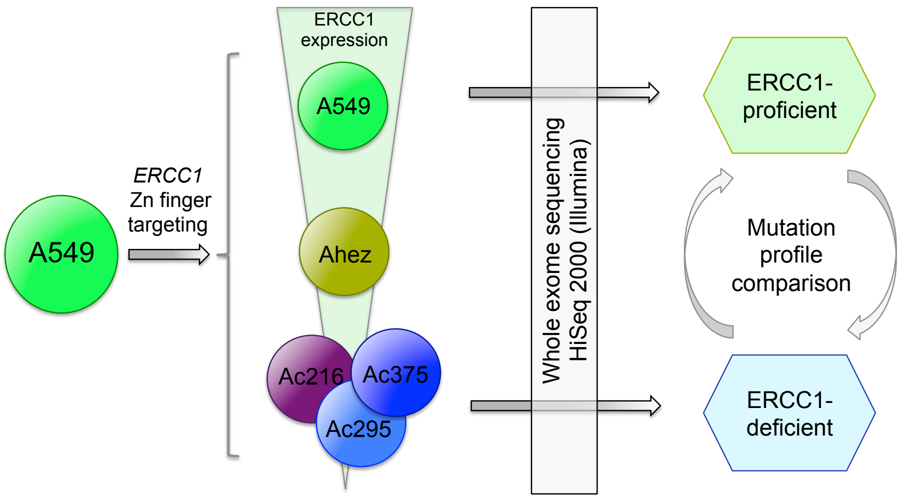


Figure 5. 7. Flowchart of whole exome sequencing of the ERCC1-proficient and ERCC1-deficient populations

Whole Exome Sequencing was performed on the ERCC1-proficient parental cell line and all ERCC1-deficient clones using targeted exome capture followed by high-throughput sequencing on Illumina® HiSeq 2000. BWA was used to align reads to the human reference genome (GRCh37). PCR duplicates were removed prior to further processing and variant detection. Variant calling was done using GATK Broad Best pipeline V2 with standard settings (www.broadinstitute.org/gatk/guide/best-practices). Variant calls were filtered using the following heuristic criteria: (i) variants called in regions not covered by the exome capture probes were excluded; (ii) variants marked as low quality (QUAL below 20) were excluded; (iii) variants with fewer than 10 reads covering the locus in all samples were excluded.

In order to focus on the consequences of ERCC1-deficency and by analogy with what would have been performed with a tumour sample, the parental ERCC1-WT A549 cell line was considered as “normal DNA” for the analysis, and the three ERCC1-deficient clones were labelled as “tumour DNA”. Accordingly, mutations “of interest” were variants detected at a variant allele frequency (vaf) > 0 in at least one ERCC1-deficient clone and absent (vaf=0) in the A549 DNA reference sequence. As the ultimate outcome of the present study was the definition of a genomic scar, all mutations were considered regardless of their consequence on the protein sequence.

5.2.2.2. Quality Control

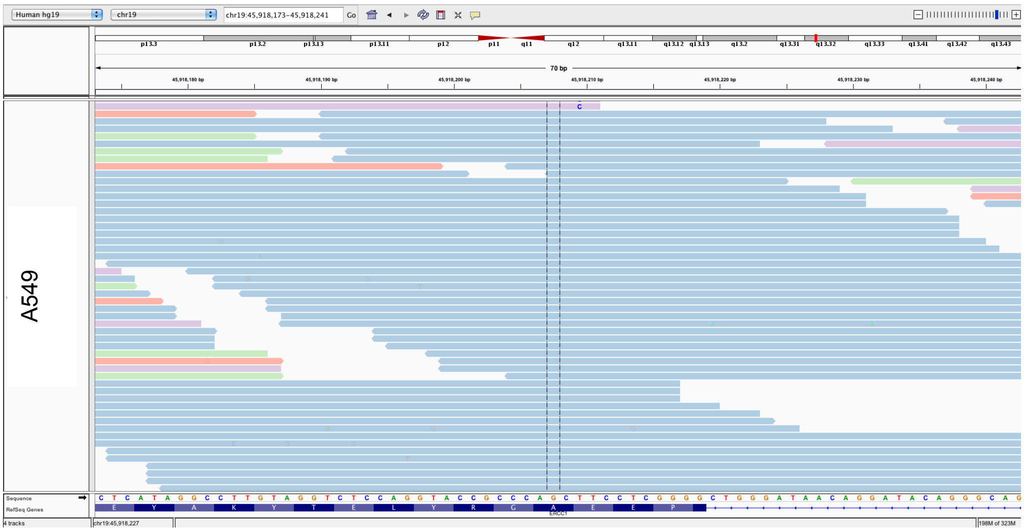
All samples were sequenced to a median depth over 30, with the A549 parental line being sequenced to a median depth of 61. The proportion of target regions covered by at least 10 reads was 88-92% (Figure 5. 8).
In order to check for the detection of expected mutations in *ERCC1*, the aligned short reads of this gene were visualised using IGV software. The region covering *ERCC1* where the ZFN-directed mutations were expected to occur is shown in Figure 5. 8. This inspection showed that the vast majority, if not all copies of *ERCC1* in these cell lines carry at least one mutation. The mutations in *ERCC1* reported by the GATK variant caller are listed in Figure 5. 8, together with the variant allele frequency at which the mutations were detected in each cell line. Two out of three mutations reported by GATK corresponded to the mutations identified by TOPO Cloning. One mutation reported by GATK in one clone (Ac295, at a variant allele frequency of 0.21) was not detected by TOPO Cloning, which may result from alignment errors by the GATK pipeline for very short deletions.

A

	A549	Ac216	Ac295	Ac375
Percent duplicates	10.0	10.4	8.6	10.7
Median depth	61	32	36	35
Percent coverage at depth of at least 10	92	88	90	90

B

B.1



B.2

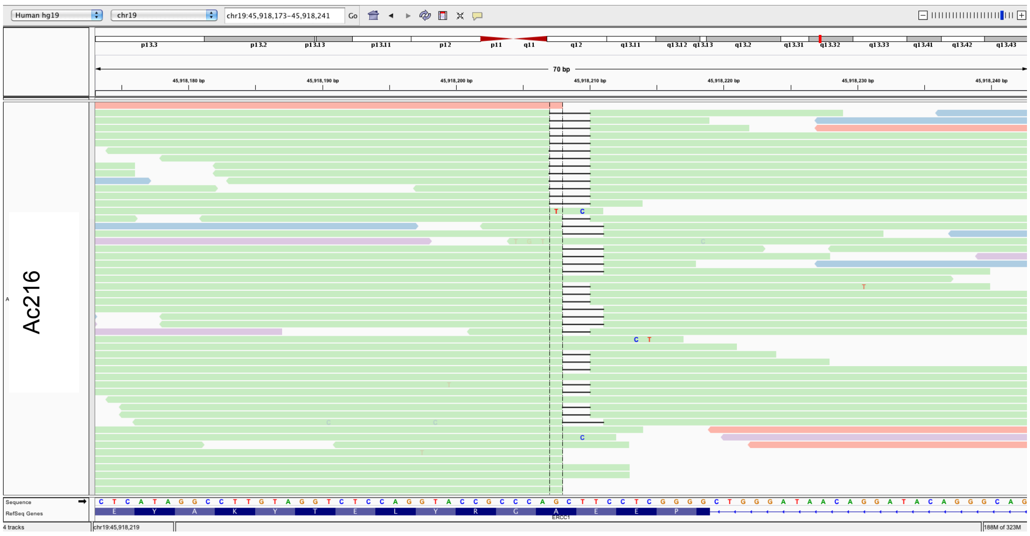


Figure 5. 8. Study of the expected mutations in the ERCC1-deficient cell lines (legend next page)

B.3

B.4

C

Mutation	Ac216	Ac295	Ac375
19:45918207 GCT/G	0.29	0.35	0.25
19:45918210 TC/T	0.00	0.21	0.00
19:45918206 AGCT/A	0.25	0.26	0.35

Figure 5. 8. Study of the expected mutations in the ERCC1-deficient cell lines

A) Summary of exome sequencing metrics; B) IGV was used to view pileups of short reads in *ERCC1* in each cell line. Mutations detected by NGS were located at the expected position, i.e. the ZFN-targeted sequence. Pileup of the short reads in the ZFN-targeted region of *ERCC1* in A549 ERCC1-proficient cell line (B1), Ac216 ERCC1-deficient clone (B2), Ac295 ERCC1-deficient clone (B3) and Ac375 ERCC1-deficient clone (B4). C) Summary of the three mutations detected in ERCC1-deficient cells and variant allele frequency at which each of them was detected. Two of the mutations reported by the GATK Unified Genotyper were consistent with those reported by TOPO Cloning®; one mutation reported by GATK was not detected by TOPO Cloning, which may be related in alignment errors from GATK when reporting very small deletions. Please note that the GATK reported mutations are on the forward strand, whereas those reported by TOPO Cloning® are on the reverse strand.

5.2.2.3. Overall description of the results

The allele frequencies for the variants detected in the parental A549 and ERCC1-null derived cell lines were estimated from the proportion of reads reporting the variants. Thus, a variant allele frequency of 0.0 (zero) would indicate an absence of reads reporting the variant and a value of 1.0 would indicate all reads reporting the variant in a given sample. Figure 5. 9. B displays the allele frequencies of all variants detected in the ERCC1-null cell lines (*y*-axes) *versus* the parental A549 cell line (*x*-axes). In each scatter plot shown in Figure 5. 9. B the majority of points fall on the diagonal line running between the origin and (1,1) where $x=y$, and are thus detected at similar allele frequencies in the parental and ERCC1-deficient cell lines, which confirms that these cell lines are indeed isogenic: the relative alignments of all points over this diagonal reflects the typical pattern of isogenic cell lines, where the majority of the genetic background is identical. Of greater interest are the large number of points representing variants that are either absent in the parental cell line or detected at an increased allele frequency in the ERCC1-null cell lines. Variants reported with an allele frequency of 0.0 in the parental A549 cell line are considered to be *de novo* mutations in the ERCC1-null cell lines (red dashed vertical box). A similar set of variants detected in the parental A549 cell line but not detected in the ERCC1-null cell lines exist and are referred to as lost (blue dashed horizontal box). A further two sets of variants were selected as either having increased allele frequency in the ERCC1-null cell lines (above red dashed diagonal line) or decreased allele frequency in ERCC1-null cell lines (below blue dashed diagonal line).

Figure 5. 9. C displays plots of the variant allele frequencies of all variants detected in each of the ERCC1-null cell lines. In contrast to the pattern observed in Figure 5. 9. B, the allele frequencies of variants compared between the ERCC1-null cells lines deviate less from the diagonal ($x=y$). This suggests that the variant calls in the ERCC1-null cell lines are more similar to one another than to the A549 ERCC1-proficient parental cell line.

When all mutations that were present in at least one clone but absent in the parental cell line (the "*de novo*" mutations) were displayed on a Venn diagram, the vast majority fell into the common centre of the diagram, suggesting that most of the *de novo* mutations were indeed shared by all clones (Figure 5. 9. D).

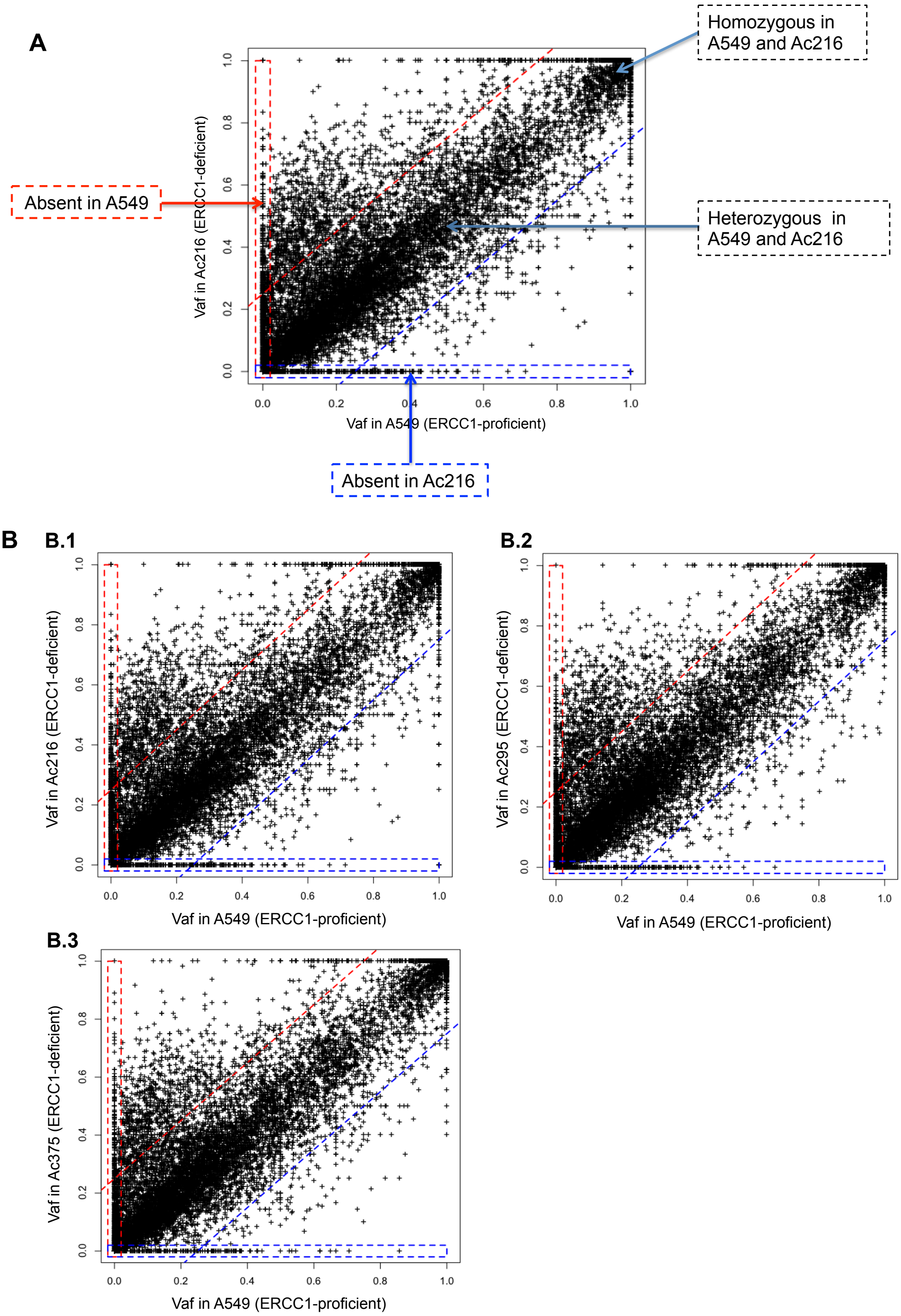


Figure 5. 9. Overall description of exome sequencing results of the ERCC1-isogenic model (legend next page)

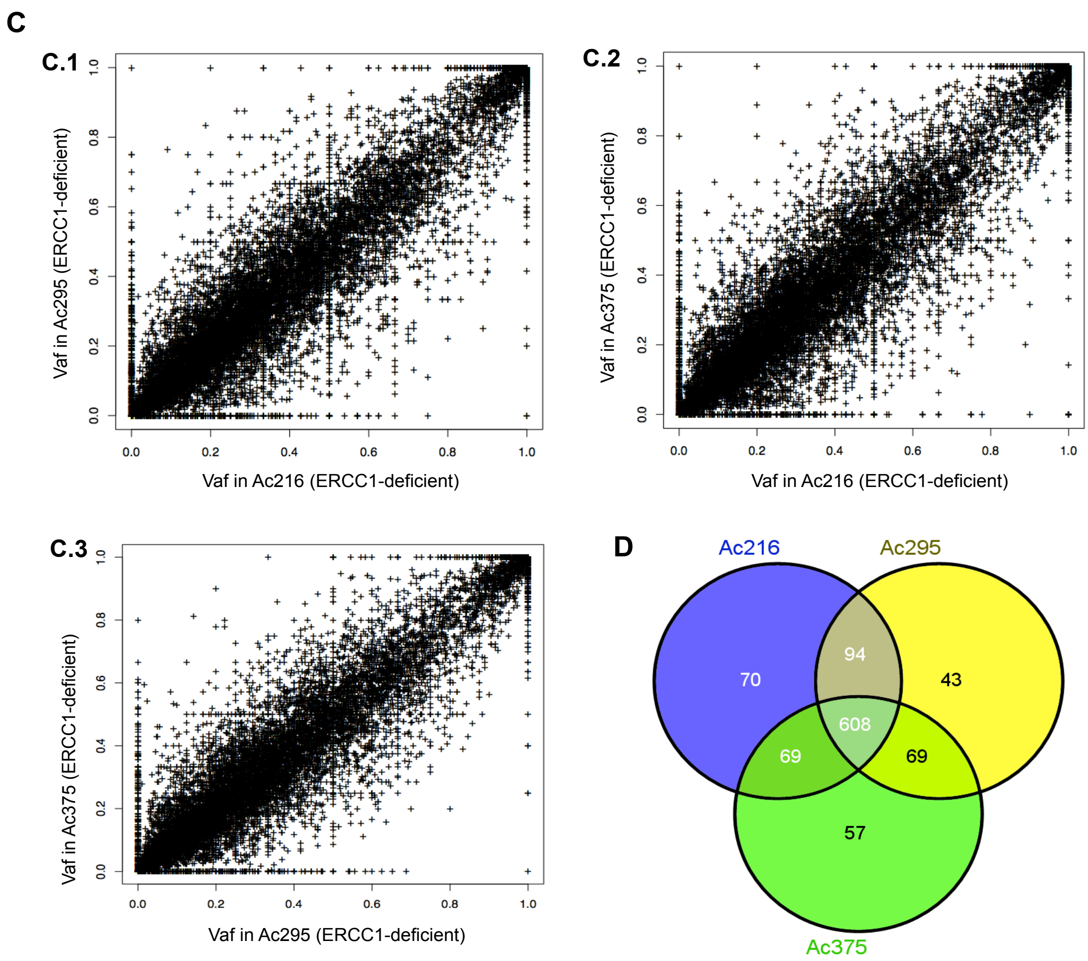


Figure 5. 9. Overall description of exome sequencing results of the ERCC1-isogenic model

A) Example graph of the comparison of allele frequencies of variants detected between all cell lines of the isogenic model. Allele frequencies of variants detected in the first cell line are displayed on the x-axis, whereas allele frequencies of variants detected in the second cell line are displayed on the y-axis. Dots on the diagonal indicate variants that are present at a similar allele frequency in both cell lines. Boxed regions indicate *de novo* (red) or lost (blue) variants. Diagonal dashed lines indicate the thresholds for calling increased allele frequency (red) or decreased allele frequency (blue). Here, most of the dots are located on the diagonal, which is a characteristic pattern from isogenic cell lines. Non-isogenic cell lines would display a "four clouds" pattern, with a cloud of dots located in each corner of the graph, as most allele frequencies for each variant would be different between cell lines. B) Comparison of A549 (*x*-axes) and the ERCC1-deficient clones (*y*-axes), with B.1) Ac216, B.2) Ac295 and B.3) Ac375. This pattern is representative of an isogenic model of cell lines, as most dots are located on the diagonal; however, variants presents only in the parental cell line, or in an ERCC1-deficient clone, can be detected (blue and red boxes, respectively). C) Comparison of allele frequencies detected in ERCC1-deficient cell lines. B.1) Ac216 compared with Ac295. B.2) Ac216 compared with Ac375. B.3) Ac295 compared with Ac375. The observation that the cloud of dots located on the diagonal in graphs displayed in B is wider than in C reflects the fact that ERCC1-deficient clones are more similar to another to their parental ERCC1-proficient cell line. C) Venn diagram of all mutations: Single Nucleotide Variants (SNVs), insertions and deletions.

5.2.2.4. Excess of A>T transversions detected in the ERCC1-deficient clones

As off-target effects of the Zinc finger nucleases could have introduced artefact insertions or deletions in the DNA of ERCC1-deficient clones, only Single Nucleotide Variants (SNVs) were considered for further analysis. Also, as most of the *de novo* mutations were shared by the three ERCC1-deficient clones – potentially derived from their common ERCC1-heterozygous ancestor - SNVs detected in these three cell lines were pooled so that each redundant mutation was counted only once; this avoided artificially enriching the pool of *de novo* mutations with SNVs that would have been counted two or three times.

The SNV mutation profile of the ERCC1-deficient clones revealed an excess of A>T (and reciprocal T>A) transversions, which was not detected on the reverse analysis (i.e. when considering SNVs present in the parental cell line and absent in all clones) (Figure 5. 10). This unusual mutation profile could also be observed when considering *de novo* SNVs that were shared by the three ERCC1-deficient clones or *de novo* SNVs shared by two ERCC1-deficient clones, but not when considering *de novo* SNVs observed in a single ERCC1-deficient clone only (Figure 5. 11. A-C).

T>A transversions are usually among the rarest transversion types in any tumour type and NSCLC cell lines usually harbour an excess of C>A transversions, which corresponds to DNA signature induced by tobacco exposure (Hainaut *et al.*, 2001; Pfeifer *et al.*, 2002; Rodin *et al.*, 2002; Pfeifer *et al.*, 2005; Alexandrov *et al.*, 2013; Lawrence *et al.*, 2013). In order to be able to compare this unusual mutation profile with other in-house generated NSCLC exome profiling, the exome sequencing of 14 NSCLC cell lines from the non-isogenic model was performed. Vcf files were generated using exactly the same methods (pipeline, quality criteria and filtering) as for the isogenic model dataset. For this non-isogenic cell line dataset, no “normal” DNA was available, rendering the definition of *de novo* SNVs more difficult. The decision of filtering of all SNVs listed in dbSNP and present at a minor allele frequency > 0 in the reference genome (GRCh37) was taken. *De novo* SNVs resulting from this analysis were classified according to the alteration type and pooled to create a unique dataset gathering data from the 14 NSCLC cell lines.

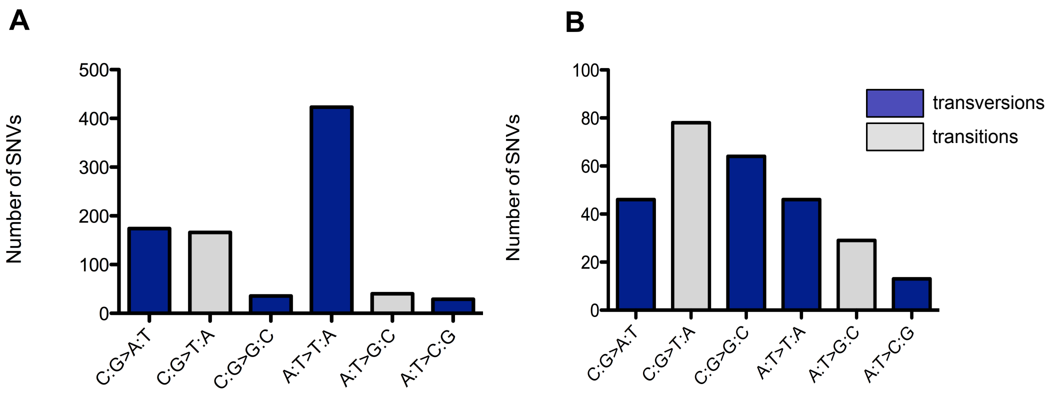


Figure 5. 10. Overall mutational profile of ERCC1-deficient clones

A) SNVs presents at a variant allele frequency (vaf) ≥ 0 in at least one clone and absent (vaf = 0) in the ERCC1-proficient parental cell line. When present in several clones, the SNV was counted only once. SNVs are classified by type, taking the forward strand as reference genome. B) Reverse analysis: SNVs present at a variant allele frequency (vaf) ≥ 0 in the parental cel line and absent (vaf = 0) in all ERCC1-deficient cells. SNVs are classified by type, taking the forward strand as reference genome. Mutational pattern in A showed an unusual excess of A:T>T:A transversions, which was not detected in the reverse analysis (B).

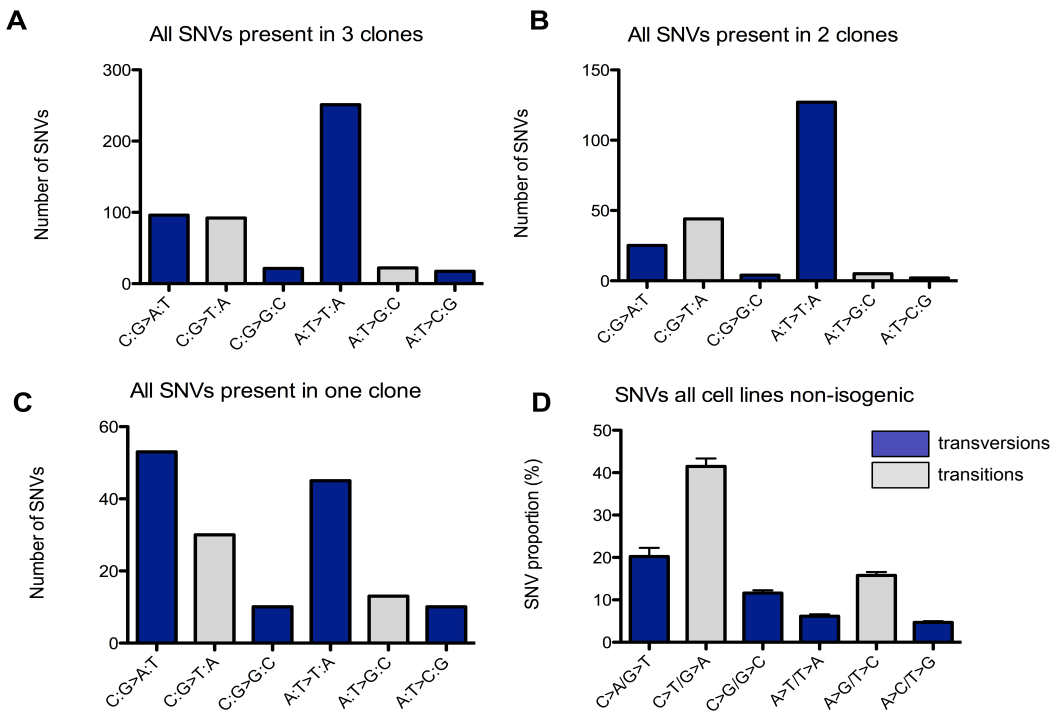


Figure 5. 11. Single nucleotide variant profile in the isogenic and non-isogenic models

A-C) SNVs presents at a variant allele frequency (vaf) ≥ 0 in at least one ERCC1-deficient clone and absent (vaf = 0) in the ERCC1-proficient parental cell line. If present in several clones, the SNV was counted only once. SNVs are classified by type, taking the forward strand as reference genome. A) SNVs detected in all three clones; B) SNVs detected in two ERCC1-deficient clones and absent in the parental cell line; C) SNVs detected in a single ERCC1-deficient clone and absent in the parental cell line. D) Overall SNVs mutational profile in the non-isogenic panel of 14 NSCLC cell lines. Are represented SNVs detected at a vaf ≥ 0 in any NSCLC cel line and absent (vaf = 0, not reported in dbSNP) in the reference genome. The transitions / transversion ratio is close to 3/1 and A:T>T:A transversions are extremely rare, which corresponds to what is expected for cancer cells.

As shown in Figure 5. 11. D, the SNV mutation profile obtained differed significantly from the one of the isogenic model, and corresponded to what would have been expected for a NSCLC exome sequencing profiling: (i) the ratio of transitions / tranversions was approximately equal to three; (ii) an excess of C>A likely resulting from tobacco exposure - the smoking status of the patients from which cell lines arose was available for only five of the 14 cell lines (4 smokers and 1 non-smoker) – was observed; and (iii) other mutation types represented a small minority of the SNVs (http://www.broadinstitute.org/gatk/guide/best-practices).

5.2.2.5. Specific context of occurrence of the A>T transversions

The knowledge of the previous and following base surrounding the lesion often allows a finer tuning of a signature definition and may bring information about the mechanism underlying the generation of the signature (Alexandrov *et al.,* 2013). In order to look for a specific context of occurrence of the A>T transversions, the 50 bases surrounding the SNV of interest were retrieved using an in-house designed Perl loop (www.perl.org/get.html). The resulting 50bp sequences – where position 25 corresponded to the mutation of interest - were aligned using Weblogo (Berkeley university, http://weblogo.berkeley.edu) to look for enrichment in specific bases at specific positions. This resulted in the identification of two specific profiles: on the forward strand, the 25bp preceding A>T transversions were significantly enriched in T repeats, whereas the 25bp preceding T>A transversions of the reverse strand were enriched in A repeats (Figure 5. 12. A). A search for kataegis (foci of localized hypermutation (Nik-Zainal *et al.*, 2012a)) in the ERCC1-deficient clones did not evidence the existence of such process in the absence of ERCC1 (Figure 5. 12. B).

Of note, the search for off-target effects of the zinc-finger nucleases (ZFN) was also performed using the same methodology: the 50 bp flanking small deletions or insertions (i.e. the type of lesions usually induced by ZFN) were aligned using Weblogo (http://weblogo.berkeley.edu). No consensus sequence logo could be found, and notably no consensus logo corresponding to the ZFN DNA-binding or DNA-cutting sequence could be detected (data not shown). This suggested that the proportion of off-target effects induced by the ZFN was minimal and undetectable by whole exome sequencing.

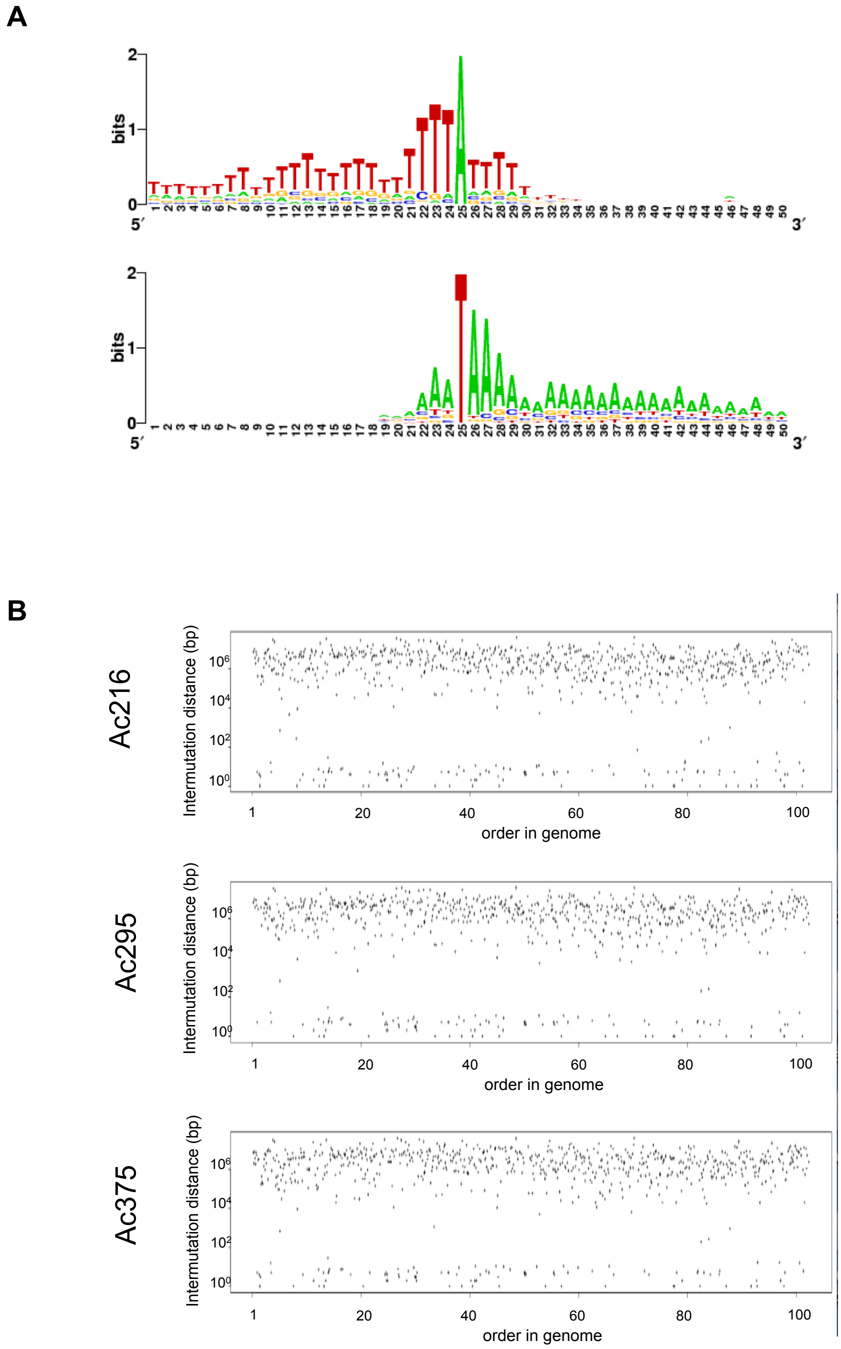


Figure 5. 12. Specific context of occurrence of T:A>A:T transversions

A) Weblogo of the 50 bases surrounding the A>T (upper panel) or T>A (lower panel) SNVs; the mutation of interest is at position 25 of the weblogo sequence. A>T transversions occur mainly after sequences enriched in T, whereas T>A transversions occur mainly after sequences enriched in A, mirroring what is observed on the complementary strand. B) ERCC1-deficient clones display no kataegis phenotype; each mutation is represented by a dot, with the Y axis being the distance (in bp) between two mutations, and the X axis being the position of the mutation in the genome. A typical kataegis pattern would appear like a continuous cloud of dots at the median inter-mutation distance (10^6 bp) associated with small “packets of rain” (i.e. several dots at a given genetic position for which the inter-mutation distance is much lower).

5.2.2.6. SNV analysis in the non-isogenic model

To investigate whether the potential genomic scar of ERCC1-deficiency detected in the isogenic model could also be found in the non-isogenic model, further analysis was performed on A:T>T:A transversions present in the exome sequencing of the 14 NSCLC cell lines. To focus on A:T>T:A transversions that would occur following T-rich (or A-rich for the reverse strand) sequences, the 50 bp flanking the SNV of interest were retrieved using a specifically designed Perl loop. A:T>T:A tranversions were then classified as “in context” according to the following criteria: any A>T transversion occurring either at TpTp<u>A</u>, TpNp<u>A</u> or NpTp<u>A</u> trinucleotides (where the underlined nucleotide is the mutated one) was considered as being “in context”; similarly, T>A tranversions were considered as being “in context” when they occurred either at ApAp<u>T</u>, ApNp<u>T</u> or NpAp<u>T</u> trinucleotides. Other A>T or T>A transversions were discarded. The proportion of A:T>T:A transversions occurring “in context” was subsequently calculated as the ratio of the SNVs of interest occurring “in context” over the total number of SNVs of the cell line. The search for a correlation with ERCC1 expression was performed using either a linear model of ERCC1 relative expression or by clustering cell lines into two groups (ERCC1-high *vs* ERCC1-low) (Chapter 3). Neither of these models showed any significant correlation (Figure 5. 13. A). Similar analysis investigating the correlation between the proportion of A:T>T:A “in context” and the sensitivity of the cell lines to cisplatin IC50 did not demonstrate any significant correlation (data not shown). Further, when pooling all A:T>T:A together, no specific context for the occurrence of this transversion was observed (Figure 5. 13. B). This suggested either that results observed in the isogenic model are the result of an artefact, or that larger panels of cell lines are required detected such signature in a non-isogenic model, given the rarity of the A:T>T:A transversions in cancer cell lines.

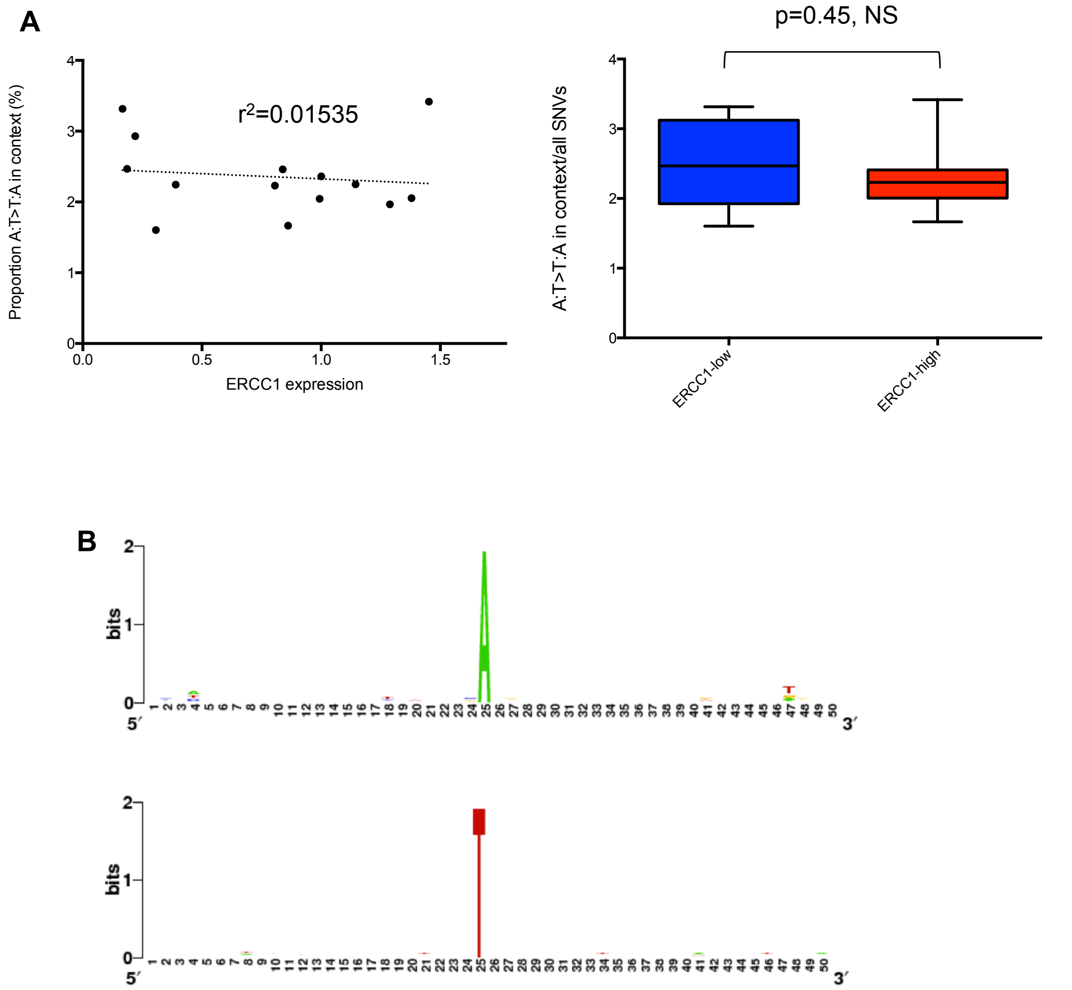


Figure 5. 13. A>T transversions in the non-isogenic model

A) The proportion of A:T>T:A transversions occurring "in context" did not correlate with ERCC1 expression, neither when considering ERCC1 relative expression in a linear model (r^2=0.01535, NS) nor when classifying cell lines in two groups (ERCC1-low vs ERCC1-high; unpaired two-tailed T-test: p=0.45, NS). For this analysis, A>T transversions were considered as being "in context" when they occurred either at TpTp<u>A</u>, TpNp<u>A</u> or NpTp<u>A</u> trinucleotides (where the underlined nucleotide is the mutated one); similarly, T>A tranversions were considered as being "in context" when they occurred either at ApAp<u>T</u>, ApNp<u>T</u> or NpAp<u>T</u> trinucleotides. The proportion of A:T>T:A transversions was calculated as the ratio of the SNVs of interest occurring "in context" over the total number of SNVs of the cell line. B) When considering al A:T>T:A SNVs, no enrichment in a specific sequence logo was detected; representative example of the H292 cell line.

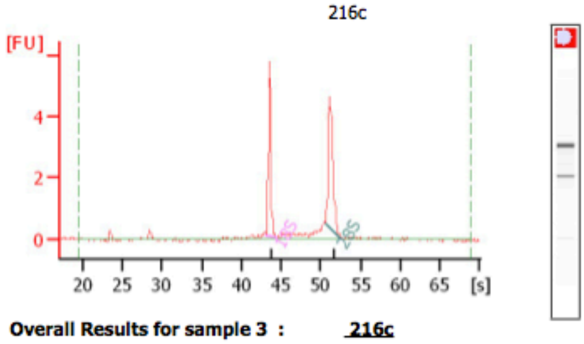
5.2.3. Transcriptomic analysis using Illumi BeadArray® HumanHT-12 v4

As ERCC1/XPF plays a key role both in GG-NER and TC-NER, performing a transcriptomic analysis of ERCC1-proficient and ERCC1-deficient cell lines was of highest interest: beyond the potential of RNA expression analysis for identifying targets for synthetic lethal relationship with ERCC1-deficiency, transcriptional modifications due to TC-NER defects could represent surrogate biomarkers of ERCC1 activity. Moreover, ERCC1/XPF and other members of the NER machinery have been demonstrated to play a role in regulating transcription, which has been suggested to be independent of their role in DNA repair (Le May *et al.*, 2010; Kamileri *et al.*, 2012). Therefore, RNA expression analysis was performed in order to investigate this potential function.

5.2.3.1. Quality control and metrics of the analysis

In order to investigate whether synthetic lethal relationships or surrogate biomarkers of ERCC1 activity could be found at the RNA level, a whole transcriptome analysis was performed on an Illumina BeadArray® HumanHT-12 v4 device. This analysis was performed in collaboration with the Wellcome Trust Centre for Human Genetics, University of Oxford.

The following samples were analysed with three biological replicates each: A459 (ERCC1-proficient parental cell line), Ahez (ERCC1-heterozygous cell line), Ac216, Ac295, Ac385 (three ERCC1-deficient clones), Ac375+isof201, Ac375+isof202, Ac375+isof203 and Ac375+isof204 (clone Ac375 in which each distinct ERCC1 isoform had been reintroduced). In order to ensure appropriate quality of the results, RNA samples had to meet the following quality criteria: RIN (RNA Intergity Number) greater than 7, 28S/18S ratio greater than 1.6, 260/280 ratio between 1.8-2.0 and 260/230 ratio close to 2.0. Representative examples of appropriate quality extraction are displayed in Figure 5. 14.

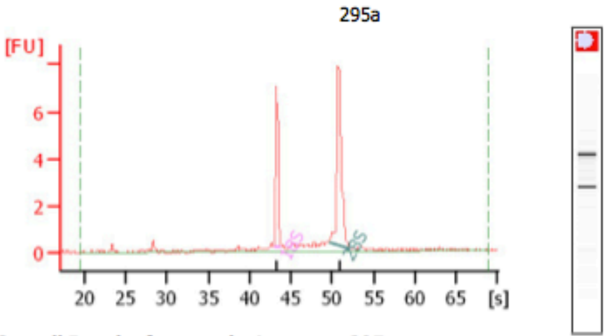


Overall Results for sample 3 : 216c

RNA Area:	17.4
RNA Concentration:	221 ng/µl
rRNA Ratio [28s / 18s]:	1.2
RNA Integrity Number (RIN):	9.4 (B.02.08)
Result Flagging Color:	
Result Flagging Label:	RIN: 9.40

Fragment table for sample 3 : 216c

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	43.10	44.50	4.6	26.7
28S	50.60	52.70	5.8	33.2



Overall Results for sample 4 : 295a

RNA Area:	24.5
RNA Concentration:	312 ng/µl
rRNA Ratio [28s / 18s]:	1.8
RNA Integrity Number (RIN):	9.8 (B.02.08)
Result Flagging Color:	
Result Flagging Label:	RIN: 9.80

Fragment table for sample 4 : 295a

Name	Start Time [s]	End Time [s]	Area	% of total Area
18S	42.90	43.85	5.5	22.4
28S	49.55	52.10	10.0	40.7

Figure 5. 14 - Representative examples of appropriate RNA samples quality

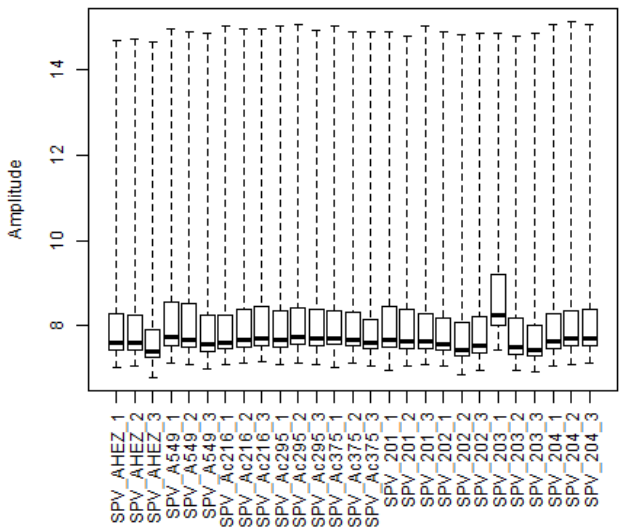
Each RNA sample quality was assessed using a 2100 Bioanalyzer® platform (Agilent® technologies). In addition to providing the RNA sample concentration, this analyzer provided comprehensive data on quality metrics of the samples.

5.2.3.2. Data analysis

Raw data were first normalized using the vst (variance stabilization) and rsn (robust spline normalisation) methods in order to minimise inter-sample variability and ensure the comparability of the results obtained from different samples (Chapter 2). Box plots of Figure 5. 15 represent the distribution of raw value (upper panel) and normalised values (lower panel) for all samples. Before normalisation, a single sample (203_1, corresponding to Ac375 + isoform 203 replicate 1) had an outlier profile, which was not detectable after normalisation.

As ERCC1 was known to be differentially expressed in different samples, it was used as a positive control of the dataset in order to evaluate the accuracy of the results. The four probes targeting ERCC1 on the Illumina BeadArray were isolated and their localisation on the ERCC1 gene was visualised using Array-Check (projects.insilico.us/SpliceCenter/ArrayCheck.jsp), as shown on Figure 5. 16. A. Two probes, ILMN_1797172 and ILMN_1652369, were located in regions that were upstream of the region targeted by the Zinc finger nucleases, and which were not shared by all ERCC1 isoforms; another probe, ILMN_2277676, was located downstream of the last *ERCC1* exon and was flagged up as "degenerated". These first three probes were consequently deemed as not appropriate for being used as positive controls. The last probe, ILMN_2377496 that was located after the cutting site of the Zinc finger nucleases and in exons that were shared by all isoforms, was therefore considered to be the most relevant probe for being used as positive control. When the heatmaps of ERCC1 expression were analysed for each probe (Figure 5. 16. B), the expression profile obtained for probe ILMN_2377496 was consistent with expectations (zoomed in Figure 5. 16. C). Indeed, high levels of ERCC1 mRNA expression were observed in the A549 ERCC1-parental and Ahez cell lines, associated with lower levels of expression in the ERCC1-deficient clones. Interestingly, clones in which each distinct isoforms had been reintroduced also displayed low levels of expression for this probe.

A



B

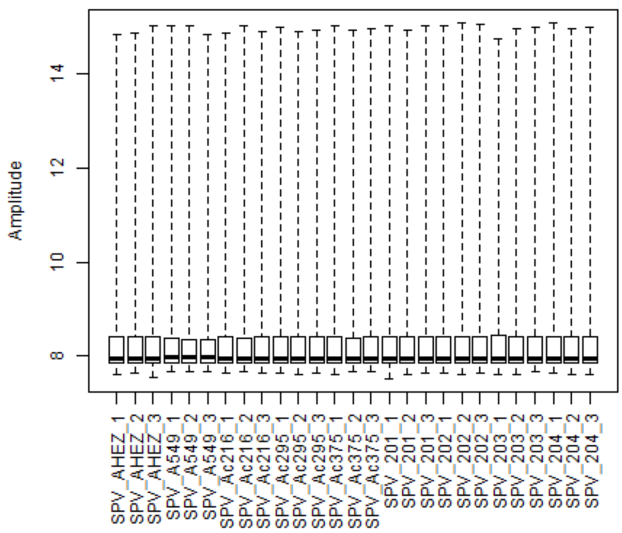


Figure 5. 15. Boxplots of Illumina BeadArray values of transcriptome analysis in the ERCC1-isogenic model

A) Boxplots of raw values prior to normalisation. B) Boxplots of normalized values. Each cell line was analysed in biological triplicates.

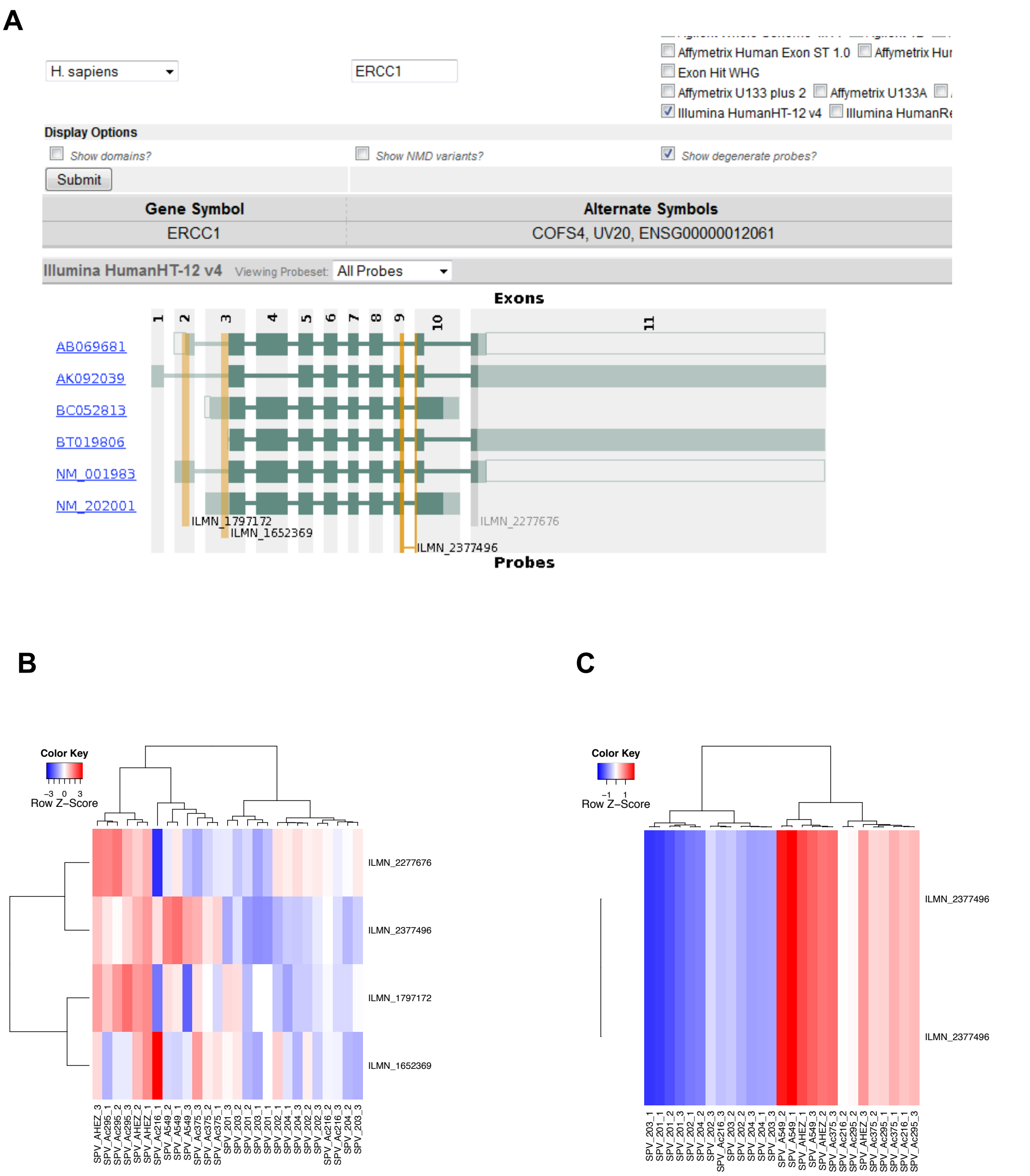


Figure 5. 16. Four ERCC1 probes used in the Illumina® HT-12 v4 BeadArray

A) Mapping of the four ERCC1 probes present in the Illumina® HT-12 v4 BeadArray with *ERCC1*. One probe, ILMN_2277676, was flagged as degenerated and two probes, ILMN_1797172 and ILMN_1652369 were located in regions that were not shared by all ERCC1 isoforms; ILMN_2377496 was therefore considered as the most relevant probe (http://projets.insilico.us/SpliceCenter/ArrayCheck). B) Heatmap of each ERCC1 probe signal; little consistency was observed, likely secondary to observations made in A). C) Heatmap of the probe ILMN_2377496, which was *a priori* (Cf A) considered as being the most relevant probe.

The relationship between samples present in the dataset was studied by hierarchical clustering and principal component analysis (PCA). Hierarchical clustering and principal component analysis were computed based on the most varying probes according to the coefficient of variation (cv = sd/mean > 0.5). Unsupervised hierarchical clustering of all samples present in the dataset was performed on the most variant 7598 probes across the array using euclidian distance and average linkage. As displayed in Figure 5. 17, A549 ERCC1-proficient cell lines clustered together, whereas all other cell lines were regrouped in a second cluster. This latter could be further divided into two main clusters: one gathering all ERCC1-deficient cell lines (including heterozygous cell line), and one containing all Ac375 clones in which each ERCC1 isoform had been reintroduced. One replicate of the Ac216 ERCC1-deficient clone, clustered independently, suggesting potential artifact or less reliable results for this replicate. This clustering was confirmed by PCA, where ERCC1-proficient and ERCC1-deficient cell lines were regrouped following two different components, with the exception of Ac216 that was an outlier (Figure 5. 17). Similar to the analysis performed on all cell lines, the relationship between ERCC1-deficient cell lines only was studied by non-supervised hierarchical clustering and PCA. The hierarchical clustering of the 5066 most variant probes of ERCC1-altered samples resulted in two main clusters: one containing all ERCC1-altered cell lines and one gathering all Ac375 clones in which each ERCC1 isoform had been stably reintroduced (Figure 5. 18). Ac216 replicate 1 was again apparent as an outlier. The PCA of these ERCC1-altered samples corroborated these results, by showing two main clusters constituted of the same groups as those identified by hierarchical clustering. These initial analyses suggested that transcriptomic data could recapitulate the expected classification of the samples according to ERCC1 status, and might as such provide meaningful results.

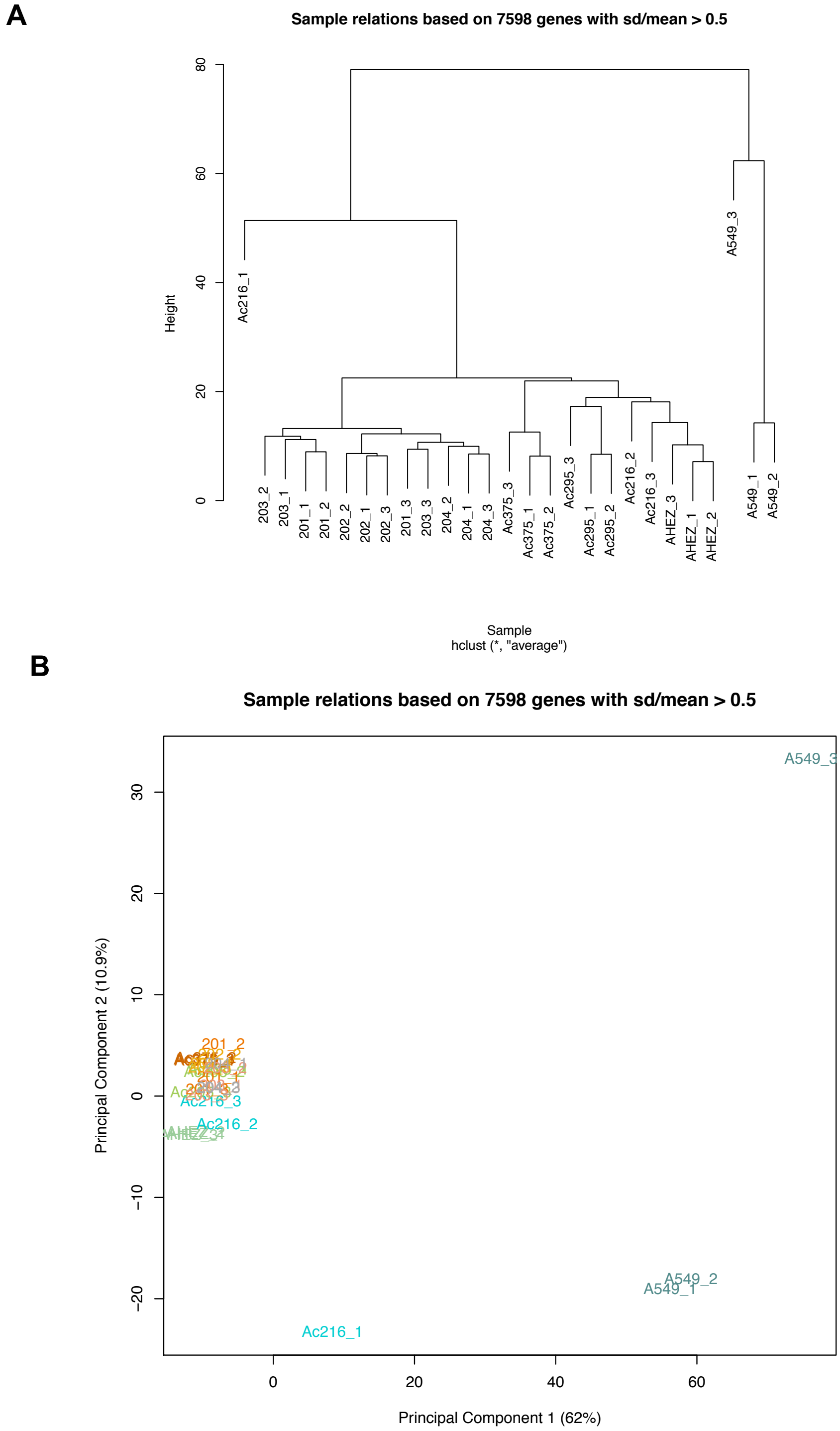


Figure 5. 17. Relationship between all samples present in the dataset

A) Hierarchical clustering (unsupervised analysis) of all samples using euclidian distance and average linkage method. Ac216 replicate 1 appears to be an outlier. A549 cell lines cluster separately, and all ERCC1-deficient cell lines (including heterozygous cell line and isoform-rescued cell lines) are regrouped within the same cluster. B) Principal component analysis of all samples present in the dataset. The most variant 7598 probes across the gene expression profiles of samples included - as computed by the coefficient of variation (cv = sd/mean > 0.5) - were used in order to generate hierarchical clustering and PCA.

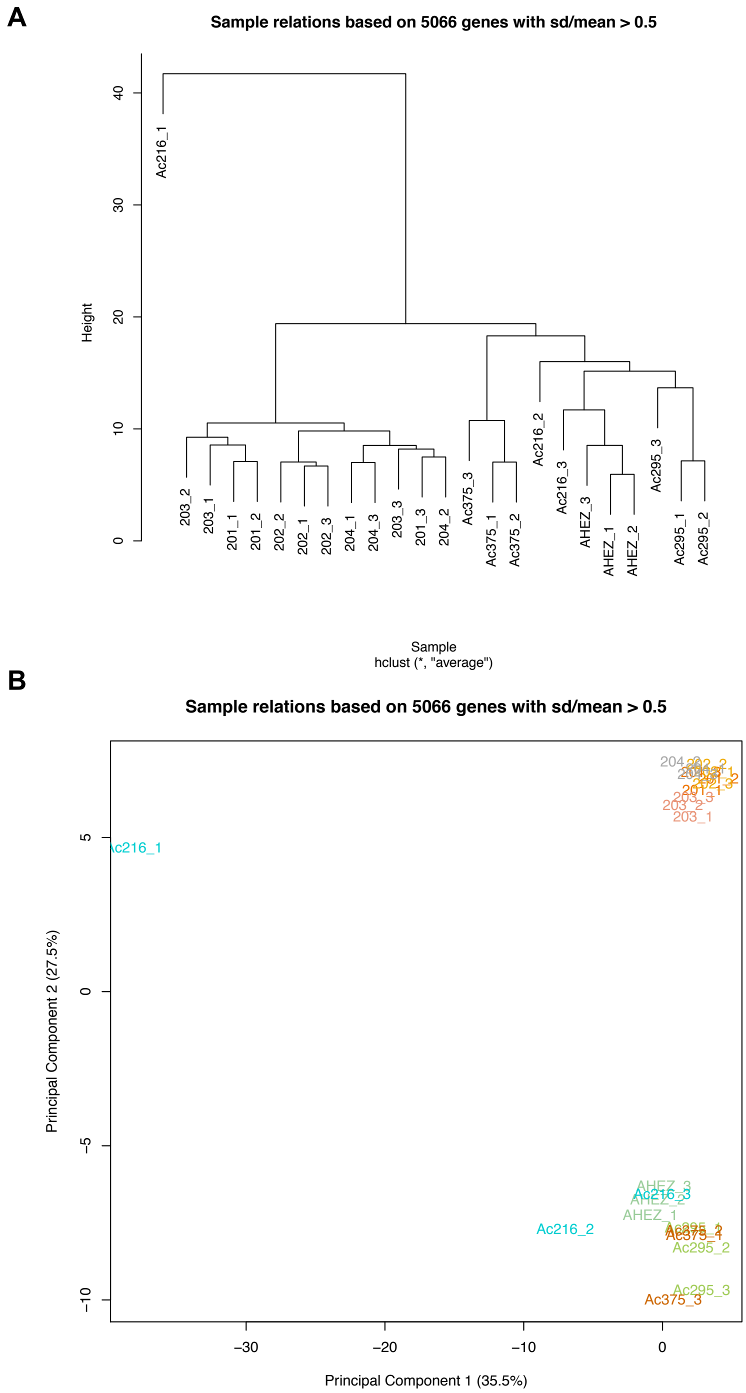


Figure 5. 18. Relationship between all ERCC1-altered samples present in the dataset

A) Hierarchical clustering of all ERCC1-altered samples, i.e. ERCC1-heterozygous cell line, ERCC1-deficient clones, and ERCC1-deficient clone Ac375 in which each distinct isoform had been re-expressed. Ac216 replicate 1 appears to be an outlier. All ERCC1-models in which one isoform had been reintroduced cluster together. B) Principal component analysis of all ERCC1-altered samples present in the dataset. The most variant 5066 probes across the gene expression profiles of samples considered - as computed by the coefficient of variation (cv = sd/mean > 0.5) - were used in order to generate hierarchical clustering and PCA.

5.2.3.3. Transcriptomic differences between ERCC1-proficient and ERCC1-deficient cells

To investigate transcriptomic differences between ERCC1-proficient and ERCC1-deficient cells, the RNA expression profile of A549 cells were compared to those of the three ERCC1-deficient clones considered as a pool (Ac216, Ac295 and Ac375). Genes significantly differentially expressed were identified based on their significant Bonferroni corrected p-value < 0.01. The heatmap of the top 296 significant genes (Figure 5. 19. A), where hierarchical clustering was performed using Pearson correlation and Ward distance, revealed that most of the differentially transcribed genes were down-regulated in the ERCC1-deficient cells. Most of the significant hits were shared between the three ERCC1-defiient clones, as illustrated by the Venn diagrams of the corresponding probes (Figure 5. 19. B).

In order to investigate specific pathway enrichment, genes that were significantly differentially expressed between ERCC1-proficient and ERCC1-deficient cells were entered into the Gene Set Enrichment Analysis (GSEA) database (www.broadinstitute.org/gsea/). When considering this dataset, pathways that were significantly enriched were (by order of significance): axon guidance (2 databases), immune system, developmental biology, neutrophils pathway, PDGFRB pathway, integrin 2 pathway, adhesion molecules, and cancer (Table 5. 1, where colours have been chosen arbitrarily to ease visualisation). As it has been previously suggested, these results supported that ERCC1/XPF function was not limited to NER or DNA repair, and that additional roles remained to be investigated.

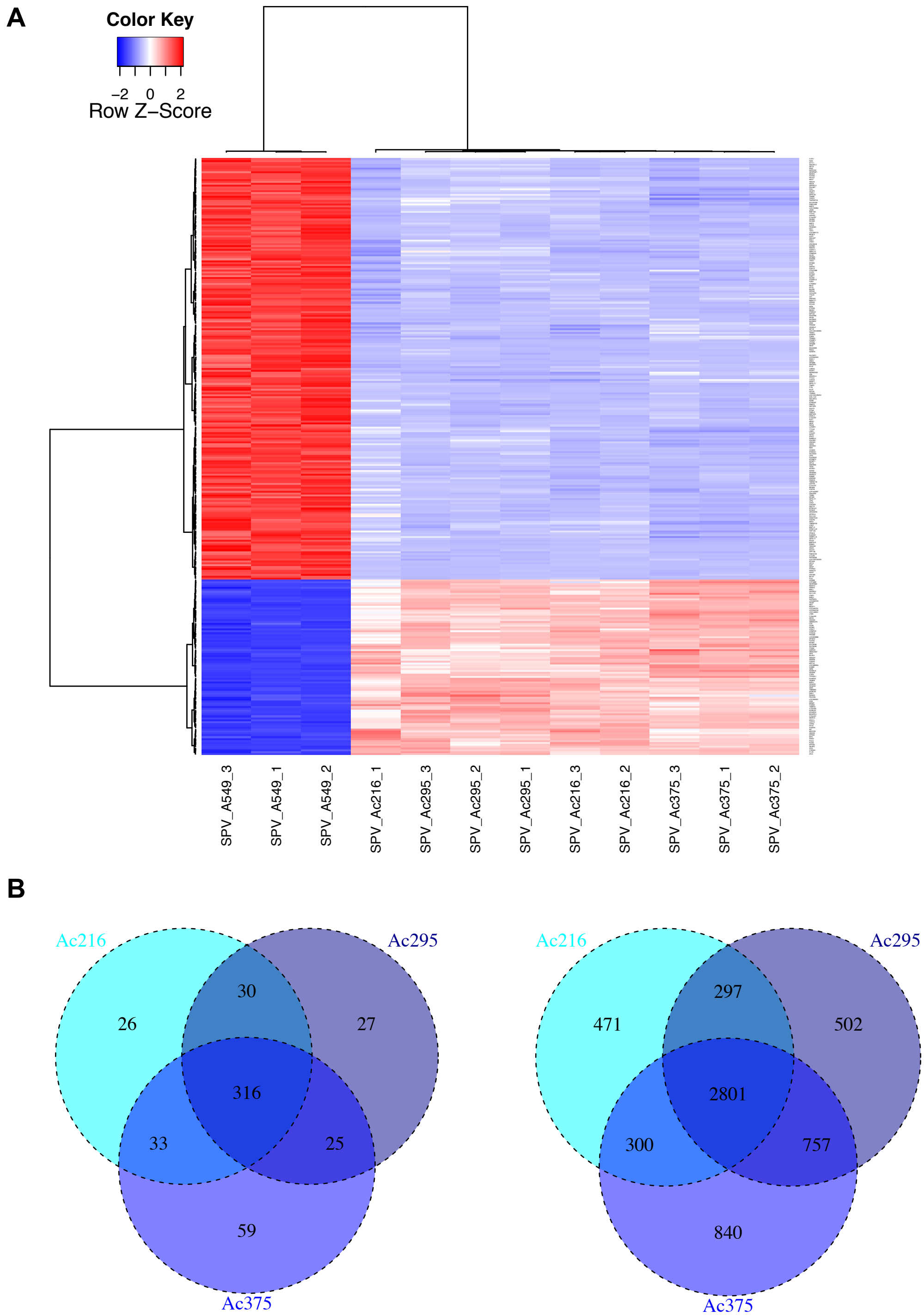


Figure 5. 19. Transcriptomic differences between ERCC1-proficient and ERCC1-deficient cell lines

A) Heatmap of the top 296 significant genes (Bonferroni corrected p-value < 0.01) B) Most differences in RNA expression between the ERCC1-proficient and ERCC1-deficient cells are shared between the three ERCC1-deficient clones. Venn diagrams of probes differentially expressed between ERCC1-proficient and ERCC1-deficient cells; left panel: significant probes with Bonferroni corrected p-value < 0.01; right panel: significant probes with FDR < 0.01

Gene Symbol	Description	KEGG	REACTOME	REACTOME	REACTOME	PID	PID	PID	KEGG	KEGG
	Description	Axon guidance	Axon guidance	Immune system	Development biology	Neutrophils pathway	PDGFRB pathway	Integrin 2 pathway	Adhesion molecules	Cancer
	# Genes in G	129	251	933	396	41	129	29	134	328
	# Genes in O	15	17	34	20	7	11	6	11	17
	k/K	0.1163	0.0677	0.0364	0.0505	0.1707	0.0853	0.2069	0.0821	0.0518
	p-value	2.30E-10	5.94E-08	1.78E-07	4.98E-07	1.09E-06	1.38E-06	1.99E-06	2.00E-06	2.50E-06
	FDR q-value	3.03E-07	3.92E-05	7.82E-05	1.64E-04	2.88E-04	3.03E-04	3.31E-04	3.31E-04	3.67E-04
FYN	FYN oncogene related to SRC, FGR, YES									
PAK1	p21/Cdc42/Rac1-activated kinase 1									
MET	met proto-oncogene									
SLIT2	slit homolog 2									
SRGAP3	SLIT-ROBO Rho GTPase activating protein 3									
ROBO2	roundabout, axon guidance receptor, homolog 2									
SEMA6A	sema domain and cytoplasmic domain (semaphorin) 6A									
EPHB1	EPH receptor B1									
EPHA4	EPH receptor A4									
EFNB2	ephrin-B2									
CXCR4	chemokine (C-X-C motif) receptor 4									
GNAI1	guanine nucleotide binding protein alpha inhibiting polypept. 1									
EPHA3	EPH receptor A3									
EFNB3	ephrin-B3									
LRRC4C	leucine rich repeat containing 4C									
NCAM1	neural cell adhesion molecule 1									
ENAH	enabled homolog									
SH3GL2	SH3-domain GRB2-like 2									
COL6A1	collagen, type VI, alpha 1									
CRMP1	collapsin response mediator protein 1									
DPYSL4	dihydropyrimidinase-like 4									
RGMA	RGM domain family, member A									
RGMB	RGM domain family, member B									
CACNB2	calcium channel, voltage-dependent, beta 2 subunit									
KCNQ2	potassium voltage-gated channel, KQT-like subfamily, mb 2									
LYN	v-yes-1 Yamaguchi sarcoma viral related oncogene homolog									
MAPK10	mitogen-activated protein kinase 10									
CBL	Cas-Br-M ecotropic retroviral transforming sequence									
ICAM3	intercellular adhesion molecule 3									
CDH1	cadherin 1, type 1, E-cadherin									
BCL2	B-cell CLL/lymphoma 2									
ITPR3	inositol 1,4,5-triphosphate receptor, type 3									
ITPR2	inositol 1,4,5-triphosphate receptor, type 2									
AP1M2	adaptor-related protein complex 1, mu 2 subunit									
AP1S2	adaptor-related protein complex 1, sigma 2 subunit									
CYBA	cytochrome b-245, alpha polypeptide									
CTSH	cathepsin H									
CANX	calnexin									
DYNC1LI2	dynein, cytoplasmic 1, light intermediate chain 2									
OSBPL1A	oxysterol binding protein-like 1A									
PSMD5	proteasome (prosome, macropain) 26S subunit, non-ATPase, 5									
FBXW7	F-box and WD-40 domain protein 7									
TRIM9	tripartite motif-containing 9									
WSB1	WD repeat and SOCS box-containing 1									
CFD	complement factor D									
CD55	CD55 molecule, decay accelerating factor for complement									
CD46	CD46 molecule, complement regulatory protein									
CFI	complement factor I									
TXN	thioredoxin									
IRS1	insulin receptor substrate 1									
IL1RAP	interleukin 1 receptor accessory protein									
CASP4	caspase 4, apoptosis-related cysteine peptidase									
SIGIRR	single immunoglobulin and toll-interleukin 1 receptor domain									
KPNA2	karyopherin alpha 2 (RAG cohort 1, importin alpha 1)									
PPARG	peroxisome proliferative activated receptor, gamma									
TCF4	transcription factor 4									
BOC	Boc homolog									
JAM3	junctional adhesion molecule 3									
PLAUR	plasminogen activator, urokinase receptor									
THY1	Thy-1 cell surface antigen									
PLAT	plasminogen activator, tissue									
JAM2	junctional adhesion molecule 2									
PDGFRB	platelet-derived growth factor receptor, beta polypeptide									
MYC	v-myc myelocytomatosis viral oncogene homolog									
ARPC1B	actin related protein 2/3 complex, subunit 1B, 41kDa									
PPP2R2B	protein phosphatase 2 regulatory subunit B beta isoform									
SRF	serum response factor									
TAGLN	transgelin									
F11R	F11 receptor									
ITGA6	integrin, alpha 6									
CLDN23	claudin 23									
NLGN1	neuroligin 1									
NLGN3	neuroligin 3									
NLGN4X	neuroligin 4, X-linked									
LAMA4	laminin, alpha 4									
FZD7	frizzled homolog 7									
FGFR3	fibroblast growth factor receptor 3									
FZD3	frizzled homolog 3									
RUNX1	runt-related transcription factor 1 (aml 1 oncogene)									
CYCS	cytochrome c, somatic									
CKS1B	CDC28 protein kinase regulatory subunit 1B									
RALGDS	ral guanine nucleotide dissociation stimulator									
PTPRZ1	protein tyrosine phosphatase, receptor-type, Z polypeptide 1									

axon guidance · developmental biology · integrins / adhesion molecules

immune system · PDGFRB pathway · cancer general

Table 5. 1. Pathway enrichment of the transcriptome hits between ERCC1-proficient and ERCC1-deficient cells (legend next page)

Table 5.1. Pathway enrichment of the transcriptome hits between ERCC1-proficient and ERCC1-deficient cells (Table previous page)

GSEA overlap pathways (Gene Set Enrichment Analysis database, Broad Institute; www.broadinstitute.org/gsea/) of hits showing statistically significant difference of Bonferroni corrected p-value < 0.01 between ERCC1-proficient and ERCC1-deficient cells (as a pool). Significantly enriched pathways are depicted using an arbitrary colour code.

5.2.3.4. Transcriptomic differences between Ac375 ERCC1-deficient cells before and after reintroduction of each ERCC1 isoform

To investigate transcriptome differences between ERCC1-deficient Ac375 and the same clone after reintroduction of each distinct ERCC1 isoform, the RNA expression profile of Ac375 cells was compared to those of the four Ac375 cell lines in which ERCC1 isoforms had been stably re-expressed, while considering them as a pool (Ac375 + isoform 201, Ac375 + isoform 202, Ac375 + isoform 203, Ac375 + isoform 204). Genes significantly differentially expressed were identified based on their significant Bonferroni corrected p-value < 0.01. The heatmap of the top 127 significant genes (Figure 5. 20. A), where hierarchical clustering was performed using Pearson correlation and Ward distance, revealed that most of the differentially transcribed genes were down-regulated even after reintroduction of one of the ERCC1 isoforms. Most of the significant hits were shared between the four cell lines in which isoforms had been reintroduced, as illustrated by the Venn diagrams of the corresponding probes (Figure 5. 20. B). Of note, clones in which isoforms 204 and 202 had been reintroduced shared the highest number of genes, consistent with similar structure of these two isoforms as compared to isoforms 201 and 203.

To investigate for specific pathways enrichment, genes that were significantly differentially expressed between ERCC1-deficient Ac375 clone and Ac375 in which each ERCC1 isoform had been re-introduced, were entered into the Gene Set Enrichment Analysis (GSEA) database. When considering this dataset, pathways that were significantly enriched were (by order of significance): cell cycle and mitosis (4 databases), immune system (2 databases), DNA replication, metabolism of aminoacids, chemokines, and valine-isoleucine degradation (Table 5. 2), where colours have been chosen arbitrarily to ease visualisation). This supported results previously presented in Chapters 3 and 4, describing a pivotal role for ERCC1 in proper cell cycle and mitotic progression.

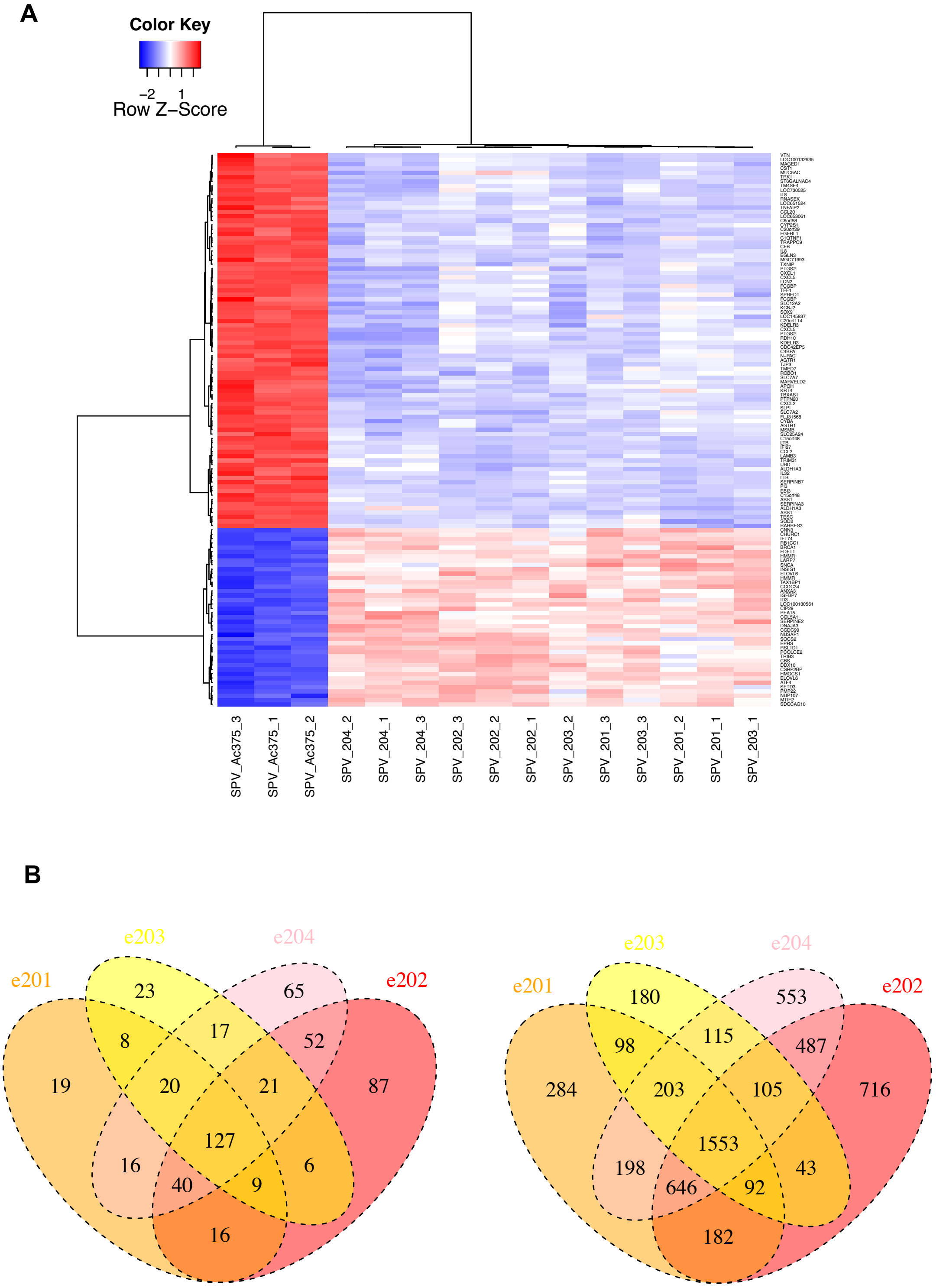


Figure 5. 20. Transcriptomic differences between Ac375 ERCC1-deficient clone and Ac375 after stable re-expression of each ERCC1 isoform

A) Heatmap of the top 127 significant genes (Bonferroni corrected p-value < 0.01) B) Most differences in RNA expression between the Ac375 cell line ERCC1-KO and after reintroduction of each isoform are shared between the four distinct isoforms. Venn diagrams of probes differentially expressed between ERCC1-proficient and ERCC1-deficient cells; left panel: significant probes with Bonferroni corrected p-value < 0.01; right panel: significant probes with FDR < 0.01

Gene Symbol	Description	Pathway	Cell cycle Mitosis	Cell cycle	Immune system	Mitosis / G1 phases	DNA replication	Mitosis / Prometaphase	Metabolism of AAcids	Chemokines	Innate Immune System	Valine - isoleucine degradation
		# Genes in Gene Set (K)	325	421	933	172	192	87	200	57	279	44
		# Genes in Overlap (k)	21	23	33	13	13	9	12	7	13	6
		k/K	0.0646	0.0546	0.0354	0.0756	0.0677	0.1034	0.06	0.1228	0.0466	0.1364
		p-value	2.38E-12	6.67E-12	2.58E-11	5.73E-09	2.14E-08	8.74E-08	2.79E-07	7.45E-07	1.58E-06	2.51E-06
		FDR q-value	3.15E-09	4.40E-09	1.14E-08	1.89E-06	5.66E-06	1.92E-05	5.25E-05	1.23E-04	2.31E-04	3.31E-04
CDC20	CDC20 cell division cycle 20 homolog											
KIF18A	kinesin family member 18A											
NUP107	nucleoporin 107kDa											
PSME1	proteasome activator subunit 1											
PSMC4	proteasome 26S subunit, ATPase, 4											
PSMC1	proteasome 26S subunit, ATPase, 1											
ANAPC1	anaphase promoting complex subunit 1											
CUL1	cullin 1											
DYNC1I2	dynein, cytoplasmic 1, intermediate chain 2											
SGOL1	shugoshin-like 1											
CENPA	centromere protein A											
CCDC99	coiled-coil domain containing 99											
KNTC1	kinetochore associated 1											
SGOL2	shugoshin-like 2											
ZWILCH	Zwilch, kinetochore associated, homolog											
MCM8	MCM8 minichromosome maintenance deficient 8											
AURKA	aurora kinase A											
CCNB2	cyclin B2											
NEK2	NIMA (never in mitosis gene a)-related kinase 2											
CEP135	centrosomal protein 135kDa											
PLK4	polo-like kinase 4											
BRCA1	breast cancer 1, early onset											
RUVBL1	RuvB-like 1											
JUN	jun oncogene											
TNFAIP3	tumor necrosis factor, alpha-induced protein 3											
ISG15	ISG15 ubiquitin-like modifier											
PCBP2	poly(rC) binding protein 2											
ATG12	ATG12 autophagy related 12 homolog											
TAX1BP1	Tax1 binding protein 1											
SAA1	serum amyloid A1											
NFKB2	nuclear factor of k-light polypeptide gene enhancer in B-cells 2											
DUSP6	dual specificity phosphatase 6											
TLR1	toll-like receptor 1											
C4BPA	complement component 4 binding protein, alpha											
CFB	complement factor B											
TXNIP	thioredoxin interacting protein											
HLA-A	major histocompatibility complex, class I, A											
EGR1	early growth response 1											
CD44	CD44 molecule											
SEC23A	Sec23 homolog A											
SEC24D	SEC24 related gene family, member D											
CYBA	cytochrome b-245, alpha polypeptide											
CUL5	cullin 5											
TRIB3	tribbles homolog 3											
CTSC	cathepsin C											
IFI27	interferon, alpha-inducible protein 27											
IFI6	interferon, alpha-inducible protein 6											
ACAT1	acetyl-Coenzyme A acetyltransferase 1											
ALDH6A1	aldehyde dehydrogenase 6 family, member A1											
DBT	dihydrolipoamide branched chain transacylase E2											
ASS1	argininosuccinate synthetase 1											
CKB	creatine kinase, brain											
CKMT1A	creatine kinase, mitochondrial 1A											
PRODH	proline dehydrogenase (oxidase) 1											
CBS	cystathionine-beta-synthase											
DIO2	deiodinase, iodothyronine, type II											
IL8	interleukin 8											
CCL2	chemokine (C-C motif) ligand 2											
CXCL1	chemokine (C-X-C motif) ligand 1											
CXCL2	chemokine (C-X-C motif) ligand 2											
CCL20	chemokine (C-C motif) ligand 20											
CXCL5	chemokine (C-X-C motif) ligand 5											
CX3CL1	chemokine (C-X3-C motif) ligand 1											
ACAT2	acetyl-Coenzyme A acetyltransferase 2											
HMGCS1	3-hydroxy-3-methylglutaryl-Coenzyme A synthase 1											
ALDH3A2	aldehyde dehydrogenase 3 family, member A2											
CYR61	cysteine-rich, angiogenic inducer, 61											

Cell cycle
Immunity
Metabolism
DNA replication
Chemokines

Table 5. 2. Pathway enrichment of the transcriptomic hits between Ac375 ERCC1-deficient cell line and Ac375 with stable re-expression of each ERCC1 isoform

GSEA overlap pathways (Gene Set Enrichment Analysis database, Broad Institute; www.broadinstitute.org/gsea/) of hits showing statistically significant difference of Bonferroni corrected p-value < 0.01 between Ac375 ERCC1-deficient cells and Ac375 with stable reintroduction of each ERCC1 isoform (as a pool). Significantly enriched pathways are depicted using an arbitrary colour code.

5.2.4. Identification of differential protein expression between ERCC1-proficient and ERCC1-deficient cells

5.2.4.1. High throughput SILAC proteomic analysis

Briefly, SILAC (Stable Isotope Labelling by Amino acid in Cell culture) relies on metabolic incorporation of a given "light" or "heavy" form of the amino acid into the proteins. The "heavy" amino acid has substituted stable isotopic nuclei (e.g. 13C and 15N labelled arginine, and13C and 15N labelled lysine (R10K8)). Thus in a SILAC experiment, two cell populations are grown in culture media that are identical except that one of them contains a "light" and the other a "heavy" form of particular amino acids. When the labelled analogue of an amino acid is supplied to cells in culture instead of the natural amino acid, it is incorporated into all newly synthesized proteins. After a number of cell divisions (at least seven, in order to ensure complete replacement of the normal amino acid by the labelled one), each instance of this particular amino acid will be replaced by its isotope labelled analogue. Since there is hardly any chemical difference between the labelled amino acid and the natural amino acid isotopes, the cells behave exactly like the control cell population grown in the presence of normal amino acid. The proteins from both cell populations are then mixed in a 1:1 ratio, and the final mix analysed by mass-spectrometry. The distance between the "light" and "heavy" labelled peptides (e.g. 10Da between the "heavy" 13C and 15N labelled arginine and normal "light" labelled arginine) allows the quantification of the ratio of each peptide in both cells populations. A reciprocal (or reverse) experiment, where the labelling is inverted, is usually performed in parallel to minimise the number of false-positive results (Ong *et al.*, 2006; Geiger *et al.*, 2011).

For the purpose of this thesis, the Ac216 clone, which expressed the lowest level of ERCC1 expression (6%, see Chapter 3, Figure 3.2), was chosen for comparison to its ERCC1-proficient isogenic counterpart, the A549 ERCC1-proficient parental cell line. The overall flowchart of the experiment is described in Figure 5. 21. Cells were labelled with "heavy" or "light" amino acids (R10K8 or null labelling, respectively) for eight doubling time prior to protein extraction. Extracted proteins were then mixed in a 1:1 ratio to allow quantitative comparison of protein abundance. Quality of the labelling was checked on the first set of four samples, which were trypsin digested prior to mass spectrometry (MS) analysis. Other samples subjected to fractionation (12 fractions each) by off- gel isoelectric focusing (IEF) using a 3100 Off-Gel Fractionator (Agilent® Technologies). Importantly, a forward and a reverse experiment (or reciprocal, where labelling was inverted) were carried out at the same time in order to improve the quality of the final analysis, which limited false-positive hits. Each fraction of mixed peptides arising from both cell lines was run on Orbitrap Velos® (Thermo Scientific) by Dr Andrew Thompson at the ICR Mass Spetrometry unit and analysed by MS/MS. Data analysis was performed by Dr Faraz Mardakheh using MaxQuant (www.maxquant.org) according to Cox et al, Nature Protocols 2009 (Cox *et al.*, 2009). Briefly, Raw MS/MS data we're searched and quantified on MaxQuant against Human IPI database (v3.68).

Hits that reached significance (Bonferroni-corrected p-value < 0.05) in both experiments with an opposite fold change direction (e.g; increased in forward and decreased in reciprocal) were selected for subsequent analysis (Figure 5.22). Hits that reached significance in both experiments with a similar fold change direction were discarded and considered as false positives. Hits that reached significance in only one experiment were not considered, given the high risk of false positive. Forty-five hits listed in Table 5. 3 met these pre-defined selection criteria and 16 of them were selected for subsequent low-throughput revalidation by western blot. Of note, the positive controls, ERCC4 (XPF) and ERCC1, were detected as significantly decreased (81% and 49%, respectively) in the reciprocal experiment; they were not detected in the forward experiment. This was likely related to the initial omission of the C18 ‘ZipTip’ clean up step of samples, normally employed with this type of experiment, with the objective to reduce sample handling steps without decreasing the sensitivity of the overall workflow.

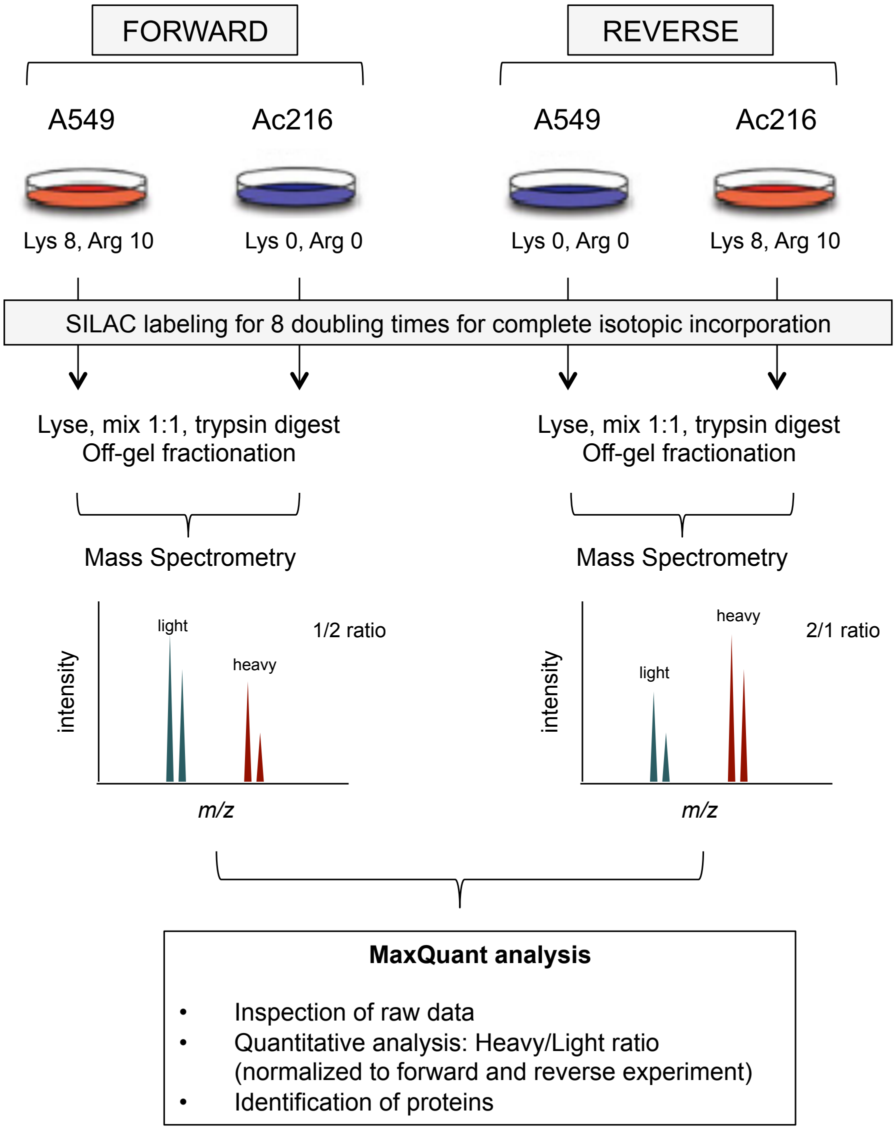


Figure 5. 21. Flowchart of SILAC experiment

Ac216, which expressed the lowest level of ERCC1, was chosen for comparison to the A549 parental cell line. Two experiments were set up and run in parallel: the forward experiment, in which A549 and Ac216 cells were labelled for 8 doubling times with heavy and light amino acids, respectively, and the reverse experiment which mirrored the forward experiment. This allowed limiting the number of false positive by having an internal control (e.g. any protein increased in a given cell population in the forward experiment should theoretically be decreased in the same proportion in the reverse experiment). After labelling, cells were harvested, lysed for protein collection, and protein lysates were mixed in a 1:1 proportion between the heavy- and light-labelled cells. Proteins were subsequently digested into peptides, prior to off-gel fractionation, which allowed obtaining 12 peptides fractions by mixed protein lysate. Fractions were analysed by MS/MS, which allowed the differentiation between the heavy- and the light-labelled peptides, and final analysis of raw data was performed using MaxQuant (www.maxquant.org) following Cox et al, Nature Protocols 2009.

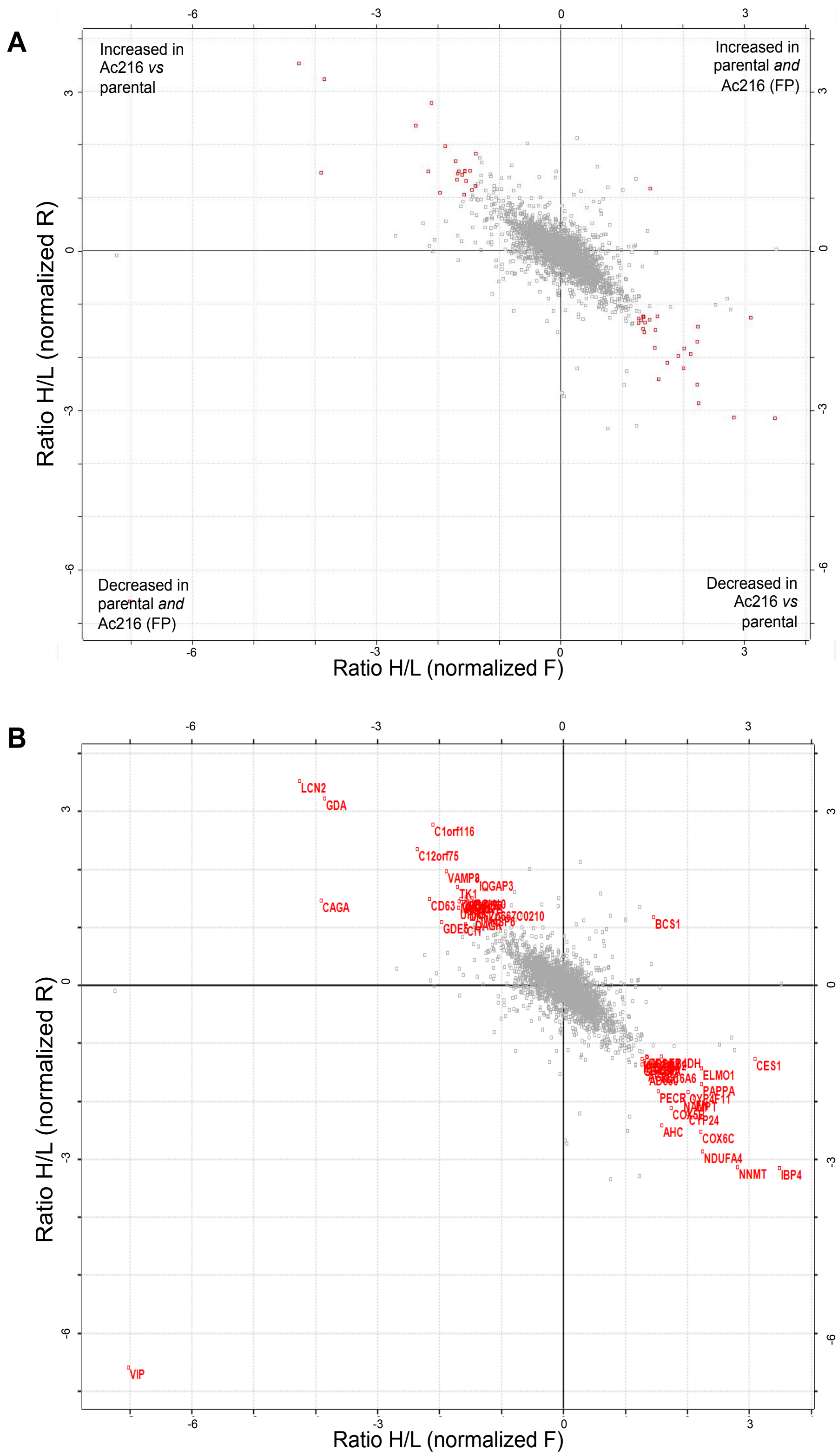


Figure 5. 22. Overall SILAC experiments results (legend next page)

Figure 5. 22. Overall SILAC experiments results (Figure previous page)

Ratio of heavy over low amino acids normalized to the reverse (R) or forward (F) experiment, expressed following a Log2 scale. Each graph contains four quadrants: top left quadrant = hits increased in the R experiment and decreased in the F experiment, i.e. increased in the ERCC1-deficient Ac216 clone as compared to the parental cell line; bottom right quadrant = hits decreased in the R and increased in the F experiment, i.e. decreased in the ERCC1-deficientAc216 clone as compared to parental cell line; top right and bottom left quadrants: hits either increased or decreased in both experiments i.e. false positives. A) Scatter plot of all SILAC results. Hits reaching significance (Bonferroni corrected p-value > 0.05) in both experiments are depicted in red; B) Scatter plot of all SILAC results annotated with the name of one of the proteins containing the identified peptide.

Table 5.3. SILAC hits that reached significance in the forward and reverse experiment (Table next page)

Summary table of the forty-eight peptides that reached significance (Bonferroni corrected p-value < 0.05) in the reverse and forward SILAC experiment. Protein names in grey and italic characters correspond to false positives that reached the significance threshold. Columns 1&2: ratio of heavy / low amino acids normalized to the forward experiment; numbers are expressed as a log2 of the fold change. Column 3: total number of counts for each peptide. Columns 4&5: Bonferroni corrected p-value. Column 6: first protein ID identified, corresponding to the peptide of interest. Column 7: first two genes names identified, corresponding to the peptide of interest. Column 8: total number of proteins identified corresponding to the peptide of interest. Column 9: succinct description of the protein function. Colour code has been arbitrarily defined and added to help visualization of proteins involved in shared cellular process.

	1 Ratio H/L Norm F	2 Ratio H/L Norm R	3 Ratio H/L Count	4 Ratio H/L Norm F Significance	5 Ratio H/L Norm R Significance	6 Protein IDs (first only)	7 Gene Names (first two only)		8 Nb of prot. identified	9 Protein / gene name overview
Peptides increased in Ac216 as compared to A549-WT cells	-7.0233	-6.5977	37	9.70E-63	3.19E-64	IPI00000959	VIP		3	
	-4.2650	3.5236	9	2.55E-24	1.93E-24	IPI00643623	LCN2	NGAL	2	lipocalin 2
	-3.9101	1.4673	4	1.02E-20	1.93E-05	IPI00007047	CAGA	S100A8	1	S100 calcium binding protein A8
	-3.8571	3.2287	8	3.32E-20	8.59E-21	IPI00873506	GDA	GAH	4	guanine deaminase
	-2.3586	2.3479	6	1.57E-08	9.60E-12	IPI00922250	C12orf75	OCC1	1	Overexpressed In Colon Carcinoma 1 Protein
	-2.1656	1.4925	10	2.03E-07	1.39E-05	IPI00215998	CD63	MLA1	7	CD63 molecule
	-2.1111	2.7740	10	4.04E-07	8.90E-16	IPI00028392	C1orf116	SARG	3	Specifically androgen-regulated gene protein
	-1.9670	1.0998	12	2.30E-06	0.00131163	IPI00099820	GDE5	GPCPD1	2	glycerolphosphodiester phosphodiesterase 5
	-1.8898	1.9718	7	5.57E-06	1.02E-08	IPI00030911	VAMP8		3	vesicle-associated membrane protein 8 (endobrevin)
	-1.7103	1.6899	20	3.85E-05	8.88E-07	IPI00791117	TK1		4	thymidine kinase 1
	-1.6913	1.3362	4	4.68E-05	9.81E-05	IPI00797279	UHRF1	ICBP90	3	ubiquitin-like with PHD and ring finger domains 1
	-1.6877	1.4514	9	4.85E-05	2.36E-05	IPI00016007	MYO5C		5	myosin VC
	-1.6572	1.4865	4	6.60E-05	1.50E-05	IPI00015104	CKS1	CKS1B	4	Cyclin-dependent kinases regulatory subunit 1
	-1.6063	1.4324	17	0.000109062	3.01E-05	IPI00010491	RAB27B		3	RAB27B, member RAS oncogene family
	-1.5772	1.0562	14	0.000144347	0.00201575	IPI00719285	CIT	KIAA0949	4	citron (rho-interacting, serine/threonine kinase 21)
	-1.5697	1.5115	9	0.000155089	1.08E-05	IPI00013002	UBCH10	UBE2C	4	ubiquitin-conjugating enzyme E2C
	-1.5697	1.4965	24	0.000155152	1.32E-05	IPI00555902	OCIAD2		2	OCIA domain containing 2
	-1.5468	1.3164	8	0.000192729	0.00012394	IPI00306532	DKFZp667C0210	MYO15B	1	Putative myosin-XVB
	-1.4772	1.5003	13	0.00036615	1.25E-05	IPI00291215	BAL2	PARP14	9	poly (ADP-ribose) polymerase family, member 14
	-1.4539	1.1498	6	0.000451212	0.000786114	IPI00333143	DAGK	DGKA	3	diacylglycerol kinase, alpha 80kDa
	-1.3972	1.2205	11	0.000741905	0.00036884	IPI00013449	TM4SF6	TSPAN6	2	tetraspanin 6
	-1.3841	1.8318	11	0.000830085	1.01E-07	IPI00328905	IQGAP3		1	IQ motif containing GTPase activating protein 3
Peptides decreased in Ac216 as compared to A549-WT cells	1.2632	-1.2715	30	0.000779877	0.00123619	IPI00553043	KIAA1102	LIMCH1	11	LIM and calponin homology domains 1
	1.2645	-1.3635	22	0.000769357	0.000526223	IPI00005724	GPR69A	LANCL1	6	LANCL1
	1.3036	-1.3110	4	0.000521345	0.000861848	IPI00744889	CDH1	CDHE	4	cadherin 1, type 1, E-cadherin
	1.3387	-1.4640	11	0.000363897	0.00019505	IPI00903328	AGFG1	HRB	6	ArfGAP with FG repeats 1
	1.3435	-1.2277	8	0.000346296	0.00182199	IPI00029260	CD14		1	CD14 molecule
	1.3494	-1.2396	31	0.000325682	0.00164059	IPI00007702	HSPA2		2	heat shock 70kDa protein 2
	1.3640	-1.5238	19	0.000279699	0.000104857	IPI00024816	AD030	MFF	11	mitochondrial fission factor
	1.3708	-1.3549	16	0.000260281	0.000571323	IPI00843789	GCSP	GLDC	1	glycine dehydrogenase (decarboxylating)
	1.4455	-1.2930	24	0.000115912	0.00101671	IPI00006579	COX4	COX4I1	2	cytochrome c oxidase subunit IV isoform 1
	1.4560	1.1711	15	0.000103064	0.00062839	IPI00003985	BCS1	BCS1L	8	
	1.5286	-1.8281	13	4.50E-05	3.15E-06	IPI00744627	PECR	PRO1004	3	peroxisomal trans-2-enoyl-CoA reductase
	1.5419	-1.4912	12	3.85E-05	0.000147617	IPI00783565	SLC6A6	hCG_1995322	23	solute carrier family 6 (neurotransmitter transporter, taurine) mb 6
	1.5736	-1.2282	84	2.64E-05	0.00181406	IPI00292657	LTB4DH	PTGR1	4	prostaglandin reductase 1
	1.5886	-2.4186	11	2.21E-05	6.46E-10	IPI00020943	NROB1	DAX1	2	nuclear receptor subfamily 0, group B, member 1
	1.7371	-2.1074	37	3.38E-06	7.49E-08	IPI00021785	COX5B		1	cytochrome c oxidase subunit Vb
	1.9144	-1.9743	62	2.94E-07	4.73E-07	IPI00018873	NAMPT	PBEF	7	nicotinamide phosphoribosyltransferase
	2.0021	-2.2151	35	8.08E-08	1.55E-08	IPI00020586	CYP24	CYP24A1	3	cytochrome P450, family 24, subfamily A, polypeptide 1
	2.0121	-1.8425	12	6.95E-08	2.63E-06	IPI00307483	CYP4F11		2	Cytochrome P450 4F11
	2.1152	-1.9446	10	1.41E-08	7.03E-07	IPI00855785	FN	FN1	18	fibronectin 1
	2.2200	-2.5235	20	2.58E-09	1.13E-10	IPI00015972	COX6C		1	cytochrome c oxidase subunit VIc
	2.2246	-1.7059	8	2.39E-09	1.38E-05	IPI00001869	PAPPA		2	Pregnancy-associated plasma protein A
	2.2272	-1.4248	13	2.28E-09	0.0002896	IPI00219532	ELMO1	KIAA0281	6	engulfment and cell motility 1
	2.2431	-2.8660	6	1.75E-09	2.32E-13	IPI00011770	NADH-ubiquinone oxidoreductase		2	NADH-ubiquinone oxidoreductase MLRQ subunit
	2.8115	-3.1326	11	3.97E-14	1.11E-15	IPI00027681	NNMT	hCG_39357	1	nicotinamide N-methyltransferase
	3.0884	-1.2622	51	9.30E-17	0.00134368	IPI00607801	CES1	CES2	8	carboxylesterase 1 (monocyte/macrophage serine esterase 1)
	3.4844	-3.1484	14	6.10E-21	7.97E-16	IPI00305380	IBP4	IGFBP4	1	insulin-like growth factor binding protein 4

DNArepair

Cell cycle and mitosis progression

Aminoacids and proteins metabolism

Mitochondrial and cytochrome pathways

Cell growth and proliferation

NAD metabolism signalling

Table 5. 3. SILAC hits that reached significance in the forward and reverse experiment (legend previous page)

5.2.4.2. Revalidation of SILAC hits

5.2.4.2.1. Revalidation of selected hits by western blot in the isogenic model

Western blot remains the most appropriate way to confirm in low throughput hits detected by the SILAC experiment. Sixteen hits were selected for further revalidation, based on the potential relevance of their biological function (notably DNA repair or cell cycle) and the enrichment observed in some pathways (such as amino acids, mitochondrial or NAD metabolism) (Table 5. 3). The expression of the proteins of interest was first evaluated in the NSCLC ERCC1 isogenic model (ERCC1-WT parental A549 cell line, heterozygous cell lines, and three ERCC1-deficient clones) in order to (i) revalidate the difference detected by the high-throughput SILAC experiment and to (ii) assess the generality of what had initially been detected by a comparison between the A549 parental cell line and one ERCC1-deficient clone (Ac216) only to all ERCC1-deficient clones. Hits that were consistently revalidated among all clones in this model were then evaluated in the Ac375 ERCC1-deficient clone in which each distinct isoform had been stably reintroduced.

Among the 16 selected hits, five did not revalidate (Figure 5. 23), either because the initial differential effect was not detected by western blot (CIT), or because of lack of homogeneity of the results among ERCC1-deficient clones (UHRF1, TK1, UBE2C and CSK1). Also, the three COX proteins (COX4I1, COX5B, COX6C, respectively 19kDa, 14kDa and 8kDa) could not be detected despite of running specific western blots for low molecular weight proteins (12% Bis-Tris gel with MESH buffer), suggesting either a need for specific optimization of the antibody, or for an alternative protein extraction method. Finally, ten hits could be consistently revalidated among all clones, including NAMPT, NNMT, PARP14, DGKA, HSP90, GDA, PAPPA and IGFBP4 (with caution for the latter two proteins for which the antibody signal was suboptimal) (Figure 5. 24).

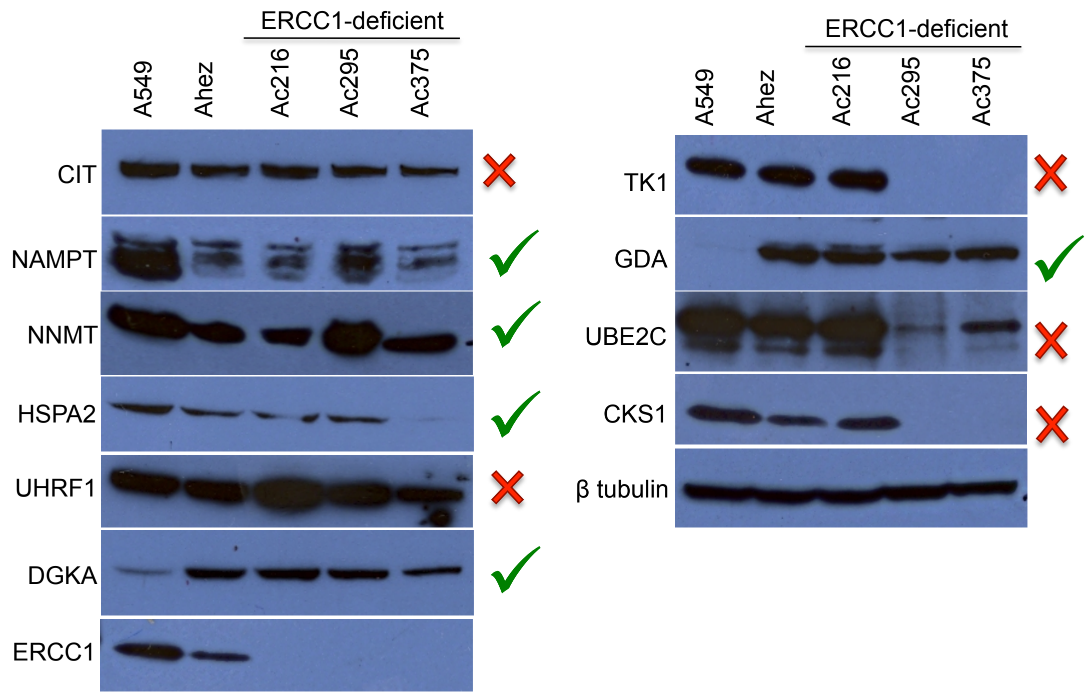


Figure 5. 23. Revalidation of SILAC hits in the ERCC1-deficient isogenic model

Western blot assessing in all three ERCC1-deficient clones some selected hits identified in the SILAC experiment. Hits that revalidated in all clones are annotated with a ✓ whereas hits that did not revalidate in all clones (even if revalidation was obtained in Ac216) are annotated with an X.

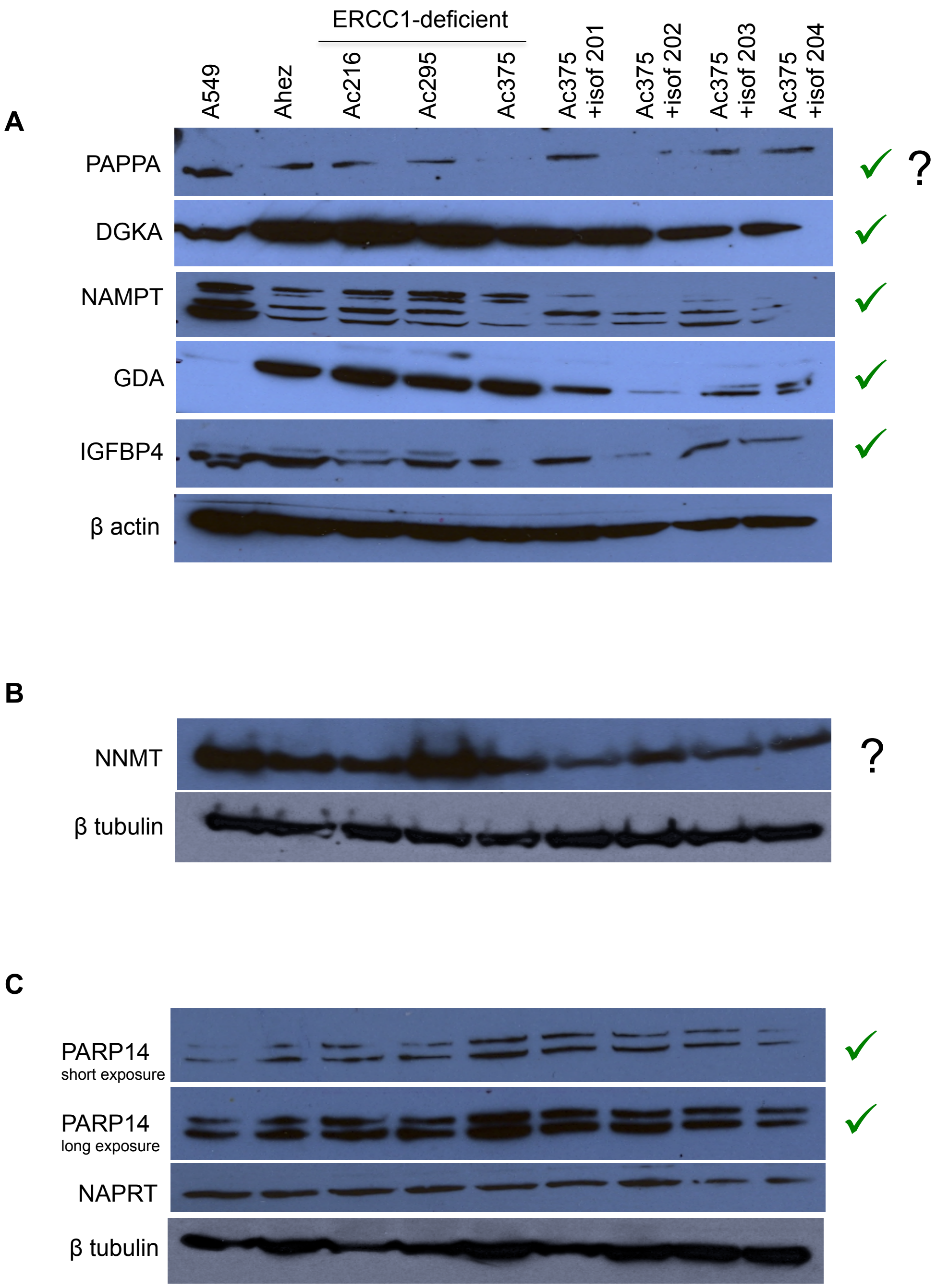


Figure 5. 24. Revalidation of SILAC hits in the ERCC1-deficient isogenic model with and without reintroduction of each ERCC1 isoform.

Western blot assessing some selected hits identified in the SILAC experiment in all three ERCC1-deficient clones and in the Ac375 ERCC1-deficient clone in which each distinct ERCC1 isoform had been stably reintroduced. Hits that revalidated in all clones are annotated with a ✓. Question marks annotate proteins for which no clear revalidation could be obtained by WB, either because of antibody issues (PAPPA) or because of inconsistencies between the three ERCC1-deficient clones (clone Ac295 for NNMT). A, B and C represent three distinct WB membranes and have therefore not been pooled to take into account loading differences. NAPRT has been added on C as this corresponded to an important determinant of the NAD pathway, of which two closely related proteins (NNMT and NAMPT) had been identified through the SILAC experiment.

5.2.4.2.2. Functional revalidation of selected SILAC hits by search for synthetic lethal interactions

In order to assess in a more functional way the potential interest of SILAC hits identified as “increased” in the ERCC1-deficient clone as compared to the ERCC1-proficient parental cell line, an alternative approach was taken in parallel to the revalidation by western blot. We hypothesized that proteins that were increased in the ERCC1-deficient clones could be “compensating” the deficiency in ERCC1 and allow cell survival in absence of ERCC1, and as such would represent potential synthetic lethal targets. Therefore, effect on cell viability of silencing these hits by siRNA was assessed.

Each cell line of the isogenic model was reverse transfected according to optimal conditions (Chapter 3) in 6-well plates, transferred into 96-well plates at day 1 and cell viability was assessed 6 days after transfection using CellTitre-Glo®. In order to limit the potential for erroneous results related to the use of the single A549 ERCC1-WT cell line from Institut Gustave Roussy, the sensitivity of A549 cells maintained at the Institute of Cancer Research was assessed in parallel. Two independent experiments were performed at the same time, each in triplicate (Figure 5. 25), where the left panel of each graph corresponds to the first experiment and the right panel to the second experiment). As shown in Figure 5. 25, an excellent transfection efficacy was obtained in all cell lines, with a similar cell kill following siPLK1 transfection (<10% cell viability). Most of the selected siRNA showed no differential effects between the ERCC1-proficient and ERCC1-deficient cells. Interestingly, two of them, UHRF1 and PARP14, had paradoxically a much more pronounced effect on cell viability of ERCC1-proficient cells than ERCC1-deficient cells. By contrast, some siRNA targeting hits that were successfully revalidated by western blot, such as GDA or DGKA, showed no differential effects.

On the basis of these revalidation results, the two most promising hits were selected for further investigation: GDA, which showed a strikingly clear difference in expression between ERCC1-deficient and ERCC1-proficient cells by western blot, and the NAMPT/NNMT couple, which were both involved in the NAD metabolism.

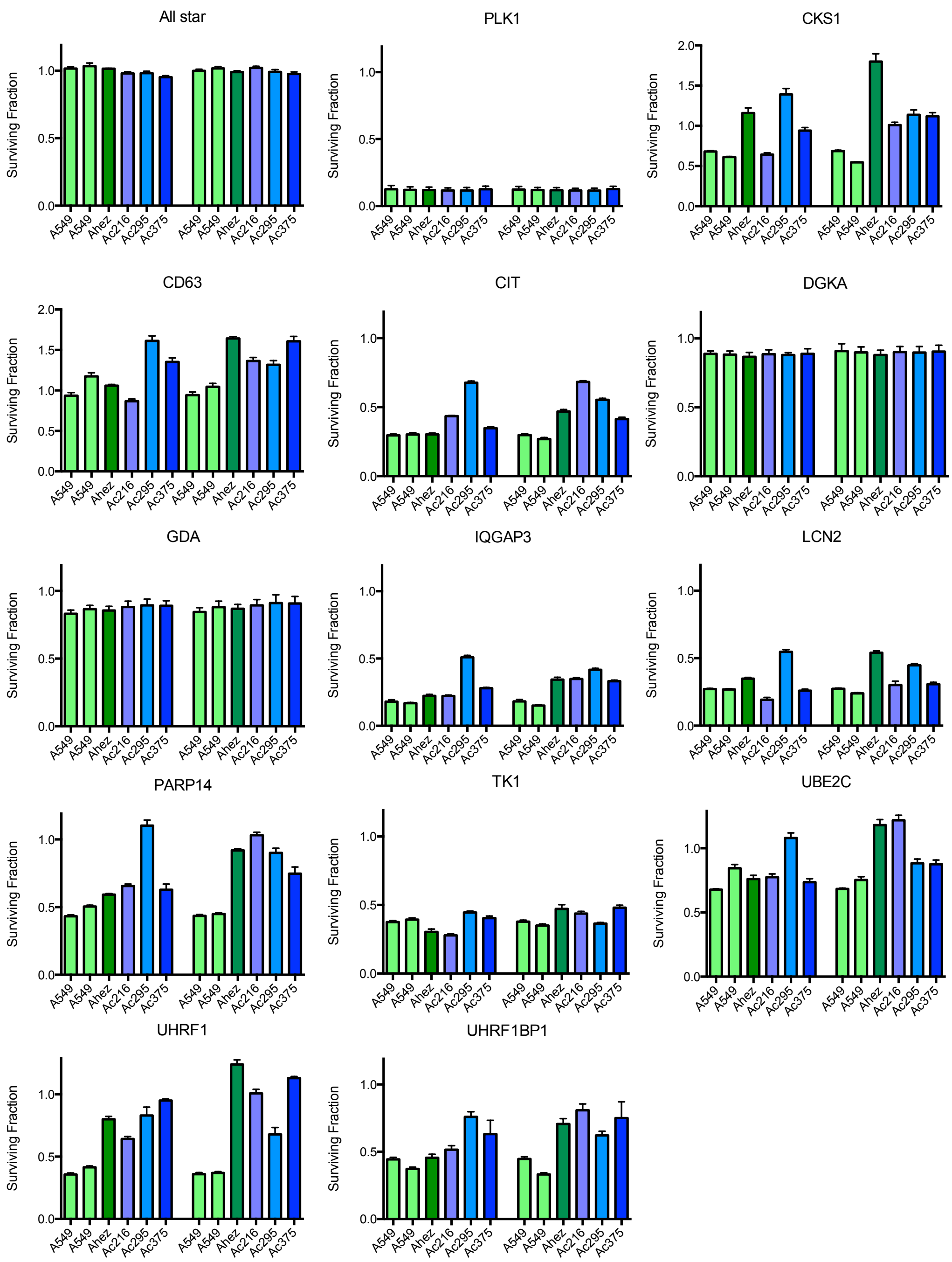


Figure 5. 25. Viability effect of silencing SILAC hits that were increased in Ac216 as compared to the parental cell line

Increased level of some SILAC hits could reflect a survival dependence of the Ac216 ERCC1-deficient clone on some of them, and increased SILAC hits therefore represented putative targets for synthetic lethal relationships with ERCC1-deficiency. In order to test this hypothesis, the effect of silencing these latter by siRNA was assessed in 2 independent experiments, performed in triplicates each (left and right panel of the bar plots). Cells were reverse transfected in 6-well plates, transferred at 12h after transferction un 96-well plates and cell viability was assessed at day 6 using CellTitre-Glo® and luminescent reading. Error bars represent the standard deviation of the mean of the triplicates.

5.2.4.2.3. Guanine Deaminase (GDA)

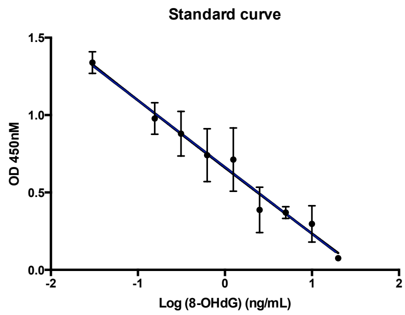
Guanine deaminase (GDA, also called guanase) is responsible for the hydrolytic deamination of guanine, thereby producing xanthine and ammonia (Yuan *et al.*, 1999), and is involved in the *de novo* synthesis of purines. This enzyme, which is highly expressed in liver, kidney, brain and placenta (Firestein *et al.*, 1999; Kubo *et al.*, 2006), is also involved in the redulation of dendrite development by modulating guanine concentrations (Firestein *et al.*, 1999).

SILAC quantitative results initially showed a 10-fold increase in GDA in the ERCC1-deficient Ac216 clone in the reciprocal experiment, consistent with a 16-fold decrease in the parental cell line in the forward experiment. Subsequent revalidation by western blot confirmed that GDA was consistently increased in all clones as compared to the parental cell line where it was undetectable (Figure 5. 23). Interestingly, GDA was also increased in the heterozygous cell line, and the reintroduction of isoform 202 only - the functional isoform with regards to all known ERCC1 functions - was associated with a decrease of GDA expression (Figure 5. 24), whereas the reintroduction of other isoforms had no effect. Although GDA silencing by siRNA had no effect on cell viability, the striking difference observed on the western blot supported further investigation for this hit.

Several hypotheses could have explained this increase in GDA, notably (i) the accumulation of unrepaired guanine lesions in the context of ERCC1-deficiency, which would in turn ultimately impair cell fitness; reducing the pool of available guanine by transforming it into xanthine would accordingly allow reducing the risk of forming the putative lesion; (ii) a metabolic switch in the absence of ERCC1, with the need for the cell to produce more xanthine, ammonia, or any downstream metabolite. Interestingly, one of the four siRNA screen hits of the ERCC1-deficient population that could be revalidated was *NUDT1* (aka *MTH1*). As described in Chapter 3, this enzyme prevents the incorporation of oxidized nucleotides into DNA / RNA by hydrolysing oxidized purine nucleoside triphosphates to monophosphates (Sakumi *et al.*, 1993; Furuichi *et al.*, 1994). The observation that ERCC1-deficient cells were more sensitive to the silencing of *NUDT1* than ERCC1-proficient cells might have suggested that ERCC1-deficient cells displayed a higher level of reactive oxygen species or that they were more sensitive to oxidative stress. 8-oxoguanine is the most commonly produced oxidized base, and, contrary to some other oxidized bases, can be incorporated into nucleic acid without blocking their synthesis (Takagi *et al.*, 2012). Transforming guanine into xanthine and ammonia using GDA would therefore allow reducing the pool of available guanine in ERCC1-deficient cells, thereby limiting the formation of 8-oxoguanine. We therefore tested whether 8-oxoguanine levels were significantly different between ERCC1-proficient and ERCC1-deficient cells, by using the Cell Biolabs® OxiSelect Oxidative DNA damage kit, which measures 8-oxoguanine levels by ELISA on 10µg of DNA. As displayed in Figure 5. 26, no significant difference could be observed between the A549 ERCC1-WT cell lines (both cell lines from IGR and ICR), the heterozygous cell line and the level of 8-oxoguanine incorporated into the DNA. Other hypotheses could not be explored as a part of the PhD because of time constraints, but further experiments are ongoing.

5.2.4.2.4. Nicotinamide Mononucleotide Phosphorybosyl-Transferase (NAMPT) and the NAD pathway

NAMPT is a rate-limiting enzyme involved in the major pathway of generation of the PARP substrate β-NAD^+. Briefly, β-NAD^+ is largely provided by a salvage pathway that utilizes nicotinamide, a by-product of the PARSylation of proteins, which is processed back to β-NAD^+ by two enzymes, NAMPT and NMNAT (Bajrami *et al.*, 2012; Chiarugi *et al.*, 2012). β-NAD^+ can also be synthesized de novo from either nutritional tryptophan (via the kynurenine pathway) or via nicotinic acid (Figure 5. 27). Two proteins of the NAD pathway, NAMPT and NNMT, came up as hits of the SILAC experiment, with a 4-fold and 10-fold decrease in the Ac216 ERCC1-deficient clone, respectively (Table 5. 3). This decrease could be consistently revalidated by western blot, and observed in all clones (Figure 5. 23). The reintroduction of each distinct ERCC1 isoform had no effect on the expression of either of these proteins (Figure 5. 24), maybe because reverting the effect required reintroducing two ERCC1-isoforms, re-expressing them in a different cellular compartment, at a higher level or with another ERCC1 partner.



	Absorbance		Interpolated C°
	rep1	rep2	(ng/mL)
A549 IGR	0.146	0.188	14.70445
A549 ICR	0.099	0.079	16.39056
Ahez	0.308	0.213	12.68327
Ac216	0.087	0.130	15.96903
Ac295	0.198	0.189	14.1316
Ac375	0.133	0.296	13.67765

Figure 5. 26. Quantification of 8-oxoguanine in the ERCC1 isogenic model

Detection of 8-oxoguanine contained in DNA of each cell line of the isogenic model. Quantification was performed by ELISA using the Cell Biolabs® OxiSelect Oxidative DNA damage kit on 10µg of DNA (excepted for Ac295 where only 8µg were available). Respective concentrations of each sample were extrapolated from the standard curve based on raw values read by spectrophotometry at 450 nm wavelength. Two reference A549 cell lines (from IGR and from ICR) were used as control population.

5.2.4.2.4.1. Evaluation of other determinants of the NAD pathway

To gain insight into potential other differences of the NAD metabolism pathway between the ERCC1-proficient and ERCC1-deficient cells, the expression of siRNA screen hits (Chapter 3) was evaluated by western blot. Three hits were selected: PARP1 (siRNA killing preferentially ERCC1-proficient cells; median difference in Z-score of 8.5), SIRT1 (siRNA killing preferentially ERCC1-proficient cells; median difference in Z-score of 2.4) and PMK2 (siRNA killing preferentially ERCC1-low cells; median difference in Z-score of 2). No difference in protein expression could be observed for PARP1 and PKM2, but SIRT1 showed lower expression in the ERCC1-deficient clones, which could be rescued by the reintroduction of isoforms 202, 203 and 204 (Figure 5. 28).

5.2.4.2.4.2. Effects of NAMPT inhibition on ERCC1-deficient cells

To investigate whether diminished expression of NAMPT in ERCC1-deficient cells led to an increased dependency on NAMPT for producing β-NAD^+, the effect of NAMPT inhibition was evaluated. FK866 is a non-competitive small molecule chemical inhibitor of NAMPT that is able to reduce cellular β-NAD^+ levels (Hasmann *et al.*, 2003). ERCC1-isogenic NSCLC models were exposed to the drug for five days prior to assessment of cell viability by CellTitre-Glo® luminescent reading. A ten-fold difference in FK866 IC_{50} was observed between ERCC1-proficient and ERCC1-deficient cells (IC_{50} = 1nM, 0.09nM, 0.1nM, 0.15nM and 0.09nM for the A549, Ahez, Ac216, Ac295 and Ac375 cell lines, respectively) (Figure 5. 29. A). The reintroduction of each distinct ERCC1 isoform did not rescue FK866 sensitivity of the ERCC1-deficient clones, both in short-term and long-term assays (Figure 5. 29. B). The supplementation of the culture media with 10µM nicotinic acid rescued FK866 sensitivity in the A549 ERCC1-WT cell lines from IGR and ICR only, but had no effect on ERCC1-deficient clones or ERCC1-deficient clones in which isoforms had been re-expressed (Figure 5. 29. B). When combined with the PARP inhibitor olaparib, low concentrations of FK866 allowed significantly increasing the therapeutic window between (i) ERCC1-proficient and ERCC1-deficient cells, (ii) ERCC1-proficient cells and the ERCC1-heterozygous cell line, and (iii) ERCC1-proficient and ERCC1-deficient cells in which the ERCC1 functional isoform 202 had been reintroduced (Figure 5. 29. C). The setting of the experiment did not allow a reliable calculation of a combination index to look for increased synergy between both drugs in the ERCC1-deficient population.

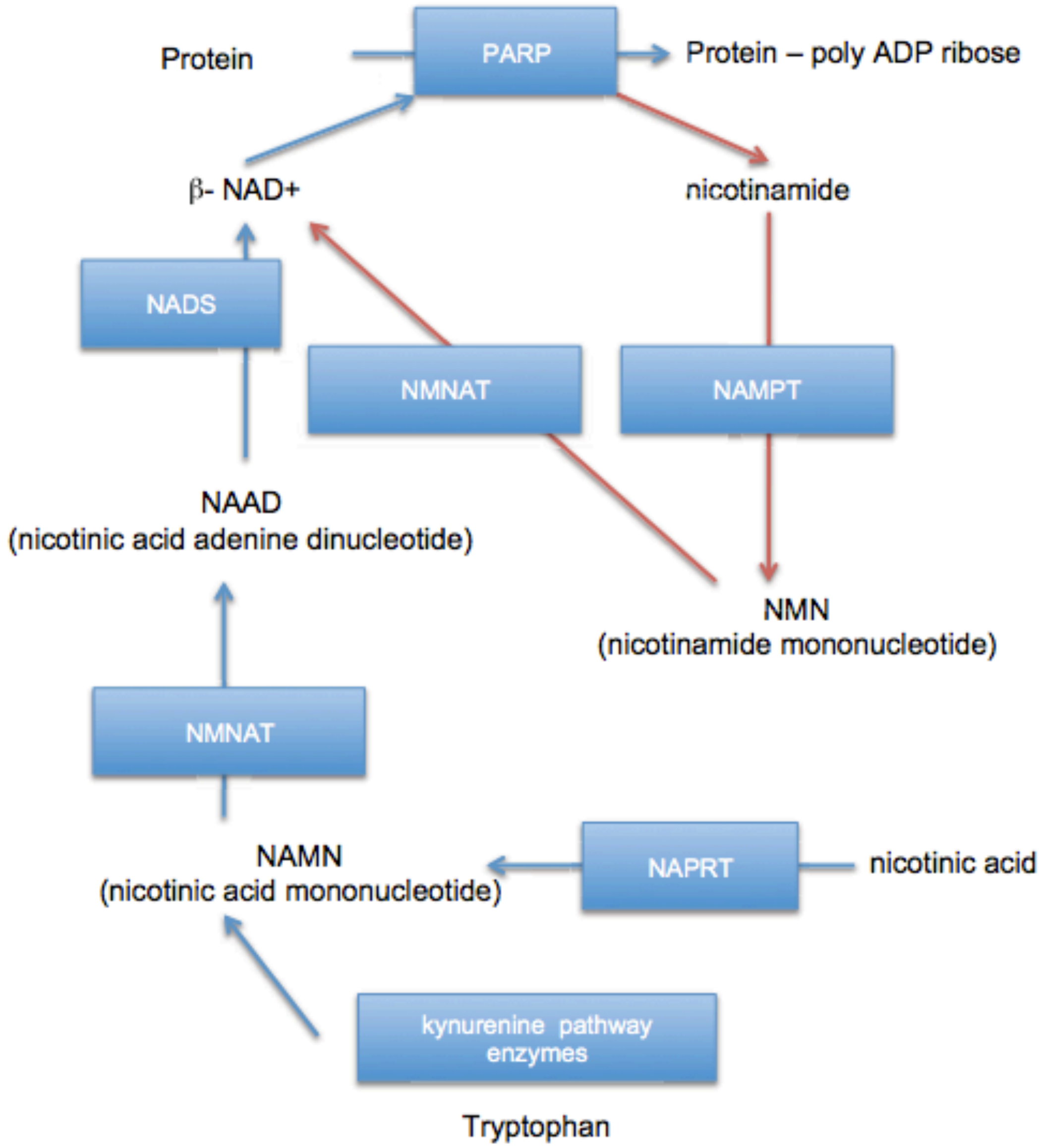


Figure 5. 27. Schematic of mammalian NAD metabolism (Figure and legends reproduced from Bajrami *et al*., EMBO 2012)

Metabolites are shown in plain text and enzymes in blue boxes. Abbreviations for enzymes are as follows: PARP, poly(ADP-ribose) polymerase; NAMPT, nicotinamide phosphoribosyl transferase; NMNAT, nicotinamide mononucleotide adenylyl transferase; NAPRT, nicotinic acid phosphoribosyl transferase; NADS, NAD synthase. PARP reactions require β-NAD+ as a substrate and generate nicotinamide as a by-product of the PARSylation of proteins. In cells where PARPs are highly active, b-NADþ is largely provided by a salvage pathway that utilizes nicotinamide. Nicotinamide is processed back to β-NAD+ by two enzymes, NAMPT and NMNAT, with NAMPT catalysis representing the rate limiting step in this process. Alternatively, β-NAD+ can be synthesized de novo from either nutritional tryptophan (via the kynurenine pathway) or via NA, which is processed by NAPRT, NMNAT and NADS.

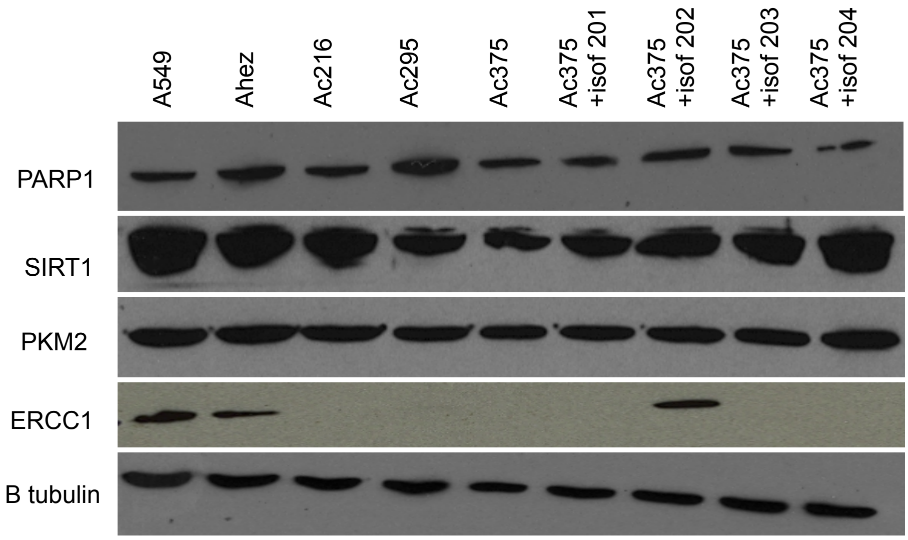


Figure 5. 28. Assessment of selected hits of the siRNA screen involved in shared cellular metabolic processes with revalidated SILAC hits

Western blot assessing selected hits of the siRNA screen that were involved in shared cellular processes with the most promising SILAC hits: PARP1 and SIRT1 (shared process with PARP14 and dependency under the NAD metabolism), and PKM2 (metabolic changes to a predominant anaerobic state, as suggested by the multiple hits from the NAD pathway or COX enzymes).

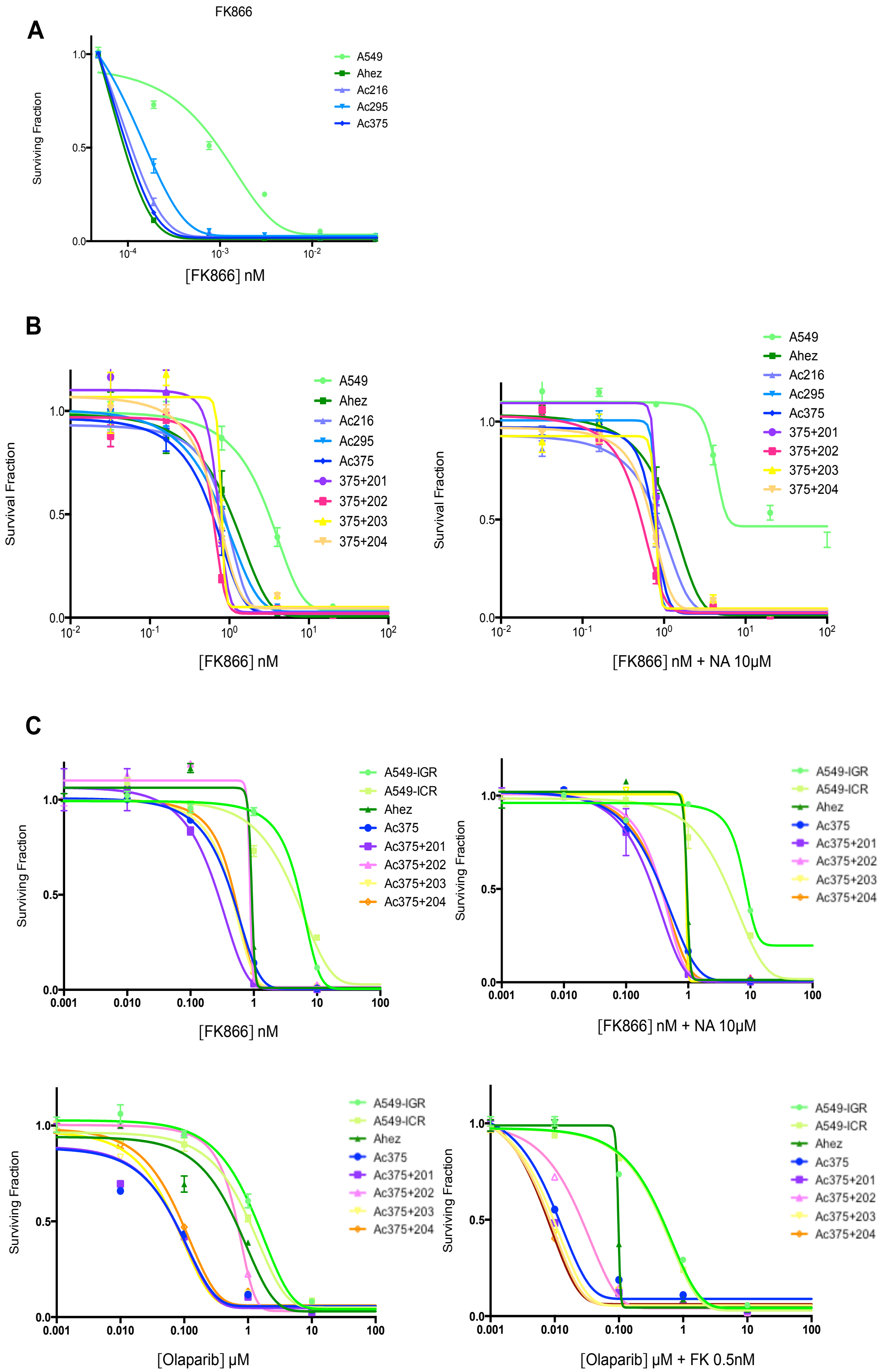


Figure 5. 29. FK866 sensitivity of the ERCC1-isogenic model (legend next page)

Figure 5. 29. FK866 sensitivity of the ERCC1-isogenic model (Figure previous page)

Assays evaluating FK866 (NAMPT inhibitor) sensitivity as monotherapy or in combination with olaparib, in A549 *ERCC1*-WT cell lines, ERCC1-deficient clones, and Ac375 cell line in which each distinct ERCC1 isoform has been stably reintroduced. The ability of nicotinic acid (NA) to rescue FK866 toxicity is also displayed. A) Short term assay displaying a therapeutic window under FK866 exposure as monotherapy between ERCC1-deficient clones and the reference *ERCC1*-WT A549 cell line. Cells were exposed to the drug for 6 days and cell viability was assessed using CellTitre-Glo® luminescent reading. B) Short term assay evaluating the ability of nicotinic acid (NA, 10µMol) to rescue FK866 sensitivity. Cells were exposed to the drug for 6 days and cell viability was assessed using CellTitre-Glo® luminescent reading. None of the ERCC1 isoforms allowed rescuing FK866 sensitivity in this setting, even in presence of NA. C) Long-term assay evaluating FK866 sensitivity in the ERCC1-isogenic model. Cells were exposed to the drug for 14 days and cell viability was assessed using CellTitre-Glo® luminescent reading. Two independent A549 cell lines (from IGR and from ICR) are displayed as ERCC1-proficient control group. FK866 allowed increasing olaparib ERCC1-selective effect at non-toxic concentrations (bottom right panel); this effect was observed even for the Ac375 cell line in which the functional ERCC1 isoform had been reintroduced (isoform 202, pink line), although to a slightly lesser extent than for the other isoforms. Error bars represent the standard deviation from the mean of three independent experiments.

However, this experiment suggests that reduced NAMPT expression is associated with selective FK866 sensitivity in ERCC1-deficient cells but that these are insensitive to FK866 rescue by nicotinic acid.

5.2.4.2.4.3. NAD/NADH ratio is decreased in ERCC1-deficient clones

To gain insight into potential changes in the NAD metabolism between ERCC1-deficient and ERCC1-proficient cells, the NAD/NADH ratio was assessed in each population of the ERCC1-isogenic model. Total NAD and decomposed NADH were quantified by enzymatic cycling reaction using the Abcam® NAD/NADH detection kit. After extraction of total NAD and decomposed NADH, completion of the enzymatic cycling reaction and adjunction of the developer, samples and standards NAD and NADH concentrations were read multiple times at OD_{450}nm using a Victor X5 plate reader. Several dilutions were performed for each sample in order to ensure being within the standard curve limits, and samples concentrations were extrapolated from the standard curve realized simultaneously. When considering only the latest reading and pooling all sample dilutions for each cell line, a significant decrease in the

NAD/NADH ratio could be detected in the ERCC1-deficient clones as compared to the A549 parental cell line (median NAD/NADH ratio of A549 cell line: 5.6 +/- 0.7; median NAD/NADH ratio of ERCC1-deficient clones: 3.1 +/- 0.4; two-sided t-test: p=0.022) (Figure 5. 30; A-B). When pooling all four time-points of reading, this difference was even more significant (median NAD/NADH ratio of A549 cell line: 5.1 +/- 0.35; median NAD/NADH ratio of ERCC1-deficient clones: 2.8 +/- 0.3; two-sided t-test: p<0.0001) (Figure 5. 30. C-D).

5.2.4.2.5. Translatability of SILAC results to the non-isogenic model

To investigate whether results obtained in the ERCC1-isogenic model could be transposed and generalized to the non-isogenic model, the following parameters were assessed in the non-isogenic panel of 15 NSCLC cell lines: (i) correlation between ERCC1 protein expression (evaluated by western blot) and SILAC hits that were revalidated in the isogenic model; (ii) correlation between ERCC1 relative expression and FK866 sensitivity; and (iii) correlation between ERCC1 relative expression and ability of nicotinic acid to rescue FK866 sensitivity.

5.2.4.2.5.1. Evaluation of SILAC hits in the non-isogenic model

The expression of the four SILAC hits that best revalidated in the isogenic model - namely NAMPT, NNMT, GDA and PARP14 - were evaluated in the non-isogenic model together with ERCC1 expression (Figure 5. 31). No obvious correlation could be observed between the cell lines ERCC1 status, for any of these hits. Of note, GDA was very rarely expressed in cell lines from the non-isogenic panel, a result that was confirmed on different western blots. The parallel evaluation of siRNA screen hits involved in the NAD pathway (PARP1, SIRT1, and NAPRT) also demonstrated an absence of correlation with ERCC1 expression (Figure 5. 31).

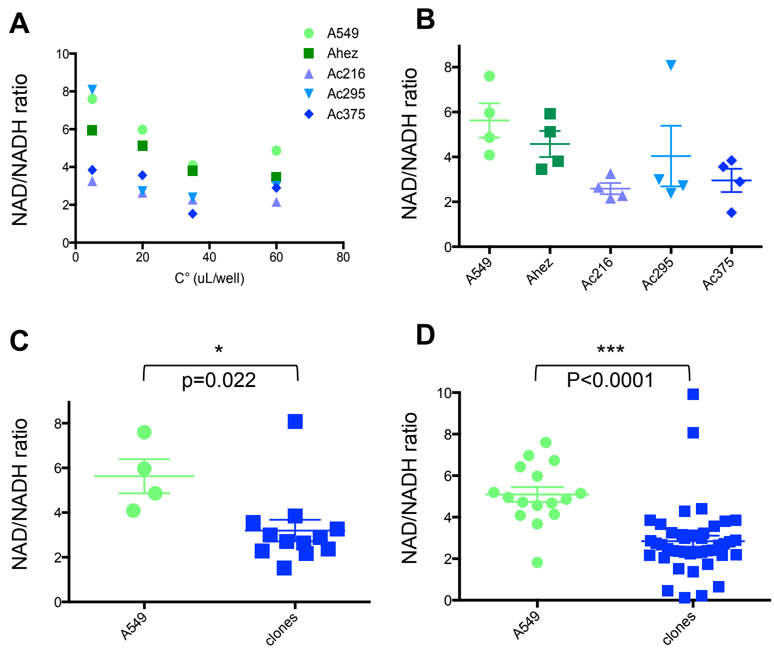


Figure 5. 30. Assessment of the NAD/NADH ratio in the ERCC1-isogenic model

Total NAD and NADH levels were measured by enzyme reaction using the Abcam® NAD/NADH kit. NAD and NADH were extracted from $2x10^5$ cells. Reading was performed at four different time points to optimize for signal detection that could vary over time, and using 4 different dilutions from the original cell extract. A) NAD/NADH ratio at different cell extract dilutions, after extrapolation from standard curve; representative example of reading at 4 hours after starting the cycling reaction. B) Median NAD/NADH ratio for each cell line; each dot represent the median NAD/NADH ratio of the different cell extract dilutions that were used for each cell line, at one time point of reading; scatter plots represent the median (with 95% standard deviation) of the four time points of reading for each cell line (1h, 2h, 4h and 5.30h after starting the cycling reaction). C) Comparison of the NAD/NADH ratio between the ERCC1-proficient cells and ERCC1-deficient clones (as a pool of Ac216, Ac295 and Ac375) at one time point of reading; representative example of reading at four hours after starting the cycling reaction; significant p-value of the two-tailed T-test between both groups: p=0.028. D) Comparison of the NAD/NADH ratio between the ERCC1-proficient cells and ERCC1-deficient clones (as a pool of Ac216, Ac295 and Ac375) at all four time points of reading (1h, 2h, 4h and 5.30h after starting the cycling reaction); significant p-value of the two-tailed T-test between both groups: p<0.0001.

5.2.4.2.5.2. ERCC1 relative expression does not correlate with FK866 sensitivity in the non-isogenic model

To investigate whether the ERCC1-selective effects observed in the isogenic model with the NAMPT inhibitor FK866 were also applicable to the non-isogenic model, the sensitivity of the panel of 13 NSCLC cell lines to FK866 was assessed. Cell lines displayed various degree of sensitivity or resistance to the drug, with IC_{50}s ranging from 0.2 to 100nM in short-term assay (6 days of drug exposure). When classifying the cell lines according to their ERCC1 expression status (ERCC1-high vs ERCC1-low), no correlation could be found (Figure 5. 32). In line with this observation, FK866 IC_{50}s could not be correlated, in a linear model, to ERCC1 relative expression (quantified by western blot; Chapter 3, Figure 3.6).

5.2.4.2.5.3. Ability of nicotinic acid to rescue FK866 sensitivity is not related to ERCC1 expression in the non-isogenic model

As FK866 sensitivity could be rescued by nicotinic acid in ERCC1-proficient cells only whereas it had no effect on ERCC1-deficient cells, we sought to determine whether this differential effect would also be observed in the non-isogenic model. As illustrated in Figure 5. 33, there was a striking difference between NSCLC cell lines that could be rescued by nicotinic acid with regards to FK866 sensitivity, and cell lines for which nicotinic acid had no effect. Overall, the sensitivity to FK866 of three out of four ERCC1-low cell lines could be rescued by nicotinic acid; rescue by nicotinic acid was also obtained in five out of eight ERCC1-high cell lines that were evaluated concomitantly, showing the absence of correlation between ERCC1 status and the potential for nicotinic acid to rescue FK866 sensitivity.

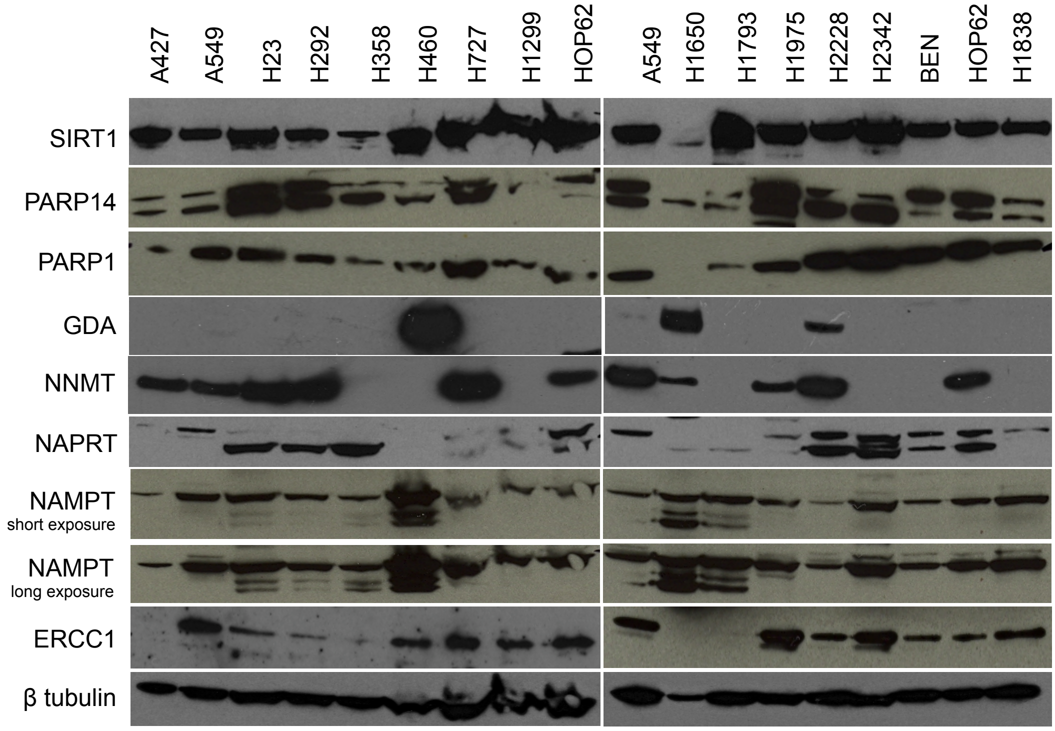


Figure 5. 31. Expression of selected SILAC hits in the non-isogenic model

Western blot assessing the expression of selected SILAC hits in the non-isogenic model. No obvious correlation could be found between any of these hits and the expression of ERCC1-isoform 202.

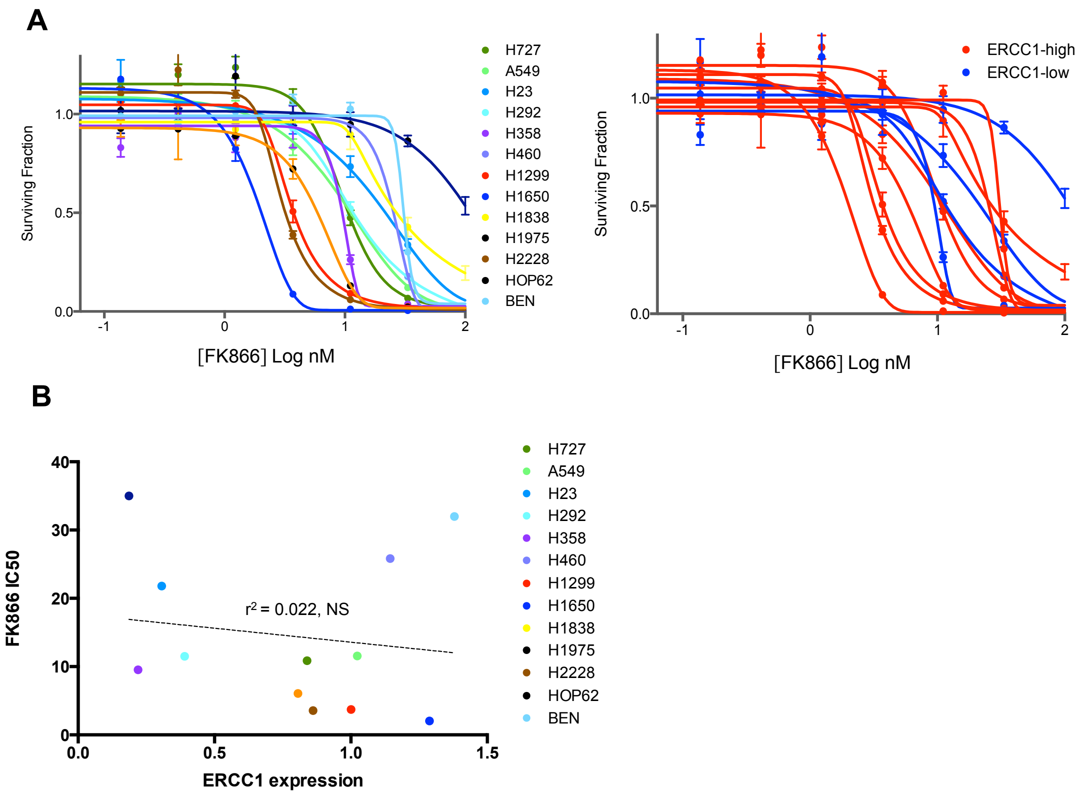


Figure 5. 32. FK866 sensitivity of the non-isogenic panel of 14 NSCLC cell lines

A) Short term assay of FK866 in the non-isogenic panel of NSCLC cell lines. Cells were exposed to the drug for 6 days and cell viability was assessed using CellTitre-Glo® luminescence reading. Non-isogenic cell lines displayed variable sensitivity to FK866 (left panel), which was not correlated to their ERCC1-isform 202 level of expression (right panel). B) Absence of significant correlation between FK866 IC_{50} and either ERCC1 relative expression (quantified by western blot; Figure 3.6).

A

Cell line	ERCC1 status	FK866 sensitivity rescue by nicotinic acid
H1975	low	no
H358	low	yes
H23	low	yes
H292	low	yes
HOP62	high	yes
H727	high	yes
H2228	high	no
A549	high	yes
H1299	high	yes
H460	high	no
H1650	high	no
BEN	high	yes

B

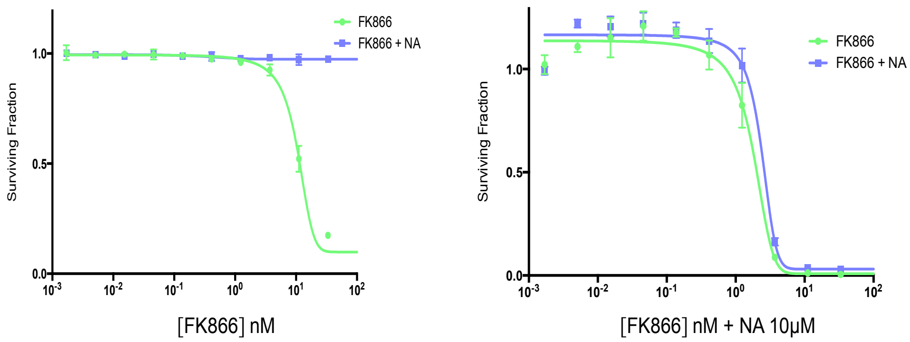


Figure 5. 33. Ability of nicotinic acid to rescue FK866 sensitivity in the non-isogenic model is not related to ERCC1 expression

A) Summary table of ERCC1 status and ability of nicotinic acid (NA, 10µMol) to rescue FK866 sensitivity. B) Representative examples of cell lines that are sensitive (left panel, H292 cell line) or insensitive (right panel, H1650 cell line) to FK866 rescue by nicotinic acid.

5.3. Discussion

Presented in this Chapter are the results of four parallel approaches aiming at investigating the presence of surrogate biomarkers of ERCC1 functionality by considering four different levels: UV-induced lesions (pyrimidine dimers), genomic scar (DNA), transcriptome (RNA) and SILAC (proteins). As illustrated by the success of RAD51 foci evaluation for assessing HR deficiency, surrogate biomarkers of proteins / pathways activity are becoming an area of growing interest (Graeser *et al.*, 2010; Lord *et al.*, 2012). Indeed, these latter offer several advantages: (i) they represent a meaningful alternative when the gene / protein of interest cannot be easily studied (e.g. long gene with several mutations of uncertain significance, such as *LKB1* (Sanchez-Cespedes, 2007); e.g. impossibility to distinguish between functional and non-functional isoforms, such as *ERCC1* (Friboulet *et al.,* 2013a); e.g. possibility of several mutations, some of which are responsible for reversion of the phenotype, such as *BRCA2* (Ashworth, 2008); (ii) they provide a direct assessment of the functionality of a pathway, which is ultimately more relevant than evaluating a single node of the pathway; this is of notable clinical importance, when the aim of developing such assay is to build a companion diagnostic / predictive test for response to chemotherapy; (iii) they allow indirectly taking into account known and unknown interacting factors accounting for modification of the pathway activity.
Results of each approach will be discussed in the following paragraphs, and potential developments of the most promising will be presented.

5.3.1. Monitoring of the repair of UV-induced cyclobutane pyrimidine dimers

The first approach for developing a surrogate biomarker of ERCC1 functionality consisted of the monitoring of the repair of UV-induced CPDs. Preliminary results obtained on cultured cells allowed identification of a difference in the accumulation and repair of CPDs between ERCC1-proficient and ERCC1-deficient cells. These findings could not be further evaluated on mice xenografts due to technical limitations that might be bypassed by using functional developments in future experiments. Although exquisite sensitivity to UV of ERCC1-deficient cells has been extensively described (Orelli *et al.*, 2010), the potential for using UV-induced lesions as a biomarker of ERCC1 activity had not been previously evaluated. Based on results obtained with cultured cells, a protocol for evaluating NER functionality through the quantification of UV-induced CPD lesions on human cancer cells was designed (Figure 5. 34). A main pitfall that had to be considered was the absence of deep tissue penetration of the UV (contrary to X-Rays used in protocols designed to evaluate HR functionality). To circumvent this problem, the use of tumour cell spreads on slides following fine needle aspiration of tumours was proposed, together with the choice of the UVB wavelength offering the best tissue penetration (Chadwick *et al.*, 1998). A pilot experiment is ongoing at the Institut Gustave Roussy using breast cancer nodule FNAs, with the help of Drs Julien Adam and Philippe Viehl, in order to evaluate (i) the feasibility of the procedure; (ii) the dose of UV that has to be delivered to detect a differential effect between NER-proficient and NER-deficient cells; (iii) the optimal incubation time required before cells fixation to best detect this difference; and (iv) the building of a semi-quantitative immunohistochemistry H-Score based on the CPDs fluorescence intensity, which would provide the most sensitive and specific classification of patients' tumours as "ERCC1-positive" or "ERCC1-negative".

5.3.2. ERCC1-associated genomic signature

All cancer genomes carry somatic mutations, a small minority of which are ''drivers'' of oncogenesis, whereas the vast majority are only ''passengers'' that have not been positively selected during the evolution of the disease (Stratton *et al.*, 2009). Driver mutations have been studied in NSCLC for more than twenty years, and the mutational landscape of key oncogenic events that becomes more and more complete every year (Oxnard *et al.*, 2013). Such mutations have come to forefront during the last decade, following the therapeutic success of targeted therapies such as EGFR or ALK inhibitors for *EGFR*-mutated (Lynch *et al.*, 2004; Paez *et al.*, 2004; Kim *et al.*, 2008; Mok *et al.*, 2009; Rosell *et al.*, 2012) or *ALK*-translocated NSCLC (Kwak *et al.*, 2010; Shaw *et al.*, 2011a; Shaw *et al.*, 2011b). Beyond these driver events, several mutational processes, resulting from DNA repair defects or exposure to carcinogens, can leave characteristic mutational signatures in cancer cells. For example, C:G > A:T transversions are characteristic of smoking-associated lung cancer, whereas C:G > T:A transitions occurring mainly at dipyrimidines and CC:GG > TT:AA double nucleotide substitutions are observed in UV light-associated skin cancers (Hainaut *et al.,* 2001; Pfeifer *et al.*, 2002; Pfeifer *et al.*, 2005).

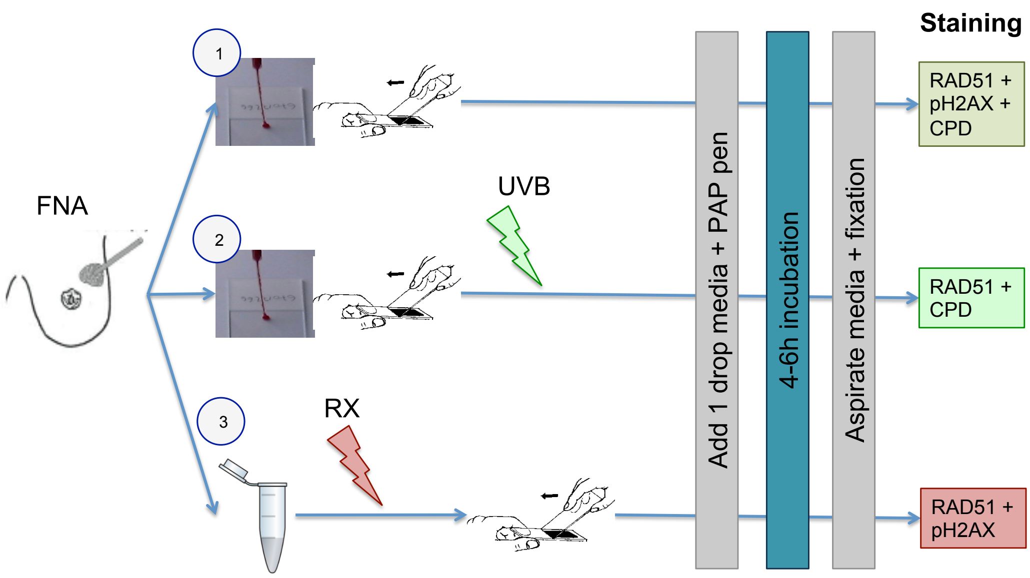


Figure 5. 34. Protocol proposal of CPD evaluation on UVB irradiated human tumour cell spreads.

After collection of human tumour cells by fine needle aspiration (FNA), three parallel processes would take place: 1) Control cells: cells spread on a glass strip, incubation for 4-6h, and triple staining with RAD51, γH2AX and CPDs. 2) Assessment of NER functionality: cell spread on a glass strip, irradiation with UVB, incubation 4-6h, and double staining with RAD51 and CPDs. 3) Assessment of HR functionality: irradiation with X-Rays followed by cells spread on a glass strip, incubation for 4-6h, and double staining with RAD51 and γH2AX.

As costs of genome sequencing are decreasing rapidly and interest for genomic scars is growing, available mutational profiling data are accumulating exponentially together with the development of computational modelling for analysing these signatures (Alexandrov *et al.,* 2013, Lawrence *et al.,* 2013; Tomasetti *et al.,* 2013). Although not directly actionable by themselves, these “fingerprints” of DNA repair deficiencies could be used as predictive biomarkers of response to DNA damaging agents.

This constituted the rationale for the second approach selected to look for a surrogate biomarker of ERCC1 functionality. The comparison of the whole exome sequencing mutational profile between ERCC1-deficient and ERCC1-proficient cells allowed highlighting a significant excess of A:T>T:A transversions in the ERCC1-deficient population, which occurred in a specific context of T-repeats or T-rich regions. Given the rarity of such mutational profile, and which is usually only associated with leukemias (acute myeloid leukemia and chronic lymphocytic leukemia (Lawrence *et al.,* 2013)), complementary analysis and revalidation in different models have to be performed to eliminate the possibility of a sequencing artefact. Ideally, revalidation in non-isogenic models should be performed in much larger datasets than the limited panel of 14 NSCLC cell lines. Also, considering the role of ERCC1 in TC-NER, the investigation of the presence of a transcriptional strand bias might bring interesting data (Stratton *et al.*, 2013). Interestingly, ERCC1 is involved in the repair of psoralen-induced lesions, which also result in similar mutation types (T>A transversions in a context of A-rich regions) (Sage *et al.*, 1993); although NSCLC cells are neither exposed to psoralen nor to UV irradiation, this observation might represent a direction for further mechanistic investigation, should this signature be confirmed. Further experiments will first aim at revalidating the signature, by studying *HPRT* mutational profile of ERCC1-deficient clones under 6-thioguanine selective pressure (Chiu *et al.*, 2006) and performing a customized exome sequencing, with deeper coverage and specific primer design, on loci presenting A:T>T:A substitutions in the first dataset. This will potentially subsequently be followed by whole genome sequencing. In parallel, analysis on NSCLC patients’ matched normal and tumour DNA will be performed, by taking as primary outcome of measure overall survival after adjuvant platinum therapy. If revalidated, such signature might then be used as predictive biomarker of platinum efficacy, and by analogy with companion diagnostic tests that have recently been developed and validated, allow customizing patient therapy to the DNA repair profile of the tumour (Figure 5. 35).

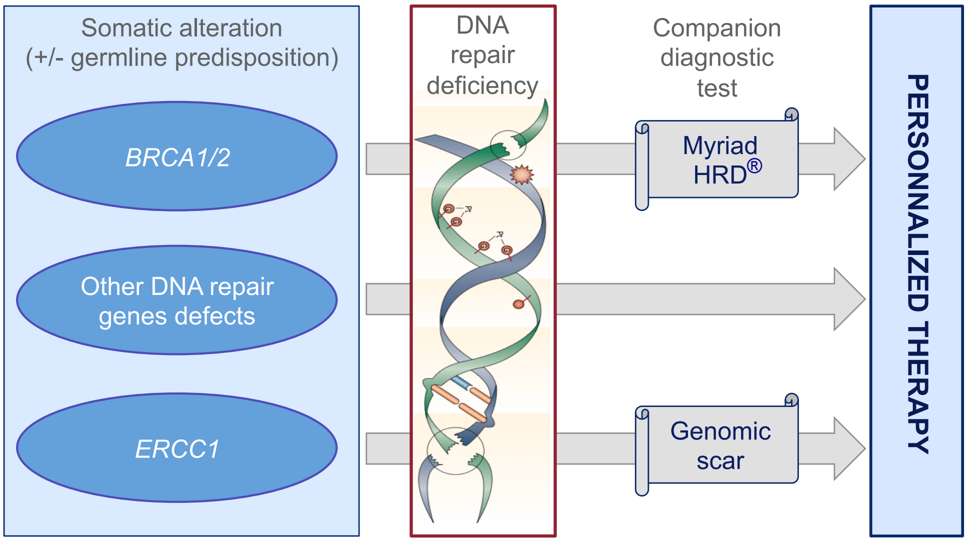


Figure 5. 35. Potential applications of an ERCC1-associated genomic scar

Specific germline or acquired somatic genetic alterations in DNA repair genes result in DNA repair defects, which in turn account for sensitivity to selected chemotherapeutic agents. When limited to a single gene mutation (ideally located ina few recurrent positions in the gene), the detection of the mutation constitutes an ideal biomarker for predicting the efficacy of the chemotherapy targeting the altered pathway. When mutations are not limited to a single gene or to a number of recurrent position in the gene (e.g. *LKB1*), or when the gene sequence is intact but its expression is altered do to epigenetic silencing (*ERCC1*), the evaluation of the genotype is much more difficult to establish. In these cases, genomic scars (or genomic signatures), which evaluate the consequence of the DNA repair defect, can constitute an alternative surrogate biomarker for predicting chemotherapy sensitivity and personalising therapy. An ERCC1 deficiency-associated genomic scar would therefore have the potential for selecting patients that would benefit from platinum-based regimens, and avoid over-treating patients that will not benefit from such therapy.

5.3.3. RNA expression changes and potential additional roles of ERCC1

The analysis of transcriptomic changes in ERCC1-deficient cells had two main aims: (i) looking for a surrogate biomarker of ERCC1 activity at the RNA level; (ii) evaluating a potential role for ERCC1 in transcription regulation which would be distinct from its role in TC-NER (Le May *et al.*, 2010; Kamileri *et al.,* 2012). As transcriptomic data were obtained at the very end of the PhD, only preliminary and descriptive results could be reported here. However, three interesting trends could be identified.

First, the comparison between ERCC1-proficient and ERCC1-deficient cells allowed the identification of several proteins involved in axon guidance and developmental processes (Table 5. 1). Interestingly, among all NER-deficient mutant mice strains, only ERCC1-mutant mice spontaneously develop neurodegeneration (Gregg *et al.,* 2011). More importantly, from a clinical point of view, there is an intriguing variability between patients presenting with germline mutations in *ERCC1*, *XPF*, or other *XP*-genes. Generally speaking, the exquisite UV sensitivity is the predominant feature of XP patients (who only present mild and late-onset symptoms of neurodegeneration), whereas half of XPF-deficient patients present severe neurodegenerative symptoms (Kraemer *et al.,* 2007). Finally, all known ERCC1-deficient patients develop severe symptoms of neurodegeneration, associated with cerebro-oculo-facio-skeletal syndrome and developmental abnormalities (reviewed in Gregg *et al.*, 2011). These differences cannot uniquely be attributed to differences in residual NER levels in these patients, suggesting additional non-DNA repair related roles for the ERCC1/XPF heterodimer.

Second, the reintroduction of all ERCC1 isoforms induced transcriptomic changes in targets involved in cell cycle and mitosis regulation. This is in line with the role of ERCC1 described in Chapter 4 (Friboulet *et al.*, 2013b), and deeper analysis is required to identify key steps that could represent therapeutic targets. Also, individual analysis of RNA expression results of each distinct isoform could bring further insight into the role of isoforms whose function remains currently unknown, i.e. isoforms 201, 202 and 203 (Friboulet *et al.,* 2013a; Friboulet *et al.,* 2013b).

Third, hierarchical clustering of the most variant genes across cell lines showed that ERCC1-heterozygous cell lines clustered with the ERCC1-deficient clones. Although this may simply result from the fact that all clones directly derived from this heterozygous cell line, it may also suggest that a lower level of ERCC1-deficiency (reduction of 35% in ERCC1 expression), which is associated with a very mild reduction in platinum sensitivity (Chapter 3), could result in significant transcriptomic changes. If validated, these changes would more sensitively correlate to variations in ERCC1 level than to DNA repair capacity measurement, and may as such represent surrogate biomarkers of ERCC1 activity.

Finally, the comparison of transcriptomic data was performed at a preliminary level (i) with SILAC data and (ii) with gene expression microarray data from untreated NSCLC samples (Friboulet *et al.*, 2011). The intersection with SILAC results evidenced 11 common hits, namely CDH1, HSPA2, CYP4F11, RAB27B, MLPH, INPP4B, COX6C, TK1, PAPPA, OCIAD2 and ALDH1A3 (Supplementary table 3). NAMPT and GDA were not identified at the transcriptomic level, which suggests that protein level differences result from differences in protein stability and degradation (e.g. at the proteasome level), rather than synthesis. The integration with transcriptomic expression results from tumour samples revealed five common hits, namely FEZ1, FAM131A, PPP2R2C, MYC and ITAG6 (Supplementary table 3). Several reasons might explain this small number of shared hits, including the current absence of reliable method for assessing the tumour ERCC1 status (i.e. a putative initial misclassification of the samples) (Friboulet *et al.*, 2013a), and the difference between results obtained in a cellular isogenic system and patient-derived tumours. Deeper analysis of RNA expression levels, intersections of available datasets as well as integration with siRNA screen results is warranted.

5.3.4. SILAC investigation: identification of GDA and changes in NAD metabolism as putative surrogate biomarkers

The study of proteomic changes through SILAC experiment allowed the identification of two main hits that prompted further investigation: guanine deaminase (GDA) and nicotinamide phosphoribosyltransferase (NAMPT).

5.3.4.1. Guanine deaminase (GDA)

Guanine deaminase catalyzes the hydrolytic deamination of guanine, producing xanthing and ammonia (according to the following reaction: Guanine + H_2O = xanthine + NH_3) (Yuan *et al.*, 1999). Initial SILAC results showed a 10-fold increase in GDA levels, which was partially erased by the reintroduction of ERCC1 functional isoform 202 (Figure 5. 24). Silencing GDA by siRNA had no effect on cell viability, suggesting either the presence of redundancy mechanisms or of potential salvage pathways to bypass the step catalysed by this enzyme; of note, appropriate silencing of the protein was not verified by western blot, and the absence of effect of the siRNA secondary to inappropriate silencing could therefore not be excluded.

Two main hypotheses were formulated to explain such increase in GDA levels. First, it was hypothesized that ERCC1-deficiency could result in an increase of oxidative lesions (or other types of bulky adducts) on guanine, that would not be repaired in the absence of ERCC1/XPF heterodimer; thus, decreasing the pool of available cellular guanine by transforming it into xanthine might allow limiting the formation of potentially toxic amino acids and, as a result, confer a selective advantage to the cell. The assessment of DNA 8-oxoguanine levels failed to demonstrate any increased in oxidized nucleotides, but the sensitivity and specificity of the kit was limited and only a single experiment could be performed. Interestingly, it has been recently reported that ERCC1/XPF could play a role in the removal of reactive oxygen species-induced 3'-blocked ends (Fisher *et al.*, 2011). Also, ERCC1-deficient cells were selectively more sensitive to the silencing of *NUDT1* (aka *MTH1*), which prevents the incorporation of oxidized nucleotides into DNA / RNA by hydrolysing oxidized purine nucleoside triphosphates to monophosphates. Whether ERCC1-deficient NSCLC cells display higher accumulation of ROS-induced DNA damage or are more sensitive to oxidative stress warrants further investigation.

The second hypothesis that may have explained an increase in GDA was the presence of metabolic changes in an ERCC1-deficient context. Notably, ERCC1-deficient cells might require increased levels of xanthine, ammonia, or any other downstream product to ensure cell survival. This is currently being investigated by performing a more comprehensive metabolomic profiling of these isogenic cell lines. If confirmed and mechanistically explained, a metabolic surrogate biomarker of ERCC1 activity – such as increase in ammonia associated with increased GDA -

could be of highest interest in the clinical setting, and would also open new therapeutic opportunities.

5.3.4.2. NAD metabolism pathway

The second most promising hit identified by the SILAC experiment were changes in the NAD metabolism pathway, notably decreased NAMPT and NNMT expression. This was associated with a selective sensitivity of the ERCC1-deficient clones to the allosteric NAMPT inhibitor FK866 (Figure 5. 29) (Bajrami *et al.*, 2012), and with a decreased NAD/NADH ratio (Figure 5. 30) in the ERCC1-deficient cells. This selective sensitivity to FK866, which was also observed in the heterozygous cell line, could not be rescued by the re-expression of any of the ERCC1 isoforms, suggesting that it might be an ERCC1-independent effect. When examining the protein-altering mutations of the exome sequencing that were present in the ERCC1-deficient clones and absent in the parental ERCC1-proficient cell line, none was related to the NAD pathway, supporting the assumption that the observed sensitivity to FK866 did not result from a secondary mutation that would have occurred during the generation process of the clones. Whether rescuing FK866 selective cytotoxicity requires the reintroduction of two ERCC1-isoforms, their re-expression in a different cellular compartment (cytoplasm, mitochondria) or at a higher level – given the observed sensitivity of the heterozygous cell line – or concomitant re-expression of another ERCC1 partner, remains to be studied.

Most published papers so far have correlated enhanced NAMPT expression to increased FK866 sensitivity, notably in breast and NSCLC cell lines (Bajrami *et al.*, 2012; Okumura *et al.*, 2012). In these publications, FK866 toxicity could be rescued by the reintroduction of nicotinic acid into the cell. The correlation observed in our ERCC1-isogenic model between increased FK866 sensitivity and lower NAMPT expression, together with the absence of rescue by nicotinic acid, contrasts with these previous reports. A first hypothesis would be that FK866 sensitivity in ERCC1-deficient cells is related to an off-target effect of the drug; this could be tested by evaluating the effects on cell viability of silencing *NAMPT* by siRNA. More interestingly, a recent publication reported that a transcriptional-dependent (TFEB) and independent (PI3K/mTORC1) activation of autophagy mediated FK866 cytotoxicity in multiple myeloma cells (Cea *et al.*, 2012). Whether the increased sensitivity of ERCC1-deficient cells to FK866 would result from a transcriptional-dependent activation autophagy (and not apoptosis), which would be facilitated in absence of ERCC1 (i.e. TC-NER), warrants further investigation. As FK866 has already been evaluated in Phase I trials in humans, with thrombocytopenia as main dose-limiting toxicity (Holen *et al.*, 2008), therapeutic applications for ERCC1-deficient tumours might be considered, providing a robust revalidation of *in vitro* preliminary results and the development of a predictive biomarker of response.

Metabolic changes described so far in liver of ERCC1-deficient mouse by transcriptomic analysis included increased cell death and anti-oxidant defences, a shift towards anabolism and reduced growth hormone/insulin-like growth factor 1 (IGF1) signalling - corresponding to a metabolic reprogramming in response to stress (Niedernhofer *et al.*, 2006). Modifications observed in NSCLC ERCC1-deficient clones included notably a decrease in IGFBP-3 and PAPPA (negative regulators of the IGF-axis; Figure 5. 24), a decreased NAD/NADH ratio (Figure 5. 30), an increased sensitivity to *PKM2* silencing (Figure 3. 20. A) and a decreased sensitivity to *HIF1AN* silencing (Figure 3.24). Taken together, these results suggest that metabolic modifications observed in malignant ERCC1-deficient cells do not recapitulate what has been described in non-malignant ERCC1-deficient mouse models, and that the Warburg effect might be increased in ERCC1-deficient cells. The comprehensive metabolomic profiling that will be performed on ERCC1-deficient cells should bring interesting complementary data.

To conclude, results of four parallel approaches seeking a surrogate biomarker of ERCC1 activity did not bring equivalent solutions and had different advantages and drawbacks. If validated, DNA sequencing and an ERCC1-associated genomic signature probably represents the first choice for several reasons: 1) the advantage of working on DNA (stability and accessibility of the material on tumour samples, circulating tumour cells, circulating DNA etc.; requirement of very low amounts); 2) the huge knowledge that has been accumulated on DNA sequencing techniques and results analysis in the last decade; 3) the price of sequencing samples, which is reducing exponentially; 4) the exportability and reproducibility of the technique; 5) the possibility of designing customized arrays evaluating only the regions of interest. The evaluation of CPD removal following UV-irradiation also represents a promising technique, but limitations such as invasiveness, accessibility of tumour to FNA, quantity of tumour cells in the sample and inter-operator variability in samples staining or immunofluorescence reading, need to be considered. Finally, metabolic changes identified by SILAC results, notably GDA increase or modifications of the NAD pathway, might represent very promising and innovative approaches, but the results obtained here are preliminary and require much more investigation and validation.

CHAPTER 6

Identification of genetic dependencies in NSCLC cell lines by functional profiling: *KRAS* as an example

6.1. Introduction

In the last decade, lung cancer landscape has moved from a simplistic binary classification distinguishing small cell lung cancer (SCLC) from non-small cell lung cancer (NSCLC), to a molecular portrait based on several driver genetic alterations, including *EGFR*, *KRAS*, *BRAF*, and *PIK3CA* mutations, *ALK* and *ROS* translocations, as well as *MET* and *ERBB2* amplifications. Recent data from the Lung Cancer Mutation Consortium, based on the analysis of nearly 1000 lung adenocarcinoma tumours, reported mutations in 54% of the samples (Kris *et al.*, 2011). As illustrated by the therapeutic success of EGFR inhibitors in *EGFR*-mutated NSCLC (Lynch *et al.*, 2004; Paez *et al.*, 2004; Kim *et al.*, 2008; Mok *et al.*, 2009; Rosell *et al.*, 2012) or ALK inhibitors in *ALK*-translocated NSCLC (Kwak *et al.*, 2010; Shaw *et al.*, 2011a; Shaw *et al.*, 2011b), there is a need to develop targeted therapies specifically adapted to the tumour molecular profile. Although the number of described alterations is growing rapidly, few customized targeted therapies have been successfully identified so far. High throughput drug and siRNA screens represent a powerful way to examine and identify genetic dependencies in cancer cells (Schlabach *et al.*, 2008; Barbie *et al.*, 2009; Iorns *et al.*, 2009; Brough *et al.*, 2011), in particular when integrated with molecular profiling data. Therefore, a comprehensive functional analysis was performed on the panel of 15 non-isogenic NSCLC cell lines, which included siRNA screens, drug screens, and whole exome sequencing.

This has produced a resource that was only partially exploited during the time of this PhD, but rather to be produced during the PhD in order to generate hypothesis and long-term developments beyond the PhD: they were as such only a satellite project of ERCC1-deficiency in NSCLC, and preliminary results that will be presented below only illustrate one of the potential developments enabled by such approach. Full analysis will require considerable additional efforts. As *KRAS* mutations are the most common change in NSCLC but no targeted therapy has been found so far (Roberts *et al.*, 2013), a pilot analysis of the screen dataset was performed after stratifying cell lines according to their *KRAS* mutational status, in order to look for *KRAS*-mutant genetic dependencies (Figure 6. 1).

6.2. Results

6.2.1. Cell lines classification according to KRAS mutational status

In order to look for oncogenic mutations in key NSCLC driver genes, whole exome sequencing was performed on 14 of the 15 non-isogenic NSCLC cell lines as previously described (Chapter 5). Cell lines were subsequently classified according to the presence of a *KRAS* activating mutation. Although there are some functional differences, mutations in *KRAS* and *NRAS* share considerable characteristics. Therefore, the *NRAS*-mutant cell line (H1299) was included in the *KRAS*-mutant group (Table 6. 1. A). Mutation profile obtained by the in-house generated exome sequencing was subsequently compared to publicly available data, using the COSMIC (cancer.sanger.ac.uk) and CCLE (Cancer Cell Line Encyclopedia, www.broadinstitute.org) databases. For most cell lines, the reported mutational status was similar in the three databases (Table 6. 1. B).

6.2.2. High throughput drug screen in non-isogenic NSCLC cell lines

To identify *KRAS*-selective agents, we performed a drug sensitivity screen using a library of 80 drugs either already used in oncology or in late stage development. Each cell line of the isogenic model was plated in 384-well plates and exposed to the drug library for five days, with each drug represented at four concentrations (Chapter 2, Material and Methods). In total, each cell line was screened in triplicate, and replica data were combined in the final analysis. CellTitre Glo® raw luminescence values were analysed using the CellHTS2 package on R v3.0.1 (www.r-project.org) software (Boutros *et al.,* 2006).

Only screens that met pre-defined quality criteria (Z' factor > 0.5 and Spearman rank correlation > 0.8, Chapter 3) were further considered. All screens were performed in triplicates with all triplicate data used for the final analysis. As described in Chapter 3, drug screen results were analysed based on the calculation of the survival fraction (SF) at each drug concentration for each cell line.

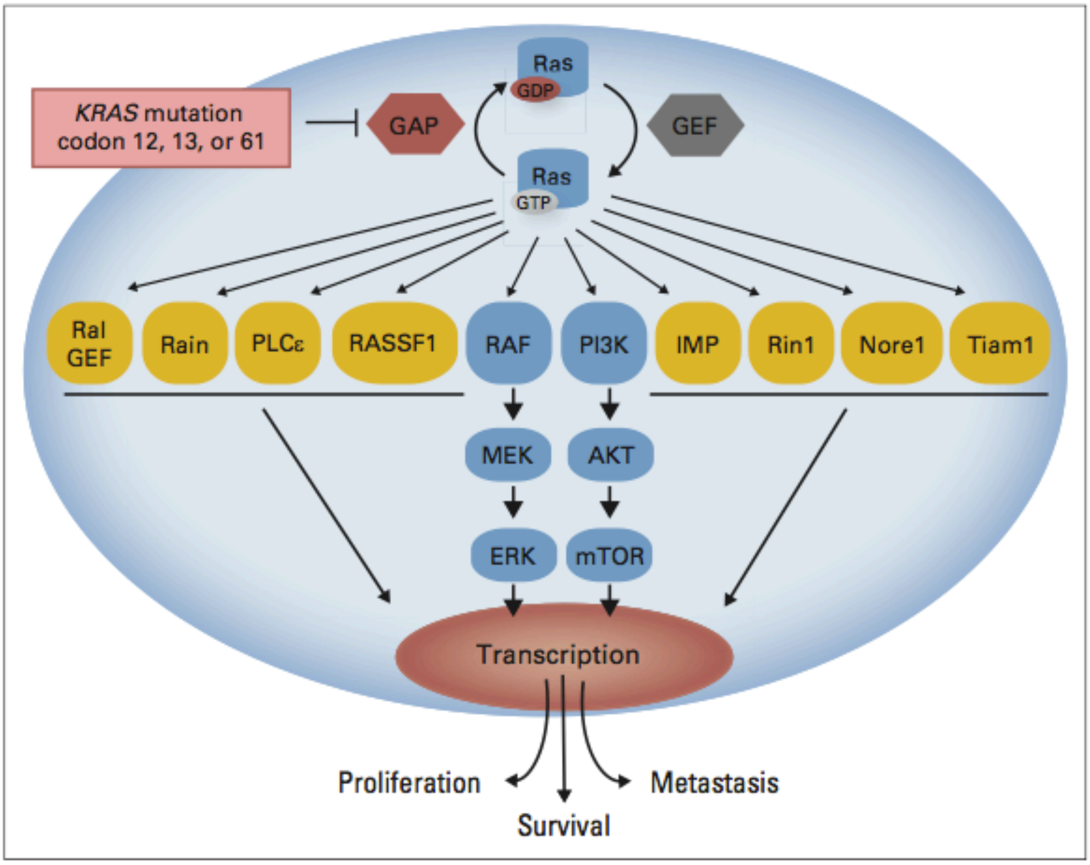


Figure 6. 1. RAS signalling cascade (Figure and legend reproduced from Roberts *et al*, Journal of Clinical Oncology 2013).

RAS proteins function as guanosine diphos-phate (GDP)/guanosine triphosphate (GTP)–regulated binary on-off switches, where GDP/GTP cycling is regulated by guanine nucleotide exchange factors (RasGEFs) that promote formation of active RAS-GTP, and GTPase-activating proteins (GAPs) stimulate GTP hydrolysis and formation of inactive RAS-GDP. In normal quiescent cells, RAS is GDP bound and inactive until extracellular stimuli cause transient formation of the active, GTP- bound form of RAS. Both wild type and mutant, activated RAS-GTP binds to a spectrum of downstream effector pathways. Mutant RAS proteins are GAP insensitive, rendering the proteins constitutively GTP bound and activated, leading to stimulus-independent, persistent activation of downstream effectors, including RAF, mitogen-activated protein kinase kinase (MEK), ERK and phosphatidylinositol 3-kinase (PI3K), AKT, mammalian target of rapamycin (mTOR) to promote cell proliferation, survival, and metastasis.

A

Cell line	KRAS status
H23	p.G12C
H358	p.G12C
HOP62	p.G12C
A427	p.G12D
A549	p.G12S
H727	p.G12V
H292	p.Q61H
H460	p.Q61H
H1299	p.Q61K
H1650	WT
H1793	WT
H1975	WT
H2228	WT
H2342	WT
BEN	WT

B

	KRAS mutation		
Cell line	COSMIC	CCLE	ICR
A427	X	NA	NA
A549	X	X	
H23	X	X	X
H460	X	X	X
H727	X	X	X
H292			X
H358	X	X	X
HI229	NRAS	NRAS	NRAS
H1650		X	
H1793			
H1975			
H2228			
H2342			
H0P62	X	NA	X
BEN			

Table 6. 1. Classification of the non-isogenic NSCLC cell lines according to KRAS mutation status

A) Final classification was obtained after merging of in-house generated data (ICR exome sequencing) with publicly available data from the COSMIC (cancer.sanger.ac.uk) and CCLE (Cancer Cell Line Encyclopedia, www.broadinstitute.org) databases. The mutation localisation and corresponding change in amino acid is displayed; B) Comparison of in-house generated data with publicly available databases. In case of discrepancies between databases, priority was given to in-house generated exome sequencing. All cell lines were STR-typed prior to exome sequencing.

Cell lines were subsequently split into two groups according to their *KRAS* mutational status, with group 1 being composed of *KRAS* and *NRAS* mutant cell lines (regardless the nature of the mutation) and group 2 encompassing all *KRAS* wild-type cell lines. In order to focus our analysis on KRAS-selective effects, we identified those drugs where there was a > 15% difference in median SF between the two above described groups.

This approach identified 13 drugs for subsequent validation (Figure 6. 2), including four different HER1/2 inhibitors. Two agents inhibiting downstream targets of the RAS pathway (one MEK inhibitor and one RAF inhibitor - sorafenib) also resulted in the list, supporting the idea that the *KRAS* mutated group of NSCLC cell lines might display some degree of pathway addiction. In addition, since comparing the median of two groups of cell lines resulted in comparing two individual cell lines, a similar analysis was performed by comparing the mean of each group. This second approach led to similar results (data not shown). Finally, a two-sided T-test was calculated in order to look for statistically significant differences in SF between the two groups; the relatively low number of cell lines (9 *RAS*-mutant *vs* 6 *RAS*-wild type) did not allow detecting highly significant effects.

6.2.3. High throughput siRNA screens in non-isogenic NSCLC cell lines

6.2.3.1. Optimisation, quality control and performance of the screen

In order to identify genetic dependencies in non-isogenic NSCLC cell lines, high throughput siRNA screens were performed, as detailed in Chapter 3. Optimal transfection conditions were first assessed for all 15 non-isogenic NSCLC cell lines: similar to what had been performed with the *ERCC1*-isogenic model, three cell plating densities and three transfection reagents were cross-evaluated for each cell line. Optimal transfection conditions were subsequently selected, which associated (i) minimal toxicity of the transfection reagent on its own; (ii) minimal toxicity of non-targeting siRNAs; (iii) more than 80% cell kill induced by siPLK1 (positive control) and (iv) cell confluency between 80-100% in mock wells at the time of reading (see Chapter 3, where more details are provided). Table 6. 2 summarises the optimal transfection conditions that were found for each cell line, and illustrates that the selected NSCLC cell lines did not transfect equally well. In order to investigate potential dependencies upon a large number of genes, several siRNA libraries were evaluated, including the kinome - tumour suppressor - DNA repair (1008 siRNAs) and the metabolic – phosphatase (978 siRNAs) libraries. Table 6. 3 recapitulates libraries that were evaluated for each cell line, together with the quality control of each corresponding screen (as evaluated by the Z' factor and Spearman rank correlation – Cf Chapter 3). For some cell lines, different libraries were run at different times, which explains that quality controls are displayed separately.

6.2.3.2. High throughput siRNA screen analysis

In order to investigate *KRAS* synthetic lethal genetic dependencies, cell lines were combined into two groups according to their *KRAS* mutational status (see Table 6. 1) and results of all screens were computed to create a matrix containing the results from all cell lines results. This approach has been previously described and successfully used by our group to identify genetic functional dependencies (Brough *et al.*, 2011).

Briefly, normalized Z-scores adjusted to the median absolute deviation (MAD) of the screen were calculated for each siRNA and for each cell line. Of note, all siRNA libraries could not be screened at the same time for each cell line (Table 6. 3); for example, the kinome and tumour suppressor libraries were run separately from the DNA repair library for the H23 cell line, as opposed to the A549 cell line where both libraries were run together. Screens that are not processed at the same time, even if performed on the same cell line, should theoretically be considered as two independent screens, as the variance and distribution of the results may vary between screens; as such, these screens should theoretically not be merged.

Figure 6.2. Results of the high-throughput drug screen – Hits meeting pre-defined selectivity criteria (figure next page)

Names in the grey box on the bottom right corner of each plot represent display the targets of each drug. A) Four hits belong to the HER inhibitors family; B) Three hits corresponded to downstream targets of pathways sharing common steps with the RAS/RAF/MEK pathway, as well as pathways responsible for functional redundancy in driving tumour growth. These were considered as positive controls of the quality of the screen. B) Six hits that reached significance were not previously described as being selective for KRAS-mutant cells in NSCLC, with the exception of IKKε, which belongs to the NF-κB pathway recently identified as essential for *KRAS* mutant tumours survival (Barbie *et al.*, Nature 2009). Error bars represent the standard deviation of the median of each group.

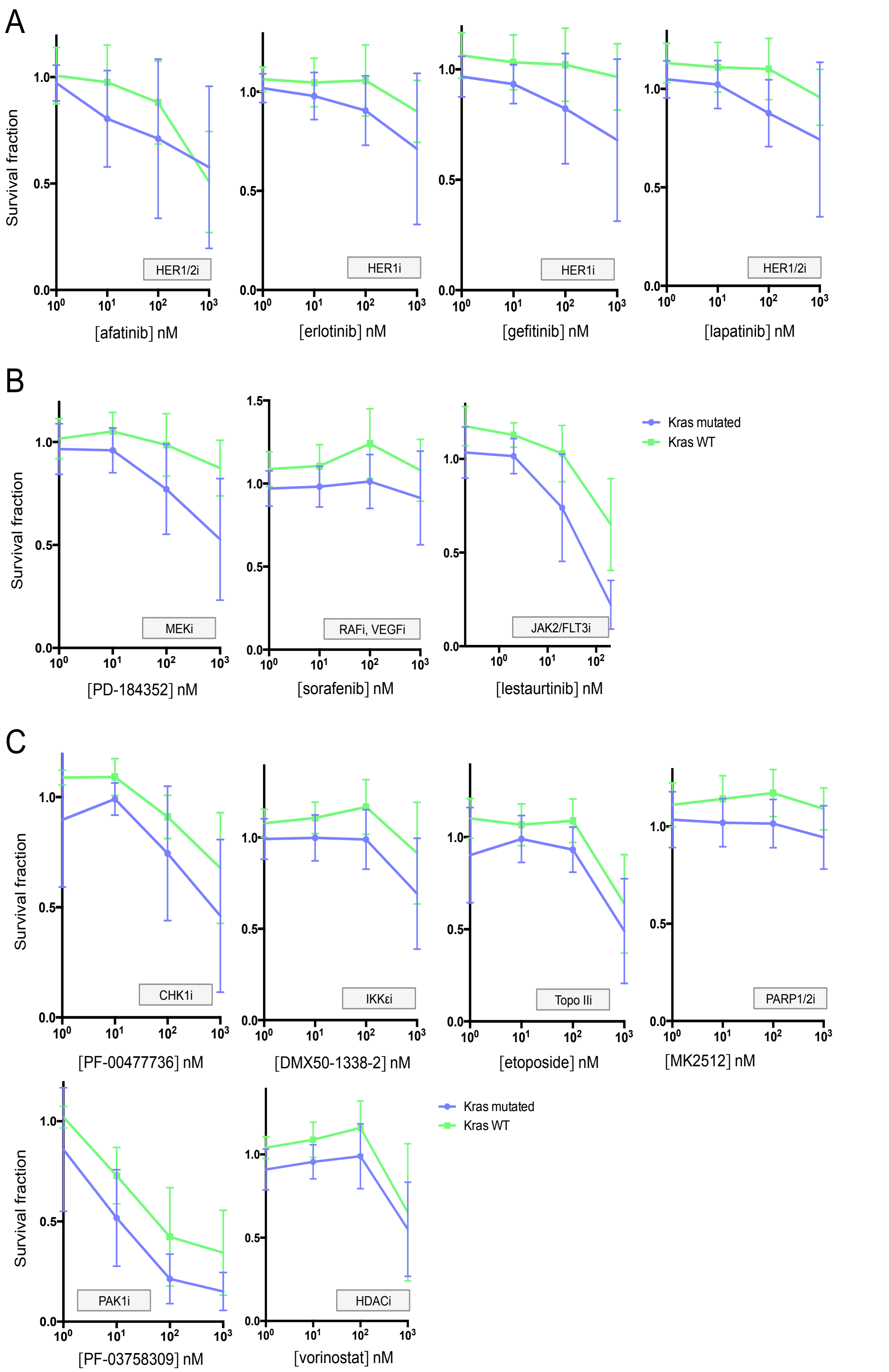


Figure 6. 2. Results of the high-throughput drug screen – Hits meeting pre-defined selectivity criteria (legend previous page)

Cell line	# of cells/w	Lipofectamine 2000 0.05µl/w	RNAimax 0.05µl/w	Dharmafect 3 0.05µl/w
A427	250			
	500	✓✓		
	1000			
A549	250			
	375	✓✓✓	✓✓✓	
	1000			
H23	250			✓✓
	500			
	1000			
H292	250			
	500			✓✓
	1000			
H358	250			
	500			✓✓
	1000			
H460	250	✓✓✓	✓✓✓	
	500			
	1000			
H727	250			
	500		✓✓	
	1000		✓✓	
H1299	250	✓✓✓		
	500	✓✓✓		
	1000			
H1650	250			
	500	✓		
	1000			
H1793	250	✓		
	500			
	1000			
H1975	250			
	500		✓	
	1000			
H2228	250			
	500	✓		
	1000			
H2342	250			
	500		✓✓	
	1000			
BEN	250			
	500		✓✓	
	1000			
HOP62	250			
	500	✓✓✓		
	1000			

Table 6. 2. Optimal reverse-transcription conditions summary of the 15 non-isogenic NSCLC cell lines

Non-isogenic NSCLC cell lines were transfected either with one the negative siRNA controls (siCON1, siCON2 or Allstar), with the positive control (PLK1), or were plated in wells where no siRNA was added (Mock). For each cell line, three different transfection reagents and three different cell densities at plating were assessed. Cell viability was assessed after 7 days using CellTitre-Glo and luminescence reading on a Victor X5 plate reader, and mean raw luminescence values were normalised to the Mock. Optimal conditions were further selected as described in Figure 3.17. Selected optimal reverse-transfection conditions are notified by a ✓, with the number of ✓ reflecting the ability of the cell line to be transfected (✓✓✓ = easily transfected cell line; ✓ = cell line very difficult to transfect). Cell line, H2228, was particularly difficult to transfect and optimal conditions could hardly be obtained.

		Spearman's rank correlation				Z prime		
Library	Cell line	plate 1	plate 2	plate 3	plate 4	rep 1	rep 2	rep 3
KTSDR	A427	0.94-0.98	0.94-0.98	0.94-0.96		0.61	0.46	0.71
	A549 Fr	0.9-0.94	0.82-0.88	0.86-0.94		0.58	0.63	0.45
	BEN	0.93-0.94	0.91-0.93	0.92-0.94		0.65	0.72	0.62
	H727	0.96-0.97	0.87-0.93	0.96-0.97		0.8	0.63	0.72
	H1650	0.88-0.92	0.93-0.94	0.87-0.92		0.18	0.13	0.09
	H1793	0.73-0.79	0.84-0.99	0.8-0.87		0.24	0.03	0.31
	H1838	0.9-0.92	0.92-0.93	0.92-0.93		0.49	0.58	0.36
	H1975	0.89-0.93	0.89-0.93	0.91-0.92		0.42	0.4	0.27
	H2228	0.92-0.94	0.69-0.95	0.91-0.95		0.48	0.72	0.71
	H2342	0.81-0.86	0.76-0.79	0.75-0.77		0.59	0.56	0.58
	HOP62	0.65-0.68	0.75-0.82	0.68-0.72		0.54	0.49	0.61
	H460	0.87-0.91	0.77-0.83	0.75-0.87		0.1	0.53	0.14
	H1299	0.93-0.96	0.94-0.95	0.81-0.87		0.7	0.77	0.58
KTS	H23	0.86-0.87	0.82-0.87	0.71-0.83		0.34	0.4	0.45
	H292	0.87-0.96	0.84-0.95	0.71-0.94		0.63	0.61	0.61
	H358	0.93-0.95	0.93-0.95	0.72-0.83		0.54	0.37	0.16
DR	H23	0.96-0.97				0.57		
	H292	0.84-0.9				0.75		
	H358	0.94-0.95				0.8		
Mbolptase	A427	0.89-0.91	0.88-0.91	0.92-0.92	0.89-0.93	0.17	0.42	0.25
	A549 Fr	0.86-0.9	0.94-0.94	0.85-0.9	0.91-0.93	0.58	0.66	0.69
	A549 UK	0.88-0.91	0.93-0.94	0.93-0.94	0.93-0.94	0.57	0.66	0.73
	BEN	0.82-0.84	0.85-0.9	0.85-0.91	0.9-0.94	0.15	0.34	0.19
	H23	0.75-0.77	0.74-0.82	0.73-0.85	0.85-0.91	0.18	0.06	0.43
	H292	0.68-0.81	0.67-0.7	0.44-0.69	0.72-0.77	0.63	0.03	0.19
	H358	0.62-0.63	0.6-0.79	0.71-0.8	0.8-0.82	0.11	0.18	0.23
	H460	0.7-0.74	0.89-0.89	0.7-0.83	0.77-0.82	0.38	0.46	0.56
	H727	0.63-0.73	0.71-0.78	0.79-0.8	0.77-0.83	0.21	0.06	0.4
	H1299	0.88-0.93	0.96-0.97	0.91-0.95	0.96-0.97	0.8	0.79	0.77
	H1650	0.77-0.8	0.78-0.83	0.58-0.69	0.76-0.8	0.24	0.28	0.69
	H1793	0.55-0.63	0.55-0.68	0.51-0.64	0.85-0.88	0.28	0.5	0.1
	H1975	0.67-0.77	0.81-0.92	0.87-0.92	0.94-0.96	0.64	0.7	0.3
	H2342	0.58-0.64	0.64-0.78	0.57-0.63	0.7-0.71	0.03	0.07	0.14
	HOP62	0.7-0.72	0.58-0.65	0.59-0.78	0.86-0.89	0.72	0.28	0.38

Table 6. 3. High-throughput non-isogenic siRNA screens Quality Control

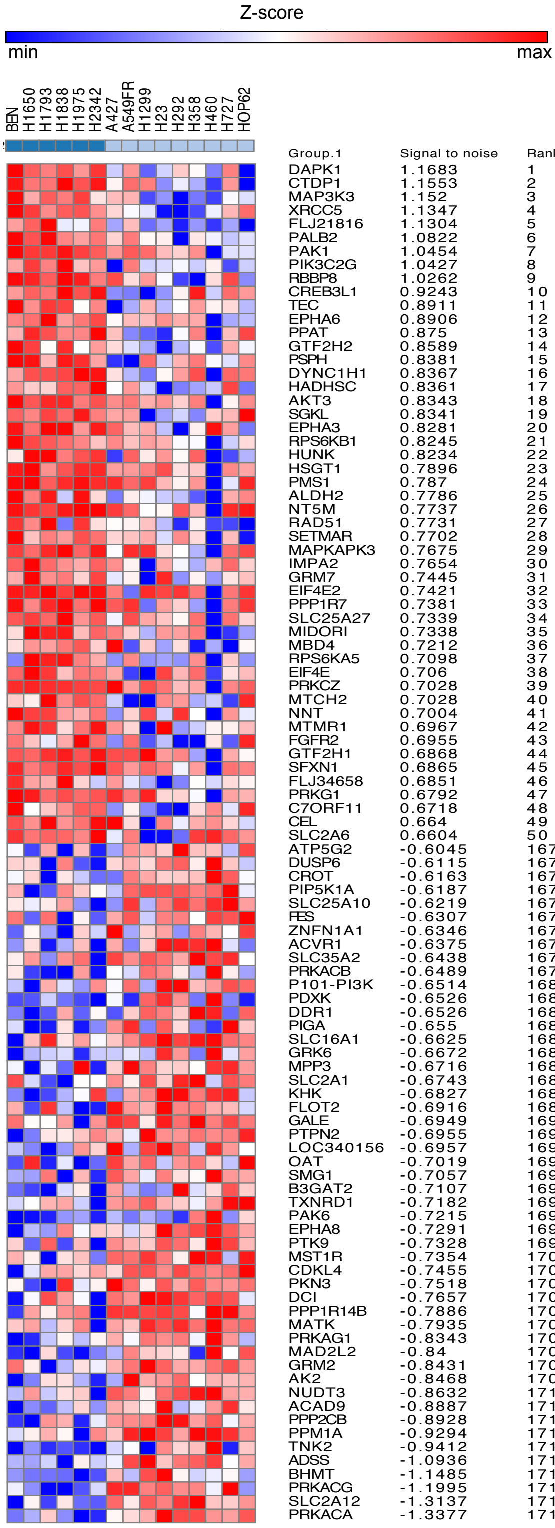
Quality Control of the siRNA screens performed in the non-isogenic model, which were included in further analysis. Most screens met pre-defined excellent quality criteria, including Spearman rank correlation > 0.8 and Z' > 0.5. A minimal median Spearman rank correlation of 0.6 and Z prime factor of 0 were considered as acceptable for cell lines that were very difficult to transfect (see example of H1793). Screens are separated according to their library of origin, as different libraries were sometimes run at two different times for some cell lines (e.g. the KTS library was run separately from the DR library for three cell lines, precluding from merging the QC results). KTSDR: Kinome – Tumour suppressors – DNA repair; KTS: Kinome – Tumour suppressors; DR: DNA repair; Mbolptase: Metabolic – phosphatase libraries

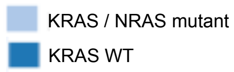
However, the observation that (i) PLK1 siRNA produced equally toxic effects in two independent screens of the same cell line, and that (ii) raw values were comparable, which allowed some confidence in the comparability of the independent screens and consequently the integration of both datasets in the final analysis. This allowed the construction of a siRNA screen matrix, in which each Z-score value corresponded to the effect of a unique siRNA on a given cell line.

Two parallel approaches were subsequently undertaken to analyse the whole dataset: first, Z-scores were processed into GENE-E (www.broadinstitute.org/cancer/software/GENE-E) where the comparison of the two groups (*KRAS* and *NRAS*-mutant *vs* *KRAS*-wild type) allowed the generation of a heatmap of all Z-score values (Figure 6. 3), together with accompanying statistical analysis (significance, false discovery rate and family-wise error rate); second, the difference between the median Z-score of each group was calculated and hits were subsequently selected based on a median Z-score difference > 1.5. This was completed by the calculation of a two-sided T-test between both groups of cells to ease identification of potentially interesting hits (Table 6. 4).

Figure 6. 3. Heatmap of non-isogenic siRNA screen hits after marker selection according to *KRAS* and *NRAS* mutational status (figure next page)

Heatmap of the top 100 hits from the siRNA screen after marker selection (see below) according to *KRAS* and *NRAS* mutational status, generated using the GENE-E software (Broad Institute). By analogy with expression arrays, each square of the heatmap represents the Z-score value obtained after analysis and normalisation of raw data values (with a blue square representing a low Z-score, i.e. a killing effect of the siRNA; red squares reflect the absence of effect on cell survival of the siRNA of interest). Marker selection identifies objects that are differentially expressed between two classes. For each object, the analysis uses a test statistic to calculate the difference in expression between the classes and then estimates the significance (p-value) of the test score. It then corrects for multiple hypotheses testing (MHT) by computing both the false discovery rate (FDR) and the family-wise error rate (FWER). The output of marker selectionprovided by GENE-E consists of: Score: The calculated value of the test statistic; p-value: the estimated significance of the test statistic for this row (not yet corrected for MHT); FDR (BH): the expected proportion of non-marker genes (false positives) within the set of genes declared to be differentially expressed. It is estimated using the Benjamini and Hochberg procedure; FWER: the probability of having any false positives.





Group.1	Signal to noise	Rank	p-value	FDR(BH)	FWER	maxT	Fold Change
DAPK1	1.1683	1	0.001996	0.2144	0.544	0.543	3.3868
CTDP1	1.1553	2	0.001996	0.2144	0.566	0.565	-35.9475
MAP3K3	1.152	3	0.003992	0.2214	0.575	0.574	-20.1181
XRCC5	1.1347	4	0.003992	0.2214	0.632	0.631	0.339
FLJ21816	1.1304	5	0.001996	0.2144	0.643	0.641	-3.3522
PALB2	1.0822	6	0.001996	0.2144	0.77	0.768	-4.5294
PAK1	1.0454	7	0.001996	0.2144	0.856	0.855	-5.5422
PIK3C2G	1.0427	8	0.001996	0.2144	0.861	0.86	-11.5942
RBBP8	1.0262	9	0.003992	0.2214	0.883	0.882	-2.7594
CREB3L1	0.9243	10	0.01198	0.3743	0.986	0.986	-2.7931
TEC	0.8911	11	0.003992	0.2214	0.993	0.993	2.4511
EPHA6	0.8906	12	0.001996	0.2144	0.993	0.993	-2.5037
PPAT	0.875	13	0.007984	0.305	0.998	0.998	-14.667
GTF2H2	0.8589	14	0.001996	0.2144	0.998	0.998	0.8157
PSPH	0.8381	15	0.01397	0.3921	0.998	0.998	1.4741
DYNC1H1	0.8367	16	0.01796	0.401	0.998	0.998	-3.8365
HADHSC	0.8361	17	0.005988	0.2709	0.998	0.998	3.3953
AKT3	0.8343	18	0.001996	0.2144	0.999	0.999	-46.2307
SGKL	0.8341	19	0.01198	0.3743	0.999	0.999	0.1908
EPHA3	0.8281	20	0.003992	0.2214	0.999	0.999	-8.9675
RPS6KB1	0.8245	21	0.003992	0.2214	0.999	0.999	-31.6475
HUNK	0.8234	22	0.01996	0.4236	0.999	0.999	-2.6649
HSGT1	0.7896	23	0.003992	0.2214	1	1	7.1016
PMS1	0.787	24	0.007984	0.305	1	1	3.9823
ALDH2	0.7786	25	0.01597	0.3921	1	1	3.7967
NT5M	0.7737	26	0.003992	0.2214	1	1	-5.2288
RAD51	0.7731	27	0.01796	0.401	1	1	-1.8784
SETMAR	0.7702	28	0.005988	0.2709	1	1	0.5203
MAPKAPK3	0.7675	29	0.02595	0.4664	1	1	8.2162
IMPA2	0.7654	30	0.005988	0.2709	1	1	9.1632
GRM7	0.7445	31	0.01397	0.3921	1	1	0.4044
EIF4E2	0.7421	32	0.001996	0.2144	1	1	6.179
PPP1R7	0.7381	33	0.02794	0.4664	1	1	16.8436
SLC25A27	0.7339	34	0.00998	0.3574	1	1	-2.9223
MIDORI	0.7338	35	0.02395	0.4626	1	1	2.6695
MBD4	0.7212	36	0.02395	0.4626	1	1	3.3238
RPS6KA5	0.7098	37	0.01597	0.3921	1	1	42.9575
EIF4E	0.706	38	0.01597	0.3921	1	1	-7.3902
PRKCZ	0.7028	39	0.003992	0.2214	1	1	-14.9696
MTCH2	0.7028	40	0.04192	0.51	1	1	0.3425
NNT	0.7004	41	0.02994	0.4765	1	1	-13.2076
MTMR1	0.6967	42	0.01597	0.3921	1	1	0.4751
FGFR2	0.6955	43	0.02994	0.4765	1	1	-3.0818
GTF2H1	0.6868	44	0.01796	0.401	1	1	3.6594
SFXN1	0.6865	45	0.003992	0.2214	1	1	0.8289
FLJ34658	0.6851	46	0.00998	0.3574	1	1	0.9745
PRKG1	0.6792	47	0.01198	0.3743	1	1	4.9087
C7ORF11	0.6718	48	0.02794	0.4664	1	1	-2.2633
CEL	0.664	49	0.02595	0.4664	1	1	0.147
SLC2A6	0.6604	50	0.06986	0.6096	1	1	-5.0792
ATP5G2	-0.6045	1670	0.04391	0.51	1	1	9.8466
DUSP6	-0.6115	1671	0.03792	0.5015	1	1	0.2227
CROT	-0.6163	1672	0.03792	0.5015	1	1	6.9946
PIP5K1A	-0.6187	1673	0.03992	0.51	1	1	0.1559
SLC25A10	-0.6219	1674	0.02994	0.4765	1	1	49.1643
FES	-0.6307	1675	0.0519	0.5473	1	1	3.7546
ZNFN1A1	-0.6346	1676	0.03593	0.5015	1	1	1.0139
ACVR1	-0.6375	1677	0.05389	0.5581	1	1	0.6775
SLC35A2	-0.6438	1678	0.02794	0.4664	1	1	1.0279
PRKACB	-0.6489	1679	0.02595	0.4664	1	1	0.7944
P101-PI3K	-0.6514	1680	0.02395	0.4626	1	1	5.3642
PDXK	-0.6526	1681	0.04391	0.51	1	1	-2.2309
DDR1	-0.6526	1682	0.04391	0.51	1	1	2.5191
PIGA	-0.655	1683	0.03792	0.5015	1	1	12.8727
SLC16A1	-0.6625	1684	0.01597	0.3921	1	1	1.5726
GRK6	-0.6672	1685	0.02994	0.4765	1	1	2.6542
MPP3	-0.6716	1686	0.03792	0.5015	1	1	2.7909
SLC2A1	-0.6743	1687	0.01796	0.401	1	1	6.1473
KHK	-0.6827	1688	0.02196	0.4389	1	1	0.2355
FLOT2	-0.6916	1689	0.01597	0.3921	1	1	1.3328
GALE	-0.6949	1690	0.01597	0.3921	1	1	0.5588
PTPN2	-0.6955	1691	0.01796	0.401	1	1	0.169
LOC340156	-0.6957	1692	0.01597	0.3921	1	1	12.4763
OAT	-0.7019	1693	0.03593	0.5015	1	1	0.2096
SMG1	-0.7057	1694	0.02994	0.4765	1	1	2.1472
B3GAT2	-0.7107	1695	0.01996	0.4236	1	1	-3.1572
TXNRD1	-0.7182	1696	0.01397	0.3921	1	1	0.9189
PAK6	-0.7215	1697	0.02196	0.4389	1	1	22.9667
EPHA8	-0.7291	1698	0.01796	0.401	1	1	0.8902
PTK9	-0.7328	1699	0.02794	0.4664	1	1	1.1773
MST1R	-0.7354	1700	0.01597	0.3921	1	1	3.0529
CDKL4	-0.7455	1701	0.001996	0.2144	1	1	5.9471
PKN3	-0.7518	1702	0.01198	0.3743	1	1	0.05265
DCI	-0.7657	1703	0.005988	0.2709	1	1	0.244
PPP1R14B	-0.7886	1704	0.01796	0.401	1	1	0.1403
MATK	-0.7935	1705	0.003992	0.2214	1	1	0.5629
PRKAG1	-0.8343	1706	0.007984	0.305	0.999	0.999	1.2176
MAD2L2	-0.84	1707	0.01397	0.3921	0.998	0.998	0.01933
GRM2	-0.8431	1708	0.003992	0.2214	0.998	0.998	0.7766
AK2	-0.8468	1709	0.007984	0.305	0.998	0.998	11.3008
NUDT3	-0.8632	1710	0.007984	0.305	0.998	0.998	23.1432
ACAD9	-0.8887	1711	0.003992	0.2214	0.993	0.993	0.2898
PPP2CB	-0.8928	1712	0.005988	0.2709	0.993	0.993	0.1924
PPM1A	-0.9294	1713	0.003992	0.2214	0.983	0.983	92.482
TNK2	-0.9412	1714	0.01198	0.3743	0.974	0.974	0.1576
ADSS	-1.0936	1715	0.001996	0.2144	0.755	0.753	0.05648
BHMT	-1.1485	1716	0.001996	0.2144	0.588	0.587	2.7801
PRKACG	-1.1995	1717	0.003992	0.2214	0.484	0.484	0.2845
SLC2A12	-1.3137	1718	0.001996	0.2144	0.283	0.283	0.5207
PRKACA	-1.3377	1719	0.001996	0.2144	0.253	0.253	0.5317

Figure 6. 3. Heatmap of non-isogenic siRNA screen hits after marker selection according to *KRAS* and *NRAS* mutational status (legend previous page)

	Kras mutated									Kras WT						Difference WT	T-test WT
Gene ID	A427	A549	H23	H292	H358	H727	H460	H1299	HOP62	H1838	BEN	H1650	H1975	H1793	H2342	minus mutant	vs mutant
CREB3L1	-5.8749	-6.1024	-5.7624	-4.0039	-0.8919	-2.0399	-4.9054	-7.9518	-2.7065	-0.5128	-2.7178	-1.8804	-1.6313	-2.1562	-0.7060	**3.1495**	0.01125131
GNPDA1	-2.2266	-5.6531	-4.1638	-3.5816	0.5068	-2.5434	-8.1442	-4.6792	-1.2448	0.5205	0.2257	-2.2098	-3.7240	-0.7029	-0.1647	**3.1478**	0.05265966
RRM2	-7.7476	-8.8967	-10.1422	-3.0757	-4.7436	-2.8948	-21.7739	-19.4414	-2.7761	-6.0470	-1.7623	-4.4612	-3.6386	-5.8062	-6.7088	**2.6139**	0.17373807
RPS6	-5.9447	-7.6550	-6.8538	-4.8191	-6.0992	-3.6948	-23.3545	-12.8855	-2.7275	-6.7023	-2.8546	-5.0811	-3.8977	-3.0898	-2.1216	**2.6054**	0.13680438
DYNC1H1	-3.6255	-5.4030	-4.0019	-2.5180	-1.1616	-1.5691	-8.5382	-6.4087	-1.8610	-1.0344	-2.6293	-1.2033	0.1060	-1.1816	-0.1545	**2.5175**	0.01861862
PPAT	-1.7287	-4.5760	-6.0113	-2.7752	-0.7711	-1.7065	-6.6831	-5.0861	0.1708	-0.9022	-1.4668	0.6546	-0.8130	0.0042	1.1974	**2.3708**	0.01362834
SOS1	-2.9874	-4.9431	-3.9061	-0.6958	0.2659	-0.5796	-7.1802	-7.7852	-0.7656	-0.7024	-0.3889	-3.8477	-2.3859	-0.7057	-0.7795	**2.2447**	0.21658673
MAPKAPK3	-2.4281	0.3545	-1.9288	-2.3885	-3.3604	-0.4951	-5.7400	0.2900	-0.0030	0.8926	0.1418	-0.4392	-0.3436	0.4037	0.6185	**2.2016**	0.03905298
DHFR	-0.1530	-4.4180	-1.0410	-2.8295	-2.2020	-0.5632	-7.8704	-2.6177	-0.5627	-2.6961	0.7380	0.0028	-0.0242	0.7496	-0.0072	**2.1997**	0.05865832
ACO2	-2.0030	-1.7449	-2.9808	0.5652	-1.5493	-1.0261	-2.3932	-1.7311	0.7416	0.6595	0.1425	-0.1424	0.5214	0.9164	-3.4993	**2.0630**	0.16040887
PGD	-1.9920	-1.8016	-3.2669	-4.2652	-3.7771	-1.1652	-10.2277	-3.1330	-1.9604	-3.0905	-1.1572	-4.7210	-0.1185	-0.9970	-0.6113	**2.0559**	0.1944181
RAD51	-3.0799	-2.9907	-3.6969	-5.0417	-2.9884	-4.7670	-4.7600	-2.5500	-5.4248	-4.3109	-1.2690	-2.0305	-1.2787	-0.8309	-2.8079	**2.0423**	0.01055329
NDUFAB1	-0.5448	-2.6578	-2.3734	-3.7176	-1.8947	-2.8693	-6.7170	-18.0945	-0.3830	0.2832	-0.9847	-1.9370	-0.4663	-0.7847	-0.1416	**2.0324**	0.12945434
PIGK	-4.2496	-4.0769	-4.8971	-1.1052	0.8282	-1.2290	-2.3696	-5.1561	-2.9101	-0.1480	-2.0056	0.0756	-3.7627	-1.4456	-0.3507	**2.0119**	0.13638095
PALB2	-2.3889	-1.2269	-2.7629	-6.4384	-2.6595	-3.7763	-3.1900	-2.8700	-3.4092	-1.3591	0.3645	-0.6848	-1.7222	0.3369	-1.1629	1.9462	0.00215409
NTHL1	-0.7829	-0.0210	-1.4150	-1.9622	-2.1389	-5.0088	0.3300	-2.4700	-2.5477	0.0953	1.2584	-0.1440	-1.7897	0.1336	-1.1875	1.9379	0.06478973
RPS6KA5	-1.6055	-0.0596	-1.4828	-1.6035	-0.1057	-0.5013	-2.5800	-1.6000	-1.5924	0.7029	-1.6822	0.8228	0.0050	0.6257	-0.3014	1.9078	0.01727191
PPAP2B	-1.6012	-4.7331	-5.1714	-2.2403	-1.3225	-4.1600	-18.1417	-11.3214	-3.1160	-2.1796	-1.8940	-1.8848	-6.2660	-2.4420	-2.5178	1.8491	0.2405054
FLJ21816	-2.4295	-1.1013	-2.4413	-3.5848	-2.5512	-2.4439	-3.3850	-2.9150	-3.8453	-1.8511	-0.8457	-0.3263	-1.7727	0.4553	-0.5711	1.8428	0.0008255
ALDRL6	-4.4742	-2.8584	-11.3639	-4.0576	-1.4336	-1.9091	-13.1995	-9.5626	-2.0490	-0.7334	-1.8307	-1.5499	-3.7913	-3.1927	-2.7304	1.7771	0.10018784
EIF4E2	-1.4754	-6.2957	-0.4350	-0.1592	-1.3115	-1.3400	-8.3367	0.0373	-0.1566	-0.2887	0.3850	0.1629	0.6253	0.7157	0.5007	1.7544	0.06632001
NT5M	-1.3768	-2.2348	-5.2630	-1.9599	-3.9113	-0.9954	-10.8273	-5.4287	-0.6141	-0.6284	-0.4503	-1.2135	-0.1742	-1.5112	-0.1803	1.6954	0.04891651
KRAS2	-3.0448	-1.2700	-2.6479	-3.3724	-3.2768	-1.7060	-8.0008	-2.1625	-3.0507	-0.5191	-1.7290	-6.2532	-1.6420	-1.0800	1.2555	1.6838	0.21151282
HUNK	-4.4782	-1.2313	-1.7044	-2.6295	-1.2269	-4.3352	-5.0400	-2.4800	-3.3740	-1.4778	-2.4751	-0.3951	-1.0952	-0.3641	-0.8221	1.6709	0.01414365
CTDP1	-1.4353	-0.4190	-2.3864	-1.6584	-2.1264	-0.7382	-3.0099	-0.5393	-3.4570	0.3841	0.2562	-0.4853	-0.2135	-0.4716	0.2376	1.6704	0.002919
KSR	-1.6122	-0.1924	-1.2763	-0.6702	-0.8538	-2.0382	-3.4600	-3.0100	-0.1794	0.0061	-4.5592	0.7397	-0.5281	1.0695	1.2131	1.6492	0.2097524
H2AFX	-3.7890	-0.1110	-4.5321	-2.6730	-1.5882	-2.8557	-1.8700	-2.1800	-2.4846	-3.0442	-0.3357	-0.9643	-2.0819	0.1854	-0.7180	1.6435	0.06989678
IRAK3	-5.3162	-0.0313	-2.0809	-3.0436	-1.9969	-3.6957	0.3900	-1.3200	-2.3116	0.0081	-0.5878	-3.0685	-0.3091	-1.2026	0.1918	1.6324	0.13289597
GNE	0.3532	-0.0429	-1.4169	-2.3696	-2.0241	0.2533	-2.7000	-1.3800	0.8353	0.3587	-0.2053	-2.4036	0.9812	0.1451	0.7189	1.6319	0.21576678
ERCC6	-3.2260	-0.1042	-2.8083	-3.3769	-2.1928	-1.1493	0.0200	-0.4900	-2.0846	-2.0609	0.6707	-0.8315	0.4068	-0.1939	-1.7982	1.5718	0.1268523
DKFZP761P0	-2.0968	0.0309	-1.7738	-2.4420	-1.2569	-0.6372	1.7700	-1.1500	0.1513	0.6674	0.2644	-0.0825	-1.8161	1.0444	0.5525	1.5584	0.1694145
SLC25A27	-3.6605	-0.9211	-2.6162	-3.3533	-1.4031	-1.4354	-7.1967	-2.2033	-2.8860	0.1569	-1.0453	-0.7047	-1.9107	-1.2680	-1.0856	1.5508	0.03674088
DGKQ	-2.9949	-2.9899	-1.2014	-2.6333	-1.9868	-4.5279	-3.5100	-3.4200	-5.6609	-0.8227	-0.5818	-2.7278	-1.9452	-6.6004	-0.9719	1.5363	0.32560674
PR48	-2.6645	-5.7280	-2.8052	-2.8055	-0.2418	-3.0927	-4.0902	-7.9699	-3.1430	-0.3076	-0.4643	-1.6986	-1.4214	-2.5874	-3.2434	1.5327	0.06132153
KIAA1361	0.0551	-0.7299	-1.1250	-0.6995	-1.1827	-0.6818	-9.0600	0.5000	-1.9076	-0.3570	0.0499	1.0622	0.7501	1.0021	0.8236	1.5168	0.08923273
HSGT1	-1.0782	-1.0079	-0.0618	-1.9036	-1.5980	-0.8016	-5.5191	-0.6962	-1.1551	-0.1498	0.6091	0.8724	0.4075	-0.8850	0.4434	1.5037	0.02414879

median Z-score difference > 2

median Z-score difference > 1.5 and < 2

Table 6. 4. Median Z-score differences between *KRAS*-mutant and *KRAS*-wild type groups

Hits are sorted by order of decreasing Z-score difference. A two-sided T-test was also calculated between both groups.

6.2.3.3. High throughput siRNA screen results

These two parallel approaches allowed the identification of 50 and 36 putative targets using GENE-E and the Z-score comparison methods, respectively. Fourteen genes, listed in Table 6. 5, were common to both datasets. Of note, MAPKAPK3 and EIF4E2, both downstream proteins of KRAS signalling, came up as hits using both methods of analysis. In order to be able to detect false positive results that were driven by outliers, the siRNA Z-score of the 14 shared hits were represented using waterfall plots (Figure 6. 4). This method provides a direct visualisation of the siRNA effect on each cell line, and allows straightforward detection of results driven by a very low value of a single cell line only in one group, or artefact positive values in the other group. Waterfall plots allowed the identification of HSGT1 as a potential false positive, as results were driven by one cell line (H460, which presented a very low Z-score) whereas the siRNA had no major effect on most of the other *KRAS*-mutant cell lines.

Finally, pathway enrichment was investigated through analysis of the 36 hits detected with the Z-score comparison method using GSEA (Broad Institute, see Chapter 3). Among eight pathways that showed significant enrichment, three were MAPKinase-related pathways, one mTOR-related pathway and one DNA repair pathway (Table 6. 6). This suggested that *KRAS*-mutant NSCLC cell lines included in the present panel displayed some degree of addiction on the MAPKinase pathway.

Genes shared in both datasets	Gene full name
CREB3L1	cAMP responsive element binding protein 3-like 1
CTDP1	CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) phosphatase, subunit 1
DYNC1H1	dynein, cytoplasmic 1, heavy chain 1
EIF4E2	eukaryotic translation initiation factor 4E family member 2
FLJ21816	
HSGT1	ecdysoneless homolog
HUNK	hormonally up-regulated Neu-associated kinase
MAPKAPK3	mitogen-activated protein kinase-activated protein kinase 3
NT5M	5',3'-nucleotidase, mitochondrial
PALB2	partner and localizer of BRCA2
PPAT	phosphoribosyl pyrophosphate amidotransferase
RAD51	RAD51 homolog
RPS6KA5	ribosomal protein S6 kinase polypeptide 5
SLC25A27	solute carrier family 25, member 27

Table 6. 5. Genes whose silencing showed significance in both methods of analysis

SiRNA screen hits identified using Gene-E were matched with siRNA screen hits identified when performing a difference in Z-scores between *KRAS* mutant and *KRAS* wild-type cells. List of hits that were identified in both datasets are shown here, together with the corresponding full gene name.

Figure 6. 4. Waterfall plots of Z-scores of siRNA showing a differential effect between the *KRAS*-WT and *KRAS* mutant populations in the non-isogenic model (figure next page)

Hits were selected on the basis of a >1.5 difference in median Z-score between *KRAS*-WT (green) and *KRAS*-mutant (purple) cells, and on a p-value < 0.05 on GENE-E analysis. Waterfall plots of Z-scores allow a straightforward visualization of the cell killing effect of each significant siRNA, as well as results driven by outliers. Of note, MAPKAPK3 and EIF4E2 scored as candidate hit using both methods, and could as such be considered as examples of positive controls of the screen

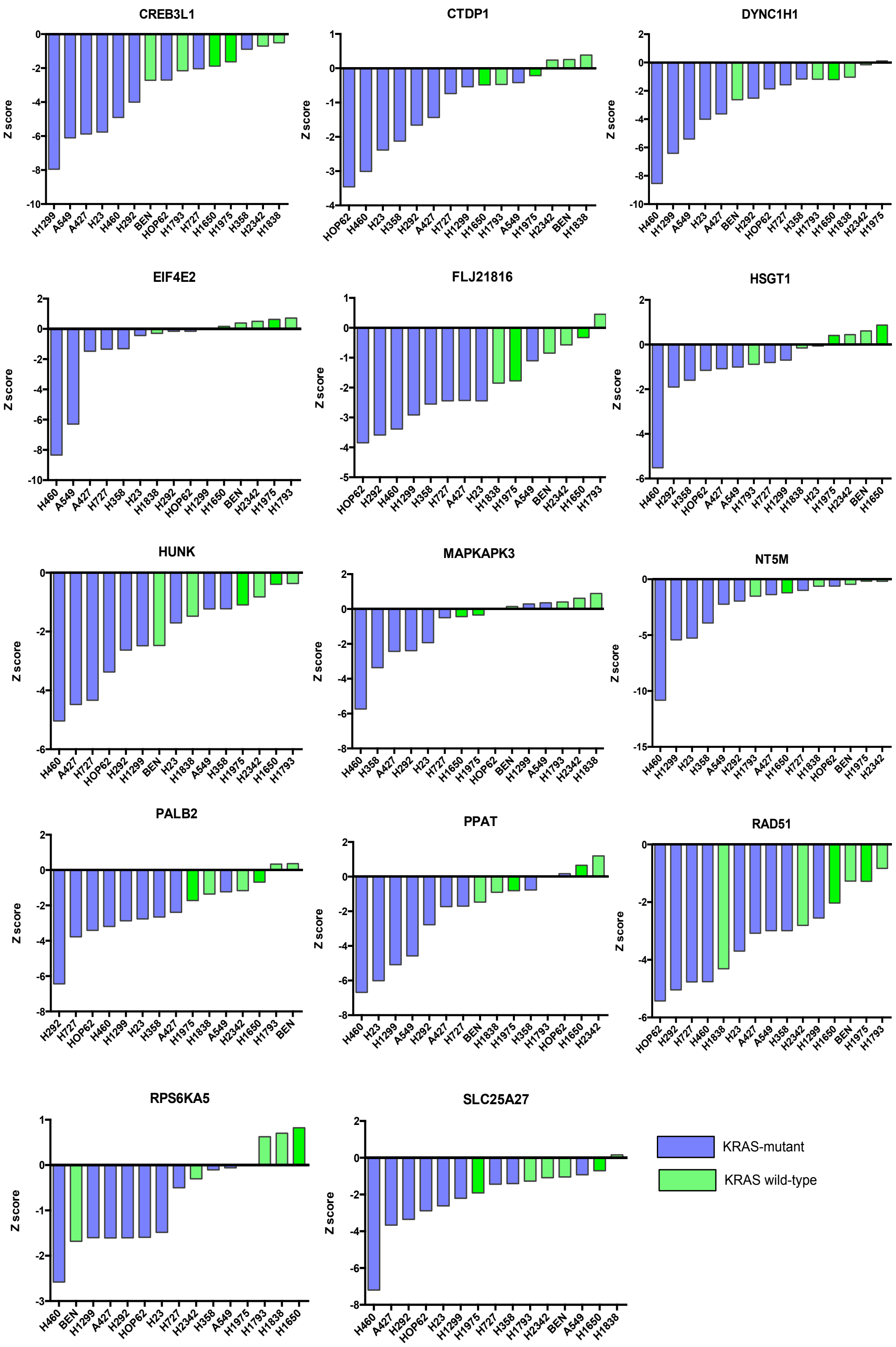


Figure 6. 4. Waterfall plots of Z-scores of siRNA showing a differential effect between the *KRAS*-WT and *KRAS* mutant populations in the non-isogenic model (legend previous page).

Gene Symbol	Description	REACTOME DNA repair	REACTOME Immune	BIOCARTA MAPK	KEGG mTOR	KEGG MAPK	KEGG Insulin	PID BARD1	PID P38
	# Genes in Gene Set (K)	112	933	87	52	267	137	29	31
	# Genes in Overlap (k)	6	10	5	4	6	5	3	3
	k/K	0.0536	0.0107	0.0575	0.0769	0.0225	0.0365	0.1034	0.0968
	p-value	1.20E-09	1.51E-08	2.22E-08	1.91E-07	2.13E-07	2.16E-07	2.94E-06	3.61E-06
	FDR q-value	1.58E-06	9.76E-06	9.76E-06	4.76E-05	4.76E-05	4.76E-05	5.54E-04	5.96E-04
XRCC5	X-ray repair complementing 5								
RAD51	RAD51 homolog								
GTF2H2	general transcription factor IIH, polypeptide 2								
GTF2H1	general transcription factor IIH, polypeptide 1								
PALB2	Partnet and localized of BRCA2								
MBD4	methyl-CpG binding domain protein 4								
PAK1	p21/Cdc42/Rac1-activated kinase 1								
MAP3K3	mitogen-activated protein kinase kinase kinase 3								
RPS6KA5	ribosomal protein S6 kinase, 90kDa, polypeptide 5								
MAPKAPK3	mitogen-activated protein kinase-activated protein kinase 3								
AKT3	v-akt murine thymoma viral oncogene homolog 3								
EIF4E	eukaryotic translation initiation factor 4E								
EIF4E2	eukaryotic translation initiation factor 4E member 2								
PRKG1	protein kinase, cGMP-dependent, type I								
TEC	tec protein tyrosine kinase								
DYNC1H1	dynein, cytoplasmic 1, heavy chain 1								
RPS6KB1	ribosomal protein S6 kinase, 70kDa, polypeptide 1								
FGFR2	fibroblast growth factor receptor 2								
PRKCZ	protein kinase C, zeta								
RBBP8	retinoblastoma binding protein 8								
CTDP1	CTD phosphatase, subunit 1								

DNA repair | MAPK | Insulin | Immune system | mTOR

Table 6. 6. Pathways enrichments of the siRNA screen hits of the non-isogenic NSCLC cell lines classified acording to *KRAS* mutation status.

GSEA overlap pathways (Gene Set Enrichment Analysis database, Broad Institute) of hits showing a >2 difference in median Z-scores between *KRAS*-mutant (including *NRAS*-mutant) and *KRAS*-WT populations. Significantly enriched pathways are depicted using an arbitrary colour code.

6.3. Discussion

Presented in this Chapter are the results of the high throughput drug and siRNA screen on a panel of 15 non-isogenic NSCLC cell lines. Exome sequencing was performed in parallel of the screens, in order to be able to integrate functional viability profiles and genetic dependencies with cell lines mutational profiles. This "satellite" project aimed at generating long-term developments of this PhD, and preliminary results of the pilot analysis performed on *KRAS* mutational status presented in this Chapter only aimed at illustrating such integrated approach.

KRAS mutations, detected in up to 25% of the patients presenting with NSCLC, are the most common molecular change in NSCLC. Although the presence of the mutation has been associated with poor prognosis, it has demonstrated little clinical utility so far (Roberts *et al.*, JCO 2013). Since high throughput screens represent a relatively unbiased way to discover genetic dependencies or novel therapeutic applications of drugs, *KRAS* was chosen to perform a pilot integrated analysis of siRNA screen and drug screen results in the panel of 15 NSCLC cell lines. The cell line classification used in this Chapter was pragmatic, since all different *KRAS* substitutions were aggregated together, and grouped with one *NRAS* mutation. Indeed, although all *KRAS* mutations listed above (Table 6. 1) are described as "activating" mutations and occur in the same 3-dimensional space the folded protein, they are distinct substitutions in amino acids (Span *et al.*, 1996; Al-Mulla *et al.*, 1999; Al-Mulla *et al.*, 2001). It has been suggested that different mutations could result in functionally different mutant KRAS proteins, notably resulting from different binding affinities for the various downstream effector molecules (Ihle *et al.*, 2012). For example, G12D-mutated KRAS preferentially activates the PI3K and MEK signalling pathways, whereas G12C-mutated KRAS rather activates RalGDS. Accordingly, a preliminary and mandatory requirement prior to any further investigation is the evaluation of the degree of KRAS addiction (and potentially of some of its downstream effectors) of the cell lines; *KRAS* siRNA was not present in any of the siRNA libraries. Similarly, all drug screen and siRNA screen hits discussed below require revalidation in a low-throughput format and using deconvoluted siRNAs and alternative cellular models, prior to any further investigation.

Results of the drug screen analysis showed a selective effect of one MEK and one RAF inhibitors, which are part of the RAS signalling pathway and could as such be considered as positive controls of the model. Surprisingly, four distinct HER1/2 inhibitors also scored as hits. *KRAS* mutant status has been consistently associated with resistance to anti-EGFR monoclonal antibodies in colo-rectal cancer, and the benefit of these agents is reported to be restricted to patients with *KRAS* wild-type tumours (Allegra *et al.*, 2009). In analysis performed in NSCLC patients treated with cetuximab, there was no indication that *KRAS* mutational status was predictive of response (Khambata-Ford *et al.*, 2010; O'Byrne *et al.*, 2011). Data are more controversial regarding *KRAS* mutation as a predictive marker of EGFR TKI benefit: two meta-analyses have reported an association between the presence of a *KRAS* mutation and absence of response, but no correlation could be found regarding a potential interaction between *KRAS* mutation status and PFS or OS (Mao *et al.*, 2010; Linardou *et al.*, 2011). Regarding results obtained with the panel of cell lines, it was hypothesised that the presence of sensitising or resistant *EGFR* mutations may have introduced biases in the screen results. Among the 15 cell lines, only two cell lines of the *KRAS* wild-type group (H1650 and H1975), harboured *EGFR* mutations (L858R sensitizing mutation, and T790M resistant mutation, respectively). It was therefore unlikely that *EGFR* mutational status would have acted as a confounding factor in this series. It can however not be excluded that another mutational imbalance between the two groups of cell lines could have driven these results, such as *LKB1* that was mutated in 4/8 cell lines of the KRAS mutant group and none of the KRAS wild-type group. *In vivo* testing has shown that *KRAS*-mutant / LKB1-deficient NSCLC tumors are resistant to MEK inhibition, whereas *KRAS*-mutant / *P53*-mutant NSCLC are not (Carretero *et al.*, 2010; Chen *et al.*, 2012). Therefore, future analysis may require further stratification of *KRAS*-mutant NSCLCs based on other recurrently mutated genes, such as *LKB1*. Of note, one IKKε (inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase epsilon inhibitor) also demonstrated differential effects between *KRAS*-mutant and *KRAS*-wild type cells; this was in line with the recently reported results of a siRNA screen performed by Barbie *et al* (Barbie *et al.*, 2009) which identified TBK1 as selectively essential in cells that contain mutant *KRAS*, and that more generally NF-κB signaling was essential in *KRAS* mutant tumours.

SiRNA screens performed in the non-isogenic model identified a number of hits as synthetically lethal with *KRAS*-mutant status, three of which were downstream effectors of the RAS pathway (MAPKAPK3, EIF4E2 and RPS6KA5) (Karachaliou *et al.*, 2013; Roberts *et al.*, 2013). More unexpectedly, two key players of the homologous recombination pathway, namely PALB2 and RAD51, showed selective effects on *KRAS*-mutant cell lines. Whether *KRAS* mutant NSCLC cells display a high level of replicative stress, that would support an enhanced sensitivity to HR inhibition, warrants further investigation. The selective effects on *KRAS* mutant cells observed with a CHK1 inhibitor on the drug screen might also support this hypothesis.

To conclude, this Chapter provides an illustration of the potential of high throughput screens performed in non-isogenic NSCLC cell lines models. Although more cell lines would ideally be required, results obtained on a panel of 15 NSCLC may still allow hypothesis to be generated that could be subsequently revalidated and investigated in low-throughput formats. Multiple other outputs could be derived from this dataset, including functional profiling through hierarchical clustering of the cell lines based on the screen results (an approach that has been successfully used previously in the Gene Function Laboratory on a panel of 34 breast cancer cell lines; Brough *et al.,* 2011), or integration of NSCLC data to other data from other tumour types (in order to identify shared determinants of specific tumour subtypes and genetic dependencies across multiple histologies). Finally, examining siRNA screen data from a subgroup of cell lines displaying an unusual sensitivity to a drug may also help identifying mechanisms underlying this enhanced sensitivity.

CHAPTER 7

Final discussion and future directions

The development of targeted therapies, catalysed by an increasing knowledge of cancer biology and tumour mutational profiles, has been transforming the treatment of cancer in the last decade (Reck *et al.*, 2013; Rosell *et al.*, 2013). The "one size fits all" approach is not longer valid and is being gradually replaced by "personalized medicine", which aims at customizing the treatment according to the tumour molecular profile. The majority of targeted therapies have initially exploited the phenomenon of oncogene addiction (Weinstein, 2002), exemplified first by the targeting of BCR-ABL fusion protein in chronic myeloid leukaemia by the tyrosine kinase inhibitor imatinib. In non-small cell lung cancer, therapeutic successes include targeting *EGFR*-mutated and *ALK*-translocated tumours by EGFR (Lynch *et al.*, 2004; Paez *et al.*, 2004; Kim *et al.*, 2008; Mok *et al.*, 2009; Rosell *et al.*, 2012) and ALK tyrosine kinase inhibitors, respectively (Kwak *et al.*, 2010; Shaw *et al.*, 2011a; Shaw *et al.*, 2011b). Unfortunately, resistance inevitably arises (Ohashi *et al.*, 2013) and complementary or alternative approaches need to be developed. A promising approach is synthetic lethality, as illustrated by the successful development of PARP inhibitors in BRCA-deficient tumours (Fong *et al.*, 2006; Farmer *et al.*, 2005; Bryant *et al.*, 2005). Synthetic lethality exploits the relationship that exists between two genes, which are individually non-essential to cell survival, but whose concomitant abrogation causes cell death (Farmer *et al.,* 2005; Rehman *et al.,* 2011). Therefore, targeting a gene product whose synthetic lethal partner is selectively deficient in tumour cells (i.e. proficient in normal cells) allows the creation of a therapeutic window between normal and cancer cells, thereby selectively killing tumour cells and the limitation of systemic toxicity of a therapy. ERCC1-deficiency has been described in approximately 50% of NSCLC (Postel-Vinay *et al.*, 2012), and as such represents a potential target for exploiting synthetic lethality in NSCLC.

The first part of this thesis focused on identifying a synthetic lethal relationship with ERCC1 deficiency in NSCLC models, using high-throughput drug and siRNA screens. PARP1 catalytic inhibition was identified as selectively toxic for ERCC1-deficient cells, and dissection of the mechanism underlying this selective sensitivity was performed. The working model proposed in Chapter 4 suggests the creation of a lesion by PARP1 itself trapped on the DNA, the repair of which is dependent on

ERCC1 be involved. Although several observations detailed in Chapter 4 supported this model, no evidence was found for a direct increase of DNA-bound PARP1 after PARP inhibitor exposure could not be evidenced. Several cell fractionation experiments with evaluation by western blot of PARP1 quantity in the chromatin fraction were performed, but no significant increase of chromatin-bound PARP1 could be detected in the ERCC1-deficient clones (data not shown). Previous similar experiments have required the addition of the alkylating agent methyl methanesulfonate (MMS) to demonstrate an increase in DNA-bound PARP1 (Murai *et al.*, 2012). To further investigate this hypothesis, chromatin immunoprecipitation experiments could be performed on ERCC1-isogenic models following olaparib and niraparib treatment, in particular: (i) immunoprecipitating γH2AX and blotting PARP1; this would examine whether persisting γH2AX foci observed in ERCC1-deficient clones are PARP1-related; and (ii) immunoprecipitating PARP1 and subsequently quantifying PARP1-bound DNA (Kedar *et al.*, 2012). This experiment would require one selection of DNA sequence for amplification that would be representative of bulk DNA. Previous similar experiments have used GC-rich isochores (Kedar *et al.*, 2012) or DNA polβ, but repetitive elements may also be used for this purpose. Alternative techniques to bypass this technical hurdle would include performing an overall sequence-unspecific DNA quantification, using dyes, fluorescence or radio-labelling of the DNA (Schoppee Bortz *et al.*, 2011). Another hypothesis that might explain the ERCC1 deficiency-PARP1 inhibition synthetic lethal relationship would be the presence of a “pure” synthetic lethal relationship, i.e. without the creation of a DNA-PARP1 lesion. In this latter case, the ERCC1-selective toxicity of PARP1 inhibitors would result from a similar mechanism as the BRCA-PARP synthetic lethal relationship, i.e. the persistence of SSB that would, when PARP is inhibited, turn into DSB that would not be repaired in ERCC1-deficient cells. The epistasis observed between some HR genes and ERCC1 in presence of a PARP inhibitor is compatible with this hypothesis. ERCC1 in DSB repair has mainly been implicated in single strand annealing, non-homologous-end joining, and micro-homology-directed end joining (Niedernhofer *et al.*, 2004; Zhang *et al.*, 2007; Al-Minawi *et al.*, 2009; Bhagwat *et al.*, 2009a). A putative role for ERCC1 in the resolution of double Holliday junctions has also been proposed, which could fit with the present hypothesis (Al-Minawi *et al.*, 2008). This could be examined through recombination assays using plasmids containing specific substrates for each pathway, which would allow the evaluation of the functionality of each of the above-mentioned DSB repair pathways. The putative involvement of ERCC1 in the resolution of Holliday junctions may also be investigated through specific Holliday junction visualisation by electron microscopy in the presence and absence of the PARP inhibitor. Finally, the participation of a metabolic component to the PARP inhibitors ERCC1-selective effects should be considered. The SILAC results presented in Chapter 5 demonstrated that NAMPT levels were decreased in ERCC1-deficeint clones as compared to their ERCC1-proficient counterpart, and that ERCC1-deficient models were 10 times more sensitive to the NAMPT inhibitor FK866 (Figure 5.32). High NAMPT expression level has been correlated with resistance to PARP inhibitors, due to the increased cellular capacity to recycle βNAD+ from nicotinic acid through the NAD salvage pathway (Bajrami *et al.*, 2012). Whether this differential expression of NAMPT may have contributed to the PARP1 selective effects in ERCC1-deficient cells was evaluated, by adding nicotinic acid into the media of ERCC1-deficient models exposed to olaparib (data not shown). As nicotinic acid did not allow rescuing the ERCC1-selective effects of olaparib, the contribution of NAMPT to olaparib sensitivity, if any, is probably minimal. Overall, these last two hypotheses seemed less likely than the creation of a lesion by PARP1 itself, considering the “rescue” of PARP inhibitor effect and the restoration of PARP inhibitor resistance following PARP1 silencing. Interestingly, a recent publication from Michels et al (Michels *et al.*, 2013) reported that a model of CDDP-resistant NSCLC cells displayed elevated intracellular levels of poly(ADP-ribosyl)ated (PAR) proteins, which was predictive of higher sensitivity to PARP1 inhibition, both chemically and by siRNA. Although these results contrast with most preclinical and clinical observations that correlate platinum- sensitising to PARP inhibitor-sensitising DNA repair defects (Turner *et al.*, 2008; Rehman *et al.*, 2010; Johnson *et al.*, 2011; Ledermann *et al.*, 2012; Murai *et al.*, 2012), PARylation levels have not been evaluated in ERCC1-deficient cells; this might merit further investigation to better understand the link between ERCC1, platinum sensitivity and PARP inhibitor sensitivity.

Based on preclinical results produced in this thesis, a phase II clinical protocol was designed in order to evaluate olaparib versus placebo as maintenance therapy in platinum-sensitive patients with advanced NSCLC. The clinical study proposal was subsequently approved by AstraZeneca and the protocol, which is currently being written, will be supported by a comprehensive translational work aiming at understanding the determinants of PARP inhibitor sensitivity in NSCLC. As the proposal protocol was being set up, it was noted that a similar study was about to start from another collaborative group, based on observations performed in BRCA1-deficient NSCLC (Paul *et al.*, 2011). Although an attempt to merge the clinical part of both studies unfortunately failed, samples of both trials may hopefully still be exploited as a part of the translational studies planned within our protocol. Several questions had to be asked prior to designing the protocol. First, which patients with advanced NSCLC should be eligible for such trial? Patients presenting with *EGFR*-mutated or *ALK*-translocated tumours currently benefit from effective targeted therapies that are licenced as first line therapy, and offer better activity and tolerability profiles than platinum-based therapy (e.g. erlotinib and crizotinib) (Rosell *et al.*, 2012). Further, the biology of oncogene-driven diseases is different from NSCLC arising from DNA repair defects and accumulation of tobacco-induced damage. It was therefore deemed that patients presenting with *EGFR*-mutated or *ALK*-translocated tumours should not be offered this trial. Second was asked the question of the optimal drug schedule administration: should PARP inhibitors be combined with cisplatin, or should both drugs be administered sequentially, i.e. platinum-based doublets as per institutional recommendations first, followed by PARP inhibitor maintenance? Considering the excellent tolerability profile of PARP inhibitors as monotherapy and the fact that none of the PARP inhibitor – chemotherapy combinations that have been evaluated so far in early phase trial has proven to be readily tolerable, it seems that the maintenance setting was the most appropriate one (Rajan *et al.*, 2012; Samol *et al.*, 2012; Dent *et al.*, 2013). Third arose the question of how should patients that would be eligible for randomisation be selected: should patient selection be based on ERCC1 status (i.e. ERCC1-low *vs* ERCC1-high patients), or on a surrogate biomarker of ERCC1 activity? As described in Chapter 5, there is currently no consensus on the optimal method and threshold of expression that should be used to classify a tumour as ERCC1-proficient or ERCC1-deficient; moreover, the existence of several isoforms that cannot be discriminated by current techniques, and of which only one is functional, can lead to false positives and misclassification. Efforts to develop a reliable assay evaluating ERCC1 isoform 202 (e.g. ERCC1/XPF heterodimer evaluation by PLA®) or a surrogate biomarker of ERCC1 activity are currently being established, but results are still too preliminary and insufficiently validated to be used in the clinical setting. Finally, some DNA damage response deficiencies underlying platinum-sensitivity have also been associated with PARP inhibitor sensitivity (Norquist *et al.*, 2011); patients presenting with tumours harbouring such ERCC1-independent DNA repair defects may equally benefit from PARP inhibitor therapy, and should therefore not be excluded from randomisation based on the assessment of ERCC1 status only. It was therefore considered that platinum-sensitivity would be a better surrogate selection biomarker than a molecular biomarker, and that all platinum-sensitive patients should be eligible for randomisation. Finally, whether patients whose disease was stable under platinum therapy should be eligible for randomisation, was considered. In the era of personalized medicine and customized therapies, patient selection is absolutely crucial: inappropriate definition of the right patient population to target can lead to trials failure, and subsequent halting of drug development although a population of patients would still derive benefit from the drug. Retrospective biomarker or subgroup analysis offer an alternative way of examining which population would respond to the drug, but such analysis is always questionable if not initially planned and therefore always requires subsequent prospective validation. It was therefore deemed that the population of interest should be restricted to patients with platinum-sensitive disease, and that patients presenting stable disease after 4-6 cycles of chemotherapy should be treated according to the best institutional procedures. Interestingly, a large phase III trial using a similar design was recently performed in patients suffering from serous ovarian carcinoma. A significant progression-free survival benefit was observed in patients receiving olaparib (8.4 vs 4.8 months; HR = 0.35 (0.25-0.49); $p<0.001$), which was mainly, but not only, driven by the BRCA1/2 status of the patients (Ledermann *et al.*, 2012). This highlights the potential for using PARP inhibitors in platinum-sensitive patients.

As discussed above, the functional relationships between platinum-sensitivity, PARP inhibitor-sensitivity and ERCC1 activity are still not fully elucidated and further *in vitro* experiments are required to further investigate this question. Most importantly, although the clinical protocol has been designed based on a “bench-to-bedside” approach, the reciprocal work should now take place, thanks to comprehensive translational studies that will be performed in parallel to the clinical trial (Figure 7. 1). These translational studies will have 5 main objectives:

1. Identifying predictive biomarkers of olaparib sensitivity in NSCLC
2. Identifying predictive biomarkers of primary and acquired resistance to olaparib in NSCLC
3. Investigating the threshold and method of assessment of ERCC1-deficiency that correlates with platinum sensitivity and olaparib sensitivity, respectively
4. Establishing a novel predictive molecular and functional classification of NSCLC patients according to their DNA damage response profile
5. Studying the clonality, correlation genotype-phenotype and changes over time of advanced / metastatic NSCLC tumours

To achieve these aims, tumour material (mandatory at inclusion only), blood samples, eyebrow hair follicles and skin biopsies will be regularly collected over the course of the trial. The following studies will be performed on tumour material: multiplex ultra-deep sequencing and methylation profiling of genes of interest, i.e (i) genes involved in HR or DSB-repair (including *BRCA1, BRCA2, ERCC1, PTEN, CHEK1, ATM, ATR*, the *FANC*-family of genes, *PARP1* and *PARP2)*; (ii) genes implicated as drivers in NSCLC (including *SLC34A2/* or *CD74/ROS* (translocation), *KRAS*, *NRAS*, *HRAS*, *BRAF, PI3K*, *AKT1*, *P53*, *PTEN*, *FGFR1 (*amplification*)*, *PDGFRA*, *DDR2*, *ERCC1*, *LKB1*, *CTNNB1*, *RET*, *HER2*, *MET* (amplification), *RB*, *NF1*, *MGMT*, *SMARCA4*, *CDKN2A*); genes for which treatments have been approved or are being evaluated in early phase trials (including *FGFR2, FGFR3, HER2 amp, HER3, HGF, IGF1-R, BH3*). Pharmacogenetic profiling and pharmacodynamic evaluation of PARP1 inhibition will be performed on peripheral blood mononuclear cells (PBMNCs). When possible, RAD51 and γH2AX foci staining, as well as targeted sequencing of genes of interest will also be performed on circulating tumour cells. Finally, RAD51 and γH2AX foci staining will be performed on eyebrow hair follicles and skin biopsies (Fong *et al.*, 2009; Graeser *et al.*, 2010).

In the long-term and in parallel to the trial, a comprehensive programme of laboratory research aiming at understanding response to PARP inhibitors and DNA-damaging agents in NSCLC will be carried out, including:

- Generation and profiling of olaparib resistant clones
- Identification of additive or synergistic combinations with olaparib
- Generation of mouse avatars carrying NSCLC xenografts to be treated with olaparib and molecularly characterized to identify determinants of sensitivity / resistance to olaparib (to be carried out at the Institute of Cancer Research Tumour Profiling Unit)
- Study of genomic DNA from patients to evaluate the presence of a genomic scar for DNA damage signature
- Deep whole genome sequencing and CGH (comparative genomic hybridization) analysis of “exceptional responders” (i.e. patients presenting an unexpected complete response or prolonged partial response on treatment).

Taken together, these studies should provide important data regarding the determinants of platinum and PARP inhibitor sensitivity in NSCLC, as well as provide further insight into surrogate biomarkers of ERCC1 activity to allow appropriate patient selection and personalised medicine.

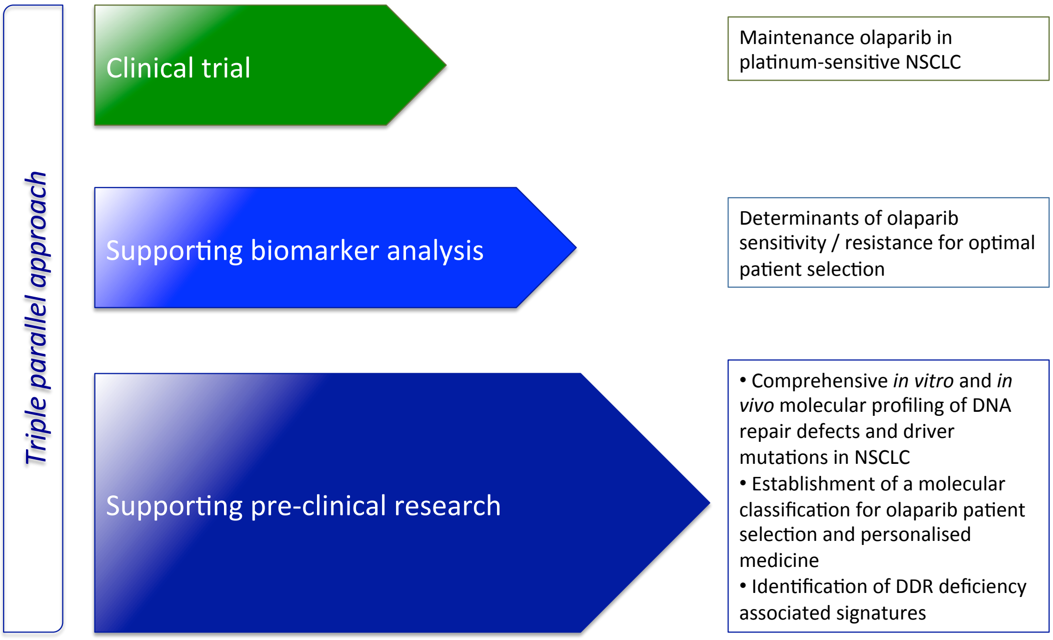


Figure 7. 1. Supporting translational studies

Several translational studies will take place in parallel to the trial, in order to examine the determinants of olaparib sensitivity and allow a comprehensive study of DNA repair in NSCLC. Figure 7.1 summarizes the parallel approaches that will be undertaken, with short-term, mid-term and long-term investigations.

A significant issue in translating *in vitro* observations into clinical practice is the identification of the appropriate patient population. The search for surrogate biomarkers of ERCC1 activity was consequently the focus of the second part of this PhD. Four parallel approaches were examined, which provided some promising results, as discussed in Chapter 5. Of note, since no reliable method or assay evaluating ERCC1 functionality was available, results found in the isogenic model were difficult to translate into the non-isogenic model. This could either be explained by the inappropriate classification of the non-isogenic cell lines, the presence of confounding factors (such as other unknown DNA repair defects), or that very low ERCC1 levels are required to give the ERCC1-deficient phenotype. This difficulty in comparability was a major limitation of the work, and the creation of alternative models is certainly required to better study ERCC1 deficiency. These would include other isogenic models derived from different NSCLC cell lines, using alternative gene targeting techniques (such as CRISPR or TALEN targeting) (Gaj, Trends Biotechn 2013) or inducible models in which *ERCC1* gene expression could be regulated so as to allow the study of the threshold of deficiency that correlates with the appearance of the deficient phenotype.

Genomic signatures are probably one of the most promising biomarkers currently being developed as DNA is a stable and reliable material to work with, publicly available data are accumulating exponentially, and the knowledge in analysing and interpreting high-throughput sequencing data is now well matured. Moreover, high throughput sequencing machines currently allow the rapid generation of big genome scale data at a relatively low cost. Most genomic signatures described so far arise from model systems exposed to known mutagens or perturbations of the DNA repair machinery, which have been compared to mutational profiles found in human tumours. Consequently, genomic signatures have so far mostly been associated with mutational processes underlying carcinogenesis and have not been used as predictive biomarkers (Pfeifer *et al.*, 2003; Pfeifer *et al.*, 2005; Stratton *et al.*, 2009; Pfeifer, 2010; Nik-Zainal *et al.*, 2012a; Alexandrov *et al.*, 2013; Lawrence *et al.*, 2013; Tomasetti *et al.*, 2013). Where intrinsic DNA maintenance machinery deficiencies are associated with treatment sensitivity, such genomic signatures could be used as surrogate predictive biomarkers of response. Results obtained with the ERCC1-isogenic model showed an excess of A>T transversions located in T-rich regions, at the intron-exon boundaries. However, several additional analysis and experiments need to be performed in order to try to confirm the validity of the signature. First, the specific coverage of the regions of interest and the presence of a transcriptional strand bias should be examined. Second, a customised exome sequencing, focused on the regions of interest, potentially followed by whole genome sequencing, should be performed. Third, the *HPRT* coding sequences mutational profile on ERCC1-deficient clones after 6-thioguanine selection (Tomita-Mitchell *et al.*, 2000; Chiu *et al.*, 2006) should be investigated. Moreover, as the ultimate outcome is to predict sensitivity to platinum-based therapy, a parallel “reciprocal” approach, in which mutational profiles will be stratified according to response to chemotherapy, is also required. Importantly, focusing on platinum sensitivity would allow several interacting mechanisms that underlie platinum response to be taken into account, rather than limiting the analysis to ERCC1 as a single biomarker. In this “reciprocal” approach, overall survival of NSCLC patients treated with adjuvant platinum-based chemotherapy would be taken as the ultimate measure outcome, and tumour mutational profiles would be stratified according to patient’s OS. The major limitation of such approach is currently the absence of genomic DNA for most patients, but patient series will be created and publicly available data that could be used in this purpose should be released soon.

The second approach used to investigate the presence of a surrogate biomarker of ERCC1 activity was the evaluation of NER functionality through the repair of UV-induced cyclobutane pyrimidine dimers (CPDs). An excess of CPDs could be detected by immunohistochemistry in ERCC1-deficient cells as compared to their ERCC1-proficient counterparts, and the feasibility of such approach in the clinical setting on fine needle aspirations (FNAs) is currently being evaluated. There are nevertheless several limitations to the use of UV-induced CPDs as a surrogate biomarker of ERCC1 activity. First, live tumour cells are required and therefore the tumour (or an associated lymph node) needs to be accessible to FNA. Second, although FNA can be performed during the clinic appointment of an outpatient, it is an invasive procedure that needs to be done by an experienced physician. Third, cells have to be irradiated immediately prior to incubation and time considerations are therefore important to ensure the feasibility of the procedure. Fourth, pathologists would need to be trained to provide a reliable interpretation of the results (such as an H-score that would be built based on the fluorescence intensity). Finally, it should be kept in mind that the ultimate objective of evaluating ERCC1 functionality is to predict platinum sensitivity. Platinum salts cause both intrastrand crosslinks (approximately 80-90% of the lesions, repaired through NER), and interstrand crosslinks (approximately 10-20% of the lesions, repaired through ICL-R) (Knox, CCR 1986).

Although little data is available on which lesion underlies platinum cytotoxicity, there is a general consensus that interstrand crosslinks, even if in minority compared to intrastrand crosslinks, are responsible of most of the toxicity of platinum salts. UV irradiation almost exclusively results in lesions that can be assimilated to intrastrand crosslinks (CPDs, 4-6PPs, Dewar isomers), and only occasionally causes interstrand crosslinks in particular DNA contexts (Love *et al.*, 1986; Rastogi *et al.*, 2010). It is therefore uncertain whether an isolated evaluation of NER could be used as a surrogate biomarker of platinum sensitivity.

The identification of Guanine deaminase (GDA) and Nicotinamide phosphoribosyl transferase (NAMPT) by SILAC experiment demonstrated that proteomic analysis could identify novel cellular changes unexpectedly associated with DNA repair alterations. Transcriptomic analysis could unfortunately not be fully exploited during this PhD – in the interest of time - but preliminary results supported the hypothesis that ERCC1 functions were not limited to DNA repair. It has been known for years that the hallmarks of cancer are not independent, but indeed interacting together to promote cell growth and survival (Hanahan *et al.*, 2000; Hanahan *et al.*, 2011). It is therefore conceivable that changes in DNA maintenance capacity might result in modifications in other hallmarks of cancer, such as metabolic pathways, to ensure cell survival. No previous interaction between ERCC1-deficiency and changes in nucleotide metabolism has been described so far, although it has been consistently reported that tumours that display low ERCC1 levels also display low RRM1 levels (Rosell *et al.*, 2004; Rosell *et al.*, 2007; Zheng *et al.*, NEJM 2007). Guanine Deaminase (GDA) catalyses the irreversible deamination of guanine to xanthine (Yuan *et al.*, 1999) and is involved in the *de novo* synthesis of purines. This enzyme, which is highly expressed in liver, kidney, brain and placenta (Firestein *et al.*, 1999; Kubo *et al.*, 2006), is also involved in the regulation of dendrite development by modulating guanine concentrations (Firestein *et al.*, 1999). Contrary to cytidine deaminase, which is involved in the pyrimidine metabolism and response to cytosine arabinoside used notably in the treatment of childhood leukemias, little is currently known about GDA and cancer. The mechanism explaining why ERCC1 deficiency would have an effect on GDA has not been elucidated. As discussed in Chapter 5, a role for GDA into decreasing the pool of guanine available, in order to limit the formation of ROS-induced lesions on guanine that would be toxic in the absence of ERCC1, has been hypothesized. The evaluation of ROS metabolism in ERCC1-deficient cells would allow the investigation of this possibility. A synthetic viability relationship between ERCC1-deficiency and one of the downstream metabolites of

GDA might be an alternative explanation; the metabolic profiling that will be performed on ERCC1-deficient clones should allow examination of this hypothesis. Interestingly, a recent interaction profiling of FDA-approved nucleoside and nucleobase analogs identified several compounds that induced a concentration-dependent stabilisation of GDA, including the antiretroviral molecules valaciclovir, acyclovir, penciclovir, ganciclovir, as well as mercaptopurine, and thioguanine (Egeblad *et al.*, 2012). Silencing GDA by siRNA had no viability effect on ERCC1-deficient cells (Figure 5. 25), but silencing of the protein was not verified and results are as such incomplete. More interestingly, obtaining colonies from ERCC1-deficient cells under 6-thioguanine selective pressure (for the *HPRT* mutagenesis experiment, Chapter 5) was much more difficult than obtaining colonies from ERCC1-proficient cells. Whether this is related to GDA has not been examined, but assessing the effects of the above-cited antiretroviral drugs and thioguanine on ERCC1-deficient cells viability would be of interest. The mechanism of the observed relationship between FK866 sensitivity, decreased expression of NAMPT and ERCC1 deficiency also could not be elucidated. One concern was the absence of rescue by any of the distinct ERCC1 isoforms. As discussed in Chapter 5, several factors (requirement of two isoforms, higher expression level, expression in a specific cellular compartment etc.) could explain this absence of rescue, but revalidation in an alternative isogenic or non-isogenic model is required prior to further investigation. Finally, the intersection of transcriptomic and SILAC data, as well as the integration with other expression datasets (Friboulet *et al.*, 2011) and siRNA screen results, should be further explored. Overall, the above results highlight the potential of integrated high-throughput approaches in identifying novel therapeutic targets or biomarkers. The SILAC and transcriptomic experiment results are still preliminary, and alternative approaches, such as RIME (rapid immunoprecipitation mass spectrometry of endogenous proteins) (Mohammed *et al.*, 2013), are currently being tested in order to identify ERCC1 partners that are involved in response to platinum. Once combined together, the results of the above experiments should allow the identification of several biomarkers that could be combined to generate a signature of ERCC1 functionality, similar to Oncotype DX (Paik *et al.*, 2006; Harris *et al.*, 2007).

To conclude, it is clear that considerable research is required not only to define novel therapeutic targets for subgroups of cancers, such as ERCC1-deficient NSCLC, but also to build companion diagnostic tests that allow the identification of the appropriately selected patient population. Parallel approaches are required to successfully reach these goals, which are the bench-to-bedside and bedside-to-bench paths. It will be equally important to assess the response to PARP inhibitors of patients with platinum-sensitive NSCLC patients, and to concomitantly best exploit patient's material collected during the trial in order to ensure a meaningful drug development for future patients' benefit.

REFERENCES

Ahmad A, Robinson AR, Duensing A, van Drunen E, Beverloo HB, Weisberg DB*, et al.* (2008). ERCC1-XPF endonuclease facilitates DNA double-strand break repair. *Molecular and cellular biology* **28**(16)**:** 5082-5092.

Akamatsu Y, Jasin M (2010). Role for the mammalian Swi5-Sfr1 complex in DNA strand break repair through homologous recombination. *PLoS genetics* **6**(10)**:** e1001160.

Al-Minawi AZ, Lee YF, Hakansson D, Johansson F, Lundin C, Saleh-Gohari N*, 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 research* **37**(19)**:** 6400-6413.

Al-Minawi AZ, Saleh-Gohari N, Helleday T (2008). The ERCC1/XPF endonuclease is required for efficient single-strand annealing and gene conversion in mammalian cells. *Nucleic acids research* **36**(1)**:** 1-9.

Al-Mulla F, MacKenzie EM (2001). Differences in in vitro invasive capacity induced by differences in Ki-Ras protein mutations. *The Journal of pathology* **195**(5)**:** 549-556.

Al-Mulla F, Milner-White EJ, Going JJ, Birnie GD (1999). Structural differences between valine-12 and aspartate-12 Ras proteins may modify carcinoma aggression. *The Journal of pathology* **187**(4)**:** 433-438.

Albert JM, Cao C, Kim KW, Willey CD, Geng L, Xiao D*, et al.* (2007). Inhibition of poly(ADP-ribose) polymerase enhances cell death and improves tumor growth delay in irradiated lung cancer models. *Clin Cancer Res* **13**(10)**:** 3033-3042.

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV*, et al.* (2013). Signatures of mutational processes in human cancer. *Nature* **500**(7463)**:** 415-421.

Allegra CJ, Jessup JM, Somerfield MR, Hamilton SR, Hammond EH, Hayes DF*, et al.* (2009). American Society of Clinical Oncology provisional clinical opinion: testing for KRAS gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy. *J Clin Oncol* **27**(12)**:** 2091-2096.

Altaha R, Liang X, Yu JJ, Reed E (2004). Excision repair cross complementing-group 1: gene expression and platinum resistance. *International journal of molecular medicine* **14**(6)**:** 959-970.

Ame JC, Spenlehauer C, de Murcia G (2004). The PARP superfamily. *BioEssays : news and reviews in molecular, cellular and developmental biology* **26**(8)**:** 882-893.

Arora S, Kothandapani A, Tillison K, Kalman-Maltese V, Patrick SM (2010). Downregulation of XPF-ERCC1 enhances cisplatin efficacy in cancer cells. *DNA repair* **9**(7)**:** 745-753.

Arriagada R, Bergman B, Dunant A, Le Chevalier T, Pignon JP, Vansteenkiste J*, et al.* (2004). Cisplatin-based adjuvant chemotherapy in patients with completely resected non-small-cell lung cancer. *The New England journal of medicine* **350**(4)**:** 351-360.

Ashworth A (2008). Drug resistance caused by reversion mutation. *Cancer research* **68**(24)**:** 10021-10023.

Auckley DH, Crowell RE, Heaphy ER, Stidley CA, Lechner JF, Gilliland FD*, et al.* (2001). Reduced DNA-dependent protein kinase activity is associated with lung cancer. *Carcinogenesis* **22**(5)**:** 723-727.

Audeh MW, Carmichael J, Penson RT, Friedlander M, Powell B, Bell-McGuinn KM*, et al.* (2010). Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial. *Lancet* **376**(9737)**:** 245-251.

Bajrami I, Kigozi A, Van Weverwijk A, Brough R, Frankum J, Lord CJ*, et al.* (2012). Synthetic lethality of PARP and NAMPT inhibition in triple-negative breast cancer cells. *EMBO molecular medicine* **4**(10)**:** 1087-1096.

Balmana J, Domchek SM, Tutt A, Garber JE (2011). Stumbling blocks on the path to personalized medicine in breast cancer: the case of PARP inhibitors for BRCA1/2-associated cancers. *Cancer discovery* **1**(1)**:** 29-34.

Barber LJ, Rosa Rosa JM, Kozarewa I, Fenwick K, Assiotis I, Mitsopoulos C*, et al.* (2011). Comprehensive genomic analysis of a BRCA2 deficient human pancreatic cancer. *PloS one* **6**(7)**:** e21639.

Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF*, et al.* (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**(7269)**:** 108-112.

Bartek J, Lukas J (2003). Chk1 and Chk2 kinases in checkpoint control and cancer. *Cancer cell* **3**(5)**:** 421-429.

Bartolucci R, Wei J, Sanchez JJ, Perez-Roca L, Chaib I, Puma F*, et al.* (2009). XPG mRNA expression levels modulate prognosis in resected non-small-cell lung cancer in conjunction with BRCA1 and ERCC1 expression. *Clinical lung cancer* **10**(1)**:** 47-52.

Bassi C, Ho J, Srikumar T, Dowling RJ, Gorrini C, Miller SJ*, et al.* (2013). Nuclear PTEN controls DNA repair and sensitivity to genotoxic stress. *Science (New York, N.Y* **341**(6144)**:** 395-399.

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* **57**(1)**:** 289-300.

Bepler G, Olaussen KA, Vataire AL, Soria JC, Zheng Z, Dunant A*, et al.* (2011). ERCC1 and RRM1 in the international adjuvant lung trial by automated quantitative in situ analysis. *The American journal of pathology* **178**(1)**:** 69-78.

Bepler G, Sommers KE, Cantor A, Li X, Sharma A, Williams C*, et al.* (2008). Clinical efficacy and predictive molecular markers of neoadjuvant gemcitabine and pemetrexed in resectable non-small cell lung cancer. *J Thorac Oncol* **3**(10)**:** 1112-1118.

Bergman AM, Eijk PP, Ruiz van Haperen VW, Smid K, Veerman G, Hubeek I*, et al.* (2005). In vivo induction of resistance to gemcitabine results in increased expression of ribonucleotide reductase subunit M1 as the major determinant. *Cancer research* **65**(20)**:** 9510-9516.

Besse B, Massard C, Haddad V, Andre F, Dunant A, Pirker R*, et al.* (2011). ERCC1 influence on the incidence of brain metastases in patients with non-squamous NSCLC treated with adjuvant cisplatin-based chemotherapy. *Ann Oncol*.

Besse B, Olaussen KA, Soria JC (2013). ERCC1 and RRM1: ready for prime time? *J Clin Oncol* **31**(8)**:** 1050-1060.

Bessho T, Sancar A, Thompson LH, Thelen MP (1997). Reconstitution of human excision nuclease with recombinant XPF-ERCC1 complex. *The Journal of biological chemistry* **272**(6)**:** 3833-3837.

Bhagwat N, Olsen AL, Wang AT, Hanada K, Stuckert P, Kanaar R*, et al.* (2009a). XPF-ERCC1 participates in the Fanconi anemia pathway of cross-link repair. *Molecular and cellular biology* **29**(24)**:** 6427-6437.

Bhagwat NR, Roginskaya VY, Acquafondata MB, Dhir R, Wood RD, Niedernhofer LJ (2009b). Immunodetection of DNA repair endonuclease ERCC1-XPF in human tissue. *Cancer research* **69**(17)**:** 6831-6838.

Birch JM, Alston RD, McNally RJ, Evans DG, Kelsey AM, Harris M*, et al.* (2001). Relative frequency and morphology of cancers in carriers of germline TP53 mutations. *Oncogene* **20**(34)**:** 4621-4628.

Birkelbach M, Ferraiolo N, Gheorghiu L, Pfaffle HN, Daly B, Ebright MI*, et al.* (2013). Detection of impaired homologous recombination repair in NSCLC cells and tissues. *J Thorac Oncol* **8**(3)**:** 279-286.

Birmingham A, Anderson EM, Reynolds A, Ilsley-Tyree D, Leake D, Fedorov Y*, et al.* (2006). 3' UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nature methods* **3**(3)**:** 199-204.

Boukovinas I, Papadaki C, Mendez P, Taron M, Mavroudis D, Koutsopoulos A*, et al.* (2008). Tumor BRCA1, RRM1 and RRM2 mRNA expression levels and clinical response to first-line gemcitabine plus docetaxel in non-small-cell lung cancer patients. *PloS one* **3**(11)**:** e3695.

Boulton S, Kyle S, Durkacz BW (2000). Mechanisms of enhancement of cytotoxicity in etoposide and ionising radiation-treated cells by the protein kinase inhibitor wortmannin. *Eur J Cancer* **36**(4)**:** 535-541.

Boutros M, Bras LP, Huber W (2006). Analysis of cell-based RNAi screens. *Genome biology* **7**(7)**:** R66.

Brenner JC, Feng FY, Han S, Patel S, Goyal SV, Bou-Maroun LM*, et al.* (2012). PARP-1 inhibition as a targeted strategy to treat Ewing's sarcoma. *Cancer research* **72**(7)**:** 1608-1613.

Brough R, Frankum JR, Sims D, Mackay A, Mendes-Pereira AM, Bajrami I*, et al.* (2011). Functional viability profiles of breast cancer. *Cancer discovery* **1**(3)**:** 260-273.

Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E*, et al.* (2005). Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**(7035)**:** 913-917.

Buettner R, Wolf J, Thomas RK (2013). Lessons learned from lung cancer genomics: the emerging concept of individualized diagnostics and treatment. *J Clin Oncol* **31**(15)**:** 1858-1865.

Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B*, et al.* (2013). APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494**(7437)**:** 366-370.

Caiafa P, Guastafierro T, Zampieri M (2009). Epigenetics: poly(ADP-ribosyl)ation of PARP-1 regulates genomic methylation patterns. *FASEB J* **23**(3)**:** 672-678.

Camps C, Sirera R, Iranzo V, Taron M, Rosell R (2007). Gene expression and polymorphisms of DNA repair enzymes: cancer susceptibility and response to chemotherapy. *Clinical lung cancer* **8**(6)**:** 369-375.

Carreras CW, Santi DV (1995). The catalytic mechanism and structure of thymidylate synthase. *Annual review of biochemistry* **64:** 721-762.

Carretero J, Shimamura T, Rikova K, Jackson AL, Wilkerson MD, Borgman CL*, et al.* (2010). Integrative genomic and proteomic analyses identify targets for Lkb1-deficient metastatic lung tumors. *Cancer cell* **17**(6)**:** 547-559.

Cea M, Cagnetta A, Fulciniti M, Tai YT, Hideshima T, Chauhan D*, et al.* (2012). Targeting NAD+ salvage pathway induces autophagy in multiple myeloma cells via mTORC1 and extracellular signal-regulated kinase (ERK1/2) inhibition. *Blood* **120**(17)**:** 3519-3529.

Ceppi P, Volante M, Novello S, Rapa I, Danenberg KD, Danenberg PV*, et al.* (2006a). ERCC1 and RRM1 gene expressions but not EGFR are predictive of shorter survival in advanced non-small-cell lung cancer treated with cisplatin and gemcitabine. *Ann Oncol* **17**(12)**:** 1818-1825.

Ceppi P, Volante M, Saviozzi S, Rapa I, Novello S, Cambieri A*, et al.* (2006b). Squamous cell carcinoma of the lung compared with other histotypes shows higher messenger RNA and protein levels for thymidylate synthase. *Cancer* **107**(7)**:** 1589-1596.

Chabalier C, Lamare C, Racca C, Privat M, Valette A, Larminat F (2006). BRCA1 downregulation leads to premature inactivation of spindle checkpoint and confers paclitaxel resistance. *Cell cycle (Georgetown, Tex* **5**(9)**:** 1001-1007.

Chadwick CA, Potten CS, Nikaido O, Patsunaga T, Proby C, Young AR (1995). The detection of cyclobutane thymine dimers, (6-4)photolesions and the Dewar photoisomers in sections of UV-irradiated human skin using specific antibodies, and the demonstration of depth penetration effects. *Journal of Photochemistry and Photobiology* **28:** 163-170.

Chang P, Jacobson MK, Mitchison TJ (2004). Poly(ADP-ribose) is required for spindle assembly and structure. *Nature* **432**(7017)**:** 645-649.

Chen CC, Kennedy RD, Sidi S, Look AT, D'Andrea A (2009). CHK1 inhibition as a strategy for targeting Fanconi Anemia (FA) DNA repair pathway deficient tumors. *Molecular cancer* **8:** 24.

Chen CY, Chang YL, Shih JY, Lin JW, Chen KY, Yang CH*, et al.* (2011). Thymidylate synthase and dihydrofolate reductase expression in non-small cell lung carcinoma: the association with treatment efficacy of pemetrexed. *Lung cancer (Amsterdam, Netherlands)* **74**(1)**:** 132-138.

Chen S, Zhang J, Wang R, Luo X, Chen H (2010). The platinum-based treatments for advanced non-small cell lung cancer, is low/negative ERCC1 expression better than high/positive ERCC1 expression? A meta-analysis. *Lung cancer (Amsterdam, Netherlands)* **70**(1)**:** 63-70.

Chen Z, Cheng K, Walton Z, Wang Y, Ebi H, Shimamura T*, et al.* (2012). A murine lung cancer co-clinical trial identifies genetic modifiers of therapeutic response. *Nature* **483**(7391)**:** 613-617.

Cheng H, Zhang Z, Borczuk A, Powell CA, Balajee AS, Lieberman HB*, et al.* (2013). PARP inhibition selectively increases sensitivity to cisplatin in ERCC1-low non-small cell lung cancer cells. *Carcinogenesis* **34**(4)**:** 739-749.

Chiappori A, Simon G, Williams C, Haura E, Rocha-Lima C, Wagner H*, et al.* (2005). Phase II study of first-line sequential chemotherapy with gemcitabine-carboplatin followed by docetaxel in patients with advanced non-small cell lung cancer. *Oncology* **68**(4-6)**:** 382-390.

Chiarugi A, Dolle C, Felici R, Ziegler M (2012). The NAD metabolome--a key determinant of cancer cell biology. *Nature reviews* **12**(11)**:** 741-752.

Chiu RK, Brun J, Ramaekers C, Theys J, Weng L, Lambin P*, et al.* (2006). Lysine 63-polyubiquitination guards against translesion synthesis-induced mutations. *PLoS genetics* **2**(7)**:** e116.

Cobo M, Isla D, Massuti B, Montes A, Sanchez JM, Provencio 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. *J Clin Oncol* **25**(19)**:** 2747-2754.

Cooper WA, Kohonen-Corish MR, Chan C, Kwun SY, McCaughan B, Kennedy C*, et al.* (2008). Prognostic significance of DNA repair proteins MLH1, MSH2 and MGMT expression in non-small-cell lung cancer and precursor lesions. *Histopathology* **52**(5)**:** 613-622.

Cox J, Matic I, Hilger M, Nagaraj N, Selbach M, Olsen JV*, et al.* (2009). A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics. *Nature protocols* **4**(5)**:** 698-705.

Dabholkar M, Vionnet J, Parker R, Bostickbruton F, Dobbins A, Reed E (1995). Expression of an alternatively spliced ercc1 messenger-RNA species, is related to reduced DNA-repair efficiency in human T-lymphocytes. *Oncol Rep* **2**(2)**:** 209-214.

Das D, Folkers GE, van Dijk M, Jaspers NG, Hoeijmakers JH, Kaptein R*, et al.* (2012). The structure of the XPF-ssDNA complex underscores the distinct roles of the XPF and ERCC1 helix- hairpin-helix domains in ss/ds DNA recognition. *Structure* **20**(4)**:** 667-675.

Das D, Tripsianes K, Jaspers NG, Hoeijmakers JH, Kaptein R, Boelens R*, et al.* (2008). The HhH domain of the human DNA repair protein XPF forms stable homodimers. *Proteins* **70**(4)**:** 1551-1563.

Davidson JD, Ma L, Flagella M, Geeganage S, Gelbert LM, Slapak CA (2004). An increase in the expression of ribonucleotide reductase large subunit 1 is associated with gemcitabine resistance in non-small cell lung cancer cell lines. *Cancer research* **64**(11)**:** 3761-3766.

de Laat WL, Jaspers NG, Hoeijmakers JH (1999). Molecular mechanism of nucleotide excision repair. *Genes & development* **13**(7)**:** 768-785.

de las Penas R, Sanchez-Ronco M, Alberola V, Taron M, Camps C, Garcia-Carbonero R*, et al.* (2006). Polymorphisms in DNA repair genes modulate survival in cisplatin/gemcitabine-treated non-small-cell lung cancer patients. *Ann Oncol* **17**(4)**:** 668-675.

Deans AJ, West SC (2011). DNA interstrand crosslink repair and cancer. *Nature reviews* **11**(7)**:** 467-480.

Dent RA, Lindeman GJ, Clemons M, Wildiers H, Chan A, McCarthy NJ*, et al.* (2013). Phase I trial of the oral PARP inhibitor olaparib in combination with paclitaxel for first- or second-line treatment of patients with metastatic triple-negative breast cancer. *Breast Cancer Res* **15**(5)**:** R88.

Donawho CK, Luo Y, Luo Y, Penning TD, Bauch JL, Bouska JJ*, et al.* (2007). ABT-888, an orally active poly(ADP-ribose) polymerase inhibitor that potentiates DNA-damaging agents in preclinical tumor models. *Clin Cancer Res* **13**(9)**:** 2728-2737.

Du P, Kibbe WA, Lin SM (2008). lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**(13)**:** 1547-1548.

Dzagnidze A, Katsarava Z, Makhalova J, Liedert B, Yoon MS, Kaube H*, et al.* (2007). Repair capacity for platinum-DNA adducts determines the severity of cisplatin-induced peripheral neuropathy. *J Neurosci* **27**(35)**:** 9451-9457.

Echeverri CJ, Beachy PA, Baum B, Boutros M, Buchholz F, Chanda SK*, et al.* (2006). Minimizing the risk of reporting false positives in large-scale RNAi screens. *Nature methods* **3**(10)**:** 777-779.

Edelman MJ, Le Chevalier T, Soria JC (2012). Maintenance therapy and advanced non-small-cell lung cancer: a skeptic's view. *J Thorac Oncol* **7**(9)**:** 1331-1336.

Edwards SL, Brough R, Lord CJ, Natrajan R, Vatcheva R, Levine DA*, et al.* (2008). Resistance to therapy caused by intragenic deletion in BRCA2. *Nature* **451**(7182)**:** 1111-1115.

Egeblad L, Welin M, Flodin S, Graslund S, Wang L, Balzarini J*, et al.* (2012). Pan-pathway based interaction profiling of FDA-approved nucleoside and nucleobase analogs with enzymes of the human nucleotide metabolism. *PloS one* **7**(5)**:** e37724.

Engelman JA, Chen L, Tan X, Crosby K, Guimaraes AR, Upadhyay R*, et al.* (2008). Effective use of PI3K and MEK inhibitors to treat mutant Kras G12D and PIK3CA H1047R murine lung cancers. *Nature medicine* **14**(12)**:** 1351-1356.

Eriksson A, Yachnin J, Lewensohn R, Nilsson A (2001). DNA-dependent protein kinase is inhibited by trifluoperazine. *Biochemical and biophysical research communications* **283**(4)**:** 726-731.

Fagbemi AF, Orelli B, Scharer OD Regulation of endonuclease activity in human nucleotide excision repair. *DNA repair*.

Fagbemi AF, Orelli B, Scharer OD (2011). Regulation of endonuclease activity in human nucleotide excision repair. *DNA repair* **10**(7)**:** 722-729.

Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB*, et al.* (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**(7035)**:** 917-921.

Fedorov Y, Anderson EM, Birmingham A, Reynolds A, Karpilow J, Robinson K*, et al.* (2006). Off-target effects by siRNA can induce toxic phenotype. *Rna* **12**(7)**:** 1188-1196.

Firestein BL, Firestein BL, Brenman JE, Aoki C, Sanchez-Perez AM, El-Husseini AE*, et al.* (1999). Cypin: a cytosolic regulator of PSD-95 postsynaptic targeting. *Neuron* **24**(3)**:** 659-672.

Fisher LA, Samson L, Bessho T (2011). Removal of reactive oxygen species-induced 3'-blocked ends by XPF-ERCC1. *Chemical research in toxicology* **24**(11)**:** 1876-1881.

Fong PC, Boss DS, Yap TA, Tutt A, Wu P, Mergui-Roelvink M*, et al.* (2009). Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *The New England journal of medicine* **361**(2)**:** 123-134.

Friboulet L, Barrios-Gonzales D, Commo F, Olaussen KA, Vagner S, Adam J*, et al.* (2011). Molecular Characteristics of ERCC1-Negative versus ERCC1-Positive Tumors in Resected NSCLC. *Clin Cancer Res* **17**(17)**:** 5562-5572.

Friboulet L, Olaussen KA, Pignon JP, Shepherd FA, Tsao MS, Graziano S*, et al.* (2013a). ERCC1 isoform expression and DNA repair in non-small-cell lung cancer. *The New England journal of medicine* **368**(12)**:** 1101-1110.

Friboulet L, Postel-Vinay S, Sourisseau T, Adam J, Stoclin A, Ponsonnailles F*, et al.* (2013b). ERCC1 function in nuclear excision and interstrand crosslink repair pathways is mediated exclusively by the ERCC1- 202 isoform. *Cell cycle (Georgetown, Tex* **12**(20).

Furuichi M, Yoshida MC, Oda H, Tajiri T, Nakabeppu Y, Tsuzuki T*, et al.* (1994). Genomic structure and chromosome location of the human mutT homologue gene MTH1 encoding 8-oxo-dGTPase for prevention of A:T to C:G transversion. *Genomics* **24**(3)**:** 485-490.

Gaillard PH, Wood RD (2001). Activity of individual ERCC1 and XPF subunits in DNA nucleotide excision repair. *Nucleic acids research* **29**(4)**:** 872-879.

Gaj T, Gersbach CA, Barbas CF, 3rd (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in biotechnology* **31**(7)**:** 397-405.

Galluzzi L, Senovilla L, Vitale I, Michels J, Martins I, Kepp O*, et al.* (2012). Molecular mechanisms of cisplatin resistance. *Oncogene* **31**(15)**:** 1869-1883.

Galvani E, Peters GJ, Giovannetti E (2011). Thymidylate synthase inhibitors for non-small cell lung cancer. *Expert opinion on investigational drugs* **20**(10)**:** 1343-1356.

Geiger T, Wisniewski JR, Cox J, Zanivan S, Kruger M, Ishihama Y*, et al.* (2011). Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nature protocols* **6**(2)**:** 147-157.

Giovannetti E, Lemos C, Tekle C, Smid K, Nannizzi S, Rodriguez JA*, et al.* (2008). Molecular mechanisms underlying the synergistic interaction of erlotinib, an epidermal growth factor receptor tyrosine kinase inhibitor, with the multitargeted antifolate pemetrexed in non-small-cell lung cancer cells. *Molecular pharmacology* **73**(4)**:** 1290-1300.

Giovannetti E, Mey V, Nannizzi S, Pasqualetti G, Marini L, Del Tacca M*, et al.* (2005). Cellular and pharmacogenetics foundation of synergistic interaction of pemetrexed and gemcitabine in human non-small-cell lung cancer cells. *Molecular pharmacology* **68**(1)**:** 110-118.

Gorlova OY, Weng SF, Zhang Y, Amos CI, Spitz MR, Wei Q (2008). DNA repair capacity and lung cancer risk in never smokers. *Cancer Epidemiol Biomarkers Prev* **17**(6)**:** 1322-1328.

Graeser M, McCarthy A, Lord CJ, Savage K, Hills M, Salter J*, et al.* (2010). A marker of homologous recombination predicts pathologic complete response to neoadjuvant chemotherapy in primary breast cancer. *Clin Cancer Res* **16**(24)**:** 6159-6168.

Gregg SQ, Robinson AR, Niedernhofer LJ (2011). Physiological consequences of defects in ERCC1-XPF DNA repair endonuclease. *DNA repair* **10**(7)**:** 781-791.

Gurubhagavatula S, Liu G, Park S, Zhou W, Su L, Wain JC*, et al.* (2004). XPD and XRCC1 genetic polymorphisms are prognostic factors in advanced non-small-cell lung cancer patients treated with platinum chemotherapy. *J Clin Oncol* **22**(13)**:** 2594-2601.

Hainaut P, Pfeifer GP (2001). Patterns of p53 G-->T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke. *Carcinogenesis* **22**(3)**:** 367-374.

Haince JF, McDonald D, Rodrigue A, Dery U, Masson JY, Hendzel MJ*, et al.* (2008). PARP1-dependent kinetics of recruitment of MRE11 and NBS1 proteins to multiple DNA damage sites. *The Journal of biological chemistry* **283**(2)**:** 1197-1208.

Hammerman PS, Sos ML, Ramos AH, Xu C, Dutt A, Zhou W*, et al.* (2011). Mutations in the DDR2 kinase gene identify a novel therapeutic target in squamous cell lung cancer. *Cancer discovery* **1**(1)**:** 78-89.

Hanahan D, Weinberg RA (2000). The hallmarks of cancer. *Cell* **100**(1)**:** 57-70.

Hanahan D, Weinberg RA (2011). Hallmarks of cancer: the next generation. *Cell* **144**(5)**:** 646-674.

Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S*, et al.* (2007). American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol* **25**(33)**:** 5287-5312.

Hasmann M, Schemainda I (2003). FK866, a highly specific noncompetitive inhibitor of nicotinamide phosphoribosyltransferase, represents a novel mechanism for induction of tumor cell apoptosis. *Cancer research* **63**(21)**:** 7436-7442.

Hassler M, Ladurner AG (2012). Towards a structural understanding of PARP1 activation and related signalling ADP-ribosyl-transferases. *Current opinion in structural biology* **22**(6)**:** 721-729.

Helleday T, Petermann E, Lundin C, Hodgson B, Sharma RA (2008). DNA repair pathways as targets for cancer therapy. *Nature reviews* **8**(3)**:** 193-204.

Hewish M, Lord CJ, Martin SA, Cunningham D, Ashworth A (2010). Mismatch repair deficient colorectal cancer in the era of personalized treatment. *Nature reviews. Clinical oncology* **7**(4)**:** 197-208.

Hoeijmakers JH (2001). Genome maintenance mechanisms for preventing cancer. *Nature* **411**(6835)**:** 366-374.

Holen K, Saltz LB, Hollywood E, Burk K, Hanauske AR (2008). The pharmacokinetics, toxicities, and biologic effects of FK866, a nicotinamide adenine dinucleotide biosynthesis inhibitor. *Investigational new drugs* **26**(1)**:** 45-51.

Hsu HS, Lee IH, Hsu WH, Kao WT, Wang YC (2007). Polymorphism in the hMSH2 gene (gISV12-6T > C) is a prognostic factor in non-small cell lung cancer. *Lung cancer (Amsterdam, Netherlands)* **58**(1)**:** 123-130.

Hsu HS, Wen CK, Tang YA, Lin RK, Li WY, Hsu WH*, et al.* (2005). Promoter hypermethylation is the predominant mechanism in hMLH1 and hMSH2 deregulation and is a poor prognostic factor in nonsmoking lung cancer. *Clin Cancer Res* **11**(15)**:** 5410-5416.

Hu Y, Liu J, Huang H (2013). Recent agents targeting HIF-1alpha for cancer therapy. *Journal of cellular biochemistry* **114**(3)**:** 498-509.

Ihle NT, Byers LA, Kim ES, Saintigny P, Lee JJ, Blumenschein GR*, et al.* (2012). Effect of KRAS oncogene substitutions on protein behavior: implications for signaling and clinical outcome. *Journal of the National Cancer Institute* **104**(3)**:** 228-239.

Imoto K, Boyle J, Oh K, Khan S, Busch D, Jaspers NG*, et al.* (2007). Patients with defects in the interacting nucleotide excision repair proteins ERCC1 or XPF show xeroderma pigmentosum with late onset severe neurological degeneration. *Journal of Investigative Dematology* **127**(Supp. 92).

Iorns E, Lord CJ, Grigoriadis A, McDonald S, Fenwick K, Mackay A*, et al.* (2009). Integrated functional, gene expression and genomic analysis for the identification of cancer targets. *PloS one* **4**(4)**:** e5120.

Iorns E, Lord CJ, Turner N, Ashworth A (2007). Utilizing RNA interference to enhance cancer drug discovery. *Nature reviews. Drug discovery* **6**(7)**:** 556-568.

Izzard RA, Jackson SP, Smith GC (1999). Competitive and noncompetitive inhibition of the DNA-dependent protein kinase. *Cancer research* **59**(11)**:** 2581-2586.

Jackson AL, Bartz SR, Schelter J, Kobayashi SV, Burchard J, Mao M*, et al.* (2003). Expression profiling reveals off-target gene regulation by RNAi. *Nature biotechnology* **21**(6)**:** 635-637.

Jackson AL, Burchard J, Leake D, Reynolds A, Schelter J, Guo J*, et al.* (2006). Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. *Rna* **12**(7)**:** 1197-1205.

Jackson D, Dhar K, Wahl JK, Wold MS, Borgstahl GE (2002). Analysis of the human replication protein A:Rad52 complex: evidence for crosstalk between RPA32, RPA70, Rad52 and DNA. *Journal of molecular biology* **321**(1)**:** 133-148.

Jalal S, Earley JN, Turchi JJ (2011). DNA repair: from genome maintenance to biomarker and therapeutic target. *Clin Cancer Res* **17**(22)**:** 6973-6984.

Jasin M (2002). Homologous repair of DNA damage and tumorigenesis: the BRCA connection. *Oncogene* **21**(58)**:** 8981-8993.

Jaspers NG, Raams A, Silengo MC, Wijgers N, Niedernhofer LJ, Robinson AR*, et al.* (2007). First reported patient with human ERCC1 deficiency has cerebro-oculo-facio-skeletal syndrome with a mild defect in nucleotide excision repair and severe developmental failure. *American journal of human genetics* **80**(3)**:** 457-466.

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D (2011). Global cancer statistics. *CA: a cancer journal for clinicians* **61**(2)**:** 69-90.

Jin G, Kim MJ, Jeon HS, Choi JE, Kim DS, Lee EB*, et al.* (2010). PTEN mutations and relationship to EGFR, ERBB2, KRAS, and TP53 mutations in non-small cell lung cancers. *Lung cancer (Amsterdam, Netherlands)* **69**(3)**:** 279-283.

Jiricny J (2006). The multifaceted mismatch-repair system. *Nature reviews. Molecular cell biology* **7**(5)**:** 335-346.

Johnson N, Li YC, Walton ZE, Cheng KA, Li D, Rodig SJ*, et al.* (2011). Compromised CDK1 activity sensitizes BRCA-proficient cancers to PARP inhibition. *Nature medicine* **17**(7)**:** 875-882.

Kaelin WG, Jr. (2005). The concept of synthetic lethality in the context of anticancer therapy. *Nature reviews* **5**(9)**:** 689-698.

Kaklamani VG, Pasche B (2004). Role of TGF-beta in cancer and the potential for therapy and prevention. *Expert review of anticancer therapy* **4**(4)**:** 649-661.

Kamal NS, Soria JC, Mendiboure J, Planchard D, Olaussen KA, Rousseau V*, et al.* (2010). MutS homologue 2 and the long-term benefit of adjuvant chemotherapy in lung cancer. *Clin Cancer Res* **16**(4)**:** 1206-1215.

Kamileri I, Karakasilioti I, Sideri A, Kosteas T, Tatarakis A, Talianidis I*, et al.* (2012). Defective transcription initiation causes postnatal growth failure in a mouse model of nucleotide excision repair (NER) progeria. *Proceedings of the National Academy of Sciences of the United States of America* **109**(8)**:** 2995-3000.

Kanellis G, Chatzistamou I, Koutselini H, Politi E, Gouliamos A, Vlahos L*, et al.* (2006). Expression of DNA mismatch repair gene MSH2 in cytological material from lung cancer patients. *Diagnostic cytopathology* **34**(7)**:** 463-466.

Karachaliou N, Mayo C, Costa C, Magri I, Gimenez-Capitan A, Molina-Vila MA*, et al.* (2013). KRAS mutations in lung cancer. *Clinical lung cancer* **14**(3)**:** 205-214.

Kashiyama K, Nakazawa Y, Pilz DT, Guo C, Shimada M, Sasaki K*, et al.* (2013). Malfunction of nuclease ERCC1-XPF results in diverse clinical manifestations and causes Cockayne syndrome, xeroderma pigmentosum, and Fanconi anemia. *American journal of human genetics* **92**(5)**:** 807-819.

Kastan MB, Bartek J (2004). Cell-cycle checkpoints and cancer. *Nature* **432**(7015)**:** 316-323.

Kedar PS, Stefanick DF, Horton JK, Wilson SH (2012). Increased PARP-1 association with DNA in alkylation damaged, PARP-inhibited mouse fibroblasts. *Mol Cancer Res* **10**(3)**:** 360-368.

Khambata-Ford S, Harbison CT, Hart LL, Awad M, Xu LA, Horak CE*, et al.* (2010). Analysis of potential predictive markers of cetuximab benefit in BMS099, a phase III study of cetuximab and first-line taxane/carboplatin in advanced non-small-cell lung cancer. *J Clin Oncol* **28**(6)**:** 918-927.

Kim ES, Hirsh V, Mok T, Socinski MA, Gervais R, Wu YL*, et al.* (2008). Gefitinib versus docetaxel in previously treated non-small-cell lung cancer (INTEREST): a randomised phase III trial. *Lancet* **372**(9652)**:** 1809-1818.

Kirschner K, Melton DW (2010). Multiple roles of the ERCC1-XPF endonuclease in DNA repair and resistance to anticancer drugs. *Anticancer research* **30**(9)**:** 3223-3232.

Koh SH, Chang DI, Kim HT, Kim J, Kim MH, Kim KS*, et al.* (2005). Effect of 3-aminobenzamide, PARP inhibitor, on matrix metalloproteinase-9 level in plasma and brain of ischemic stroke model. *Toxicology* **214**(1-2)**:** 131-139.

Kohno T, Ichikawa H, Totoki Y, Yasuda K, Hiramoto M, Nammo T*, et al.* (2012). KIF5B-RET fusions in lung adenocarcinoma. *Nature medicine* **18**(3)**:** 375-377.

Korbel JO, Campbell PJ (2013). Criteria for inference of chromothripsis in cancer genomes. *Cell* **152**(6)**:** 1226-1236.

Kouso H, Yoshino I, Miura N, Takenaka T, Ohba T, Yohena T*, et al.* (2008). Expression of mismatch repair proteins, hMLH1/hMSH2, in non-small cell lung cancer tissues and its clinical significance. *Journal of surgical oncology* **98**(5)**:** 377-383.

Kraemer KH, Patronas NJ, Schiffmann R, Brooks BP, Tamura D, DiGiovanna JJ (2007). Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: a complex genotype-phenotype relationship. *Neuroscience* **145**(4)**:** 1388-1396.

Kraus WL (2008). Transcriptional control by PARP-1: chromatin modulation, enhancer-binding, coregulation, and insulation. *Current opinion in cell biology* **20**(3)**:** 294-302.

Kris MG, Johnson BE, Kwiatkowsky DJ, Iafrate AJ, Wistuba I, Aronson SL*, et al.* (2011). Identification of driver mutatons in tumor specimen from 1000 patients with lung adenocarcinoma: the NCI's Lung Cancer Mutation Consortium (LCMC). *Journal of Clinical Oncology* **29**(18_suppl).

Kubo K, Honda H, Honda H, Sannomiya K, Aying Y, Mei W*, et al.* (2006). Histochemical and immunohistochemical investigation of guanase and nedasin in human tissues. *The journal of medical investigation : JMI* **53**(3-4)**:** 264-270.

Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG*, et al.* (2010). Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *The New England journal of medicine* **363**(18)**:** 1693-1703.

Kwon WS, Rha SY, Choi YH, Lee JO, Park KH, Jung JJ*, et al.* (2006). Ribonucleotide reductase M1 (RRM1) 2464G>A polymorphism shows an association with gemcitabine chemosensitivity in cancer cell lines. *Pharmacogenetics and genomics* **16**(6)**:** 429-438.

Langelier MF, Planck JL, Roy S, Pascal JM (2011). Crystal structures of poly(ADP-ribose) polymerase-1 (PARP-1) zinc fingers bound to DNA: structural and functional insights into DNA-dependent PARP-1 activity. *The Journal of biological chemistry* **286**(12)**:** 10690-10701.

Langelier MF, Planck JL, Roy S, Pascal JM (2012). Structural basis for DNA damage-dependent poly(ADP-ribosyl)ation by human PARP-1. *Science (New York, N.Y* **336**(6082)**:** 728-732.

Langer CJ (2012). Exploring biomarkers in head and neck cancer. *Cancer* **118**(16)**:** 3882-3892.

Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A*, et al.* (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**(7457)**:** 214-218.

Le May N, Mota-Fernandes D, Velez-Cruz R, Iltis I, Biard D, Egly JM (2010). NER factors are recruited to active promoters and facilitate chromatin modification for transcription in the absence of exogenous genotoxic attack. *Molecular cell* **38**(1)**:** 54-66.

Ledermann J, Harter P, Gourley C, Friedlander M, Vergote I, Rustin G*, et al.* (2012). Olaparib maintenance therapy in platinum-sensitive relapsed ovarian cancer. *The New England journal of medicine* **366**(15)**:** 1382-1392.

Lee MN, Tseng RC, Hsu HS, Chen JY, Tzao C, Ho WL*, et al.* (2007). Epigenetic inactivation of the chromosomal stability control genes BRCA1, BRCA2, and XRCC5 in non-small cell lung cancer. *Clin Cancer Res* **13**(3)**:** 832-838.

Lehmann AR (2001). The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases. *Genes & development* **15**(1)**:** 15-23.

Leung AK, Vyas S, Rood JE, Bhutkar A, Sharp PA, Chang P (2011). Poly(ADP-ribose) regulates stress responses and microRNA activity in the cytoplasm. *Molecular cell* **42**(4)**:** 489-499.

Li L, Elledge SJ, Peterson CA, Bales ES, Legerski RJ (1994). Specific association between the human DNA repair proteins XPA and ERCC1. *Proceedings of the National Academy of Sciences of the United States of America* **91**(11)**:** 5012-5016.

Lieber MR, Ma Y, Pannicke U, Schwarz K (2003). Mechanism and regulation of human non-homologous DNA end-joining. *Nature reviews. Molecular cell biology* **4**(9)**:** 712-720.

Linardou H, Briasoulis E, Dahabreh IJ, Mountzios G, Papadimitriou C, Papadopoulos S*, et al.* (2011). All about KRAS for clinical oncology practice: gene profile, clinical implications and laboratory recommendations for somatic mutational testing in colorectal cancer. *Cancer treatment reviews* **37**(3)**:** 221-233.

Lindahl T, Wood RD (1999). Quality control by DNA repair. *Science (New York, N.Y* **286**(5446)**:** 1897-1905.

Lipson D, Capelletti M, Yelensky R, Otto G, Parker A, Jarosz M*, et al.* (2012). Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. *Nature medicine* **18**(3)**:** 382-384.

Lisy K, Peet DJ (2008). Turn me on: regulating HIF transcriptional activity. *Cell death and differentiation* **15**(4)**:** 642-649.

Liu SK, Coackley C, Krause M, Jalali F, Chan N, Bristow RG (2008). A novel poly(ADP-ribose) polymerase inhibitor, ABT-888, radiosensitizes malignant human cell lines under hypoxia. *Radiother Oncol* **88**(2)**:** 258-268.

Liu X, Erikson RL (2003). Polo-like kinase (Plk)1 depletion induces apoptosis in cancer cells. *Proceedings of the National Academy of Sciences of the United States of America* **100**(10)**:** 5789-5794.

Liu X, Shi Y, Maag DX, Palma JP, Patterson MJ, Ellis PA*, et al.* (2012). Iniparib nonselectively modifies cysteine-containing proteins in tumor cells and is not a bona fide PARP inhibitor. *Clin Cancer Res* **18**(2)**:** 510-523.

Lord CJ, Ashworth A (2008). Targeted therapy for cancer using PARP inhibitors. *Curr Opin Pharmacol* **8**(4)**:** 363-369.

Lord CJ, Ashworth A (2012). The DNA damage response and cancer therapy. *Nature* **481**(7381)**:** 287-294.

Lord CJ, Garrett MD, Ashworth A (2006). Targeting the double-strand DNA break repair pathway as a therapeutic strategy. *Clin Cancer Res* **12**(15)**:** 4463-4468.

Lord CJ, Martin SA, Ashworth A (2009). RNA interference screening demystified. *J Clin Pathol* **62**(3)**:** 195-200.

Lord RV, Brabender J, Gandara D, Alberola V, Camps C, Domine M*, et al.* (2002). Low ERCC1 expression correlates with prolonged survival after cisplatin plus gemcitabine chemotherapy in non-small cell lung cancer. *Clin Cancer Res* **8**(7)**:** 2286-2291.

Love JD, Nguyen HT, Or A, Attri AK, Minton KW (1986). UV-induced interstrand cross-linking of d(GT)n.d(CA)n is facilitated by a structural transition. *The Journal of biological chemistry* **261**(22)**:** 10051-10057.

Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V*, et al.* (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**(9859)**:** 2095-2128.

Lu Y, Mani S, Kandimalla ER, Yu D, Agrawal S, States JC*, et al.* (2001). The Cockayne syndrome group B DNA repair protein as an anti-cancer target. *International journal of oncology* **19**(6)**:** 1089-1097.

Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW*, et al.* (2004). Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *The New England journal of medicine* **350**(21)**:** 2129-2139.

Ma D, Baruch D, Shu Y, Yuan K, Sun Z, Ma K*, et al.* (2012). Using protein microarray technology to screen anti-ERCC1 monoclonal antibodies for specificity and applications in pathology. *BMC biotechnology* **12:** 88.

Makowski L, Hayes DN (2008). Role of LKB1 in lung cancer development. *British journal of cancer* **99**(5)**:** 683-688.

Mao C, Qiu LX, Liao RY, Du FB, Ding H, Yang WC*, et al.* (2010). KRAS mutations and resistance to EGFR-TKIs treatment in patients with non-small cell lung cancer: a meta-analysis of 22 studies. *Lung cancer (Amsterdam, Netherlands)* **69**(3)**:** 272-278.

Marsit CJ, Liu M, Nelson HH, Posner M, Suzuki M, Kelsey KT (2004). Inactivation of the Fanconi anemia/BRCA pathway in lung and oral cancers: implications for treatment and survival. *Oncogene* **23**(4)**:** 1000-1004.

Martin SA, Lord CJ, Ashworth A (2006a). Therapeutic targeting of the DNA mismatch repair pathway. *Clin Cancer Res* **16**(21)**:** 5107-5113.

Martin SA, McCabe N, Mullarkey M, Cummins R, Burgess DJ, Nakabeppu Y*, et al.* (2006b). DNA polymerases as potential therapeutic targets for cancers deficient in the DNA mismatch repair proteins MSH2 or MLH1. *Cancer cell* **17**(3)**:** 235-248.

Martin SA, McCarthy A, Barber LJ, Burgess DJ, Parry S, Lord CJ*, et al.* (2009). Methotrexate induces oxidative DNA damage and is selectively lethal to tumour cells with defects in the DNA mismatch repair gene MSH2. *EMBO molecular medicine* **1**(6-7)**:** 323-337.

Matakidou A, el Galta R, Webb EL, Rudd MF, Bridle H, Eisen T*, et al.* (2007). Genetic variation in the DNA repair genes is predictive of outcome in lung cancer. *Human molecular genetics* **16**(19)**:** 2333-2340.

Matsunaga T, Park CH, Bessho T, Mu D, Sancar A (1996). Replication protein A confers structure-specific endonuclease activities to the XPF-ERCC1 and XPG subunits of human DNA repair excision nuclease. *The Journal of biological chemistry* **271**(19)**:** 11047-11050.

McCabe N, Turner NC, Lord CJ, Kluzek K, Bialkowska A, Swift S*, et al.* (2006). Deficiency in the repair of DNA damage by homologous recombination and sensitivity to poly(ADP-ribose) polymerase inhibition. *Cancer research* **66**(16)**:** 8109-8115.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A*, et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**(9)**:** 1297-1303.

McNeil EM, Melton DW (2012). DNA repair endonuclease ERCC1-XPF as a novel therapeutic target to overcome chemoresistance in cancer therapy. *Nucleic acids research* **40**(20)**:** 9990-10004.

McWhir J, Selfridge J, Harrison DJ, Squires S, Melton DW (1993). Mice with DNA repair gene (ERCC-1) deficiency have elevated levels of p53, liver nuclear abnormalities and die before weaning. *Nature genetics* **5**(3)**:** 217-224.

Meira LB, Graham JM, Jr., Greenberg CR, Busch DB, Doughty AT, Ziffer DW*, et al.* (2000). Manitoba aboriginal kindred with original cerebro-oculo- facio-skeletal syndrome has a mutation in the Cockayne syndrome group B (CSB) gene. *American journal of human genetics* **66**(4)**:** 1221-1228.

Mendeleyev J, Kirsten E, Hakam A, Buki KG, Kun E (1995). Potential chemotherapeutic activity of 4-iodo-3-nitrobenzamide. Metabolic reduction to the 3-nitroso derivative and induction of cell death in tumor cells in culture. *Biochemical pharmacology* **50**(5)**:** 705-714.

Mendes-Pereira AM, Martin SA, Brough R, McCarthy A, Taylor JR, Kim JS*, et al.* (2009). Synthetic lethal targeting of PTEN mutant cells with PARP inhibitors. *EMBO molecular medicine* **1**(6-7)**:** 315-322.

Metzger R, Bollschweiler E, Holscher AH, Warnecke-Eberz U (2010). ERCC1: impact in multimodality treatment of upper gastrointestinal cancer. *Future Oncol* **6**(11)**:** 1735-1749.

Michels J, Vitale I, Galluzzi L, Adam J, Olaussen KA, Kepp O*, et al.* (2013). Cisplatin resistance associated with PARP hyperactivation. *Cancer research* **73**(7)**:** 2271-2280.

Miknyoczki SJ, Jones-Bolin S, Pritchard S, Hunter K, Zhao H, Wan W*, et al.* (2003). Chemopotentiation of temozolomide, irinotecan, and cisplatin activity by CEP-6800, a poly(ADP-ribose) polymerase inhibitor. *Molecular cancer therapeutics* **2**(4)**:** 371-382.

Mogi A, Kuwano H (2011). TP53 mutations in nonsmall cell lung cancer. *Journal of biomedicine & biotechnology*.

Mohammed H, D'Santos C, Serandour AA, Ali HR, Brown GD, Atkins A*, et al.* (2013). Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell reports* **3**(2)**:** 342-349.

Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N*, et al.* (2009). Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *The New England journal of medicine* **361**(10)**:** 947-957.

Morgensztern D, Goodgame B, Govindan R (2010). Vaccines and immunotherapy for non-small cell lung cancer. *J Thorac Oncol* **5**(12 Suppl 6)**:** S463-465.

Moskwa P, Buffa FM, Pan Y, Panchakshari R, Gottipati P, Muschel RJ*, et al.* (2011). miR-182-mediated downregulation of BRCA1 impacts DNA repair and sensitivity to PARP inhibitors. *Molecular cell* **41**(2)**:** 210-220.

Munoz IM, Hain K, Declais AC, Gardiner M, Toh GW, Sanchez-Pulido L*, et al.* (2009). Coordination of structure-specific nucleases by human SLX4/BTBD12 is required for DNA repair. *Molecular cell* **35**(1)**:** 116-127.

Munoz P, Blanco R, Flores JM, Blasco MA (2005). XPF nuclease-dependent telomere loss and increased DNA damage in mice overexpressing TRF2 result in premature aging and cancer. *Nature genetics* **37**(10)**:** 1063-1071.

Murai J, Huang SY, Das BB, Renaud A, Zhang Y, Doroshow JH*, et al.* (2012). Trapping of PARP1 and PARP2 by Clinical PARP Inhibitors. *Cancer research* **72**(21)**:** 5588-5599.

Murphy CG, Moynahan ME (2010). BRCA gene structure and function in tumor suppression: a repair-centric perspective. *Cancer journal (Sudbury, Mass* **16**(1)**:** 39-47.

Naim V, Wilhelm T, Debatisse M, Rosselli F (2013). ERCC1 and MUS81-EME1 promote sister chromatid separation by processing late replication intermediates at common fragile sites during mitosis. *Nature cell biology* **15**(8)**:** 1008-1015.

Nel I, Gauler TC, Eberhardt WE, Nickel AC, Schuler M, Thomale J*, et al.* (2013). Formation and repair kinetics of Pt-(GpG) DNA adducts in extracted circulating tumour cells and response to platinum treatment. *British journal of cancer* **109**(5)**:** 1223-1229.

Niedernhofer LJ, Bhagwat N, Wood RD (2007). ERCC1 and non-small-cell lung cancer. *The New England journal of medicine* **356**(24)**:** 2538-2540; author reply 2540-2531.

Niedernhofer LJ, Essers J, Weeda G, Beverloo B, de Wit J, Muijtjens M*, et al.* (2001). The structure-specific endonuclease Ercc1-Xpf is required for targeted gene replacement in embryonic stem cells. *The EMBO journal* **20**(22)**:** 6540-6549.

Niedernhofer LJ, Garinis GA, Raams A, Lalai AS, Robinson AR, Appeldoorn E*, et al.* (2006). A new progeroid syndrome reveals that genotoxic stress suppresses the somatotroph axis. *Nature* **444**(7122)**:** 1038-1043.

Niedernhofer LJ, Odijk H, Budzowska M, van Drunen E, Maas A, Theil AF*, et al.* (2004). The structure-specific endonuclease Ercc1-Xpf is required to resolve DNA interstrand cross-link-induced double-strand breaks. *Molecular and cellular biology* **24**(13)**:** 5776-5787.

Niida H, Tsuge S, Katsuno Y, Konishi A, Takeda N, Nakanishi M (2005). Depletion of Chk1 leads to premature activation of Cdc2-cyclin B and mitotic catastrophe. *The Journal of biological chemistry* **280**(47)**:** 39246-39252.

Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K*, et al.* (2012a). Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**(5)**:** 979-993.

Nik-Zainal S, Van Loo P, Wedge DC, Alexandrov LB, Greenman CD, Lau KW*, et al.* (2012b). The life history of 21 breast cancers. *Cell* **149**(5)**:** 994-1007.

Norquist B, Wurz KA, Pennil CC, Garcia R, Gross J, Sakai W*, et al.* (2011). Secondary somatic mutations restoring BRCA1/2 predict chemotherapy resistance in hereditary ovarian carcinomas. *J Clin Oncol* **29**(22)**:** 3008-3015.

Nouspikel T, Lalle P, Leadon SA, Cooper PK, Clarkson SG (1997). A common mutational pattern in Cockayne syndrome patients from xeroderma pigmentosum group G: implications for a second XPG function. *Proceedings of the National Academy of Sciences of the United States of America* **94**(7)**:** 3116-3121.

Nunez F, Chipchase MD, Clarke AR, Melton DW (2000). Nucleotide excision repair gene (ERCC1) deficiency causes G(2) arrest in hepatocytes and a reduction in liver binucleation: the role of p53 and p21. *FASEB J* **14**(9)**:** 1073-1082.

O'Byrne KJ, Gatzemeier U, Bondarenko I, Barrios C, Eschbach C, Martens UM*, et al.* (2011). Molecular biomarkers in non-small-cell lung cancer: a retrospective analysis of data from the phase 3 FLEX study. *The lancet oncology* **12**(8)**:** 795-805.

Ohashi K, Maruvka YE, Michor F, Pao W (2013). Epidermal growth factor receptor tyrosine kinase inhibitor-resistant disease. *J Clin Oncol* **31**(8)**:** 1070-1080.

Okumura S, Sasaki T, Minami Y, Ohsaki Y (2012). Nicotinamide phosphoribosyltransferase: a potent therapeutic target in non-small cell lung cancer with epidermal growth factor receptor-gene mutation. *J Thorac Oncol* **7**(1)**:** 49-56.

Olaussen KA, Dunant A, Fouret P, Brambilla E, Andre F, Haddad V*, et al.* (2006). DNA repair by ERCC1 in non-small-cell lung cancer and cisplatin-based adjuvant chemotherapy. *The New England journal of medicine* **355**(10)**:** 983-991.

Olaussen KA, Fouret P, Kroemer G (2007). ERCC1-specific immunostaining in non-small-cell lung cancer. *The New England journal of medicine* **357**(15)**:** 1559-1561.

Olaussen KA, Soria JC (2010). Validation of ERCC1-XPF immunodetection--letter. *Cancer research* **70**(9)**:** 3851-3852; author reply 3852.

Olivier M, Hollstein M, Hainaut P TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harbor perspectives in biology* **2**(1)**:** a001008.

Ong SE, Mann M (2006). A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature protocols* **1**(6)**:** 2650-2660.

Orelli B, McClendon TB, Tsodikov OV, Ellenberger T, Niedernhofer LJ, Scharer OD (2010). The XPA-binding domain of ERCC1 is required for nucleotide excision repair but not other DNA repair pathways. *The Journal of biological chemistry* **285**(6)**:** 3705-3712.

Orlow I, Park BJ, Mujumdar U, Patel H, Siu-Lau P, Clas BA*, et al.* (2008). DNA damage and repair capacity in patients with lung cancer: prediction of multiple primary tumors. *J Clin Oncol* **26**(21)**:** 3560-3566.

Oxnard GR, Binder A, Janne PA (2013). New targetable oncogenes in non-small-cell lung cancer. *J Clin Oncol* **31**(8)**:** 1097-1104.

Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S*, et al.* (2004). EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science (New York, N.Y* **304**(5676)**:** 1497-1500.

Paik S, Tang G, Shak S, Kim C, Baker J, Kim W*, et al.* (2006). Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* **24**(23)**:** 3726-3734.

Pao W, Girard N (2011). New driver mutations in non-small-cell lung cancer. *The lancet oncology* **12**(2)**:** 175-180.

Pathania S, Nguyen J, Hill SJ, Scully R, Adelmant GO, Marto JA*, et al.* (2011). BRCA1 is required for postreplication repair after UV-induced DNA damage. *Molecular cell* **44**(2)**:** 235-251.

Paul I, Savage KI, Blayney JK, Lamers E, Gately K, Kerr K*, et al.* (2011). PARP inhibition induces BAX/BAK-independent synthetic lethality of BRCA1-deficient non-small cell lung cancer. *The Journal of pathology* **224**(4)**:** 564-574.

Paz-Ares L, de Marinis F, Dediu M, Thomas M, Pujol JL, Bidoli P*, et al.* (2012). Maintenance therapy with pemetrexed plus best supportive care versus placebo plus best supportive care after induction therapy with pemetrexed plus cisplatin for advanced non-squamous non-small-cell lung cancer (PARAMOUNT): a double-blind, phase 3, randomised controlled trial. *The lancet oncology* **13**(3)**:** 247-255.

Pears CJ, Couto CA, Wang HY, Borer C, Kiely R, Lakin ND (2012). The role of ADP-ribosylation in regulating DNA double-strand break repair. *Cell cycle (Georgetown, Tex* **11**(1)**:** 48-56.

Peasland A, Wang LZ, Rowling E, Kyle S, Chen T, Hopkins A*, et al.* (2011). Identification and evaluation of a potent novel ATR inhibitor, NU6027, in breast and ovarian cancer cell lines. *British journal of cancer* **105**(3)**:** 372-381.

Peltomaki P (2003). Role of DNA mismatch repair defects in the pathogenesis of human cancer. *J Clin Oncol* **21**(6)**:** 1174-1179.

Pena-Diaz J, Bregenhorn S, Ghodgaonkar M, Follonier C, Artola-Boran M, Castor D*, et al.* (2012). Noncanonical mismatch repair as a source of genomic instability in human cells. *Molecular cell* **47**(5)**:** 669-680.

Petesch SJ, Lis JT (2008). Rapid, transcription-independent loss of nucleosomes over a large chromatin domain at Hsp70 loci. *Cell* **134**(1)**:** 74-84.

Pettitt SJ, Rehman FL, Bajrami I, Brough R, Wallberg F, Kozarewa I*, et al.* (2013). A genetic screen using the PiggyBac transposon in haploid cells identifies Parp1 as a mediator of olaparib toxicity. *PloS one* **8**(4)**:** e61520.

Pfeifer GP (2010). Environmental exposures and mutational patterns of cancer genomes. *Genome medicine* **2**(8)**:** 54.

Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, Hainaut P (2002). Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene* **21**(48)**:** 7435-7451.

Pfeifer GP, Hainaut P (2003). On the origin of G --> T transversions in lung cancer. *Mutation research* **526**(1-2)**:** 39-43.

Pfeifer GP, You YH, Besaratinia A (2005). Mutations induced by ultraviolet light. *Mutation research* **571**(1-2)**:** 19-31.

Postel-Vinay S, Vanhecke E, Olaussen KA, Lord CJ, Ashworth A, Soria JC (2012). The potential of exploiting DNA-repair defects for optimizing lung cancer treatment. *Nature reviews. Clinical oncology* **9**(3)**:** 144-155.

Powell C, Mikropoulos C, Kaye SB, Nutting CM, Bhide SA, Newbold K*, et al.* (2010). Pre-clinical and clinical evaluation of PARP inhibitors as tumour-specific radiosensitisers. *Cancer treatment reviews* **36**(7)**:** 566-575.

Qu YY, Hu SL, Xu XY, Wang RZ, Yu HY, Xu JY*, et al.* (2013). Nimotuzumab enhances the radiosensitivity of cancer cells in vitro by inhibiting radiation-induced DNA damage repair. *PloS one* **8**(8)**:** e70727.

Quinn JE, Kennedy RD, Mullan PB, Gilmore PM, Carty M, Johnston PG*, et al.* (2003). BRCA1 functions as a differential modulator of chemotherapy-induced apoptosis. *Cancer research* **63**(19)**:** 6221-6228.

Rageul J, Fremin C, Ezan F, Baffet G, Langouet S (2011). The knock-down of ERCC1 but not of XPF causes multinucleation. *DNA repair* **10**(9)**:** 978-990.

Rahman L, Voeller D, Rahman M, Lipkowitz S, Allegra C, Barrett JC*, et al.* (2004). Thymidylate synthase as an oncogene: a novel role for an essential DNA synthesis enzyme. *Cancer cell* **5**(4)**:** 341-351.

Rajan A, Carter CA, Kelly RJ, Gutierrez M, Kummar S, Szabo E*, et al.* (2012). A phase I combination study of olaparib with cisplatin and gemcitabine in adults with solid tumors. *Clin Cancer Res* **18**(8)**:** 2344-2351.

Rastogi RP, Richa, Kumar A, Tyagi MB, Sinha RP (2010). Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *Journal of nucleic acids* **2010:** 592980.

Reck M, Heigener DF, Mok T, Soria JC, Rabe KF (2013). Management of non-small-cell lung cancer: recent developments. *Lancet* **382**(9893)**:** 709-719.

Reck M, von Pawel J, Zatloukal P, Ramlau R, Gorbounova V, Hirsh V*, et al.* (2010). Overall survival with cisplatin-gemcitabine and bevacizumab or placebo as first-line therapy for nonsquamous non-small-cell lung cancer: results from a randomised phase III trial (AVAiL). *Ann Oncol* **21**(9)**:** 1804-1809.

Rehman FL, Lord CJ, Ashworth A (2010). Synthetic lethal approaches to breast cancer therapy. *Nature reviews. Clinical oncology* **7**(12)**:** 718-724.

Roberts PJ, Stinchcombe TE (2013). KRAS mutation: should we test for it, and does it matter? *J Clin Oncol* **31**(8)**:** 1112-1121.

Rodin SN, Rodin AS (2002). On the origin of p53 G:C --> T:A transversions in lung cancers. *Mutation research* **508**(1-2)**:** 1-19.

Rodon J, Dienstmann R, Serra V, Tabernero J (2013). Development of PI3K inhibitors: lessons learned from early clinical trials. *Nature reviews. Clinical oncology* **10**(3)**:** 143-153.

Rosell R, Bivona TG, Karachaliou N (2013). Genetics and biomarkers in personalisation of lung cancer treatment. *Lancet* **382**(9893)**:** 720-731.

Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E*, et al.* (2012). Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. *The lancet oncology* **13**(3)**:** 239-246.

Rosell R, Danenberg KD, Alberola V, Bepler G, Sanchez JJ, Camps C*, et al.* (2004a). Ribonucleotide reductase messenger RNA expression and survival in gemcitabine/cisplatin-treated advanced non-small cell lung cancer patients. *Clin Cancer Res* **10**(4)**:** 1318-1325.

Rosell R, Felip E, Taron M, Majo J, Mendez P, Sanchez-Ronco M*, et al.* (2004b). Gene expression as a predictive marker of outcome in stage IIB-IIIA-IIIB non-small cell lung cancer after induction gemcitabine-based chemotherapy followed by resectional surgery. *Clin Cancer Res* **10**(12 Pt 2)**:** 4215s-4219s.

Rosell R, Perez-Roca L, Sanchez JJ, Cobo M, Moran T, Chaib I*, et al.* (2009). Customized treatment in non-small-cell lung cancer based on EGFR mutations and BRCA1 mRNA expression. *PloS one* **4**(5)**:** e5133.

Rosell R, Scagliotti G, Danenberg KD, Lord RV, Bepler G, Novello S*, et al.* (2003). Transcripts in pretreatment biopsies from a three-arm randomized trial in metastatic non-small-cell lung cancer. *Oncogene* **22**(23)**:** 3548-3553.

Rosell R, Skrzypski M, Jassem E, Taron M, Bartolucci R, Sanchez JJ*, et al.* (2007). BRCA1: a novel prognostic factor in resected non-small-cell lung cancer. *PloS one* **2**(11)**:** e1129.

Rouleau M, Patel A, Hendzel MJ, Kaufmann SH, Poirier GG (2010). PARP inhibition: PARP1 and beyond. *Nature reviews* **10**(4)**:** 293-301.

Sage E, Drobetsky EA, Moustacchi E (1993). 8-Methoxypsoralen induced mutations are highly targeted at crosslinkable sites of photoaddition on the non-transcribed strand of a mammalian chromosomal gene. *The EMBO journal* **12**(2)**:** 397-402.

Sakumi K, Furuichi M, Tsuzuki T, Kakuma T, Kawabata S, Maki H*, et al.* (1993). Cloning and expression of cDNA for a human enzyme that hydrolyzes 8-oxo-dGTP, a mutagenic substrate for DNA synthesis. *The Journal of biological chemistry* **268**(31)**:** 23524-23530.

Samol J, Ranson M, Scott E, Macpherson E, Carmichael J, Thomas A*, et al.* (2012). Safety and tolerability of the poly(ADP-ribose) polymerase (PARP) inhibitor, olaparib

(AZD2281) in combination with topotecan for the treatment of patients with advanced solid tumors: a phase I study. *Investigational new drugs* **30**(4)**:** 1493-1500.

Sanchez-Cespedes M (2007). A role for LKB1 gene in human cancer beyond the Peutz-Jeghers syndrome. *Oncogene* **26**(57)**:** 7825-7832.

Sandler A, Gray R, Perry MC, Brahmer J, Schiller JH, Dowlati A*, et al.* (2006). Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. *The New England journal of medicine* **355**(24)**:** 2542-2550.

Sandler AB (2006). Advanced non-small-cell lung cancer: new data, therapy choices, and challenging decisions. *Oncology (Williston Park, N.Y* **20**(6)**:** 626-628.

Sargent RG, Rolig RL, Kilburn AE, Adair GM, Wilson JH, Nairn RS (1997). Recombination-dependent deletion formation in mammalian cells deficient in the nucleotide excision repair gene ERCC1. *Proceedings of the National Academy of Sciences of the United States of America* **94**(24)**:** 13122-13127.

Scagliotti GV, Parikh P, von Pawel J, Biesma B, Vansteenkiste J, Manegold C*, et al.* (2008). Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. *J Clin Oncol* **26**(21)**:** 3543-3551.

Scartozzi M, Franciosi V, Campanini N, Benedetti G, Barbieri F, Rossi G*, et al.* (2006). Mismatch repair system (MMR) status correlates with response and survival in non-small cell lung cancer (NSCLC) patients. *Lung cancer (Amsterdam, Netherlands)* **53**(1)**:** 103-109.

Schiewer MJ, Goodwin JF, Han S, Brenner JC, Augello MA, Dean JL*, et al.* (2012). Dual roles of PARP-1 promote cancer growth and progression. *Cancer discovery* **2**(12)**:** 1134-1149.

Schiller JH, Harrington D, Belani CP, Langer C, Sandler A, Krook J*, et al.* (2002). Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. *The New England journal of medicine* **346**(2)**:** 92-98.

Schlabach MR, Luo J, Solimini NL, Hu G, Xu Q, Li MZ*, et al.* (2008). Cancer proliferation gene discovery through functional genomics. *Science (New York, N.Y* **319**(5863)**:** 620-624.

Schoppee Bortz PD, Wamhoff BR (2011). Chromatin immunoprecipitation (ChIP): revisiting the efficacy of sample preparation, sonication, quantification of sheared DNA, and analysis via PCR. *PloS one* **6**(10)**:** e26015.

Schreiber V, Dantzer F, Ame JC, de Murcia G (2006). Poly(ADP-ribose): novel functions for an old molecule. *Nature reviews. Molecular cell biology* **7**(7)**:** 517-528.

Shaheen M, Allen C, Nickoloff JA, Hromas R (2011). Synthetic lethality: exploiting the addiction of cancer to DNA repair. *Blood* **117**(23)**:** 6074-6082.

Sharma S, Stumpo DJ, Balajee AS, Bock CB, Lansdorp PM, Brosh RM, Jr.*, et al.* (2007). RECQL, a member of the RecQ family of DNA helicases, suppresses chromosomal instability. *Molecular and cellular biology* **27**(5)**:** 1784-1794.

Shaw AT, Solomon B (2011a). Targeting anaplastic lymphoma kinase in lung cancer. *Clin Cancer Res* **17**(8)**:** 2081-2086.

Shaw AT, Yeap BY, Solomon BJ, Riely GJ, Gainor J, Engelman JA*, et al.* (2011b). Effect of crizotinib on overall survival in patients with advanced non-small-cell lung cancer harbouring ALK gene rearrangement: a retrospective analysis. *The lancet oncology* **12**(11)**:** 1004-1012.

Shiraishi K, Kohno T, Tanai C, Goto Y, Kuchiba A, Yamamoto S*, et al.* Association of DNA repair gene polymorphisms with response to platinum-based doublet chemotherapy in patients with non-small-cell lung cancer. *J Clin Oncol* **28**(33)**:** 4945-4952.

Shrivastav M, De Haro LP, Nickoloff JA (2008). Regulation of DNA double-strand break repair pathway choice. *Cell research* **18**(1)**:** 134-147.

Sijbers AM, van der Spek PJ, Odijk H, van den Berg J, van Duin M, Westerveld A*, et al.* (1996). Mutational analysis of the human nucleotide excision repair gene ERCC1. *Nucleic acids research* **24**(17)**:** 3370-3380.

Simon G, Sharma A, Li X, Hazelton T, Walsh F, Williams C*, et al.* (2007). Feasibility and efficacy of molecular analysis-directed individualized therapy in advanced non-small-cell lung cancer. *J Clin Oncol* **25**(19)**:** 2741-2746.

Simon GR, Sharma S, Cantor A, Smith P, Bepler G (2005). ERCC1 expression is a predictor of survival in resected patients with non-small cell lung cancer. *Chest* **127**(3)**:** 978-983.

Slade D, Dunstan MS, Barkauskaite E, Weston R, Lafite P, Dixon N*, et al.* (2011). The structure and catalytic mechanism of a poly(ADP-ribose) glycohydrolase. *Nature* **477**(7366)**:** 616-620.

Soria JC, Mauguen A, Reck M, Sandler AB, Saijo N, Johnson DH*, et al.* (2013). Systematic review and meta-analysis of randomised, phase II/III trials adding bevacizumab to platinum-based chemotherapy as first-line treatment in patients with advanced non-small-cell lung cancer. *Ann Oncol* **24**(1)**:** 20-30.

Souglakos J, Boukovinas I, Taron M, Mendez P, Mavroudis D, Tripaki M*, et al.* (2008). Ribonucleotide reductase subunits M1 and M2 mRNA expression levels and clinical outcome of lung adenocarcinoma patients treated with docetaxel/gemcitabine. *British journal of cancer* **98**(10)**:** 1710-1715.

Span M, Moerkerk PT, De Goeij AF, Arends JW (1996). A detailed analysis of K-ras point mutations in relation to tumor progression and survival in colorectal cancer patients. *International journal of cancer* **69**(3)**:** 241-245.

Steffensen KD, Waldstrom M, Jakobsen A (2009). The relationship of platinum resistance and ERCC1 protein expression in epithelial ovarian cancer. *Int J Gynecol Cancer* **19**(5)**:** 820-825.

Stephens PJ, McBride DJ, Lin ML, Varela I, Pleasance ED, Simpson JT*, et al.* (2009). Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462**(7276)**:** 1005-1010.

Stordal B, Davey R (2009). A systematic review of genes involved in the inverse resistance relationship between cisplatin and paclitaxel chemotherapy: role of BRCA1. *Current cancer drug targets* **9**(3)**:** 354-365.

Stratton MR, Campbell PJ, Futreal PA (2009). The cancer genome. *Nature* **458**(7239)**:** 719-724.

Strauss GM, Herndon JE, 2nd, Maddaus MA, Johnstone DW, Johnson EA, Harpole DH*, et al.* (2008). Adjuvant paclitaxel plus carboplatin compared with observation in stage IB non-small-cell lung cancer: CALGB 9633 with the Cancer and Leukemia Group B, Radiation Therapy Oncology Group, and North Central Cancer Treatment Group Study Groups. *J Clin Oncol* **26**(31)**:** 5043-5051.

Su Y, Orelli B, Madireddy A, Niedernhofer LJ, Scharer OD (2012). Multiple DNA binding domains mediate the function of the ERCC1-XPF protein in nucleotide excision repair. *The Journal of biological chemistry* **287**(26)**:** 21846-21855.

Su Z, Dias-Santagata D, Duke M, Hutchinson K, Lin YL, Borger DR*, et al.* (2011). A platform for rapid detection of multiple oncogenic mutations with relevance to targeted therapy in non-small-cell lung cancer. *J Mol Diagn* **13**(1)**:** 74-84.

Subramanian J, Govindan R (2007). Lung cancer in never smokers: a review. *J Clin Oncol* **25**(5)**:** 561-570.

Subramanian J, Govindan R (2008). Molecular genetics of lung cancer in people who have never smoked. *The lancet oncology* **9**(7)**:** 676-682.

Sun JM, Han J, Ahn JS, Park K, Ahn MJ (2011). Significance of thymidylate synthase and thyroid transcription factor 1 expression in patients with nonsquamous non-small cell lung cancer treated with pemetrexed-based chemotherapy. *J Thorac Oncol* **6**(8)**:** 1392-1399.

Sun Y, Li T, Ma K, Tian Z, Zhu Y, Chen F*, et al.* (2009). The impacts of ERCC1 gene exon VIII alternative splicing on cisplatin-resistance in ovarian cancer cells. *Cancer investigation* **27**(9)**:** 891-897.

Svendsen JM, Smogorzewska A, Sowa ME, O'Connell BC, Gygi SP, Elledge SJ*, et al.* (2009). Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair. *Cell* **138**(1)**:** 63-77.

Takagi Y, Setoyama D, Ito R, Kamiya H, Yamagata Y, Sekiguchi M (2012). Human MTH3 (NUDT18) protein hydrolyzes oxidized forms of guanosine and deoxyguanosine diphosphates: comparison with MTH1 and MTH2. *The Journal of biological chemistry* **287**(25)**:** 21541-21549.

Takeuchi K, Soda M, Togashi Y, Suzuki R, Sakata S, Hatano S*, et al.* (2012). RET, ROS1 and ALK fusions in lung cancer. *Nature medicine* **18**(3)**:** 378-381.

Taron M, Rosell R, Felip E, Mendez P, Souglakos J, Ronco MS*, et al.* (2004). BRCA1 mRNA expression levels as an indicator of chemoresistance in lung cancer. *Human molecular genetics* **13**(20)**:** 2443-2449.

Tomasetti C, Vogelstein B, Parmigiani G (2013). Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation.

Proceedings of the National Academy of Sciences of the United States of America **110**(6)**:** 1999-2004.

Tomita-Mitchell A, Kat AG, Marcelino LA, Li-Sucholeiki XC, Goodluck-Griffith J, Thilly WG (2000). Mismatch repair deficient human cells: spontaneous and MNNG-induced mutational spectra in the HPRT gene. *Mutation research* **450**(1-2)**:** 125-138.

Tooker P, Yen WC, Ng SC, Negro-Vilar A, Hermann TW (2007). Bexarotene (LGD1069, Targretin), a selective retinoid X receptor agonist, prevents and reverses gemcitabine resistance in NSCLC cells by modulating gene amplification. *Cancer research* **67**(9)**:** 4425-4433.

Tripsianes K, Folkers G, Ab E, Das D, Odijk H, Jaspers NG*, et al.* (2005). The structure of the human ERCC1/XPF interaction domains reveals a complementary role for the two proteins in nucleotide excision repair. *Structure* **13**(12)**:** 1849-1858.

Tripsianes K, Folkers GE, Zheng C, Das D, Grinstead JS, Kaptein R*, et al.* (2007). Analysis of the XPA and ssDNA-binding surfaces on the central domain of human ERCC1 reveals evidence for subfunctionalization. *Nucleic acids research* **35**(17)**:** 5789-5798.

Tsodikov OV, Enzlin JH, Scharer OD, Ellenberger T (2005). Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. *Proceedings of the National Academy of Sciences of the United States of America* **102**(32)**:** 11236-11241.

Tsodikov OV, Ivanov D, Orelli B, Staresincic L, Shoshani I, Oberman R*, et al.* (2007). Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA. *The EMBO journal* **26**(22)**:** 4768-4776.

Turner N, Tutt A, Ashworth A (2004). Hallmarks of 'BRCAness' in sporadic cancers. *Nature reviews* **4**(10)**:** 814-819.

Turner NC, Lord CJ, Iorns E, Brough R, Swift S, Elliott R*, et al.* (2008). A synthetic lethal siRNA screen identifying genes mediating sensitivity to a PARP inhibitor. *The EMBO journal* **27**(9)**:** 1368-1377.

Tutt A, Robson M, Garber JE, Domchek SM, Audeh MW, Weitzel JN*, et al.* (2010). Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and advanced breast cancer: a proof-of-concept trial. *Lancet* **376**(9737)**:** 235-244.

Tutt AN, Lord CJ, McCabe N, Farmer H, Turner N, Martin NM*, et al.* (2005). Exploiting the DNA repair defect in BRCA mutant cells in the design of new therapeutic strategies for cancer. *Cold Spring Harbor symposia on quantitative biology* **70:** 139-148.

Varella-Garcia M (2010). Chromosomal and genomic changes in lung cancer. *Cell Adh Migr* **4**(1)**:** 100-106.

Vilmar AC, Sorensen JB (2011). Customising chemotherapy in advanced nonsmall cell lung cancer: daily practice and perspectives. *Eur Respir Rev* **20**(119)**:** 45-52.

Vyas S, Chesarone-Cataldo M, Todorova T, Huang YH, Chang P (2013). A systematic analysis of the PARP protein family identifies new functions critical for cell physiology. *Nature communications* **4:** 2240.

Wang B, Chu D, Feng Y, Xin Y, Myers P, Post L*, et al.* (2011a). Novel PARP inhibitors with potent antitumor activity as single-agent and combination therapies. *Molecular cancer therapeutics* **8**(12)**:** Supplement 1.

Wang L, Wei J, Qian X, Yin H, Zhao Y, Yu L*, et al.* (2008). ERCC1 and BRCA1 mRNA expression levels in metastatic malignant effusions is associated with chemosensitivity to cisplatin and/or docetaxel. *BMC cancer* **8:** 97.

Wang LE, Yin M, Dong Q, Stewart DJ, Merriman KW, Amos CI*, et al.* (2011b). 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)**:** 4121-4128.

Wang M, Wu W, Wu W, Rosidi B, Zhang L, Wang H*, et al.* (2006). PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic acids research* **34**(21)**:** 6170-6182.

Wang YC, Lu YP, Tseng RC, Lin RK, Chang JW, Chen JT*, et al.* (2003). Inactivation of hMLH1 and hMSH2 by promoter methylation in primary non-small cell lung tumors and matched sputum samples. *The Journal of clinical investigation* **111**(6)**:** 887-895.

Ward TA, Dudasova Z, Sarkar S, Bhide MR, Vlasakova D, Chovanec M*, et al.* (2012). Components of a Fanconi-like pathway control Pso2-independent DNA interstrand crosslink repair in yeast. *PLoS genetics* **8**(8)**:** e1002884.

Weeda G, Donker I, de Wit J, Morreau H, Janssens R, Vissers CJ*, et al.* (1997). Disruption of mouse ERCC1 results in a novel repair syndrome with growth failure, nuclear abnormalities and senescence. *Curr Biol* **7**(6)**:** 427-439.

Weinstein IB (2002). Cancer. Addiction to oncogenes--the Achilles heal of cancer. *Science (New York, N.Y* **297**(5578)**:** 63-64.

Weiss J, Sos ML, Seidel D, Peifer M, Zander T, Heuckmann JM*, et al.* (2010). Frequent and focal FGFR1 amplification associates with therapeutically tractable FGFR1 dependency in squamous cell lung cancer. *Science translational medicine* **2**(62)**:** 62ra93.

Westerveld A, Hoeijmakers JH, van Duin M, de Wit J, Odijk H, Pastink A*, et al.* (1984). Molecular cloning of a human DNA repair gene. *Nature* **310**(5976)**:** 425-429.

Wikenheiser-Brokamp KA (2006). Retinoblastoma regulatory pathway in lung cancer. *Curr Mol Med* **6**(7)**:** 783-793.

Willers H, Taghian AG, Luo CM, Treszezamsky A, Sgroi DC, Powell SN (2009). Utility of DNA repair protein foci for the detection of putative BRCA1 pathway defects in breast cancer biopsies. *Mol Cancer Res* **7**(8)**:** 1304-1309.

Willingham AT, Deveraux QL, Hampton GM, Aza-Blanc P (2004). RNAi and HTS: exploring cancer by systematic loss-of-function. *Oncogene* **23**(51)**:** 8392-8400.

Winton T, Livingston R, Johnson D, Rigas J, Johnston M, Butts C*, et al.* (2005). Vinorelbine plus cisplatin vs. observation in resected non-small-cell lung cancer. *The New England journal of medicine* **352**(25)**:** 2589-2597.

Yan J, Kim YS, Yang XP, Li LP, Liao G, Xia F*, et al.* (2007). The ubiquitin-interacting motif containing protein RAP80 interacts with BRCA1 and functions in DNA damage repair response. *Cancer research* **67**(14)**:** 6647-6656.

Yano T, Haro A, Shikada Y, Maruyama R, Maehara Y (2011). Non-small cell lung cancer in never smokers as a representative 'non-smoking-associated lung cancer': epidemiology and clinical features. *International journal of clinical oncology* **16**(4)**:** 287-293.

Yin M, Yan J, Martinez-Balibrea E, Graziano F, Lenz HJ, Kim HJ*, et al.* (2011). ERCC1 and ERCC2 polymorphisms predict clinical outcomes of oxaliplatin-based chemotherapies in gastric and colorectal cancer: a systemic review and meta-analysis. *Clin Cancer Res* **17**(6)**:** 1632-1640.

Yuan G, Bin JC, McKay DJ, Snyder FF (1999). Cloning and characterization of human guanine deaminase. Purification and partial amino acid sequence of the mouse protein. *The Journal of biological chemistry* **274**(12)**:** 8175-8180.

Zhang N, Liu X, Li L, Legerski R (2007). Double-strand breaks induce homologous recombinational repair of interstrand cross-links via cooperation of MSH2, ERCC1-XPF, REV3, and the Fanconi anemia pathway. *DNA repair* **6**(11)**:** 1670-1678.

Zhang YW, Regairaz M, Seiler JA, Agama KK, Doroshow JH, Pommier Y (2011). Poly(ADP-ribose) polymerase and XPF-ERCC1 participate in distinct pathways for the repair of topoisomerase I-induced DNA damage in mammalian cells. *Nucleic acids research* **39**(9)**:** 3607-3620.

Zhao Y, Thomas HD, Batey MA, Cowell IG, Richardson CJ, Griffin RJ*, et al.* (2006). Preclinical evaluation of a potent novel DNA-dependent protein kinase inhibitor NU7441. *Cancer research* **66**(10)**:** 5354-5362.

Zheng H, Wang X, Legerski RJ, Glazer PM, Li L (2006). Repair of DNA interstrand cross-links: interactions between homology-dependent and homology-independent pathways. *DNA repair* **5**(5)**:** 566-574.

Zheng Z, Chen T, Li X, Haura E, Sharma A, Bepler G (2007). DNA synthesis and repair genes RRM1 and ERCC1 in lung cancer. *The New England journal of medicine* **356**(8)**:** 800-808.

Zhou W, Gurubhagavatula S, Liu G, Park S, Neuberg DS, Wain JC*, et al.* (2004). Excision repair cross-complementation group 1 polymorphism predicts overall survival in advanced non-small cell lung cancer patients treated with platinum-based chemotherapy. *Clin Cancer Res* **10**(15)**:** 4939-4943.

Zhu XD, Niedernhofer L, Kuster B, Mann M, Hoeijmakers JH, de Lange T (2003). ERCC1/XPF removes the 3' overhang from uncapped telomeres and represses formation of telomeric DNA-containing double minute chromosomes. *Molecular cell* **12**(6)**:** 1489-1498.

Supplementary materials

Supplementary Figure A1

ERCC1 Human DNA Sequence

- Primers used for determination of the mutant sequence by TOPO Cloning® are in bold and underlined
- ZFN binding site are in bold and red
- ZFN cutting site in lowercase and red

CCGGAAGTGCTGCGAGCCCTGGGCCACGCTGGCCGTGCTGGCAGTGGGCCGCCTCGA
TCCCTCTGCAGTCTTTCCCTTGAGGCTCCAAGACCAGCAGGTGAGGCCTCGCGGCGCT
GAAACCGTGAGGCCCGGACCACAGGTGCGGGAGGCGGAGACTGCGGGTGGAGATTGG
CGCCGCGGAAGCCAATCATTGCCGAGTCTGAGAGATGGACAAGGCCAGGCGTGGGGA
GGGCGTCCAGATGCTAGCCTCGGGGGCCGGACGAACGGAAGGCGGGGGATGGTGGG
GACGGAGCCAATAGAATCCGGTGGGGGCGAGGGGCGGAGCGATGGGACTTGTGGACC
TGTAAGGGGCGGGGCGAGCCGAAGGTGGAGGTCAAAGGGGCGTGGCGTTACAGAGCC
TCTAGCGCTGGGTGTTGGGGACCTGACGCTATGGAGCTCTCGGAGTTTTGTGGGGGAC
GGCTGTGAGTGGGGGGTTCCTGCTGCGGGATGAGAACGTAGACGCCAGTGGCTCACTC
GCTCCTGGCACCTTCCCTTTCAGGCTCCAGATGGACCCTGGGAAGGACAAAGAGGGGG
TGCCCCAGCCCTCAGGGCCGCCAGCAAGGAAGAAATTTGTGATACCCCTCGACGAGGA
TGAGGTCCCTCCTGGAGTGGTAGGACAAGGAGATGCGGGGCCCCTGGGAGGCTGGGG
GCTGTTAGGACGAAGAGGATAGGATGGGGCCTGTGGGACCAGGGTGTGGGTTAGTGGA
TTTGGGGGCCACGGACGACTTGGGGAAACAGTCCTTGGTCCTCCCCAGGTCCCAGTTT
CCCCATCTGTGAAATGGATGGGTGGTTCTAAGAGAGGGCTAAGGCAGAGGCCAGACAC
TGGCATTCAGCAGGTAGCCCCTGATGTGTTTTACGAGGCCAGCACCCTGATTTTTCAAG
ACATGAATTTATTGCTCGTGTTTAAGAATCGCCAGGTTTTGCAATCTTAAAAAAATCTACA
CTTTAGACATATTTTGGTAAATGGAAGGACCTGGCAATATTGAAATCTCCTTCCCAAGGG
AAAACAATCGACGGGTGATGAGCATTGGCTATGAAGAAAAGGAAAGTGGGCTAAGGATG
TAGGGCATGATGGGGGAGAGGACCTTCCTACTCTATTTATTTATTTATTTATTTATTTTTTT
GAGACAGAGTCTCGCTCTTTTGCCCAGGCTGGAGTGCACTGGCACCATCTCGGCTCACT
GCAACCTCTGCCTCCTGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTAACTGGG
ATTACAGGAGCGCGCCTGGCTAATTTTTGTATTTTTAGTAGAGACGGGTTTCACCATGTT
GGTCAGGCTGGTCTCGAACCTCAAGTGATCCGCCCACCTCGGCCTCCCAAAGTGCTGG
GATGTACAGGCATGAGCCACCATGCCTGGCCCCTTCCTACTTTAGATAGGAAGGTCAAG
GGAGGGCCTCTCTGAGGAGGTGACATCTTGGTAGAGACGTGAAAGAGGTGAGGGAGTG
AGCCATGGCGACATCTGGGGTAGAGGGAACAGCCAGTGCCAAAGCCCTGAGGTCAGAG
AGAGCCTGCCGTGTTCAAGGGCAGAACGAAGTCTGGTGGGGCTGGAGTAGAGAAAGCA
AGGAGAAGTGGTAGAAGGTGTAGTTTCTGAGGCAAGGAGCAGGGACAGATCTAGAGCC
TTGTGGGTGATGGTGACAGCTTTGGCTTTCACTCTGAGGCAGGAGCCACCCACCAGCTC
TGAGCAGAGGAGGGACATGAGGACTTAGTTAGGGGACAGACTTCCAGGGGGCGAGGG
TGTGGGCAAGGAGCTGGTGAGGAAGTGACTTCAGTGGTCCAAGGGTGAAACGCTGGTG
GCTGGGACCAGAGTGGAGGCTGAGGAGTGGGGGAGAAGTGGCTGAATTCAGATTTTGA
GCTCTCAGCCTTCTGGAAGGGACATTAGGAGGAAGGAGGATGGGCAGCCCTGAAGAGG
GGAGAGGGGTCAAGTGCTGTGGCTCACACCTGTAATCCCAGGACTTTGGGAGGCCGAG
GCGGGCGGATCACCTGAGGTCAGGAGTTCGAGATCAGCCTGGACAACATGTTGAAACC
CTGTCTCTACTAAAAATACAAAAATTAGCCAGCGTGGTGGTACACGCCTGTAGTCCCAGC
TATTTGGGAGGCTGAGACAGGAGAATCGCTTGAACCTGGGGGGCGGAGGTTGCAGTGA
GCCGAGATCACGCCACTGCACTCCAGCCTGGGTGACAGAGCAAGACTCCATCTCAAAAA
AAAAAAAAAAAAATTAGCTGGGCGTGGTGGTGGGCACCTGTAATCCCAGCTACTTGGGA
AGCTGAGGCAGAATTGCTTGAACCCGGGAGGTGGAGGTTGCAATGAGCCAAGATAGTG
CCACTACACTCCAGTCTGGGTGACAGAGCAAGACTGTCTCAAAAAAAGAAAAAGAAAGG
GGAGAGGAACTCACAGGGCCTCAGATGTCCTCTGCTCACCCCAACCAGTCCCCTGCAA
ACTCCCTTTCTCCCCACAGGCCAAGCCCTTATTCCGATCTACACAGAGCCTTCCCACTGT
GGACACCTCGGCCCAGGCGGCCCCTCAGACCTACGCCGAATATGCCATCTCACAGCCT
CTGGAAGGGGCTGGGGCCACGTGCCCCACAGGGTCAGAGCCCCTGGCAGGAGAGACG

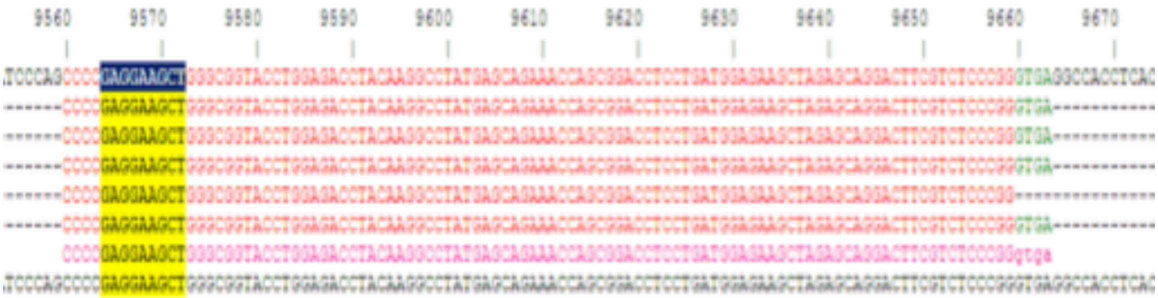
CCCAACCAGGCCCTGAAACCCGGGGCAAAATCCAACAGCATCATTGTGAGCCCTCGGC
AGGTGAGGAGGGAGACGGAGAAGTGAGGCCTTGAGGTTTTCAGTGGGAAACGCTGTTC
TAGGGATGACTCCAGTGCAAGGAGGTCTGAGTTCCAGCCACTTTGTTCTCCCCATTCATT
CATTCATTCATTCATTCATTCGACAGCAGTGTGTGGGTTGAGAGTGCCCCAGCTCTACTT
GCTGGGGCTGTGGCTGTATTGGTGTAGCTCTGTTTCCTCACCCCAGGTTTGTTAATTTAT
TCAAAATGTACTGAATGAGGCTGGGTGTGGTGGCTCACACCTGTAATCCCAGCACTTTG
GAGAGGCTGAGGCAGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGTCTGGCGAAT
ATGGTGACACCCCATCTCTACTAAAAACACAAAAAGTTGCTGGGCGTGAGGGCAGGAAG
TGTTCAGGACCACAGGACACGCAGACTGGGGCTGGAGGGAGAGGGCTGAGCTGGAGA
CAGACCCGGGGACCCTTTAGGAAAGGCCTGGTGCCAGTGGGCATGGAGGGGAGGAGC
CTGGATCAGAGGATCAGGGACTGTCCAGGGTTAGGAGGAGAGAGAAGCTGGAAAAGAC
CCTGCCACAGGTGGGAGGAGGGAGGAGGTGTGGGAAGAGGTGCGAGGAGGCAGGAG
GTGTGGGAAGAGGTGGGAGGAGGGAGGAGGTGCAAGAAGAGGTGGAGGAGGGCCCT
GTGGTTATCAAGGGTCATCCCTATTGATGGCTTCTGCCCTTCGTCCCTCCCCAGAGGGG
CAATCCCGTACTGAAGTTCGTGCGCAATGTGCCCTGGGAATTTGGCGACGTAATTCCCG
ACTATGTGCTGGGCCAGAGCACCTGTGCCCTGTTCCTCAGGTGAGCTCTGCGGCGCCA
CCCCAGACTTCAGGAAGGGCACCCCACTGGCCTGGGAGGGTCATGTCCCAGTGTTCTG
GACTGTTCTATGAGAAGCCTCTGGGCATGGTCCTGCAGGTTCTGGGCCTTGGTGGGAC
CCTGTGTTGGAGGCAGGGGGTCTTGGCCTGGAGGTGCCCAGTGCAACAAGAGCTGGA
GCCCGAACTCCTGGGTCTCAGGAAGGAGGGGCTGGGGGCCGGACTCCTGGGTCCGGG
GGAGGAGGGGCTGAGGGCTGGACTCCTGGGGATCTGAGGGAGGAGGGGCTGGGGCC
TAGACTCCTGGGTCTGAGGCAGGAGGGGCTGGGGGCCTGGACTCCTGGGGTCTCAGG
GAGGAGGGGCTGGGGGCCTGGACTCCTGGGGTCTCAGGGAGGAGGGGCTGGGGGCC
TGGACTCCTGGGTCTAAGGGAGGAGGGGCTGGGGCCAGATTCCTGGGTCTAAGGGAG
GAGGGGCTGGGCCCAGACTCCTGGATCTGACGGAGGAGGGACTGGGTCACTCATGTC
CCAGACATCCTTGACTGGAGTTGATGAGAAGCAGGATTTCTCCCATCTCTACCCTAGTTA
AACAACTTTCCCCTCCAACCTAGGGGCCCCTCTGCAGACAGGGCCAGACAGAGGAAGT
GAAACCTCTGAGTTCCTGTGCTCCCTGCAGGCTGTGACAGCTGCAGACTGCTGGGCAG
AGCCAGTCCTAGTTCCCTGGGTCATGAAGCCCTCACCGTAGCCCTCCAAGGCCAGAGC
AGATGGTTTGGCAGTGGGCCCTGCTTCCCACCGCGAGCCATGCAGTGTCCCCGTGCCC
ACACTCTCTAGGGCCCCTGGCACCTGTTCTCCCCTTGAAGCCCAAGTCTCTTCCCCTCTT
CATGAAGAAGCCAGGGGTCTGGTCTTCTAGGTAAGCACAGCCATGTGGCCCTGGCTGC
CAATCCCTGCGGAGCATCTTCTTGGAAGAGTGACCTGGGGCAAGTGACTTACCCTCTGT
GGGCCTCAGGACCCTCTGTCCCCCATGCCGTACTCTGAGCCCTTAGTATTCCAGTGAGA
GGGAAAAGGGCTTTGGGGGTCTTGAAGAGACTGAGACCTCTCAACTCCCGCTCCCTGC
CCCCACCAGCCTCCGCTACCACAACCTGCACCCAGACTACATCCATGGGCGGCTGCAG
AGCCTGGGGAAGAACTTCGCCTTGCGGGTCCTGCTTGTCCAGGTGGATGTGGTAAGCA
GGGGCTGCTCCCTAGCCAGCCTCACCGGGCTTTGAGGTTGTGCAGTGGGCTGTTGAAT
CCCTTCCAAGAGGAAATGCAGCTAAAACCTCAGCGAGGAACGTTTTTCACATCTGAGAT
GACATGGGTCACCAAGACATTATGTTGGCAAGGAGAGGGGAGGGCAAAATTGTCTATTT
CCTCTTTCCCCACATCCCCTTCAATTAGTCCCAGATCTCTGCCACCTAAATATCTCTTCTT
TTTGTTTTGTGTTTGTTTGTTTGTTTGTTTGAGACGGAGTTTCGCTCTTGTTGCCCAGGCT
GGAGTGCAATGACTTGATCTTGGCTCACTGCAGCCTTCACCTCTGGGGTTCAAGCGATT
CTCCTGCCTCAGCCTTCTGAGTAGCTGGGATTATAGGCATGCACCACCACGCCCGGCTG
ATTTTTGTATTTTTAGTAGAGACTGGTTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCC
TGACCTCAGGTGATCCACCCACCTCAGCCTCCCAAAGGCTGGGATTACAGGCATGAGC
CACCGTGCCTGGCCATATCTCTCTCTTTTATTTTTTTAAGACGGAGTCTCCTCTGTCGCC
CAGGCTGGAGTGCAATGGCCAGATCTTAGCCCACTGCAACCTCTGCCTCCTGGGTTCAA
ACGATTCTCCTGCCTCAGCCTCTCGAGCAACTGGGATGACACCATGCCTTGCTAATTTTT
ATTGATTGATTGATTGATTGATTGTTTTGAGACAGAGTCTCGCTCTTTCGCCCAGACTGG
AGTGTAGTGGTGCAATCTTGGCTCACTGCAACCTCCACCTCCTGGATTCAGGTGATTCTC
CTGACTCAGCCTCCTGAATAGCTGGGATTACAGGTGCGTACCACCATGACCGGCTAATT
TTTTGTATTTTTAGTAGAGATGCGGTTTCACCATGTTGGTCAGGCTGGTCTTGAACTTCTG
ATCTCAGGTGATCCTAAAGTGTTGGGATTACAGGTGTGAGCCACCCCCACCAGGTGATT
TTTTTTTTTAATTTTTAGAGACAGGGTCTTGCTATGTTACCCAGTCTGGTCTTGAACTCCT
GGCCTCAAGCATTCCTCTTCCATTGGGTTGGGTGCAGTGGCTCATGCCTGTAATCCTAG
CTCTTTGAGAGGCTGAGGCAGGAGGATCACTTGAGGCCAAGAGTTCGAGACTAGTCTG
GGCAACATAGTGAAGACCCCCATCTCTACAAAATACAAAAAATATGGTGGTGGGCACCT
ATAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCCGGGAGGTGGA
GGTTGCAATGAGCCAAGATCATGTCACTGCACTCCAGCCTGGGCAACAGAGCGAGACT

CCGTCTGAAAAAAATTAAAAAGATCATAAAACTCACATGGCTCCTAAGCACCCTAGGAGA
AAGTTCTGCTCCCTTCCATGACCTATGGGGCCCTGTGTGACCTCCAACCTCTACCCAGTT
CTCCAGGCCCGTCTCTAAGATTCCATGCTCCTCATGGAATTTCACATCCTCTCTCCCGTA
GGGAGCGAATTATAATTCCTCAAGCCCTCTTTCCCCTCCCATCCTTTGCCCATGCTGTTC
TTTATGTCTGGAATTCCCTTTTCTGGTCTCCTTTTCATCCTTTCCTCCTCCTCATCCTCCTT
TGGGTCTCAGCTCTAATGTCCCCTCCTCCAAGAAGCCCTCCCTGATCCCCCAGGCAGGC
TCAGTCCCTCCTCTGGCTCCCACATCGCAGCCCTGGCCACTCCAGATCATCAGTGTCTG
GGGACAGGTCTGTCCTGCTCATCAGACTGTGAGCCCTGGGAGGGCAGAGTCAGGGCTG
TCTCAGCCACTGCTGGGTCCTAGCACTGCTTAGCACGGGCCAGGCACCCAGTAGGAGC
TTAGAGAGTGACAATTGGGTGAGAGGCTCTTGGCATCCTGAGCTACTGTGTTTTCAGGC
TTTAATTCTTTTTTTGTTCTTCTTTTTTTTTTTGAGACAGAGTCTCACCCTGTCGCCCAGGT
TGGAGTGCAGTGGCGCGATCTTGGCTCACTGCAACGTCTGCCTCCCGGGTTCAAGCAG
TTCCCCCACCTCAGCTTCCCTAGTAGCTGGGATTACAGGCACCTGCCACCACGCCTGGC
TAATTTTTGTAATTCCTGGCTTCTAGGTTTCTAATTCTGATTTTCTCCTCCAGAAAGATCCC
CAGCAGGCCCTCAAGGAGCTGGCTAAGATGTGTATCCTGGCCGACTGCACATTGATCCT
CGCCTGGAGGTGAGATGAGGGCTTCCCTGCCTCATTCAGGCTCCACCTGGGATTGGTC
CCTGCTCCCCCATCCCACCCTGCCCAGCCCTTCCCTTCTCCTTCCTGTGTGGGCCTGGT
GCAGGCACTGCCTCTCTCTGGCCTTCAGTTTCCTCATCTGTAAAACAAGGTCTTAATTTTT
TTTTTTTTTTTTTTGATACAGAGCCTCACTCTGTCCCCTAGGCTGAAGTGCAGTGGCACG
ATCTCGGCTCTGCCTTGGGGGTTCAAGCAATTCTGCCTCAGCCTCCCAAGTAGCTGGGA
TTACAGGTGTGCACCACCACGCCCAGCTAATTTTTTTGTATTTTTAGTAGGGACGGGTTT
TGCCATGTTGGCCAGGCTGGTCTTGAAATCCTGACCTCAGGTGATCCTCCCGCCTTACC
CTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACTGTGCCCGGTTGGAAGCTGAACATT
TCATGAGGTTGGCTTCTTGCTGCAGCGGGAAAATAAGCAGTGTTTCTGGTTTTCCTCCTG
TCAAACCCAGTACAAACACTTTGTGTGCTGACTTCCCTTCCAAAATAGCTAACATTTCTTC
AGCAGTGCCAGGCGTGGCTTTTGGCACTTACAGAGCATTCCAGAGTATTCTCTGATCTTC
AAAGTCATACCATCCTCCCCACTTTATTATTATTTTTTTGAGACAGGGTCTTGCTCTTTTG
CTGAGGCTGGAGTGCAGTGGTGCTATCACAGGTTACTGAAGCCTCAACTTGACAGGCTC
AGGCAGTCCTCCTGCCTCAGCCTCCCGAGCAGCTGAGACTAAAGGCACGCGCCACCAC
ACCTGGCTAATTTTTAATTTTTTGTGGAGACAACACCTTGCTATGTTGTCCCGGCTGGTCT
TGAACTCCTGAGCTCAAGTGATCCTCCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAA
GCATGAGCCACTATGCCTGACCACCCCACTTTACAGTGAGGAAACAGGCCCAGAGAGC
AAACCACTGGCCCAGGGTTACCCAGGGAGAAGTGGGGTTGGGCTCCCAAGAGCTGGCC
TGTAGCCAGAAAAGATATCAAGATGCGCCCAGCAGTGGTTACTACCAGGTTGTATTTGG
CGGGGGATGGGCTGGAATTTTGATTTTTAGCAACCACAATTGTTTTTTTTTTTTTTTTTTTT
TTTGAGACGGAGTCTGGCTCTGTCGCCCAGGCTGGAGTGCAGTGGCGTGATCTCGGCT
CATTGCAAACTCCACCTCCCGGGTTCACGCCATTCTCATGCCTCAGCCTCCTGAAAAGC
TGGGACTACAGGCGCCCGCTACCACGCCCGGCTAATTTTTTGTATTTTTAGTAGAGATG
GGGTTTCACCATGTTAGTCGGGATGGTCTCGATCTCCTGACCTGGTGATCTGCTCACCT
CAGCCTCCCAAAGTGCTGGGATTATAGGCGTGAGCCACCACACCCGGCCTTTTTTTTTTT
TTTTTTCTGAAATGGAGTTTCACTCTTGTTGCCCAAGCGGGAGTGCAATGGCGTGATCTT
GGCTCACTGCAACCTCCGCCTCCCGGGTTCAGGCGATTCTCCTGCCTCAGCCTCCTGA
GTAGCTGGGACTACAGGCACGCGCCACCAAGCCCAGCTACTTTTTTGTATTTTTAGTAGA
AACGAGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGATGATCTGC
CTGCCTCATCATCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCGCCCGGCCTAG
TATTGGCAATTCTTATGACTGACCATAATGAAGCAAATCTATAATCTGATACCTCACCTCC
CGGCTGCCCTGTATC**CTGTTATCCCAGCCCCGA**ggaagc**TGGGCGGTACCTGGA**GACCT
ACAAGGCCTATGAGCAGAAACCAGCGGACCTCCTGATGGAGAAGCTAGAGCAGGACTT
CGTCTCCCGGGTGAGGCCACCTCACCTCCCATCCCTGCCTGGGCCTCCCCGCAGCTCC
TGGGTGGTTGAGCTTCAGTTCCAATTTTAAGGCACTGTCCATTCATCCCTGGGCCTCGTC
TTTCCTGGCCCATCCC**ACCGAAACTGCAGCTAGAGG**GGTTTTCCTGAGCACAAATCTGC
TGCTACCCTCCCTAGCTCAAAATCCTCCCATGGCTCCCCAGTACCCCGAGAGAAAGCCC
AGACTTCTCACCAGGCTCAACAAGGGCCCGTCAGGCTCAGAAGCCCCGTGGCCTCCAC
TTCACACTTCACTTCTCCCAGGCCTAGCTCGCTTCTCCAAGCACTTGCCCCTTCTGTTCC
CTCTGCCCAATTACCTGCCTAATTCCTTTTTTTTTTTTTTTGAGATGGAGTCTCACTCTCTTG
CCCAGGCTGGAGTGCAATGACATGATCTCGGCTCACTGCAACCTCTGCCTCCCAGGTTC
AAGTGATTCTCCTGCCTGAGCCTCCTGAGTAGCTGGGACTACAGGTGCCCACAACCATG
CCTGTCTAATTTTTGTATTTTTTAGTAAAGATGGGGTTTCAGTGTGTTGGCCAGGCTGGT
CTTGAACTCCTGGCCTCAGGTGATCCTCCCGCCTCGGCCTCGAAAGTGCTGGGATTATT
AGGCATGAGCCACCACGCCCAGCCTGAGATGACTTTTTAATCAGTTGGGAAAAAAAGCG

TTTATTGAGTGCTTCCTGCACGCCAGGCAGTCTGGGGACACAGCTGTGACCCAGCCCTG
CCCTCCTAGGCTCAGTCTTGCTGGCCTTTCTTCTTCCGACACACTCCTGCCTCCACCCTT
TCCAGGTGACTGAATGTCTGACCACCGTGAAGTCAGTCAACAAAACGGACAGTCAGACC
CTCCTGACCACATTTGGAGTAAGGAATGGCTCCCCTGCCCCATAGGCATTTCTGTCCCC
TTGCCTTTTTTAGAAAGCCTGTCACCCATTTCCTTCCATCTCCCTGCCCACCCCTTGCCC
GTCCTGCCTTCCCACAGGCCCTCATCTCCCCTGGGGAATATCTGAGGCCCCTCCTCAGC
TGGGACAGGGGAGGCTTTTGTGCTCAACTGCCCTGACCCCTCGCTTTCACCTTTCAGTC
TCTGGAACAGCTCATCGCCGCATCAAGAGAAGATCTGGCCTTATGCCCAGGCCTGGGC
CCTCAGAAAGTAAGAGCTCTGGGAAAGAACCCAAGGAGTTGGGGGAAGGAGAGAGCCC
CAAATAAACACAACCTGAGACCCCAAAGTTTTAAGGTGAAAAAAGAACCAAAGACCAGAC
ACAGTGGCTTCCGCCTGTAATCCCAACATTTTGGGAGGCCAAGGCGGGAGGACTGCTT
GAGGCCAGAAGTTGGAGACCAGCCTGGGCAAGTGGACACCTCATTTTTACTAAAAATAA
AAAAAACTAGCTGGGCATGGTGGTGTGAGCTTGCTACGTGCCTGTAGTCTCAGCTACTG
GGGAGACTGAGGTGGGAGGATCGCTTGAGCCCAGGAGGTTGAGGCTGCAGTGAGCTAT
AAGCATGTTACTACGCTCCAGCCTGGGTGACAGAGCGATCCTGTCTCAAAGCAGAAGCA
ATCCAATGAGGAAGAATGAAACTAGATAGCTATCTCTCGCCATATACAAAAATCAAATTAA
AATGGATTACAGACTTAAATCTAAGACCTCAATGCATGAAACGTCTACAAGAAAACATTG
AGAAGCTCTCCAGGACATTGAACTGGGCAGAGATTTCTTGAGTAATGCCCTATAAGCATA
GGAAACTAAAGCAAAATGGACAAATGGGATCACATCAAATTAAAAAGCTTCTAAGCCAGG
TGCGGTGGCTTGGGCCTGTAATCCCAGCACTTTGAGAGGCTGAGGCGGCTGGATCACC
TGAGGTCAGGAGTTTGGGACCAGCCTGGCCAACATGGAGACACCCCATCTTTACTAAAA
ATACAAAAAAGTAGCCAGGTATGGTGGCACATGCCTGTAATCCCAGCTACTTGGAAGGC
TGAGGCAGGACAATCGTTTGAACCCGGGAGGTGGAGGTTGCAGTGAGCCAAGATAGCG
CCATTGCACTCCAGCCTGGACAATAAGAGTGAAACTCTATCTCAAAAATTAAAAAAAAAA
AAAAGAATAAAAGGCTTCTGCACAGCAAACATTCCACAAAGTGAAGAGACAACACACAG
GATGGGAGAAAATATCTGCAAACTAACCATCTGACAAGGAATCAATAATCAGAAGATAAG
AGTTCTAACAACTCTATAGGATAAAATATAATAATCCAATCAAAAAGTGGGCAAAATATTT
GAATTAACATTTCTCAAAAGACATACAAATGGCAAGCAGGCATATGAAAAGGTGCTCAGC
ATCATTGATCATCAGAAAAATGCAAATCAAAACTACAGTGAGGCTGGGTGCAGTGGCTCA
TGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCAGATCACTTGAGGCCAGGAG
TTCAAGATCAGCCTGGCCAACATGGCGAAACCCCATCTCTACTAAAAATATAAAAATTAG
TTGGGTGTGGTGGCGCATGCTTGTAATCCCAGCTACTTGAGTGGCTGAGGCAGGAGAAT
CTCTTGAACCCGGGAGACGGAGGTTGCAGTGAGCTGAAATCACACCACTGCCCTCCAG
CCTGGACAACAGAAGCAAAACTCCATCTCAAAAAAAACCAAAAATAGGCCGGGCATGGT
GGCTCACGCCTGTAATCCCAACACTTTGGGAGGCCGAGGCGGGTGGATCGTGAGGTCA
GGAGATTGAGACCATCATGGCTAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAA
AAAATTAGCTGGGTGTGGTGGTGGGCACCTGTAGTCCCAGCTACTCGGGAGGCTGAGG
CAGGAGAATGGCATGAACCTGGGAGGCAGAGCTTGCAGTGAGCCGAGATCATACCACT
GCACTCCAGCCTGGGTGACAGAGTGAGACTCCATCTCAAAAAAGAAAACCAAAAACCAA
AAATAAACTATAATGAGATACCATCTCAGCTCTATTAAAATGGCTTATTATTATTTTTTGAG
ATGGAGTCTCACTCTGTTGCCCAGGCTGGAGTGCAATGGCATGATCTCAGCTCACTGCA
ACCTCCGTCTCCTGGATCCAAGCGATTCTCCTGCCTCAGCCTCCTGAGTAGATGGGATT
ACAGGCACTCGCTACCATGCCCAGCTAGTTTTTGTATTTTTAGTAGAGATGGGGTTTCCC
TAGGTTGGTCAGGCTGGTCTTGAACTCCTGACCTCTGGTGATCCACCCACCTCGGTCTC
CCAAAGTGCCGGGATTACAGGTGTGAGCCATGGTGCCCGGCCTTAAAATGGCTTTTATC
CAAAGGACAGGCAATAACAAATGGTGGACAGGCAATAACAAATGGTGGCGAGGATGTGA
AGAAAAGGGAACCCTCATATACTTTTGATGGGAATGTAAATCAGTACAACCACTATGGAA
AACAGTTTGGAGGCTCCTCTAAAAACTGAAAAGTTGAGGTTCCACATGATCCAGCAATCC
CACTGCTGGGTATATCCCCAAAAAAAGGAAATCAGTGTATGGAACAGGTATCTGCACTC
CTGTTTGTTGCAGCACAGTTCACATTGCCAAAACTTAGAACCAACCTGTGTCCATCAGCA
GATGACCAGGCAAAGAAAATGTGGTACTTACACACAATGGATACTCTTCAGCCATAAAAA
AGAATGAGATCCTGGCTGGGCGCGGTGGCTCACACCTCTAATCCCAGCACTTTGGGAG
GCCGAGGCGGGCAGATCACTTGAGGCCAGAAGTTCAAGACCAGCCTGGCCAACGTGGT
GAAACCTTGTCTCTACTAAAAATGAAAAAATGAGCTGGGTGTGGTGGCGCGTGCCTGTA
ATCACAACTACTTGGGAGGCTGAGGCAGGAGAATCGCTTGAATCCGGGAGGCGGAGGT
TGCAGTGAGCCAAGATTGTGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACCCTT
TTCAAAACAAAACAAAACAAAACAAAAAAAAACATGAGATCCTGTCAATTACAACATAATGA
ACTGGGGATTATTATGTTAAGTGAAATAAGCCAGGCACAGAAAGACAAACATTGCATGTT
CTTATTTATTTGTGTATCTAAAAATCAAAACAATTGAACTCATGGAGATAAGAGTGGATGG
ATAGTTCCTGGAGGCTGGGGAGGGTAGTTAGGGGCTGGCAGGGAGATGGGGAAGGTT

AATGGGTACCACACACACACACACACACACACACACACACACAAAGAAAGAATGAATAAGA
CATACTATTTGATAACACAACCAGGTGACTATAGTCAATAATAATGGTACATTTTAAAATA
ACTTGGCAGGGTGCGGTGGCTCACACCTGTTATCCTAGCACTTTGGGAGGTCAAGGCA
GGTGGATCACGAGGTCAGGAGTTCGAGACCAGCCTGGCCAATATGGTGAAACCCCTGT
CTCTACTAAAAATACAAAAATTAGCCGGGTATAGTGGTGGGCACCTTTAGGCCCAGCTAC
TCGGGAGGCTGAGGCAGGAGAATCGTTGGAACCCAGGAGGCGGAGGTTGCAGTGAGC
CAAGATCACACCACTGCACTCCAGCCTGGGTGACACCAGGAGACTGTCTCAAAAAAATA
ATAATAACTTAGCCAGGCGCAGTGGCTCACACCTATAATCCCAGCACTGGGAGGCTGAG
GCAGGCGCATCAGGAGGTCAGGAGATCGAGACCATCCTGGCTAACGTGGTGAAACCCC
ATCTCTACTAAAAAAAAATACAAAAAATTAGCTGGGCAGGGGCCTGTAGTCCCAGCTACT
CAGGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGGAGGCGGAGCTTGCAGTGAGCC
GAGATTGCGCCACTGCACTCCAACCTGGGCGACAGAGCGAGACTCCATCTCAAAAAAAA
AAAAAATAATAATAATAATAATAACTTAAAAAGATTGTAACTGGATTGTTTGTAACTCAATG
GATAAATGCTTGAGGGTATAGGGGAAAATAAAGAACCAAAACCCCACTCTAGATTTACCC
AGGAAACAGCTCCTTTAATGACTGGTTCCTCATTTTTTTCTCCCTCCCATCCAGGCCCGG
AGGCTGTTTGATGTCCTGCACGAGCCCTTCTTGAAAGTACCCTGATGACCCCAGCTGCC
AAGGAAACCCCCAGTGTAATAATAAATCGTCCTCCCAGGCCAGGCTC

Supplementary Figure A2



Supplementary Figure A3

cDNA sequences

> isoform 201.
AAAGGGGCGTGGCGTTACAGAGCCTCTAGCGCTGGGTGTTGGGGACCTGACGCTATGG
AGCTCTCGGAGTTTTGTGGGGGACGGCTGTGAGTGGGGGGTTCCTGCTGCGGGATGAG
AACGTAGACGCCAGTGGCTCACTCGCTCCTGGCACCTTCCCTTTCAGGCTCCAGATGGA
CCCTGGGAAGGACAAAGAGGGGGTGCCCCAGCCCTCAGGGCCGCCAGCAAGGAAGAA
ATTTGTGATACCCCTCGACGAGGATGAGGTCCCTCCTGGAGTGGCCAAGCCCTTATTCC
GATCTACACAGAGCCTTCCCACTGTGGACACCTCGGCCCAGGCGGCCCCTCAGACCTA
CGCCGAATATGCCATCTCACAGCCTCTGGAAGGGGCTGGGGCCACGTGCCCCACAGGG
TCAGAGCCCCTGGCAGGAGAGACGCCCAACCAGGCCCTGAAACCCGGGGCAAAATCCA
ACAGCATCATTGTGAGCCCTCGGCAGAGGGGCAATCCCGTACTGAAGTTCGTGCGCAAT
GTGCCCTGGGAATTTGGCGACGTAATTCCCGACTATGTGCTGGGCCAGAGCACCTGTG
CCCTGTTCCTCAGCCTCCGCTACCACAACCTGCACCCAGACTACATCCATGGGCGGCTG
CAGAGCCTGGGGAAGAACTTCGCCTTGCGGGTCCTGCTTGTCCAGGTGGATGTGAAAG
ATCCCCAGCAGGCCCTCAAGGAGCTGGCTAAGATGTGTATCCTGGCCGACTGCACATTG
ATCCTCGCCTGGAGCCCCGAGGAAGCTGGGCGGTACCTGGAGACCTACAAGGCCTATG
AGCAGAAACCAGCGGACCTCCTGATGGAGAAGCTAGAGCAGGACTTCGTCTCCCGGGT
GACTGAATGTCTGACCACCGTGAAGTCAGTCAACAAAACGGACAGTCAGACCCTCCTGA
CCACATTTGGATCTCTGGAACAGCTCATCGCCGCATCAAGAGAAGATCTGGCCTTATGC
CCAGGCCTGGGCCCTCAGAAAGTAAGAGCTCTGGGAAAGAACCCAAGGAGTTGGGGGA
AGGAGAGAGCCCCAAATAAACACAACCTGAGACCCCAAAGTTTTAAGGTGAAAAAAGAA
CCAAAGACCAGACACAGTGGCTTCCGCCTGTAATCCCAACATTTTGGGAGGCCAAGGCG

GGAGGACTGCTTGAGGCCAGAAGTTGGAGACCAGCCTGGGCAAGTGGACACCTCATTT
TTACTAAAAATAAAAAAAACTAGCTGGGC

>isoform 202.
CCGGAAGTGCTGCGAGCCCTGGGCCACGCTGGCCGTGCTGGCAGTGGGCCGCCTCGA
TCCCTCTGCAGTCTTTCCCTTGAGGCTCCAAGACCAGCAGGTGAGGCCTCGCGGCGCT
GAAACCGTGAGGCCCGGACCACAGGCTCCAGATGGACCCTGGGAAGGACAAAGAGGG
GGTGCCCCAGCCCTCAGGGCCGCCAGCAAGGAAGAAATTTGTGATACCCCTCGACGAG
GATGAGGTCCCTCCTGGAGTGGCCAAGCCCTTATTCCGATCTACACAGAGCCTTCCCAC
TGTGGACACCTCGGCCCAGGCGGCCCCTCAGACCTACGCCGAATATGCCATCTCACAG
CCTCTGGAAGGGGCTGGGGCCACGTGCCCCACAGGGTCAGAGCCCCTGGCAGGAGAG
ACGCCCAACCAGGCCCTGAAACCCGGGGCAAAATCCAACAGCATCATTGTGAGCCCTC
GGCAGAGGGGCAATCCCGTACTGAAGTTCGTGCGCAATGTGCCCTGGGAATTTGGCGA
CGTAATTCCCGACTATGTGCTGGGCCAGAGCACCTGTGCCCTGTTCCTCAGCCTCCGCT
ACCACAACCTGCACCCAGACTACATCCATGGGCGGCTGCAGAGCCTGGGGAAGAACTT
CGCCTTGCGGGTCCTGCTTGTCCAGGTGGATGTGAAAGATCCCCAGCAGGCCCTCAAG
GAGCTGGCTAAGATGTGTATCCTGGCCGACTGCACATTGATCCTCGCCTGGAGCCCCGA
GGAAGCTGGGCGGTACCTGGAGACCTACAAGGCCTATGAGCAGAAACCAGCGGACCTC
CTGATGGAGAAGCTAGAGCAGGACTTCGTCTCCCGGGTGACTGAATGTCTGACCACCGT
GAAGTCAGTCAACAAAACGGACAGTCAGACCCTCCTGACCACATTTGGATCTCTGGAAC
AGCTCATCGCCGCATCAAGAGAAGATCTGGCCTTATGCCCAGGCCTGGGCCCTCAGAA
AGCCCGGAGGCTGTTTGATGTCCTGCACGAGCCCTTCTTGAAAGTACCCTGATGACCCC
AGCTGCCAAGGAAACCCCCAGTGTAATAATAAATCGTCCTCCCAGGCCAGGCTC

>isoform 203.
TGGCAGTGGGCCGCCTCGATCCCTCTGCAGTCTTTCCCTTGAGGCTCCAAGACCAGCA
GGTGAGGCCTCGCGGCGCTGAAACCGTGAGGCCCGGACCACAGGCTCCAGATGGACC
CTGGGAAGGACAAAGAGGGGGTGCCCCAGCCCTCAGGGCCGCCAGCAAGGAAGAAAT
TTGTGATACCCCTCGACGAGGATGAGGTCCCTCCTGGAGTGGCCAAGCCCTTATTCCGA
TCTACACAGAGCCTTCCCACTGTGGACACCTCGGCCCAGGCGGCCCCTCAGACCTACG
CCGAATATGCCATCTCACAGCCTCTGGAAGGGGCTGGGGCCACGTGCCCCACAGGGTC
AGAGCCCCTGGCAGGAGAGACGCCCAACCAGGCCCTGAAACCCGGGGCAAAATCCAAC
AGCATCATTGTGAGCCCTCGGCAGAGGGGCAATCCCGTACTGAAGTTCGTGCGCAATGT
GCCCTGGGAATTTGGCGACGTAATTCCCGACTATGTGCTGGGCCAGAGCACCTGTGCC
CTGTTCCTCAGCCTCCGCTACCACAACCTGCACCCAGACTACATCCATGGGCGGCTGCA
GAGCCTGGGGAAGAACTTCGCCTTGCGGGTCCTGCTTGTCCAGGTGGATGTGAAAGAT
CCCCAGCAGGCCCTCAAGGAGCTGGCTAAGATGTGTATCCTGGCCGACTGCACATTGAT
CCTCGCCTGGAGCCCCGAGGAAGCTGGGCGGTACCTGGAGACCTACAAGGCCTATGAG
CAGAAACCAGCGGACCTCCTGATGGAGAAGCTAGAGCAGGACTTCGTCTCCCGGTCTC
TGGAACAGCTCATCGCCGCATCAAGAGAAGATCTGGCCTTATGCCCAGGCCTGGGCCC
TCAGAAAGCCCGGAGGCTGTTTGATGTCCTGCACGAGCCCTTCTTGAAAGTACCCTGAT
GACCCCAGCTGCCAAGGAAACCCCCAGTG

>isoform 204.
TGGCAGTGGGCCGCCTCGATCCCTCTGCAGTCTTTCCCTTGAGGCTCCAAGACCAGCA
GGTGAGGCCTCGCGGCGCTGAAACCGTGAGGCCCGGACCACAGGCTCCAGATGGACC
CTGGGAAGGACAAAGAGGGGGTGCCCCAGCCCTCAGGGCCGCCAGCAAGGAAGAAAT
TTGTGATACCCCTCGACGAGGATGAGGTCCCTCCTGGAGTGAGGGGCAATCCCGTACT
GAAGTTCGTGCGCAATGTGCCCTGGGAATTTGGCGACGTAATTCCCGACTATGTGCTGG
GCCAGAGCACCTGTGCCCTGTTCCTCAGCCTCCGCTACCACAACCTGCACCCAGACTAC
ATCCATGGGCGGCTGCAGAGCCTGGGGAAGAACTTCGCCTTGCGGGTCCTGCTTGTCC
AGGTGGATGTGAAAGATCCCCAGCAGGCCCTCAAGGAGCTGGCTAAGATGTGTATCCT
GGCCGACTGCACATTGATCCTCGCCTGGAGCCCCGAGGAAGCTGGGCGGTACCTGGA
GACCTACAAGGCCTATGAGCAGAAACCAGCGGACCTCCTGATGGAGAAGCTAGAGCAG
GACTTCGTCTCCCGGGTGACTGAATGTCTGACCACCGTGAAGTCAGTCAACAAAACGGA
CAGTCAGACCCTCCTGACCACATTTGGATCTCTGGAACAGCTCATCGCCGCATCAAGAG
AAGATCTGGCCTTATGCCCAGGCCTGGGCCCTCAGAAAGCCCGGAGGCTGTTTGATGT
CCTGCACGAGCCCTTCTTGAAAGTACCCTGATGACCCCAGCTGCCAAGGAAACCCCCAG
TG

Protein sequences

>isoform 201.
MDPGKDKEGVPQPSGPPARKKFVIPLDEDEVPPGVAKPLFRSTQSLPTVDTSAQAAPQTYA
EYAISQPLEGAGATCPTGSEPLAGETPNQALKPGAKSNSIIVSPRQRGNPVLKFVRNVPWEF
GDVIPDYVLGQSTCALFLSLRYHNLHPDYIHGRLQSLGKNFALRVLLVQVDVKDPQQALKEL
AKMCILADCTLILAWSPEEAGRYLETYKAYEQKPADLLMEKLEQDFVSRVTECLTTVKSVNKT
DSQTLLTTFGSLEQLIAASREDLALCPGLGPQKVRALGKNPRSWGKERAPNKHNLRPQSFK
VKKEPKTRHSGFRL

>isoform 202.
MDPGKDKEGVPQPSGPPARKKFVIPLDEDEVPPGVAKPLFRSTQSLPTVDTSAQAAPQTYA
EYAISQPLEGAGATCPTGSEPLAGETPNQALKPGAKSNSIIVSPRQRGNPVLKFVRNVPWEF
GDVIPDYVLGQSTCALFLSLRYHNLHPDYIHGRLQSLGKNFALRVLLVQVDVKDPQQALKEL
AKMCILADCTLILAWSPEEAGRYLETYKAYEQKPADLLMEKLEQDFVSRVTECLTTVKSVNKT
DSQTLLTTFGSLEQLIAASREDLALCPGLGPQKARRLFDVLHEPFLKVP

>isoform 203.
MDPGKDKEGVPQPSGPPARKKFVIPLDEDEVPPGVAKPLFRSTQSLPTVDTSAQAAPQTYA
EYAISQPLEGAGATCPTGSEPLAGETPNQALKPGAKSNSIIVSPRQRGNPVLKFVRNVPWEF
GDVIPDYVLGQSTCALFLSLRYHNLHPDYIHGRLQSLGKNFALRVLLVQVDVKDPQQALKEL
AKMCILADCTLILAWSPEEAGRYLETYKAYEQKPADLLMEKLEQDFVSRSLEQLIAASREDLA
LCPGLGPQKARRLFDVLHEPFLKVP

>isoform 204.
MDPGKDKEGVPQPSGPPARKKFVIPLDEDEVPPGVRGNPVLKFVRNVPWEFGDVIPDYVLG
QSTCALFLSLRYHNLHPDYIHGRLQSLGKNFALRVLLVQVDVKDPQQALKELAKMCILADCTL
ILAWSPEEAGRYLETYKAYEQKPADLLMEKLEQDFVSRVTECLTTVKSVNKTDSQTLLTTFG
SLEQLIAASREDLALCPGLGPQKARRLFDVLHEPFLKVP

Supplementary Figure 1:

A1) *ERCC1* DNA human reference sequence. Primers used for determination of the mutant sequence by TOPO Cloning® are in bold and underlined; ZFN binding site are in bold and red; ZFN cutting site in lowercase and red. The genomic position of the ZFN cutting site is 19:45918207-19:45918212 **(**GRCh37**).**
A2) Multalin alignment of the isoforms sequence and the ZFN cutting site. Screen shot of the alignment of the isoforms sequences (beginning of *ERCC1* exon 7) using Multalin; ZFN cutting site is highlighted in yellow.
A3) ERCC1 cDNA and protein sequence of isoforms 201, 202, 203 and 204

Molecule	C°	A549	Ahez	Ac216	Ac295	Ac375
S+camptothecin	1000nM	0.06436	0.06924	0.06814	0.06305	0.05930
S+camptothecin	100nM	1.00439	1.00326	1.00037	0.97038	1.00857
S+camptothecin	10nM	0.92190	0.86461	0.93576	0.80871	0.69399
S+camptothecin	1nM	0.90687	0.86051	0.92676	0.87173	0.70391
17-AAG	1000nM	0.04379	0.03180	0.03250	0.03112	0.04336
17-AAG	100nM	0.07385	0.03961	0.06469	0.04849	0.04891
17-AAG	10nM	0.87400	0.72263	0.85257	0.69801	0.32049
17-AAG	1nM	0.92327	0.86067	0.89188	0.78534	0.75570
2-methoxyestradiol	1000nM	0.47789	0.28897	0.39143	0.38961	0.30081
2-methoxyestradiol	100nM	1.11530	1.05709	0.98585	0.97187	1.07877
2-methoxyestradiol	10nM	0.97223	0.97013	0.94509	0.96856	0.99332
2-methoxyestradiol	1nM	0.99700	0.97446	0.92902	0.91479	1.06200
4-OH-tamoxifen	1000nM	1.09446	1.04754	1.01648	0.98678	0.91135
4-OH-tamoxifen	100nM	1.12564	1.03203	1.04123	0.96301	1.04872
4-OH-tamoxifen	10nM	0.94361	0.96489	0.92000	0.95326	1.03469
4-OH-tamoxifen	1nM	0.96132	0.95576	0.91053	0.95938	1.03399
5-FU	1000nM	0.48519	0.45191	0.48070	0.30606	0.25828
5-FU	100nM	1.04934	0.93772	1.06059	0.87345	0.92629
5-FU	10nM	0.91321	0.90263	0.94887	0.94409	1.07154
5-FU	1nM	0.91294	0.91780	0.94521	0.91503	0.96281
6-thioguanine	1000nM	1.00397	0.99709	0.91091	0.85581	0.93442
6-thioguanine	100nM	1.00890	1.02358	0.94528	0.94565	0.97905
6-thioguanine	10nM	0.96076	0.95489	0.90234	0.95047	0.99074
6-thioguanine	1nM	0.97834	0.97105	0.91885	0.93794	1.00919
ABT-737	1000nM	1.14877	1.12039	1.06926	1.06000	1.15435
ABT-737	100nM	1.15162	1.09715	1.05196	1.08587	1.09513
ABT-737	10nM	0.99843	0.99237	0.92695	0.98904	1.04296
ABT-737	1nM	1.01199	1.00233	0.94619	1.04594	1.08471
AG-14699	**1000nM**	**0.89518**	**0.67205**	**0.74249**	**0.57763**	**0.44556**
AG-14699	**100nM**	**0.93051**	**0.79784**	**0.85740**	**0.71176**	**0.63730**
AG-14699	**10nM**	**0.95841**	**0.87776**	**0.93880**	**0.90764**	**0.79513**
AG-14699	**1nM**	**0.97692**	**0.88991**	**0.94286**	**0.95208**	**0.95038**
AZ4547	1000nM	1.01204	0.93181	0.99441	1.10859	0.99121
AZ4547	100nM	1.02320	1.01913	1.00662	1.08858	1.16433
AZ4547	10nM	0.99731	1.00203	0.95413	1.03138	1.05310
AZ4547	1nM	0.98833	1.01478	0.97547	1.03756	1.05933
BEZ-235	1000nM	0.14471	0.07565	0.08504	0.05742	0.08932
BEZ-235	100nM	0.28607	0.17706	0.20505	0.13109	0.14152
BEZ-235	10nM	0.96746	0.78651	0.84807	0.67706	0.42604
BEZ-235	1nM	0.99953	1.01199	0.92987	0.94382	1.00313
BI-2536	1000nM	0.04530	0.04468	0.08697	0.11810	0.11427
BI-2536	100nM	0.06099	0.05690	0.12460	0.14526	0.16772
BI-2536	10nM	0.15588	0.05045	0.16164	0.18081	0.17516
BI-2536	1nM	0.89350	0.91105	0.88588	0.86585	0.88594
BIBW2992	1000nM	0.99625	0.42974	0.56210	0.38997	0.29209
BIBW2992	100nM	1.06377	0.68551	0.72901	0.63317	0.57597
BIBW2992	10nM	0.99242	0.78309	0.81025	0.82923	0.74184
BIBW2992	1nM	1.00289	0.99704	0.94505	0.99719	0.97342
bleomycinsulfate	1000nM	0.58306	0.41146	0.57216	0.49918	0.42927
bleomycinsulfate	100nM	1.02563	0.85181	0.90221	0.84931	0.83976

bleomycinsulfate	10nM	0.90723	0.85183	0.90532	0.77598	0.68190
bleomycinsulfate	1nM	0.90964	0.82393	0.89434	0.87892	0.65855
MSC2358705A	1000nM	1.08212	0.94267	0.90281	0.82433	0.81668
MSC2358705A	100nM	1.11445	1.04958	1.01142	0.95588	1.08153
MSC2358705A	10nM	0.95780	0.98290	0.89580	0.94594	0.99817
MSC2358705A	1nM	0.97244	0.98291	0.91646	0.91509	1.03557
canertinib	1000nM	0.96749	0.63507	0.68689	0.53267	0.31565
canertinib	100nM	1.07069	0.86677	0.87146	0.80614	0.81272
canertinib	10nM	0.90541	0.81110	0.87678	0.71744	0.67641
canertinib	1nM	0.89880	0.83032	0.88064	0.79334	0.69803
carboplatin	1000nM	1.14519	1.05720	1.05158	1.08184	1.08478
carboplatin	100nM	1.17393	1.08403	1.11713	1.06155	1.14702
carboplatin	10nM	0.91404	0.95462	0.96228	0.97447	0.98215
carboplatin	1nM	0.93936	0.94115	0.93082	1.01425	1.00963
celecoxib	1000nM	1.13171	1.08028	1.03380	1.05372	1.06889
celecoxib	100nM	1.16025	1.11637	1.00100	1.02300	1.09912
celecoxib	10nM	1.01132	1.02714	0.93769	0.95471	1.04239
celecoxib	1nM	1.02065	0.99437	0.93791	0.97784	1.02819
CT241533	1000nM	1.04658	1.02644	1.02294	1.05693	1.17390
CT241533	100nM	1.03693	1.01345	0.98480	0.97200	1.07127
CT241533	10nM	0.98955	0.96912	0.94883	0.95455	1.07516
CT241533	1nM	1.00082	0.98001	0.94694	0.99441	1.04831
curcumin	1000nM	1.04363	1.01030	0.99543	0.96750	0.92643
curcumin	100nM	1.04821	0.97093	0.96985	0.93353	0.97363
curcumin	10nM	0.91770	0.83293	0.90675	0.85849	0.81947
curcumin	1nM	0.91415	0.88970	0.87475	0.84329	0.72482
cyclophosphamideH20	1000nM	1.18814	1.09005	1.09145	1.16966	1.13135
cyclophosphamideH20	100nM	1.16991	1.11744	1.11090	1.08447	1.09980
cyclophosphamideH20	10nM	0.96086	0.98788	0.95953	1.06211	1.05880
cyclophosphamideH20	1nM	0.95638	0.96230	0.94417	0.99986	1.02468
dasatinib	1000nM	0.42383	0.31882	0.34310	0.29628	0.23162
dasatinib	100nM	0.49952	0.64093	0.46995	0.50668	0.35280
dasatinib	10nM	0.75780	0.76105	0.70258	0.67426	0.50554
dasatinib	1nM	0.90659	0.86913	0.86590	0.75047	0.72211
decitabine	1000nM	1.13570	1.02741	1.07681	1.12837	1.01875
decitabine	100nM	1.16340	1.08209	1.08370	1.07055	1.10300
decitabine	10nM	0.95601	0.97282	0.93849	1.01575	1.01540
decitabine	1nM	0.94400	0.96514	0.93312	0.93565	1.08873
DMX50-1281	1000nM	0.42621	0.27301	0.39507	0.35647	0.26035
DMX50-1281	100nM	1.07691	0.91616	0.95494	0.90880	0.84983
DMX50-1281	10nM	0.98990	0.96364	0.92343	1.03098	1.03963
DMX50-1281	1nM	0.98016	0.96801	0.94523	0.91074	1.07235
DMX50-1338-2	1000nM	0.76931	0.53731	0.67597	0.55537	0.54668
DMX50-1338-2	100nM	1.09944	1.03651	0.98468	1.00231	1.04874
DMX50-1338-2	10nM	1.02677	1.01696	0.92928	0.92776	1.01070
DMX50-1338-2	1nM	1.01851	1.00348	0.92962	0.95409	1.00926
DMX501331-2	1000nM	0.64426	0.39728	0.53434	0.36802	0.34104
DMX501331-2	100nM	1.08933	1.04976	1.01159	1.05428	1.01501
DMX501331-2	10nM	0.99457	1.01091	0.93937	0.98296	1.00283
DMX501331-2	1nM	0.99415	1.00907	0.91844	0.94976	0.99176
doxorubicinHCl	1000nM	0.03918	0.02638	0.04135	0.03809	0.02673
doxorubicinHCl	100nM	0.40638	0.25411	0.37111	0.23332	0.24441
doxorubicinHCl	10nM	0.80524	0.50436	0.64296	0.53082	0.43543

doxorubicinHCl	1nM	0.95819	0.91586	0.89809	0.91545	0.94525
EB-47	1000nM	1.14049	1.06769	1.02637	1.09391	1.05083
EB-47	100nM	1.14269	1.07759	0.99163	1.08161	1.00646
EB-47	10nM	0.94442	0.92573	0.93461	0.90455	0.95124
EB-47	1nM	0.94537	0.94992	0.90260	0.93882	0.96252
erlotinib	1000nM	1.07107	0.76035	0.77959	0.64798	0.64773
erlotinib	100nM	1.11778	0.97973	0.94036	0.94247	0.84929
erlotinib	10nM	0.97224	0.97273	0.91110	0.92003	1.02074
erlotinib	1nM	0.97224	0.98617	0.89768	0.92203	1.01086
etoposide	1000nM	0.64223	0.38642	0.54724	0.46167	0.36442
etoposide	100nM	1.09065	0.91060	0.96734	0.88386	0.91414
etoposide	10nM	1.00637	1.00108	0.92813	0.93596	0.97241
etoposide	1nM	1.00930	1.01497	0.92225	0.96037	1.03269
everolimus	1000nM	0.60655	0.43638	0.38846	0.30305	0.28848
everolimus	100nM	0.63869	0.46110	0.43995	0.37172	0.34371
everolimus	10nM	0.66207	0.48654	0.47007	0.33605	0.33942
everolimus	1nM	0.81519	0.65553	0.59924	0.45683	0.42732
Flavopiridol	1000nM	0.05945	0.04343	0.04897	0.02793	0.03147
Flavopiridol	100nM	0.50834	0.22573	0.37434	0.24005	0.20479
Flavopiridol	10nM	0.98294	0.98342	0.96847	0.96300	1.03633
Flavopiridol	1nM	0.95498	0.98471	0.95409	1.03368	1.04504
GDC-0449	1000nM	0.90490	0.86089	0.93324	0.85270	0.81795
GDC-0449	100nM	0.95920	0.90254	0.94004	0.83818	0.71879
GDC-0449	10nM	0.95803	0.91941	0.95273	0.88152	0.95065
GDC-0449	1nM	0.95100	0.93030	0.97077	1.02730	0.90196
gefitinib	1000nM	1.04659	0.76576	0.79789	0.67975	0.61300
gefitinib	100nM	1.11263	0.99054	0.96670	0.94015	0.74882
gefitinib	10nM	0.96944	0.96360	0.90528	0.94090	0.95308
gefitinib	1nM	0.97722	0.94821	0.91587	0.99551	1.00010
gemcitabineHCl	1000nM	0.17954	0.12272	0.12237	0.14482	0.16106
gemcitabineHCl	100nM	1.00476	0.46163	0.88592	0.50036	0.73989
gemcitabineHCl	10nM	0.99245	0.98635	0.94495	0.92979	1.00543
gemcitabineHCl	1nM	0.99884	0.99279	0.93974	0.96685	1.01476
GSK-2334470A	1000nM	0.63986	0.46959	0.43354	0.40536	0.29929
GSK-2334470A	100nM	0.92539	0.73496	0.79316	0.60697	0.49317
GSK-2334470A	10nM	0.90879	0.86000	0.93813	0.92808	0.93340
GSK-2334470A	1nM	0.88593	0.89336	0.90655	0.89528	0.91220
GSK1904529A	1000nM	0.98820	0.88343	0.85003	0.83044	0.86797
GSK1904529A	100nM	0.99818	1.00223	0.94387	0.97708	1.05477
GSK1904529A	10nM	1.00968	0.99585	0.95620	1.03517	1.05357
GSK1904529A	1nM	0.98476	0.98394	0.97358	0.94837	1.04744
GSK2194069A	1000nM	1.04664	0.94667	0.90862	0.88902	0.84369
GSK2194069A	100nM	1.08531	0.94512	0.95403	0.93924	0.80715
GSK2194069A	10nM	1.03565	0.97661	0.93076	0.92319	1.04257
GSK2194069A	1nM	1.00237	1.02132	0.95813	0.98989	1.08464
imatinibmesylate	1000nM	1.17111	1.14147	1.07328	1.14803	1.07370
imatinibmesylate	100nM	1.14399	1.14656	1.06160	1.13170	1.11630
imatinibmesylate	10nM	0.97828	0.98721	0.95306	0.96657	1.10898
imatinibmesylate	1nM	0.98230	1.00224	0.95314	1.00773	1.11268
RO-3306	2000nM	0.33233	0.20627	0.22202	0.28583	0.18282
RO-3306	1000nM	0.54243	0.47951	0.59235	0.50004	0.22774
RO-3306	500nM	0.91013	0.75926	0.84561	0.83599	0.68714
RO-3306	200nM	0.93561	0.87606	0.92079	0.84255	0.91526

KU-59652	1000nM	0.92639	0.78651	0.99592	0.88742	0.80865
KU-59652	100nM	1.04936	0.97887	1.06862	1.05366	1.03081
KU-59652	10nM	0.98099	0.93607	0.97233	1.03661	1.05409
KU-59652	1nM	0.96845	0.92015	0.95540	0.95849	0.99391
KU0057788	1000nM	0.93640	0.84444	0.86510	0.77620	0.63571
KU0057788	100nM	1.03942	1.03049	0.98716	1.01006	1.03584
KU0057788	10nM	1.00288	0.98331	0.96963	0.99920	1.05071
KU0057788	1nM	1.01465	1.00293	0.96676	0.96346	0.92830
lapatinib	1000nM	0.95883	0.71531	0.72161	0.66006	0.63421
lapatinib	100nM	1.01434	0.88572	0.86369	0.86501	0.75369
lapatinib	10nM	0.94697	0.92570	0.93008	0.96756	0.98796
lapatinib	1nM	0.98740	0.96525	0.93665	1.01785	1.05747
lestaurtinib	200nM	0.33116	0.27905	0.38981	0.35424	0.20479
lestaurtinib	20nM	1.10087	0.50656	0.68286	0.59274	0.50388
lestaurtinib	2nM	0.99734	0.99817	0.93317	1.01066	1.00956
lestaurtinib	0.2nM	1.00612	1.01114	0.93546	1.03376	1.06823
MDV-3100	1000nM	1.12141	1.04575	1.05024	1.01553	1.07476
MDV-3100	100nM	1.12836	1.07097	1.06160	1.03421	1.06126
MDV-3100	10nM	0.90205	0.90563	0.95031	0.98975	0.97150
MDV-3100	1nM	0.89265	0.93696	0.92146	0.97047	0.99527
MK-1175	1000nM	0.28590	0.11186	0.26009	0.18398	0.20025
MK-1175	100nM	1.11654	0.69450	0.74246	0.41884	0.48546
MK-1175	10nM	0.99785	0.97754	0.91275	0.99778	0.99645
MK-1175	1nM	1.01133	1.00041	0.90697	0.99856	0.96139
MK-2512	**1000nM**	**1.13499**	**0.92053**	**0.92073**	**0.81718**	**0.73019**
MK-2512	**100nM**	**1.14129**	**1.05707**	**0.97557**	**0.97216**	**0.99396**
MK-2512	**10nM**	**0.97810**	**0.99191**	**0.92484**	**0.96931**	**0.96105**
MK-2512	**1nM**	**1.01837**	**1.01787**	**0.94754**	**1.03477**	**1.10993**
MK-4827	**1000nM**	**1.14628**	**0.72293**	**0.74040**	**0.61415**	**0.35634**
MK-4827	**100nM**	**1.12694**	**1.00061**	**1.00897**	**0.96203**	**0.93744**
MK-4827	**10nM**	**0.96846**	**0.94744**	**0.90861**	**0.91861**	**0.96146**
MK-4827	**1nM**	**0.98521**	**0.96908**	**0.93112**	**0.96523**	**1.07433**
nilotinib	1000nM	1.04293	0.93438	0.94898	0.73739	0.82142
nilotinib	100nM	1.08725	0.99948	0.96802	0.84970	0.93668
nilotinib	10nM	0.97159	0.95935	0.93306	0.95891	0.92794
nilotinib	1nM	0.96074	0.95038	0.96008	0.94407	0.94080
Nutlin3	1000nM	1.01757	0.79782	0.63930	0.42263	0.29923
Nutlin3	100nM	1.14177	1.06581	0.98780	0.91919	0.96011
Nutlin3	10nM	0.96038	0.94245	0.91437	0.88718	0.98904
Nutlin3	1nM	0.96611	0.94726	0.91949	0.93508	1.00875
OL-PIX-A17A	**1000nM**	**1.03788**	**0.54889**	**0.63931**	**0.50866**	**0.32049**
OL-PIX-A17A	**100nM**	**1.07923**	**0.90719**	**0.92198**	**0.83633**	**0.75394**
OL-PIX-A17A	**10nM**	**0.98477**	**0.94131**	**0.91270**	**0.91909**	**0.95605**
OL-PIX-A17A	**1nM**	**0.98517**	**0.95498**	**0.94633**	**0.90027**	**0.99121**
OL-PIX-F3-B	**1000nM**	**0.94847**	**0.50279**	**0.57972**	**0.45207**	**0.36782**
OL-PIX-F3-B	**100nM**	**1.09718**	**0.88768**	**0.87509**	**0.86980**	**0.70453**
OL-PIX-F3-B	**10nM**	**1.00195**	**0.97610**	**0.89658**	**0.84137**	**0.95675**
OL-PIX-F3-B	**1nM**	**1.00624**	**1.03509**	**0.92321**	**0.92625**	**1.03216**
olaparib	**1000nM**	**0.86636**	**0.63475**	**0.68377**	**0.55010**	**0.39935**
olaparib	**100nM**	**1.02335**	**0.86818**	**0.95404**	**0.74458**	**0.71937**
olaparib	**10nM**	**0.95033**	**0.90926**	**0.92528**	**0.90076**	**0.91896**
olaparib	**1nM**	**0.96195**	**0.89067**	**0.95244**	**0.99046**	**0.91933**
paclitaxel	1000nM	0.11822	0.07564	0.10144	0.10353	0.10017

paclitaxel	100nM	0.15213	0.10278	0.12220	0.13898	0.12005
paclitaxel	10nM	0.28207	0.19935	0.26386	0.25195	0.16956
paclitaxel	1nM	0.96671	0.72688	0.77055	0.55868	0.32923
PBS-1086	1000nM	1.06607	0.97620	0.94120	0.83067	1.05432
PBS-1086	100nM	1.05483	1.02555	0.98012	1.00544	1.04373
PBS-1086	10nM	0.95486	0.96495	0.90930	0.89525	0.97488
PBS-1086	1nM	0.97393	0.96165	0.91252	0.92031	1.14054
PBS-1169	1000nM	1.10503	1.06372	1.00904	0.98148	0.99574
PBS-1169	100nM	1.05629	1.03773	0.95370	0.98196	1.02751
PBS-1169	10nM	0.98324	1.00524	0.93380	0.93405	1.01886
PBS-1169	1nM	0.98383	0.95420	0.91500	0.91862	0.99858
PD-0332991	1000nM	0.98044	0.80512	0.88965	0.71367	0.62803
PD-0332991	100nM	1.12349	0.96077	1.04887	0.88064	0.82149
PD-0332991	10nM	0.96804	0.98586	0.95395	0.96961	1.00485
PD-0332991	1nM	0.93910	0.96840	0.95177	1.01311	1.02321
PD-184352	1000nM	0.69089	0.58260	0.62830	0.48717	0.34535
PD-184352	100nM	0.95883	0.84910	0.87745	0.78171	0.71531
PD-184352	10nM	0.98508	0.95810	0.94768	0.94228	0.97858
PD-184352	1nM	0.97629	0.95912	0.95797	0.98135	1.03270
PD173074	1000nM	0.97850	0.99468	0.97074	0.94762	0.85702
PD173074	100nM	1.00063	1.01303	0.98622	0.94875	1.10956
PD173074	10nM	0.96107	0.95140	0.94836	0.99728	1.02934
PD173074	1nM	0.96673	0.94322	0.95785	1.00322	0.97430
PF-00299804	1000nM	0.92392	0.52416	0.58009	0.37690	0.36489
PF-00299804	100nM	1.00774	0.66916	0.72216	0.61308	0.53155
PF-00299804	10nM	0.99846	0.81098	0.82476	0.77707	0.71805
PF-00299804	1nM	1.00304	0.94542	0.93169	0.90284	0.91684
PF-00477736	1000nM	0.84368	0.28674	0.36559	0.33364	0.38955
PF-00477736	100nM	0.97672	0.84656	0.97639	0.82436	0.79513
PF-00477736	10nM	0.97844	0.90891	0.93920	0.89483	0.91370
PF-00477736	1nM	0.96367	0.91393	0.95625	0.93553	0.91902
PF-02341066	1000nM	0.40089	0.21453	0.34041	0.21612	0.19664
PF-02341066	100nM	1.04379	0.93159	0.95893	0.92006	0.81593
PF-02341066	10nM	0.91643	0.88004	0.90925	0.78173	0.72164
PF-02341066	1nM	0.93728	0.85070	0.91855	0.83510	0.65033
PF-03758309	1000nM	0.23403	0.08786	0.24353	0.16164	0.15825
PF-03758309	100nM	0.30727	0.24074	0.34859	0.22240	0.24935
PF-03758309	10nM	0.43182	0.32287	0.43280	0.29802	0.25616
PF-03758309	1nM	1.02989	0.85847	0.92389	0.90407	0.81744
PF-03814735	1000nM	0.34694	0.21714	0.29281	0.21797	0.18211
PF-03814735	100nM	0.51451	0.33978	0.41628	0.36272	0.38264
PF-03814735	10nM	0.98144	0.93034	0.92818	0.97266	0.98269
PF-03814735	1nM	0.99628	0.96416	0.94482	0.94256	0.91153
PF-04691502	1000nM	0.09977	0.07042	0.05169	0.04534	0.04840
PF-04691502	100nM	0.66510	0.45915	0.42187	0.36608	0.31917
PF-04691502	10nM	0.97827	0.88278	0.77020	0.78771	0.76231
PF-04691502	1nM	0.98933	0.97078	0.91301	0.91946	1.01732
PF-04929113	1000nM	0.04712	0.03584	0.03631	0.03263	0.04842
PF-04929113	100nM	0.21676	0.15503	0.20103	0.21268	0.16861
PF-04929113	10nM	0.97550	0.93683	0.93493	1.05691	1.13261
PF-04929113	1nM	0.97679	0.94166	0.93771	0.92052	1.01241
PF-332991	1000nM	0.82411	0.73119	0.83353	0.65317	0.52308
PF-332991	100nM	0.98477	0.80586	0.89210	0.86339	0.61942

PF-332991	10nM	1.01043	0.96780	0.93105	0.87421	0.94777
PF-332991	1nM	1.02918	1.00279	0.93321	1.05248	1.04644
MK2206	1000nM	0.90976	0.80832	0.66003	0.59495	0.48455
MK2206	100nM	1.10693	0.94079	0.82700	0.83414	0.75458
MK2206	10nM	1.00622	0.98823	0.94096	0.92005	1.04687
MK2206	1nM	1.01312	0.97553	0.95353	0.95486	1.02885
PLX-4720	1000nM	0.98523	1.00180	0.92444	0.92816	0.87931
PLX-4720	100nM	1.03472	1.00860	0.98569	0.99825	1.02728
PLX-4720	10nM	0.92183	0.86033	0.95069	0.83048	0.75000
PLX-4720	1nM	0.91997	0.84944	0.92188	0.82263	0.73293
PRIMA1	1000nM	1.10629	1.07261	0.99859	0.99853	1.03166
PRIMA1	100nM	1.10752	1.03847	1.02304	0.99862	1.03488
PRIMA1	10nM	0.97644	0.96226	0.91632	0.96158	0.95237
PRIMA1	1nM	0.99640	0.97082	0.92985	0.97567	1.03594
resveratrol	1000nM	1.12493	1.03769	0.98987	0.91825	0.98696
resveratrol	100nM	1.04719	1.04124	1.01393	0.99273	1.02800
resveratrol	10nM	1.00692	0.97734	0.95981	1.04169	1.12049
resveratrol	1nM	1.00064	1.00810	0.96754	1.05016	1.15342
salinomycin	1000nM	0.51362	0.26264	0.51667	0.35612	0.18982
salinomycin	100nM	1.01665	0.99188	0.95804	0.99608	1.00165
salinomycin	10nM	0.95675	0.95696	0.92801	0.96701	0.94778
salinomycin	1nM	0.93381	0.92693	0.91311	0.89109	0.95350
sapacitabine	1000nM	0.38906	0.44180	0.39767	0.38268	0.26875
sapacitabine	100nM	0.80406	0.87341	0.72404	0.55133	0.53838
sapacitabine	10nM	0.91197	0.84512	0.89565	0.75437	0.68695
sapacitabine	1nM	0.89885	0.82597	0.87569	0.81226	0.71438
SAR-20106	1000nM	0.49305	0.34292	0.31285	0.30628	0.38852
SAR-20106	100nM	1.11708	1.05303	0.98584	0.88201	0.95256
SAR-20106	10nM	0.90808	0.91037	0.91644	0.96525	0.96797
SAR-20106	1nM	0.91869	0.94534	0.90448	0.93238	0.96401
sorafenib	1000nM	1.15146	1.04078	1.00303	1.00315	1.07810
sorafenib	100nM	1.20783	1.12927	1.04459	1.09520	1.14289
sorafenib	10nM	0.95713	1.00818	0.92965	1.05581	1.01153
sorafenib	1nM	0.96402	0.98222	0.92539	0.98728	1.02480
sunitinib	1000nM	0.32919	0.20796	0.24872	0.20441	0.18027
sunitinib	100nM	1.08072	1.00649	1.04180	1.08867	1.08249
sunitinib	10nM	0.91432	0.88026	0.86346	0.86999	0.76489
sunitinib	1nM	0.87898	0.83627	0.91639	0.78841	0.76332
sutent	1000nM	0.28739	0.17719	0.26696	0.20070	0.15504
sutent	100nM	1.04052	0.90805	0.98679	0.93346	0.87897
sutent	10nM	0.91364	0.90588	0.93244	0.90844	0.91554
sutent	1nM	0.93086	0.88390	0.93417	0.92217	0.94857
temozolomide	1000nM	1.08550	1.05091	0.99974	0.96300	1.06854
temozolomide	100nM	1.04450	1.05492	0.98227	1.04370	1.08917
temozolomide	10nM	1.00986	1.00610	0.97090	0.98317	0.96846
temozolomide	1nM	1.02434	1.02321	0.95851	1.00187	1.09096
vinorelbine	1000nM	0.14915	0.08106	0.15523	0.10947	0.12501
vinorelbine	100nM	0.33038	0.16691	0.27447	0.21487	0.21481
vinorelbine	10nM	0.95402	0.67119	0.88098	0.75215	0.38945
vinorelbine	1nM	0.95174	0.96315	0.92676	0.96189	0.96268
vorinostat	1000nM	0.61639	0.40208	0.48525	0.36408	0.38094
vorinostat	100nM	1.17813	1.12205	1.01139	0.97716	1.09804
vorinostat	10nM	0.92413	0.93994	0.90299	0.87263	1.02758

vorinostat	1nM	0.91291	0.91712	0.88394	0.86773	0.92531
XAV-939	1000nM	1.02783	0.86897	0.87626	0.77126	0.74663
XAV-939	100nM	1.00257	0.98694	0.96990	0.97659	1.04307
XAV-939	10nM	0.96106	0.99701	0.95676	1.03044	1.06214
XAV-939	1nM	0.97871	0.96386	0.93504	0.99947	1.03602

Supplementary table 1. Complete results of the drug screen in the ERCC1-isogenic model

Survival fractions of all cell lines are displayed for every concentration of each compound. PARP1/2 inhibitors that came up as “hits” of the screen are highlighted in bold.

Figure	drug	Cell line	EC_{50} (μM)	EC_{50} SEM (μM)	SF_{50} (μM)
Figure 3. 2. E	cisplatin	A549	6.185	1.298	8.114
		Ahez	1.652	1.445	1.737
		Ac216	0.015	3.79	0.014
		Ac295	0.08	1.205	0.054
		Ac375	0.017	3.665	0.021
Figure 3. 12. A	olaparib	A549	15.84	1.211	17.621
		Ac216	1.46	1.189	2.250
		Ac295	0.287	3.968	2.587
		Ac375	0.388	2.46	2.545
	niraparib	A549	3.863	1.031	4.061
		Ac216	0.913	1.059	0.978
		Ac295	0.128	1.274	1.195
		Ac375	1.122	1.109	1.288
Figure 3. 13. A	olaparib	A549	6.283	1.228	6.866
		Ac216	0.09	1.989	0.166
		Ac295	0.582	1.274	0.717
		Ac375	0.018	1.602	0.110
		DLD1 B2+/+	6.586	ND	6.884
		DLD1 B2-/-	0.015	ND	0.019
Figure 4. 1. B	olaparib	U2OS siCON	1.621	1.307	1.807
		U2OS siBRCA2	0.006	ND	0.079
		U2OS siERCC1	0.007	ND	0.069
Figure 4. 1. D	BMN673 (nM)	A549	9.828	1.122	11.664
		Ac216	0.953	1.742	2.303
		Ac375	0.726	2.552	1.415
		DLD1 B2+/+	2.272	2.451	2.531
		DLD1 B2-/-	0.79	ND	1.316
Figure 4. 5	olaparib	A549	15.84	1.211	17.621
		Ac216 + isof201	0.564	1.978	1.447
		Ac216 + isof202	19.9	1.128	23.914
		Ac216 + isof203	0.16	3.963	1.364
		Ac216 + isof204	0.36	1.5	1.770
	niraparib	A549	3.863	1.031	4.061
		Ac216 + isof201	0.698	1.11	0.700
		Ac216 + isof202	5.4	1.032	5.855
		Ac216 + isof203	0.511	2.295	0.638
		Ac216 + isof204	0.138	2.487	0.630
Figure 4. 7. A	olaparib	A549 siCON	17.06	1.824	19.110
		A549 siBRCA2	0.061	1.153	1.491
		Ac216 siCON	2.295	1.17	3.702
		Ac216 siBRCA2	0.586	ND	2.064
		Ac295 siCON	3.87	2.154	4.591
		Ac295 siBRCA2	2.466	ND	5.923
		Ac375 siCON	2.341	1.495	3.566
		Ac375 siBRCA2	4.5	ND	11.541
Figure 4. 11. A	niraparib	A549 siCON	2.22	1.022	2.353
		A549 siPARP1	3.214	1.025	3.196
		Ac216 siCON	0.45	1.17	0.933
		Ac216 siPARP1	3.243	1.37	3.884
		Ac295 siCON	0.187	2.65	1.697
		Ac295 siPARP1	4.652	1.193	4.085
		Ac375 siCON	0.074	ND	0.796
		Ac375 siPARP1	1.351	1.428	2.680

Supplementary table 2. EC_{50} and SF_{50} values

EC_{50} and SF_{50} of each cell line for all experiments displayed in the manuscript, with experimental error (Standard Error of the Mean, SEM). All values are displayed in μM, unless otherwise notified (BMN673). Abbreviations: ND = Not Determined

Abbreviation	Full gene / protein name
ALDH1A3	Aldehyde dehydrogenase 1 family, member A3
CDH1	cadherin 1
COX6C	Cytochrome c oxidase subunit VIc
CYP4F11	Cytochrome P450 4F11
FAM131A	Family with sequence similarity 131
FEZ1	Fasciculation and elongation protein zeta 1
HSPA2	Heat shock protein A2
INPP4B	Inositol polyphosphate-4-phosphatase
ITGA6	integrin, alpha 6
MLPH	melanophilin
MYC	v-myc myelocytomatosis viral oncogene homolog
OCIAD2	OCIA domain containing 2 (154 aa)
PAPPA	Pregnancy-associated plasma protein 1
PPP2R2C	protein phosphatase 2, regulatory subunit B
RAB27B	RAB27B, member RAS oncogene family
TK1	thymidine kinase 1

Supplementary table 3. Full genes names of hits shared between the trancriptomic and SILAC analysis, or transcriptomic analysis and gene expression dataset from Friboulet *et al.*, Clin Cancer Res 2011

Publications from this thesis

Publications directly related to the thesis

The potential of exploiting DNA-repair defects for optimizing lung cancer treatment
Postel-Vinay S, Vanhecke E, Olaussen KA, Lord CJ, Ashworth A, Soria JC.
Nat Rev Clin Oncol. 2012 Feb 14;9(3):144-55.

A high-throughput screen identifies PARP1/2 inhibitors as a potential therapy for ERCC1-deficient non-small cell lung cancer.
Postel-Vinay S, Bajrami I, Friboulet L, Elliott R, Fontebasso Y, Dorvault N, Olaussen KA, André F, Soria JC, Lord CJ, Ashworth A.
Oncogene. 2013 Nov 21;32(47):5377-87

ERCC1 function in nuclear excision and interstrand crosslink repair pathways is mediated exclusively by the ERCC1- 202 isoform.
Friboulet L, Postel-Vinay S, Sourisseau T, Adam J, Stoclin A, Ponsonnailles F, Dorvault N, Commo F, Saulnier P, Salome-Desmoulez S, Pottier G, André F, Kroemer G, Soria JC, Olaussen KA.
Cell Cycle. 2013 Oct 15;12(20):3298-306

Publication indirectly related to the thesis

AXL and acquired resistance to EGFR inhibitors.
Postel-Vinay S, Ashworth A.
Nat Genet. 2012 Jul 27;44(8):835-6.