Predictive Biomarkers in Muscle-Invasive Bladder Cancer

Alexandra K Walker

Linacre College

Hilary Term 2017

University of Oxford

Supervisor

Professor Anne Kiltie

Co Supervisor

Judith Nicholson

Thesis submitted to the University of Oxford

in accordance with the requirements for the degree

of Doctor of Philosophy

The study presented in this thesis was performed at the Cancer Research UK/ Medical Research Council Oxford Institute for Radiation Oncology, University of Oxford. This work was supported by a scholarship from the Medical Research Council.

Abstract

Organ-confined muscle-invasive bladder cancer (MIBC) is treated with either radical cystectomy (RC) or radiotherapy (RT) and carries survival rates of only 40-60%. There is no clear survival advantage between the two therapies and therefore biomarkers that could aid in the choice of treatment would be of use in this disease. Both the abundance of the DNA repair protein, MRE11, and the genotype of a single nucleotide polymorphism in the *MRE11* gene, have been found to have potential as predictive biomarkers of RT outcome in MIBC.

The aim of this thesis was to investigate biomarkers relevant to MIBC cancer-specific survival (CSS). To achieve this aim, an automated MRE11 IHC biomarker was investigated, with the goal of obtaining the technical validation required to progress this potential biomarker toward a prospective study. Following on from this the specificity of MRE11 IHC as a biomarker for MIBC was tested by performing MRE11 IHC on a cohort of anal cancer samples. Next the rs1805363 G>A germline SNP in *MRE11* was investigated, with the aim of validating the previously reported association between the presence of the rs1805363 A minor allele and increased *MRE11* isoform 2 expression. The role of *MRE11* isoform 2 in the repair of radiation-induced DNA damage was then examined. Finally, the crowdsourcing of IHC scoring, through a smartphone gaming app, was investigated as a tool to reliably score potential IHC biomarkers with increased efficiency.

The re-optimisation of an automated MRE11 IHC protocol was found to increase the reliability of IHC scoring and significantly alter the distribution of scores for patients within a MIBC TMA tissue cohort when compared to an insufficiently optimised assay.

However, this study did not have the statistical power to validate MRE11 as a predictive biomarker of RT response in MIBC. Additionally MRE11 IHC was not found to be predictive of patient relapse-free survival in anal cancer.

In the analysis of the rs1805363 G>A germline MRE11 SNP, the carriage of the rs1805363 A minor allele was found to correlate with a significant increase in the expression of *MRE11* isoform 2 in patient tumour biopsies. Cells which expressed an *MRE11* isoform 2 construct were also found to have a significant reduction in survival following irradiation, increased levels of endogenous DNA damage and slower kinetics of DSB repair compared to cells which expressed the *MRE11* isoform 1 construct.

In the investigation into crowdsourcing as a method of scoring IHC, the crowdsourced scores for six of the eleven IHC stains analysed were found to have a strong correlation with scores produced by trained researchers. The crowdsourced scores were also found to be of sufficient accuracy to identify MRE11 and CK20 as potential predictive and prognostic biomarkers in MIBC. However, as implemented by this study, the use of crowdsourcing did not result in an increase in the speed of data analysis.

In conclusion, this thesis provided more evidence for MRE11 as a key target for investigations into MIBC biomarkers. The finding that *MRE11* isoform 2 expression results in a DNA repair defect is an important consideration for future studies of this protein as a biomarker. Crowdsourcing was identified as a potentially useful tool for screening biomarkers. However, to reach its full potential further investigation into the optimal application of this methodology is needed.

Acknowledgements

I would like to express my whole-hearted appreciation to my supervisors Professor Anne Kiltie and Dr Judith Nicholson. Thank you for all of your advice, guidance, encouragement and support. You have made my DPhil experience hugely rewarding.

I would also like to thank all of the members of the Kiltie lab (past and present) for making the duration of my DPhil an enjoyable one. Particularly I would like to thank Katalin for imparting the ways of IHC to me during my first month. Jude and Martin thank you for letting me talk my ideas at you even when they were just incoherent rambles. Helen you are responsible for getting me to go to the only two salsa classes I will probably ever take and you were always there to share dog walking stories with (and mojitos).

Thank you to all the people in the ORCRB building and the department of Oncology for their friendship, knowledge and support. I will always be grateful for my time in Dr Peter O'Neill's lab, during my Masters, which was my first experience in the department and made me want to come back. Nicola Sibson, who was my director of graduate studies, was an amazing support when I experienced a tough period of illness and I am appreciative of her help and understanding. For keeping me well supplied with coffee and hilarity I'd like to thank Katie, Judith (Oehlar) and Mike.

I would like to express my appreciation to Linacre College for the encouragement they have given me, specifically Catherine Walters, who I will always remember as the most kind and selfless women I've ever met.

To my family, friends, and Oxford Lindy Hoppers I will always be indebted. Four years is a long time. I would not have reached this stage without the unwavering belief of my

family and Anna (honorary family member). Oxford Lindy Hoppers have been friends and my life outside the lab. Thank you for keeping me sane and keeping me dancing.

Many aspects of my DPhil involved large collaborative projects. I would like to thank CRUK and everyone involved in Reverse the Odds, particularly Rupesh Robinson-Vyas and Dr Peter Smittenaar for all the time and effort they put into the gaming side of the project and its promotion, along with collating all of the public scores. I would also like to express my thanks to Catherine West, Ananya Choudhury, Shaun McGill and Dr Victoria Strauss for their work on the BCON project. For statistical assistance I would also like to thank Dr Christiana Kartsonaki.

Declaration

I declare that this thesis and the work described in it is my work, aside from the following areas:

In Chapter 3, scoring of MRE11 IHC on BCON TMAs was conducted in Oxford by Anne Kiltie, Shaun McGill and Alexandra K Walker. In Manchester, MRE11 IHC scoring was conducted by Catherine West and Ananya Choudhury. Kaplan-Meier curves with log rank statistics and Cox proportional hazard analyses were conducted by Dr Victoria Strauss from the Centre for Statistics in Medicine (University of Oxford).

In Chapter 4, Kaplan-Meier curves, Cox proportional hazards analysis and the correlation matrix were generated by Dr Christiana Kartsonaki from the Department of Population Health (University of Oxford). All analysis and interpretation of these statistics were conducted by Alexandra K Walker.

For Chapter 6, the Reverse the Odds mobile app was a joint collaboration between Cancer Research UK, Channel 4, Maverick Television, CHUNK and Zooniverse. Collation, aggregation and correction of public scoring was managed by the CRUK Citizen Science team, specifically Dr Peter Smittenaar. Spearman's rank correlation, quadratic-weighted kappa, Figure 6.4 and Figure 6.8 were produced by Dr Peter Smittenaar (CRUK Citizen Science statistical team). Figure 6.7, Kaplan-Meier curves, Cox proportional hazards analysis and the correlation matrix were generated by Dr Christiana Kartsonaki from the Department of Population Health (University of Oxford). All analysis and interpretation of these statistics were conducted by Alexandra K Walker.

Table of Contents

Table of Figures

Table of Tables

Abbreviations

53BP1: p53 binding protein 1

5-FU: 5-flurouracil

aa: amino acid

ADMA: asymmetric dimethylated arginine

AEC: 3-amino-9-ethylcarbazole

AP: alkaline phosphatase

APTX: aprataxin

ASYM25: Anti-dimethyl-Arginine Antibody, asymmetric

ATM: Ataxia telangiectasia mutated protein

ATP: adenosine triphosphate

BCG: Bacillus Calmette-Guerin

BCON: bladder carbogen nicotinamide trial

BLM: Bloom syndrome protein

bp: base pair

BRCA1: breast cancer susceptibility gene 1

BRCA2: breast cancer susceptibility gene 2

BRCT1: BRCA1 C-terminal domain

BRCT2: BRCA2 c terminus domain

BSA: bovine serum albumin

CDK2: cyclin dependent kinase 2

cDNA: copy DNA

95% CI: 95% confidence interval

CIS: carcinoma *in situ*

CK20: cytokeratin 20

CK5/6: cytokeratin 5/6

CO_2: carbon dioxide

CRT: chemoradiotherapy

CSS: cancer-specific survival

CtIP: C-terminal binding protein 1

DAB: 3,3'-diaminobenzidine tetrahydrochloride

ddH2O: double distilled water

DDR: DNA damage response

dH2O: distilled water

DMSO: dimethylsulphoxide

DNA: deoxyribonucleic acid

DNA2: DNA replication helicase/nuclease 2

DNA-PKcs: DNA-dependent protein kinase catalytic subunit

dNTP: deoxyribonucleotide triphosphate

DPX: dibutylphthalate polystyrene xylene

DSB: double-stranded DNA

E2F1: E2F transcription factor 1

E2F2: E2F transcription factor 2

E6AP: E6 associated protein

EAU: European association of urology

EDTA: ethylene diamine tetra acetic acid

EMT: epithelial-mesenchymal transition

ER: eostrogen receptor

ERCC1: ERCC excision repair 1, endonuclease non-catalytic subunit

ERCC2: excision repair cross-complementation group 2

Exo1: exonuclease 1

FBS: foetal bovine serum

FFPE: formalin fixed paraffin embedded

FGFR3: fibroblast growth factor receptor 3

FHA: forkhead-associated domain

FISH: fluorescent in situ hybridisation

GAR: glycine-arginine-rich

HCl: hydrochloric acid

HDAC: Histone deacetylase

HEPES: N-2-Hydroxyethylpiperazine-N'-2'-ethanesulfonic acid

HER2: human epidermal growth factor 2

HPV: Human papilloma virus

hr: hazard ratio

HR: homologous recombination

hrE6: high risk E6 protein

hrE7: high risk E7 protein

HRP: horseradish peroxidase

ICC: intraclass correlation coefficient

IHC: immunohistochemistry

IMRT: intensity modulated radiotherapy

KRAS: Kirsten rat sarcoma viral oncogene homolog

LOH: loss of heterozygosity

MIBC: muscle-invasive bladder cancer

MMA: monomethylated arginine

MMEJ: microhomology-mediated end-joining

MRE11: meiotic recombination 11

MRN: MRE11-RAD50-NBS1 complex

NBS: Nijmegen breakage syndrome

NHEJ: non-homologous end-joining

NICE: National Institute for Health and Care Excellence

NMIBC: non-muscle-invasive bladder cancer

O_2: Oxygen

OS: overall survival

PBS: phosphate buffered saline

PBST: PBS tween

PCR: polymerase chain reaction

PNKP: polynucleotide kinase 3'-phosphatase

PR: progesterone receptor

pRb: retinoblastoma protein

PRMT1: protein arginine methyltransferase type I

PTEN: phosphatase and tensin homolog

QOL: Quality of life

qPCR: quantitative polymerase chain reaction

RC: radical cystectomy

RFS: relapse free survival

RNA: ribonucleic acid

RNA-seq: RNA sequencing

ROS: reactive oxygen species

RPA: replication protein A

rpm: revolutions per minute

RT: radiotherapy

SCCA: squamous cell carcinomas of the anus and anal canal

SDMA: symmetric dimethylated arginine

SDS: sodium dodecyl sulfate

SDSA: synthesis-dependent strand annealing

SDSS: Sloan Digital Sky Survey

shRNA: small hairpin RNA

siRNA: small interfering RNA

SNP: single nucleotide polymorphism

SPARE: Selective Bladder Preservation Against Radical Excision Trial

SQS: semi-quantitative score

SSB: single-strand break

ssDNA: single-stranded DNA

TBE: tris-borate EDTA

TIL: tumour infiltrating lymphocyte

Tip60: Tat interactive protein 60kDa

TMA: tissue microarray

TMT: trimodality therapy/treatment

TNM: Tumour Node Metastasis

TP53: tumour protein p53

TURBT: transurethral resection of the bladder tumour

UC: urothelial carcinoma

WRN: Werner

WSR: Wilcoxon signed rank test

WT: wild type

XLF: XRCC4-like factor

XPF: excision repair cross-complementing rodent repair deficiency, complementation group 4

XRCC4: X-ray cross complementing protein 4

Chapter 1 Introduction

1.1 Bladder cancer

Bladder cancer is the 9th most commonly diagnosed cancer worldwide, with the highest incidence rates evident in Europe (Olive & Banáth, 2006). It is a disease more common in men compared to women and is placed as the 4th most common cancer in men and the 11th most common cancer in women (Kamat *et al*, 2016). However, although bladder cancer occurrence is lower in women, females tend to present with more advanced disease and have a worse prognosis (Fajkovic *et al*, 2011). There is a higher incidence of bladder cancer in older patients, which poses issues with regard to treatment due to the prevalence of comorbidities. In the UK between 2012-2014 over half of all bladder cancer cases were aged 75 or over (Bladder cancer statistics | Cancer Research UK).

Bladder cancer neoplasms are defined according to their histological type. Approximately 90-95% of bladder cancers are classed as urothelial carcinomas (UC) (alternatively termed transitional cell carcinomas) and originate from the urothelial cells of the bladder lining (Martin *et al*, 2016).

UC neoplasms display a range of variant histologies which are identified by their distinctive cellular morphologies. Examples of morphological variant histologies in UC are micropapillary, nested, plasmacytoid, neuroendocrine and sarcomatoid. In addition to morphological variants, a proportion of UC show cellular differentiation within the neoplasm towards different epithelial types. This is termed divergent differentiation and is present in up to 20% of UC (Gofrit *et al*, 2016). In the tumours that show foci indicative of divergent differentiation, squamous differentiation is the most common

with glandular differentiation occurring in a lower proportion of cases (Amin, 2009). UC with squamous divergent differentiation should not be confused with a pure histology type of squamous cell carcinoma, where the tumour contains predominantly squamous features. In a UC with squamous differentiation, the predominant cell type is UC and only foci displays squamous features. Histological variants in UC have been found to associate with poor prognosis and increased risk of recurrence and progression when compared to non-variant counterparts (Seisen *et al*, 2014; Hsieh *et al*, 2015; Gofrit *et al*, 2016). However, a recent study by Krasnow *et al* (2017) demonstrated that in UC, where the tumour has invaded into muscle of the bladder, the presence of variants had no influence on patient outcome following treatment with bladder-sparing chemoradiation.

Several studies have been conducted into the molecular characterisation of UC and have identified recurring chromosomal alterations, point mutations and inactivating mutations in the disease. Furthermore, recurrent mutations have been found to occur frequently in genes involved in cell cycle regulation, chromatin regulation, and kinase signalling pathways (Gui *et al*, 2011; Lindgren *et al*, 2012; Weinstein *et al*, 2014).

Characterising the molecular profile of bladder UC has enabled the identification of two divergent pathways of tumorigenesis. One pathway is associated with low grade papillary tumours which are unlikely to progress to invade the muscle of the bladder. This pathway is marked by activating mutations in *FGFR3* (a fibroblast growth factor receptor involved in kinase signalling) and gene mutations that lead to a constituently active receptor tyrosine kinase-Ras pathway. The second pathway is associated with high grade, invasive tumours and is characterised by defects in the p53 and the

retinoblastoma protein (pRb) pathways (Wu, 2005; Goebell & Knowles, 2010). Recent work in the molecular profiling of bladder cancer has also led to the discovery of distinct bladder cancer subtypes in the muscle-invasive form of the disease; these are discussed in 1.1.2.

Apart from UC pure histological types of bladder cancer include squamous cell carcinoma that accounts for 2-5% of malignancies, adenocarcinoma (accounting for 0.5-2% of malignancies) and small cell carcinoma (accounting for <1% of bladder cancer cases) (Martin *et al*, 2016).

Bladder cancer tumours are classified according to TNM (tumour, node, metastasis) staging (Table 1.1), World Health Organisation (WHO) grading classifications (Table 1.2) and are stratified according to risk (Table 1.3). The grade of a tumour is based on its pattern of growth and cytological characteristics such as cell polarity and nuclear size. The growth of a tumour through the lining of the bladder into the muscularis propria is used to stage a neoplasm and is also utilised to separate bladder cancer into two types: non-muscle-invasive (NMIBC) and muscle-invasive (MIBC) (Table 1.1, Figure 1.1)

Table 1.1: **TNM staging of bladder cancer**. A description of the features used to categorise neoplasms to a specific T (TX-T4), N (Nx-N3) and M (M0-M1) stage. The growth the tumour through the bladder determines the T stage of a bladder neoplasm and is used to separate patients into NMIBC and MIBC types. The N and M stages assess the presence of lymph node invasion and distant metastasis (taken from Babjuk *et al*, 2017).

TNM staging classification of urinary bladder cancer		
T	pT	Stage determined through pathological examination
	NMIBC	
	TX	Primary tumour cannot be assessed
	T0	No evidence of primary tumour
	CIS (Carcinoma *in situ*)	Early high-grade cancer on the innermost lining of the bladder. A flat lesion that is always high-grade. High-risk of progression to muscle-invasive disease
	T1	Tumour invades subepithelial connective tissue has not yet invaded through to the muscle
	MIBC	
	T2	Tumour invades muscle T2a: Tumour invades superficial muscle (inner half) T2b:Tumour invades deep muscle (outer half)
	T3	Tumour invades perivesical fat tissue T3a: Tumour contained within perivesical fat tissue is only detectable under the miscroscope T3b: Tumour within the perivesical fat tissue is evident with macroscopic testing
	T4	Tumour has invaded surrounding organs: prostate, uterus, vagina, pelvic wall or abdominal wall T4a: Tumour invades prostate uterus or vagina T4b: Tumour invades pelvic wall or abdominal wall
N	NX	Regional lymph nodes cannot be assessed
	N0	No regional lymph node metastasis
	N1	Metastasis in a single lymph node in the true pelvis (hypogastric, obturator, external iliac, or presacral)
	N2	Metastasis in multiple lymph nodes in the true pelvis (hypogastric, obturator, external iliac, or presacral)
	N3	Metastasis in one or more common iliac lymph node(s)
M	M0	No distant metastasis
	M1	Distant metastasis

Table 1.2: **World Health Organisation (WHO) grading classifications**. A description of the criteria used in assessing the grade of a bladder neoplasm. Cytological characteristics and morphological features are used for grading a bladder neoplasm. The morphological pattern of growth of a neoplasm is described by papillary or flat lesions. (taken from Babjuk *et al*, 2017)

Grade 1 (G1)	Well differentiated cells (low-grade)	
Grade 2 (G2)	Moderately differentiated (low to high-grade)	
Grade 3 (G3)	Poorly differentiated (high-grade)	
Papillary lesions	-urothelial papilloma	-Non-cancerous (benign) tumour -Papillary urothelial neoplasm of low malignant potential (PUNLMP): slow growing cancer unlikely to spread -Low-grade papillary urothelial carcinoma: slow growing cancer that is unlikely to spread -High-grade papillary urothelial carcinoma: quicker growing cancer that is more likely to spread
Flat lesions		-Urothelial proliferation of uncertain malignant potential -Reactive atypia -Atypia of unknown significance -Urothelial dysplasia -Urothelial *CIS* (always high-grade)

Table 1.3: **Risk categorisation of NMIBC**. A description of the features that are used to categorise bladder neoplasms as low, intermediate or high risk (information obtained from Veeratterapillay *et al*, 2016 and Babjuk *et al*, 2017).

Low-risk	Primary, solitary, Ta, low-grade /G1, < 3 cm, no *CIS*
Intermediate-risk	All tumours not defined as low or high-risk
High-risk	T1/Tis, with high-grade/G3 tumours, concomitant *CIS* or large (>3 cm) Ta/T1 G1/G2 tumours

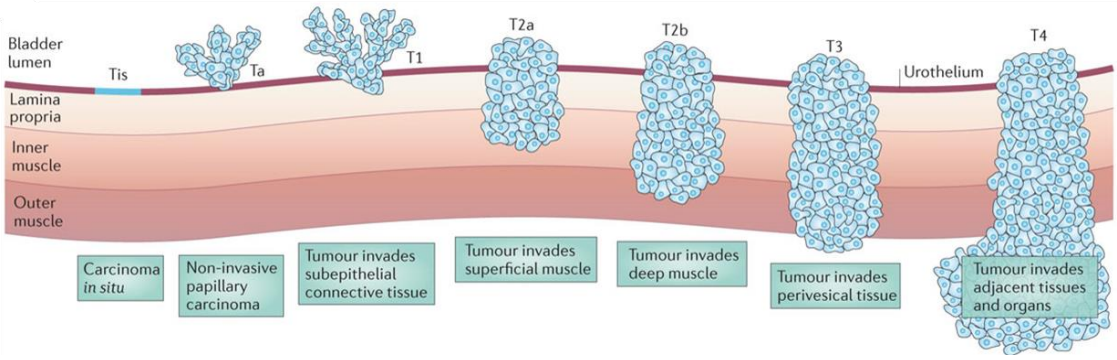


Figure 1.1: Bladder cancer staging. An illustration of the growth of a bladder neoplasm through the lining of the bladder and its corresponding T stage. CIS, Ta and T1 tumours are classed as non-muscle invasive as they are confined to the bladder lining. CIS is an early high grade, high risk flat lesion retained in the innermost lining of the bladder. Ta lesions are papillary carcinomas confined to the innermost lining of the bladder. T1 tumours show growth into the connective tissue beneath the bladder lining. T2, T3 and T4 tumours are classed as muscle-invasive. T2 tumours have grown into the muscle of the bladder but not through to the perivesical fat tissue. T3 tumours have invaded through to the perivesical fat tissue but are still confined to the bladder. T4 tumours have invaded through the layers of the bladder and have growth outside of the bladder in the prostate, uterus, vagina, pelvic wall or abdominal wall (reprinted from Knowles & Hurst, 2014).

NMIBC has a good prognosis with 5 year survival rates between 80 and 90%. By comparison MIBC has a significantly worse prognosis with only a 40-60% 5 year survival rate for disease stage T2-3 disease. Roughly 75% of bladder cancer patients will present at diagnosis with NMIBC with the remaining 25% of patients having either MIBC or metastatic disease. Only a small proportion of patients (10-20%) with MIBC are those who have previously been diagnosed with NMIBC and progressed to MIBC. Therefore, 80-90% of MIBC patients will be diagnosed with the muscle-invasive form of the disease at primary presentation.

1.1.2 Subtypes of MIBC

Recently studies have revealed distinct MIBC subtypes that are associated with genetic and clinicopathological features. Three independent groups have identified basal and luminal subtypes within bladder cancer which are analogous to the basal and luminal subtypes found in breast cancer (Choi *et al*, 2014b; Damrauer *et al*, 2014; Weinstein *et al*, 2014).

The basal subtype occurs more frequently in women and such tumours are intrinsically aggressive, having decreased rates of cancer-specific survival (CSS) and overall survival (OS). Related to this, basal MIBC are also associated with advanced disease stage and metastasis at presentation. Additionally, basal MIBCs are enriched with squamous features and sarcomatoid differentiation, a characteristic associated with epithelial-mesenchymal transition (EMT) (Choi *et al*, 2014a). The luminal MIBC subtype tumours have both high levels of the fibroblast growth factor receptor 3 (FGFR3) protein and activating mutations in the *FGFR3* gene, features frequently found in NMIBCs. This, in addition to an enrichment of papillary features in the histology of luminal MIBCs, has led to the idea that luminal MIBCs correspond to papillary NMIBCs that have progressed to become muscle-invasive. In contrast to the basal subtype, that displays EMT plasticity, luminal cancers are considered more epithelial in nature. Yet despite this, luminal tumours are still capable of lymph node invasion and metastasis (Choi *et al*, 2014a).

The basal and luminal subtypes can be further separated into two more subgroups that display a decreased sensitivity to chemotherapy. A basal claudin low subgroup has been identified that has intermediate sensitivity to chemotherapy. This group is characterised

by the expression of markers of EMT and tumours that have a strong mesenchymal signature compared to other basal tumours (Damrauer *et al*, 2014). Within the luminal subtype, a p53-like group has been found which is resistant to neoadjuvant chemotherapy. p53-like MIBCs have similar frequencies of p53 mutation to other subgroups of MIBC but are characterised by a gene expression signature similar to that seen with active p53 expression. p53-like tumours are not stable. Indeed a p53-like status was found to be induced in some luminal subtype tumours after exposure to neoadjuvant chemotherapy. p53-like MIBCs also show high levels of stromal and fibroblast infiltration (Choi *et al*, 2014b).

The distinct gene expression profiles and characteristics of recently defined MIBC subgroups have the potential to more accurately describe the heterogeneous features and behaviour of MIBC (Dadhania *et al*, 2016) and to inform treatment. It is already possible to distinguish basal and luminal MIBC using a simple technique called immunohistochemistry (IHC). Antibodies against two related cytoskeletal proteins, cytokeratin 5/6 (CK5/6) and cytokeratin 20 (CK20), are found to be expressed differentially in basal and luminal MIBC. Levels of CK5/6 expression are significantly higher in basal tumours, whereas, CK20 is more highly expressed in luminal tumours, with the abundance of the two proteins being inversely correlated (Choi *et al*, 2014b).

1.2 Treatment of Bladder Cancer

The treatment of NMIBC and MIBC differs. However, both NMIBC and MIBC are usually treated with curative intent (Figure 1.2)

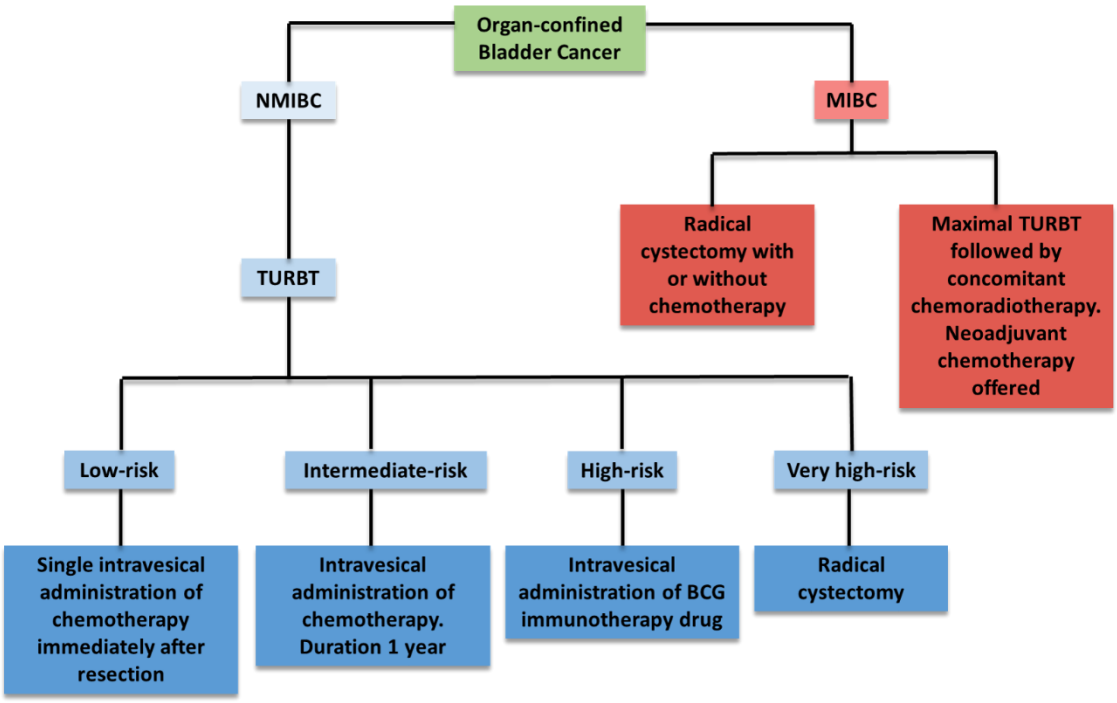


Figure 1.2: Management of organ-confined bladder cancer. Non-muscle invasive bladder cancer is treated with transurethral resection of the bladder tumour (TURBT) followed by either intravesical administration of chemotherapy or BCG, depending on risk categorisation. Very high-risk non-muscle invasive bladder cancer or patients who have failed BCG are offered radical cystectomy. Muscle-invasive bladder cancer is treated with either radical cystectomy or maximal TURBT followed by concomitant chemoradiotherapy (CRT), with or without neoadjuvant chemotherapy.

1.2.1 Management of NMIBC

NMIBC is treated initially with a transurethral resection of the bladder tumour (TURBT) using adequate surgical depth to include the muscularis propria. If high-grade pT1 disease is found, a second TURBT is recommended 4-6 weeks after the initial resection to remove residual tumour and for restaging, as the tumour may in fact be pT2 (Herr *et al*, 1999; Divrik *et al*, 2006; Hashine *et al*, 2016; Kamiya *et al*, 2016). In addition to

TURBT, an intravesical administration of chemotherapy after resection is also recommended. The frequency of chemotherapy is dependent on how the tumour is classified: low, intermediate or high-risk (Table 1.3). Low-risk NMIBC, such as low-grade Ta tumours, receive a single intravesical treatment of mitomycin (or more rarely epirubicin or gemcitabine) immediately after resection. For intermediate-risk NMIBC, such as multifocal or multi-recurrent low-grade Ta tumours, it is advocated that chemotherapy is maintained over the first year after TURBT. For high-risk NMIBC the recommended treatment is intravesical administration of Bacillus Calmette-Guerin (BCG) immunotherapy. For very high-risk NMIBCs, such as those with multiple and/or large high-grade T1 tumours, micropapillary histological variants, concomitant *CIS* present in bladder, prostatic urethra, the presence of lymphovascular invasion or patients who have failed BCG, radical cystectomy is offered (Kamat *et al*, 2016).

1.2.2 Management of MIBC Bladder Cancer

MIBC is routinely treated with either radical cystectomy (RC) or concomitant chemoradiotherapy (CRT), with or without neoadjuvant chemotherapy. Historically RC has been considered the gold standard treatment for MIBC. However, in recent years there has been an increased interest in the role of radiotherapy (RT). Both RC and RT show comparable rates of CSS (Kotwal *et al*, 2008; Booth *et al*, 2014). However, this comparison of treatment efficacy has not yet been verified within a randomised phase III trial.

In 2010 a multicentre feasibility pilot study (SPARE: Selective bladder Preservation Against Radical Excision (cystectomy)) was set up. The study was designed to determine if a phase III randomised clinical trial into the standard of care of patients with muscle-

invasive, T2/T3 transitional cell carcinoma of the bladder was possible. Three potential treatments, namely neoadjuvant chemotherapy, CRT and RC, were to be investigated in this study. Although the study had wide support within the uroncology community, it was unfortunately forced to close after 30 months due to insufficient patient recruitment (Huddart *et al*, 2010). The failure of the SPARE trial is a setback in directly comparing the effectiveness of RC to radiotherapy-based treatments and it may be that it will never be possible to compare these two treatment modalities in this manner.

1.2.3 Radical cystectomy

RC is the surgical removal of the bladder. In men, RC for MIBC includes removal of the prostate, seminal vesicles, and distal ureters. In women RC for MIBC involves anterior pelvic exenteration including the bladder, uterus, entire urethra, adjacent vagina and distal ureters (Gakis *et al*, 2013). The inclusion of bilateral pelvic lymphadenectomy is an important element in RC for MIBC and improves local tumour control. The extent of lymph node dissection is debated (Bruins *et al*, 2014), but a growing body of evidence suggests that extended lymph node dissection is associated with improved staging and outcome (Dhar *et al*, 2008; Tilki *et al*, 2013; Bi *et al*, 2014; Kiss *et al*, 2016).

Chemotherapy is offered in conjunction with RC. Neoadjuvant chemotherapy with cisplatin combination therapy is recommended for patients where possible, due to proven improvements in disease progression and OS (Vale, 2005b; Yin *et al*, 2016). The use of adjuvant chemotherapy with RC has also been found to be associated with improved survival in MIBC (Leow *et al*, 2014). Indeed, studies comparing outcomes following neoadjuvant or adjuvant platinum-based chemotherapy find no difference in OS or CSS (Millikan *et al*, 2001; Wosnitzer *et al*, 2012). Despite this, neoadjuvant

chemotherapy remains the preferred method of treatment based on evidence from phase III clinical trials, comprehensive meta-analysis and the general ability of patients to tolerate the effects of the treatment pre-surgery. Adjuvant chemotherapy is offered to patients who did not receive neoadjuvant chemotherapy, or were identified with invasive disease at the time of surgery (NICE guideline [NG2], 2015).

Although RC has been viewed as the definitive treatment for MIBC in the past, it has a number of disadvantages. Bladder cancer is often a disease of older patients, with nearly three quarters of patients aged >65 years at presentation (Manoharan *et al*, 2009). As a result, patients often carry a heavy burden of comorbidities which may preclude the use of cystectomy as a viable treatment option (Roth & Thalmann, 2015). Additionally, despite a significant reduction in the mortality resulting from RC over the years, the 90 day rates of mortality are still reportedly as high as 8-11% (Chahal *et al*, 2003; Liedberg, 2010; Aziz *et al*, 2014; Schiffmann *et al*, 2014; Larcher *et al*, 2015) and this rate is increased in elderly patients (Berger *et al*, 2014; Schiffmann *et al*, 2014). Another disadvantage of RC is the resulting significant peri- and postoperative complications (Liedberg, 2010; Lavallée *et al*, 2014; Djaladat *et al*, 2016; Krajewski *et al*, 2016; Sood *et al*, 2017). The most common complications of RC include: blood loss and the need for blood transfusions during surgery, gastrointestinal problems such as diarrhoea, constipation and ileus, deep venous thrombosis, pulmonary embolism and infection. A notable reduction in quality of life (QOL) is also reported by patients following RC mainly due to the presence of incontinence, sexual dysfunction and the need for a stoma, thereby altering body image (Caffo *et al*, 1996; Zahran *et al*, 2014; Kretschmer *et al*, 2016a, 2016b; Yang *et al*, 2016).

1.2.4 Radical Radiotherapy

The current standard regime of bladder preservation for MIBC comprises trimodality treatment (TMT) consisting of maximal TURBT followed by concurrent chemoradiation with 3D conformal or intensity modulated radiotherapy (IMRT) external beam radiation. Radiation regimes of 64 Gy in 32 fractions or 55 Gy in 20 fractions, administered in a 2 Gy per day fractionation and a hypofractionated moderately accelerated schedule, respectively, are considered standard in the UK.

The use of concomitantly administered chemotherapeutic agents in RT has been shown to improve local region control and OS following RT. Two of the largest and most recent randomised controlled phase III clinical trials into the impact of concomitant administration of chemotherapeutics on the efficacy of RT are BCON (Hoskin *et al*, 2010) and BC2001 (James *et al*, 2012). The key features and findings of these and other selected trials are summarised in Table 1.4.

Table 1.4 Clinical phase III trials into the inclusion of chemotherapy in RT-based treatment of MIBC. The number of participants, TNM staging of patients eligible for inclusion, aim and key findings of the trials are summarised.

Trial	Number of participants	Patient eligibility	Aim	Key findings
BCON (Hoskin *et al*, 2010)	333	T2-T4a, T1G3, N0M0	Compared the efficacy of RT plus concurrent carbogen and nicotinamide (RT+CON) to RT alone	OS at 3 years- RT+CON: 59% RT alone: 46% p=0.04 CR- RT+CON: 81% RT alone: 76% p=0.3
BC2001 (James *et al*, 2012)	360	T2-T4a, N0M0	Compared the efficacy of RT plus mitomycin C and 5-fluorouracil (RT+MF) to RT alone	LRC at 2 years RT+MF: 67% RT alone: 54% p=0.03 OS at 5 years- RT+MF: 48% RT alone: 35% p=0.16
BA06 30894 (International Collaboration of Trialists *et al*, 2011)	976	T2G3, T3, T4a, N0-NX, or M0	Investigated the use of neoadjuvant cisplatin, methotrexate, and vinblastine (CMV) chemotherapy in patients treated with either cystectomy and/or RT	OS at 10 years- With CMV: 49% No CMV: 30% p=0.037 Risk of death: 16% reduction with CMV
RTOG 89-03 (Shipley *et al*, 1999)	123	T2 to T4a, NXM0	Investigated the addition of neoadjuvant chemotherapy with CMV in patients receiving CRT (RT with concurrent cisplatin)	OS at 5 years- CRT+ neoadjuvant MCV: 48% CRT alone: 49% p>0.05 Percentage of patients with evidence of distant metastasis- CRT+ neoadjuvant CMV: 33% CRT alone: 39% p>0.05 Closed early due to high levels of severe complications

LRC: local regional control, OS: overall survival, RT: radiotherapy, CON: carbogen and nicotinamide, MF: mitomycin C and 5-fluorouracil, CMV: cisplatin, methotrexate, and vinblastine, CRT: chemoradiotherapy, CR: complete response rate

In contrast to concomitant chemotherapy, the benefit of the addition of neoadjuvant or adjuvant chemotherapy in TMT treatment of MIBC is debated. Although results from meta-analyses (Advanced Bladder Cancer Meta-analysis Collaboration, 2003; Vale, 2005b; Yin *et al*, 2016) and the BA06 30894 trial (International Collaboration of Trialists *et al*, 2011) (Table 1.4) have demonstrated a clear benefit for the addition of neoadjuvant chemotherapy in the treatment of MIBC treated with RT alone. There is currently no consensus whether neoadjuvant chemotherapy has any benefit in patients receiving concomitant CRT, as opposed to RT alone. Unfortunately the RTOG 89-03 phase III trial, which was aimed at investigating this, was closed early due to high levels of severe complications (Shipley *et al*, 1999). However, initial results found no significant benefit in survival when neoadjuvant chemotherapy was added to CRT treatment (Table 1.4). Furthermore, an the analysis of long term outcomes of TMT MIBC patients, treated at the Massachusetts General Hospital, identified no benefit to the addition of neoadjuvant chemotherapy to concomitant CRT treatment (Efstathiou *et al*, 2012).

The benefit of adjuvant chemotherapy after CRT of MIBC is less well defined than the use of neoadjuvant chemotherapy. Indeed a meta-analysis conducted into the benefit of adjuvant chemotherapy following CRT concluded that there was insufficient evidence on which to reliably base treatment decisions (Vale, 2005a). Consequently, the 2016 European association of urology (EAU) guidelines for the treatment of MIBC with TMT state that neither neoadjuvant or adjuvant chemotherapy seem to improve outcome (Alfred Witjes *et al*, 2017). Nevertheless, the UK National Institute for Health and Care Excellence (NICE) guidelines recommend offering neoadjuvant chemotherapy to all newly diagnosed MIBC patients for whom it is suitable (NICE, 2015)

The main benefit of using TMT in the treatment of MIBC is the improvement in patient QOL compared to RC (Henningsohn *et al*, 2002; Zietman *et al*, 2003; Mak *et al*, 2015, 2016). Patient QOL should not be forgotten in treatment choice. Indeed, in one QOL assessment of MIBC treatment, 46% of patients reported that they would be willing to accept some reduction in survival to become symptom-free (Henningsohn *et al*, 2002). Using TMT, the majority of patients retain their bladder, and normal bladder function is maintained in approximately 75% of patients (Henningsohn *et al*, 2002; Zietman *et al*, 2003). In addition to this, bowel function has either been found to be improved (Mak *et al*, 2016) or comparable following TMT compared to RC (Henningsohn *et al*, 2002). Sexual function is also significantly improved with TMT (Henningsohn *et al*, 2002; Mak *et al*, 2016).

1.3 Biomarkers in cancer

Cancer is a heterogeneous disease characterised by high levels of genomic instability, driving mutation and malignancy. It can safely be stated that each cancer patient will present with a primary tumour with its own unique array of mutations that have facilitated its progression. The mutations within the cells of a tumour will determine its response to treatment. Yet despite our understanding of this, the majority of cancer patients are often treated with a one-size-fits-all therapy.

The radiosensitivity of tumours is particularly determined by the genetic and molecular makeup of the neoplasm (Hennequin *et al*, 2008; Williams *et al*, 2010b; Skvortsov *et al*, 2014; Morgan & Lawrence, 2015). With 60% of all cancer patients receiving RT as part of their treatment (Orth *et al*, 2014), it is clear that establishing markers to predict patient response would be advantageous.

Prognostic biomarkers are able to determine the likely outcome of a patient, such as the probability of long-term survival or cancer recurrence, regardless of the type of treatment received. Conversely, predictive biomarkers are able to stratify patient populations into groups that are more likely to respond to a particular therapy (Mehta *et al*, 2010).

Predictive biomarkers have a clearly defined use in treatment decision making. However, as prognostic biomarkers offer information on outcome independent of treatment, their clinical use in treatment decisions is not as clear-cut. Prognostic biomarkers have been used to guide the development of targeted therapies. In breast cancer the HER2 biomarker was, historically, associated with higher mortality in early stage disease, reduced time to relapse, and an increased incidence of metastases (Slamon *et al*, 1987; Gonzalez-Angulo *et al*, 2009). The advent of targeted therapies against the HER2 tyrosine kinase receptor, such as trastuzumab, improved the prognosis of HER2 positive patients and consequently HER2 status is now a predictive biomarker for the efficacy of HER2 targeted therapy (Perez *et al*, 2014). Prognostic biomarkers could also be used to direct patients with a good prognosis away from toxic, aggressive treatments which would not be of benefit to these patients (Yue *et al*, 2016).

Synthetic lethality is an important concept in the development of anticancer therapeutics and its optimal application is reliant on biomarkers. Synthetic lethality occurs between two genes when a mutation in either gene alone is not sufficient to cause lethality, but simultaneous mutation in both genes does (Kaelin, 2005). In the context of cancer therapeutics, this means that if an oncogenic mutation occurs within a gene with a synthetically lethal interaction, then cytotoxicity within the neoplasm can

be achieved by using a targeted treatment against the interacting gene product. Chemical agents can be developed against interacting genes but for synthetically lethal drugs to be used optimally there is a need for markers, which can identify patients with the prerequisite mutations within their neoplasm.

Biomarkers can be DNA, RNA or protein-based and can be analysed within a variety of clinically relevant substrates such as blood, urine or tissue samples. Examples of DNA-based biomarkers are single nucleotide polymorphisms (SNP), chromosomal aberrations and methylation profiles. RNA-based biomarkers often involve evaluation of RNA transcript expression or the presence/abundance of regulatory RNAs. Protein-based biomarkers are represented by the presence/abundance of particular proteins and protein modifications such as phosphorylation or isoform expression.

A number of techniques can be used to quantify biomarkers. *In situ* methodologies, such as fluorescent *in situ* hybridisation (FISH) for DNA and IHC for proteins, are highly advantageous as they retain the morphological information contained in tissue samples. Furthermore, as *in situ* techniques are already in clinical use new findings, using these methods, are often more easily transferable to the clinic. Aside from *in situ* approaches, a variety of assays exist for assessing biomarkers. Such techniques include array-based hybridisation technologies, whole genome sequencing, RNAseq, mass spectrometry and quantitative PCR (qPCR).

A concern with biomarkers is the rate of false-positive or false-negative identifications they produce. False identifications occur due technical issues with the assays employed to evaluate the marker and produce results that do not accurately describe the biology of the patient. High rates of false classifications can lead to the over or under-treatment

of patients, which can negatively impact their outcome and increase exposure to unnecessary toxicities (Zhang & Chan, 2010; Garrison *et al*, 2015). Indeed, false classifications are a particular area of concern for the current clinical IHC-based biomarkers (Seferina *et al*, 2013; Kaufman *et al*, 2014).

The false classifications from IHC can arise from: delays in tissue fixation after sample acquisition, extended fixation times, the use of an inappropriate fixative, sample quality/size, batch effects in reagents, the use of different kits to conduct the assay and the reliance on the subjective and biased human-eye for analysis (Gown, 2008; True, 2008; Nuovo, 2016).

In the clinic, steps are taken to reduce the rate of false classifications arising from IHC. For example, in the assessment of the HER2 IHC biomarker in breast cancer, it is recommended that samples are placed in fixative less than one hour after acquisition. In addition, fixatives containing alcohol or Bouin's fixative are not used, as they can result in staining of normal tissue and preclude fluorescence testing, respectively (Rakha *et al*, 2015). It is also advised that if a negative result is found in a biopsy from a primary tumour, that the assay be conducted again on a corresponding surgical sample, to avoid false negatives resulting from small sample size (Duffy *et al*, 2017). In an effort to reduce the subjectivity and bias in the human-based analysis of IHC, HER2 IHC samples are assessed according to clear guidelines, with example control images to refer to (Duffy *et al*, 2017). In the UK laboratories performing clinical HER2 assessment are required to participate in the UK National External Quality Assessment Scheme for Immunocytochemistry and In Situ Hybridisation (UK NEQAS ICC & ISH) HER2 IHC and ISH modules (Rakha *et al*, 2015).

Although a concerted effort is being placed into the research of biomarkers for use in cancer, the vast proportion of markers identified in the research setting never make it to clinic. In the initial phases of biomarker development, the factors that contribute to this lack of translation are: the limited progression of retrospective studies to prospective analysis, the use of small sample numbers in studies and a lack of robust assay validation.

Studies into candidate biomarkers for clinical use can either be prospective or retrospective. Prospective randomized clinical trials are the gold standard for evaluating the utility of a biomarker. However, such studies are costly and take a considerable amount of time to achieve a sufficient length of patient follow up. Due to this, retrospective sample collections are often used to evaluate candidate biomarkers. The limitations of retrospective sample collections are that there is no control over the patient population being assessed or method of sample collection. Additionally, sample size is predetermined and reported patient data can be incomplete or inconsistent (Euser *et al*, 2009). What is more, differences in treatment procedures, such as the use of chemotherapeutics, can vary between retrospective cohorts, which may affect results.

Small sample numbers are a large source of stagnation in biomarker studies. The majority of the literature into potential biomarkers are from statistically underpowered studies on small sample numbers. This can lead to an inability to identify statistically significant results or, conversely, an over-interpretation of significant results occurring through chance (Hernández *et al*, 2014).

A lack of assay validation and standardisation also impedes the progression of potential biomarkers to clinic. This is because clinically relevant assays need to be proven to have specificity and to generate reproducible results (Drucker & Krapfenbauer, 2013). The failure to validate candidate biomarkers in validation studies is considered to be one of the main points of failure for many potential biomarkers (Hernández *et al*, 2014).

After a putative biomarker has been validated as reliable and technically robust, its adoption into clinical use is not ensured. The low uptake of a validated biomarker in the clinic may arise, if it is found to provide no additional information to a clinician over standard diagnostics or provide information which does not aid in treatment decisions (Alymani *et al*, 2010; Diamandis, 2012). Furthermore, the practicality of performing the assay to measure the biomarker may also affect its clinical utility. Procedures that demand a high level of resources in terms of labour, expensive reagents or specialised expertise not found within the health care system, may not be feasible for routine use (Sturgeon *et al*, 2010; Thariani *et al*, 2012). As medical costs continue to rise, in a landscape of decreasing resources, a biomarker needs to be cost effective or offer a distinct benefit to a patient, which would be unethical to ignore, for it to be accepted (Thariani *et al*, 2012).

1.3.1 DNA-based biomarkers

Detection of germline or somatic mutations can be used as DNA-based biomarkers. Mutations frequently occur within cells with little or no effect. However, if a mutation confers a selective advantage to the cell, in terms of survival or proliferation, this is can lead to tumorigenesis (Stratton *et al*, 2009). Oncogenic mutations are often used as

biomarkers, as they can be associated with the risk of developing cancer, the prognosis of the disease and the response of patients to therapies.

a) Germline mutations

Germline mutations occur in germ cells, which produce the gametes required for sexual reproduction. Therefore, a feature of a germline mutation is that it is heritable to the offspring of an individual (Stratton *et al*, 2009). As with all mutations, the effect of a germline mutation may be innocuous. However, some mutations, such as those seen within the breast cancer susceptibility genes (*BRCA1* and *2*), can predispose individuals to cancer.

Germline mutations within *BRCA1* and *2* are recognised biomarkers for the risk of developing breast and ovarian cancer (Berliner *et al*, 2007; Domchek *et al*, 2010). Mutations in either *BRCA1* or *BRCA2* are associated with a 70% increase in cumulative risk for the development of breast cancer (Giannakeas & Narod, 2017). In ovarian cancer carriage of inheritable *BRCA1* or *BRCA2* mutations are associated with a 44-63% and 30% increased cumulative risk for developing the disease, respectively (Dreyer, 2012). Due to these increased risks, women with germline *BRCA* mutations are offered preventative surgery, which reduces the risk associated with developing these cancers (Domchek *et al*, 2010). Additionally, the heritability of germline mutations enables this biomarker to be used to screen family members that may also carry greater risk of developing the disease if positive for the mutation.

For women where *BRCA* mutation is identified after the development of a primary breast cancer, many cancer centres offer immediate or post-chemotherapy bilateral mastectomy. This recommendation is made to reduce the risk of developing a

secondary primary breast cancer and is given regardless of the size or focality of the primary tumour the patient presented with. However, a recent publication by Copson *et al* (2018) into the outcome of 2733 young-onset breast patients, found that there was no significant difference in overall survival between *BRCA*-positive and *BRCA*-negative patients at 2, 5 or 10 year follow up. This result highlights the importance of using a biomarker optimally as it implies that patients with a good prognosis for their treated breast cancer, could delay preventative surgery until they are psychologically and physically recovered from their cancer treatment.

In recent years there has been particular interest in the of poly ADP ribose polymerase (PARP) inhibitors in the treatment of *BRCA*-associated cancers. The rational for this is based on the principal of synthetic lethality. Cells which are defective in BRCA1 and/or 2 are more sensitive to PARP inhibitors, compared to cells proficient for these proteins (Bryant *et al*, 2005; Farmer *et al*, 2005). PARP inhibitors are therapeutic agents that trap PARP on damaged DNA (Strom *et al*, 2011). The precise mechanism of synthetic lethality between PARP and *BRCA* mutation is still debated. However, one model proposes that the trapping of PARP on DNA causes the perturbation of replication forks. To avoid the toxicity of these lesions the activities of BRCA1 and BRCA2 are then required for the resolution of obstructed replication forks (Helleday, 2011; Strom *et al*, 2011)

The results of preclinical and clinical studies have demonstrated the efficacy of PARP inhibitors as monotherapeutic agents in hereditary *BRCA*-mutated ovarian and breast cancers. Phase I and II trials have identified improvements in progression free survival in these patients treated with PARP inhibitors, compared to the placebo, and phase III studies are ongoing (Evans & Matulonis, 2017; Ponec *et al*, 2017). For breast cancer, the

findings of the OlympiAD phase III trial were recently reported and found significantly improved progression free survival in *BRCA*-mutated, HER2 negative metastatic breast cancer patients treated with Olaparib compared to standard single agent chemotherapy (Robson *et al*, 2017).

b) Somatic mutations

Somatic mutations occur in somatic cells, which are not involved in sexual reproduction. As such somatic mutations are not inherited by the offspring of the individual. However, through mitosis, somatic mutations can be perpetuated within the individual into a population of cells which carry the mutation. Through the process of tumourogenesis tumours accumulate a range of somatic mutations (Stratton *et al*, 2009), which can be exploited as biomarkers.

Metastatic colorectal cancer (mCRC) provides an example of how somatic mutations within a cancer can have predictive value for targeted molecular therapies. Monoclonal antibody based anti-EGFR therapies, such as cetuximab and panitummab, prevent the activation of EFGR, inhibiting the stimulation of the downstream RAS/RAF/MAPK and PI3K/PTEN/AKT signalling pathways. In mCRC activating mutations in codons 12 or 13 within exon 2 of the *Kirsten rat sarcoma (KRAS)* gene occur in 33-48% of patients (Waring *et al*, 2016). *KRAS* is a pivotal downstream effector of EGFR and hence the constitutive activation of this protein negates the need for EGFR stimulation for signal transduction to occur. Consequently, in these tumours there is a reduced response to EGFR targeted therapies. Indeed, evidence from clinical trials have demonstrated that mutated *KRAS* is a predictor of non-response to anti-EGFR based therapies in mCRC (Amado *et al*, 2008; Karapetis *et al*, 2008; Douillard *et al*, 2010).

Recently, retrospective analyses of a number of randomised control trials have provided evidence supporting an extension of the KRAS test, to include a larger number of codons and analysis of the *neuroblastoma rat sarcoma (NRAS)* gene. The Association of Clinical Pathologists Molecular Pathology and Diagnostics Group state that RAS analysis should include at least *KRAS* codons 12, 13, 59, 61, 117 and 146 and *NRAS* codons 12, 13, 59 and 61 (Wong *et al*, 2014). However, despite these updated recommendations, a recent assessment of four UK National External Quality Assessment Service (NEQAS) schemes for genetic analysis of colorectal cancer, found delayed implementation of extended RAS testing in clinical application (Richman *et al*, 2017).

Somatic mutations in *KRAS* predict resistance to anti-EGFR antibodies in mCRC, and is one of the most successful examples of a biomarker that had been adopted into routine clinical use (Waring *et al*, 2016). That being said, there remains a large proportion of wild type *RAS* patients that do not benefit from anti-EGFR based therapies. Mutations in *BRAF*, which encode a major effector of RAS proteins, have been investigated for their potential to predict patient response to targeted EGFR therapies. However, while *BRAF* has been identified as having a prognostic role in mCRC, its use as a predictive marker for anti-EGFR is less clear (Friedrich *et al*, 2016; Van Cutsem *et al*, 2016; Loree & Kopetz, 2017). The factors in mCRC that provide resistance to anti-EGFR therapy are currently under investigation.

c) Epigenetics

In addition to genomic alterations, cancer cells exhibit aberrant epigenetic profiles. Global changes in DNA methylation and altered patterns of histone modifications are present in the cancer epigenome, resulting in the abnormal repression or expression of

genes (Esteller *et al*, 1998; Corn *et al*, 2000; Paz *et al*, 2003; Fraga *et al*, 2005; Seligson *et al*, 2005; Brait *et al*, 2008; Hon *et al*, 2012; Zelic *et al*, 2015; Karsli-Ceppioglu *et al*, 2017). As such, epigenetics offers an exciting area for the discovery of cancer biomarkers. DNA methylation, particularly promoter-associated hypermethylation, has been the most extensively researched form of epipgenetic alteration in cancer to date. Aberrant methylation of DNA occurs in two forms, hypomethylated or hypermethylated. Cancer cells exhibit a global hypomethylation of 5-methyl-cytosine as well as hypermethylation in specific promoter regions of DNA (Rodríguez-Paredes & Esteller, 2011).

The abnormal hypermethylation of DNA in the promoter regions of genes involves CpG islands, which are short (typically 1000bp) interspersed DNA sequences of CG-rich DNA. In vertebrates 70% of genes are associated with CpG islands in their promotor regions, and in an un-methylated state CpG islands are linked to a transcriptionally permissive chromatin state (Deaton & Bird, 2011). In cancer the hypermethylation of CpG islands is associated with transcriptional repression, frequently occurring in tumour-suppressor genes (Sandoval & Esteller, 2012).

Investigations have identified DNA methylation as having prognostic or predictive value in a range of cancers including, but not limited to, prostate (Strand *et al*, 2014; Geybels *et al*, 2015), pancreas (Jiao *et al*, 2007; Kisiel *et al*, 2015), bladder (Kandimalla *et al*, 2013; Xylinas *et al*, 2016) and ovary (Chaudhry *et al*, 2009; Giannopoulou *et al*, 2017).

In glioblastoma, methylation of the *O6-methylguanine-DNA methyltransferase (MGMT)* promoter is of clinical importance. MGMT is a DNA repair enzyme that removes alkyl

adducts from the O^6 position of guanines and its expression is heavily regulated by methylation-dependent epigentic silencing (Everhard *et al*, 2009).

Clinical studies have reported *MGMT* promotor methylation to be a strong prognostic marker in glioblastoma, associated with prolonged survival, and also to be a predictive biomarker for patient response to DNA alkylating agents (Hegi *et al*, 2005; Stupp *et al*, 2009; Wick *et al*, 2012; Zhang *et al*, 2013; Barbagallo *et al*, 2014; Binabaj *et al*, 2018).

1.3.2 RNA-based biomarkers

RNA has great potential as a source material for biomarker identification as it sits between the genetic information encoded at the DNA level and functional information retained on the protein level. Therefore, RNA holds information about transcriptional regulation and, through microRNA or isoform analysis, can also hold clues as to post-transcriptional regulation and protein expression. Thanks to techniques such as PCR, RNA analysis can be conducted on a very small quantity of starting material. Furthermore, next generation sequencing technology has enabled researchers to more accurately detect subtle variations in RNA transcripts, for both protein coding and non-coding RNAs, such as small nuclear RNA, miRNA and long noncoding RNA. Consequently, over recent years there has been an increase in the number of studies investigating RNA biomarkers in cancer. However, a major limitation of using RNA as a biomarker is its instability.

Oncotype DX is an approved RNA-based biomarker to help guide decisions about chemotherapy after surgery in some patients with oestrogen receptor positive (ER+), lymph node negative (LN−) and human epidermal growth factor receptor 2 negative (HER2−) early breast cancer (Paik *et al*, 2004).

The test uses DNA microarray-based technology that hybridises RNA extracted from FFPE tissue to assess the expression of 21 genes. Using this data Oncotype DX is able to assign a recurrence score to the patients, which relates to the likelihood of distant recurrence at 10 years (Paik *et al*, 2004). In retrospective and prospective analysis, information provided by Oncotype DX has been found to successfully assess patient prognosis and identify patients who are likely to benefit from adjuvant chemotherapy (Paik *et al*, 2006; Albain *et al*, 2010; Sparano *et al*, 2015; Gluz *et al*, 2016; Petkov *et al*, 2016). When used in combination with traditional histopathology and established protein biomarkers Oncotype DX provides additional information for the treatment of breast cancer which can aid in treatment decisions and prevent the overtreatment of patients.

Until recently, it was not possible to analyse RNA *in situ* and hence RNA would have to be extracted from tissue for analysis. Extraction removes valuable morphological information and, if not controlled for by microdissection, can be contaminated by unwanted tissue artefacts such as fibrosis, necrosis or infiltrative immune cells. The development of RNAscope has enabled RNA to be assessed within the context of a tissue section (Wang *et al*, 2012). Although, currently prices are prohibitive for wide spread use of this technology in biomarker screening.

1.3.3 Protein-based biomarkers

Proteins are highly useful for biomarker development. They are diverse, have a dynamic turnover and some can be secreted into blood and bodily fluids (Drabovich *et al*, 2015).

The most basic analysis of a protein as a biomarker is the evaluation of protein abundance. Although simplistic, this type of biomarker determination has provided a

number of validated biomarkers currently used in the diagnosis and treatment of cancer patients. In addition to protein abundance, post-translational protein modifications, usually evaluated by mass spectrometry, can also be used in the identification of biomarkers and dramatically increases the number of potential targets for investigation. It is estimated that there are up to 1.8 million different proteoforms when accounting for modifications (Nørregaard Jensen, 2004). Such a large pool of potential candidate molecules increases the chances of finding relevant biomarkers. However, such a volume may also provide analytical challenges.

The treatment of breast cancer is one of the best examples of the utility of protein biomarkers in the management and diagnosis of cancer. Expression of oestrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor 2 (HER2) and the proliferation marker Ki67 are routinely used in breast cancer. The determination of the protein abundance of these biomarkers is performed using automated IHC. A request to the UK NEQAS ICC & ISH confirmed that within their external quality assessment scheme, all UK clinically testing laboratories used automated IHC platforms for the determination of breast cancer biomarkers.

In the clinic, the determination of ER and PR status by IHC is used to predict the response of breast cancer patients to hormonal therapies such as tamoxifen (Duffy *et al*, 2017). HER2 IHC is used to predict patient response to HER2-targeted therapies such as trastuzumab. It is noteworthy that although HER2 can be assessed through protein abundance or gene amplification, in the UK, HER2 testing is usually performed using IHC to detect protein abundance, with assessment of gene amplification reserved for equivocal cases (Rakha *et al*, 2015).

ER, PR and HER2 status are all predictive biomarkers for targeted therapy options. However, the absence of all of these markers in a tumour can also be used as a prognostic biomarker. The lack of all three of these receptor molecules defines triple negative breast cancer (TNBC). Approximately 10-20% of breast cancers are triple negative and are associated with poor prognosis, high risk of relapse, short progression-free survival (PFS) and overall survival (OS) (Costa & Gradishar, 2017). Due to the lack of target receptors TNBC does not respond to hormonal or HER-2 targeted therapies.

Ki67 is a marker of cellular proliferation present in all phases of the cell cycle except G_0 (Scholzen & Gerdes, 2000). It is a prognostic biomarker in breast cancer with high levels of Ki67 associated with poor outcome in breast cancer patients (de Azambuja *et al*, 2007; Stuart-Harris *et al*, 2008; Petrelli *et al*, 2015). The optimal method and interpretation of Ki67 in clinical practice is still the subject of debate with a high level of inter-laboratory variability. Despite these issues a consistent relationship between higher Ki67 values and poor patient outcome have been identified in studies (Petrelli *et al*, 2015). Due to this and the low cost and availably of the IHC assay for testing Ki67, in breast cancer Ki67 IHC is frequently conducted (Duffy *et al*, 2017).

Genetic tests and profiling assays for breast cancer are currently being researched. However, to date the IHC of ER, PR, HER2 and Ki67 protein biomarkers in tissue biopsies remains the gold standard to guide treatment plans for this disease (Wu *et al*, 2015).

DNA and RNA biomarkers are clearly useful in aiding cancer diagnostics and guiding treatment decisions. However, it can be argued that the effects of alterations at the DNA or RNA level often lead to changes in protein expression, form or modification.

Additionally, DNA and RNA-based biomarkers cannot detect post-translational protein modifications which are important factors in protein turnover and function.

1.4 IHC in biomarker identification

IHC is a widely used technique for biomarker identification. The method uses antibodies to specifically recognise proteins within sections of either FFPE or frozen tissue. It results in stained tissue sections where the quantity of staining directly relates to the amount of protein in the tissue. IHC can either be chromogenic or fluorescent. This thesis will focus on chromogenic IHC.

There are five main steps in chromogenic IHC staining: epitope retrieval, blocking of nonspecific endogenous reactivity within the tissue, primary antibody incubation, amplification of the primary antibody signal and immunodetection (Figure 1.3). The three principal methods used to amplify the primary antibody signal are: the avidin-biotin complex (ABC), labelled streptavidin-biotin (LSAB) and polymer methods. The ABC method uses secondary antibodies that are conjugated to biotin molecules. Biotinylated antibodies link the primary antibody with tetrameric avidin-biotin-complexes that contain a reporter enzyme. The reporter enzyme within the avidin-biotin complex is then able to catalyse the conversion of a chromogenic substrate to a coloured precipitate. The LSAB method used the same principle as the ABC method with streptavidin in place of avidin. The polymer method uses a polymer backbone that is conjugated to secondary antibodies and reporter enzymes to enhance the primary antibody signal. The most common reporter enzymes used in chromogenic IHC are horseradish peroxidase (HRP) and alkaline phosphatase (AP). HRP catalyses the oxidation reaction of 3,3'-diaminobenzidine tetrahydrochloride (DAB) into a brown

insoluble precipitate whereas AP catalyses the hydrolysis 3-amino-9-ethylcarbazole (AEC) into a red insoluble precipitate.

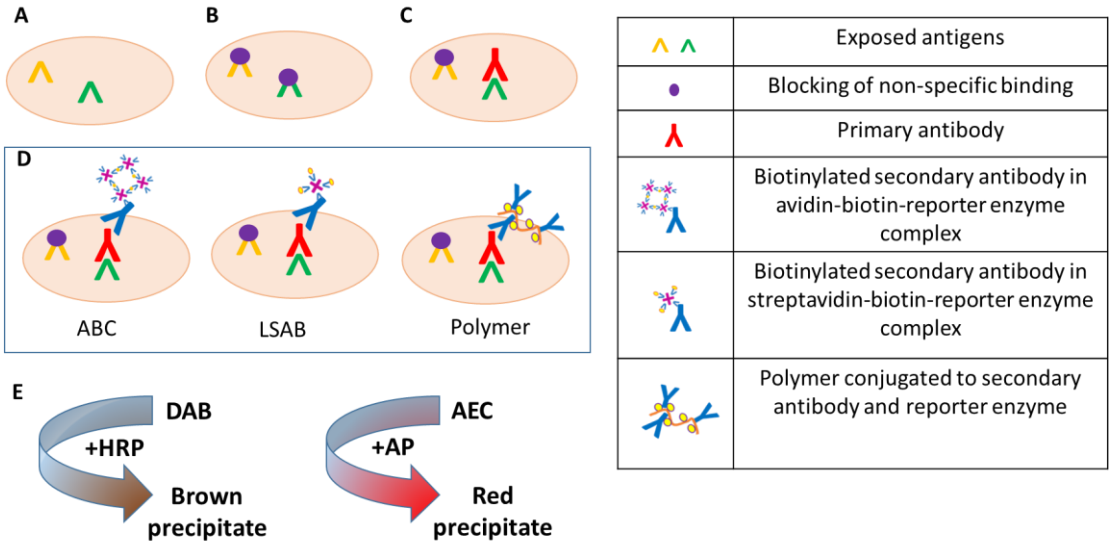


Figure 1.3: **A schematic description of chromogenic IHC**. (A) Epitope retrieval step. Heat or enzymatic action is used to remove methylene bridges formed during fixation, exposing antigenic sites; (B) Blocking step of antigenic sites to prevent non-specific binding or peroxidase activity; (C) Incubation with primary antibody; (D) Amplification of the primary antibody signal. ABC, LSAB and polymer signal amplification are schematically represented; (E) Immunodetection. HRP or AP are applied to samples where they catalyse the oxidation of 3,3'-diaminobenzidine tetrahydrochloride (DAB) into a brown insoluble precipitate or the hydrolysis 3-amino-9-ethylcarbazole (AEC) into a red insoluble precipitate, repectively.

IHC can be conducted manually or using automated platforms. Manual IHC protocols are still widely used in the research context. However, they are time consuming, laborious and the number of samples stained at any one time is limiting. In recent years the use of automated IHC has grown in popularity. The automation of IHC has dramatically increased the efficiency of the technique enabling many more samples to be easily stained simultaneously. In addition to increased efficiency, automated IHC can also increase the reproducibility of results and reduce staining variation due to batch effects (Warford *et al*, 2004). As a consequence, in the clinical setting automated IHC is

primarily used for performing such analysis. Improvements in the reliability of IHC are highly important for the use of this technique in a clinical setting, as erroneous staining in this context could lead to suboptimal treatment of a patient.

Another advancement in IHC efficiency has come in the form of tissue microarrays (TMA). In classical IHC, a section of embedded tissue is cut, placed on a glass slide and stained. However, with TMAs small cylindrical cores of tissue are removed from multiple embedded tissue samples and re-embedded in an array within a blank paraffin cassette. Sections of the array can be cut and, as for whole tissue sections, placed on a glass slide for staining (Figure 1.4). However, instead of one sample being contained on the slide, it is possible for hundreds of samples to be represented. As IHC is a multistep process that is sensitive to changes in incubation times, temperature and PH, unless it is rigorously performed there is a risk of staining variability between sample batches (O'Hurley *et al*, 2014). The simultaneous staining of multiple cores on the same slide ensures that conditions between tissue samples are standardised and reduces the need to conduct IHC in batches. Hence TMA's increase not only the throughput of IHC, thus increasing the feasibility of large scale studies, but also the reliability of results (Camp *et al*, 2008; Voduc *et al*, 2008).

The size of tissue cores used in TMAs range between 0.6 mm to 2.0 mm (Kampf *et al*, 2012). Due to these small sizes, when TMAs first came into use concerns were raised over how representative the results would be compared to those obtained from whole tissue sections. Since then numerous studies have demonstrated that results from IHC on two TMA cores sized ≥0.6 mm are comparable to those obtained from whole tissue sections (Camp *et al*, 2008; Voduc *et al*, 2008). Although, it has been identified that

tissues and markers with heterogeneous or location-dependent expression may require an increased number of tissue cores per patient to be representative (Iakovlev *et al*, 2007; Linderoth *et al*, 2007; Permuth-Wey *et al*, 2009; Allott *et al*, 2016).

A potential problem with TMAs is loss of usable tissue cores. The physical loss of cores can occur during sectioning, the transfer of sections to glass slides and staining procedures. It is estimated that 10–30% of cores are lost from TMAs is due to these technical reasons (Voduc *et al*, 2008).

The presence of artefacts, resultant from tissue extraction, handling or processing, or the presence of only a small amount of tumour pathology within a neoplasm, can also reduce the number of usable cores within a TMA. In a study comparing breast cancer whole tissue sections and TMAs, containing 157 cores, Gillett *et al* (2000) observed that 12% and 13% of cores floated off during IHC staining for PR and ER receptors, respectively. Additionally, 10% and 12% of the cores did not contain invasive tumour. A study by Hager *et al* (2007) specifically investigated the rate of core loss in renal cell tumour TMAs and the factors responsible for core exclusion. They identified that core exclusion was predominantly due to physical core loss, <25% tumorous tissue per core, core folding and the presence of necrosis in the core. Moreover, it was found that percentage of core loss and reduction of unusable cores tripled, from the first slice to the last slice of the TMA block.

The loss of tissue cores from a TMA reduces the number of tissue cores per patient. This can reduce the representativeness of the TMA and its ability to be used to accurately determine associations with outcome in biomarker research (Eckel-Passow *et al*, 2010;

Khouja *et al*, 2010). Though there is a need to be aware of the limitations when using TMAs, their use is now well accepted for research purposes.

The ability of TMAs to be representative is a consideration when studying IHC biomarkers in bladder cancer. Bladder cancer displays intra-tumour, genetic, histologic and functional heterogeneity (Amin, 2009; Morrison *et al*, 2014; Nordentoft *et al*, 2014; Gerlinger *et al*, 2015). Currently, there is a limited amount of research into the heterogeneity of protein expression within bladder cancer neoplasms. However, proteins have been reported to show heterogeneity within tumours of the bladder, by either IHC or through analysis of mRNA transcripts (Agerbaek *et al*, 2006; Laé *et al*, 2010; Taylor *et al*, 2014; Zhang *et al*, 2016; Thomsen *et al*, 2017).

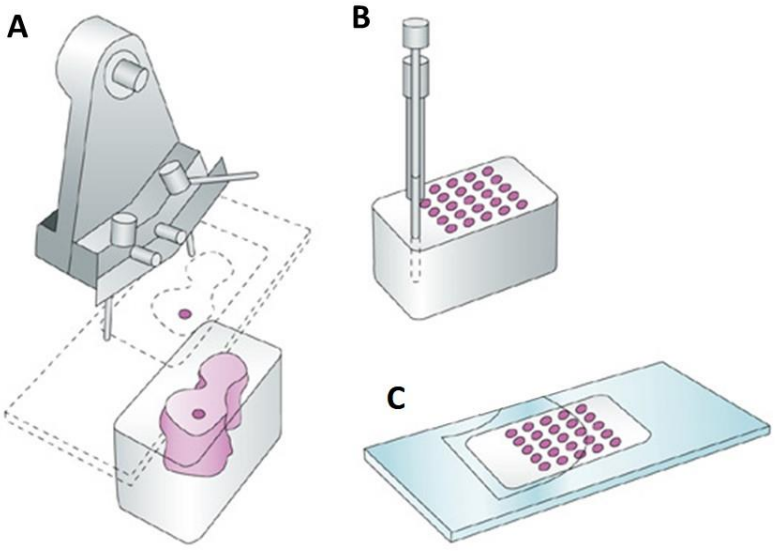


Figure 1.4: **The manufacture of a TMA**. (A) Cylindric tissue cores 0.6-2.0 mm in diameter are removed from a conventional FFPE tissue block; (B) Tissue cores are placed into pre-made holes of a blank paraffin block in a predesigned array; (C) Sections of the TMA block are cut and placed on glass slides for staining procedures (from Sauter *et al*, 2003).

There are many advantages of using IHC for biomarker identification. Firstly, IHC is already employed in clinical histopathology departments. Therefore, the equipment, expertise and quality control regulations required for IHC analysis of biomarkers are already established, and newly characterised IHC biomarkers are arguably more easily translatable than biomarkers identified using other methods (O'Hurley *et al*, 2014). Another factor in favour of IHC in the clinical environment is that it is relatively inexpensive to conduct, making it more feasible for regular use (Howat *et al*, 2014). IHC biomarkers use tissue as their substrate. While the resection of cancerous lesions is invasive, it is routinely carried out for the purposes of diagnosis. Hence, acquiring the starting material for IHC is not a major issue. For the purposes of biomarker discovery, large archives of FFPE tissue are available (Voduc *et al*, 2008).

One criticism of traditional manual IHC procedures is that they are laborious, time consuming and not suited to large sample volumes. However, advances in the automation of IHC staining protocols have dramatically increased the efficiency and reproducibility of the procedure. The diversity of antibodies available for specific protein epitopes and protein modifications has also influenced the use of IHC for biomarker discovery. The biggest advantage of using an IHC-based modality for biomarker discovery is the capacity for IHC to retain the morphological context of a sample (de Matos *et al*, 2010). Using IHC, protein expression can be determined and observed in relation to pathological features of the tissue, which adds valuable information about the cancer.

As with all techniques, IHC has its disadvantages. Automation of IHC staining has vastly improved the efficiency and consistency of the staining procedure. However, the gold

standard of IHC analysis is still conducted by eye (manual scoring), by trained researchers or clinicians. The manual scoring of IHC data is time consuming and is a significant bottleneck in the efficiency of IHC analysis. It is also liable to bias due to the innate limitations of human judgement and can never be absolutely quantitative (Okoye & Nnatuanya, 2015).

The use of automated digital image analysis in histopathology is an area of considerable interest for cancer therapeutics and IHC biomarkers. Computer assisted image analysis has the potential to address many of the long-standing limitations associated with the current human-based analysis of histopathology. Digital image analysis is by its nature objective and not subject to ocular phenomena that can trick the human eye (Aeffner *et al*, 2017). It can also be standardised and contain algorithms to correct for variances that may occur in tissue staining between batches, or reagents and improve contrast (Hoffman *et al*, 2014; Kothari *et al*, 2014; Kumar *et al*, 2015). Furthermore, due to the use of colour deconvolution in automated image analysis, to quantify the contribution of colours to staining intensity or saturation, it can be truly quantifiable as opposed to semi-quantitative (Pham *et al*, 2007; Gurcan *et al*, 2009). Importantly, considering the increased demand for trained histopathologists, digital image analysis has the potential to improve the efficiency of pathology diagnostics.

Basic computer algorithms are available to analyse digitally scanned tissue specimens. However, these are often instrument-specific, which limits their universal utility. Basic algorithms for the automated analysis of histopathology require human input, to configure the algorithm for a specific tissue type and biomarker under investigation. This human input can reintroduce the biases associated with manual analysis and can

be time consuming. As of yet there is no basic algorithm that does not require human intervention, of a comparable standard to human scoring (Di Cataldo *et al*, 2012; Levenson *et al*, 2015).

In recent years deep learning and artificial neural networks have been used in the development of digital image analysis programmes. Deep learning and artificial neural networks, in computer programming, refer to a machine learning technique whereby a computer model learns and adapts to new data. Due to the complexity of microscopic morphology and need for objectivity in image analysis, deep learning is ideally suited to digital pathology, as it can use complex computer model architectures to learn complex features from data in an unsupervised manner (Madabhushi & Lee, 2016). The results of deep learning methods in digital pathology have been encouraging generating tissue classifications concordant with those produced by pathologists. However, the need for large datasets on which to train deep learning algorithms currently limit the development of this technology (Robertson *et al*, 2017). Additionally the knowledge gap between the fields of deep machine learning and digital histology also creates a burden of entry (Janowczyk & Madabhushi, 2016) .

The work conducted by Levenson *et al* (2015), using pigeons to assess pathology and radiology images, could be a solution to the large datasets required to train machine learning programmes. Pigeons share many visual system properties with humans and it was found that, when trained, the birds rapidly learned to discriminate the appearance of benign from malignant breast tissue histology to a high level of accuracy. This led the proposition that pigeons could be used as surrogates for human observers in certain medical image perception studies. The use of pigeons would avoid the need to recruit,

pay and retain already heavily burdened clinicians. Pigeons would also be able to examine many more cases for a study than humans, which would increase the statistical power.

Crowdsourcing is another possible solution to the training data required for machine learning programmes in digital pathology. A number of projects have reported success in using members of the public to perform simple scientific tasks, such as morphological classifications. These studies have been conducted in a range of scientific disciplines, including biochemistry and biomedical research (Lintott *et al*, 2008; Cooper *et al*, 2010; Kawrykow *et al*, 2012; Lee *et al*, 2014; Candido Dos Reis *et al*, 2015; Rallapalli *et al*, 2015; Lawson *et al*, 2017) and have led to both *bona fide* scientific discoveries and generated improvements in computational algorithms (Curtis, 2014). Two crowdsourcing studies, Cell Slider and Trailblazer, have been conducted to determine the ability of members of the public to analyse IHC data. Both studies reported encouraging results, demonstrating an ability for members of the public to accurately score IHC samples (Candido Dos Reis *et al*, 2015; Lawson *et al*, 2017). The use of crowdsourcing for IHC scoring is still in need of further research and validation but has potential to produce the large sets of data required for advancement of digital image analysis programmes.

The development of deep learning image analysis programmes for pathology diagnostics is still in its infancy and to date, the standard of care in the clinic is to continue human-based specimen analysis. For computer assisted image analysis to be adopted within the clinical environment, the technique requires validation on clinically relevant samples in collaboration with clinical pathologists.

An important consideration for IHC in biomarker discovery is the requirement for standardisation and robust validation. Each of the major steps in IHC can significantly alter staining if modified. While this allows for effective optimisation, if staining protocols have previously been insufficiently optimised, validated or standardised, erroneous results that are not representative of protein abundance can be obtained. Therefore, although IHC is a powerful tool for biomarker discovery, to be clinically relevant, antibodies and protocols should be robustly validated. Additionally, if using FFPE tissue it is important to understand that for some epitopes the duration of fixation in formaldehyde can be critical (de Matos *et al*, 2010; Dunstan *et al*, 2011; Howat *et al*, 2014; O'Hurley *et al*, 2014).

New targets for cancer biomarker screening are based on the molecular characterisation of tumours and cell lines. The DNA damage response (DDR) pathway is a large source of cancer biomarkers candidates. This is due to the dysregulation of DNA damage signalling pathways that is required by cells for tumourigenesis to occur (Li *et al*, 2013). The mutations that lead to this dysregulation can directly affect tumour aggression and hence prognosis. Additionally, as the majority of non-surgical cancer therapies are based on the induction of DNA damage, changes within the DDR pathway can also determine the efficacy of a treatment. Therefore, understanding the complex interactions of the DDR pathway is highly important in biomarker discovery.

1.5 Cellular repair of double-stranded DNA damage

DNA double strand breaks (DSB) are harmful to genomic integrity and can lead to genomic rearrangements, mutation and cell death. Indeed, for radiation induced DNA damage, while the majority of breaks produced are single stranded, it is the DSB that are responsible for cell death (Jackson & Bartek, 2009). Cells have evolved complex interconnected pathways to respond to and repair DSB. The first responders to DSB are ataxia telangiectasia mutated (ATM) and the MRN protein complex (Lavin, 2007). The MRN complex is a multifunctional protein complex composed of meiotic recombination 11 (MRE11), Nijmegen breakage syndrome 1 (NBS1) and RAD50. The MRN complex detects DSB and is responsible for enhancing the rate of ATM activation (Carson *et al*, 2003; Lee & Paull, 2004, 2005; Dupré *et al*, 2006). Upon activation, ATM induces the signalling cascade that mobilises DNA repair pathways, cell cycle checkpoints and apoptosis (Matsuoka *et al*, 1998; Weber & Ryan, 2015) (Figure 1.5).

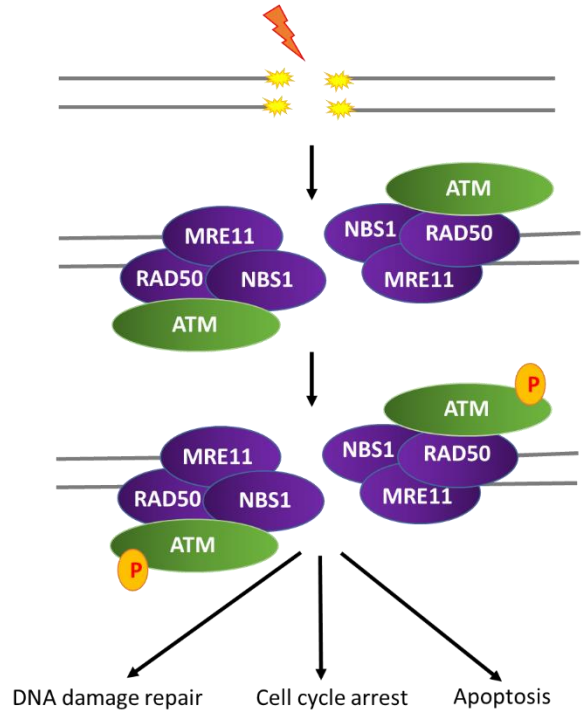


Figure 1.5: **A schematic representation of the recognition of a DSB**. ATM and MRE11 are recruited to DSB where upon MRE11 binds DNA ends and enhances the rate of ATM activation. Activation of ATM induces signalling cascades resulting in the activation of DNA damage repair pathways, cell cycle checkpoints and apoptotic pathways.

A key target of activated ATM is the tumour suppressor protein p53. p53 responds to cellular stress such as unauthorized cell proliferation, growth factor or nutrient deprivation and DNA damage. *TP53* (the gene that encodes p53) is the most commonly mutated gene in human cancer and mutations have been identified in all regions of the gene. The most common forms of mutation in *TP53* are missense substitutions which lead to a stably expressed protein, albeit with a loss of wild-type activity. Indeed, tumours with mutated p53 often display an abnormally high level of cellular p53. In addition to the loss of the p53 tumour suppressor function, cancer associated p53 mutations have been found to have a gain of function that aid tumour progression (Yue *et al*, 2017).

p53 functions as a transcription factor that binds as a tetramer to response elements in the promotor region of multiple genes, activing or repressing their transcription (Wang *et al*, 2010). Through this function p53 regulates a diverse group of biological activities including cell cycle progression, apoptosis, senescence, DNA metabolism, energy metabolism, angiogenesis, immune response, cell differentiation, motility, migration and cell–cell communication (Nag *et al*, 2013). Although, the primary targets of p53, as a tumour suppressor, are those involved with the regulation of the cell cycle, senescence and apoptosis such as *CDKN1A* (*p21, CIP1, WAF1), p53 upregulated modulator of apoptosis (PUMA),* and *Promyelocytic leukemia protein (PML)* (el-Deiry *et al*, 1993; Nakano & Vousden, 2001; Yu *et al*, 2001; de Stanchina *et al*, 2004; Menendez *et al*, 2009; Campisi, 2013; Rodriguez-Brenes *et al*, 2016).

Under basal conditions, p53 levels are maintained at low levels in cells by a number of proteins. However, the most important negative regulator of p53 is the E3 ubiquitin ligase MDM2, which exists in a negative feedback loop with p53. MDM2 binds the N terminal transactivation domain of p53, preventing its activity as a transcription factor. It is also responsible for ubiquitinating the protein, leading to its proteasomal degradation (Momand *et al*, 1992; Rodriguez *et al*, 2000). Upon the induction of cellular stress p53 is stabilised and activated by a series of post-translational modifications including phosphorylation, acetylation, methylation, ubiquitination and sumoylation at multiple sites by a variety of proteins (Harris & Levine, 2005).

The activation of ATM in response to double strand breaks and ionising radiation, results in the stabilisation of p53. ATM causes the direct and indirect phosphorylation of p53 at multiple sites, including Ser 6, 9, 15, 20, 46 and Thr 18(Khanna. *et al*, 1998; Hirao *et*

al, 2000; Saito *et al*, 2002; Kodama *et al*, 2010). The phosphorylation of p53 Ser15 in response to DNA damage, has been found to prime the protein for subsequent modifications (Lavin & Gueven, 2006). ATM also phosphorylates MDM2 (Khosravi *et al*, 1999; Maya *et al*, 2001; Chen *et al*, 2005; Gannon *et al*, 2012). Phosphorylation of p53 and MDM2 by ATM act to destabilise the interaction between p53 and MDM2 and consequently results in the stabilisation of p53.

One of the most investigated p53 targets is the cyclin dependent kinase (CDK) inhibitor p21. p53 binds the promoter of the *CDKN1A* (*p21, CIP1, WAF1)* gene that encodes p21 and activates its transcription (el-Deiry *et al*, 1993; Macleod *et al*, 1995; Elbendary *et al*, 1996) The upregulated p21 then inhibits the activity of cyclin-CDK2, cyclin-CDK1, and cyclin-CDK4 complexes, which prevents the phosphorylation of proteins required for the activation of E2F transcription factors that facilitate cell cycle progression (He *et al*, 2005; Satyanarayana *et al*, 2008). p21 also binds to proliferating cell nuclear antigen (PCNA) inhibiting PCNA-dependent DNA replication and modulating PCNA-dependent DNA repair processes (Luo *et al*, 1995; Abbas & Dutta, 2009). As such, p53 mediated upregulation of p21 is a critical activator of cell cycle arrest in response to stress. However, p53-independent mechanisms of regulation also exist (Macleod *et al*, 1995; Loignon *et al*, 1997; Li *et al*, 2002; Tanaka *et al*, 2007; Jeong *et al*, 2010; Xu *et al*, 2014)

The two major pathways involved in DSB repair are homologous recombination (HR) and non-homologous end joining (NHEJ). There is also less preferentially used microhomology-mediated end joining (MMEJ) pathway.

1.5.1 Homologous recombination

HR requires 5'-3' resection of double-stranded DNA (dsDNA) at the site of the break to form single-stranded 3' DNA overhangs. The initial short range resection of DNA, of up to a few hundred bases, is facilitated by MRE11 and C-terminal binding protein 1 (CtIP) (Limbo *et al*, 2007; Sartori *et al*, 2007). After initial resection, BRCA1 and CtIP promote longer range resection of DNA with exonuclease 1 (Exo1), Bloom syndrome protein (BLM) and DNA replication helicase/nuclease 2 (DNA2) (Garcia *et al*, 2011; Nimonkar *et al*, 2011; Hoa *et al*, 2015). Replication protein A (RPA) rapidly coats and stabilises the newly formed length of single-stranded DNA (ssDNA) preventing the formation of secondary structures. The RPA coating is subsequently replaced by RAD51, which is promoted by RAD52 and BRCA2 along with the partner and localizer of BRCA2 (PALB2) (Sugiyama & Kowalczykowski, 2002; San Filippo *et al*, 2006; Xia *et al*, 2006). This loading of RAD51 onto ssDNA generates a presynaptic filament and is a crucial step in HR as it allows for homology search, enables strand invasion of duplex DNA and D-loop formation (Short *et al*, 2016). DNA synthesis occurs using the invaded duplex DNA as a template. This forms a DNA intermediate which is resolved through the resolution of a double Holliday junction (Figure 1.6). The double Holliday junction structure is resolved by structure-selective endonucleases such as the mus81-eme1 endonuclease and GEN1 (Ciccia *et al*, 2003; Rass *et al*, 2010). Resolution of the double Holliday junction generates crossover and non-crossover products. Alternatively, in the synthesis-dependent strand annealing (SDSA) pathway of resolution, the formation of a double Holliday junction is suppressed and the invading strand is displaced after DNA synthesis where it can then anneal with the broken DNA ends (Renkawitz *et al*, 2014) (Figure 1.6). HR is a high-fidelity form of DNA damage repair and often termed error free, although

due to the use of a sister chromatid as the template for repair, HR can lead to loss of heterozygosity (Liu & Huang, 2014).

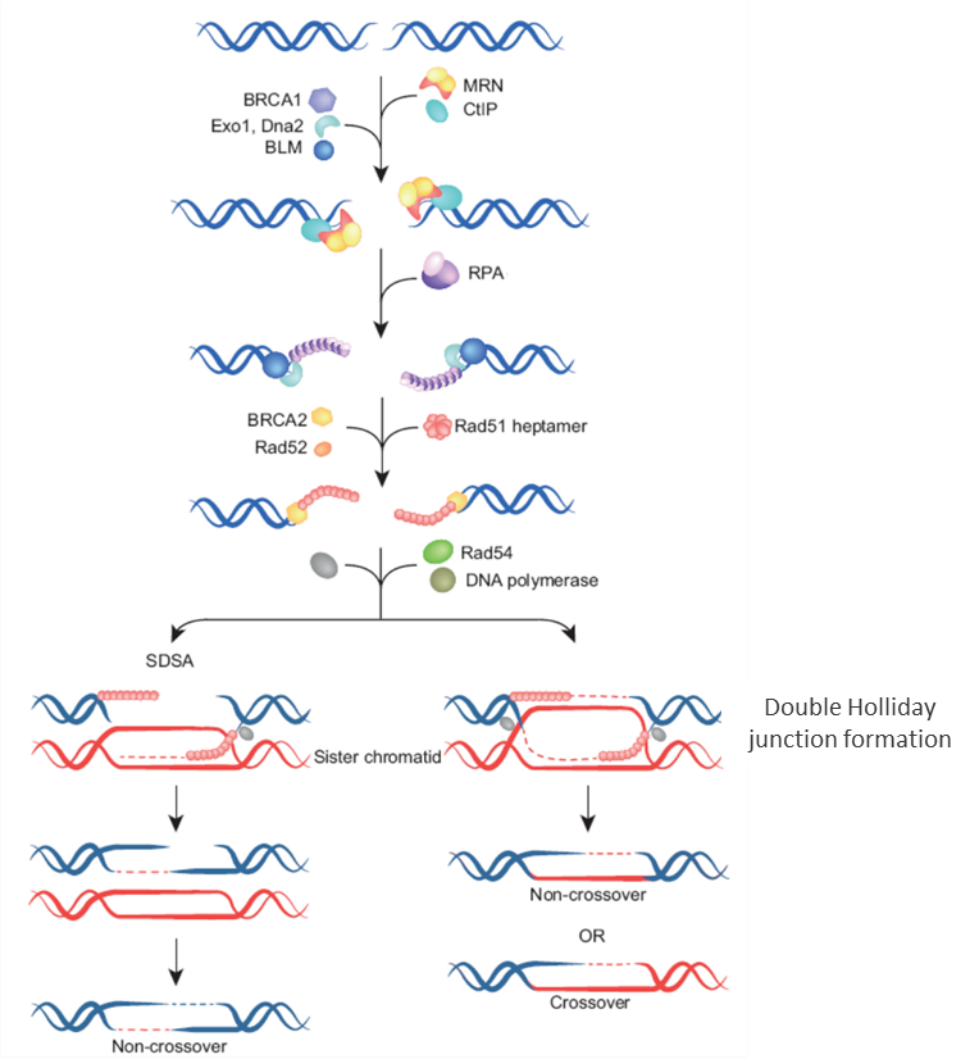


Figure 1.6: **A schematic representation of DSB repair via homologous recombination**. Repair is initiated by MRE11 and CtIP, which perform an initial short range resection of the DSB. This short range processing facilitates the extended resection of the DNA by Exo1, BLM, DNA2, which is promoted by BRCA1 and CtIP. Resection of the break site produces ssDNA overhangs that are rapidly stabilised and protected from degradation by RPA. RAD52, BRCA2 and PALB2 then promote the displacement of RPA for RAD51. The loading of RAD51 onto the ssDNA overhang catalyses homology search and strand invasion of the homologous sister chromatid. The invaded homologous sister chromatid is then used as a template for DNA synthesis. This forms a DNA intermediate which is resolved via the processes of synthesis-dependent strand annealing (SDSA) or double Holliday junction resolution. In the resolution of the double Holliday junction, structure-selective endonucleases cleave the junctions to produce crossover or non-crossover products. Whereas in SDSA the formation of the double Holliday junction is suppressed and the invading strand is displaced after synthesis. This displaced strand can then anneal with the broken DNA ends to complete repair (reprinted from Dueva & Iliakis, 2013).

1.5.2 Non-homologous end joining

Canonical NHEJ is composed of three steps: bridging and stabilising of DNA ends, end processing and ligation. DNA end bridging and stabilisation is carried out by the Ku70/80 heterodimer and the DNA-dependent protein kinase catalytic subunit (DNA-PKcs). The Ku70/80 heterodimer encircles and protects DNA ends and then acts as a scaffold to recruit the additional factors required for NHEJ (Walker *et al*, 2001; Fell & Schild-Poulter, 2015). DNA-PKcs is recruited to the DNA break by the Ku70/80 heterodimer where it is activated. This activation induces a translocation of the Ku70/80 heterodimer from the ends of DNA inward along the DNA (Yoo & Dynan, 1999) and additionally activates other molecules involved in NHEJ (Liu & Huang, 2014). As well as having a role in activation, the binding of DNA-PKcs also acts to tether DNA ends (Spagnolo *et al*, 2006).

Damaged DNA ends have different structures which need to be processed to 5'-phosphate and 3'-hydroxyl termini for repair to continue (Povirk, 2012). DNA end processing is carried out by a variety of different molecules containing nuclease activity, with the structure of the DNA ends determining which end processor is used (Chang *et al*, 2016). Artemis, Werner (WRN), polynucleotide kinase 3′-phosphatase (PNKP), aprataxin (APTX) , the MRN complex, and even Ku are some of the proteins used to process DNA ends (Strande *et al*, 2012).

Once processed, DNA polymerases μ and λ are used to fill the nucleotide gap (Pryor *et al*, 2015). Repair is completed by the ligation of DNA with DNA Ligase IV. The action of Ligase IV is stimulated by X-ray cross complementing protein 4 (XRCC4) and XRCC4-like factor (XLF), which promote a mobile bridging between DNA molecules (Ahnesorg *et al*, 2006; Brouwer *et al*, 2016) (Figure 1.7). Interestingly the requirement for XLF in DNA

repair has been found to be variable and not absolutely required for Ligase IV stimulation (Roy *et al*, 2012, 2015). As NHEJ does not use a sister chromatid to fill in nucleotides it is an inherently more error prone method of DSB repair compared to HR. Despite this, in recent years, NHEJ has undergone an image change. Studies have found repair by NHEJ to be surprisingly precise with the structure of the DNA ends being the limiting factor in repair accuracy rather than the NHEJ machinery (Bétermier *et al*, 2014).

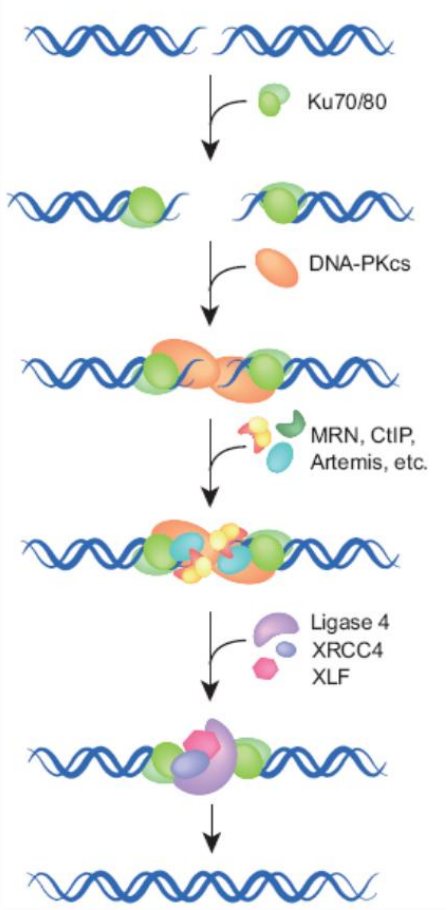


Figure 1.7: **A schematic representation of DSB repair via non-homologous end joining.** The Ku70/80 complex encircles DNA ends, protecting them from nucleolytic degradation, and recruits DNA-PKcs to the DSB. Upon recruitment to the DSB the kinase activity of DNA-PKcs is activated. The interaction between DNA-PKcs and the Ku70/80-DNA complex causes the translocation of the Ku heterodimer along the DNA, away from the break site. The damaged DNA ends can then be processed by structure specific nucleases. End processing is followed by the filling of the residual nucleotide gap, which is performed by DNA polymerases μ and λ. XRCC4 and XLF stimulate DNA ligase IV to complete repair (reprinted from Dueva & Iliakis, 2013).

1.5.3 Microhomology-mediated end joining

In addition to canonical NHEJ there is also alternative NHEJ or microhomology-mediated end joining (MMEJ). MMEJ is a Ku-independent alternative form of NHEJ. It is an error prone form of DSB repair and has been linked to oncogenic rearrangements of DNA such as chromosomal translocations and telomere fusions (Rai *et al*, 2010; Villarreal *et al*, 2012). MMEJ is a "cut and paste" method of DNA repair, dependent on annealing short regions of microhomology to "glue" DNA back together. MMEJ is composed of five fundamental steps: resection of DNA ends, annealing of areas of microhomology, removal of heterologous flaps, fill in synthesis and ligation. As opposed to HR, in MMEJ the resection of the DNA ends is limited and no strand invasion occurs.

Resection of the break site is conducted in a 5'-3' direction exposing a short length of ssDNA with an area of microhomology. Similar to HR, initial end resection in MMEJ is conducted by MRE11 and CtIP (Truong *et al*, 2013). Areas of microhomology anneal with homologous areas flanked by a 3' flap of unpaired DNA and a gap. Heterologous DNA flaps need to be removed for repair to continue, which is performed by structure-specific endonucleases such as the ERCC1-XPF endonuclease. Once flaps are removed, gaps in the DNA are co-operatively filled by DNA poymerases δ and λ (Meyer *et al*, 2015). DNA is finally ligated by DNA ligase I/III to complete the repair (Figure 1.8).

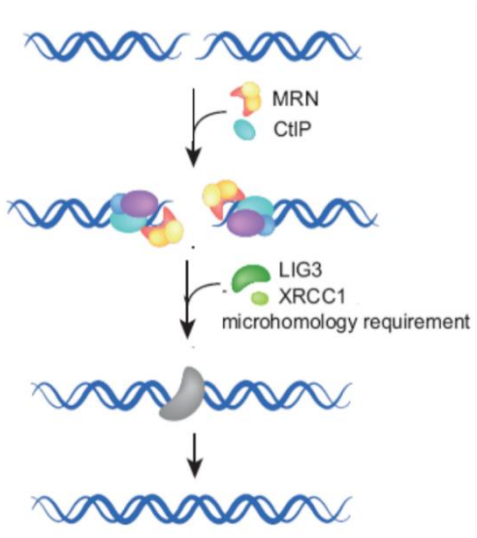


Figure 1.8: **A schematic representation of DSB repair via microhomology-mediated end joining**. Repair is initiated by MRE11 and CtIP, which perform limited 5'-3' resection of the break. This nucleolytic processing of the DNA ends exposes a short length of ssDNA with an area of microhomology. Areas of microhomology can then anneal but are flanked by a 3' flap of unpaired DNA and a gap. 3' flaps are removed by structure-specific endonucleases before the gaps in the DNA are filled in co-operatively by DNA poymerases δ and λ. DNA ligase I/III ligates the DNA to complete repair (reprinted from Dueva & Iliakis, 2013).

1.6 MRE11: a significant DNA damage response protein

MRE11 is a critical gene and murine models have found that the knock out of the *MRE11* gene or mutations inactivating MRE11 nuclease activity are embryonic lethal (Xiao & Weaver, 1997; Buis *et al*, 2008). In humans, individuals with inherited hypomorphic mutations in *MRE11,* which result in MRE11 deficiency, have ataxia-telangiectasia-like disease (ATLD).

Patients with ATLD display progressive cerebellar ataxia, hypersensitivity to ionising radiation, genomic instability, defective S phase checkpoint and defective ATM activation in cells (Taylor *et al*, 2004). Due to the rarity of ATLD there is too little evidence to firmly state whether the disease predisposes patients to cancer. However,

there is evidence that some ALTD associated MRE11 mutations are associated with cancer (Regal *et al*, 2013). Specifically, the *MRE11A*^{ATLD1} mutation has been found to associate with a familial cancer syndrome (Bartkova *et al*, 2008). In 2009 a report was published on two brothers with ALTD caused by novel *MRE11A* mutations that developed lung carcinoma at the young ages of 9 and 16 (Uchisaka *et al*, 2009).

The biological importance of MRE11, in normal cell functioning, has been demonstrated through the identification of mutations or aberrant protein expression in a range of cancers. For example, *MRE11* has been found to be frequently mutated in mismatch repair deficient cancers resulting in a loss of protein expression, mutated in lymphoid tumours and aberrantly expressed and mutated in ovarian and breast cancers (Fukuda *et al*, 2001; Giannini *et al*, 2002; Heikkinen *et al*, 2003; Miquel *et al*, 2007; Bartkova *et al*, 2008; Wen *et al*, 2008; Brandt *et al*, 2017). There is also evidence that germline mutations in *MRE11* associate with hereditary susceptibility to breast and ovarian cancer (Heikkinen *et al*, 2003; Bartkova *et al*, 2008).

In bladder cancer clinically relevant MRE11 variants have been described. The rs2155209 SNP in the *MRE11* 3'UTR has been found to associate with a marginal increase in bladder cancer risk (Choudhury *et al*, 2008). Additionally, the rs1805363 SNP in the *MRE11* 5'UTR has been associated with poor outcome following RT in MIBC patients and an alteration in *MRE11* isoform expression (Teo *et al*, 2014). On the protein level a novel C-terminally truncated form of MRE11 has also been detected which is less stable then WT MRE11 (Nicholson *et al*, 2017).

MRE11 is a key player in the cellular response to DNA damage and its biological function is dependent on its ability to form the MRN complex with RAD50 and NBS1. The MRN

complex is one of the first responders to be recruited to sites of DNA damage where it binds to and tethers DNA ends (Lavin, 2007). This binding of DNA is facilitated through the DNA binding domains of MRE11 and the coiled-coil domains of RAD50 (de Jager *et al*, 2001b; Lee *et al*, 2013). Once bound to DNA, ATP hydrolysis stimulates the partial unwinding of the DNA duplex, which is promoted by MRN, and also triggers a large conformational change in the MRN complex itself (Paull & Gellert, 1999; Liu *et al*, 2016).

MRN recruits ATM to sites of damage and promotes the transition of inactive dimeric ATM into an active monomeric form (Lavin, 2007). The presence of NBS1 in the MRN complex and the binding of ATP are both critical for the activation of ATM. However, it has been found that ATP hydrolysis and MRE11 nuclease function is not (Lee *et al*, 2013). Although MRE11 has no direct role in activating ATM, mutations in the N-terminal domain of MRE11 have been found to impact ATM signalling (Regal *et al*, 2013). The activation of ATM initiates the downstream signalling initiating cell cycle checkpoints and DNA damage repair.

In the repair of DSB through HR and MMEJ, the endonuclease and exonuclease activity of MRE11 is required to initiate the resection of the break site. Firstly, phosphorylated CtIP promotes the endonuclease activity of MRE11 to make an initial single strand nick in the DNA upstream of the damage site. DNA resection then proceeds bidirectionally as the exonuclease activity of MRE11 digests DNA in a 3'-5 direction, toward the exposed DNA end and the 5'-3' exonuclease activity of EXO1 and BLM resected DNA away from the end (Shibata *et al*, 2014; Anand *et al*, 2016). The resection of the DSB by MRE11 is a key step in repair pathway choice as it directs repair away from NHEJ (Shibata *et al*, 2014).

Another role for MRE11 in DNA damage repair is the processing of DNA ends, such as the hairpin-capped DSBs and the removal of bulky adducts from DNA (Lobachev *et al*, 2002; Sacho & Maizels, 2011; Deshpande *et al*, 2016). Bulky DNA adducts can arise endogenously from DNA protein crosslinks (DPCs) or can arise from exposure to chemical agents or ionising radiation (Tretyakova *et al*, 2015). 3' bulky DNA adducts require the endonuclease activity of MRE11 for their removal. Whereas for the removal of 5' bulky DNA adducts, there is a requirement for the endonuclease and exonuclease activity of MRE11 (Liao *et al*, 2016). In the processing of DPCs MRE11 generates a clean DSB that can be repaired by classical repair pathways. It has been found that this role of MRE11 is of particular importance for the repair of the potentially lethal sites where topoisomerase I or II are covalently bound to DNA (Hartsuiker *et al*, 2009).

MRE11 also plays a role in the repair and restart of stalled or collapsed replication forks. Under conditions of replicative stress or DNA damage, replication forks can stall or collapse. Stalled replication forks can continue replication, once the inhibition to fork progression is removed. However, through the dissociation of replication machinery or the formation of a DSB, a collapsed fork cannot perform DNA synthesis upon the removal of inhibition (Petermann *et al*, 2010; Cortez, 2015).

Upon stalling, replication forks can be remodelled into a regressed chicken-foot (four-way junction) structure by annealing of newly synthesized (nascent) DNA and re-annealing of parental strands (Sogo *et al*, 2002; Ray Chaudhuri *et al*, 2012). Fork reversal stabilises the stalled fork allowing time for the removal of the replicative inhibition. In recent years, it has been found that if the nascent DNA of a reversed fork is not protected, it undergoes degradation by MRE11. This can lead to cell lethality and

genomic instability (Hashimoto *et al*, 2010; Schlacher *et al*, 2011; Ying *et al*, 2012; Yang *et al*, 2015; Kolinjivadi *et al*, 2017). Indeed, MRE11-dependent fork resection underlies the increased chromosome breakage exhibited by BRCA2 null cells (Schlacher *et al*, 2011). A number of proteins, including RAD51, BRCA2, PARP1 and REV1 have been identified as factors that protect the DNA at damaged forks and prevent excessive degradation by MRE11 (Hashimoto *et al*, 2010; Schlacher *et al*, 2011; Ying *et al*, 2012; Yang *et al*, 2015).

MRE11 is also required at collapsed replication forks. At collapsed replication forks, subunits of the GINS complex and DNA polymerase epsilon dissociate from the replisome, preventing the restart of replication even upon the removal of replication inhibition. It has been found that MRE11 and RAD51 facilitate the reloading of these components back onto the collapsed fork, promoting fork restart and preventing the formation of DSBs (Trenz *et al*, 2006; Hashimoto *et al*, 2011). However, if a DSB is induced by replication fork collapse, the major pathway of repair has been identified as homologous recombination, which requires MRE11 (Arnaudeau *et al*, 2001; Saleh-Gohari *et al*, 2005; Petermann *et al*, 2010; Groth *et al*, 2012).

As described above, the function of MRE11 relies on its ability to bind DNA and its nuclease function. The MRE11 DNA binding motifs are contained within two distinct sites in the C-terminal of the protein. MRE11 has the capacity to bind ssDNA and dsDNA. However, the affinity for MRE11 to bind ssDNA is greater than that for dsDNA (Williams *et al*, 2008). Exonuclease function is defined by four highly conserved phosphoesterase motifs in the N terminal of the protein (Krogh *et al*, 2005). However, nuclease activity is

regulated by a number of factors. A schematic representation of MRE11 with the relative positions of important features can be seen in Figure 1.9.

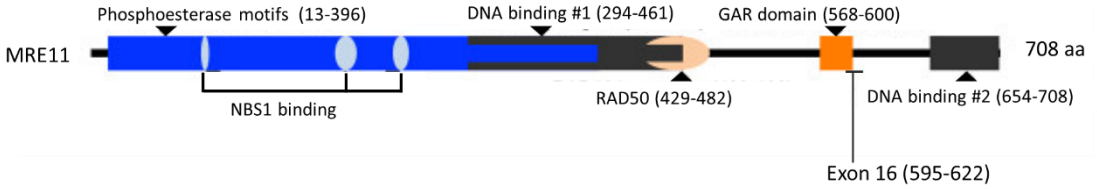


Figure 1.9: **The relative positions of the functional motifs and domains within MRE11**. The location of the C terminal phoshoeterase motifs (required for MRE11 nuclease function), the two DNA binding motifs, regulatory GAR domain and sites of RAD50 and NBS1 interaction are depicted along with the position of *MRE11* exon 16 (reprinted from Damiola *et al* (2014)).

1.6.1 MRE11 GAR domain

MRE11 contains a glycine-arginine-rich (GAR) domain which is subject to post-translational methylation on arginine residues by protein arginine methyltransferase 1 (PRMT1) (Boisvert *et al*, 2005a). Methylation of arginine residues occurs on the guanidino group (CH_4N_3) of the arginine side chain. Multiple methyl groups can be added to an arginine residue changing its conformation, charge and also its ability to interact with hydrogen bond donors. Hence, the methylation of arginine within a protein can disrupt molecular interactions with binding partners or protein activity.

Arginine methylation can occur in three forms. Monomethylated arginine (MMA) contains a single methyl group attached to the N terminal atom of the guanidino group. Two forms of dimethylated arginine are possible, asymmetric dimethylated arginine (ADMA) and symmetrically dimethylated arginine (SDMA). ADMA holds two methyl groups on one of the terminal nitrogen atoms of the guanidino group whereas SDMA contains two methyl groups placed either side of two of the N terminal guanidino

nitrogens (Figure 1.10) (Bedford & Clarke, 2009). Methylation of the MRE11 GAR domain has been found to have large impacts on protein function including DNA binding, nuclease activity (Boisvert *et al*, 2005a; Déry *et al*, 2008) and re-localisation from PML bodies to sites of DNA damage (Boisvert *et al*, 2005b). Loss of MRE11 methylation results in cell cycle checkpoint defects, hypersensitivity to gamma irradiation and chromosomal instability (Yu *et al*, 2012).

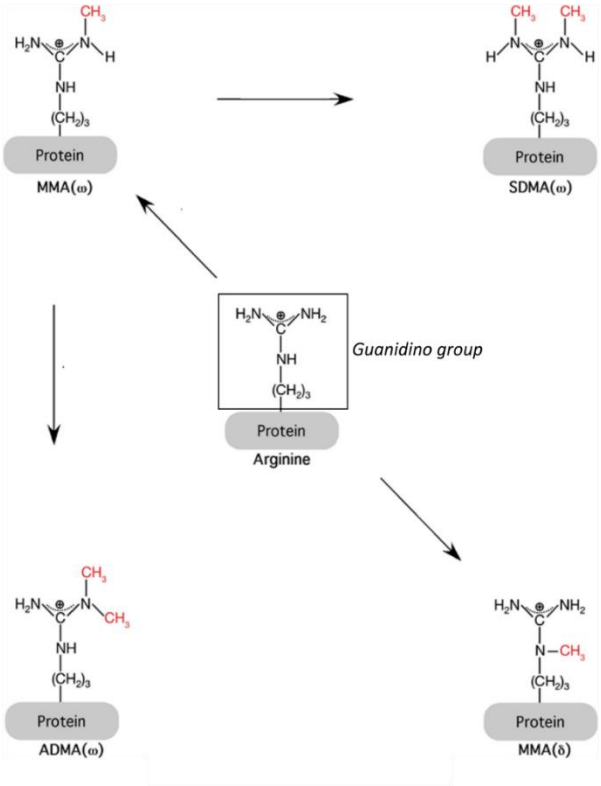


Figure 1.10: **The forms arginine methylation**. Protein arginine methyltransferases (PRMTs) transfer a single methyl group onto one of the terminal guanidino group (CH_5N_3) nitrogen atoms of arginine to form monomethylated arginine (MMA). The addition of a second methyl group can then result in asymmetric dimethylated arginine or symmetric dimethylated arginine (SDMA). In ADMA two methyl groups are attached to one of the terminal nitrogen atoms of the guanidino group, whereas in SDMA two methyl groups are attached to two of the N terminal guanidino nitrogens (reprinted from Bedford & Clarke, 2009).

1.6.2 Alternative splicing of MRE11

The alternative splicing of *MRE11* results in two transcript variants encoding two distinct isoforms, *MRE11* isoform 1 and *MRE11* isoform 2. The two *MRE11* isoforms are highly similar (Figure 1.11), with the only difference between the two isoforms at the protein level being the absence of a short 27 amino-acid-long exon, exon 16, in *MRE11* isoform 2. At the transcript level *MRE11* isoform 2 has a 5'UTR which is 107 nucleotides longer than *MRE11* isoform 1 but there is currently an absence of knowledge as to any difference in the function of *MRE11* isoform 1 and isoform 2.

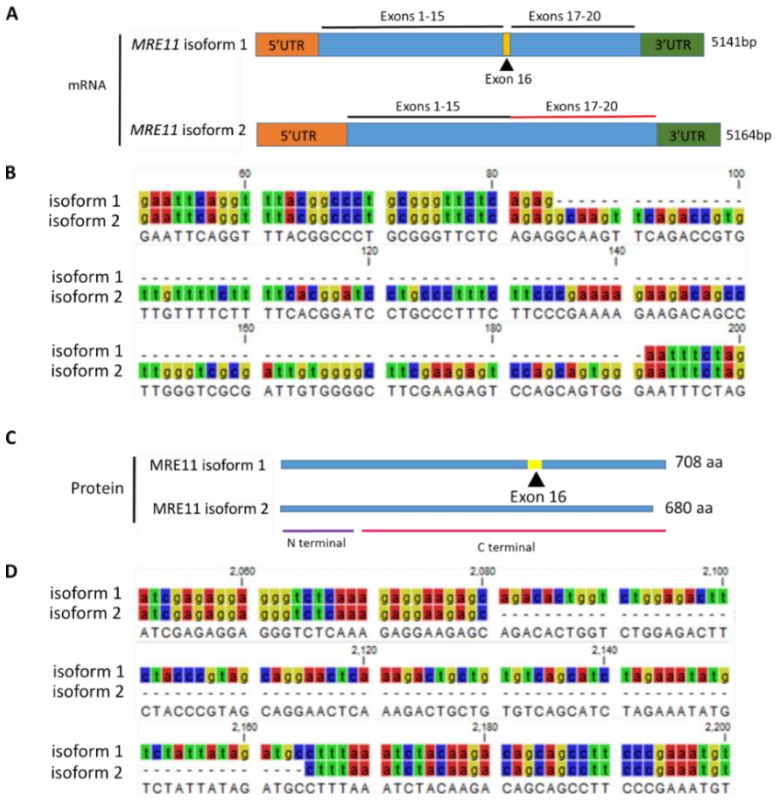


Figure 1.11: ***MRE11* isoforms 1 and 2**. (A) Schematic representations of *MRE11* isoforms 1 and 2 mRNA. *MRE11* isoform 1 has a shorter 5'UTR compared to isoform 2 and contains all exons 1-20, MRE11 isoform 2 has an addition of 107 nucleotides in its 5'UTR and is missing exon 16; (B) The extended 5' UTR nucleotide sequence of *MRE11* isoform 2; (C) Schematic representation of *MRE11* isoforms as proteins. *MRE11* isoform 1 is a full length 708aa long version of the protein containing all *MRE11* exons and functional domains. *MRE11* isoform 2 is a 680aa long version of the protein without exon 16, however all reported functional domains are present; (D) The nucleotide sequence of *MRE11* exon 16, absent in *MRE11* isoform 2.

1.6.3 Regulation of MRE11 activity by RAD50 and NBS1

MRE11 functions within the MRN complex and many of its functions are reliant on its ability to interact with its complex partners, RAD50 and NBS1. MRE11 is the core of the MRN complex and exists as a homodimer, providing a frame on which RAD50 and NBS1 can bind (Williams *et al*, 2008) (Figure 1.12). Recently the flexibility of the MRN complex has been highlighted. Large conformational changes in the complex have been identified and found to be important for the functioning of the complex (Lafrance-Vanasse *et al*, 2015).

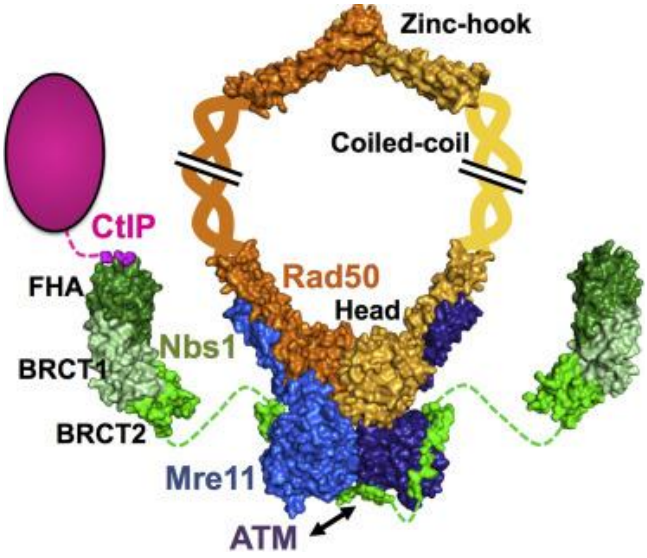


Figure 1.12: **The MRN complex**. The MRN complex is a heterohexamer composed of 3 protein dimers. The MRE11 dimer is depicted in blue and dark blue (for each monomer). MRE11 is the core of the complex providing a frame on which RAD50 and NBS1 can bind. The RAD50 dimer is represented in orange and yellow (for each monomer) with the positions of the coiled-coil region and Zinc-hook highlighted. The coiled-coil region folds back on itself bringing together the two nucleotide (NTP)-binding motifs to create a functional ATP-binding cassette ATPase domain. The Zinc-hook facilitates dimerisation of RAD50. NBS1 is shown in green with different shades representing the different domains (FHA, BRCT1 and BRCT2) for protein interaction. The region of ATM interaction with NBS1 is shown by a black arrow. The dotted green lines represent disordered protein linkers. The binding of CtIP to the MRN complex is also shown (reprinted from Lafrance-Vanasse *et al*, 2015).

The RAD50-MRE11 interaction is critical for proper functioning of MRE11. The MRE11-RAD50 complex is a heterotetrameric assembly consisting of head, coil and hinge domains (de Jager *et al*, 2004).

A coiled-coil domain is a secondary structure motif composed of α helices wound together in a superhelical bundle. This structure is maintained through the interaction between the hydrophobic residues within the core of the bundle (Hartmann *et al*, 2016). The RAD50 coiled-coil folds back upon itself to bring together the Walker A and Walker B nucleotide (NTP)-binding motifs, which creates a functional ATP-binding cassette (ABC) ATPase domain in the protein (de Jager *et al*, 2001b; Park *et al*, 2017). It has been found that this domain in RAD50 can facilitate in the binding of the MRN complex to DNA, through the binding of internal sites on DNA (Lee *et al*, 2013; Rojowska *et al*, 2014). Furthermore, the coiled-coil imparts flexibility to RAD50 dimers and has been observed to adopt multiple conformations (de Jager *et al*, 2001b). Interestingly, it has also been proposed that the RAD50 coiled-coil may communicate allosteric changes in the head domain of the protein to the Zn hook, through the sliding of the two coiled-coil α helices (Lafrance-Vanasse *et al*, 2015).

At the apex of the RAD50 coiled-coil is a conserved Cys-X-X-Cys motif (Zn hook motif). This motif creates one-half of a composite Zn^{2+}-binding site. Two Zn hooks from oppositely protruding RAD50 molecules interlock to bind one Zn^{2+} ion, which facilitates the dimerisation of the protein (Hopfner *et al*, 2002; Kochańczyk *et al*, 2016).

MRE11 binds RAD50 through its C-terminal helix-loop-helix domain, through a flexible linker, allowing for large conformational changes in the protein complex (Figure 1.13) (Lammens *et al*, 2011). The RAD50 subunits of the heterotetramer are bound to ATP. In

its unhydrolysed form ATP negatively regulates MRE11 nuclease activity by blocking its active site and promoting binding and tethering of DNA ends (Paull & Gellert, 1999; Lim *et al*, 2011). However, MRE11 stabilises RAD50 dimerisation, locking residues around the RAD50 ATP-binding site, promoting ATP hydrolysis. In turn, the hydrolysis of ATP causes large conformational changes in the MRE11-RAD50 complex exposing the nuclease active sites of MRE11 and increasing non-end specific DNA binding (Paull & Gellert, 1999; Lammens *et al*, 2011; Lim *et al*, 2011; Seifert *et al*, 2016).

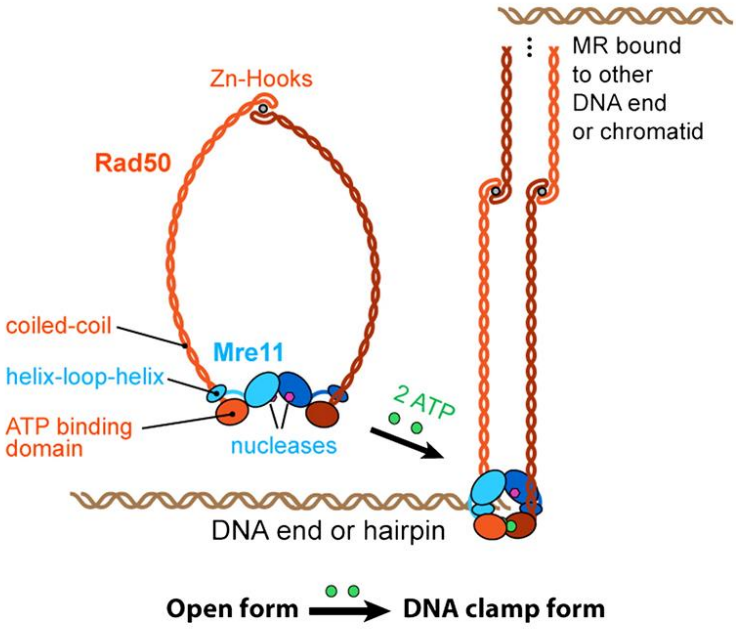


Figure 1.13: **The ATP dependent conformations of MRE11 in complex with RAD50**. The MRE11 dimer is depicted in blue and dark blue (for each monomer). The RAD50 dimer is represented in orange and dark orange (for each monomer). MRE11 binds RAD50 through a flexible helix-loop-helix to allow for large conformational changes in the complex. The zinc hooks of RAD50 facilitate RAD50 dimerisation. The RAD50 coiled-coils add flexibility to the MRE11-RAD50 complex and bring together the two nucleotide (NTP)-binding motifs. This creates a functional ATP-binding cassette ATPase domain. When ATP is bound to the RAD50 the MRE11-RAD50 complex adopts a "closed" DNA clamp conformation promoting binding and the tethering of DNA ends. The hydrolysis of ATP results in a change to an "open" complex conformation. In the "open" complex MRE11 nuclease sites are exposed and there is an increasing non-end specific DNA binding (reprinted from Lammens *et al*, 2011).

NBS1 acts to regulate the activities of MRE11 and RAD50 and also influences the subcellular localization of the MRN complex (Desai-Mehta et al, 2001). Structural studies have identified that MRE11 in complex with NBS1 has large amounts of asymmetry and that molecules of NBS1 stretch around the outside of the MRE11 nuclease domains in a highly extended conformation. One of the NBS1 subunits is also involved in binding across the MRE11 dimer interface, bridging and locking the MRE11 dimer (Figure 1.12) (Park *et al*, 2011; Schiller *et al*, 2012). NBS1 stabilises the ATP bound

form of the MRE11-RAD50 complex and is necessary for the ATP dependent actions of MRN, such as partial unwinding of duplex DNA (Paull & Gellert, 1999; Lee *et al*, 2013). Interestingly NBS1 is also responsible for the substrate specificity of MRE11 nuclease activity. NBS1 promotes MRE11 exonuclease at DNA ends containing adducts but inhibits its activity at 'clean' DNA ends (Deshpande *et al*, 2016).

In addition to directly affecting the function of MRE11 and RAD50, NBS1 also recruits a number of proteins to sites of DNA damage and hence facilitates the progression of the DDR. NBS1 recruits ATM to sites of DNA damage and causes its activation (You *et al*, 2005; Lee & Lim, 2006). It is also responsible for the recruitment of CtIP through a forkhead associated (FHA) domain and BRCA1 and BRCA2 through BRCA1 C-terminal (BRCT1) and BRCA2 C-terminal (BRCT2) domains located at the N terminus of the protein The binding of CtIP causes large conformational changes in NBS1, thus altering its ability to interact with BRCA1 and BRCA2 and hence affects repair pathway choice (Yun & Hiom, 2009; Lafrance-Vanasse *et al*, 2015).

1.7 MRE11: a potential biomarker in MIBC

There are currently no clinically approved biomarkers for use in bladder cancer, although a large body of potential biomarkers have been identified. While numerous studies into bladder cancer biomarkers exist, only a relatively small number focus on MIBC. Furthermore, due to the historical bias toward RC as the gold standard treatment of MIBC, there is a significant lack of evidence into predictive MIBC biomarkers for outcome following RT.

1.7. 1 High levels of MRE11 protein are predictive of improved survival following RT in MIBC

The DNA repair protein MRE11 has been identified as a potential predictive biomarker for outcome following RT in MIBC but not RC. Choudhury *et al* (2010) determined MRE11 protein abundance in retrospective FFPE biopsies of MIBC, using IHC, and found that high levels of MRE11 (>25th percentile) significantly associated with improved CSS in patients who underwent RT but had no association with outcome in patients who received RC.

This result was independently validated by Laurberg *et al* (2012). However, as opposed to Choudhury *et al* (2010) where the majority of patients received RT alone, 83% of the patients in the Laurberg *et al* (2012) patient cohorts received CRT. Similar to Choudhury *et al* (2010), this study used IHC to determine MRE11 protein abundance in retrospective FFPE MIBC samples and identified high levels of MRE11 (>25th percentile) were significantly associated with improved outcome in the radiotherapy-based cohort but had no association with outcome in RC cohorts.

The finding that high levels of MRE11 associate with improved RT response in MIBC patients is counterintuitive, as cells with more MRE11 would be predicted as having the capability to efficiently repair the DNA damage induced by radiation. However, within the Anne Kiltie lab a novel C-terminally truncated variant of MRE11 has been identified in MIBC cell lines and patient samples, which may explain the result (Nicholson *et al*, 2017). This truncated MRE11 variant is formed through post-translational cleavage of WT MRE11 and its formation can be induced by treating cells with HDAC inhibitors. It is hypothesised that this is the overexpressed form of MRE11 detected in IHC studies and that it is either functionally impaired or as suggested by Nicholson *et al* (2017) is a degradation intermediate which is processed by the proteasome.

As discussed in section 1.3 of this thesis, the large number of cancer biomarkers identified by researchers never end up in the clinic. One of the main reasons for this is lack of validation. For a cancer biomarker to progress towards the clinical setting, prospective validation in a randomised clinical trial is the gold standard. However, before this, the assay used to evaluate the biomarker must be found to be robust and reproducible, giving concordant results when performed by different laboratories. This step of establishing the technical validity of a biomarker is often unaddressed, which prevents its further development.

The results from investigations into the reliability biomarker assays are important and required for the progression of a biomarker into a prospective setting. Consequently, the primary aim of Chapter 3 of this thesis was to evaluate the concordance in staining and scoring of the MRE11 IHC assay between different research laboratories. Furthermore, this work transferred the previously manual MRE11 IHC protocol onto

automated staining platforms, as automated IHC is the primary method of performing IHC in the clinic.

1.7.2 A germline *MRE11* SNP (rs1805363) is predictive of poor survival following RT in MIBC

Using next generation sequencing, a SNP in MRE11 has also been found to be predictive of RT outcome in MIBC (Teo *et al*, 2014). The rs1805363 SNP is a germline transition mutation in *MRE11* of the G major allele to an A minor allele. It is located intronically in *MRE11* isoform 1, five bases 3' to the exon 1 AG donor splice site and in *MRE11* isoform 2 is located within the 5'UTR. It was found that MIBC patients who carried the rs1805363 A minor allele had a reduced 5 year CSS following RT (p=0.001) compared to patients who were homozygous carriers for the G major allele. Patients who were homozygous for the A minor allele were identified as having the worst CSS with none of the patients surviving at 5 years (n=6), whereas patients who were heterozygous for the A minor allele (AG) had a 5 year CSS of 42% (n=28). Patients homozygous for the G major allele (GG) displayed the highest CSS (58%, n=152) (Figure 1.16A). There was no association between carriage of the A minor allele and CSS following RC (p= 0.88).

In addition to associating with CSS, the presence of the rs1805363 A minor allele was also found to correlate with changes in the relative expression of *MRE11* isoforms 1 and 2 (Teo et al, 2014). The expression of *MRE11* isoforms 1 and 2 were assessed in nine samples, consisting of bladder cancer cell lines and primary bladder tumours identified as positive or negative for the rs1805363 A minor allele. Samples which were genotyped, as containing the rs1805363 A minor allele, were found to have higher ratio of *MRE11* isoform 2 expression to *MRE11* isoform 1 expression, compared to samples which were homozygous for the G major allele (Figure 1.16C). However, due to the small sample size and use of only 3 tumour samples in this analysis, there is a need to confirm this result in a larger number of patient samples.

The finding that *MRE11* isoform 2 expression could be higher in a group of MIBC patients that respond poorly to RT is interesting and leads one to question whether the increase in *MRE11* isoform 2 expression, relative to isoform 1, in these patients is responsible for the poor survival of patients following RT. Although this hypothesis would imply that the function of *MRE11* isoform 2 was not equivalent to *MRE11* isoform 1, as currently assumed. This theory is tested in Chapter 5 of this thesis.

It is noteworthy that the association between *MRE11* rs1805363 A minor allele carriage and CSS is contrary to what is seen in MIBC patients, when MRE11 protein levels are assessed indiscriminately using IHC. Therefore, it is possible that in MRE11 IHC studies, the presence of patients with the rs1805363 A minor allele may confound results. A confounding factor such as this would reduce the predictive capacity of a MRE11 IHC biomarker. Hence, it may be that the using both of these biomarkers to stratify patients toward RT or RC may be more powerful than using either biomarker in isolation. Furthermore, if *MRE11* isoforms are found to have a different efficacy or function in DNA damage repair, patients with differential *MRE11* isoform expression may also confound the results of MRE11 IHC in MIBC.

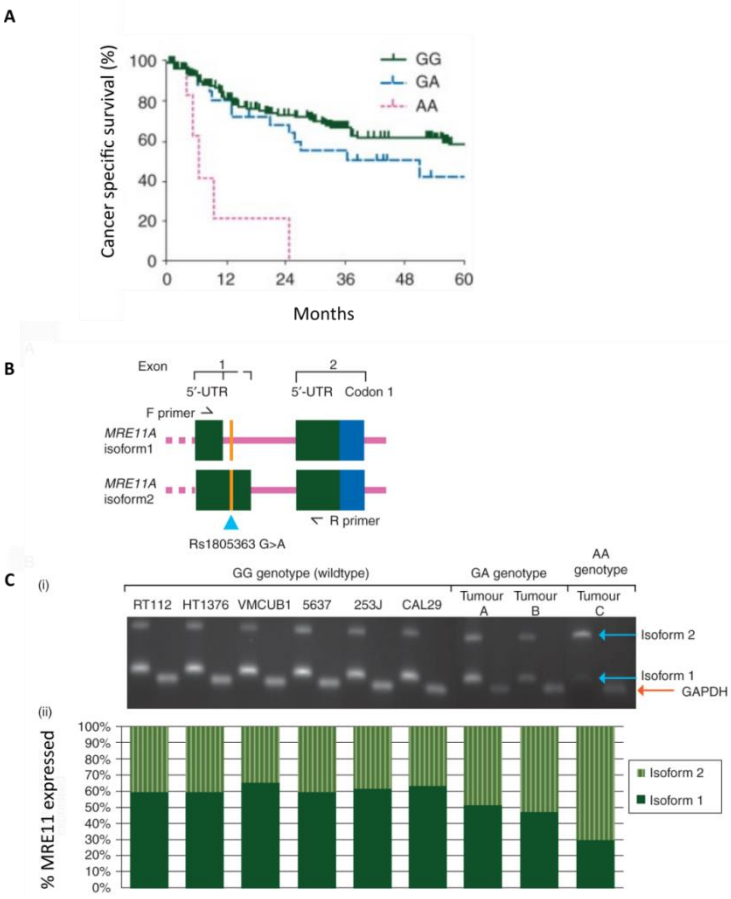


Figure 1.14: **The germline *MRE11* rs1805363 G>A SNP as a predictive biomarker in MIBC**. (A) Kaplan-Meier CSS survival for patients who received RT, stratified by their rs1805363 genotype. Carriage of the A minor allele is associated with worse 5 year CSS following RT p=0.001. (GG hr=1, AG patients hr= 1.49, 95% CI: 0.80-2.78, AA patients hr= 8.00, 95% CI: 2.93–21.90); (B) A schematic representation of the 5'UTR of *MRE11* isoforms 1 and 2. The relative position of rs1805363 SNP is shown. The position of the forward and reverse primers that were used in the PCR amplification of the *MRE11* 5'UTR from cDNA are also shown; (C) (i) Gel electrophoresis bands, of PCR amplified fragments, from six bladder cancer cell lines (RT112, HT1376, VMCUB1, 5637, 253J, CAL29) and three primary bladder tumours. Lanes containing *MRE11* isoform products contain 2 bands corresponding to *MRE11* isoform 1 and 2, and are loaded next to a lane containing an amplified GAPDH product for the sample, which was used as a positive control. In the lanes containing amplified fragments of *MRE11* cDNA the lower band contains amplified cDNA from the *MRE11* isoform 1 5'UTR and the upper band contains amplified cDNA from the longer *MRE11* isoform 2 5'UTR. Samples are grouped according to rs1805363 genotype (ii) Quantification of the percentage of *MRE11* isoform 1 and 2 expression relative to total *MRE11* expression in the lane, for each sample. The percentage of *MRE11* isoform 1 decreased in samples positive for the rs1805363 A minor allele (GG: 61.7%, GA: 49.6% and AA: 30.3%) (taken from Teo *et al*, 2014.).

1.8 Anal cancer

Biomarkers are, in the majority of cases, initially identified in one type of cancer. However, there is often the question as to whether a biomarker is specific to one cancer type or whether it can be of used in other cases. There is evidence that some cancer biomarkers have utility in more than one cancer type. For example, *BRCA* mutational status is used to assess the risk of developing both breast and ovarian cancer (Domchek *et al*, 2010). Additionally, HER2 expression and/or gene amplification has been assessed in a range of disparate cancers, including MIBC, to determine its prognostic or predictive utility, beyond breast cancer (Hirsch *et al*, 2009; Seo *et al*, 2014; Abrahao-Machado & Scapulatempo-Neto, 2016; Kiss *et al*, 2017). MRE11 protein levels have been identified as a potential biomarker for RT outcome in MIBC patients. However, the question of whether MRE11 protein levels could predict patient response following RT in other cancers remains unanswered.

Table 1.5 A comparison between anal cancer and bladder cancer. The factors associated with increased risk of developing bladder cancer and anal cancer are listed. The male:female ratio, predominant histological type and incidence for each disease is also compared.

	Bladder cancer	**Anal Cancer**
Predominant risk factor	Smoking ○ accounts for approximately 50% of cases [1] [2]	Infection with high risk HPV subtypes (predominately HPV16) ○ accounts for 85-90% of cases [3]
Other risk factors	• Occupational exposure to carcinogens such as aromatic amines and polycyclic hydrocarbons • Inherited genetic factors- polymorphisms in N-acetyltransferase 2 (NAT2) • Glutathione S-transferase mu 1 (GSTM1) null genotype • Schistosomiasis (chronic endemic cystitis) from recurrent infection with a parasitic trematode • Exposure to arsenic in drinking water [1] [2]	• HIV infection • Sexual activity – -Increased number of sexual partners -Anal intercourse • Immune suppression • Long term use of immunosuppressants • Autoimmune disorders • Smoking [4,5]
Male:female ratio	3:1 [2]	1:5 [6]
Predominant histological type	Urothelial cell carcinoma [7]	Squamous cell carcinoma [8]
Incidence	Tenth most common cancer in the UK. Ninth most commonly diagnosed cancer worldwide [9,10]	<1% total cancer cases in the UK [11]

Information obtained from: [1] (Burger *et al*, 2013), [2] (Knowles & Hurst, 2014), [3] (Baricevic *et al*, 2015), [4] (Glynne-Jones *et al*, 2014), [5] (Shridhar *et al*, 2015), [6] (Salati & Al Kadi, 2012), [7] (Kamat *et al*, 2016), [8] (Nelson *et al*, 2013), [9] Cancer Research UK, a), [10] (Antoni *et al*, 2017), [11] (Cancer Research UK, b)

Bladder cancer and anal cancer are both diseases that occur more frequently in older patients. In bladder cancer, over half of all cases are aged >75 (Cancer Research UK, a). Whereas in anal cancer, over half of all cases are ages >65 (Cancer Research UK, b). However, in terms of aetiology, histological characterisation and male to female presentation, these two diseases are dissimilar. Table 1.5 compares and contrasts the features of bladder and anal cancer.

The major difference between the aetiologies of these two cancers is the link between infection with the human papilloma virus (HPV) and cancer development. HPV infection is not a factor in the development of bladder cancer, but infection with high-risk subtypes of HPV, particularly HPV16, has been found to be the main driver of anal cancer tumorigenesis. Eighty five to ninety percent of all anal cancers are positive for HPV infection (Baricevic *et al*, 2015). In contrast, for western cultures, the main aetiological factor for bladder cancer is tobacco smoking and it has been calculated that approximately 50% of all bladder cancers are attributable to smoking (Burger *et al*, 2013).

Unlike bladder cancer, which is common, anal cancer is relatively rare accounting for less than 1% of the total cancer cases in the UK in 2014 (Cancer Research UK, b). Although, its incidence is increasing (Shiels *et al*, 2015). Additionally, the primary histological type of anal cancer (approximately 85%) is squamous cell carcinoma, with adenocarcinoma (10%) or other types comprising the remaining 15% of cases (Nelson *et al*, 2013). This is opposed to UC that is the most prevalent form of bladder cancer.

The treatment of anal cancer may consist of either surgical excision alone for small tumours measuring ≤2cm across (staged as T1), but for majority of patients the standard of care is combination CRT (Glynne-Jones *et al*, 2009). CRT for anal cancer achieves good responses with complete tumour regression in 80-90% of cases (Vinayan & Glynne-Jones, 2016). However, despite this good response, acute toxicities resulting from current CRT treatment regimens for anal cancer are common. Patients can suffer long term effects including lower limb venous thrombosis, bowel, bladder and sexual dysfunction (Ghosn *et al*, 2015). Moreover, recent findings have identified a subgroup of anal cancer patients who do not respond to current treatment (Gilbert *et al*, 2013; Doll *et al*, 2014; Koerber *et al*, 2014; Serup-Hansen *et al*, 2014; Mai *et al*, 2015; Meulendijks *et al*, 2015).

Although bladder and anal cancer, show dissimilarity to one another, there are reasons for investigating MRE11 IHC as a predictive biomarker in anal cancer. Firstly, there is a need for biomarkers in anal cancer. Although CRT is the standard of care for the majority of anal cancers, not all patients respond and toxicities are high. MRE11 IHC may identify groups of patients where an alternative treatment may be more beneficial. Secondly, in HPV infected cells the DDR is exploited to enable amplification and integration of the

viral genome (Anacker & Moody, 2016). The MRN complex has been found to be important for productive viral replication, and aberrant expression of the components of the MRN complex have been identified in HPV infected cells (Anacker *et al*, 2014). The altered expression or utilisation of MRE11 in viral replication may change the efficacy of the repair of radiation induced DNA damage affecting tumour response.

1.9 Aims and objectives

The aim of this thesis was to investigate biomarkers relevant to MIBC CSS. To achieve this goal, two previously reported MRE11 biomarkers for patient response following RT were further investigated and the use of crowdsourcing as a method of reliably and efficiently scoring IHC was studied.

Aim 1: To re-optimise an automated MRE11 IHC protocol so that it could be used to determine the reliability of MRE11 scoring and validate MRE11 as a predicative biomarker for RT outcome in MIBC.

To achieve this aim the following objectives were set:

1) Analyse the distribution of scores from the MRE11 staining of a MIBC tissue cohort, stained using an insufficiently optimised automated MRE11 IHC protocol.

2) Conduct a thorough re-optimisation of the automated MRE11 IHC protocol and compare the variation between scoring distributions before and after re-optimisation.

3) Use statistical analysis to determine the reliability between the MRE11 IHC scores, produced in either the same or two different institutions.

4) Conduct a survival analysis to determine the association between MRE11 IHC and CSS in this cohort.

Aim 2: To investigate the specificity of MRE11 IHC as a biomarker for MIBC by assessing MRE11 IHC in a cohort of anal cancer tissue samples.

To achieve this aim the following objectives were set:

1) Perform IHC of MRE11 IHC in a cohort of anal cancer tissue and determine the association between MRE11 IHC and relapse-free survival in this cohort.

2) Use the existing data for this cohort on p16 and tumour infiltrating lymphocyte status, to determine if MRE11 IHC can increase the prognostic value of either of these biomarkers in anal cancer.

Aim 3: To validate the correlation between the increased expression of *MRE11* isoform 2 in MIBC patients who carry the rs1805363 A minor allele (reported by Teo et al (2014) and to characterise the radiation response of cells expressing *MRE11* isoform 2.

To achieve this aim the following objectives were set:

1) Genotype germline DNA samples, extracted from MIBC patients, for the rs1805363 SNP before quantifying the ratio of *MRE11* isoform 1 to 2 in the corresponding patient tumour biopsies.

2) Generate stable MIBC cell lines which re-express a construct containing the full length coding sequence of *MRE11* isoform 1 or the *MRE11* coding sequence with exon 16 removed (representative of *MRE11* isoform 2).

3) Use the stable cell lines in clonogenic and comet and assays to determine the effect of *MRE11* isoform 2 expression on cellular radiosensitivity and DSB repair.

4) Use immunoprecipitation assays to determine if the variations in response between the cell lines is due to the methylation of the MRE11 GAR domain or an impairment in the formation of the MRN complex.

Aim 4: To determine if crowdsourcing could be used as a reliable and efficient method of scoring IHC biomarkers.

The following objectives were set to achieve this aim:

1) Optimise IHC staining for 14 antibodies to stain 7 MIBC TMAs

2) Input digitally scanned images of the stained tissue into a smartphone gaming app that allowed members of the public to score the data.

3) Assess the accuracy of crowdsourced scoring by conducting statistical tests, for the correlation between crowdsourced scores and scores produced by trained researchers.

4) Perform survival analyses to determine the association between each of IHC stains and CSS, using the crowdsourced scores.

Chapter 2 Materials and Methods

2.1 Solutions and Buffers

1% Acid Alcohol

990 ml 70% ethanol

10 ml concentrated hydrochloric acid (Millipore)

Scott's Tap Water

42 mM $NaHCO_3$ (Sigma)

166 mM $MgSO_4$ (Scientific Laboratory Supplies)

dH_2O up to 1 litre

0.5% lithium carbonate

1 g lithium carbonate (Sigma)

200 ml dH_20

Citrate buffer

10 mM citric acid (Fisher Scientific)

0.05% v/v Tween 20 (Promega)

dH_2O up to 1 litre

adjusted to pH 6.0

Tris-EDTA buffer

10 mM Trizma (Sigma)

1 mM EDTA (Fisher Scientific)

0.05% v/v Tween 20 (Promega)

dH_2O up to 1 litre

adjusted to pH 9.0

1x Tris buffered saline (TBS)

20 mM Tris-HCL (Sigma)

150 mM NaCl (Sigma)

pH 7.4

Manual IHC protein block

3% bovine serum albumin (BSA) (VWR International)

10% v/v normal goat serum (Dako)

PBS up to final volume (ThermoFisher Scientific)

Antibody diluent

1 g BSA (VWR International)

100 ml PBS (ThermoFisher Scientific)

Tris/Boric Acid/EDTA (TBE)

Ready-made 10x Tris/Boric Acid/EDTA (Bio-Rad) was diluted to 1x concentration with dH_2O.

Cell freezing solution

10% DMSO (Sigma)

40% foetal bovine serum (FBS) (GIBCO)

50% growth medium

1x Phosphate buffered saline (PBS)

PBS tablet (GIBCO) was dissolved in required volume of dH_20 to obtain 137 mM NaCl, 2.7 mM KCl and 10 mM phosphate buffer solution.

pH 7.4

Protein lysis buffer

50 mM Hepes (Sigma Aldrich)

100 mM NaCl (Sigma Aldrich)

10 mM EDTA (Fisher Scientific)

1% v/v Triton x100 (Sigma Aldrich)

4 mM sodium pyrophosphate (Sigma Aldrich)

2 mM sodium orthovanadate (Sigma Aldrich)

10 mM sodium fluoride (Sigma Aldrich)

50 mM β-glycerophosphate (Sigma Aldrich)

pH7.5

10 ml of solution made to completion with 1 cOmplete, Mini, EDTA-free Protease Inhibitor Cocktail tablet (Roche).

8% SDS PAGE resolving gel

46% v/v dH_2O

27% v/v ProtoGel 30% acrylamide (National Diagnostics)

25% v/v 1.5 M TrisHCl pH 8.8

10% SDS (Fisher Scientific)

10% ammonium persulfate (Sigma)

0.06% v/v TEMED (Sigma)

SDS PAGE stacking gel

68% v/v dH_2O

17% v/v ProtoGel 30% acrylamide (National Diagnostics)

13% v/v 1.0 M TrisHCl pH 6.8

10% SDS (Fisher Scientific)

10% ammonium persulfate (Sigma)

0.1% v/v TEMED (Sigma)

1x Tris/HEPES/SDS electrophoresis buffer

100 mM Trizma (Sigma Aldrich)

100 mM HEPES (Sigma)

3 mM SDS (Fisher Scientific)

dH_20 up to 1 litre

Transfer buffer

25 mM Trizma (Sigma Aldrich)

192 mM glycine (Fisher Scientific)

made to 1 litre with dH_2O and 20% v/v methanol (Fisher Scientific)

PBST

0.05% v/v Tween 20 (Promega) into 1x PBS

Western blot blocking buffer

50% Odyssey Blocking buffer (LI-COR Biosciences)

50% PBST

Comet lysis buffer

2.5 M NaCl (Sigma Aldrich)

100 mM EDTA (Fisher Scientific)

10 mM Trizma (Sigma Aldrich)

1% N-lauroylsarcosine sodium salt solution (Sigma Aldrich)

0.5% Triton x100 (Sigma Aldrich)

10% DMSO (MP Biomedicals)

2.2 Acquisition of patient samples

BCON TMAs used in Chapter 3 of this thesis were constructed in Manchester University in 2011. TMAs contained 1mm tissue cores extracted from muscle-invasive areas of BCON FFPE tumour biopsies. Specimens were collected between November 2000 and April 2006. Ethics approval was obtained from Tameside and Glossop Local Research Ethics Committee (project 09/H1013/24).

Anal cancer tissue was obtained from Duncan Gilbert. FFPE tumour biopsies were retrieved from patients treated with CRT for non-metastatic SCCA from 2004 to 2009. Ethical approval was obtained (project 11/LO/1032)

Germline DNA, extracted tumour DNA and FFPE bladder cancer biopsies for the work contained in Chapter 5 of this thesis were obtained from the Bladder Cancer Prognosis Programme (University of Birmingham) with the help of Richard Bryan. Ethical approval was obtained (project 06/MRE04/65)

MIBC TMAs included in Chapter 6 of this thesis contained 0.6mm tissue cores extracted from muscle-invasive areas of FFPE TURBT chips, treated with RT at Leeds Teaching Hospitals NHS Trust from January 2006 to February 2009. Ethics approval was obtained from Leeds (East) Research Ethics Committee (project 04/Q1206/62) and NRES Committee London-Bromley (project 13/LO/0540).

2.3 Histology and IHC

2.3.1 Sectioning FFPE tissue

FFPE tissue blocks were incubated in an ice water slush for 30-60 minutes before cutting. Sections of appropriate thickness were cut using a Leica RM2125 RTS Microtome and placed in a 42^{o}C Leica HI1210 histology water bath. Sections were removed from the water bath by picking them up on SuperFrost Plus microscope slides. Slides were placed in racks on a 52^{o}C Leica HI1220 Flattening Table until excess water had disappeared. Sections for IHC or Haematoxylin & Eosin staining were baked in a 50^{o}C incubator overnight. Cut tissue sections were stored at 4^{o}C for a maximum of 1 month.

2.3.2 Manual deparaffinisation, rehydration and dehydration of FFPE tissue sections

Sections were warmed in a 50^{o}C incubator for 15 minutes. Deparaffinisation in xylene (Fisher Scientific) was performed for 5 minutes twice before rehydration in 100%, 100%, 70% and 50% ethanol for 3 minutes each. Complete rehydration was achieved by placing in dH_2O for 3 minutes twice. Dehydration of stained tissue was conducted by incubating slides sequentially in 50%, 70%, 100%, 100% ethanol for 3 minutes each. Slides were then immersed in xylene for 3 minutes twice. Dehydrated tissue was covered using dibutylphthalate polystyrene xylene (DPX) mountant (Sigma Aldrich) and glass coverslips.

2.3.3 Haematoxylin and eosin (H&E) staining

Deparaffinised and rehydrated 4 µm thick tissue sections were placed in filtered Harris haematoxylin (VWR International) for 4 minutes. Slides were washed in circulating water for 1 minute, followed by 1% acid alcohol for 10 seconds. Samples were subsequently placed in Scott's tap water and 0.5 % lithium carbonate for 10 seconds, each. Tissue was washed for 1 minute in circulating water. Slides were briefly dehydrated in 50% ethanol and 70% ethanol before placing in alcoholic eosin Y (Leica) for 2 minutes. Residual eosin was removed by three 20 second 90% ethanol washes. Slides were taken to complete dehydration in absolute ethanol for 3 minutes twice and xylene for 3 minutes twice. Tissue was covered with glass coverslips using DPX mountant.

2.3.4 Automated IHC of FFPE tissue

Four micron thick tissue sections were used in automated staining. A Leica Bondmax system was used with Bond Polymer Refine Detection kit (Leica) and assay-specific reagents: Bond dewax solution (AR9222) (Leica), Bond wash solution x10 concentrated (AR9590) diluted 1:10 with dH_2O, Bond epitope retrieval solution 1 (AR9961), Bond epitope retrieval solution 2 (AR9640). Specific conditions for each antibody stain can be found in appendix A. For an extended generalised protocol see appendix B. Stained tissue was dehydrated and covered using the procedure outlined in 2.3.2.

2.3.5 Manual IHC of FFPE tissue with Dako REAL Detection System, Peroxidase/DAB+, Rabbit/Mouse kit

Deparaffinised and rehydrated 4 µm thick tissue sections underwent heat induced epitope retrieval at 110^{o}C for 45 seconds using a decloaking chamber DC2002 (BioCare Medical) in either citrate or Tris-EDTA buffer. Samples were allowed to cool to room temperature then washed briefly in dH_2O. Samples then underwent a standard wash consisting of two 3 minute washes in 1x TBS. Tissue was peroxidase blocked for 20 minutes, at room temperature, in 3% H_2O_2 (Fisher Scientific). Sections were blocked at room temperature for 15 minutes in avidin block (Vector labs), 15 minutes in biotin block (Vector Labs) and 20 minutes in manual IHC protein block. Each block was followed by standard wash. Primary antibodies were diluted in antibody diluent and incubated on tissue under optimised conditions. For specific antibody conditions see appendix C. Primary antibody was removed and the samples washed in 1x TBS. Samples were incubated at room temperature with biotinylated secondary antibody solution for 15 minutes, washed, and then incubated at room temperature with streptavidin peroxidase (HRP) solution for 15 minutes. Tissue was washed before development of staining with DAB solution for up to 10 minutes. DAB solution was immediately removed from the tissue at the end of development by washing in circulating water for 3 minutes.

2.3.6 Manual IHC of FFPE tissue with Novolink polymer detection system (Novocastra Laboratories)

Deparaffinised and rehydrated 4 µm thick tissue sections underwent heat induced epitope retrieval as outlined for Dako REAL Detection System. Tissue was blocked at room temperature with peroxidase block (Novocastra laboratories) and protein block (Novocastra laboratories) for 20 minutes and 10 minutes respectively with TBS washes after each block. Primary antibodies were diluted in antibody diluent and incubated on tissue under optimised conditions. For specific antibody conditions see appendix C. Primary antibody was removed and the tissue washed with TBS. Novolink polymer (Novocastra laboratories) was applied to sections for 50 minutes at room temperature, and the slides washed in TBS. DAB solution was added to the tissue for up to 10 minutes before immediate removal by washing in circulating water for 3 minutes.

2.3.7 Manual counterstaining DAB stained tissue sections for haematoxylin

Manually DAB stained tissue sections were counterstained with filtered haematoxylin for 15 seconds. Excess haematoxylin was removed from the slide by rinsing in circulating water for 1 minute followed by a 10 second wash in 1% acid alcohol. Haematoxylin was blued by placing slides in Scott's tap water followed by 0.5 % lithium carbonate for 10 seconds each. Tissue was washed in circulating water for 1 minute. Stained tissue was dehydrated and covered as outlined in 2.3.2.

2.3.8 Microdissection of FFPE tissue

Ten micron thick tissue sections plus an additional 4 µm thick tissue section were cut using a Leica RM2125 RTS microtome and placed on SuperFrost Plus microscope slides (Thermo Scientific). Four micron sections were H&E stained using a standard H&E protocol outlined in 2.3.3 and sent to a Consultant Uropathologist to demarcate muscle-invasive areas of tissue. Ten micron sections were placed at -20^{o}C and stored until microdissection. Storage at -20^{o}C was limited to a maximum of 3 weeks. Microdissection of muscle-invasive tissue sections was performed using a 15C carbon steel sterile scalpel blade (Swann Morton). Using the marked up H&E section as a guide, muscle-invasive tissue was scraped from 10 µm tissue sections and placed in a sterile Eppendorf tube. Tissue from all 10 sections was pooled to collect enough material from which to extract RNA.

2.3.9 Scoring of IHC

The Aperio ScanScope^{TM} CS2 digital slide scanner was used to image stained tissue at 400x magnification. Scanned tissue was viewed with Aperio Image Scope viewing software. For whole tissue sections, invasive cancerous areas of stained tissue were marked on H&E sections of corresponding tissue by a trained pathologist. Up to 10 high-magnification images of these areas were captured for scoring.

The standard method of IHC scoring was conducted by at least two independent trained scorers. The intensity of staining was assessed according to a range of 0+ (negative), 1+ (weak staining), 2+ (moderate staining), 3+ (strong staining). To aid intensity scoring, a tutorial package, of images was collated which encompassed the range of staining

present within the tissue cohort. The percentage of cancer cells stained was estimated. Independent scores were collated and a consensus score was reached for each tissue section. For whole tissue sections, a consensus score for the 10 images was achieved by taking the modal intensity value and mean value for percentage of cancer cells stained. In TMA scoring, where multiple cores were present for the same patient, a consensus score for the patient was achieved by taking the modal intensity value and mean value for percentage of cancer cells stained. Semi-quantative scores (SQS) were generated by combining scores for intensity and percentage of cancer cells stained [SQS=intensity × percentage].

For the public scoring of Reverse the Odds (RTO) TMAs, each TMA core was divided into segments, and across all segments was scored between 81 and 3676 times (mean: 405, median: 124). Public scores were aggregated by taking the mean scores for the percentage of cancer cells stained and intensity for a complete core across all segments, ignoring any responses where the participant indicated there was no cancer. As the dataset contained a number of cores obtained from the same patient, core scores were combined by taking the mean for percentage of cancer cells stained and intensity. Percentage and intensity scores were combined to produce either a SQS score or H-score. The H-score was generated by adding the percentage of strongly stained nuclei multiplied by 3, the percentage of moderately stained nuclei multiplied by 2, and the percentage of weakly stained nuclei multiplied by 1 H score=1[1(% cells 1+) + 2(% cells 2+) + 3(% cells 3+)] giving a possible range of 0–300.

To remove biases from taking the mean across multiple users and segments, a correction with clipping was applied to the data. For the cores that had no expert scores,

the correction was based on the intercept and slope between cores scored by both citizen scientists and experts. To correct cores scored by both citizen scientists and experts, the machine learning library sklearn.cross_validation.cross_val_predict with a linear regression estimator was used to conduct 10-fold cross-validation. Any scores that were out-of-bounds of the original scoring range (0-300) were set to the bound (e.g. a corrected proportion of -5% was set to 0%).

Public scores were aggregated and corrected by Dr Peter Smittenaar from the CRUK Citizen Science statistical team, using Python primarily using the "scikit-learn" library.

2.4 Genotyping of rs1805363 SNP and *MRE11* isoform analysis

2.4.1 TaqMan genotyping of rs1805363 SNP on patient samples

A TaqMan SNP Genotyping assay (C_ _11474841_10) was used on patient DNA. Each sample contained 12.5 µl 2x TaqMan Universal PCR master mix, 1.25 µl 20x SNP Genotyping Assay and 10 ng DNA diluted to 11.25 µl in DNase-free water. PCR was performed with a 7500 Fast Real Time PCR system with genotyping parameters. Running conditions consisted of: pre PCR read at 60^{o}C for 1 minute, hold stage at 95^{o}C for 10 minutes then 40 cycles of 95^{o}C for 15 seconds, 60^{o}C for 1 minute followed by post PCR read at 60^{o}C for 1 minute. Analysis was conducted using 7500 software using default settings and described using an allelic discrimination plot.

2.4.2 Sanger sequencing of *MRE11* rs1805363 SNP

PCR was conducted using OneTaq Quik-load master mix with standard buffer (New England BioLabs) following the manufacturer's protocol. Previously designed primers complementary to a 160bp region of DNA flanking *MRE11* rs1805363 G>A SNP (for sequence, see appendix D - *MRE11* rs1805363) were used. Thermocycling conditions were as follows: initial denature at 94^{o}C for 30 seconds then 35 cycles of 94^{o}C for 20 seconds, 58^{o}C for 30 seconds, 68^{o}C for 10 seconds followed by a final elongation step for 5 minutes at 68^{o}C. Products were separated by electrophoresis on a 0.75% agarose (Sigma Aldrich) gel in 1x TBE at 80 V for 1 hour. The gel was visualised using an ultraviolet light box and the 160 bp fragment was cut from the gel. DNA was extracted from the gel using a QIAquick gel extraction kit (Qiagen) following the manufacturers protocol. Extracted samples were then sequenced with the rs1805363 forward primer using the Sanger sequencing service provided by Source Bioscience.

2.4.3 RNA extraction of FFPE tissue

Microdissected tissue was immediately subjected to RNA extraction using an RNeasy FFPE kit (Qiagen) with deparaffinization solution (Qiagen). RNA extraction was performed following manufacturer's guidelines. Successful RNA extraction was confirmed by running 5 µl of RNA on a 1% agarose (Sigma Aldrich) gel in 1x TBE at 80V for 1 hour. RNA was converted into cDNA using a High-Capacity cDNA reverse transcription kit with RNase inhibitor (Applied Biosystems) using the manufacturer's recommendations.

2.4.4 Detection of *MRE11* isoforms in FFPE extracted cDNA using PCR

Primers were designed using the PrimerQuest Tool from Integrated DNA Technologies (available online at https://www.idtdna.com/Primerquest/Home/Index). Primers were designed to flank the MRE11 rs1805363 G>A SNP, producing PCR fragment sizes of 144bp for MRE11 isoform 1 and 251bp for MRE11 isoform 2 (for primer sequences see appendix D - MRE11 isoform). Predesigned primers for GAPDH (for sequence see appendix D - GAPDH) were used in a concurrent PCR as a control for cDNA input and quality. PCR of extracted cDNA was carried out using OneTaq Quik-load master mix with standard buffer (New England BioLabs) following the manufacturer's protocol. Thermocycling conditions were as follows: initial denature at 94^{o}C for 30 seconds then 40 cycles of 94^{o}C for 20 seconds, 58^{o}C for 20 seconds, 68^{o}C for 15 seconds followed by a final elongation step for 5 minutes at 68^{o}C. Products were separated by electrophoresis on a 1% agarose (Sigma Aldrich) gel in 1x TBE at 80 V for 1.2 hours. The gel was visualised using a Molecular Imager Gel Doc XR system (BioRad) and analysed with Image Lab software (BioRad).

2.5 Cloning

2.5.1 MRE11 exon 16 site directed mutagenesis

A plasmid containing full length MRE11 cDNA sequence (NCBI Reference Sequence: NM_005591.3 CDS) had previously been created by Dr Sarah Jevons in the lab using the pcDNA^{TM}3.3 TOPO® cloning kit (ThermoFisher Scientific). The QuikChange II XL site-directed mutagenesis kit (Agilent) was used to specifically excise MRE11 exon 16 from the plasmid containing full length MRE11. Primers were designed using the QuikChange

Primer Design tool available from

(http://www.genomics.agilent.com/primerDesignProgram.jsp). For primer sequences see appendix D - SDMexon16. Reactions, thermal cycling, Dpn1 digestion and transformation with XL10 ultracompetent cells were conducted as per manufacturer's guidelines.

Plasmids were extracted from bacterial cultures using a Miniprep kit (Qiagen) using the standard protocol. Deletion of exon 16 in the plasmid was confirmed by Sanger sequencing carried out by Source Bioscience using the TK polyA reverse sequencing primer.

2.5.2 Amplification of MRE11 exon16 plasmid

Once deletion of exon 16 in the plasmid colonies was confirmed, one plasmid was chosen to be carried forward as the MRE11 exon16 negative plasmid. One microlitre of plasmid was incubated with 50 µl DH5α cells for 30 minutes on ice, before heat shock at 42^{o}C for 45 seconds. One millilitre of LB/SOC media (Sigma Aldrich) was added to the cells and incubated whilst shaking for 1 hour at 37^{o}C in a New Brunswick Innova incubator shaker at 200 rpm. Then 200 µl of mixture was plated under flame on LB agar plates containing 100 µg/ml ampicillin. Plates were incubated overnight at 37^{o}C.

Colonies resulting from transformations were picked using a sterile pipette tip and placed in a starting culture of 5 ml of LB containing 100 µg/ml ampicillin for 6 hours. The starting culture was transferred to 200 ml of LB containing 100 µg/ml ampicillin and incubated overnight, whilst shaking, at 37^{o}C in a New Brunswick Innova incubator shaker at 200 rpm. Plasmid was extracted from overnight cultures using a Maxiprep kit (Qiagen) following the manufacturer's guidelines.

2.6 Cell lines and culture

2.6.1 Culture conditions

All cell lines were maintained in 37^{o}C, 5% CO_2 humidified incubator. Unless otherwise stated, growth medium was RPMI 1640 growth medium, 1 mM L-glutamine (Sigma Aldrich) supplemented with 10% FBS (GIBCO). Growth medium for MRE11 knock down cells and MRE11 scrambled control cells was supplemented with 1 µg/ml puromycin (Cayman Chemical). Growth medium for stable MRE11 clones (MRE11 full length, MRE11 exon16 negative) were supplemented with 1 µg/ml puromycin and G418 solution (Roche) at 20 µl/ml. All reagents were stored at 4^{o}C except puromycin which was frozen at -20^{o}C and thawed just before use.

2.6.2 Cell culture

Fresh media were pre-warmed to 37^{o}C. Trypsin (Sigma Aldrich) and PBS were warmed to room temperature. Medium was removed from flasks and cells were washed in 5 ml of PBS. An appropriate volume of trypsin was applied to cover cells before replacing the flasks in the incubator until cells detached. Trypsin was quenched in ≥1.5 volumes of medium and the mixture transferred to a centrifuge tube before spinning in a Heraeus Labofuge 400 centrifuge (Fisher Scientific) at 1,300 rpm for 3 minutes. Supernatant was removed from the cell pellet by aspiration. Cells were resuspended in an appropriate volume of medium and plated at appropriate seeding densities.

2.6.3 Cell cryopreservation

Pelleted cells were resuspended in 1 ml freezing solution, using dropwise addition of the first 200 µl of solution and gentle pipetting with the remaining solution. Resuspended cells were transferred to 2 ml cryovials (Corning) and placed directly into a Mr Frosty freezing container (ThermoFisher Scientific). Cells were left at -80^{o}C overnight before transferring to liquid nitrogen storage.

2.6.4 MRE11 scrambled control transfection

A stable MRE11 knock down cell line from the T24 parental cell line had been created and validated by Dr Sarah Jevons. The stable MRE11 knockdown was performed using a pSilencer 2.1-U6 puro vector (LifeTechnologies) and used shRNA designed according to manufacturer's instructions. shRNA sequences were: MRE11 shRNA sense: 5'-GATCCGAACCTGGTCCCAGAGGAGTTCAAGAGACTCCTCTGGGACCAGGTTCTTTTTTGGAAA-3' and MRE11 shRNA antisense: 5'-AGCTTTTCCAAAAAAGAACCTGGTCCCA GAGGAGTCTCTTGAACTCCTCTGGGACCAGGTTCGC-3'. This cell line was puromycin resistant. A complementary non-silencing scrambled control plasmid in pSilencer 2.1-U6 puro vectors (Life Technologies) for the MRE11 knockdown cell line had also been generated by Sarah Jevons but had not been used with the T24 cell line. Scrambled control shRNA was designed using the same bases as MRE11 shRNA, but in a random order which had no homology with the human genome according to BLAST (NCBI). The scrambled control sequences were: MRE11 Scrambled shRNA sense: 5'-GATCCACGAAGTGCGAGTCGCAGCTTCAAGAGAGCTGCGACTCGCACTTCGTTTTTTTGGAAA-3' and MRE11 Scrambled shRNA antisense 5'-AGCTTTTCCAAAAAAACGAAGTGC GAGTCGCAGCTCTCTTGAAGCTGCGACTCGCACTTCGTG-3'. Oligonucleotides were

synthesized by Qiagen. Scrambled control pSilencer 2.1-U6 puro vector was transfected into T24 cells. T24 cells were grown in two 15 cm cell culture dishes to 60% confluency. Cells were then either transfected with 10 µg scrambled control plasmid using FuGENE HD transfection reagent (Promega) or were left untransfected. In the transfected plate, plasmid DNA was added to 500 µl of Optimem (ThermoFisher Scientific) and allowed to disperse at room temperature for 5 minutes. FuGENE HD transfection reagent (Promega) was added to the DNA mix in a 3:1 ratio of FuGENE (µl): DNA (µg). The mixture was incubated at room temperature for 15 minutes before adding dropwise to cells. Both transfected and untransfected cells were then incubated in a 37^{o}C 5% CO_2 humidified incubator overnight. The next day the medium was replaced with standard growth medium and returned to the incubator for a further 12 hours. The medium was then removed and replaced with selection medium (standard growth medium supplemented with 2 µg/µl puromycin). Cells were allowed to grow in selection medium until all cells on the untransfected control plate were dead. During the growth period fresh selection medium was added every two days.

2.6.5 Selecting for stable Clones

Clones of transfected cells were established by marking the location of clonal populations on culture dishes. Trypsin was gently added to each dish, which was kept still for 5 minutes. Ten microlitres of trypsin from each of the marked colony areas was carefully extracted, by pipetting, making sure not to disturb the dish, and transferred to individual wells of a 96-well plate containing 150 µl of selection medium. Cells were allowed to grow to 80% confluency before transferring to 6-well plates. Cells were transferred to a T75 flask and grown to 80% confluency. Medium was removed from

the flask and the cells washed in PBS. Two millilitres of trypsin was added to the flask before replacing it in the incubator to allow cells to detach. Trypsin was quenched in 8 ml of medium and cells were washed from the bottom of the flask by pipetting up and down. Five millilitres of cell mixture was then plated onto a 10 cm cell culture dish, from which cells could be later harvested for confirmation of stable transfection. The remaining 5 ml of cells were frozen using procedure described in section 2.5.3.

2.6.6 MRE11 re-expression clones

To create stable MRE11 re-expression clones, the T24 background MRE11 knockdown cell line was transfected with either MRE11 full length plasmid or MRE11 exon16 negative plasmid using FuGENE as per the method described above. Selection medium for MRE11 re-expression clones was standard growth medium supplemented with 1 µg/µl puromycin and 30 µl/ml G418 solution (Roche).

2.7 Molecular characterisation assays

2.7.1 Harvesting cell pellets

Cells were grown to 80-90% confluency on cell culture dishes and harvested. For harvest, cells were washed in ice cold PBS before scraping using a cell scraper. Cells were transferred to a 1.5ml Eppendorf tube and centrifuged at 4^{o}C for 8 minutes at 14,000 rpm in an Eppendorf 5430R microcentrifuge (Fisher Scientific). Supernatant was carefully removed from the cell pellet before subsequent washing in 1 ml of ice cold PBS and spinning at 4^{o}C for 8 minutes at 14,000 rpm. The PBS wash was removed from the cell pellet. Cell pellets were either snap frozen and stored short term at -20^{o}C or used immediately.

2.7.2 Cell lysates and determination of protein concentration

Cell pellets were re-suspended in approximately two volumes of protein lysis buffer before incubating on ice for 5 minutes. Samples were briefly sonicated using an Ultrasonic Homogeniser 300 VT (Biologic Inc) followed by incubation on ice for 5 minutes. Lysates were centrifuged at 4^{o}C for 8 minutes at 14000 rpm in an Eppendorf 5430R microcentrifuge.

2.7.3 Determination of protein concentration

Protein concentration of lysates was determined using Pierce^{TM} BCA Protein Assay kit (Thermofisher Scientific). Ten microlitres of protein standard containing 0-2mg of protein were loaded in duplicate into 24 wells of a 96 well plate. Then 1 µl of lysate was loaded onto the same plate in duplicate, and 200 µl of combined BCA reagents were then added to samples (BCA reagent A and BCA reagent B in a 50:1 ratio). Samples were incubated at 37^{o}C for 30 minutes before measuring their absorption. Absorption was measured at 562nm using a POLARstar Omega plate reader. The average absorption of duplicate protein standards was used to generate a standard curve from which the average lysate absorption could be used to determine protein concentration. Cell lysates were either used immediately or snap frozen and stored at -80^{o}C.

2.7.4 Western blotting

Ten to 40 µg of protein lysate were mixed with an equal volume of Laemmli 2× concentrate sample buffer (Sigma Aldrich) and made up to a standard volume between 10 and 20 µl with Protein Lysis buffer. Samples were boiled at 95^{o}C for 5 minutes and loaded into SDS-PAGE gels. All samples except those to be assessed by p21 #2947P (Cell Signalling Technology), anti-keratin 20 E16444 (Amsbio) and ASYM25 09-814 (Millipore)

were loaded onto homemade 8% SDS-PAGE gels using Mini-PROTEAN Tetra handcast systems (Bio-rad). Samples which were to be assessed with the p21 #2947P (Cell Signalling Technology), anti-keratin 20 E16444 (Amsbio) or ASYM25 antibody 09-814 (Millipore) were loaded onto 4-20% Mini-PROTEAN® TGX™ precast protein gels (Bio-Rad). Gels were run in Mini-PROTEAN Tetra vertical electrophoresis cells at 80 V for 1-1.3 hours in 1x Tris/HEPES/SDS electrophoresis buffer. Proteins were transferred from gels onto nitrocellulose membrane using a Mini Trans-Blot Cell (Bio-Rad) in ice cold transfer buffer at 40 V for 2 hours on ice. A nitrocellulose membrane containing the transferred protein was washed briefly in PBST and blocked with western blot blocking buffer for 30 minutes at room temperature on a rocking stage. Primary antibodies were diluted to the appropriate concentration (see appendix E) in western blot blocking buffer and incubated overnight at 4^{o}C on a rocking stage. Primary antibody was removed and the membrane washed three times for 10 minutes in PBST. Appropriate species of IRDye secondary antibody (see appendix E) (LI-COR Biosciences) was added to the membrane at a 1:5000 dilution in western blot blocking buffer and incubated at room temperature for 1 hour. Secondary antibody was removed and membrane was washed three times for 10 minutes with PBST. Protein binding was visualised using the Odyssey CLx imaging system and analysed with Image Studio Lite Software (LI-COR Biosciences).

2.7.5 Generation of cDNA from cell pellets

RNA was extracted from cell pellets using an RNeasy Plus Mini kit (Qiagen) following manufacturer's guidelines. Cellular disruption and homogenisation was carried out with a QIAshredder homogenizer (Qiagen). RNA was converted to cDNA using a High-Capacity cDNA reverse transcription kit with RNase inhibitor (Applied Biosystems) using the manufacturer's recommendations. cDNA was quantified using a Nanodrop 1000 spectrophotometer (Thermo Scientific).

2.7.6 qPCR of *MRE11* isoforms

Primers were designed for qPCR using the PrimerQuest tool from Integrated DNA Technologies (available online at https://www.idtdna.com/Primerquest/Home/Index). Primers were designed to either bind a central region of MRE11 (for sequence see appendix D - Central MRE11) or bind the exon15-exon16 junction (for sequence see appendix D - Exon15-16). A previously designed ATP5B primer pair was used as endogenous control (for sequence see appendix D - ATP5B). Reactions for ATP5B, central MRE11 and exon 15-16 were set up, in duplicate, using SYBR green reagents (ThermoFisher Scientific). Each 10 µl reaction consisted of 1 µg cDNA, 4 µM Forward Primer, 4 µM Reverse Primer, 50% v/v SYBR green mastermix made up to total volume with dH_20. qPCR was performed on a 7500 Fast Real Time PCR system, using the comparative CT parameters. Running conditions were: hold 50^oC for 20 seconds, hold 95^oC for 10 minutes then 40 cycles of 95^oC for 15 seconds, 60^oC for 1 minute followed by a melt curve analysis of 95^oC for 15 seconds, 60^oC for 1 minute, 95^oC for 30 seconds, 60^oC for 15 seconds. qPCR was analysed using 7500 software using default settings and qPCR was run in triplicate using cDNA obtained from three independent cell harvests.

2.7.7 Clonogenic cell survival assay

Harvested cells were counted and diluted to 700 cells per 10 ml in appropriate growth medium. Ten millilitres of cell dilution was added to 10 cm cell culture dishes and placed in a 37°C 5% CO_2 humidified atmosphere for 1 hour to settle. Plates were treated with either 0, 2, 4, 6 or 8 Gy of irradiation from a caesium source (Gamma Medical Services, Cs137). Cells were left to grow in a 37°C 5% CO_2 humidified incubator. A period of 8 days was selected as the optimal duration of growth as it enabled the identification of viable colonies without the colonies merging. The dilution of 700 cells per 10 ml of growth medium was used to enable single cell colony formation after the growth period rather than colonies comprised of multiple cells. At the end of the growth period medium was removed and plates were washed twice in cold PBS. Cells were fixed in methanol (Fisher Scientific) and acetic acid (Fisher Scientific) in a 3:1 ratio methanol: acetic acid for 10 minutes at room temperature. Fixative was removed completely and cellular colonies stained for 30 minutes with Brilliant Blue R concentrate (Sigma Aldrich). The stain was removed and the plates washed in water. Plates were left to dry overnight. Colonies were counted using a ColCount machine (Oxford Optronix). Each experiment was performed in biological triplicate. Raw colony counts were normalised to the 0 Gy treatment for each cell line and the median of triplicate experiments taken. Normalised data were plotted using a linear quadratic model using GraphPad Prism 6.0 Software. One-way ANOVA analysis for the significance of differences between all cell lines at each radiation dose and Student's t-test for significance were conducted using GraphPad Prism 6.0 Software.

2.7.8 Neutral comet assay

The comet assay or single cell gel electrophoresis assay is a technique used to detect DNA damage on the individual cell level. Cells are embedded in agarose, lysed, subjected briefly to an electric field, stained with a fluorescent DNA-binding stain, and viewed using a fluorescence microscope.

The theory of the technique is as follows: DNA damage causes relaxation of the DNA structure and exposes the charge on DNA ends, which when subjected to an electric field, migrate toward the positive anode. This produces a structure that resembles a comet with a brightly fluorescent head region, containing high-molecular-weight intact DNA, and a tail region, containing migrated DNA. To quantify the level of DNA damage, the DNA content and distribution of DNA in the migrated tail is calculated in a ratio to the DNA content and distribution of DNA in the head region. A long DNA tail with a high DNA content, relative to the head region, corresponds to a high level of DNA damage. The lysis and electrophoresis steps of the comet assay can be conducted using neutral or alkaline (>pH13) buffers. Neutral conditions are used to detect DSB, while alkaline conditions detect the combination of SSB, DSB and alkali-labile sites (Olive & Banáth, 2006). Due to the primary role of MRE11 in the resolution of DSBs, the neutral comet assay was used in this work.

Cells were plated into 6-well cell culture plates. In each case, 120,000 cells per well were plated for MRE11 knockdown cells, MRE11 full length and MRE11 exon16 clones and 100,000 cells were plated per well for MRE11 scrambled control cells. Cells were incubated overnight in a 37^{o}C 5% CO_2 humidified incubator. SuperFrost Plus microscope slides (Thermo Scientific) were pre-coated with 800 µl of 1% normal melting point

agarose (Sigma Aldrich) in dH_2O. Agarose was spread across the slide using a 22x50mm coverslip (VWR International). The coverslip was removed after approximately 1 minute and the agarose was allowed to air dry onto the slide overnight. Plated cells were irradiated on ice with 10 Gy from a caesium source and left to repair in a 37°C, 5% CO_2 humidified atmosphere for either 0, 30, 120 minutes. One set of cells was left untreated to control for basal levels of DNA damage. Once the allocated repair time was over, cells were placed on ice for harvesting. The medium was removed and cells were washed twice in ice cold PBS. The PBS was removed before the addition of 200 µl trypsin. Cells were trypsinised under microscopic control. Trypsin was quenched with 450 µl cold standard growth medium and cells were washed to obtain an even suspension and complete detachment from the well. Then 800 µl low melting point agarose (Bio-Rad) in 1x PBS, warmed to 37°C, was gently and thoroughly mixed with 200 µl of the cell suspension. The cell suspension was immediately placed on pre-coated SuperFrost Plus microscope slides (Thermo Scientific), kept on ice, and spread across each slide using a 22x50mm coverslip (VWR International). The coverslip was removed when the agarose had just set. Agarose slides were placed in cold comet lysis buffer and incubated at 4°C in the dark for 2 hours. Agarose slides were washed three times in cold 1x TBE before submerging in 1x TBE at 4°C in the dark for 1 hour. Slides were placed in a C.B.S. Scientific SGE-040-02 electrophoresis tank filled with fresh, 22-24°C 1x TBE and run at 1.3 V/cm for 32 minutes at room temperature. Slides were allowed to air dry in the dark overnight before staining with 1x SYBR Gold nucleic acid gel stain from a 10,000x concentrate stock (ThermoFisher) in dH_2O for 17 minutes. Slides were washed in dH_20 for 5 minutes and allowed to air dry in the dark overnight.

Comet Olive tail moments were measured using Komet 5.5 software (Kinetic Imaging). Comet assays were performed in duplicate for each condition mentioned. For each condition 50 comet tails were measured in each of the two duplicate slides, so a total of 100 comet tails were measured for each condition. The mean tail moment was taken for each of the conditions. Comet assays were repeated three times and the average of all three experiments taken. The mean tail moment for the scrambled control cell line at 10 Gy 0 hr was taken as representing 100% of damage and all other conditions were normalised and compared to this. Data for comet assays was plotted and Student's t-tests were used to assess the significance of differences between conditions using GraphPad Prism 6.0 Software.

2.7.9 Immunoprecipitation with Protein G and A beads

Ten microlitres of Dynabeads Protein G (ThermoFisher Scientific) or Pierce Protein A magnetic beads (ThermoFisher Scientific) slurry were washed 3x in 100µl of protein lysis buffer. A DynaMag spin magnet (ThermoFisher Scientific) was used to remove the washes. Dynabeads Protein G were used for immunoprecipitation of NBS1 all other immunoprecipitations used Pierce Protein A magnetic beads. Washed beads were resuspended in 10 µl protein lysis buffer and added to the protein lysate for preclearing. Preclearing was conducted for 1 hour at 4^{o}C on a Stuart SB3 tube rotator. Lysate was removed from the beads with a DynaMag spin magnet (ThermoFisher Scientific) and placed on ice in a clean 1.5ml Eppendorf tube. Five percent of the lysate volume was removed and placed in a separate Eppendorf tube with an equal volume of Laemmli 2× concentrate sample buffer to act as a 5% loading fraction. Load fraction samples were boiled at 95^{o}C before placing at -20^{o}C until completion of the protocol. One microlitre

of target protein antibody or 1 µl of IgG control antibody was added to the lysates and incubated overnight at 4°C on a Stuart SB3 tube rotator. The target antibodies used were: MRE11 ab214 (Abcam), RAD50 ab89 (Abcam), NBS1 NB110-143 (Novus Biologicals), anti-dimethyl-arginine antibody, asymmetric (ASYM25) 09-814 (Millipore). Anti-FLAG M2 F1804 (Sigma Aldrich) or normal rabbit IgG #2729S (Cell Signalling Technology) antibodies were used as an IgG control antibodies to complement target antibodies derived in mouse or rabbit, respectively. Ten microlitres of freshly washed Dynabeads Protein G or Pierce Protein A magnetic beads were added to the lysate-antibody mix and left to rotate at 4°C for 1 hour. Lysate was removed from the beads using the magnet. Beads were washed 6x in 100 µl cold PBST. PBST was removed and 10 µl Laemmli 2× concentrate sample buffer was added to the beads before boiling at 95°C for 5 minutes. The load fraction and immunoprecipitated samples were loaded onto SDS page gels and run under standard western blotting procedures.

2.7.10 Immunoprecipitation with anti-HA beads

Anti-HA bead immunoprecipitations followed a similar protocol to immunoprecipitations with Protein G and A beads. Ten microlitres of washed Pierce Protein A magnetic beads were used to pre-clear cellular lysates for 1 hour at 4°C on a Stuart SB3 tube rotator. Lysate was transferred from beads to a clean 1.5 ml Eppendorf tube, using a magnet. Five percent of the lysate volume was removed and placed in a separate Eppendorf tube with an equal volume of Laemmli 2× concentrate sample buffer to act as a 5% loading fraction. Load fraction samples were boiled at 95°C before placing at -20°C until completion of the protocol. Ten microlitres of washed Pierce Anti-HA magnetic beads (ThermoFisher Scientific) were added to each of the samples and

incubated overnight at 4°C rotating. Lysate was removed from the beads with the aid of a magnet. Beads were washed six times in 100 µl cold PBST. Ten microlitres of Laemmli 2× concentrate sample buffer was added to the beads before boiling at 95°C for 5 minutes. The load fraction and immunoprecipitated samples were loaded onto SDS page gels and run under standard western blotting procedures.

2.7.11 Quantification of immunoprecipitation blots

Bands of interest within immunoprecipitation blots were quantified using Image Studio Lite software (LI-COR Biosciences). The amount of immunoprecipitated protein was quantified as a percentage of the load fraction. Co-immunoprecipitated proteins or methylation status were quantified as a percentage of the load fraction and then normalised to the amount of immunoprecipitated protein. To assess the variation in co-precipitation or methylation in MRE11 full length and MRE11 exon16-ve clones, MRE11 exon16 normalised co-precipitation values were further normalised against the MRE11 full length values.

2.8 Irradiation of cells

All irradiation was delivered via a Caesium-137 γ-ray source, using a dose-rate of 1.7 Gy per minute.

2.9 Statistical analysis

For Chapter 3 of this thesis, statistical analysis for Wilcoxon signed rank test, intraclass correlation coefficients and distributions of MRE11 staining were conducted in SPSS statistics 24 (IBM) by Alexandra K Walker. Kappa analysis, Kaplan-Meier plots with log rank statistics and Cox proportional hazards analyses were conducted in Stata version

14·1 (StataCorp, College Station, TX) by Dr Victoria Strauss from the Centre for Statistics in Medicine (University of Oxford).

For Chapter 4 of this thesis, Kaplan-Meier plots, Cox proportional hazards analysis and the correlation matrix were conducted in R (R Core Team, 2015) by Dr Christiana Kartsonaki from the Department of Population Health (University of Oxford).

For Chapter 6 of this thesis, Spearman's rank correlation and quadratic-weighted kappa for the correlation between public and 'expert' scores was conducted by Dr Peter Smittenaar from the CRUK Citizen Science statistical team. The Spearman's rank and Kappa statistics for the correlation between 'expert' scorers was performed by Alexandra K Walker. Kaplan-Meier plots, Cox proportional hazards analysis and the correlation matrix were conducted in R (R Core Team, 2015) by Dr Christiana Kartsonaki from the Department of Population Health (University of Oxford).

Interpretation of statistical analysis was conducted by Alexandra K Walker. This included the identification of significant associations and the strength of correlations. The strength of a correlation or level of accuracy was determined as: very high if the correlation coefficient was between 1.0-0.9, high (correlation coefficient: 0.89-0.7), moderate (correlation coefficient: 0.69-0.5), low (correlation coefficient : 0.49-0.3) or negligible (0.29-0). These cut-off points are commonly used for medical research (McHugh, 2012; Mukaka, 2012).

In the interpretation of survival statistics, significant p values were identified. Visual assessment of Kaplan-Meier plots and 95% confidence intervals were then used to determine if the significance identified was a product of a 'real' association or resulting from chance and small sample size. If a consistent separation between the curves of a

Kaplan-Meier plot was present, over time, this would be indicative of a 'real' association in the data. Whereas overlapping curves would not imply association. Additionally, curves which contain large 'steps', such as the MRE11 C-terminal plots in Appendix K, often contain a limited number of subjects and are unlikely to be reliable (Dudley *et al*, 2016).

Kaplan-Meier curves were also used to guide the identification of more appropriate cut-off points for Cox proportional hazards modelling, such as the identification of the >75th percentile for initial MRE11 staining and use of the median value for CK20 staining performed in chapter 6.2.4.

Unless otherwise stated all other statistical analysis was conducted using GraphPad Prism 6.0 Software by Alexandra K Walker.

Chapter 3 Validating MRE11 as a predictive biomarker in MIBC using automated IHC

3.1 Introduction

MIBC is treated with either RC or RT, with both treatments achieving comparable CSS rates (Kotwal *et al*, 2008; Booth *et al*, 2014). Due to poor accrual, there is currently no randomised trial data comparing patient outcome between RC and RT (Huddart *et al*, 2010) but current UK guidelines recommend that patients with MIBC suitable for radical treatment should be offered a choice between RC and RT with a radiosensitiser (NICE, 2015). For those patients deemed suitable for either RC or RT, the choice of treatment is made by the patient after discussion with a urologist and clinical oncologist. Despite the comparable rate of efficacy between treatment options, as with all therapies, some patients will respond better to RT then others. Furthermore, RC can lead to significant peri- and postoperative complications (Kiss *et al*, 2016) and reduction in QOL (Caffo *et al*, 1996). Therefore, there is a need for predictive biomarkers in MIBC. Not only would a predictive biomarker aid in patient decision making but, by directing patients toward their 'optimal' therapy, a predictive biomarker could help increase survival rates.

There are currently no validated biomarkers for MIBC to predict patient response to either RC or RT (Forker *et al*, 2015). However, the results from two independent groups have identified the DDR protein, MRE11, as a potential predictive biomarker. Using IHC on retrospective MIBC tissue collections, high levels of MRE11 were found to associate with improved outcome in patients who received radiotherapy-based therapy but had no association with outcome in patients who underwent RC (Choudhury *et al*, 2010; Laurberg *et al*, 2012). To progress this finding into one of real clinical relevance, further

work is needed to confirm the relationship of MRE11 with clinical outcome in a large cohort of tissue samples, to standardise the MRE11 IHC assay and validate its reliability.

A multicentre study between the Anne Kiltie lab (Oxford), Catherine West lab (Manchester) and Nicholas James lab (Birmingham) was embarked upon to develop the MRE11 IHC assay to the appropriate clinical standards and to further define the relationship of MRE11 to clinical outcome in an analysis of retrospective tissue collections. In this study, tissue sections from FFPE MIBC tumours from three cohorts (two RT and one RC) of patients were used. The RT cohorts contained samples from the BCON trial (Hoskin *et al*, 2010) and the BC2001 trial (James *et al*, 2012). The RC cohort contained samples from a surgical series from the Christie NHS Foundation Trust. Details of RT treatments for the two RT cohorts (BCON and BC2001) are described in detail Hoskin *et al* (2010) and James *et al* (2012). Details of the RC technique have been described previously in Ramani *et al* (2010).

This current study of MRE11 IHC in MIBC would amount to the third analysis of MRE11 staining in retrospective MIBC tissue. However, as opposed to previous investigations this study also examined the reliability and reproducibility of the MRE11 IHC assay, which is required to progress the biomarker toward prospective validation in a randomised clinical trial. Additionally, in this investigation MRE11 IHC staining was performed on automated platforms. The reasons for this are twofold. Firstly, as discussed in section 1.3 of this thesis, automated IHC staining platforms are the primary method of performing IHC in the clinic. Secondly, it is well established that automated staining platforms increase not only the reliability but also the throughput of the technique (Warford *et al*, 2004). Both of these features would be highly advantageous

in the clinical setting and so establishment of their use in the MRE11 IHC assay would be beneficial.

Tissue staining with automated staining platforms was conducted with a standardised protocol in the different centres. The BCON tissue cohort was stained and scored in both Oxford and Manchester. The BC2001 tissue cohort was stained in Oxford and Birmingham but scored in Oxford. The RC cohort was stained and scored in Manchester only.

Levels of MRE11 in muscle-invasive sections of tissue were assessed by independent scoring of the percentage of cancer cells stained (% positivity) and the intensity of staining (intensity) by a least two scorers in both Oxford and Manchester. Consensus between independent scores was achieved in each centre and then scores were combined to produce a semi-quantitative score (SQS) for that centre.

In the scoring of stained tissue, a pathologist was utilised to demark areas of muscle-invasive pathology within tissue sections and was involved in the development of a set of 40 images used to train scorers. The pathologist was also available to provide an additional opinion for cases where there was uncertainty in scoring, such as where tissue morphology appeared to distort intensity, and to provide a quality control check on scoring. Unfortunately, due to the high demand and time pressures currently placed upon clinical pathologists, we could not find a pathologist who had the availability, to score all of stained tissue sections in this study.

SQS were assessed for association with clinical outcome. The relationship between MRE11 and CSS was analysed as in Choudhury *et al* (2010). Choudhury *et al* (2010) identified that the relationship between MRE11 and CSS was not linear, and therefore,

categorising MRE11 scores into quartiles was more appropriate then analysing the data as a continuous variable. They plotted a Kaplan-Meier curve to visually assist with the choice of the optimal cut-off point and based on this, log rank statistics were conducted bimodally. Survival was compared between the patients with MRE11 scores below the 25th percentile to those above the 25th percentile. In this current analysis of MRE11, there was no visual grouping or difference between the Kaplan-Meier curves for each of the quartiles and therefore log rank statistics were conducted between all quartiles.

The categorisation of biomarker levels into groups, is used in situations where there is a nonlinear relationship between the variable and outcome and also to aid analysis. However, the act of grouping variables can mask associations which may be present in the data, especially if only two groups are created. Consequently, when deciding on a cut-point for a study the use of several categories, at least for the initial analysis, is preferable from the statistical perspective. This is because less information is lost and variance in risk can be assessed across the range of values for the marker (Simon & Altman, 1994).

The median value is often used as a cut-point in biomarker studies. However, if this point is chosen arbitrarily it can mask results. Similarly using the significance of p values to determine the cut-point creates biased results and false positives (Simon & Altman, 1994). Therefore, the use of a Kaplan-Meier curve, to visually assist in the choice of a cut-off point is commonly conducted in biomarker analysis. The benefit of this, is that the results are generated from the data and therefore are not as arbitrary as choosing the median value outright. It is also less biased than basing a cut-off on p values.

An alternative method of determining a cut-point, is to base it on the distribution the biomarker level (or SQS) within the patient cohort. This does not use clinical outcome data, but instead depends on the marker level the majority of patient's exhibit. For example, if most patients have a marker level close to zero with remaining population having a mean level of 10 with a standard deviation of 2, then a cut-off point of 5 for positivity or high versus low expression would be appropriate, regardless of what the median value of the marker in the patient population is (Simon & Altman, 1994).

The results of the analysis between MRE11 quartiles and CSS in the current MRE11 study, found no significant association between MRE11 and outcome following RT in the BCON cohort (Oxford Log rank p=0.81, 3 yr CSS ≤25th percentile: 62%, 26-50th: 54%, 51-75th: 65% and >75th percentiles: 42%; Manchester Log rank p=0.67, 3 yr CSS ≤25th percentile: 66%, 26-50th: 52%, 51-75th: 59% and >75th percentiles: 49%) or the BC2001 cohort (Oxford Log rank p=0.93, 3 yr CSS ≤25th percentile: 60%, 26-50th: 58%, 51-75th: 59% and >75th percentiles: 50%) (Figure 3.1). However, concern was raised about the quality of MRE11 staining on the samples. The majority of samples were intensely stained and in all samples 90-100% of cells were positively stained. Unlike previous studies, tissue sections negative for staining were conspicuous by their absence and high levels of background staining were observed.

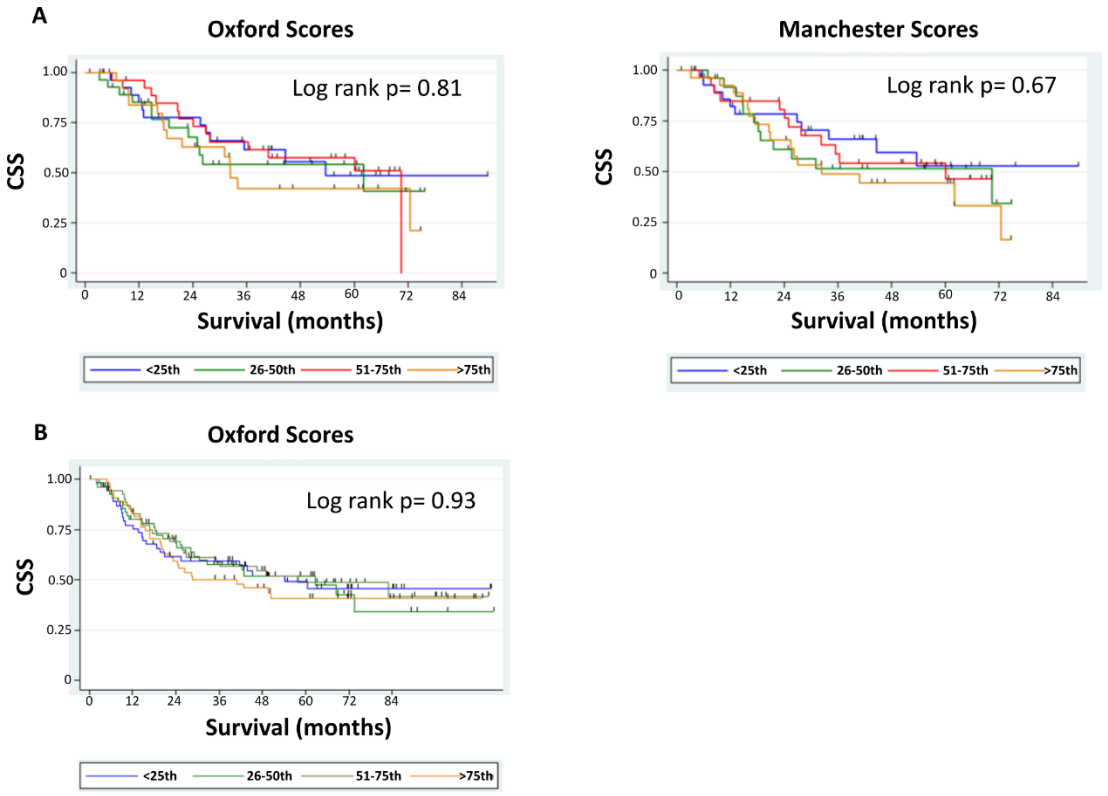


Figure 3.1: **The association between MRE11 SQS and CSS for the BCON whole section tissue cohort.** (A) Kaplan–Meier survival curves for the BCON RT tissue cohort (n=112) stratified according to the <25th, 26-50th, 51-75th and 75th percentiles of MRE11 SQS. The curve for Oxford scoring is displayed on the left of the figure. The curve for Manchester scoring is displayed on the right of the figure. No significant association was identified between MRE11 SQS and 3 year CSS for either Oxford or Manchester scoring (Oxford log rank p= 0.81, Manchester log rank p=0.67); (B) Kaplan-Meier survival curves for the BC2001 RT tissue cohort (n=221) stratified according to the <25th, 26-50th, 51-75th and 75th percentiles of MRE11 SQS. This cohort was scored in Oxford only. No significant association between MRE11 SQS and 3 year CSS was identified (Log rank p= 0.93). Statistics performed by Dr Victoria Strauss.

The IHC staining for this study was conducted using an automated system, whereas previously reported studies had utilised a manual IHC protocol. It was hypothesised that the automated protocol for MRE11 IHC was not optimal and the scoring of this staining was likely to be an inaccurate representation of MRE11 levels in the tissue. Further optimisation of automated MRE11 IHC could therefore alter the scoring of tissue and offer a more accurate assessment of the associations with clinical outcome.

It was clear that the MRE11 staining in this study was not up to the standard required to achieve the study objectives. Therefore, I set about optimising the automated MRE11 IHC protocol with three clear aims. These aims were: to determine if a difference in staining and scoring was evident after automated MRE11 IHC had been re-optimised, to assess the reliability of re-optimised IHC scoring between the Manchester and Oxford centres and to define the relationship between MRE11 and clinical outcome following RT after IHC re-optimisation.

3.2 Results

3.2.1 Assessing the MRE11 staining and scoring of BCON tissue

To assess the concordance between Oxford and Manchester scoring of the BCON tissue cohort, a Wilcoxon signed rank test (WSR) was performed for the SQS produced in each centre. This analysis identified that there was significant variation between the distributions of scores produced in Oxford and Manchester ($p \leq 0.001$). To investigate this further, histograms and boxplots were generated for the SQS produced in each centre and a table summarising the median values and range of scores was created (Figure 3.2, Table 3.1).

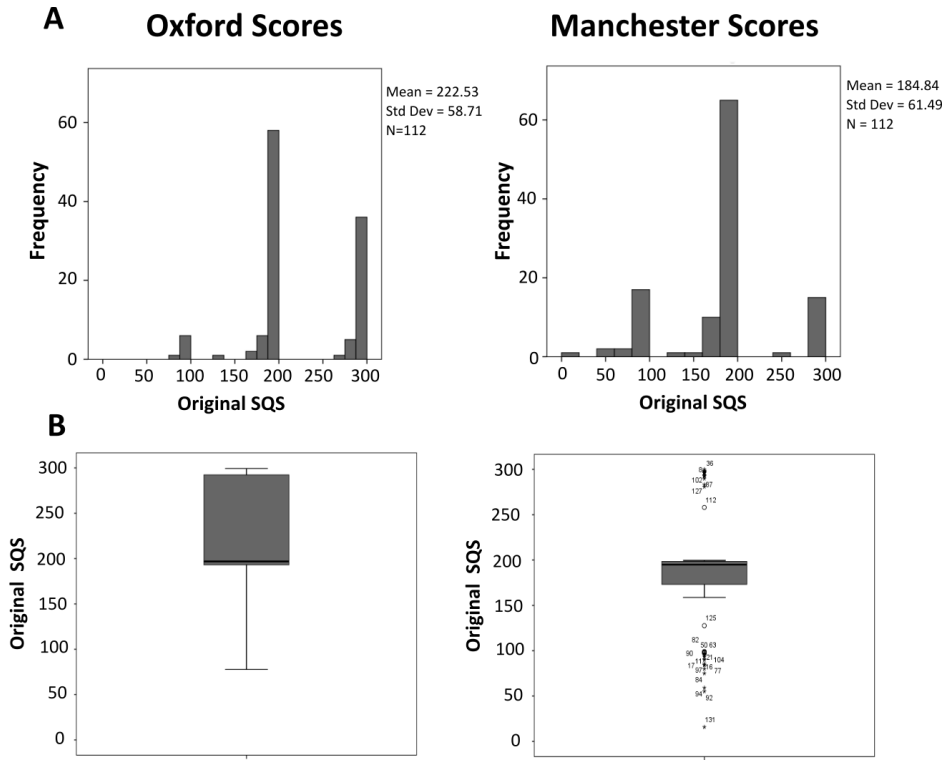


Figure 3.2: The distribution of SQS for MRE11 IHC on the BCON whole section tissue cohort. (A) Histograms for MRE11 SQS produced in Oxford and Manchester. (B) Box plots for MRE11 SQS produced in Oxford and Manchester. The median score is represented in bold. The whiskers of the plot define the maximum-minimum range of scores. The box displays the interquartile range of scores and the bold line shows the median value of the scores. The interquartile range for Oxford scores is skewed toward high values. By comparison Manchester scores have a limited interquartile range.

Table 3.1: Descriptive statistics for the scoring of MRE11 IHC on the BCON whole section tissue cohort. The median, minimum-maximum range and interquartile ranges for semi-quantitative scoring distributions and % positivity are given for the scoring performed in Oxford and Manchester

	Oxford Scores	Manchester Scores
SQS		
-Median	-197	-195
-Min-max range	-78-300	-16-300
-Interquartile range	-193-292	-173-199
% positivity		
-Median	-98	-98
-Min-max range	-68-100	-55-100
-Interquartile range	-95-99	-92-99

The median SQS for Oxford and Manchester were 197 and 195, respectively. However, compared to Manchester, Oxford scores were skewed toward high values (≥200) (Figure 3.2, Table 3.1). In the histograms of MRE11 SQS three discrete accumulations of scores are present in in both the Oxford and Manchester datasets, which are located at 75-100, 175-200 and 275-300 (Figure 3.2A).

The clustering of scores around the 100, 200 and 300 points could be the result of a limited range in the percentage of cells positively stained for MRE11 (% positivity). Indeed, the interquartile ranges for Oxford and Manchester scoring of % positivity demonstrate that the majority of the scores (ie those that lie between the first and third quartiles of the data) were between 95-99% and 92-99%, respectively (Table 3.1). Additionally, the minimum-maximum range of the % positivity scores show that none of the samples were negative for MRE11 and the lowest scores for % positivity in Oxford and Manchester were high at 68% and 55%, respectively (Table 3.1).

A comparison was made between the SQS for this study into MRE11 and the SQS values reported in the first study of MRE11 as a predictive biomarker in MIBC, Choudhury *et al* (2010). In Choudhury *et al* (2010) the values for the median and minimum-maximum range of SQS were reported for one of the cohorts (median=182, minimum-maximum range=28-296). Both Oxford and Manchester had a higher median SQS for MRE11 staining than that reported by Choudhury *et al* (2010) (Oxford median SQS = 197, Manchester median SQS = 195). Additionally, Oxford scores also had a higher valued minimum-maximum range of 78-300 compared to Choudhury *et al* (2010). The differences between the scoring distributions presented in Choudhury *et al* (2010) and this present study of MRE11 may be the result of real differences between the cohorts used. However, it is also possible that the differences in scoring may be a result of insufficient optimisation of the automated MRE11 IHC assay.

3.2.2 Re-optimisation of automated MRE11 IHC

The analysis of MRE11 scoring distributions confirmed the likelihood that the automated MRE11 IHC protocol was not optimal. A thorough process of protocol re-optimisation was therefore undertaken to investigate whether IHC staining could be improved (Table 3.2). During protocol re-optimisation, a pathologist was consulted and gave input on which conditions amounted to an improvement in staining. Every step of the IHC procedure was tested during re-optimisation. However, only three conditions were found to improve the quality of staining.

An improvement in the number of negatively stained cells, background staining and the intensity of staining was seen when a 10% BSA pre-primary antibody protein blocking step of 30 minutes was added to the automated MRE11 IHC protocol. While this step is

routinely carried out using manual methods it is frequently omitted in automated procedures using the Bondmax autostainer. This is because the manufacturer's sample washing buffer is, in most cases, sufficient to prevent non-specific protein binding. As such, the addition of a protein block is frequently omitted from automated IHC procedures.

In addition to the implementation of a protein blocking step, further improvement in the range of staining intensity was achieved by increasing the primary antibody dilution from 1:3,000 in 1% BSA to 1:6,000 in 10% BSA and reducing the primary antibody incubation time from 15 minutes to 8 minutes.

Table 3.2: **The re-optimisation of the automated MRE11 IHC protocol**. Conditions highlighted in green indicate the alterations which were chosen for the re-optimised MRE11 IHC protocol.

IHC step	Original condition	Alteration	Impact	Decision
Epitope retrieval time	20 minutes	• 10 minutes • 5 minutes • No retrieval	• Very light/ no staining • Very light/ no staining • Very light/ no staining	Keep original condition
Epitope retrieval temperature	100°C	• 97°C • 95°C • 90°C	• Very light/ no staining • Very light/ no staining • Very light/ no staining	Keep original condition
Pre-primary protein block step	No protein block step	• **10% BSA incubation for 30 minutes** • 3% BSA +20% goat serum incubation for 30 minutes	• Very good improvement on original staining • Improvement on original staining, not as good as 10% BSA for 30 minutes	Keep this condition Reject condition
Pre-primary protein block step –time	N/A	• 10% BSA incubation for 60 minutes	• No improvement on 30 minute pre-primary incubation	Reject condition
Primary antibody incubation time	15 minutes	• **8 minutes** • 5 minutes	• Improvement • Improvement on original condition but a bit light on some samples 8 minutes preferred	Keep this condition Reject condition
Antibody Dilution	1:3000	**1:6000**	• Improvement	Keep this condition
Reduce polymer incubation period	8 minutes	• 4 minutes	• No notable difference	Reject condition
Reduce post polymer incubation period	8 minutes	• 4 minutes	• No notable difference	Reject condition
Reduction of DAB incubation period	10 minutes	• 6 minutes • 2 minutes	• No notable difference • No notable difference	Reject condition Reject condition

3.2.3 Comparison of re-optimised and original MRE11 staining

The original MRE11 staining conditions and re-optimised MRE11 staining conditions were directly compared on consecutive cuts of a control TMA consisting of bladder cancer cell pellets and bladder cancer tissue cores (Figure 3.3).

Compared to the original staining conditions, where >90% of cells in all cores were positively stained, in re-optimised MRE11 staining there is an increase in the range of % positivity (0-97%) over the cores. The intensity of staining with the original protocol was strong (100% of cores scored as either 2 or 3+ intensities). However, after re-optimisation there was an increase in the range of staining intensities (0-3+ intensities present). Importantly, where it was difficult to distinguish between strong, medium and weak intensities with the original MRE11 staining, with the re-optimised MRE11 staining, strong, medium and weak intensities were reportedly easier for the scorer to differentiate. Non-specific background staining was also reduced in re-optimised conditions (Figure 3.3).

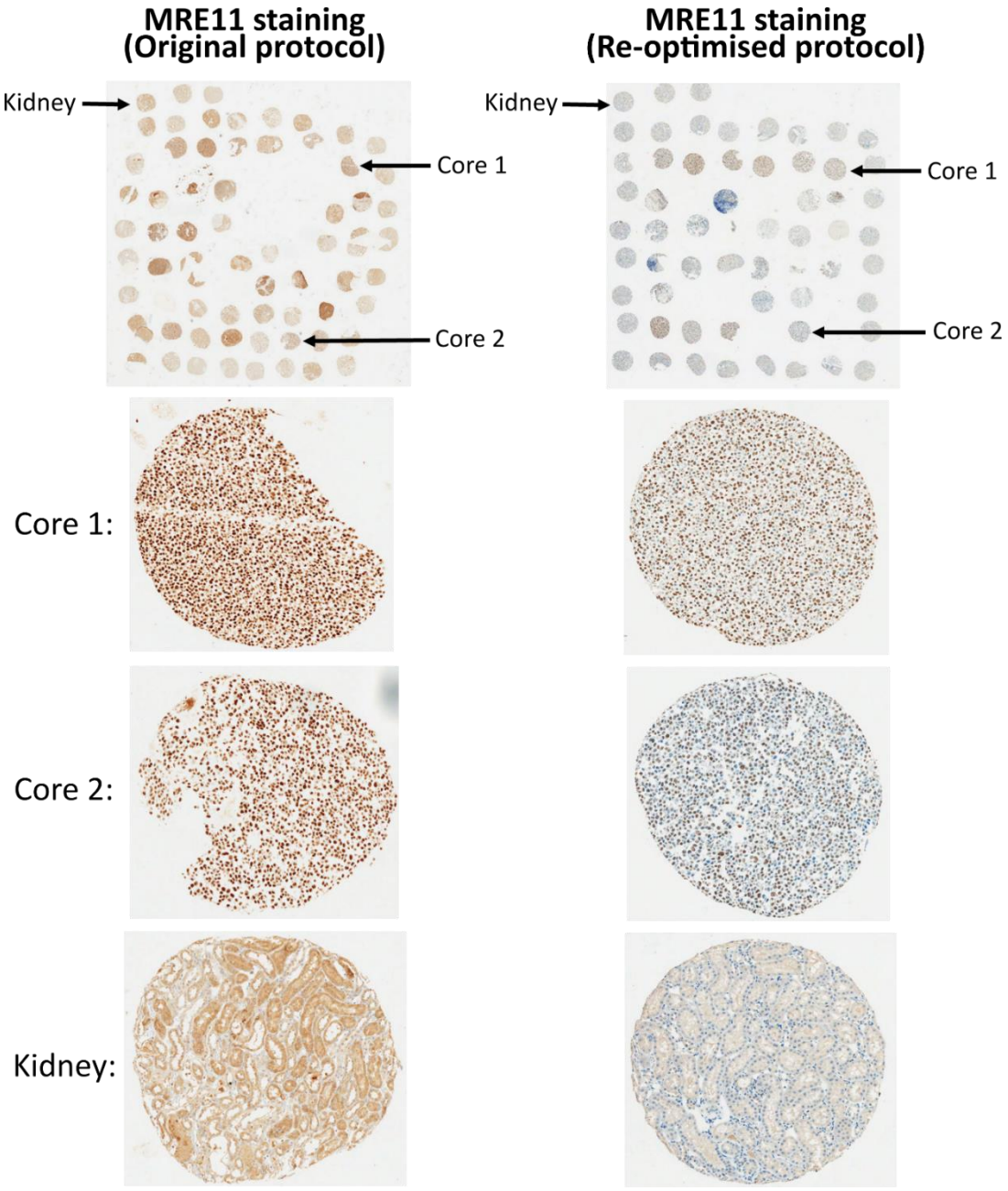


Figure 3.3: **A direct comparison between the original MRE11 and re-optimised MRE11 IHC staining on consecutive cuts from the same control TMA**. A low magnification image of the whole TMA is shown with three high magnification images of 2 bladder cancer cell pellets and one kidney core. The original MRE11 staining had non-specific background staining across all cores including the kidney samples (which are used as an indicator for nonspecific staining). Saturation in staining intensities was present, as 100% of cores were scored as either 2 or 3+ intensities. Additionally, >90% of all cells across all of the cores were positively stained. The re-optimised MRE11 staining had a reduced level of non-specific background staining in all bladder cores and kidney samples, a range of 0-97% for % positivity and no saturation in staining intensity (0-3+ intensities present).

TMAs containing 1 mm tissue cores, extracted from tissue samples from the original BCON cohort, were stained using the re-optimised MRE11 IHC protocol. TMAs contained 241 usable cores from 118 patients. Two patients were omitted from analysis due to lack of patient data available, making the total number of patients within the BCON TMA cohort 116. Patient demographics for the cohort are described in (Table 3.3).

Table 3.3: **Patient demographics and tumour characteristics of the BCON TMA cohort**. The age, sex and tumour stages of the patients in the TMA cohort are summarised, along with the tumour-stroma ratio, growth margins, growth patterns and presence of necrosis in tissue samples.

Median age (range) (n=116)	75.65 (51.5 to 90.5)
Sex (n=116)	Male- 101 (87.07%) Female- 15 (12.93%)
T stage (n=116)	T1- 5 (4.31%) T2- 83 (71.55%) T3- 23 (19.83%) T4a- 5 (4.31%)
Tumour-stroma ratio (n=113)	High- 109 (96.46%) Low- 4 (3.54%)
Growth margins (n=113)	Broad- 4 (3.54%) Infiltrative- 109 (96.46%)
Growth patterns (n=113)	Both- 47 (41.59%) Papillary- 10 (8.85%) Solid- 56 (49.56%)
Necrosis (n=113)	No- 47 (41.59%) Yes- 66 (58.41%)

Stained TMAs were scored using the standard scoring method in Oxford and Manchester. This scoring method was developed with a pathologist. It consisted of a set of 40 scored training images that encompassed the range of staining intensities present in the cohort. Before scoring, scorers would review the training images and would refer back to the images on cases that weren't clear. The pathologist was

available to provide an additional opinion for cases where there was uncertainty, such as where tissue morphology appeared to distort intensity.

The distribution of SQS resulting from the re-optimised MRE11 staining of BCON TMAs was assessed using histograms and boxplots (Figure 3.4) and a table was generated to compare the median values and range of scores before and after re-optimisation (Table 3.4).

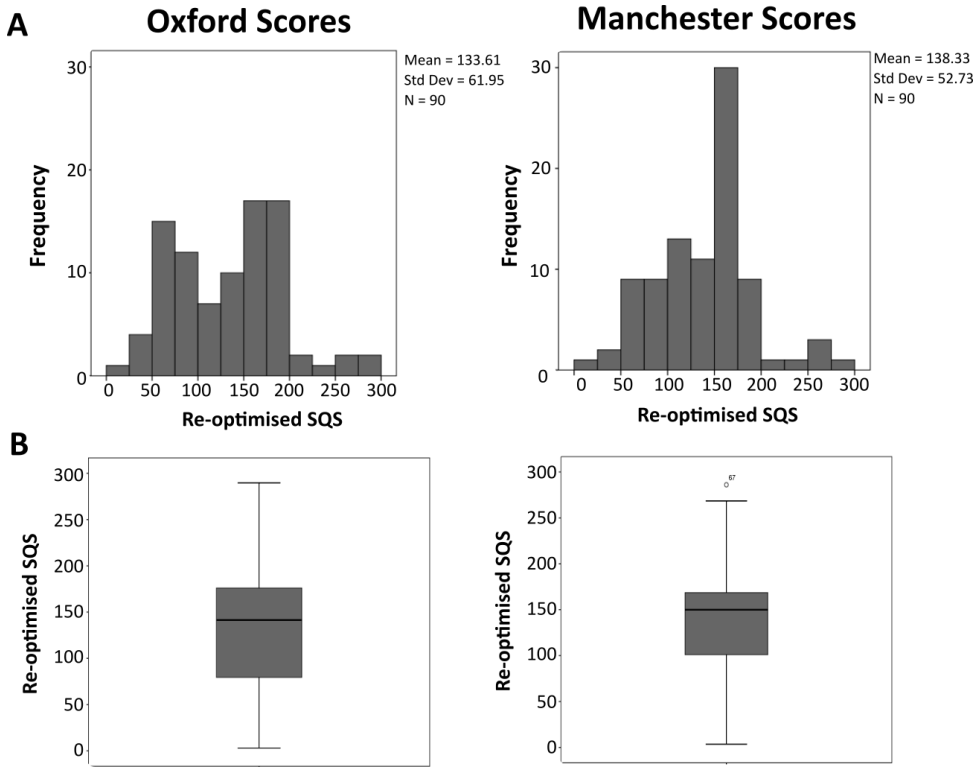


Figure 3.4: **The distribution of SQS for re-optimised MRE11 IHC on the BCON TMA cohort.** (A) Histograms for MRE11 SQS produced in Oxford and Manchester (B) Box plots for MRE11 SQS produced in Oxford and Manchester. The median score is represented in bold. The whiskers of the plot define the maximum-minimum range of scores. The box displays the interquartile range of scores and the bold line shows the median value of the scores.

Table 3.4: Descriptive statistics for the scoring of re-optimised MRE11 IHC, on the BCON TMA cohort, compared to the original MRE11 IHC, on the BCON whole section tissue cohort. The median, minimum-maximum range and interquartile ranges for MRE11 SQS and % positively.

	Oxford scoring		**Manchester scoring**	
	Re-optimised	Original	Re-optimised	Original
SQS				
-Median	-141	-197	-150	-195
-Min-max range	-3-290	-78-300	-4-286	-16-300
-Interquartile range	-79-179	-193-292	-93-169	-173-199
% positivity				
-Min-max range	-3-98	-68-100	-4-100	-55-100
-Interquartile range	-61-87	-95-99	-58-85	-92-99

After re-optimisation both Oxford and Manchester SQS distributions displayed a more normal distribution (Figure 3.4) compared to the original staining of BCON whole tissue sections (Figure 3.2). In line with this, the clustering of scores in the 50-100, 150-200 and 250-300 ranges observed in the original staining was lost after re-optimisation. Furthermore, this lack of clustering corresponded to an increased range for % positivity scores after re-optimisation (Table 3.4). Additionally, the skewing of Oxford SQS toward higher values was removed in re-optimised scoring (Figure 3.4).

A comparison was made between the SQS of re-optimised MRE11 staining to Choudhury *et al* (2010). The median values for both Oxford (141) and Manchester (149) re-optimised SQS were lower than that reported in Choudhury *et al* (2010) (182) as was the minimum-maximum range (Oxford: 3-290, Manchester: 4-286, Choudhury: 28-296). For the percentage of cancer cells stained the minimum-maximum range for Oxford (3-98) and Manchester (4-100) re-optimised scores were more comparable with Choudhury *et al* (2010) (17-99) than the original scores.

The patient scores for the original MRE11 staining of whole section BCON tissue were matched to the re-optimised MRE11 scoring of BCON tissue cores and assessed for variation. It was possible to match 90 patient tissue samples. Significant variation in MRE11 scoring was identified between the BCON original MRE11 SQS on whole tissue sections and re-optimised MRE11 SQS on TMAs, for both Oxford (WSR p≤ 0.001) and Manchester (WSR p≤0.001) scoring.

3.2.4 Reliability of Scoring Re-optimised MRE11 staining

Once we had established that the re-optimisation of MRE11 staining had altered MRE11 BCON patient scores, we addressed how reliably the tissue could be scored. Analysis was conducted on the level of agreement in % positive and intensity scores between two independent scorers within the same centre, ie. Oxford or Manchester, and the level of agreement in consensus SQS scores generated between two different centres, ie. Oxford vs Manchester.

In Oxford, scorers achieved a very high level of reliability in the scoring of the % positive scoring (Intraclass correlation coefficient (ICC): 0.95, 95% CI: 0.94-0.96) and a moderate level of agreement in intensity scoring (Kappa: 0.59). Manchester scorers also had very high levels of agreement in the scoring of % positivity (ICC: 0.99, 95% CI: 98.96-99.38). However, unlike Oxford scorers, Manchester scorers also achieved a very high level of agreement in intensity scoring (Kappa: 0.95). Comparing consensus SQS scores between Oxford and Manchester, the variation between scoring distributions were not significantly different from one another (WSR: p= 0.254) and a high level of agreement in scores was observed (ICC: 0.90, 95% CI: 0.858-0.931).

3.2.5 Association between re-optimised MRE11 staining of BCON TMAs and clinical outcome

The MRE11 SQS for Oxford and Manchester scoring of BCON TMAs was assessed for association with clinical outcome. Univariate analysis of MRE11 SQS (stratified by ≤25th, 26-50th, 51-75th and >75th percentiles) identified no significant association between MRE11 level and 3 year CSS, for either Oxford (log rank p=0.40) or Manchester (log rank p=0.33) scoring (Figure 3.5A and C). When MRE11 scores were grouped according to high (>25th percentile) or low (≤25th percentile) MRE11 level there was also no association between MRE11 and CSS. For Oxford scores the 3 year CSS for high MRE11 and low MRE11 was 60% and 52%, respectively (hr: 0.68, 95% CI: 0.37-1.26, log rank p= 0.33). For Manchester scores the 3 year CSS for high MRE11 and low MRE11 were 55% and 65%, respectively (hr: 1.44, 95% CI: 0.79-2.37, log rank p= 0.27) (Figure 3.5B and D).

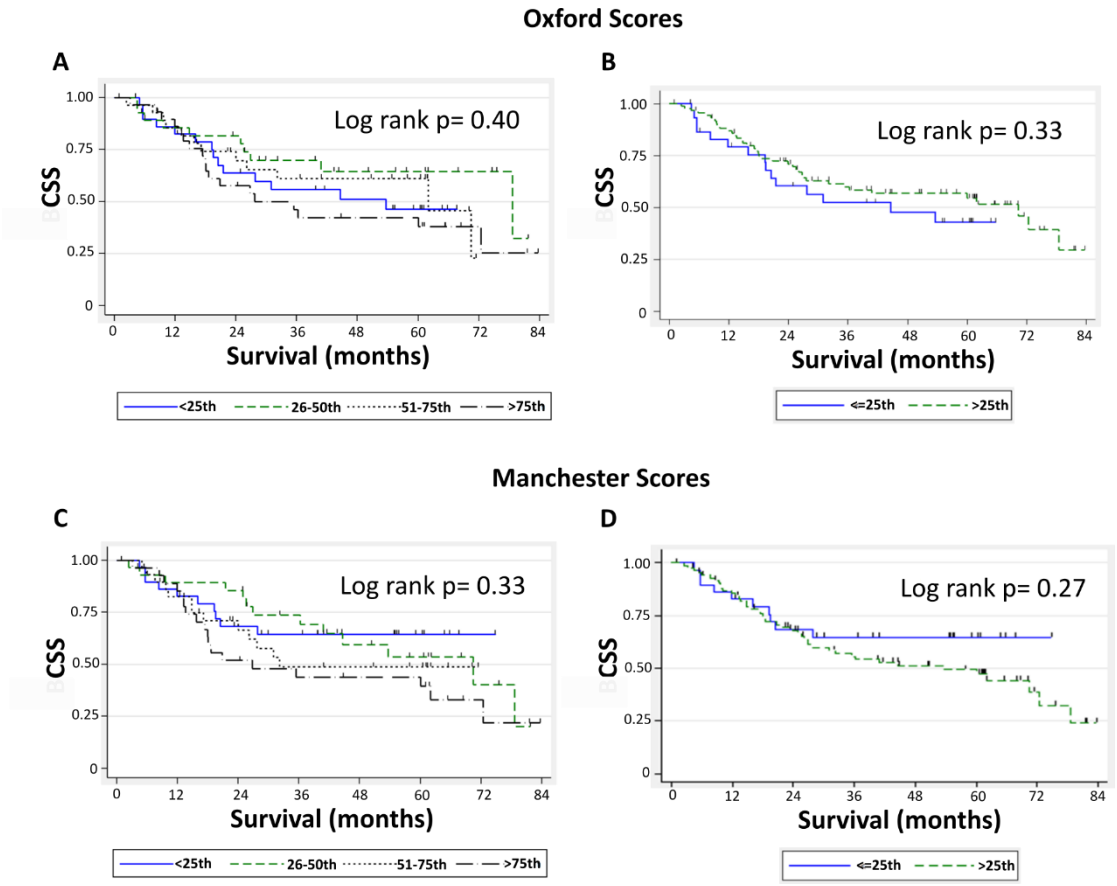


Figure 3.5: The association between re-optimised MRE11 IHC and CSS for BCON TMA cohort (n= 116). (A) Kaplan-Meier CSS survival curve for the Oxford scoring of MRE11, stratified according to the <25th, 26-50th, 51-75th and 75th percentiles. No significant association was identified between MRE11 SQS and 3 year CSS (log rank p= 0.40); (B) Kaplan-Meier survival curves for Oxford scoring of MRE11, stratified bimodally according to high MRE11 SQS (>25th percentile) or low MRE11 SQS (≤25th percentile). No significant association was identified between MRE11 SQS and 3 year CSS using bimodal grouping of MRE11 SQS (log rank p= 0.33); (C) Kaplan-Meier CSS survival curve for the Manchester scoring of MRE11, stratified according to the <25th, 26-50th, 51-75th and 75th percentiles. No significant association was identified between MRE11 SQS and 3 year CSS (log rank p= 0.33); (D) Kaplan-Meier survival curves for Manchester scoring of MRE11, stratified bimoodally for high MRE11 SQS (>25th percentile) or low MRE11 SQS (≤25th percentile). No significant association was identified between MRE11 SQS and 3 year CSS using bimodal grouping of MRE11 SQS (log rank p= 0.27). Statistics performed by Dr Victoria Strauss.

Multivariate analysis of Oxford MRE11 SQS was conducted using treatment, stage, grade, completion of TURBT, haemoglobin concentration and RT fractions as covariates. None of the covariates analysed was a significant predictor for CSS before and after controlling for other variables (Table 3.5). No significant association between Oxford MRE11 SQS and 3 year CSS was identified in multivariate analysis, either grouping scores by percentiles or by high and low MRE11 level.

Table 3.5 Results of multivariate analysis for re-optimised MRE11 IHC on the BCON TMA tissue cohort (n= 116). The hazard ratio for CSS is displayed with 95% confidence intervals (CI) for each of the ≤25th, 26-50th, 51-75th and >75th percentiles of MRE11 SQS, treatment, stage, grade, completion of TURBT, haemoglobin concentration and RT fractions. Multivariate analysis for Oxford and Manchester SQS are displayed separately. Statistics performed by Dr Victoria Strauss.

	Oxford scoring Adjusted Hazard Ratio (95% CI, p value significance)	**Manchester scoring** Adjusted Hazard Ratio (95% CI, p value significance)
MRE11 percentiles-		
<=25th	1	1
26-50th	0.58 (0.24-1.41, p>0.05)	1.17 (0.48-2.85, p>0.05)
51-75th	0.82 (0.35-1.91, p>0.05)	1.81 (0.72-4.55, p>0.05)
>75th	1.33 (0.59-3.02, p>0.05)	2.50 (1.03-6.11, p>0.05)
RT	1.32 (0.74-2.34, p>0.05)	1.32 (0.75-2.34, p>0.05)
Stage-III/IV	0.66 (0.31-1.41, p>0.05)	0.62 (0.29-1.33, p>0.05)
Grade-3/4	0.78 (0.35-1.76, p>0.05)	0.79 (0.35-1.78, p>0.05)
Completion of TURBT	1.01 (0.56-1.81, p>0.05)	1.06 (0.59-1.91, p>0.05)
Hb >=13.5 g/dl	0.67 (0.36-1.22, p>0.05)	0.66 (0.36-1.20, p>0.05)
High RT dose (64 Gy in 32 fractions)	0.72 (0.38-1.36, p>0.05)	0.69 (0.36-1.30, p>0.05)

The BCON trial was aimed at investigating the effect of the addition of carbogen and nicotiamide (CON) to RT treatment. It is possible that the impact of CON on tumour hypoxia could affect the association of MRE11 with clinical outcome. We therefore conducted an analysis on the association of MRE11 level with clinical outcome in the subgroup of patients (n=62) within the BCON TMA cohort who received RT alone.

No significant association between MRE11 SQS (stratified by ≤25th, 26-50th, 51-75th and >75th percentiles) and 3 year CSS was identified in the RT alone patient population using log rank testing (Oxford log rank p=0.19, Manchester log rank p=0.64) (Figure 3.6A and C). Using Cox proportional hazards modelling a significant association between 3 year CSS and SQS was identified for Oxford scores between the ≤25th percentile (CSS: 41%, hr: 1.00) and the 26-50th percentile (CSS: 71%, hr: 0.31, 95% CI: 0.10-0.98, p≤ 0.05). However, no significant association was found for the 51-75th (hr: 0.64, 95% CI: 0.23-1.78, p>0.05) and >75th percentiles (hr: 0.83, 95% CI: 0.32-2.17, p>0.05) compared to the ≤25th percentile.

Dividing the Oxford MRE11 SQS from the RT subgroup, into high (>25th percentile) and low (≤25th percentile) MRE11 level classifications, a small non-significant difference in 3 year CSS was observed (>25th percentile CSS: 41%, ≤25th percentile CSS: 54%, hr: 0.56 (95% CI: 0.25-1.28), log rank p=0.17) (Figure 3.6B) but this was not found with Manchester scores (>25th percentile CSS: 57%, ≤25th percentile CSS: 48%, hr: 1.21 (95% CI:0.49-2.98), log rank p=0.68) (Figure 3.6D).

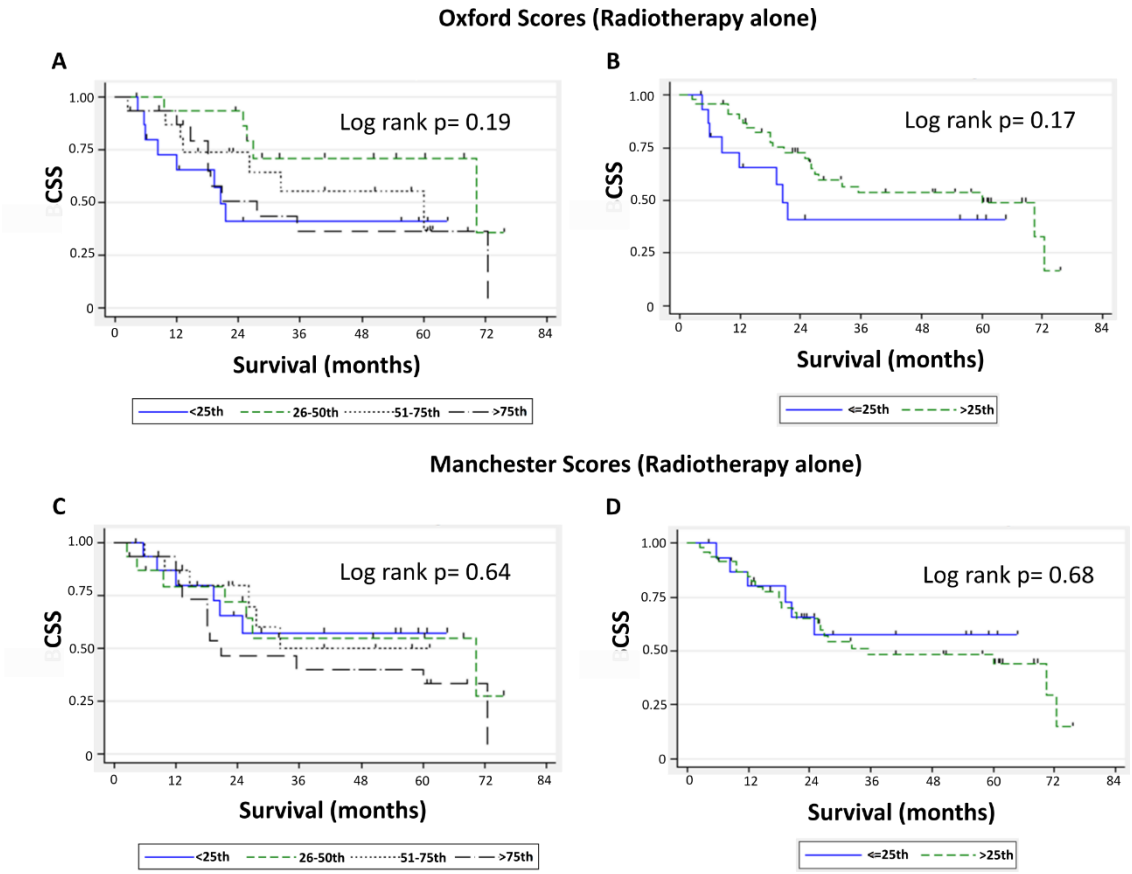


Figure 3.6: The association between re-optimised MRE11 IHC and CSS for BCON TMA patients treated with radiotherapy alone (n= 62). (A) Kaplan-Meier CSS survival curve for the Oxford scoring of MRE11, stratified according to the <25th, 26-50th, 51-75th and 75th percentiles No significant association was identified between MRE11 SQS and 3 year CSS (log rank p= 0.19); (B) Kaplan-Meier survival curves for Oxford scoring of MRE11, stratified bimodally according to high MRE11 SQS (>25th percentile) or low MRE11 SQS (≤25th percentile). No significant association was identified between MRE11 SQS and 3 year CSS using bimodal grouping of MRE11 SQS (log rank p= 0.17); (C) Kaplan-Meier CSS survival curve for the Manchester scoring of MRE11, stratified according to the <25th, 26-50th, 51-75th and 75th percentiles. No significant association was identified between MRE11 SQS and 3 year CSS (log rank p= 0.64); (D) Kaplan-Meier survival curves for Manchester scoring of MRE11, stratified bimoodally for high MRE11 SQS (>25th percentile) or low MRE11 SQS (≤25th percentile). No significant association was identified between MRE11 SQS and 3 year CSS using bimodal grouping of MRE11 SQS (log rank p= 0.68). Statistics performed by Dr Victoria Strauss.

3.3 Discussion

Investigations using IHC on clinical samples are useful for the study of potential biomarkers in cancer. However, as highlighted in this work, it is important to ensure the rigorous validation of the IHC staining and that the protocol used is optimal. For results from IHC investigations to be clinically relevant, both staining and scoring need to be consistent and reliable (O'Hurley *et al*, 2014). This work had three aims: to look at the effect of re-optimising MRE11 IHC staining on a retrospective cohort of patient tissue, to assess the reliability of re-optimised IHC scoring between two centres and to define the relationship between MRE11 and clinical outcome following RT.

The need for the re-optimisation of MRE11 IHC staining arose due to the transfer of the IHC procedure to an automated platform. During retrospective analysis of the original automated MRE11 tissue staining and scoring, the intensity of the staining was found to be saturated and percentage of cells in an image positive for MRE11 was consistently >90%. Additionally, high levels of non-specific background staining were present.

The move toward automated IHC has a number of benefits. Automated IHC can be standardised and conducted on many more samples in the same instance, improving workflow and fidelity of staining (Howat *et al*, 2014). However, it is important to note that in transferring an IHC stain to an automated platform, the changes in apparatus used can alter staining. For example, for heat induced epitope retrieval (HIER), the heating of samples using automated systems occurs via heat blocks located directly under samples. By comparison in the majority of manual IHC protocols, for HIER, tissue samples are contained in a heated buffer. The difference in heating can alter the degree of epitope retrieval and increase background staining with some antibodies.

While a major advantage of automated IHC is that it requires less hands-on input from a researcher, it is also a source of concern. With the majority of IHC steps being conducted by a machine, it is possible for a knowledge gap to develop (Prichard, 2014). In this scenario researchers may not be sufficiently familiar with the steps in an automated protocol and their relative importance. This can lead to poor troubleshooting and suboptimal staining. In the re-optimisation of automated IHC with MRE11 a pre-primary antibody protein blocking step was added which greatly reduced background staining. This may seem an obvious step and might even be assumed to be included in the basic automated protocol, especially to those familiar with manual immunostaining. However, using the Bondmax automated IHC system it is not standard. The efficiency of the manufacturer's sample washing buffer for this system is often sufficient to remove non-specific protein binding, negating the need for pre-primary antibody protein blocking. In a search of published studies utilising the Bondmax autostainer, where the protocol steps were outlined, a pre-primary antibody protein blocking step was absent in all procedures. Our findings with MRE11, however, clearly demonstrate that the incorporation of pre-primary antibody protein blocking is useful and required, at least for some stains, using automated platforms. For clinical application this finding is highly significant. If IHC protocols are not optimal, in this setting, it could lead to patients not receiving their 'optimal' therapy.

Re-optimisation of automated MRE11 IHC removed the saturation of staining intensity, decreased background staining and increased the range of scores for the percentage of cells stained positively for MRE11.

Re-staining of the BCON tissue samples, with the re-optimised MRE11 protocol, was conducted on TMAs rather than whole tissue sections. This is a limitation when directly comparing the results of re-optimised and original MRE11 staining as it is not possible to discount that changes in scoring could be due to the use of tissue cores rather than whole tissue sections. That being said, numerous studies have identified concordance between the IHC scoring of TMAs, containing cores ≥0.6mm, and scores derived from whole tissue sections, especially when multiple cores are taken for the same patient (Camp *et al*, 2008; Voduc *et al*, 2008). BCON TMAs used cores of 1 mm and contained 2-3 cores per patient, it is therefore reasonable to assume that the scores derived from the TMA cores in this study should be concordant with those derived from BCON whole tissue sections. Using this assumption it was possible to assess the impact of re-optimised MRE11 staining on the scoring of BCON tissue.

Scoring of re-optimised MRE11 staining between centres, as assessed by WSR and ICC, was found to be reliable. No statistically significant variance in the distribution of SQS between centres was identified and a high level of agreement was found. Interestingly, the concordance of SQS scoring between centres was improved in the re-optimised staining compared to originally stained BCON tissue, with original scores showing a significant variance in scores between centres.

In the analysis of independent scorer agreement in each centre for staining intensity, Oxford scorers displayed less consensus then Manchester scorers. While both Oxford and Manchester scores were generated by combining the scores of at least two independent scorers, to form a consensus score, there were two main differences in the manner of scoring between the two centres. In Oxford, independent scoring was

conducted in isolation on separate computers. Scores were then combined and a third scorer asked to score cases where there was a difference in scores. This enabled an unbiased consensus to be reached. However, in Manchester, the two 'independent' scorers scored data in the same room, at the same time, from the same computer and consensus agreement reached before a third person was involved. It is dubious how independent such scoring is. It is possible that when scoring and agreement is conducted concurrently without the addition of a third consensus scorer that there is some loss of objectivity. In this scenario, as scoring progresses scorers inadvertently influence and learn from each other. This would result in a convergence of scores and a loss of impartiality. As such, the higher level of agreement between scorers in Manchester compared to Oxford might be an artefact of the scoring practice. However, how this potential loss of impartiality in scoring might affect the reliability of consensus scores in terms of association with outcome is debatable.

Another factor that could potentially affect scoring using digital IHC images, is the device they are viewed on and the angle of viewing. Changes in either of these factors can affect the appearance of staining intensity and hence influence scoring. It is possible that part of the disparity in intensity between Oxford scorers could be due to the use of different viewing devices and screen angles. While this fact is not likely to radically affect scoring, it is worthy of consideration. Where possible it is highly advisable for scorers to be familiar with the device used for scoring digital images. Upon using a new system it is useful to familiarise the eye with a small series of standardised, pre-scored training images, as was done here. As much of the IHC analysis in clinic is still conducted using light microscopes rather than digital images, the effect of alterations in colour aspect

ratios from screens is not likely to impact the reliability of the technique in this environment.

The use of automated digital image analysis could offer an efficient method for the removal of the potential human bias present in scoring of MRE11 IHC images. Digital image analysis is not subject to ocular phenomena that can trick the human eye. It can also be standardised, contain algorithms to correct for batch effects and through the use of colour devolution provide quantitative rather than semi-quantitative results (Pham et al, 2007; Gurcan et al, 2009, Hoffman et al, 2014; Kothari et al, 2014; Kumar et al, 2015, Aeffner et al, 2017). Computer-based scoring does not require consensus scoring from multiple observers and, therefore, could increase the speed of IHC analysis. However, despite the potential of automated image analysis in IHC, currently basic algorithms are still insufficient for the task of scoring without manual intervention and although the results of deep learning algorithms are encouraging, the technology is still in its infancy and not widespread.

It was hypothesised that re-optimisation of MRE11 IHC on BCON tissue would improve scoring of the tissue and might reveal an association between MRE11 levels and clinical outcome following RT. Although there is evidence for the re-optimisation of MRE11 staining altering the scoring of the BCON dataset, compared to original staining, there was still a lack of convincing evidence for an association between MRE11 level and CSS, for both Oxford and Manchester scores. The exclusion of patients who received CON from the analysis did reveal a small but non-significant association in Oxford scores for CSS between high and low MRE11 level groupings. However, this putative association was not identified in Manchester scores. Interestingly, a significant difference in 3 year

CSS was identified for Oxford scoring of the RT alone subset for MRE11 scores between the ≤25th percentile and the 26-50th percentile but not the 51-75th and >75th percentiles. Given the small sample size, however, it is likely that the significance observed here is due to chance.

The reason for the lack of association of MRE11 to CSS in this study could be due to a number of reasons. The study was powered to detect a difference between two MRE11 groups of 43% and 70%. While our data did identify small variations between MRE11 groups, none of them reached this level. Therefore, lack of significance found in the analysis might be due to the study being underpowered (n=62). If more patient samples were available to be included in the study, it is possible that this issue could be resolved.

The use of TMAs over whole tissue sections might also be a contributing factor. As previously stated, multiple studies have found that TMAs can be representative of whole tissue sections. Nevertheless, our experience with MIBC TMAs leads us to question their usefulness in bladder cancer. Transurethral resection of bladder cancer tumours is conducted using a hot diathermy wire. The heat from the wire results in a diathermy artefact in areas of tissue proximal to the wire. The diathermy artefact reflects areas of tissue that have been damaged and are not viable for IHC analysis (Figure 3.7). As sections are cut from TMAs, different depths of the TMA cores are exposed and, with increasing depth, the number of cores with diathermy artefacts increases. This results in a substantial loss of cores from TMA cohorts or the scoring of a significantly reduced, non-representative, number of cells. Additionally, there is also a substantial loss tumour tissue (Figure 3.7A). Due to this it is reasonable to assume that results from MIBC TMAs are not representative of whole tissue sections unless there is

a significant amount of redundancy in tissue cores, ie. ≥ 3 cores per patient. This conclusion is important as it would also lead us to re-evaluate the assumption, made above, about the ability to compare original and re-optimised MRE11 scores through the comparison of stained BCON whole tissue sections and BCON TMAs.

The analysis of the high and low MRE11 groupings for Oxford scores in the RT alone patient subset suggests a small, non-significant association of MRE11 to CSS. The lack of significance in this result is most likely due to the small number of cases in the patient subset and an issue of underpower. Interestingly, the difference between survival analysis on the whole BCON cohort and the RT alone cohort might point to CON treatment being a confounding factor in the predictive capacity of MRE11 in MIBC. It is not clear why this may be. However, as CON is used to try and remove hypoxic areas within tumours, it is possible that the increase in the relative biological effect of radiation, provided by increased in oxygen perfusion of the tumour, is playing a role in this result.

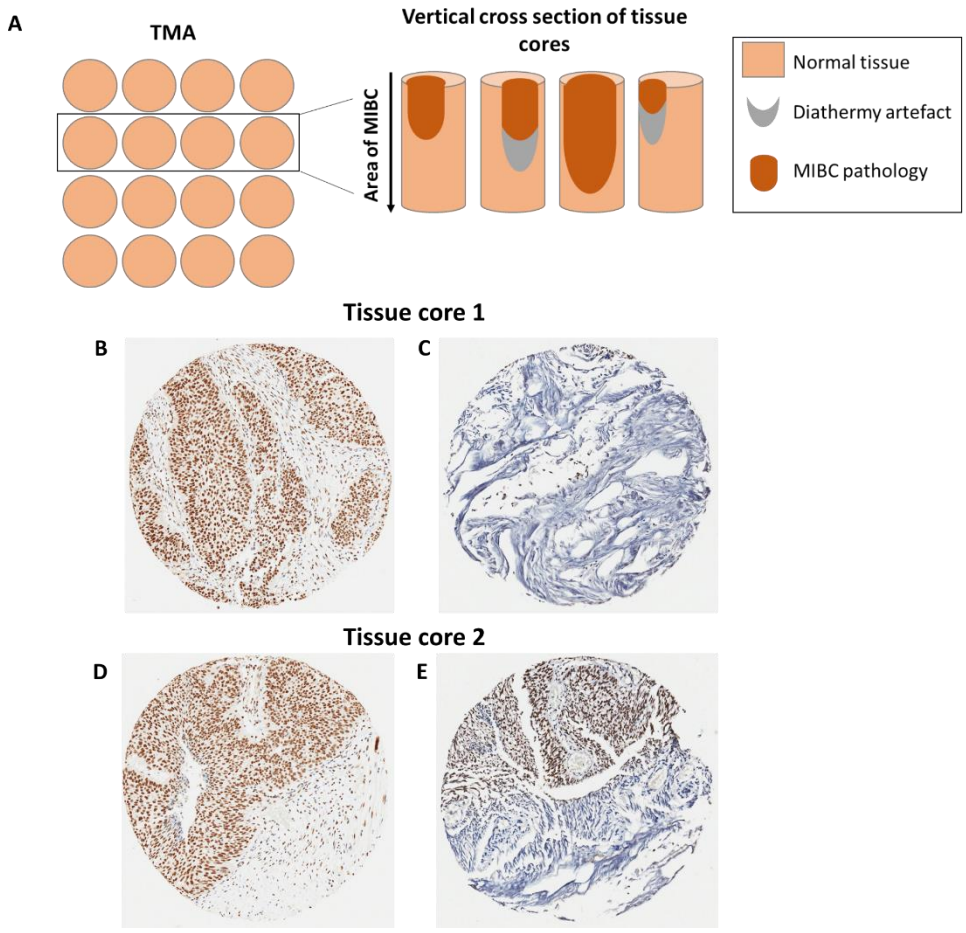


Figure 3.7: The loss of pathology and presence of diathermy in MIBC TMAs increases as the depths of cores are explored. (A) Schematic representation of how the loss of tumour tissue or diathermy in a MIBC TMA can occur as a TMA is sectioned; (B-E) Representative images of TMA cores stained with MRE11. Two sections of the same core but cut from different tissue depths are placed side by side. Sections from more depleted TMA blocks are shown in images (C) and (E); (C) Severe diathermy abolishes the MRE11 staining observed in (B); (E) Less severe diathermy present, which decreases across the tissue core. MRE11 staining returns as the level of diathermy becomes less severe.

This work has highlighted the importance of robust validation when conducting IHC and the need for standardisation of the processes of tissue staining and scoring. Despite the large number of potential biomarkers identified at bench level, due to poor validation, a disappointingly low number of biomarkers actually make it to the patient bedside (Forker *et al*, 2015; Goossens *et al*, 2015). The use of automated IHC platforms is an unequivocal advancement for the standardisation of IHC and will hopefully help assisting the progression of potential biomarkers to clinical application. However, use of automated systems should always be conducted with care and understanding.

Our experience with MIBC TMAs leads us to strongly question their use for studies aimed at identifying associations to clinical outcome in bladder cancer. The loss of tissue from MIBC TMAs due to diathermy and loss of tumour in the deeper core sections can decrease patient numbers in IHC studies. This is a significant limitation in using MIBC TMAs as it leads to issues of underpower and a reduction in the validity of IHC scoring.

Although the findings for the re-optimisation of MRE11 IHC on BCON tissue samples did not validate previously published data, this does not necessarily lead to complete rejection of the role of MRE11 as a biomarker in MIBC.

In accordance with Choudhury et al (2010) and Laurberg et al (2012), this study only evaluated the nuclear staining MRE11 in tissue. MRE11 is predominantly located in the cell nucleus but a proportion of MRE11 is also contained within the cytoplasm of the cell. The function of nuclear MRE11 in DNA repair, replication, telomere maintenance and meiosis is highly characterised (D'Amours & Jackson, 2002). The importance and activity of MRE11 in the cytoplasm is less well defined but has been reported to function

in the maintenance of mitochondrial DNA (Dmitrieva *et al*, 2011; Tadi *et al*, 2016) and participate in the innate immune response (Kondo *et al*, 2013).

Whether or not the function of MRE11 in the cytoplasm plays a role in radiosensitivity of cells or tumours, the translocation of MRE11 from the cytoplasm to the nucleus of a cell is important for MRE11 to function in nuclear DBS repair following radiation. An accumulation of MRE11 in the cytoplasm may indicate a problem in the nuclear translocation of the protein, which would affect radiosensitisation. In support of this, the increased sensitisation of mammalian cells to radiation following hyperthermia has been attributed, at least in part, to the translocation of MRE11 from the nucleus to the cytoplasm. Therefore, combining the nuclear and cytoplasmic staining of MRE11 into a nuclear to cytoplasmic ratio, may be a more informative and reliable biomarker for patient outcome following RT in MIBC, than nuclear staining alone. Indeed, Magliocco *et al* (2017) recently presented evidence on MRE11 as a biomarker in bladder cancer by quantifying MRE11 IHC using a nuclear to cytoplasmic ratio and found that low expression of MRE11 (nuclear to cytoplasmic ratio ≤1.49) associated with significantly higher cancer-specific mortality.

Chapter 4 MRE11 as a predictive biomarker in anal cancer

4.1. Introduction

In the course of investigating MRE11 as a potential predictive biomarker in MIBC the question arose as to whether MRE11 might be a possible predictive biomarker in other cancers. To begin to answer this question, we were able to obtain a number of anal cancer FFPE tumour biopsies with clinical outcome data, from Professor Duncan Gilbert, on which an analysis of MRE11 IHC staining could be conducted.

Anal cancer was chosen for this study due to the availability of the tissue samples, the standard use of radiation-based treatment for this neoplasm and the evidence for an important role for the MRN complex in the majority of anal cancers that are infected with high-risk subtypes of human papillomavirus (HPV).

The standard of care for the treatment of most anal cancer patients is combination CRT (Glynne-Jones *et al*, 2009), which achieves favourable response rates with complete tumour regression in 80-90% of cases (Vinayan & Glynne-Jones, 2016). However, acute toxicities are not uncommon and many patients suffer long term effects that significantly impact on their QOL (Bentzen *et al*, 2013; Ghosn *et al*, 2015).

Infection with high-risk HPV subtypes, predominately HPV16, are strongly associated with the development of anal cancer, with reported estimates of HPV positivity ranging from 84 to 95% (Baricevic *et al*, 2015). Low-risk HPV subtypes express E6 and E7 proteins capable of inducing benign lesions, unlikely to progress to malignancy (Ganguly & Parihar, 2009). However, expression of E6 and E7 proteins from high-risk HPV subtypes,

hrE6 and hrE7, causes increased genomic instability and drives carcinogenesis (Figure 4.1).

Expression of either HPV16 hrE6 or hrE7 proteins is sufficient to increase chromosomal instability, including centriole amplification, anaphase bridge formation and un- or misaligned chromosomal material (Duensing & Münger, 2002). The hrE6 and hrE7 viral proteins interact with a multitude of proteins to drive the host cell towards uncontrolled cellular proliferation. Arguably the most important interaction made by hrE6 is its binding to the tumour suppressor p53. hrE6 together with the ubiquitin ligase E6-associated protein (E6AP) leads to the ubiquitination of and ultimate proteasomal degradation of p53 (Thomas *et al*, 1999). This loss of p53 results in the abrogation of G1/S and G2/M checkpoints and blocking of apoptotic signalling, thus increasing chromosomal instability and enhancing the incorporation of viral DNA into the host genome (Boulet *et al*, 2007). Further interactions with hrE6 enable cellular immortalisation (Oh *et al*, 2001; Fehrmann & Laimins, 2003) and apoptotic evasion (Yuan *et al*, 2012). The hrE7 viral protein also targets the cell cycle, by binding hypophosphorylated retinoblastoma protein (pRb). hrE7 prevents pRb binding to and consequently repressing the E2F1 transcription factor. This enables the transcription of E2F1 target genes and S phase progression (Ganguly & Parihar, 2009). In addition to this, hrE7 also promotes the degradation of pRb (Boyer *et al*, 1996; Darnell *et al*, 2007). Further dysregulation of the G1/S checkpoint by hrE7 is achieved through the inhibition of the p21 cyclin dependent kinase inhibitor (Funk *et al*, 1997; Jones *et al*, 1997; Shin *et al*, 2009) and histone deacetylase inhibitors (HDACs) (Brehm *et al*, 1999; Longworth *et al*, 2005) resulting in increased CDK activity and E2F-dependent transcription, respectively. hrE6 and hrE7 are also able to induce reactive oxygen species (ROS) in cells,

increasing the level of endogenous DNA damage contributing to chromosomal instability (De Marco, 2013; Williams *et al*, 2014; Marullo *et al*, 2015).

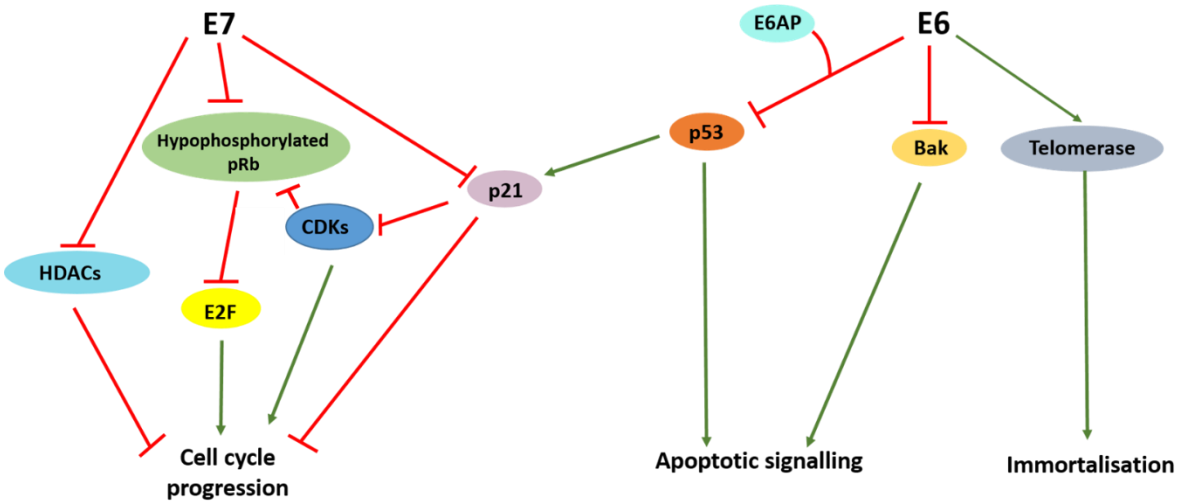


Figure 4.1: The E6 and E7 viral proteins, from high-risk HPV infection promote tumorigenesis. The E6 viral protein promotes apoptotic evasion by inhibiting p53. E6 together with the ubiquitin ligase E6AP ubiquitinate p53, resulting in its degradation. Further apoptotic evasion is achieved, by E6, through the inhibition of the apoptosis promoting protein Bak. E6 promotes immortalisation of cells by activating telomerase. The E7 viral protein promotes cell cycle progression by inhibiting the repressive binding of HDACs to the E2F2 promoter. Inhibition of p21 by E7 relieves the suppression CDKs which facilitate cell cycle progression. Activation of CDKs through p21 inhibition also results in the hyperphosphorylation of hypophosphorylated pRb which prevents pRb binding to and supressing E2F transcription factors. The E7 protein can also directly bind to hypophosphorylated pRb relieving the suppression of E2F transcription factors and can promote pRb degradation.

In anal cancer, HPV status (as measured directly or through overexpression of the surrogate marker p16) has a strong association with patient response following CRT. Numerous studies have identified that HPV-positive anal cancers show significantly improved response rates to treatment compared to their HPV-negative counterparts (Gilbert *et al*, 2013; Doll *et al*, 2014; Koerber *et al*, 2014; Serup-Hansen *et al*, 2014; Mai

et al, 2015; Meulendijks *et al*, 2015). The clear segregation in patient response highlights the need to find improved treatment strategies for HPV-negative individuals.

In order to replicate the viral genome, HPV utilises the DDR system (Anacker & Moody, 2016). ATM activation is observed throughout the viral life cycle and is recruited to nuclear viral replication foci along with other DDR proteins, including members of the MRN complex, RAD51 and BRCA1 (Moody *et al*, 2009; Sakakibara *et al*, 2011; Reinson *et al*, 2013). Indeed, the localisation of HR proteins to nuclear viral replication foci, coupled with branched structures observed in replicating viral genomes, has led to the hypothesis that HPV may replicate in a recombination-dependent manner (Anacker & Moody, 2016).

MRE11 activity is important for HPV viral replication. Unlike adenoviral infection, where viral subtypes have been found to target MRE11 for degradation (Stracker *et al*, 2002, 2005; Liu *et al*, 2005; Karen *et al*, 2009), MRE11, RAD50 and NBS1 levels have been found to be increased after expression of the HPV16 or HPV31 E7 viral protein. In addition to the increase in MRE11 protein levels with HPV infection, there is also evidence that the exonuclease function of MRE11 is required for efficient HPV viral replication. This was demonstrated by the reduction in viral replication seen when HPV31 CIN612 9E cells were treated with the MRE11 exonuclease inhibitor Mirin (Anacker *et al*, 2014).

Although functional MRE11 is important for the progression of HPV infection, there is currently no evidence on how HPV infection affects the ability of MRE11 to carry out its normal function on the hosts DNA. While high levels of functional MRE11 have been identified with HPV infection, it is unknown whether the sequestering of MRE11 to sites

of viral replication prevents the recruitment of the protein to DNA damage sites requiring repair.

MRE11 IHC is potentially useful as a predictive biomarker for radiation-based therapy in MIBC. Considering this, the use of radiation-based treatment in anal cancer and the aberrant activity of MRE11 caused by HPV infection (present in the majority of anal cancers), the hypothesis that MRE11 might be of use as a biomarker in anal cancer was tested.

The aim of this study was to identify if MRE11 IHC staining associated to patient outcome in the anal cancer tissue samples, either as a single marker or in combination with previously identified markers of anal cancer outcome (p16 or TIL). This research would:- 1) provide new information on whether MRE11 IHC might be a potential avenue of further investigation in anal cancer 2) begin to address the question as to whether MRE11 IHC might be associated with response radiation-based treatments in cancers other than MIBC.

4.2 Results

4.2.1 Patient demographics

Analysis of MRE11 in anal cancer was conducted on a subset of samples from a previously reported anal cancer cohort (Gilbert *et al*, 2013). In total, we were able to obtain tissue sections for 124/138 individual anal cancer patients. Loss of patients from the cohort at this stage was due to lack of remaining tissue in the block, absence of invasive pathology or presence of cauterisation in the sample.

Immunohistochemical staining of MRE11 was conducted on the 124 patient samples using the Bondmax autostainer with the re-optimised MRE11 staining protocol outlined in Chapter 3.2.2 of this thesis. Unfortunately, a further 42 patients were excluded from the analysis at this point. Reasons for this were a) technical issues with the autostainer washing away the smallest tissue sections, b) little tissue remaining in paraffin blocks and c) loss of invasive tissue in the cut sections. The patient demographic data for the 82 patients' evaluable samples is shown in Table 4.1.

Table 4.1: **The demographics of the anal cancer cohort.** The cohort contains a subset of 82 patients reported in Gilbert et al (2013) whose samples were stained for MRE11. The sex, age, T and N stage, p16 and p53 status, TIL score and number of cases of relapse are summarised.

Sex	Male – 39 (47.6%) Female - 43 (52.4%)
T stage (n=68)	T1- 7 (8.5%) T2- 26 (31.7%) T3- 22 (26.8%) T4- 13 (15.9%) Missing -14
N stage (n=71)	N0- 40 (48.8%) N1- 9 (11.0%) N2- 19 (23.2%) N3- 3 (3.6%) Missing- 11
p16 (n=82)	p16 +ve- 72 (87.8%) p16 –ve- 10 (12.2%)
p53 (n=82)	Strong p53 staining- 19 (23.2%) Negative-moderate p53 staining- 63 (76.8%)
TIL score (n=78)	TIL 1 (low/absent)- 18 (23.1%) TIL 2 (moderate)- 44 (56.4%) TIL 3 (high)- 16 (20.5%) Missing- 4
Number of incidences of relapse (n=82)	27

Gilbert *et al* (2013) previously determined p16 and p53 status, using IHC, on the complete (n=138) cohort patients used for this study of MRE11 in anal cancer. The Gilbert *et al* (2013) study identified that patients positive for p16 had a significantly lower risk of relapse compared to patients negative for p16 and that strong p53 staining was associated with a high risk of relapse. As the cohort used in this study only contained 82/138 patients from the Gibert *et al* (2013) cohort, we used their scoring of p16 and p53 IHC, on the 82 patients, to determine whether the p16 and p53 associations with outcome could be recapitulated in our smaller subset of patient tissue.

In our patient subset (n=82) of the Gibert et al (2013) cohort the relationships between p16 and p53 status and relapse was consistent with that previously reported. In univariate analysis, p16-positive cases showed a significantly lower risk of relapse (hr: 0.16, 95% CI: 0.07-0.40, p≤0.001) compared to p16-negative cases (Figure 4.2A), whereas strong p53 staining was associated with a significantly higher risk of relapse (hr: 2.3, 95% CI: 1.05-5.22, p=0.037) compared to negative-moderate p53 staining (Figure 4.2B).

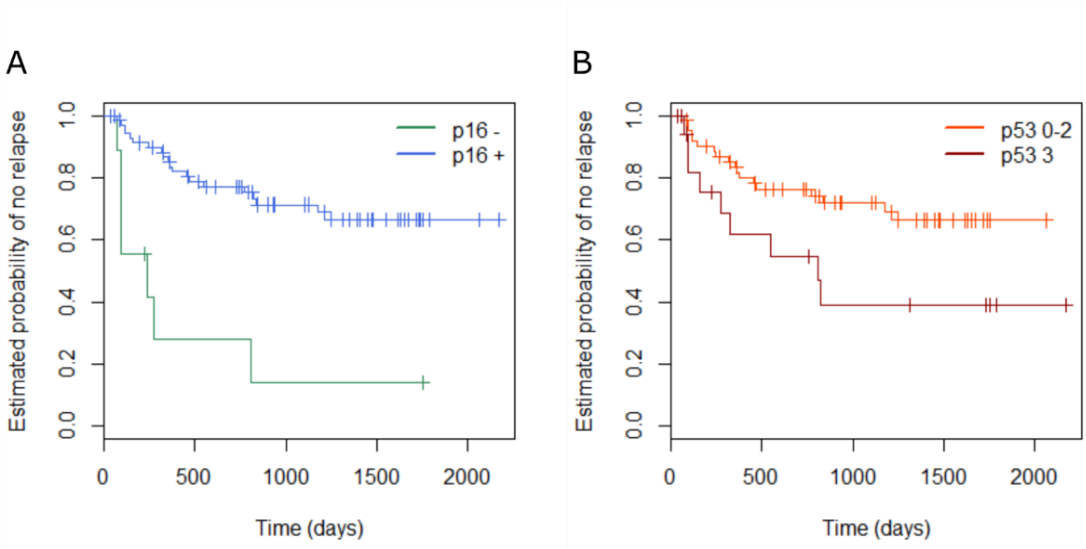


Figure 4.2: Kaplan-Meier relapse-free survival curves for patients stratified by p16 and p53 status (n=82). (A) Patients negative for p16 staining show an increased risk of relapse compared to patients positive for p16 staining (Cox proportional hazards- hr: 0.16, 95% CI: 0.07-0.40, p≤0.001) . Negative p16 staining is illustrated by the green (p16 -) curve. Positive p16 staining is illustrated in the blue (p16 +) curve; (B) Patients with strong levels of p53 staining show a significantly increased risk of relapse compared to patients with negative-moderate p53 staining (Cox proportional hazards- hr: 2.3, 95% CI: 1.05-5.22, p=0.037). Strong p53 intensity staining is illustrated by the dark red (p53=3) curve. Negative to moderate p53 intensity staining is illustrated by the orange (p53=0-2) curve. Statistics performed by Dr Christiana Kartsonaki.

Immunohistochemical staining of MRE11 on the 82 anal cancer patient samples was analysed using SQS. A broad range of staining was seen throughout the cohort (Figure 4.3 A and B).

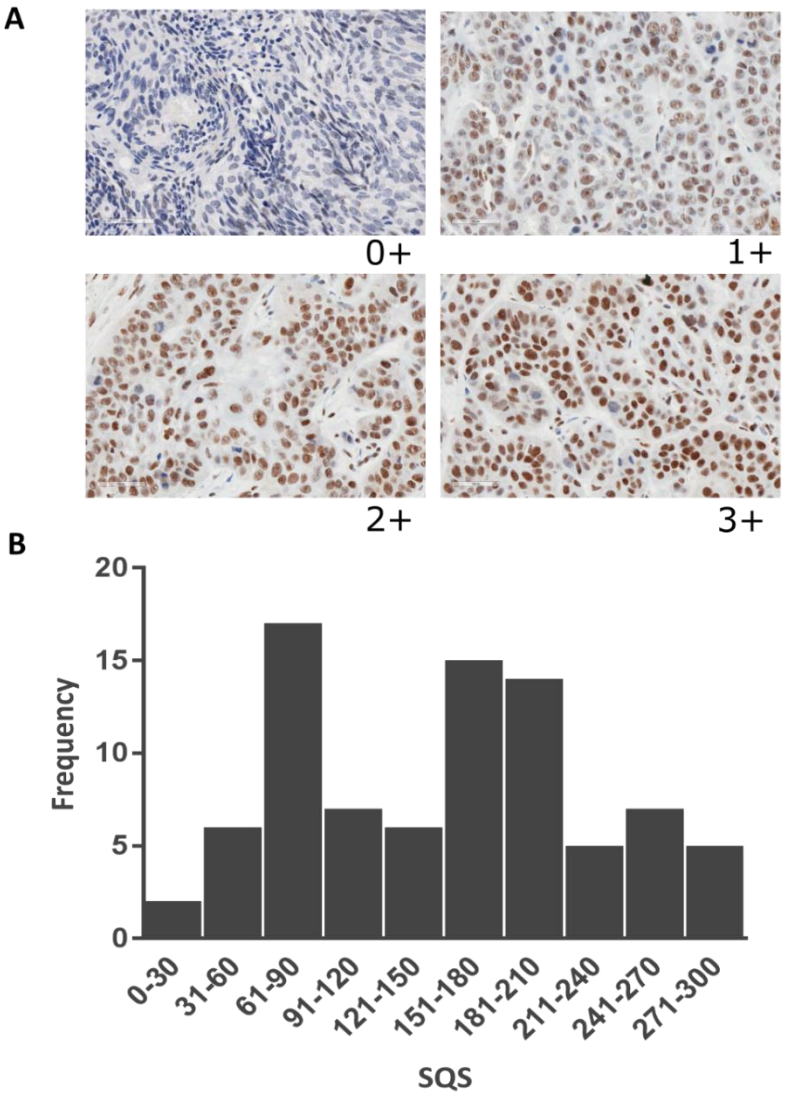


Figure 4.3: **MRE11 IHC in anal cancer specimens (n=82).** (A) Representative images for staining intensities of MRE11 in anal cancer tissue. Negative, weak, moderate and strong staining intensities are described by the values of 0, 1+, 2+ and 3+ respectively; (B) Histogram displaying the distribution of MRE11 SQS scores.

4.2.2 Association between MRE11 staining of anal cancer tissue and relapse free survival

Univariate analysis of MRE11 staining showed no association between MRE11 SQS and risk of relapse (hr: 0.999, 95% CI: 0.994-1.004, p=0.61) (Figure 4.4A). In multivariate Cox proportional hazards analysis with T stage, N stage, sex, p16, p53 and MRE11 status as covariates, only p16 was found to be of prognostic value (hr: 0.093, 95% CI: 0.024-0.359, p≤0.001), a result consistent with previously published data on this cohort from Gilbert *et al* (2013). This result was also seen when multivariate Cox regression analysis was conducted using the constituent % positive scores and intensity scores, in place of MRE11 SQS (Table 4.2).

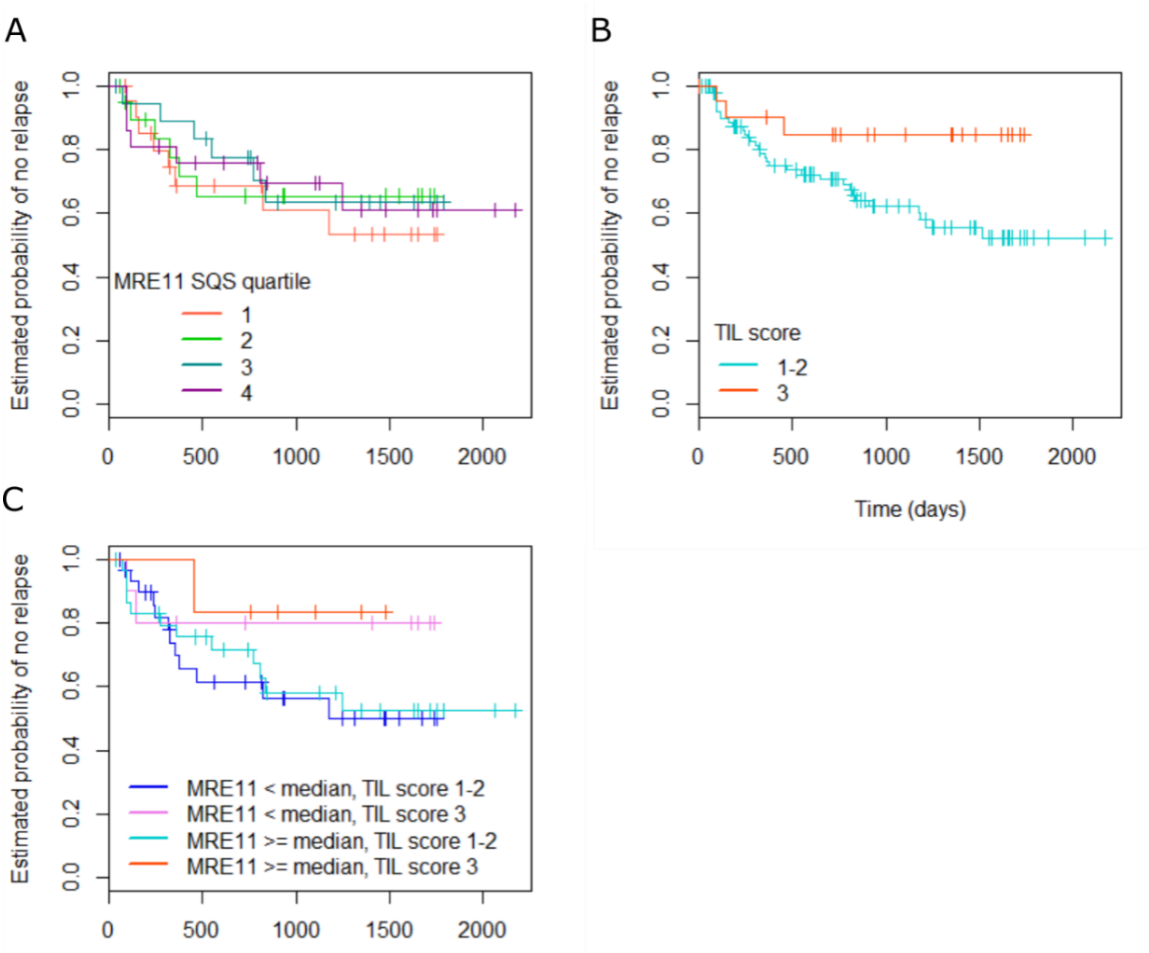


Figure 4.4: **The results of MRE11 IHC in anal cancer specimens**. (A) Kaplan-Meier relapse-free survival curves for 82 anal cancer patients stratified by MRE11 score. The first, second, third and fourth quartiles for the MRE11 SQS are illustrated by the orange (1), green (2), blue (3) and purple (4) curves, respectively. No significant association between MRE11 SQS and relapse-free survival was identified (Cox proportional hazards- hr: 0.999, 95% CI: 0.994-1.004, p=0.61); (B) Kaplan–Meier relapse-free survival curves for 115 anal cancer patients stratified by tumour infiltrating lymphocyte (TIL) score. High and low TIL scores are illustrated by the orange and blue curves, respectively. High TIL scores significantly associated with an increase in relapse-free survival (Cox proportional hazards- hr: 0.466, 95% CI: 0.261-0.832, p=0.010) (C) Kaplan–Meier relapse-free survival curves for 78 anal cancer patients stratified by MRE11 and TIL scores. The addition of MRE11 SQS to TIL score had no prognostic value over and above the use of TIL score alone (Cox proportional hazards using the interaction between MRE11 SQS and TIL score as the explanatory variable- hr: 0.83, 95% CI: 0.066- 10.449, p= 0.89). Statistics performed by Dr Christiana Kartsonaki.

Table 4.2: Results of the multivariate analysis on the anal cancer tissue cohort using the cox proportional hazards model. MRE11, T stage, N stage, sex, p16 status and p53 status are included as covariates. Multivariate analysis which used MRE11 SQS as the MRE11 variable is described separately to the model which used MRE11 % positive and intensity scores as two separate variables in the analysis. The hazard ratio for relapse-free survival is displayed with 95% confidence intervals (CI) and p values. Statistics performed by Dr Christiana Kartsonaki.

	MRE11 SQS Hazard ratio (95% CI, p value) n=67	**MRE11 % positivity and MRE11 intensity analysed as separate covariates** Hazard ratio (95% CI, p value) n=67
MRE11	MRE11 SQS: 0.998 (95% CI: 0.991-1.004, p=0.50)	MRE11 % positive: 0.37 (95% CI: 0.065-2.080, p=0.257 MRE11 intensity score: 2 vs 1: 1.98 (95% CI=0.548–7.175, p=0.297 3 vs 1: 0.65 (95% CI=0.128–3.287, p=0.602)
T stage	1.73 (95% CI: 0.953-3.129, p=3.129)	1.95 (95% CI: 1.079-3.526, p=0.069)
N stage	1.15 (95% CI: 0.711-1.864, p=0.565)	1.11 (95% CI: 0.632-1.955, p=0.713)
Sex	0.093 (95% CI: 0691-4.251)	2.130 (95% CI: 0.632-1.955, p=0.239)
p16 status	0.093, 95% CI: 0.024-0.359, p≤0.001	0.039 (95% CI: 0.006-0.230, p≤0.001)
p53 status	1.147 (95% CI: 0.407-3.233, p=0.795)	1.227 (95% CI: 0.406-3.712, p=0.717)

Analysis of MRE11 SQS in p16-negative cases was attempted. No association between MRE11 SQS and relapse was seen. However, the number of p16-negative individuals (n=9) and events were small and deemed unreliable for statistical analysis. In the individual analysis of MRE11 % positivity and intensity scores in p16-negative cases there was evidence for a non-significant protective effect for high MRE11 % positivity on the risk of relapse (hr:0.37, CI:0.1168-1.156, p=0.09), although again, due to the

limited number of cases, no strong conclusion could be drawn here. There was no substantial difference in the range of staining observed between p16-positive and p16-negative cases.

Using this cohort of patients, tumour infiltrating lymphocytes (TILs) have recently been reported to add prognostic value over and above p16 IHC in anal cancer (Gilbert et al, 2016). We therefore investigated MRE11 alongside the previously determined TIL scores on the patient subset analysed in this study. Consistent with previous findings, in this patient subset, a high TIL score was associated with improved risk of relapse (hr: 0.466, 95% CI: 0.261-0.832, p=0.010) (Figure 4.4B). The addition of MRE11 SQS to TIL score had no prognostic value over and above the use of TIL score alone. No significant association with relapse free survival was found, when Cox proportional hazards modelling was performed using the interaction between MRE11 SQS and TIL score as the explanatory variable (hr: 0.83, 95% CI: 0.066- 10.449, p= 0.89) (Figure 4.4C).

4.2.3 Correlation between variables used in the analysis

We next investigated if there was any dependence between MRE11 scores (MRE11 SQS, MRE11 % positive or MRE11 intensity scoring) and any of the other variables (Sex, T stage, N stage, p16, p53 and TIL) used in this study. Identifying correlations between variables is a useful method of identifying potential interactions in the dataset and confounding factors. If a correlation between MRE11 scores and another variable was identified in this analysis, it may suggest an interaction, which could be of biological importance and interest.

A correlation matrix plotting the Pearson's correlation coefficient between each of the variables used in the analysis of MRE11 staining on anal cancer tissue was generated

(Figure 4.5). However, MRE11 levels assessed by SQS, percentage positivity or intensity scoring, were not highly correlated with any of the other variables used in the analysis.

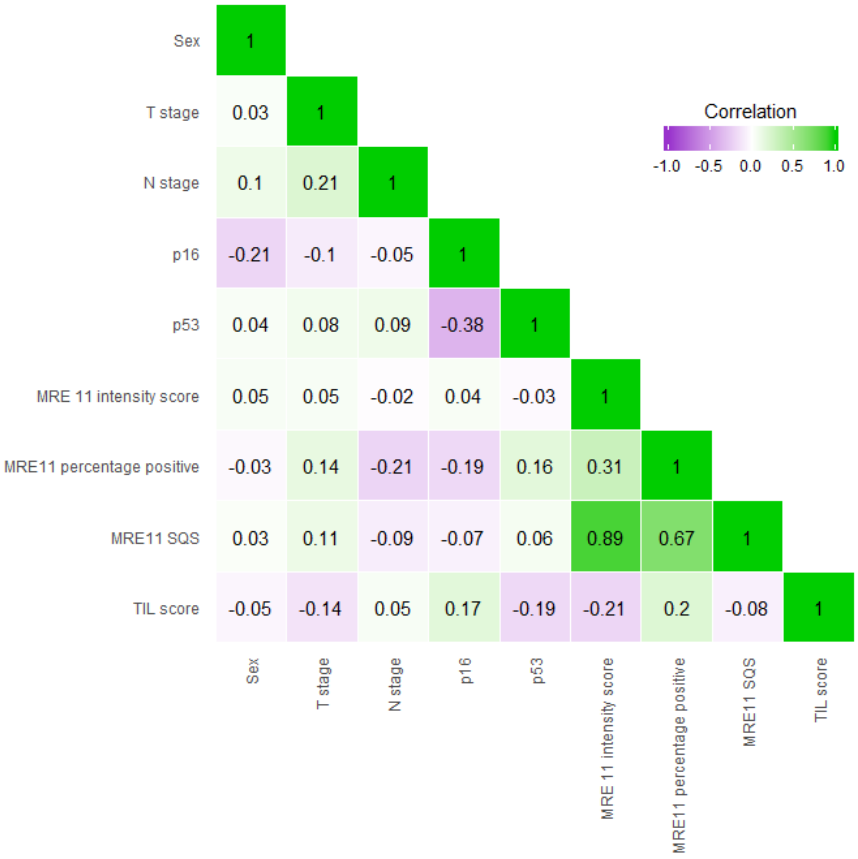


Figure 4.5: Correlation matrix of the Pearson's correlation coefficients for the variables used in this study. Negative values are highlighted in purple and indicate a negative correlation between variables. Positive values are highlighted in green and indicate a positive correlation between variables. The intensity of the purple or green colour indicates the strength of the correlation between the variables. The strength of a correlation was determined as: very high if the correlation coefficient was between 1.0-0.9, high (correlation coefficient: 0.9-0.7), moderate (correlation coefficient: 0.7-0.5), low (correlation coefficient: 0.5-0.3) or negligible (0.3-0). Statistics performed by Dr Christiana Kartsonaki.

4.3 Discussion

In contrast to published findings in MIBC, MRE11 was not identified as having predictive value in assessing outcome in anal cancer following CRT, despite the similarities in treatment technique and dose between the two tumour sites. Presumably this represents the heterogeneity that exists between cancers (Burrell *et al*, 2013). Indeed, the prevalence of HPV infection in anal cancer is a major difference between the two cancers. The hijacking of the DDR and production of ROS by HPV leaves infected cells with high levels of endogenous DNA damage and genomic instability. Indeed, the role of ROS and impaired DRR systems in HPV-positive cancer has been linked to the high response rates seen with CRT, due to the creation of clustered DNA damage (Marullo *et al*, 2015).

We assessed whether MRE11 had any role to play, over and above previously described prognostic markers (p16 and TILs (Jones *et al*, 2017)) but found no evidence of an association with outcome. It may be possible that, despite the lack of correlation to outcome in p16/HPV-positive tissue, the assessment of MRE11 levels in p16/HPV-negative tissues might yield positive findings. However, due to the rarity of these tumours this would be difficult to investigate.

Outcome for MIBC is largely assessed using CSS whereas outcome in anal cancer is generally evaluated using relapse-free survival (RFS). The reason for this difference is due to the high rate of complete tumour regression seen in anal cancer following treatment, making RFS a more appropriate outcome for analysis. It is conceivable that this alteration in endpoint may account for the difference between the predictive capacity of MRE11 in MIBC and anal cancer and should be acknowledged.

Our assessment of MRE11 levels in anal cancer was conducted using IHC. Although a powerful tool it has its limitations. While IHC can determine the level of a particular protein in a sample, it cannot be used to determine if that protein is functional. Additionally if there are alterations present in the protein such that the antibody cannot recognise its epitope, results can be unreliable. Before discounting the role of MRE11 in anal cancer, complementary molecular studies should be undertaken to explore the features of MRE11 in anal cancer which cannot be addressed by the use of IHC. It is possible that mutations within the *MRE11* gene, which do not translate to protein abundance but affect protein function, might have associations with outcome.

The treatment of anal cancer achieves good response rates. However, for certain patients, such as those with HPV-negative anal cancer, treatment outcomes are poor and toxicities can be significantly detrimental to QOL. For this 'hard to treat' population of anal cancer patients more research is needed to improve treatment outcomes. There is much evidence for altered DDR during HPV transformation but this has not yet been fully examined in the context of anal cancer, particularly in response to therapy.

Chapter 5 Characterising the effects of the germline SNP, rs1805363, on the repair of radiation induced DNA damage in MIBC

5.1 Introduction

Research conducted by Teo *et al* (2014) in the Kiltie lab, into germline *MRE11* variants as biomarkers for RT outcome in MIBC, revealed a germline SNP, rs1805363, associated with worse CSS following RT (per allele hr 2.10, 95% CI 1.34–3.28, p= 0.001) but not RC (per allele hr 0.99, 95% CI 0.61–1.60, p= 0.89). Furthermore, the association was found to follow a gene dosage effect, whereby the carriage of each A minor allele was found to contribute to a decrease in CSS. This result is counter to the association identified by Choudhury *et al* (2010) and Laurberg *et al* (2012) for MRE11 protein levels. It is possible, therefore, that rs1805363 might confound results seen in the analysis of total MRE11 protein expression. The rs1805363 finding identifies a group of MIBC patients who would not benefit from the use of total MRE11 level analysis by IHC as a predictive biomarker and highlights the need for multiple biomarkers in assessing disease.

rs1805363 is a transition mutation of the G major allele to an A minor allele. A minor allele frequency of 11% was found for the SNP in the work conducted by Teo *et al* (2014). However, bioinformatics tool SNPeadia (SNPeadia, 2015) reports a global minor allele frequency of 3.9% and an approximate 20% European minor allele frequency. rs1805363 is located intronically in *MRE11* isoform 1, five bases 3′ to the Exon 1 AG donor splice site. In *MRE11* isoform 2, rs1805363 is located within the 5′ UTR (Figure 1.16B). Due to the proximity of the SNP to the donor splice site it was hypothesised that rs1805363 might affect gene splicing and/or relative isoform expression of *MRE11*. *MRE11* isoform analysis on nine MIBC samples (six bladder cancer cell lines and three

primary tumours) supported this hypothesis with samples genotyped as GG showing the highest *MRE11* isoform 1: isoform 2 ratio and AA genotyped samples showing the lowest isoform 1: isoform 2 ratio (Figure 1.16C) (isoform 1 relative percentage: GG 61.7%, GA 49.6% and AA 30.3%) (Teo *et al*, 2014).

On a protein level, the only identified difference in *MRE11* isoform 2 compared to *MRE11* isoform 1 is the absence of exon 16, a 27 amino-acid-long exon located at position 595-622 on the amino acid chain. *MRE11* isoform 1 is approximately 80 kDa in size and consists of 708 amino acids whereas *MRE11* isoform 2 is approximately 77 kDa and is composed of a 680 amino-acid-long chain.

MRE11 exon 16 is partially encompassed by the MRE11 glycine-arginine-rich (GAR) domain, located between the 566-600 aa region of the protein. The MRE11 GAR domain undergoes ADMA by PRMT1 (Figure 1.10) and is critical for the proper functioning of the protein. Removal of GAR methylation in MRE11 causes increased genomic instability and hypersensitivity to gamma irradiation (Yu *et al*, 2012). Further characterisation of MRE11 lacking GAR methylation has identified defects in MRE11 exonuclease activity, DNA binding, DNA end resection, ATR-dependent CHK1 activation (Yu *et al*, 2012), recruitment of MRE11 to DNA damage foci (Déry *et al*, 2008) and impairment of DNA damage checkpoints (Boisvert *et al*, 2005a). The *MRE11* GAR domain has also been implicated in regulating the nuclear compartmentalisation of MRE11 (Boisvert *et al*, 2005b). While exon 16 lies partially within the *MRE11* GAR domain, it lies outside the 570-594 region directly implicated in the regulation of exonuclease activity (Boisvert *et al*, 2005a) and is located in an area of the GAR domain which shows least conservation in sequence (Déry *et al*, 2008).

MRE11 is part of the MRN complex and requires RAD50 and NBS1 for full functionality (D'Amours & Jackson, 2002). The MRN complex is a flexible assembly that undergoes large conformational changes that regulate its activity, assists in DNA damage repair pathway choice and alters the ability of the complex to bind either DNA or interacting proteins (Lafrance-Vanasse *et al*, 2015). If the loss of exon 16 results in conformational changes in MRE11, it is possible that the interactions of MRE11 with the members of the MRN complex could be affected. Additionally the absence of exon 16 may also affect protein-protein interactions in larger complexes, and DNA binding.

The findings of Teo *et al* (2014) highlight interesting areas for investigation. In this work, the association between increased *MRE11* isoform 2 expression and the presence of the rs1805363 A minor allele was validated, in a collection of MIBC tumours.

It was hypothesised, that the increased expression of *MRE11* isoform 2 could be responsible for the poor response to RT seen in patients positive for the rs1805363 A minor allele and that cells expressing high levels of *MRE11* isoform 2 may be more radioresistant than cells with high *MRE11* isoform 1 expression. To test this, cell lines were generated that expressed an MRE11 construct containing the coding sequence of *MRE11* isoform 1 or *MRE11* isoform 2. These cells were then used in assays to determine cellular radiosensitivity and DSB repair efficiency. Further to this, the cell lines were also used in immunoprecipitation experiments, to investigate whether the *MRE11* isoform 2 construct was positive for the methylation of the MRE11 GAR domain and able to bind the members of the MRN complex.

5.2 Results

5.2.1 rs1805363 is associated with increased expression of *MRE11* isoform 2 in MIBC

To validate the result of Teo *et al* (2014), 189 samples of germline DNA, extracted from the blood of MIBC patients, were obtained from Richard Bryan (University of Birmingham). The germline DNA was genotyped for the rs1805363 G>A SNP. Of the 189 samples, 22 were identified as heterozygous (AG) for the rs1805363 SNP with the remaining samples identified as WT (GG). A table of all genotyped samples can be found in appendix G.

Loss of heterozygosity (LOH) within the tumours was investigated by genotyping DNA extracted from tumour samples of patients that were germline heterozygous for the rs1805363 A minor allele. Extracted tumour DNA was available for 17 of the germline genotyped AG cases. In order to assess LOH, the results of tumour genotyping were compared to germline genotype. LOH was identified in 7 of the 17 samples (AA= 4/17, GG= 3/17, AG= 10/17) (Table 5.1).

Table 5.1: **Genotyping of the *MRE11* rs1805363 SNP on 17 samples of tumour extracted DNA**. The 17 tumour extracted DNA samples were from patients germline genotyped as AG. Samples where a loss of heterozygosity was identified in tumour DNA are highlighted in grey. A loss of heterozygosity toward the A minor allele is highlighted in pink. A loss of heterozygosity toward the G major allele is highlighted in orange.

Sample ID	Loss of Heterozygosity	Genotype
s0136	Yes	**AA**
s0168	Yes	**AA**
S0237	Yes	**AA**
s0295	Yes	**GG**
S0472	No	AG
S0521	Yes	**GG**
S0637	No	AG
S0643	No	AG
S0724	No	AG
S0729	No	AG
S0890	No	AG
S1126	No	AG
S1214	No	AG
S1434	Yes	**AA**
S1829	Yes	**GG**
S2200	No	AG
S2209	No	AG

FFPE tumour blocks were obtained for RNA extraction for 36 of the germline genotyped samples. Twelve of the tumour blocks were not suitable for RNA extraction. This was due to either lack of invasive pathology in cut sections (6/12), the area of tissue being too small for sufficient material to be collected (5/12) or the tissue being identified as highly inflammed (1/12). Eleven of the blocks were from heterozygous (AG) germline patients with the remaining 14 corresponding to WT (GG). The RNA extracted from patient tumours was converted to cDNA and PCR primers for the 5'UTR of *MRE11* were used to distinguish the expression of *MRE11* isoform 1 and 2 in the samples. Due to the difference in the length of the 5'UTR between *MRE11* isoform 1 and 2, the PCR product

amplified from *MRE11* isoform 1 was 144bp long. The PCR product amplified from *MRE11* isoform 2 was 251bp long. The proportion of *MRE11* isoform 1 and 2 present in the tissue was analysed for each patient and grouped according to germline genotype (Figure 5.1). One sample was excluded from the analysis due to a failure in isoform amplification.

The results from the analysis of *MRE11* isoform expression, in this study, was in agreement with the findings of Teo *et al* (2014). Using a two-sample t test *MRE11* isoform 2 expression was found to be significantly higher ($p=0.007$) in the AG germline genotype patient group (mean: 19.44) compared to the GG germline genotype patient group (mean: 2.29).

One of the AG germline genotyped patient samples (S0521) was identified as negative for *MRE11* isoform 2 expression. However, this sample had also been tumour genotyped and had been found to have undergone a LOH within the tumour to become GG. It is, therefore, possible that the change in tumour genotype from the germline AG to the tumour GG could be the reason for the absence of *MRE11 i*soform 2 within this sample.

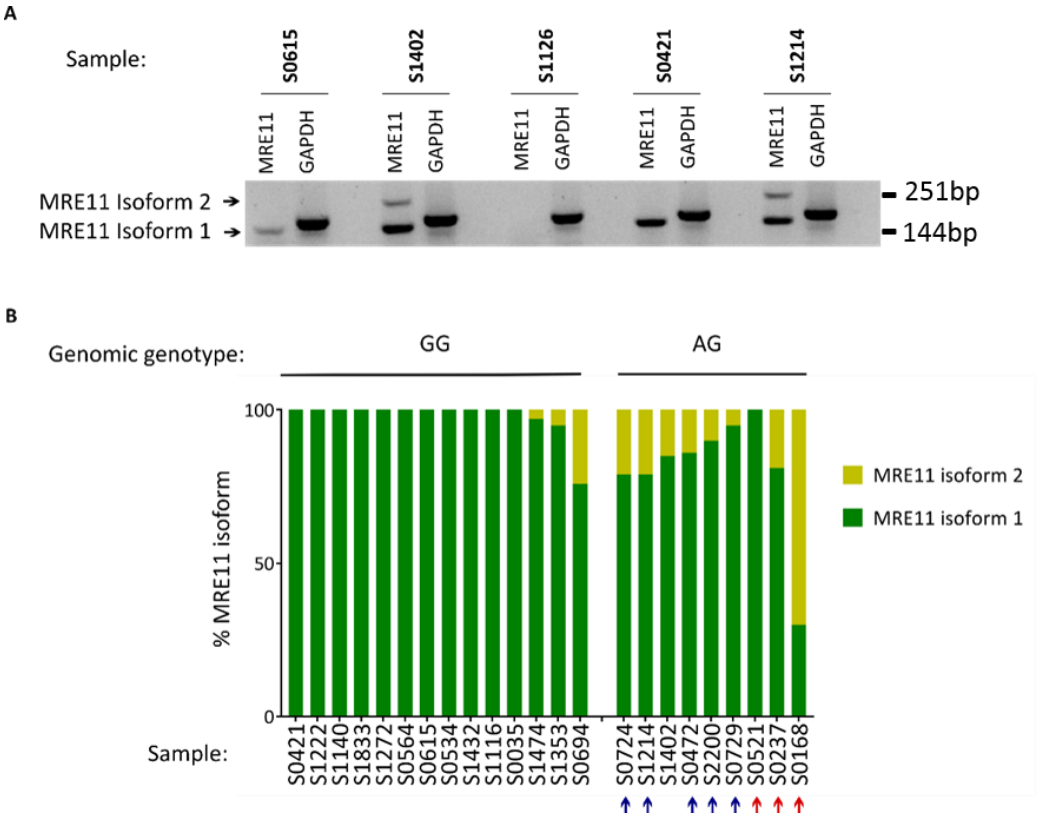


Figure 5.1: The results of the analysis into *MRE11* isoform expression in MIBC patient tissue. (A) A representative image of the electrophoresis bands produced from PCRs conducted on patient cDNA. The 144bp product is amplified from the *MRE11* isoform 1 5'UTR and the 251bp product is amplified from longer *MRE11* isoform 2 5'UTR. A PCR for *GAPDH* was run on all of the cDNA samples as a positive control for RNA extraction and cDNA conversion; (B) Quantification of *MRE11* isoform expression. *MRE11* isoform 1 and 2 expression is displayed as a percentage of the total amount of *MRE11* in each sample. Samples are grouped according to germline genotype. *MRE11* isoform 2 expression is significantly increased in the AG genotype group compared to the GG genotype group (t test p=0.007). Blue arrows indicate samples which were tumour genotyped as AG. Red arrows indicate samples which were tumour genotyped as either AA or GG.

5.2.2 The generation of MRE11 stable cell lines

Having found evidence in support of Teo *et al*'s (2014) findings, the characterisation of the potential effect of increased expression of *MRE11* isoform 2 on DNA damage repair and MRE11 function was undertaken. Two MRE11 constructs were stably re-expressed in an MRE11 knockdown cell line. *MRE11* isoform 1 expression was represented by a full length MRE11 plasmid (MRE11 full length) whereas *MRE11* isoform 2 expression was represented by an MRE11 plasmid lacking exon 16 (MRE11 exon16-ve).

Putative stable MRE11 re-expression clones were analysed using western blotting to confirm MRE11 re-expression (Figure 5.2). The level of MRE11 re-expression was calculated by quantifying the signal intensity of the blotted protein bands using Image Studiolite software (LI-COR). Using the signal intensity values for MRE11 from three biological repeat experiments, the mean fold increase in MRE11 protein abundance in the MRE11 re-expression clones could be calculated relative to the MRE11 knockdown cell line. To control for variation in gel loading, the signal intensity values for each of the samples were normalised to β actin before the fold increase in MRE11 protein expression was calculated.

MRE11 protein levels were raised in both the MRE11 full length and MRE11 exon16-ve re-expression clones, compared to the MRE11 knockdown cell line (Figure 5.2 B and E). The mean fold increase in MRE11 band intensity for the full length re-expression clones 1 and 2 were 2.36 and 2.58 respectively (n=3). The mean fold increase in MRE11 band intensity for the exon16-ve re-expression clones 1 and 2 were 2.65 and 2.53, respectively.

Although an increase in the protein levels of MRE11 was observed in the clones, compared to the MRE11 knockdown cell line, the degree of re-expression was not sufficient to rescue MRE11 levels up to that seen in the parental T24 cell line (Figure 5.2 C and F).

The MRE11 full length clone 2 (Figure 5.2A) was used as the MRE11 full length cell line and the MRE11 exon16-ve clone 1 (Figure 5.2D) was used as the MRE11 exon16-ve cell line in all subsequent experiments.

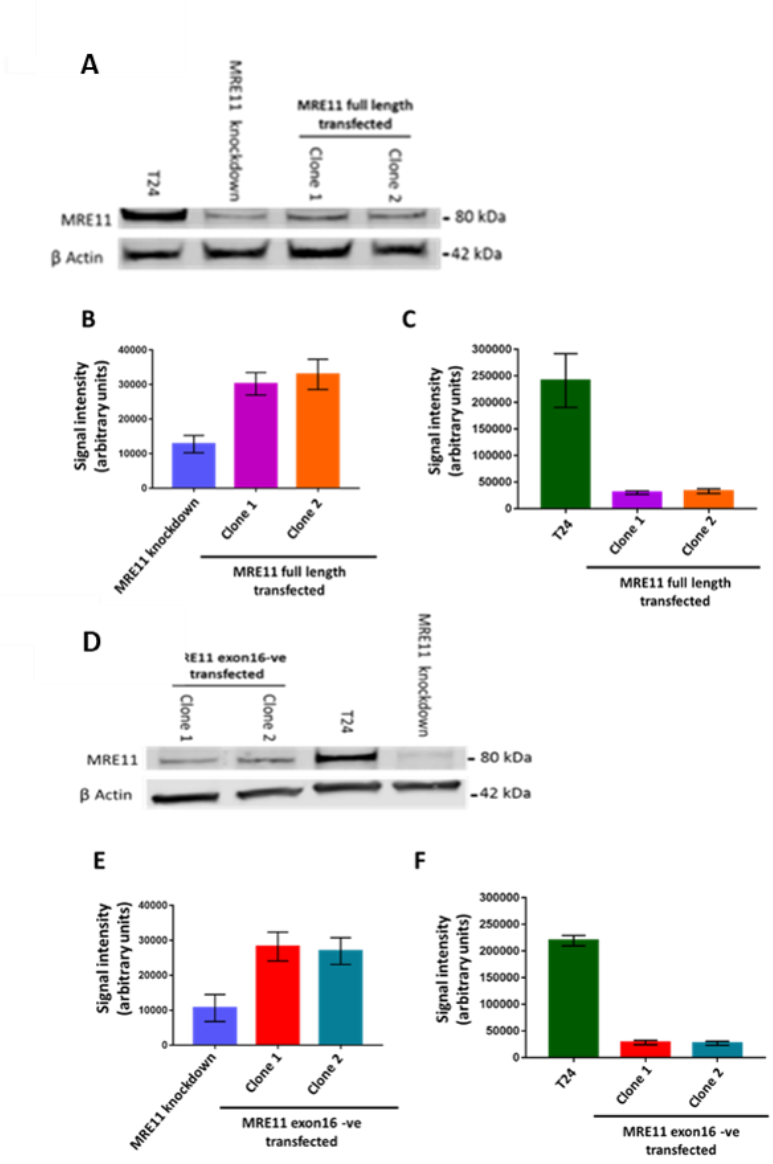


Figure 5.2: Western blot analysis of MRE11 re-expression clones. (A) A representative western blot probed for MRE11. Two MRE11 full length re-expression clones are shown with T24 and MRE11 knockdown cell lines. β Actin is blotted as a sample loading control; (B) Quantification of the mean densiometric signal intensity of MRE11 in the MRE11 knockdown cell line and MRE11 full length clones, normalised to the β Actin loading control (n=3); (C) Quantification of the mean densiometric signal intensity of MRE11 in the T24 cell line with the MRE11 full length clones normalised to the β Actin loading control (n=3); (D) A representative western blot probed for MRE11. Two MRE11 exon16-ve re-expression clones are shown with T24 and MRE11 knockdown cell lines. β Actin is blotted as a sample loading control; (E) Quantification of the mean densiometric signal intensity of MRE11 in the MRE11 knockdown cell line and MRE11 exon16-ve clones, normalised to the β Actin loading control (n=3); (C) Quantification of the mean densiometric signal intensity of MRE11 in the T24 cell line with MRE11 exon16-ve clones normalised to the β Actin loading control (n=3).

To confirm that the MRE11 exon16-ve clone was lacking exon 16 a range of techniques were employed. An antibody for MRE11 exon 16 was generated using a peptide aptamer corresponding to exon 16 as the antigen (Moravian Biotechnology). Although, unfortunately this antibody proved to be unsuitable for western blot analysis. Low percentage SDS page gels were also utilised to try and physically separate full length MRE11 and MRE11 lacking exon 16. However, the difference in size between the two proteins was found to be too small to detect in this manner.

qPCR was finally used to confirm the absence of MRE11 exon 16 in the MRE11 exon16-ve clones (Figure 5.3). Amplification of a central region of MRE11 confirmed a significant increase in MRE11 level in both MRE11 full length and MRE11 exon16-ve clones compared to MRE11 knockdown cells (MRE11 full length $p=0.021$, MRE11 exon16-ve $p=0.019$). In addition to this it was identified that there was no significant difference in MRE11 expression between the between MRE11 full length and MRE11 exon16-ve clones ($p=0.164$).

The presence of exon 16 was confirmed in the MRE11 full length clone by the amplification of a region flanking the exon 15-exon 16 junction. In the MRE11 full length clone no significant difference was identified between the amplification of exon 16 and the central region of MRE11 ($p=0.164$). However, in the MRE11 exon16-ve clone the amplification of exon 16 was significantly reduced compared to that seen when using primers to amplify the central region of MRE11 ($p=0.026$). Additionally, no significant difference was identified between the level of exon 16 amplification detected in the MRE11 exon16-ve clone compared to that detected in the MRE11 knockdown cell line

(p=0.758) (Figure 5.3). qPCR analysis of MRE11 re-expression clones confirmed both the re-expression of MRE11 and the absence of exon 16 in the MRE11 exon16-ve clone.

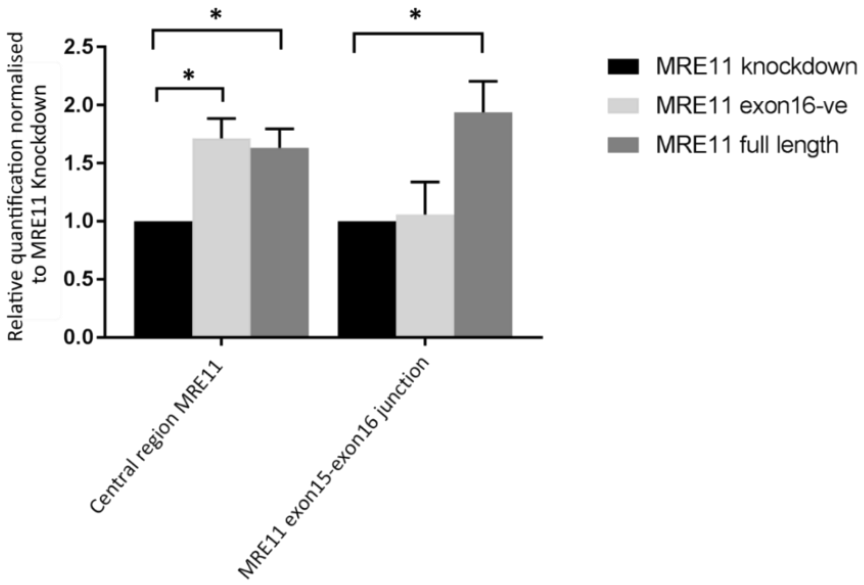


Figure 5.3: qPCR of MRE11 re-expression clones. The mean quantification of MRE11 amplification products are normalised to MRE11 knockdown cells (n=3). Total MRE11 is quantified by the amplification of a central region of MRE11 upstream of exon 16. The presence of MRE11 exon 16 is assessed by the amplification of a target region flanking the exon 15-exon 16 junction. Significant variation in MRE11 amplification products in the MRE11 re-expression clones relative to the MRE11 knockdown cells are indicated by asterisk (*= $p<0.05$).

5.2.3 Determining the radiosensitivity of MRE11 clones

To determine the radiosensitivity of MRE11 re-expression clones, a clonogenic survival assay was conducted. Analysis with one-way ANOVA revealed significant differences in survival between cell lines at 4 Gy (p=0.003) and 6 Gy (p=0.004) doses of IR (Figure 5.4). The MRE11 full length re-expression cell line showed a comparable response to radiation to the scrambled control cell line with no significant differences in cell survival at any dose. A significant increase in cell survival was seen at 4 Gy and 6 Gy doses for the MRE11 full length cell line compared to the MRE11 knockdown cell line (4 Gy

p=0.050, 6 Gy p=0.023) and for the MRE11 full length cell line compared to the MRE11 exon16-ve cell line (4 Gy p=0.045, 6 Gy p=0.016).

MRE11 exon16-ve cells showed greater radiosensitivity than the MRE11 full length cell line, with significantly different survival at 4 Gy and 6 Gy doses from MRE11 scrambled control cells (4 Gy p=0.006, 6 Gy p=0.015) and a similar response to radiation compared to MRE11 knockdown cells (4 Gy p=0.367, 6 Gy p=0.315). At 10% survival the enhancement ratio between the MRE11 full length and MRE11 exon 16-ve survival curves was 1.3 (Figure 5.4).

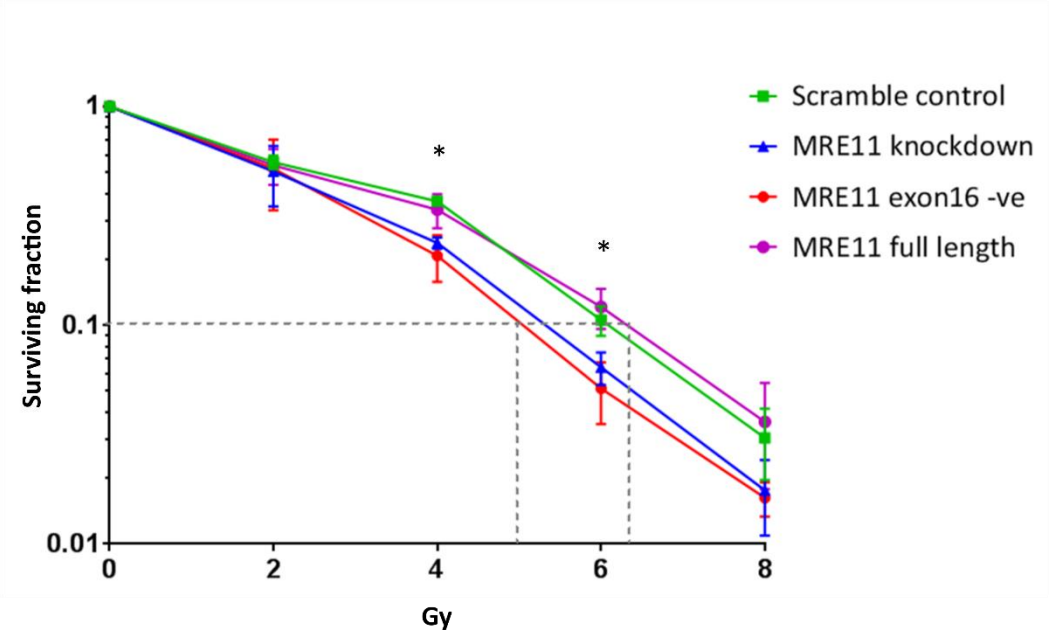


Figure 5.4: The clonogenic survival of MRE11 re-expression clones (n=3). Significant variation in the survival of cell lines at 4 Gy (p=0.003) and 6 Gy (p=0.004) is highlighted by asterisks (* = p<0.05). The dotted line indicates the enhancement ratio between MRE11 full length and MRE11 exon16-ve clones at 10% survival.

The results from the clonogenic survival assay were counter to the initial hypothesis that *MRE11* isoform 2/MRE11 lacking exon 16 might confer resistance to radiation-induced damage. To explore this findings further, neutral comet assays were conducted using MRE11 re-expression clones. Initial results using a general protocol for the neutral comet assay elicited poor quality comets in the cell lines used. This made analysis difficult and potentially unreliable. Therefore, a process of assay optimisation was undertaken to improve the clarity of comet heads and to optimise comet tails produced by the assay. Steps taken to optimise the comet assay are shown in Table 5.2.

Table 5.2: **Optimisation of the conditions used in the Neutral Comet Assay, for T24 derived cell lines**. Conditions highlighted in green indicate the alterations which were chosen for the final assay protocol.

Stage	Original Condition	Alterations	Outcome
Radiation dose	**20 Gy**	**-10 Gy** **-5 Gy**	**-Improved number of comets for analysis, sufficient tail** **-Poor tail formation in damaged cells**
Low melt point agarose %	**1%**	**0.8%**	**-Increased halo around comet head, reduced definition of tail**
Lysis conditions	**2 Hr Lysis in 1% DMSO**	**-10% DMSO** **-Overnight lysis**	**-Improved tail formation in irradiated cells** **-Increased tail formation in all conditions including untreated**
Running conditions	**TBE room temp 1.2 V/cm 25 minutes**	**-Cold TBE** **-Cold TAE** **-1.5 V/cm** **-1.3 V/cm** **-25,30,35,40 minutes** **-45min(0.8 V/cm)**	**-No tail formed** **-No tail formed, increased halo around comet head** **-Increased tail but poor clarity, increased tail also observed in 0gy** **-Good production of tail, head and tail well defined** **-Good tail formation at 30 and 35 minutes** **-Very reduced tail formation**

Once the optimised conditions for performing the comet assay were found, analysis of radiation-induced DNA damage repair could be performed on MRE11 re-expression clones (Figure 5.5). MRE11 exon16-ve clones showed significant differences in terms of endogenous and initial DNA damage and subsequent repair compared to the remainder of the cell lines used in the analysis.

MRE11 exon16-ve cells had significantly increased levels of endogenous DNA damage compared to all other cell lines (scrambled control p=0.046, MRE11 knockdown p=0.025, MRE11 full length p=0.044). Surprisingly, in contrast, MRE11 knockdown cells exhibited no increase in endogenous DNA damage compared to scrambled control cells (p=0.220). Additionally, no significant difference was seen in endogenous DNA damage in MRE11 full length cells compared to scrambled control cells. A heightened level of DNA damage in MRE11 exon16-ve cells was observed throughout the time course of repair, likely reflecting the increased level of DNA damage in untreated cells. Evidence for a delay in DNA damage repair was also present in MRE11 exon16-ve cells. In control, knockdown and MRE11 full length cell lines there was a significant decrease in Olive tail moment between 0 hour and 30 minute time points post-irradiation (scrambled control p=0.016, MRE11 knockdown p<0.001, MRE11 full length p=0.038). However, no significant difference in tail moment was seen in MRE11 exon16-ve comets between 0 hour and 30 minutes post-irradiation (p=0.549). This points toward a delay in the resolution of DNA damage in MRE11 exon16-ve cells. Levels of DNA damage in MRE11 exon16-ve comets remained high after 2 hours of repair.

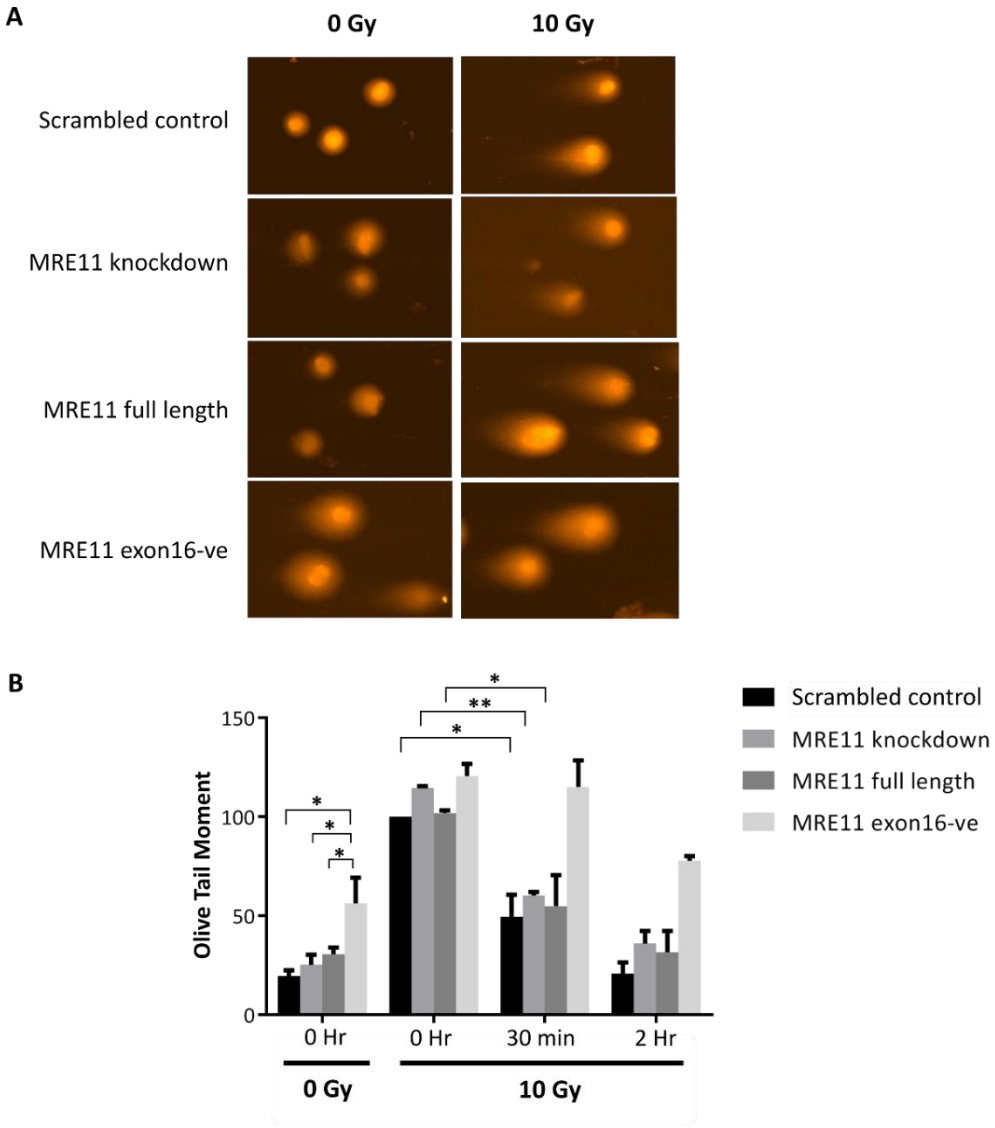


Figure 5.5: The neutral comet assay on the MRE11 re-expression clones (n=3). (A) Representative images of comets in the scrambled control, MRE11 knockdown, MRE11 full length and MRE11 exon16-ve cell lines for 0 Gy (untreated) cells and 10 Gy 0 Hr irraditated cells; (B) Bar chart of the mean Olive tail moment for each of the conditions used in the comet assay. Data is shown for untreated, 0 Gy, cells and cells treated with 10 Gy of irradiation. Irradiated cells were allowed to repair for 0 hr, 30 minutes or 2hrs before performing the comet assay. Significant variation in Olive tail moment between results are indicated by asterisks (* = $p<0.05$, ** = $p<0.01$).

In accordance with the results from the clonogenic survival assay, the results from the MRE11 full length clones in the comet assay showed no significant variation from those seen in the scrambled control cell line, at any point in the time course.

Interestingly, the MRE11 knockdown cell line did not show significant repair defects. There was no significant difference in the mean Olive tail moment between untreated MRE11 knockdown cells and MRE11 knockdown cells 2 hours post-irradiation. However, it was observed that in the MRE11 knockdown cell line more cells underwent cell death upon irradiation treatment. This was evidenced by the reduced number of cells available for harvesting in radiation-treated samples compared to untreated MRE11 knockdown samples.

5.2.4 The impact of MRE11 exon 16 on MRE11 methylation and the MRN complex

Having established an altered response to radiation in MRE11 exon16-ve cells, preliminary investigations into the underlying mechanisms behind this were made. Exon 16 is partially contained within the MRE11 GAR domain. To determine if methylation of the MRE11 GAR domain was affected by the absence of exon 16, immunoprecipitations of MRE11 were conducted from whole cell protein lysates and an antibody against asymmetrically dimethylated (ADMA) proteins was used to assess methylation of the MRE11 GAR domain.

The plasmids used to stably re-express MRE11 full length or MRE11 exon16-ve contained an HA tag. Initial immunoprecipitations were attempted using an antibody against this HA tag. Interestingly, the HA antibody was able to immunoprecipitate MRE11 in the MRE11 full length samples but not in the MRE11 exon16-ve samples.

Despite not being able to immunoprecipitate MRE11 using HA in MRE11 exon16-ve samples, it was possible to immunoprecipitate MRE11 in all samples with an MRE11 antibody. MRE11 immunoprecipitations were run on SDS page gels and blotted with the ADMA antibody. MRE11 full length and MRE11 exon16-ve immunoprecipitations were positive for asymmetric dimethylation using the ADMA antibody (AYM25) (n=3) (Figure 5.6).

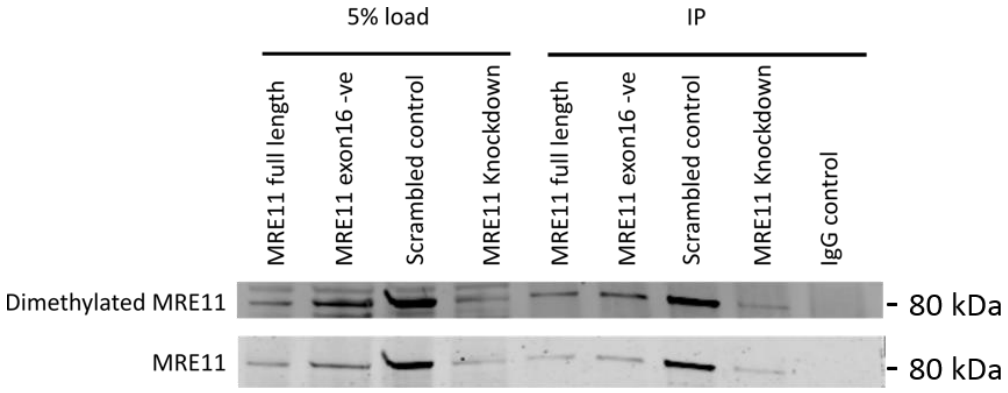


Figure 5.6: **A representative western blot of immunoprecipitated MRE11 and probed with an antibody against asymmetrically dimethylated proteins.** Blots were counter stained for MRE11.

MRE11 functions within the MRN complex. It is possible that the loss of exon 16 could impair the ability of MRE11 to bind RAD50 and/or NBS1, leading to functional defects. To determine if MRE11 lacking exon 16 was able to bind RAD50 and/or NBS1, an MRE11 antibody was used to immunoprecipitate MRE11 along with any MRE11 bound proteins. In MRE11 full length and MRE11 exon16-ve MRE11 immunoprecipitations, RAD50 and NBS1 were found to co-immunoprecipitate with MRE11 (Figure 5.7).

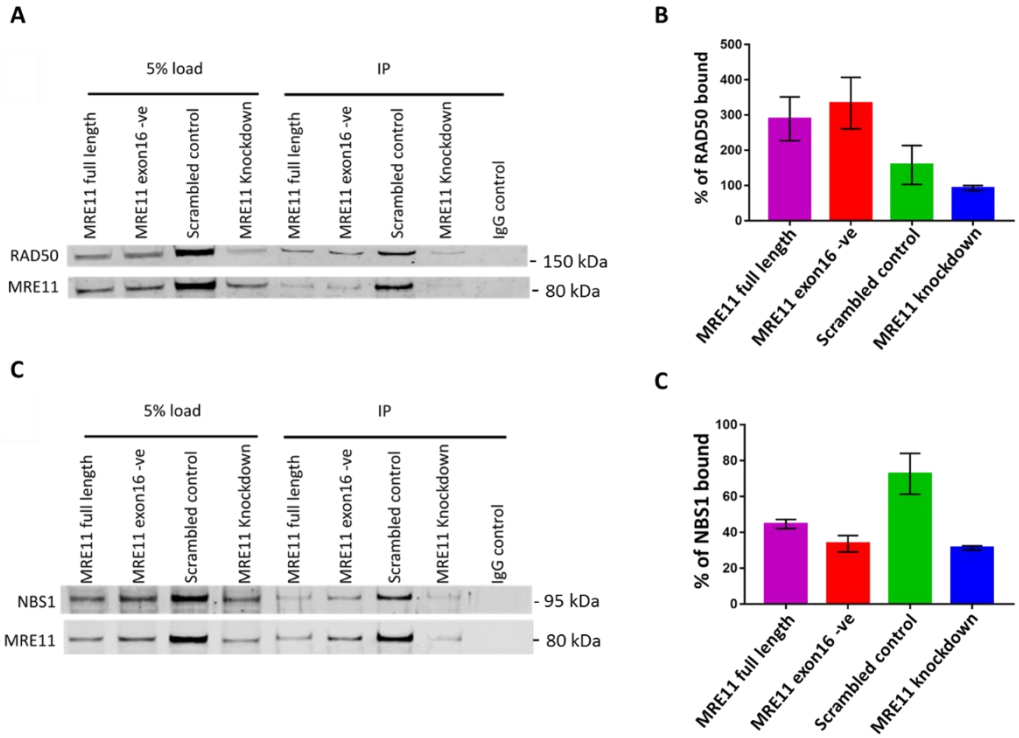


Figure 5.7: **MRE11 immunoprecipitations blotted for RAD50 and NBS1**. Blots were counter stained for MRE11. (A) A representative western blot of MRE11 immunoprecipitations blotted for RAD50 and counterstained for MRE11; (B) Quantification of the amount the RAD50 co-immunoprecipitated from the whole cell lysate with MRE11, expressed as a percentage of the 5% load (n=3); (C) A representative western blot of MRE11 immunoprecipitations blotted for NBS1 and counterstained for MRE11; (D) Quantification of the amount of the NBS1 co-immunoprecipitated from the whole cell lysate with MRE11, expressed as a percentage of the 5% load (n=3).

5.3 Discussion

The results from patient genotyping and isoform analysis were in agreement with the conclusions of Teo *et al* (2014). However, in this assessment of *MRE11* isoform expression, higher levels of *MRE11* isoform 1 were seen in samples compared to Teo *et al* (2014). Indeed, for GG samples 11 out of the 14 cases contained 100% *MRE11* isoform 1, whereas in Teo *et al* (2014) *MRE11* isoform 1 was only 61.7% of the total MRE11 in GG samples. The reason for this discrepancy in results is likely attributable to the use of alternative PCR primers, with fewer off target hits, used in the amplification of the *MRE11* isoforms in this study.

Four samples in this analysis did not follow the association for high levels of *MRE11* isoform 2 and the presence of the rs1805363 A minor allele in germline genotype. However, the results of the tumour genotyping on one of these samples might explain why these outlying results. For sample S0521, germline genotyped as AG, no *MRE11* isoform 2 expression was seen. However, tumour genotyping of this sample revealed that within the patient's tumour a reversion of genotype had occurred to GG. This result might suggest that the genotype of a patient's tumour is more important than germline genotype when determining *MRE11* isoform expression within the tumour. The next step in investigating this hypothesis would be to try and obtain the tumour-extracted DNA for the three GG samples that exhibited anomalously high levels of *MRE11* isoform 2 and to genotype these tumours. If these samples had become positive for the rs1805363 A minor allele in the tumour, then there would be further evidence for the genotype of the tumour being more influential on *MRE11* isoform expression then germline genotype.

Unfortunately no homozygous germline rs1805363 cases were identified so the gene dosage hypothesis for *MRE11* rs1805363 could not be assessed. One germline AG case with significantly higher levels of *MRE11* isoform 2 compared to all others was identified and indeed this had undergone LOH in the tumour to become AA for rs1805363. However, LOH to a homozygous AA genotype had occurred in an additional case and this did not exhibit significantly raised levels of *MRE11* isoform 2 compared to other AG rs1805363 samples.

Teo *et al* (2014) found no difference in MRE11 protein level in rs1805363 patients. However, that does not exclude the possibility that the protein level do occur due to the alternative splicing of the *MRE11* transcript. Alternative splicing has been linked to increases in translation efficiency, cytoplasmic trafficking and mRNA turnover (Mockenhaupt & Makeyev, 2015).

Stable re-expression of either full length MRE11 or MRE11 lacking exon 16, in the MRE11 knockdown cells was achieved. Although, the level of MRE11 re-expression in the MRE11 full length and MRE11 exon16-ve clones, was not up to that of the parental T24 cell line. It is, therefore, possible that these clones might have an impairment in the DNA damage repair pathways utilising MRE11. There was no significant difference in results from the clonogenic and comet assays between the MRE11 full length and scrambled control cell line however. This implies that a sufficient level of MRE11 was re-expressed in these clones to reconstitute MRE11-dependent DNA damage repair.

Results from clonogenic survival and comet assays revealed an increase in radiosensitivity and a delay in DNA repair, at 30 minutes post irradiation, in MRE11 exon16-ve clones compared to MRE11 full length clones. It was notable that MRE11

exon16-ve cells showed an equally poor survival in clonogenic assays to MRE11 knockdown cells lines while MRE11 full length cells had comparable survival to the MRE11 scrambled control cell line. Such a result indicates a functional defect in MRE11 lacking exon 16, increasing the radiosensitivity of the MRE11 exon16-ve cell line. Evidence from comet assays added weight to the clonogenic assay findings. Not only was a delay in repair suggested in MRE11 exon16-ve clones but also an accumulation of endogenous DNA damage. An increase in endogenous DNA damage is likely to result in higher levels of genomic instability leading to an accumulation of mutations that could result in a more aggressive or resistant tumour.

It was surprising that the MRE11 knockdown cell line did not show significant repair defects in the comet assays. However, the reduced number of cells available for harvesting in radiation-treated samples compared to untreated MRE11 knockdown samples could indicate that there was a high level of cell death in this cell line post IR. The main mechanisms by which radiation induces cell death are apoptosis and mitotic catastrophe (Eriksson & Stigbrand, 2010). The comet assay is able to detect the earliest signs of apoptotic cell death, but in isolation, the data from a comet assay cannot accurately be used to identify the prevalence of apoptosis in a cell population (Lorenzo *et al*, 2013). Therefore, to investigate whether apoptotic cell death was responsible for the lack of a radiation repair defect in MRE11 knockdown cells, it would be beneficial to conduct apoptosis assays, such as the flow cytometry based assays for annexin V and propidium iodide, on the cell lines used in the comet assay.

An increased radiosensitivity in the MRE11 exon16-ve cell line was counter to the initial hypothesis based on the findings of Teo *et al* (2014). In this current work, MRE11 re-

expression clones contained either an *MRE11* isoform 1 (MRE11 full length) or *MRE11* isoform 2 (MRE11 exon16-ve) construct. This situation is a simplified view of how *MRE11* is expressed in a cell and potentially is not representative of the complex interactions made by MRE11.

MRE11 exists as a homodimer and this dimerisation of MRE11 is critical for the correct assembly of the MRN complex and the binding and aligning of DNA ends (Williams *et al*, 2008). It is possible that interaction between and dimerisation of *MRE11* isoform 1 and isoform 2 has an effect on the function of MRE11 and the MRN complex not seen when only one isoform is present. Hence, it could be that the decrease in radioresponsiveness seen in rs1805363 samples reported in Teo *el al* (2014) is a result of both *MRE11* isoform 1 and isoform 2 being present in the cell. In a situation where only one isoform of *MRE11* is present, an MRE11 dimer can only contain one isoform. However, in an environment where multiple *MRE11* isoforms are present, different forms of MRE11 dimer could occur with varying levels of function. An MRE11 dimer containing two copies of MRE11 isoform 2 may lead to a heavy defect in DNA damage processing, causing high levels of basal DNA damage and increased radiosensitivity. However, it is feasible that a dimer consisting of *MRE11* isoform 1 and isoform 2 might only be partially defective and have a dominant negative effect, leading to increased mutability and possible radioresistance.

A limitation of the work reported here is that it was performed in only one cell line. Cell lines often differ from primary cells in phenotype, native function and responsiveness to stimuli. Additionally, passaging of cells can further perpetuate genetic differences between cell lines and primary cells (Kaur & Dufour, 2012). Validating findings in more

than one cell line could help in confirming these results. The best scenario for validating our findings with MRE11 re-expression clones would be in a primary MIBC cell line positive for the rs1805363 A minor allele. However, deriving primary cell lines is a difficult and lengthy process and not always possible.

Another limitation of this work is that it was performed in MRE11 knockdown cells rather than knockout cells. In an ideal world, a CRISPR-generated knockout cell line would have been used. However, knockout of MRE11 proved to be lethal to cells and hence a cell line could not be generated. MRE11 not only functions in DNA damage repair but is also involved DNA replication and as such is highly important to cellular survival.

The inability to immunoprecipitate MRE11 lacking exon 16 using the HA tag incorporated into the MRE11 exon16-ve re-expression vector might provide rudimentary evidence for a structural change in MRE11 lacking exon 16. This is an interesting finding as a conformational shifts are critical for the full functioning of, not only, MRE11 but the whole of the MRN complex. A change in the structure of MRE11 may impede the large conformational alterations required within the MRN complex and could explain the radiosensitivity of MRE11 exon16-ve cells. Further investigation is required to establish if the absence of exon 16 confers a structural alteration in MRE11.

MRE11 re-expression clones both showed methylation in MRE11 immunoprecipitations blotted for ADMA. This suggests that exon 16 does not have an inhibitory effect on the methylation of the MRE11 GAR domain. However, it is possible that the level of methylation of MRE11 may differ between MRE11 full length and MRE11 exon16-ve proteins but is beyond the sensitivity of this assay. To further investigate the

methylation status of MRE11 lacking exon 16, methylation assays using purified PRMT1, the protein arginine methyltransferase responsible for the methylation of the MRE11 GAR domain, and MRE11, plus or minus exon 16, should be conducted. Investigation of MRE11 methylation with mass spectrometry, as reported in Bremang *et al* (2013) would also be highly informative. In addition to this, immunoprecipitations with PRMT1 and analysis of DNA damage repair in cells treated with methyl-thio-adenosine methylase inhibitor might elicit further information on the methylation of MRE11 lacking exon 16.

The formation of the MRN complex is critical for the function of MRE11. Therefore, it was possible that the radiosensitivity seen in the MRE11 exon16-ve clone was due to an inability of MRE11 lacking exon 16 to interact with RAD50 and/or NBS1. RAD50 and NBS1 were found to co-immunoprecipitate with MRE11 extracted from MRE11 full length and MRE11 exon16-ve lysates, indicating that MRE11 lacking exon could bind the members of the MRN complex.

Although MRE11 minus exon 16 was found to bind members of the MRN complex, the stability of the interactions was not assessed. It is possible that, with exon 16-negative MRE11, the stability of the interactions in the MRN complex is compromised. Conducting immunoprecipitations with high ionic strength buffers and thermal stability shift assays would address this question. Furthermore, MRE11 interacts with a large number of DDR proteins, such as RPA, FANCJ and CDK2, affecting their recruitment to sites of DNA damage and/or activity. It would, therefore, be useful to establish if there is any defect in MRE11 lacking exon 16 to bind or recruit these factors.

Another area to investigate with regard to the MRN complex is the hydrolysis of ATP by RAD50. This reaction is highly important in the functioning of the MRN complex and

might also be significant here. If the interaction of MRE11 with RAD50 is not sufficiently stable for ATP hydrolysis, both nuclease activity of MRE11 and DNA binding capability are compromised.

With no evidence for an alterations in the MRN complex or MRE11 methylation, it is still unknown why a lack of MRE11 exon 16 conferred radiosensitivity to the cells analysed. Further work is required to fully understand this effect. Although MRE11 exon 16 is not associated with DNA binding or nuclease domains of MRE11. It is possible that loss of exon 16 may still impair these functions. With this in mind it would be interesting to assess the DNA binding capability of MRE11 lacking exon 16 with an electrophoretic mobility shift assay and evaluate its endonuclease and exonuclease activity using endonuclease and exonuclease assays as described in Shibata *et al* (2014). Another avenue of investigation would be to analyse the downstream effects of MRE11 minus exon 16, such as the activation of ATM.

There is a significant lack of research into the role, function and expression of *MRE11* isoform 2. However, our results from an MRE11 exon 16 negative plasmid indicate that perhaps *MRE11* isoform 2 is not as benign as currently assumed. While this work is preliminary, fully understanding the function and interactions of *MRE11* isoform 2 in the processing of DNA damage and stalled replication forks is an interesting novel avenue of research and a potential target and/or biomarker for cancer therapy.

Chapter 6 The utilisation of crowdsourcing as a novel approach in the discovery and screening of biomarkers in MIBC

6.1 Introduction

Using IHC to evaluate potential protein biomarkers in cancer can provide valuable data for the development of personalised medicine. However, analysis of IHC staining, by the current gold standard method of manual scoring, can be prohibitively time consuming. The delay in obtaining the results of IHC due to the use of manual scoring may impede treatment relevant data entering the public domain. Although automated analysis has had some success in resolving this issue, it has not been widely accepted and requires manual intervention. Therefore, until the issues associated with automated analysis are resolved, it is worthwhile considering alternative methods of manual scoring that could reduce the time burden imposed upon the researcher.

In this work a mobile gaming application was developed with the aim of investigating if crowdsourcing could increase the speed of scoring IHC, whilst maintaining a high level of scoring. The accuracy of the crowdsourced scores, for a range of IHC markers, was determined and the scores were used in survival analyses.

Scoring IHC data in its most basic form is a task involving pattern recognition and determination of colour gradients. It was therefore hypothesised, that given a short tutorial, members of the public would be able to accurately assess the IHC staining of cancerous tissue. Due to the number of potential individuals free to analyse IHC using crowdsourcing, it was also theorised that crowdsourcing would be able to increase the speed of scoring large IHC datasets.

Advances in scientific techniques have generated high volumes of data requiring analysis. With the use of TMAs and automated IHC procedures, it is possible to generate a large volume of IHC data on sizeable sample cohorts relatively quickly. Such advances increase the efficiency and practicality of using IHC to screen potential biomarkers and permit the investigation of multiple proteins on consecutive tissue sections in parallel.

However, despite these advances the analysis of IHC stained tissue is still, in the majority of cases, conducted by the naked eye. This type of scoring is time consuming and requires two or three trained researchers or histopathologists to independently assess tissue staining before a consensus score for a patient can be reached.

Software for the automated analysis of IHC data exists but has not been widely adopted and has limitations. The basic and most widely available automated IHC scoring algorithms are currently not applicable to all samples and changes in cellular morphology or a high prevalence of infiltrative immune cells can cause errors in scoring. For these challenging or ambiguous cases, manual intervention is required (Theodosiou *et al*, 2007; Di Cataldo *et al*, 2012; Levenson *et al*, 2015). Machine learning algorithms for IHC have the potential to solve the need for manual intervention in automated image analysis. However, to develop these programmes large datasets of scored tissue are required for the training of the algorithm. Furthermore, to increase confidence in the use of automated IHC scoring, a large scale validation of the technique is required (Theodosiou *et al*, 2007). As a consequence the current scoring of IHC is a major bottleneck and limits the use and efficiency of this technique in large-scale studies.

Crowdsourcing has emerged as a potential solution to the problem of 'big data' and has been employed in the collection and analysis of scientific material. A number of projects

have been conducted within a range of scientific disciplines, including biochemistry and biomedical research, which have reported success when using crowdsourcing to analyse samples (Lintott *et al*, 2008; Cooper *et al*, 2010; Kawrykow *et al*, 2012; Lee *et al*, 2014; Candido Dos Reis *et al*, 2015; Rallapalli *et al,* 2015; Lawson *et al*, 2017).

The advantage of using the public to score IHC over an image analysis programme, lies in their innate ability to recognise patterns and shapes and their ability to learn during the course of performing a task (DiCarlo *et al*, 2012). Unlike deep learning programmes, humans do not need large training datasets to distinguish between cellular morphologies and identify individual cells. Furthermore, the criteria used to programme non-neural network the image analysis software can generate bias (Lintott *et al*, 2008). Crowdsourcing relies on the collection of data from a large number of individuals to generate accuracy and therefore is potentially less prone to bias.

The disadvantage of using crowdsourcing for the analysis of IHC, is that the accuracy of results relies on obtaining a large number of individuals to participate (Lawson *et al*, 2017). Therefore, a successful crowdsourcing venture needs to be able to recruit individuals to perform in the study. Additionally, crowdsourcing assumes that the task given to the public is not beyond their capability and that cognitive diversity will ensure reliable results (Candido Dos Reis *et al*, 2015).

As with all methodologies the crowdsourcing of IHC has advantages and disadvantages. However, the results of previous studies have found this method of analysing data to be scientifically valid (Lintott *et al*, 2008; Cooper *et al*, 2010; Kawrykow *et al*, 2012; Lee *et al*, 2014; (Khatib *et al*, 2011).Ultimately, if the results of publicly generated IHC scores are found to be consistently concordant with those produced by experts then this is of

scientific use. It can be implemented immediately and used to generate the data required for the development of machine learning digital analysis programmes.

Cell Slider was a crowdsourcing project aimed at addressing the rate-limiting step of IHC manual scoring. It was demonstrated that untrained members of the public could accurately score IHC data, with participants achieving similar results to trained pathologists in cancer cell identification, ER status assignment and associations of ER status with clinical outcome in breast cancer tissue (Candido Dos Reis *et al*, 2015). However, the participants demonstrated a bias for the overestimation of the number of cancer cells in an image, thus compromising the accuracy of IHC scoring (Candido Dos Reis *et al*, 2015).

Trailblazer was developed in the wake of Cell Slider and was aimed at identifying improvements that could be made to the crowdsourcing of IHC, which might increase the accuracy of publicly generated scores. It was hypothesised that the public overestimation of the number of cancer cells in an image, identified by Cell Slider, was due an insufficient level of instruction in the tutorials provided to users and a restrictive interface showing only a small portion of a TMA core. To test this theory Trailblazer trialled three types of tutorial and showed users a whole TMA core, where segments could be zoomed into. The tutorials consisted of: a basic tutorial (comprised of text and images), an annotated tutorial (containing annotated images present throughout participation), and a feedback-based tutorial (consisting of a training set of five images to be scored before commencement of real data analysis). It was found that when given more comprehensive tutorials, the public was highly accurate in their estimation of the number of cancer cells in an image (Lawson *et al*, 2017).

One of the problems in using crowdsourcing to analyse scientific data, is the rate of drop-off and inactivity in user participation. Recent crowdsourcing projects have therefore tried to broaden the net of users participating by integrating scientific tasks into games. Examples of such ventures are Foldit (Cooper *et al*, 2010), Phylo (Kawrykow *et al*, 2012), EteRNA (Lee *et al*, 2014) and Fraxinus (Rallapalli *et al*, 2015). Results from these crowdsourcing games have been positive, finding *bona fide* scientific discoveries and generating improvements to existing computational algorithms (Curtis, 2014). Remarkably, in Foldit, two teams of players were able to generate a model for the folding of a protein in three weeks. This model was of sufficient quality for a successful crystal structure of the protein to be defined, a task that, for this protein, had proved elusive for years before the input of Foldit users (Khatib *et al*, 2011b).

In our investigations into MIBC, a number of proteins have been highlighted as being worthy of assessment for potential association with clinical outcome. IHC was conducted for proteins of interest on TMAs containing tissue from tumours resected from MIBC patients. This generated a large amount of data which needed to be analysed. The positive results reported from Cell slider and other crowdsourcing projects were encouraging, and so we decided to employ this resource to analyse our IHC datasets.

Unlike previous crowdsourcing efforts in IHC scoring, which were website based, we input our IHC data into a mobile gaming app (Reverse the Odds) available to members of the public. We first assessed the accuracy of the crowdsourced work. Then, when the crowdsourced scores were found to be accurate, used these to look for associations between protein staining and clinical outcome.

6.2 Results

6.2.1 Immunohistochemistry of MIBC TMAs

Fourteen antibodies were successfully optimised for the IHC staining of seven MIBC TMAs (Figure 6.1). TMAs contained 1356 tissue cores extracted from tumour sections resected from 273 MIBC patients treated at Leeds Teaching Hospitals between 2006 and 2009. The presence of diathermy in tissue cores, absence of cores from cut sections and presence of partial cores with insufficient tissue reduced the number of usable cores in our analysis. The maximum number of useable cores available for the analysis of an IHC stain ranged from 376-850.

The antibodies used in IHC staining encompassed a range of proteins involved in the DDR (MRE11, MRE11 C-terminal, RAD50, NBS1, p21, p53, 53BP1, RPA, Tip60, HDAC2, HDAC4), bladder cancer subgroup classification (cytokeratin 5/6 (CK5/6), cytokeratin 20 (CK20) and cellular proliferation (Ki67). An additional re-optimised staining of MRE11 was conducted after the initial MRE11 staining to reduce the non-specific background staining of tissue and eliminate the saturation of staining intensity (see chapter 3.2.2).

The specificity of all the antibodies except Ki67 were assessed with western blot analysis on a panel of bladder cancer cell lines (Figure 6.2). The Ki67 antibody was not evaluated in this way as it is a highly established IHC antibody that has previously undergone rigorous validation in collaboration with pathologists.

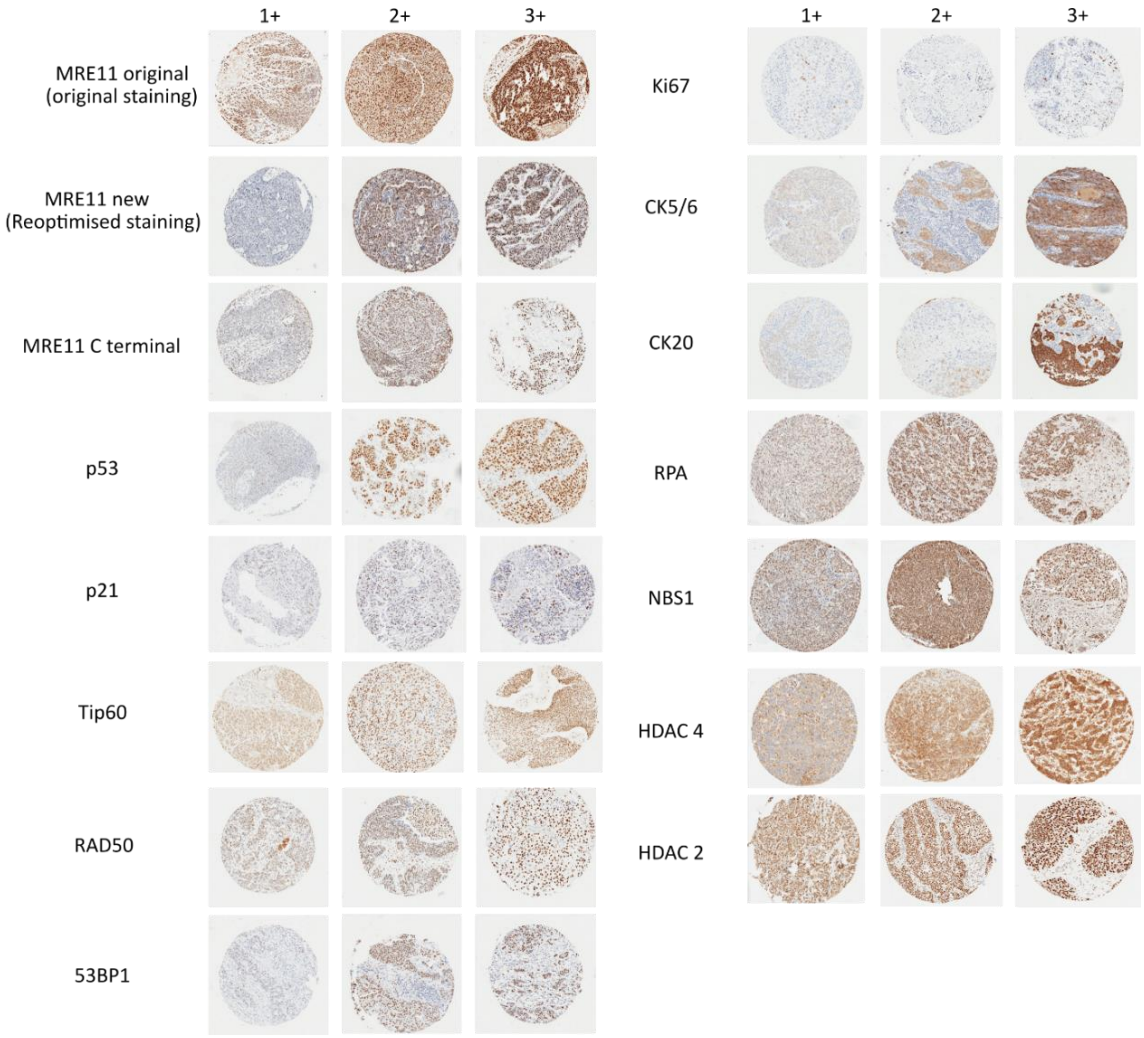


Figure 6.1: **Representative images of optimised IHC staining for each of the antibodies used to stain MIBC tissue cores**. Weak, moderate and strong staining intensity is represented by the values 0, 1+, 2+ and 3+ respectively.

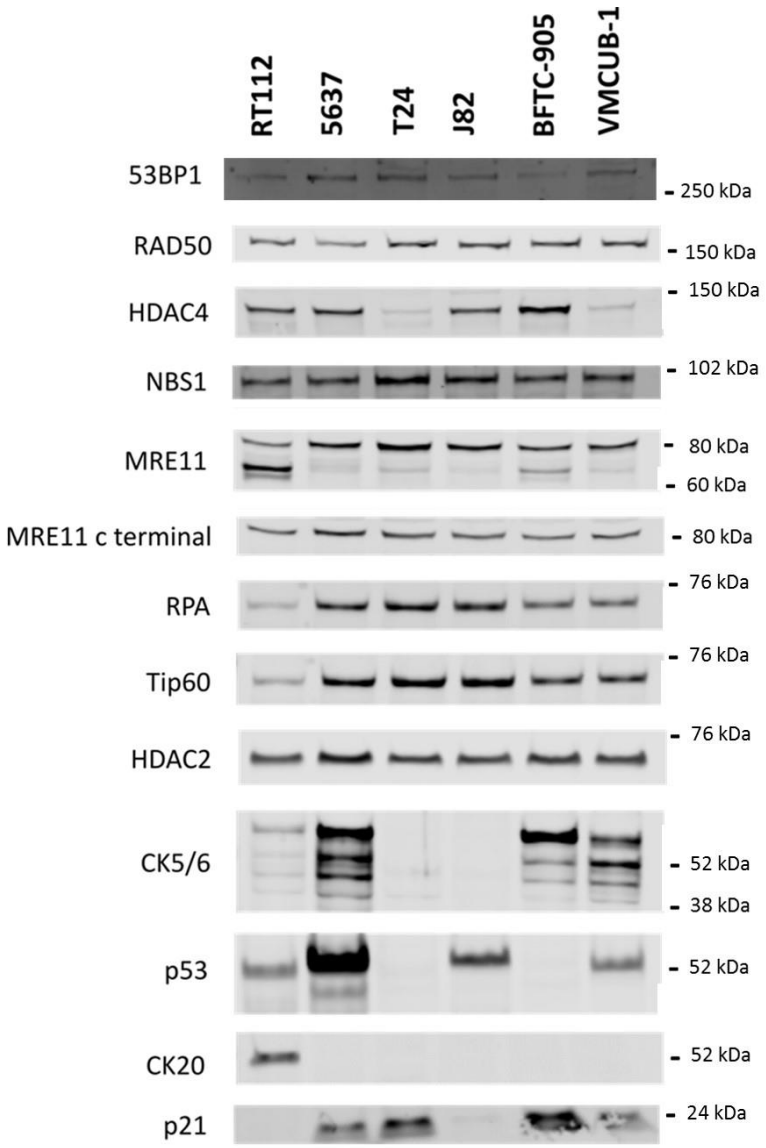


Figure 6.2: **Western blots for antibody specificity, conducted on six bladder cancer cell lines: RT112, 5637, T24, J82, BFTC-905, VMCUB-1**. Images are cropped but show multiple bands where present.

6.2.2 Reverse the Odds

Due to the large amount of IHC data to be analysed we collaborated with CRUK, the Zooinverse citizen science project and CHUNK game developers to generate a mobile gaming app called Reverse the Odds (RTO). The app gave untrained members of the public the task of scoring IHC data before they were able to play the game-based feature of the app.

High resolution images of stained TMA cores were colour transformed to make them more appealing and split into 36 segments. The public were given one of the 36 segments to score at any one time to allow for comfortable viewing of individual cells on a smartphone without zooming. A short tutorial was produced and placed into the app to educate members of the public on how to recognise which cells within an image to score, how to score the number of positively stained cells and the intensity of staining (Figure 6.3). The app was released to members of the public in October 2014 and ceased operation in September 2016.

During the period of the app's operation, the crowdsourced scores were collated so that they could later be tested for accuracy and used in survival analyses.

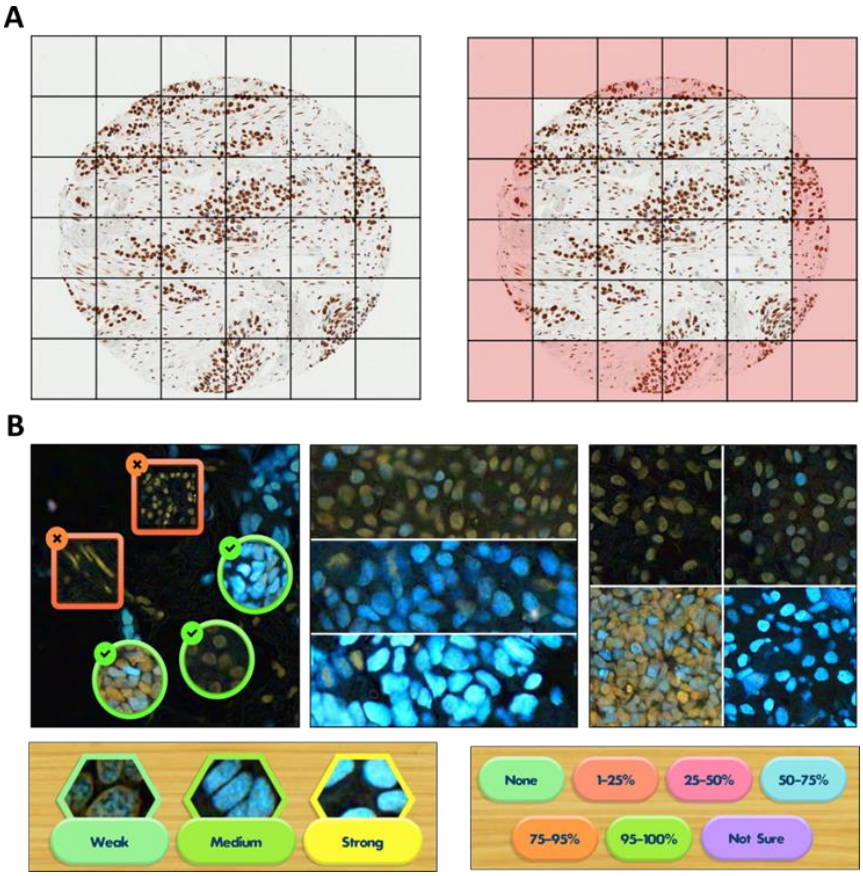


Figure 6.3: **The segmentation of the TMA cores input into RTO and example images of the IHC scoring tutorial supplied to users of the game**. (A) Representative images describing the segmentation of a TMA core. The image to the left shows a core segmented into 32 segments using a 6x6 grid. The image to the right depicts the outer 16 segments of the grid which were excluded from analysis after testing the accuracy of using only the central 16 segments in scoring; (B) An example of the tutorial given to RTO users to aid scoring.

During the course of the app's operation, a test was conducted to analyse the effect of excluding the 20 outermost segments of the TMA core from the analysis (Figure 6.3A). These outer segments of TMA cores often contained little tissue with few cells. It was hypothesised that the inclusion of these segments could skew scores generated by the public and their exclusion might potentially improve the scoring accuracy of the total TMA core or have no effect. Results obtained from the test on segmentation found that

accuracy of public scores was maintained when the outer 20 segments of the core image were excluded. Due to this finding, the later scoring of TMA cores was conducted on the central 16 segments of a core image. This also had the added benefit of increasing the speed at which a TMA cores could be analysed, as under half the number of images were required to achieve a total score for a core.

In total, RTO was able to analyse the data from 11 of our immunostains on MIBC TMAs in addition to classifying three immunostains on four sets of non-small cell lung cancer tissue (input into the game by Prof Gareth Thomas, University of Southampton). A table summarising the MIBC IHC stains analysed by RTO, with core numbers and numbers of classifications per core, can be found in appendix H. The data presented in this chapter will focus solely on the results from RTO on our 11 stains on MIBC TMAs. For our 11 stains, 3,349,937 images were analysed by members of the public relating to 8297 patient cores stained for each of the 11 antibodies.

6.2.3 Comparison of crowdsourced IHC scoring to 'expert' IHC Scoring

To assess the accuracy of crowdsourced scoring, the aggregated public scores for 10-21% of the tissue cores for each of the immunostains were compared to 'expert' scores generated by trained scientific researchers. Expert scoring was conducted using a standard method of manual scoring, ie. the individual assessment of the percentage of cancer cells stained (% positivity) and the intensity of staining with a final consensus agreement. The scores for the % of cancer cells stained and intensity were then combined into an H-score= 1[1(% cells 1+) + 2(% cells 2+) + 3(% cells 3) (Figure 6.4, Table 6.1). The primary outcome of this IHC analysis was H-score. However, examining the

accuracy of the public to score staining intensity and the % positivity is instructive, as it allows consistent biases in the scoring of these factors to be identified.

A range of accuracies for the scoring of % positivity and intensity were observed. In the scoring of % positivity the public had an accuracy that ranged from 0.17 to 0.87, for the scoring of intensity the range of scoring accuracy was from 0.19 to 0.86 (Table 6.1). It was also noted that the public achieved the lowest scoring accuracy for IHC stains that experts had classed as more challenging to score (TIP60 and MRE11).

There was no bias for the public to more accurately score intensity over % positivity or % positivity over intensity. For example, the % positivity of RAD50 was scored by the public with a low accuracy (Spearman correlation: 0.39) but the intensity scoring for this stain was high in accuracy (quadratic-weighted Kappa: 0.73). Conversely the intensity of staining for Ki67 was poorly scored by the public (quadratic-weighted Kappa: 0.19) but the accuracy of % positivity scoring was high (0.80).

RTO participants were no better at analysing nuclear IHC stains compared to membranous staining. Spearman r values for public vs expert scores of membranous stains CK5/6 and CK20 were within the range of Spearman r values calculated for nuclear stains (Table 6.1).

Table 6.1: **The accuracy of crowdsourced scoring for each of the markers scored by RTO**. The correlation coefficients between public and expert scores for each of the stains score by RTO are shown with 95% confidence intervals. Spearman correlation was used for the analysis of H scoring and % positivity. Quadratic-weight kappa analysis was conducted on intensity scores. Correlation coefficients between 1.0-0.9, 0.89-0.7, 0.69-0.5 or 0.49-0.3 were classed as very high, high, moderate and low in accuracy. Statistics performed by Dr Peter Smittenaar.

	Percentage of cores scored by expert	H-score, Spearman correlation with experts (95% CI)	% positivity, Spearman correlation with experts (95% CI)	Intensity of staining, Quadratic-weighted kappa with experts (95% CI)
MRE11	11	0.67 (0.52, 0.80)	0.44 (0.22, 0.63)	0.47 (0.30, 0.61)
RAD50	13	0.81 (0.71, 0.88)	0.39 (0.19, 0.57)	0.73 (0.56, 0.85)
p21	11	0.90 (0.84, 0.93)	0.87 (0.80, 0.91)	0.57 (0.39, 0.72)
53BP1	10	0.70 (0.55, 0.80)	0.53 (0.34, 0.68)	0.67 (0.52, 0.79)
p53	11	0.92 (0.86, 0.95)	0.85 (0.77, 0.91)	0.69 (0.57, 0.79)
CK5/6	13	0.82 (0.71, 0.89)	0.66 (0.50, 0.78)	0.86 (0.77, 0.92)
CK20	12	0.88 (0.83, 0.92)	0.83 (0.75, 0.88)	0.87 (0.79, 0.92)
Tip60	11	0.66 (0.52, 0.76)	0.17 (-0.07, 0.39)	0.59 (0.46, 0.70)
MRE11 re-optimised	16	0.65 (0.49, 0.78)	0.61 (0.44, 0.75)	0.58 (0.41, 0.72)
MRE11 C-terminal	21	0.79 (0.66, 0.86)	0.66 (0.49, 0.78)	0.70 (0.55, 0.80)
Ki67	11	0.80 (0.72, 0.86)	0.80 (0.71, 0.87)	0.19 (-0.01, 0.37)

Expert scores were generated by the independent scoring of stained tissue by two experienced scorers. In cores where independent scorers did not agree, a third scorer was used to reach a consensus score for the core. An analysis into the agreement between the independent scores generated by expert scorers was conducted on a subset of the tissue stains. This was to establish if a low accuracy in crowdsourced scoring was related to a low concordance in the scoring between experts (Table 6.2). The accuracy of crowdsourced scores was not related to the concordance in scoring between experts.

Table 6.2: A comparison between the accuracy of crowdsourced scores and the concordance of scoring between experts. Data is shown for 6 of the 11 IHC stains scored by RTO. The statistics for public scoring accuracy were produced by Dr Peter Smittenaar. The statistics for the level of agreement between experts were produced by Alexandra Walker. Correlation coefficients between 1.0-0.9, 0.89-0.7, 0.69-0.5 or 0.49-0.3 were classed as very high, high, moderate and low in accuracy and are highlighted in green, yellow, orange and red, respectively. Correlation coefficients between 0.29-0 were classed as inaccurate and are highlighted in blue.

	Public scoring vs consesus expert score		Expert 1 scores vs Expert 2 scores	
	% positivity (Spearman correlation)	Intensity of staining (Quadratic-weighted kappa)	% positivity (Spearman correlation)	Intensity of staining (Kappa)
53BP1	0.53	0.67	0.8	0.7
CK5/6	0.66	0.86	0.73	0.71
MRE11 re-optimised	0.61	0.58	0.78	0.68
Ki67	0.8	0.19	0.72	0.45
CK20	0.83	0.87	0.79	0.57
TIP60	0.17	0.59	0.9	0.71

1.0-0.9	Very high level of accuracy
0.89-0.7	High level of accuracy
0.69-0.5	Moderate level of accuracy
0.49-0.3	Low level of accuracy
0.29-0	Inaccurate

The accuracy of public scoring was only partly affected by the number of users included in the image analysis (Figure 6.7). The greater the number of classifications per image the higher the agreement between the crowdsourced and expert scores. However, for all stains a plateau was reached whereby the inclusion of more users in the analysis did not increase the accuracy of crowdsourced score. CK20, p53 and p21 reached the highest level of H scoring accuracy, with CK20 achieving the highest accuracy with the least number of users (Figure 6.7). While p53 and p21 achieved comparable levels of accuracy in H score to CK20, a larger number of users was required to reach this point. p21 and p53 required approximately 100 users to achieve maximum accuracy compared to 10 users required for CK20. Tip60, MRE11, MRE11 re-optimised and 53BP1 showed the lowest levels of public accuracy which could not be improved by the addition of more users into the analysis.

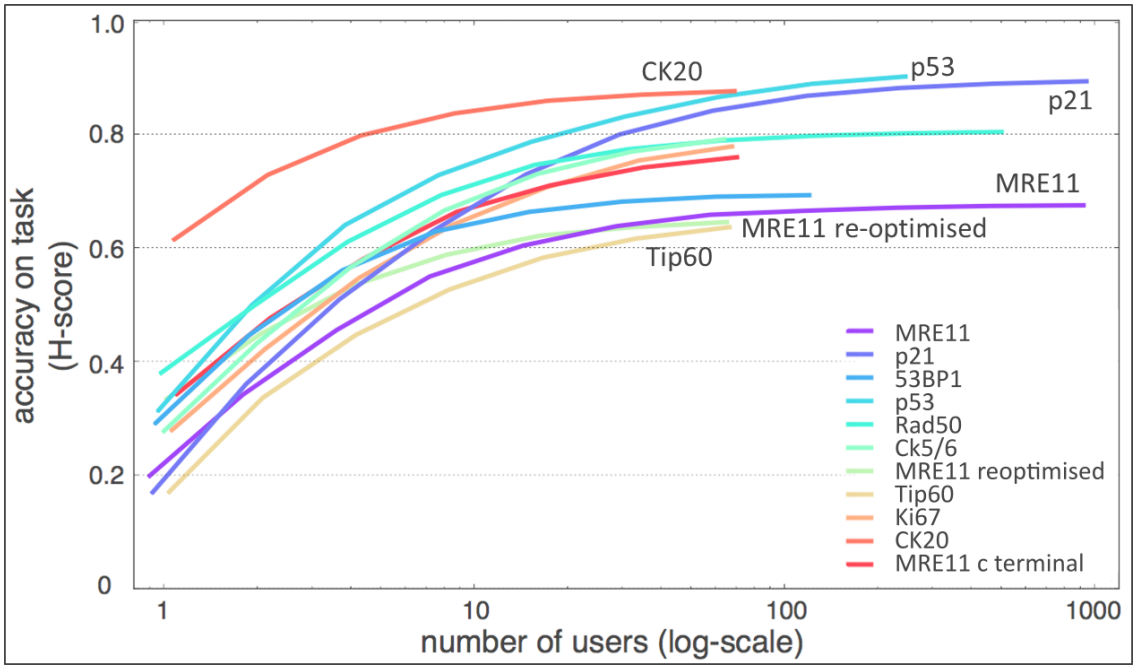


Figure 6.4: **The number of users included in a crowdsourced score plotted against the Spearman r value for H-scores**. IHC stains which achieved the highest and lowest overall accuracy are labelled. Statistics performed by Dr Peter Smittenaar.

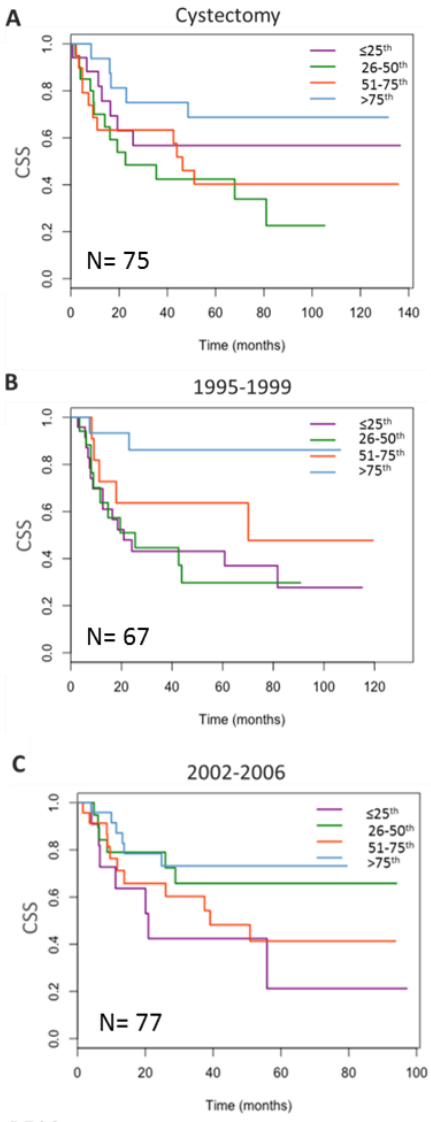
6.2.4 Associations between crowdsourced IHC scoring and clinical outcome

Association of public scores to clinical outcome was performed using Kaplan-Meier survival curves and Cox proportional hazards models. The association between H-score and CSS was assessed using either the quartiles for the H-scores or the numerical unit increase in H-score (Appendices I and J). Four cohorts of patients were scored by RTO, three RT cohorts and one cystectomy cohort. The three RT cohorts were: 1995-1999, 2002-2006 and 2006-2009. It became apparent early on in the analysis that in the 2006-2009 RT cohort a significant number (>2/3rds) of cores were unusable and hence patient scores from this cohort were unlikely to be representative. This TMA had been previously sectioned for other projects and was already quite depleted by the time we started RTO. As discussed in Chapter 3.3, as the depths of a TMA are explored, areas of diathermy are exposed and there is often loss of tumour which reduces the number of cores viable for analysis. Due to the substantial loss of cores from the 2006-2009 cohort, this was excluded from formal analysis of association with clinical outcome.

Univariate Cox proportional hazards analysis was performed for each of the stains scored by RTO. Statistically significant associations between H-score and CSS were found for MRE11, CK20, p21, 53BP1, p53 and Ki67 IHC. With large amounts of multiple testing, it is likely that some significant results were due to chance. However, MRE11 and CK20 displayed consistent results indicative of a *bona fide* association (Appendices H, I and K).

Similar to previously reported findings (Choudhury et al, 2010; Laurberg et al, 2012) high MRE11 levels, from the initial MRE11 IHC staining dataset, were found to be significantly associated with CSS in the RT cohorts but not the cystectomy cohort (Figure 6.5).

Significance was observed in both the 1995-1999 (hr: 0.141, 95% CI: 0.032-0.620, p= 0.009, n=67) and 2002-2006 (hr: 0.283, 95% CI: 0.095- 0.847, p= 0.024, n=75) RT cohorts when comparing the ≤25th percentile of the H-scores to the <75th percentile. Furthermore, when using a numeric H-score, rather than comparing quartiles, a significant association between MRE11 staining and CSS in the 1995-1999 cohort (hr per unit increase in H score: 0.991, 95%CI: 0.986-0.997, p=0.004) and borderline-significant association in the 2002-2006 cohort (hr: 0.994, 95% CI: 0.987-1.000, p=0.060) were identified.



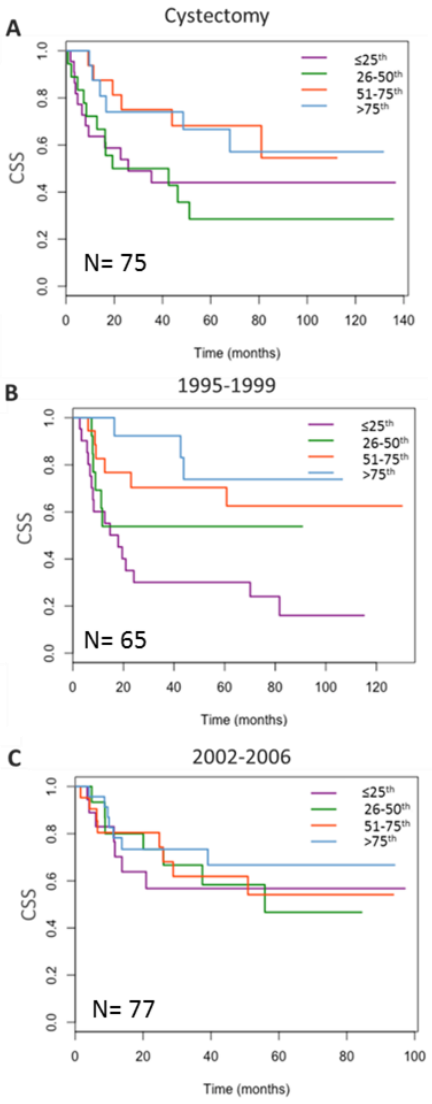
	Cystectomy Hazard ratio [95% confidence interval, p value]
≤25th vs 26-50th percentile	1.808 [0.721- 4.534, 0.207]
≤25th vs 51-75th percentile	1.4124 [0.547- 3.649, 0.476]
≤25th vs >75th percentile	0.5646 [0.1788- 1.783, 0.330]

	1995-1999 RT cohort Hazard ratio, [95% confidence interval, p value]
≤25th vs 26-50th percentile	1.041 [0.477- 2.273, 0.919]
≤25th vs 51-75th percentile	0.550 [0.200- 1.516, 0.248]
≤25th vs >75th percentile	0.141 [0.032- 0.620, 0.009**]

	2002-2006 RT cohort Hazard ratio, [95% confidence interval, p value]
≤25th vs 26-50th percentile	0.374 [0.125- 1.119, 0.079]
≤25th vs 51-75th percentile	0.6814 [0.263- 1.764, 0.429]
≤25th vs >75th percentile	0.283 [0.095- 0.847, 0.024*]

Figure 6.5: **Kaplan-Meier CSS survival curves for MRE11 IHC stratified by ≤25th (purple), 26-50th (green), 51-75th (red) and >75th percentiles (blue)**. (A, B an C) The Kaplan-Meier curves for the cystectomy cohort (A), 1995-1999 RT cohort (B) 2002-2006 RT cohort (C). Statistics from univariate Cox proportional survival analysis for each of the cohorts are displayed in the table to the right of the curve for each cohort. Statistics performed by Dr Christiana Kartsonaki.

For CK20 staining, the 51-75th and >75th percentiles were significantly associated with improved CSS in the 1995-1999 cohort but did not significantly associate with outcome in the 2002-2006 or the cystectomy cohort (Figure 6.6). Compared to the ≤25th percentile the hr for the 51-75th percentile was 0.291 (95% CI: 0.113-0.746, p=0.01, n=65) and >75th percentile was 0.1687 (95% CI: 0.049-0.581, p=0.005, n=65). Additionally, analysis of scores using per unit increase in score rather than quartiles identified a significant association between CK20 levels and CSS in the cystectomy (hr: 0.996, 95% CI: 0.995-1.00, p=0.05, n=75) and 1995-1999 cohorts (hr: 0.992, 95% CI: 0.988-0.997, p=0.002).



	Cystectomy Hazard ratio [95% confidence interval, p value]
≤25th vs 26-50th percentile	1.209 [0.542- 2.695, 0.643]
≤25th vs 51-75th percentile	0.4941 [0.185- 1.319, 0.159]
≤25th vs >75th percentile	0.4997 [0.187- 1.333, 0.166]

	1995-1999 RT cohort Hazard ratio, [95% confidence interval, p value]
≤25th vs 26-50th percentile	0.521 [0.203- 1.339, 0.176]
≤25th vs 51-75th percentile	0.291 [0.114- 0.746, **0.010***]
≤25th vs >75th percentile	0.169 [0.049- 0.581, **0.005****]

	2002-2006 RT cohort Hazard ratio, [95% confidence interval, p value]
≤25th vs 26-50th percentile	0.982 [0.344- 2.805, 0.973]
≤25th vs 51-75th percentile	0.921 [0.333- 2.543, 0.873]
≤25th vs >75th percentile	0.686 [0.241- 1.958, 0.482]

Figure 6.6*:* **Kaplan-Meier CSS survival curves for CK20 IHC stratified by ≤25th (purple), 26-50th (green), 51-75th (red) and >75th percentiles (blue)**. (A, B and C) The Kaplan-Meier curves for the cystectomy cohort (A), 1995-1999 RT cohort (B) 2002-2006 RT cohort (C). Statistics from univariate Cox proportional survival analysis for each of the cohorts are displayed in the table to the right of the curve for each cohort. Statistics performed by Dr Christiana Kartsonaki.

The curves in the Kaplan-Meier plots for the cystectomy and 1995-1999 cohort show a grouping of the ≤25th and 25-50th percentiles and the 51-75th and >75th percentiles, ie. a split at the median value (Figure 6.11A and B). Due to this, a Cox proportional hazards analysis was conducted on CK20 scores grouped according to ≤ median score and > median score. CK20 levels above the median H-score were associated with improved survival in the cystectomy (hr: 0.454, 95% CI: 0.227-0.909, p: 0.026, n=75) and the 1995-1999 (hr: 0.292, 95% CI: 0.134-0.638, p: 0.002, n=65) cohorts.

6.2.5 Correlations between IHC stains used in analysis

The scoring generated by the public was used to look for correlations between each of the IHC stains (Figure 6.7). Identifying correlations between variables is a useful method of identifying potential interactions in the dataset. A correlation between stains would suggest an association between the two factors, which could be of biological interest or useful for the combining of stains into an IHC panel. There were no strong positive or negative correlations between the IHC stains for patient scores generated by the public.

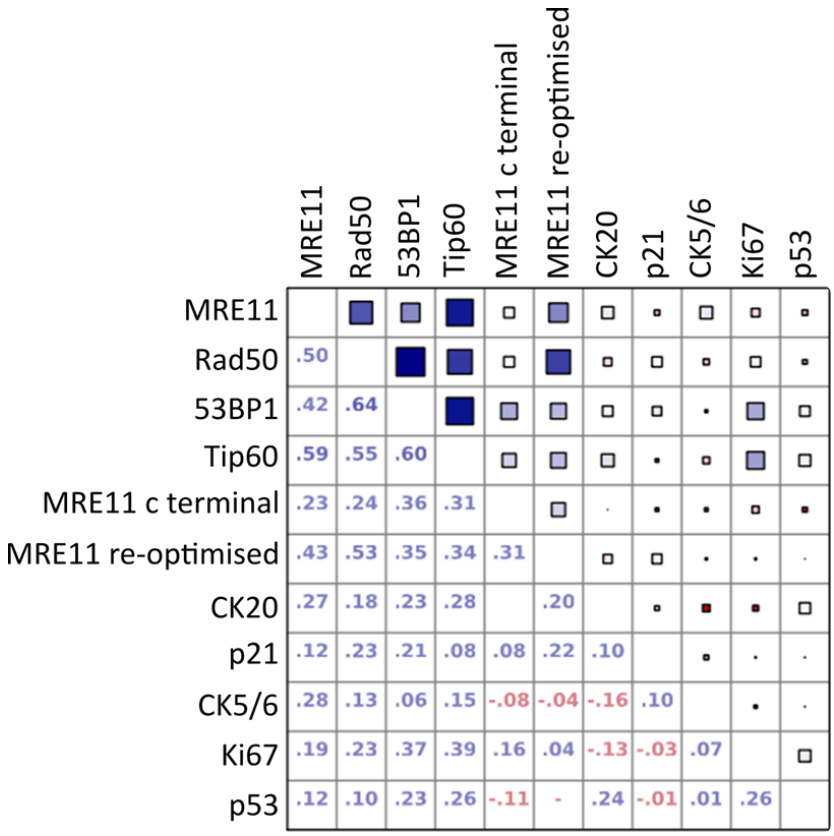


Figure 6.7: A Correlation matrix for the IHC stains scored by RTO. The correlation coefficient between the scores for each of the staining included in RTO is plotted. Negative values indicate a negative correlation between stains. Positive values indicate a positive correlation between stains. The strength of a correlation was determined as: very high if the correlation coefficient was between 1.0-0.9, high (correlation coefficient: 0.9-0.7), moderate (correlation coefficient: 0.7-0.5), low (correlation coefficient: 0.5-0.3) or negligible (0.3-0). Statistics performed by Dr Christiana Kartsonaki.

.

6.3 Discussion

RTO was aimed at improving the speed of IHC scoring. Modern games technology was combined with crowdsourcing to bring citizen science to a wider user-base than other projects such as CRUK's Cell Slider (Candido Dos Reis *et al*, 2015) and Trailblazer (Lawson *et al*, 2017). Arguably this aim of the study was not achieved. It took just under two years to score eleven out of the fourteen IHC stains, far longer than it would have taken researchers. However, this form of crowdsourcing is still in its infancy and throughout the course of RTO many lessons were learnt, which could improve its usefulness as a tool in scientific research.

Reducing the number of segments to be scored for each core to the 16 central segments did not affect the accuracy of scoring and increased the speed at which a core could be scored by the public. Additionally, during the interim analysis of the project we found that accurate results could be achieved with fewer users than initially thought. With these adjustments it is estimated that RTO could have analysed all 11 stains in the first 2 weeks after release of the game, which would amount to a significant increase in scoring efficiency over traditional scoring methods.

One problem identified in RTO was the drop-off of user participation over time. Indeed, half of all image analysis was conducted in the first 4 months of the game's release (Figure 6.17). This pattern for participation drop off has been observed in other crowdsourcing ventures such as Galaxy Zoo, Milky Way Project, Fraxinus, EteRNA, Foldit and Phylo (Curtis, 2014; Ponciano *et al*, 2014; Rallapalli *et al*, 2015). Reviews of user participation from these previously reported crowdsourcing efforts find that a small

group of dedicated individuals contribute to the bulk of classifications, with the majority of users only contributing transiently, and with some registered users never actually participating (Curtis, 2014; Ponciano *et al*, 2014; Rallapalli *et al*, 2015). To exploit the high uptake and number of analyses conducted in crowdsourcing applications after initial release, good systems need to be in place upfront. Continuous advertising of crowdsourcing applications throughout the duration of their operation is also needed to replace inactive users. It is, therefore, important that a robust user engagement plan for the promotion of a project is in place from the outset.

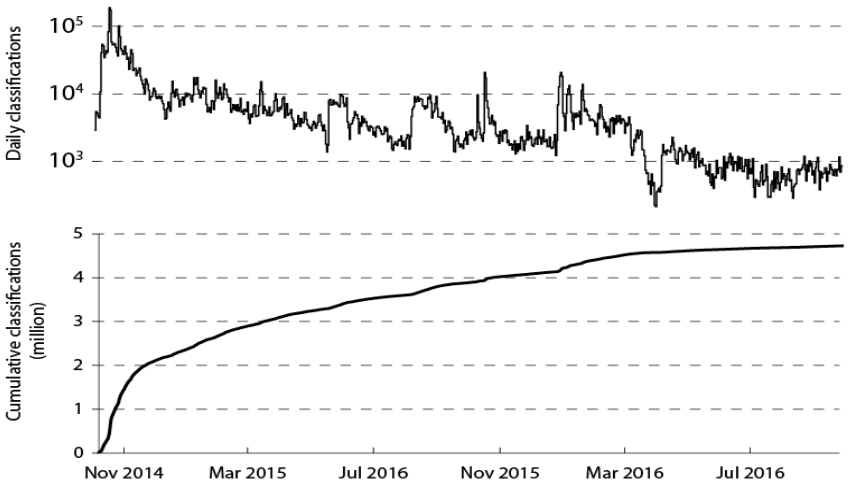


Figure 6.8: **The user participation of RTO over time.** The top panel shows the number of daily classifications (scores) conducted by the public between November 2015 and September 2016. The bottom panel shows the cumulative number of classifications conducted by the public between November 2015 and September 2016. Figure created by Dr Peter Smittenaar.

Retaining active users in crowdsourcing projects would increase the accuracy of results obtained, whilst also increasing the speed of data analysis. Therefore, efforts into extending participation are important for the future of data analysis by crowdsourcing. RTO could be considered as having two discrete elements of game play. The scoring of IHC data was required to be able to play gaming elements of the application but apart from this the two tasks remained distinct. Better integration of scoring tasks and game play could potentially help increase user enjoyment of the application and prolong active participation. An example of a highly integrated crowdsourcing game for data analysis is Phylo where much of the science behind the game was intentionally hidden from users.

The integration of crowdsourcing tasks into games has a number benefits that may decrease the rate of user drop-off. Games feature short term and long term rewards which, if well targeted, can incentivise continued participation. Due to the infancy of crowdsourcing games there are few studies into the motivations for participation. However, the limited evidence available suggests that social praise and a strong gaming community can help in the retention of users over time (Curtis, 2015). The inclusion of multiplayer features, allowing for collaboration or introducing a competitive element to gaming, has also been suggested as a feature that sustains engagement over time (Iacovides *et al*, 2013; Curtis, 2015). Further to promoting active participation, the addition of collaborative working into a game, such as RTO, could assist in improving the accuracy of users through peer training. Another avenue worthy of exploration, in order to limit participant inactivity and to assist in user uptake, would be to utilise the fan base of a pre-existing game. Project Discovery is a citizen science project that

employs such a strategy, existing as a mini game within the pre-existing game EVE online (EVE Online: Project Discovery).

A major issue in using crowdsourcing in molecular pathology studies is its reliability. In RTO we found the accuracy of public scoring to vary between immunostains. The lowest accuracy in public scores was seen in stains which experts classed as more difficult. Our data suggest that for more difficult IHC scoring tasks, the experience of 'experts' cannot be compensated for by the number of users and perhaps these are not applicable for analysis by crowdsourcing.

In this study the public demonstrated a reduced scoring accuracy in IHC stains that experts has identified as 'difficult' to score. It is possible that the accuracy of scoring these 'difficult' stains may be improved by the inclusion of more comprehensive tutorials for crowdsourcing projects assessing IHC. The tutorials included in RTO incorporated some elements of annotation and could be referred to whilst scoring tasks were being performed. However, overall, tutorials were short and fairly basic. The findings from Trailblazer (Lawson *et al*, 2017) highlight areas in which RTO tutorials could be improved. The use of feedback and test images would be strongly advisable for future applications. However, as the Trailblazer study was run simultaneously with RTO, and their findings only published as RTO was coming to a close, we were unable to implement the improvements highlighted by Trailblazer in RTO.

Similar to Cell Slider (Candido Dos Reis *et al*, 2015), in RTO the public were asked to score an isolated image of a segment of a TMA core. This step was taken to make viewing individual cells on a smartphone easier and eliminated the poor user experience of having to pinch-zoom. Segmentation was successful in achieving this objective.

However, the use of segmentation has the potential to limit the accuracy of scores generated by participants. This is due to the variation in the density of cancer cells across a TMA core. Using segmentation, the score from an area containing relatively few cancer cells is equal to the score derived from an area with a large number of cancer cells and has the potential to skew results, especially in cores where staining is heterogeneous. Furthermore, the viewing a whole tissue core can help a novice scorer in distinguishing cancer from normal tissue and infiltrative lymphocytes and allow a scorer to obtain a 'global' view of the staining across the whole core. As such studying an unsegmented TMA core can aid in the accuracy of scoring.

Ultimately the reliability of IHC scoring generated from crowdsourcing projects is still questionable. Further work is required to determine if the accuracy of crowdsourced analysis of molecular pathology data is reliable enough for continued use.

In the analysis of the association of public scores with clinical outcome (CSS) multiple unusable cores were observed in the 2006-2009 RT cohort. This was due to the previous sectioning of this TMA block, which had not occurred for the 1995-1999, 2002-2006 or cystectomy cohort. The 2006-2009 RT cohort suffered from a loss of total cores from the TMA, an increase in cores lacking tumour tissue and an increase in cores affected by diathermy artefacts, arising from TURBT. The effect of TMA block depletion was also seen when MRE11 staining was re-optimised and newly cut TMA sections stained. Compared to the 831 cores available for analysis for the initial MRE11 staining there were only 524 cores suitable for analysis for the re-optimised MRE11 staining.

In the association of public scores with clinical outcome, the initial MRE11 and CK20 stains showed the strongest associations with CSS. This is an encouraging result as these two proteins have been linked to CSS in MIBC previously.

MRE11 IHC has been directly reported to associate to CSS in MIBC by two independent research groups (Choudhury *et al*, 2010; Laurberg *et al*, 2012). Both studies identified high levels of MRE11 (greater than the 1st quartile of the data) to be associated with improved CSS in MIBC following radiotherapy-based treatment but with no association with outcome following cystectomy. While the results of this present analysis only found MRE11 levels above the 4th quartile (>75th percentile) to be significantly associated with improved CSS, the public scores were accurate enough to identify MRE11 as a candidate marker for further investigation. Furthermore, this association was identified despite this stain achieving a low scoring accuracy by the public, in comparison to other stains.

Two MRE11 stains (MRE11 re-optimised and initial MRE11) using the same antibody but a different protocol were included in RTO. The MRE11 re-optimised stain had reduced non-specific background staining and saturation of staining intensity compared to the initial MRE11 stain and so was hypothesised to be easier for the public to score and produce more reliable results (see chapter 3.2.2). Unfortunately, no association between MRE11 re-optimised and CSS was identified in this analysis. However, the likely reason for this, is a loss of patient tissue heterogeneity due to previous sectioning. It is important to note that the initial MRE11 staining in this thesis chapter differs from the original MRE11 staining described in Chapter 3 of this thesis. In this chapter MRE11 staining was conducted using a 1:6,000 antibody dilution rather than 1:3,000 antibody

dilution. As reported in the Chapter 3 (Table 3.1) changing the antibody dilution from a 1:3,000 to a 1:6,000 dilution was one of the features found to improve the quality of MRE11 staining and hence the initial MRE11 staining reported in this chapter is improved compared to the original MRE11 staining described in Chapter 3.

CK20 has been used to identify the luminal subtype of MIBC. Luminal MIBC is associated with better outcome compared to basal MIBC (Choi *et al*, 2014) and hence may be a potential prognostic biomarker for MIBC. In this study, the public CK20 scores identified a significant association between higher levels of CK20 and improved patient outcome in both the cystectomy and the 1995-1999 RT cohort. This result would be in keeping with CK20 being a prognostic marker for MIBC. However, in contrast to previous reports where CK5/6 has been found to associate with poor outcome in MIBC (Dadhania *et al*, 2016; Hayashi T *et al*, 2017; Zhang R *et al*, 2017), the public scores for the basal marker CK5/6 were not associated with CSS in this analysis. Furthermore, this was in spite of a high level of agreement between experts and crowdsourced scores for this marker.

In addition to the lack of association between CK5/6 and CSS, in this study a negligible correlation between CK5/6 and CK20 (r=-0.16) was identified. This is counter to the significant negative correlation between the two stains reported by Choi *et al* (2014). It would therefore be interesting to conduct a survival analysis for CK5/6 staining in a patient subset that is negative for CK20 staining.

Further analysis is being conducted on the public scores for each of the stains and their association with CSS. For MRE11 staining, Cox proportional hazard models are being performed comparing the $\leq 25^{th}$ percentile to $>25^{th}$ percentile. This is to enable a direct comparison to be made between the data reported here and that previously reported

by Choudhury *et al* (2010) and Laurberg *et al* (2012). Additionally, for all of the IHC stains the visual assessment of Kaplan-Meier plots is being used to determine alternative cut-off points for Cox-proportional hazards modelling, where appropriate. For CK20 staining, Cox proportional hazards modelling is being performed on scores grouped by less than or greater than median score. Multivariate analysis for MRE11 and CK20 staining is also being performed.

Throughout the course of RTO insight has been gained into the potential advantages and disadvantages of using crowdsourcing for the analysis of molecular pathology studies. A major advantage of using crowdsourcing to analyse IHC data is the potential time a well-planned and optimised method could save for skilled researchers. For crowdsourcing, researchers would only be required to score a small subset of data to generate tutorial images and 10% comparison data, thus freeing up time for other work. Although RTO did not achieve improvements in scoring efficiency, if all datasets had been immediately ready to input into the game and the final required numbers of users per image and images had been applied, all datasets could have been analysed within the first 14 days of the game going live on the internet. If the lessons learnt from RTO and Trailblazer are applied to future projects, crowdsourcing has the potential to accurately screen IHC data and greatly increase the speed of data analysis.

Chapter 7 Discussion

The aim of this thesis was to investigate potential biomarkers in MIBC. To do this, two previously identified potential MRE11 biomarkers were evaluated and a novel method of IHC screening was embarked upon.

One of the major areas of discussion in this work was the often underappreciated need for robustness in the conduct of IHC based biomarker studies. The practical simplicity of IHC is deceptive. Changes within any of the steps in the IHC procedure, or reagents used, can cause alterations in the quality and specificity of immunostaining, effecting the reliability of results (Henson, 2005; O'Hurley *et al*, 2014). While automation of the procedure should help in reducing erroneous staining due to practical mishandling and batch effects (Warford *et al*, 2004; Howat *et al*, 2014), in the transposition of manual methods to automated platforms the change in reagents and subtle differences in the steps, such as epitope retrieval, should be considered and rigorous assay validation should be conducted (Prichard, 2014).

In the MRE11 IHC staining of BCON tissue reported in Chapter 3 of this thesis, the need for IHC protocol re-optimisation was due to the transfer of the previously reported manual MRE11 IHC assay to an automated platform. The result of assay re-optimisation was found to significantly alter IHC staining both qualitatively and quantitatively. While the need for extensive assay validation for any technique may seem obvious, it cannot be overstated. Indeed, before the extensive staining conducted on BCON and BC2001 tissue, tests of the automated MRE11 IHC assay were conducted and optimised in terms of epitope retrieval and to an extent antibody dilution. Sadly, in practice this limited

optimisation of IHC protocols is often the extent of assay validation. In line with these findings, many reviews into the usefulness and reliability of IHC for biomarker identification have also reiterated these points (Henson, 2005; de Matos *et al*, 2010; Dunstan *et al*, 2011; O'Hurley *et al*, 2014; Carvajal-Hausdorf *et al*, 2015). This thesis has taken a critical view of IHC and its practice. However, the advantages of IHC such as cost effectiveness, an established pre-existing clinical skill base and the *in situ* nature of the technique are strongly in its favour. If IHC is conducted to a high standard then the assay is sensitive, reliable and useful.

The analysis of MRE11 IHC in anal cancer was an attempt at addressing the question of whether the ability of MRE11 to predict response to RT was isolated to MIBC. Anal cancer was chosen for the investigation due to the similarities in treatment and patient age between anal cancer and MIBC. Additionally, there is strong evidence for MRE11 having an important role in HPV-infected anal cancers that make up the majority of cases. No association between MRE11 and clinical outcome was found in this analysis, which provides evidence for the predictive capacity of MRE11 being MIBC specific. In accordance with this, within the Anne Kiltie lab, a C-terminally truncated form of MRE11 has been identified, which is specific to bladder cancer patients. This truncated MRE11 variant is less stable than full length MRE11 and has a reduced efficiency for HR. It is thought that this MRE11 variant is responsible for the counter intuitive increased response of MIBC patients with high levels of MRE11 to CRT.

In addition to IHC based biomarkers, this thesis also investigated the functional effect of a predictive DNA-based MIBC biomarker. Using clinical samples, the correlation between the *MRE11* rs1805363 germline SNP and increased expression of *MRE11*

isoform 2 was corroborated. However, interestingly there was evidence that the genotype of the tumour, rather than the germline genotype, was potentially more influential on the increase of *MRE11* isoform 2 expression identified within MIBC tissue.

Despite the appreciation of the existence of *MRE11* isoform 2, there is a significant lack of investigation into its function. As exon 16 is not associated with any functional domain of MRE11, it is perhaps not surprising that *MRE11* isoform 2 has been under-investigated. However, results reported here, although preliminary, imply that the absence of MRE11 exon 16, which is found in *MRE11* isoform 2, may result in radiosensitivity and a DNA damage repair defect. In recent years our understanding of the mechanism of action of the MRN complex has grown. It is now clear that the MRN complex is a highly flexible assembly that relies on large conformational changes for proper functioning (Lafrance-Vanasse *et al*, 2015). Therefore, the absence of exon 16 in *MRE11* isoform 2 may impair conformational shifts within the MRN complex thus limiting its activity.

Considering the association between increased *MRE11* isoform 2 expression and the rs1805363 SNP, the radiosensitising effect of MRE11 exon 16 was counter to our initial hypothesis. However, this work is preliminary and requires further analysis. The complexity of the interactions of *MRE11* isoform 2 within the MRN complex and with other DDR proteins needs to be more sensitively and extensively assayed. Importantly, as *MRE11* isoform 2 is present in conjunction with *MRE11* isoform 1, the functional implications of isoform interaction need to be assessed. Williams *et al* (2010) propose that, as the MRN complex is a heterohexamer composed of 3 protein dimers it can exist in multiple states and argue that different dimer assemblages within the MRN complex

affect how the complex responds to DNA damage. The best way of assessing the role of *MRE11* isoform 2 and its link to the rs1805363 SNP would be in rs1805363 positive bladder cancer cell line. However, currently we have not identified a cell line with this genotype.

The use of TMAs in IHC biomarker discovery has gained much popularity over the past decade. Evidence confirming the ability of TMAs to be representative of whole tissue sections has aided the acceptance of this technology along with the advantages of increased throughput and reliability. An unexpected finding of the work reported here was the limited utility of TMAs for MIBC. Diathermy and loss of tumour in samples resulted in significant loss of analysable tissue cores from all of the MIBC TMA cohorts used in this thesis. Additionally, due to these factors, some cores contained highly reduced areas suitable for scoring. While initial sections of MIBC TMAs are minimally or completely unaffected by diathermy and tumour loss, as more sections of tissue are taken and the depths of tissue cores are explored, the implications of these features become more obvious and severely limit the usefulness of the TMA. Indeed, in analysis of RTO scoring, one of the TMA cohorts was removed from formal analysis as over 2/3rd of cores were found to be unusable due to diathermy or tumour loss. Whole tissue sections also suffer from diathermy and loss of invasive tumour. However, as whole tissue sections are considerably larger than 0.6 mm, when areas of diathermy are present it is often still possible to assess tissue distal to the artefact. Furthermore, in whole tissue sections, as invasive areas are lost, new regions often emerge. These findings lead to the questioning of whether TMAs are a worthwhile use of MIBC FFPE patient tissue or whether this tissue would be better left in whole sections.

A limitation to using IHC to screen for potential biomarkers is the rate limiting step of traditional scoring. RTO was created to try and address this issue. Unfortunately RTO did not achieve its aim in increasing the speed of IHC scoring. However, over the course of the project a number of efficiencies were identified, which if implemented in future projects could increase their output. The use of crowdsourcing for data analysis is still in its infancy, especially for molecular pathology studies. The findings of RTO join those of Cell Slider and Trailblazer to add to the evidence as to whether this form of data analysis is suitable for biomarker screening. Taking the results of Cell Slider, Trailblazer and RTO together, the use of crowdsourcing as a screening tool in molecular pathology studies does seem viable. All of the studies found the crowdsourced pathological assessment to be fairly reliable and Trailblazer was useful in identifying a number of areas which may increase public scoring accuracy. The correlation of RTO scoring to clinical outcome was one of the distinguishing features of the project over Cell Slider and Trailblazer and provided further evidence on the ability of the public to screen for potential biomarker targets. RTO scoring was able to identify a number of trends in the datasets which have been previously reported and, if followed up using larger cohorts and whole tissue sections these trends may well be found to be significantly associated with clinical outcome.

Future work

The role of MRE11 as a predictive biomarker in MIBC still needs to be validated in a large retrospective tissue collection. Although the results from the re-optimised MRE11 staining on BCON TMA tissue did not find a significant association between MRE11 staining and improved survival following CRT, there were limitations in the power of the

experiment. Therefore, to obtain the evidence that automated MRE11 IHC can be used as a predictive biomarker of RT response in MIBC, a larger cohort of patients needs to be analysed. Furthermore, due to the limitations of TMAs in MIBC, to be completely reliable this work needs to be completed on whole tissue sections. Although IHC scoring between the Oxford and Manchester centres was found to be reliable in the re-optimised MRE11 staining of BCON tissue, the two scorers in Manchester did not conduct an independent analysis of the dataset before finding a consensus score. Therefore, to further assess the reliability of IHC scoring between centres a standardised protocol, highlighting the need for independence when scoring, should be implemented and inter-centre agreement re-evaluated.

Magliocco *et al* (2017) found that determining the nuclear to cytoplasmic ratio of MRE11 IHC was predictive of RT outcome in MIBC. Due to the nuclear function of MRE11 in this thesis only the nuclear staining of MRE11 IHC was scored. However, it may be that more reliable results are obtained from scoring the nuclear to cytoplasmic ratio of MRE11 staining. To test this, the nuclear to cytoplasmic ratio of MRE11 staining should be determined on bladder cancer tissue cohorts, the results of survival analyses with these scores should then be compared to the results survival analysis that used only nuclear scores.

To establish if the C terminally truncated form of MRE11, reported in Nicholson *et al* (2017), is associated with the predictive capacity of MRE11 IHC in MIBC, it would be useful to conduct a study where the level the C-terminally truncated form of MRE11 is determined in biopsies, and patient survival following RT or RC are followed up. Furthermore, due to the results of Teo *et al* (2014) and the findings reported in 5.2.1 of

this thesis, it would also be useful to determine the genotype of the rs1805363 SNP or level of *MRE11* isoform 2 expression in the samples in the previous or future studies of MRE11 IHC in MIBC. Performing these tests would distinguish if the use of these potential biomarkers in combination was a better predictor of response, than either marker alone.

The radiosensitising effect and decreased repair of radiation-induced DNA damage seen in a cell line expressing an exon 16 negative form of MRE11 requires further validation in additional cell lines. As MRE11 knock out is lethal to cells, a conditional MRE11 knock out cell line is currently being developed to assess the role of MRE11 lacking exon 16. In addition to this, it would be informative to screen pre-existing MIBC cell lines for the rs1805363 SNP and *MRE11* isoform 2 expression. The identification and use of a cell line with 'naturally' high levels of *MRE11* isoform 2 would have a number of benefits, for example, no MRE11 knockdown or knockout would be required in this scenario and so the potential off target effects of siRNAs would be ablated. Additionally, as *MRE11* isoform 2 has to date only been found to be expressed in a ratio with *MRE11* isoform 1, performing experiments with a highly expressing *MRE11* isoform 2 cell line may provide results which are more applicable to the physiological conditions in which *MRE11* isoform 2 functions.

An alternative to finding a pre-existing MIBC cell line with high levels of *MRE11* isoform 2 would be to establish primary cell lines, from patients identified as positive for the rs1805363 A minor allele and who express high levels of *MRE11* isoform 2.The advantage of using a primary cell culture over an established cell line, is that the results from primary cells are more closely related to the physiological response of cells in a

patient. However, the drawbacks of this approach are that primary cells have a limited lifespan and without sequencing the genetic background the makeup of the cells is unknown and it requires the ability to obtain fresh tumour biopsies.

If further work in additional cell lines confirms that MRE11 without exon 16/ *MRE11* isoform 2 confers sensitivity to DNA damage, there a number of areas to investigate to establish the function of this form of MRE11 in the DDR. These areas are: recruitment to sites of DNA damage, DNA binding, nuclease activity, determination of ATM activation, HR proficiency and resolution and restart of perturbed replication forks.

MRE11 needs to be recruited to sites of DNA damage to function. The localisation of exon 16 negative MRE11 to sites of DNA damage should, therefore, be assessed using immunofluorescent microscopy in cells expressing tagged MRE11 full length and exon 16 negative constructs.

Following from this, the binding of MRE11 lacking exon 16 to sites of DNA damage should also be investigated. This could be determined by performing an *in vitro* DNA binding assay where radiolabelled dsDNA oligonucleotides, containing either a blunt-end or a 3' ssDNA overhang, are incubated with a purified recombinant exon 16 negative form of MRE11. When the reaction products from this incubation are separated on a non-denaturing polyacrylamide gel, protein binding to the DNA can be identified. This assay has been previously described by de Jager *et al* (2001a) to determine the DNA binding of a full length MRE11 protein.

An alternative method to study the DNA binding of MRE11 without exon 16, would be to perform a chromatin immunoprecipitation (CHIP) in cells expressing a tagged MRE11 exon16-ve construct, after irradiation or treatment with a DNA damaging agent. As

opposed to a classical CHIP experiment, where probes for specific DNA sequences are used to detect the DNA immunoprecipitating with the protein, in this experiment the immunoprecipitated DNA would be labelled with [γ-32P]ATP to enable the visualisation of DNA without prior knowledge of its sequence. This method for measuring the binding of proteins to DNA has been formerly described by Wang *et al* (2009) and has also been used by Xie *et al* (2015) in the measurement of MRE11-DNA complexes.

The localisation and binding of MRE11 to sites of DNA damage enhances the rate of activation and retention of ATM at damaged DNA (Uziel *et al*, 2003; Cerosaletti & Concannon, 2004; So *et al*, 2009). Therefore, to confirm that this function of MRE11 is retained in MRE11 exon16 negative cells, or cells expressing high levels of *MRE11* isoform 2, ATM phosphorylation and ATM recruitment to sites of DNA damage should be assessed. The phosphorylation of ATM on S1981 is a marker for ATM activation. Therefore, the level of ATM activation after DNA damage can be determined by collecting protein lysates from cells at time points after damage induction and probing a western blot of the lysates with a phospho-anti-ATM (S1981) antibody. To determine if ATM is recruited efficiently to DNA damage in a MRE11 exon 16 negative background, the localisation of ATM to laser-induced DNA double-strand breaks could be assessed using immunofluorescent microscopy. The application of a time-course or use of live cell imaging in this experiment, would also provide data on the retention of ATM at the damaged site.

The nuclease activity of MRE11 is critical for many of its functions DNA damage repair. *In vitro* endonuclease and exonuclease assays with purified MRE11 have been previously described by Shibata et al (2014) and Yu *et al*, (2012). These assays could be

used to study the nuclease activity of MRE11 lacking exon 16. However, *in vitro* studies overlook the complex environment of the cell. Consequently, to confirm the results of *in vitro* experiments, the role of the MRE11 nuclease activity in cells should be assessed. However, to do this, the repair pathways in which MRE11 functions need to be investigated.

In HR, both the endo and exonuclease activity of MRE11 are required for the short range resection of the DNA break site. This short range resection is followed by long range resection of the DNA, which produces ssDNA overhangs that are coated in RPA and RAD51. The use of MRE11 nuclease inhibitors has been found to impair the formation of both RPA and RAD51 foci, after IR in G2 cells (Shibata *et al*, 2014). Therefore, the formation of RAD51 foci in replicating cells can be used as an indirect method to determine the nuclease activity of MRE11 after DNA damage. The resolution of RAD51 foci can also be used to determine the proficiency of HR. Conducting immunofluorescent microscopy to assess RAD51 foci in MRE11 exon 16 negative or high expressing *MRE11* isoform 2 cells would provide information on the nuclease activity of the MRE11 in these cells and whether a HR defect was evident. Another method for determining the proficiency of HR in MRE11 exon 16 negative cells or cells expressing high levels of *MRE11* isoform 2, would be to use the I-Sce1 homologous recombination reporter assay (Gunn & Stark, 2012).

The functionality of MRE11 lacking exon 16/*MRE11* isoform 2 in the restart and resolution of collapsed and stalled replication forks should also be determined. After IR the presence of complex ssDNA damage can result in the stalling and collapse of replication forks (Harper *et al*, 2010). Hydroxyurea (HU) induces replication fork stalling

and collapse, depending on the duration of exposure (Petermann *et al*, 2010). Performing the DNA fibre technique (Nieminuszczy *et al*, 2016) in MRE11 exon 16 negative cells or cells expressing high levels of *MRE11* isoform 2, treated with either HU or IR, would provide information on whether exon 16 negative MRE11 could function at replication forks and whether *MRE11* isoform 2 conferred a defect in the resolution of perturbed replication forks.

Our analysis of the public scores produced in RTO is ongoing. After assessing Kaplan-Meier curves for quartile grouping, univariate analysis with the Cox proportional hazard model is currently being conducted using appropriately grouped quartiles. Multivariate analysis of all the IHC stains, with the Cox proportional hazard model, is also being performed. In addition to this, it would also be interesting to analyse scores to see if any of the stains could be combined in a model, similar to TIL and p16 staining in anal cancer, where the use of two stains adds greater prognostic or predictive information over the use of one. As discussed in Chapter 6.3, CK20 and CK5/6 are good candidates for this. CK20 and Ki67 are also interesting for this type of analysis due to the findings in pT1 bladder cancer where Ki67 has been found to stratify tumours with abnormal (≥5% or complete absence staining) CK20 expression (Bertz *et al*, 2014).

Optimal conditions for automated IHC staining with antibodies on the Bondmax autostainer.

Antibody	Dilution	Heat Induced Epitope retrieval at 100°C	Conditions
Mre11 (Abcam) ab214 (original)	1:6000	Epitope retrieval solution 1 - 20 minutes	Primary antibody incubation – 15 minutes
Mre11 (Abcam) ab214 (new)	1:6000	Epitope retrieval solution 1 - 20 minutes	Pre-primary protein block - 10% BSA for 30 minutes Primary antibody incubation - 8 minutes
Tip60 nb100-87055 (Novus Biologicals)	1: 400,000	Epitope retrieval solution 1 - 20 minutes	Primary antibody incubation – 15 minutes
Rad50 ab89 (Abcam)	1:500	Epitope retrieval solution 1 - 20 minutes	Primary antibody incubation – 15 minutes
53BP1 #4937 (Cell Signalling Technology)	1:50	Epitope retrieval solution 1 - 40 minutes	Primary antibody incubation – 30 minutes
Ki67 M7240 (Dako)	1:1000	Epitope retrieval solution 1 - 20 minutes	Primary antibody incubation – 30 minutes
CK5/6 M7237 (Dako)	1:500	Epitope retrieval solution 2 - 20 minutes	Primary antibody incubation – 30 minutes
CK20 E16444 (Immunologic)	1:1000	Epitope retrieval solution 1 - 20 minutes	Primary antibody incubation – 30 minutes
NBS1 nb100-143 (Novus Biologicals)	1:500	Epitope retrieval solution 1 - 40 minutes	Primary antibody incubation – 30 minutes
RPA #22675 (Cell Signalling Technology)	1:100	Epitope retrieval solution 2 - 45 minutes	Primary antibody incubation – 45 minutes
HDAC4 #7628P (Cell Signalling Technology)	1:1000	Epitope retrieval solution 2 - 20 minutes	Primary antibody incubation – 30 minutes
HDAC2 #5113 (Cell Signalling Technology)	1:1000	Epitope retrieval solution 1 - 40 minutes	Primary antibody incubation – 30 minutes

Appendix B

Bondmax autostainer general protocol.

Primary antibody: 150µl/slide

Dewax:

1. Bond dewax solution – 30 secs -72°C
2. Bond dewax solution – 0 secs -72°C
3. Bond dewax solution – 0 secs – Ambient

Antigen retrieval: Variable

IHC Protocol:

1. Peroxide Block – 5 mins
2. Bond wash solution – 0 min
3. Bond wash solution – 0 min
4. Bond wash solution – 0 min
   -*Optional* Pre-Primary Block- 10% Albumin Bovine Serum -30mins
5. Bond wash solution – 0 min
6. Bond wash solution – 0 min
7. Bond wash solution – 0 min
8. MARKER – Primary Antibody – Variable
9. Bond wash solution – 0 min
10. Bond wash solution – 0 min
11. Bond wash solution – 0 min
12. Post Primary – 8 min
13. Bond wash solution – 2 mins
14. Bond wash solution – 2 mins
15. Bond wash solution – 2 mins
16. Polymer – 8 mins
17. Bond wash solution – 2 mins
18. Bond wash solution – 2 mins
19. Deionized water – 0 min
20. Mixed DAB Refine – 0 min
21. Mixed DAB Refine – 10 mins
22. Deionized water – 0 min
23. Deionized water – 0 min
24. Deionized water – 0 min
25. Haematoxylin – 1 mins
26. Deionized water – 0 mins *optional* 2 mins
27. Bond wash solution – 0 mins *optional* 2 mins
28. Deionized water – 0 mins *optional* 2 mins

Optimal conditions for manual IHC staining with antibodies.

Manual IHC Antibody	Primary Antibody Dilution	Heat Induced Epitope retrieval	Antibody specific conditions
p21 #2947P (Cell Signalling Technology)	1: 20	Citrate buffer	Manual Kit - Dako REAL Detection System, Peroxidase/DAB+, Rabbit/Mouse kit Primary antibody incubation – Overnight 4°C Dilution of DAB – 1:50
p53 sc-126 (Santa Cruz)	1:50,000	Tris-EDTA buffer	Manual Kit - Dako REAL Detection System, Peroxidase/DAB+, Rabbit/Mouse kit Primary antibody incubation – Overnight 4°C Dilution of DAB - 1:200
MRE11 c-terminal ab30725 (Abcam)	1:1000	Citrate buffer	Manual Kit - Novolink Polymer Detection System Primary antibody incubation – 1 hr room temperature Dilution of DAB – 1:50

Primer sequences.

Primer name	Use	Sequence
MRE11 rs1805363	PCR -Sanger sequencing of rs1805363 SNP	Forward- GGACTTGAAGCATCTACGTT Reverse- ATGCGATTCCTAAATTACCC
MRE11 isoform	PCR - Identification of *MRE11* isoform 1 and 2	Forward- CCCGCGGAATTCAGGTTTA Reverse- TTTATGGTCAGTCAAGCTCCTC
GAPDH	PCR – Control for cDNA input and quality in identification of *MRE11* isoform 1 and 2	Forward- ACAGTCAGCCGCATCTTCTT Reverse- AATGAAGGGGTCATTGATG
SDMexon16	Site directed mutagenesis - splice exon 16 from MRE11 plasmid	Forward- TACTAAGAATTATTCAGAGGTGGTCCAGCACATCATCCA Reverse- GGATGATGTGCTGGACCACCTCTGAATAATTCTTAGTAG
ATP5B	qPCR - Endogenous control	Forward- ACCATCAAAGGATTCCAGCA Reverse- GCTTTTGCCACAGCTTCTTC
Central MRE11	qPCR - Amplification of all MRE11 in sample	Forward- TCCTAAAGTAACCCAAGCCATAC Reverse- TTCTCTGGCTGGTGAGAATTAC
Exon 15-16	qPCR – Amplification of only MRE11 containing exon 16	Forward- GCCTTCCCGAAATGTCACTA Reverse- GGACCACCTTTGATCTGTCTT

Appendix E

The antibody dilutions used for western blotting and immunoprecipitations.

Antibody	Species	Antibody dilution	Western blot secondary
MRE11 ab214 (Abcam)	Mouse	Immunoprecipitation - 1:500 Western blot - 1:1,000	IRDye 800CW Goat anti-Mouse IgG (LI-COR Biosciences)
RAD50 ab89 (Abcam)	Mouse	Immunoprecipitation - 1:1,000 Western blot - 1:1,000	IRDye 800CW Goat anti-Mouse IgG (LI-COR Biosciences)
NBS1 NB110-143 (Novus Biologicals)	Rabbit	Immunoprecipitation - 1:500 Western blot - 1:1,000	IRDye 680RD Goat anti-Rabbit IgG (LI-COR Biosciences)
anti-dimethyl-arginine antibody, asymmetric (ASYM25) 09-814 (Millipore)	Rabbit	Immunoprecipitation 1:1,000	IRDye 680RD Goat anti-Rabbit IgG (LI-COR Biosciences)
MRE11 c-terminal ab30725 (Abcam)	Rabbit	Western blot - 1:1,000	IRDye 680RD Goat anti-Rabbit IgG (LI-COR Biosciences)
HDAC4 #7628P (Cell Signalling Technology)	Rabbit	Western blot - 1:2,000	IRDye 680RD Goat anti-Rabbit IgG (LI-COR Biosciences)
HDAC2 #5113 (Cell Signalling Technology)	Mouse	Western blot - 1:3,000	IRDye 800CW Goat anti-Mouse IgG (LI-COR Biosciences)
RPA #22675 (Cell Signalling Technology)	Rabbit	Western blot - 1:1,000	IRDye 680RD Goat anti-Rabbit IgG (LI-COR Biosciences)
Cytokeratin 5/6 (CK5/6) M7237 (Dako)	Mouse	Western blot - 1:500	IRDye 800CW Goat anti-Mouse IgG (LI-COR Biosciences)
Anti-keratin 20 (CK20) E16444 (Amsbio)	Rabbit	1:1,000	IRDye 680RD Goat anti-Rabbit IgG (LI-COR Biosciences)
53BP1 #4937 (Cell Signalling Technology)	Rabbit	1:500	IRDye 680RD Goat anti-Rabbit IgG (LI-COR Biosciences)
p53 sc-126 (Santa Cruz)	Mouse	Western blot - 1:1,000	IRDye 680RD Goat anti-Rabbit IgG (LI-COR Biosciences)
p21 #2947P (Cell Signalling Technology)	Rabbit	Western blot - 1:500	IRDye 680RD Goat anti-Rabbit IgG (LI-COR Biosciences)
Tip60 nb100-87055 (Novus Biologicals)	Mouse	Western blot - 1:1,000	IRDye 800CW Goat anti-Mouse IgG (LI-COR Biosciences)

Appendix F

Walker AK, Kartsonaki C, Collantes E, *et al.* (2017) No additional prognostic value for MRE11 in squamous cell carcinomas of the anus treated with chemo-radiotherapy. *Br J Cancer* **117**: 322–325, doi:10.1038/bjc.2017.188.

Correction to: *British Journal of Cancer* (2017) **117**, 322–325; doi:10.1038/bjc.2017.188; published online 22 June 2017

SHORT COMMUNICATION

BJC

British Journal of Cancer (2017) 117, 322–325 | doi: 10.1038/bjc.2017.188 BJC OPEN

Keywords: anal cancer; human papillomavirus; immunohistochemistry; MRE11; p16; tumour-infiltrating lymphocytes

No additional prognostic value for MRE11 in squamous cell carcinomas of the anus treated with chemo-radiotherapy

This article has been corrected since publication and a Corrigendum has been published

Alexandra K Walker[1], Christiana Kartsonaki[2,3], Elena Collantes[4], Judith Nicholson[1], Duncan C Gilbert[5] and Anne E Kiltie*[,1]

[1]CRUK/MRC Oxford Institute for Radiation Oncology, University of Oxford, Oxford OX3 7DQ, UK; [2]Nuffield Department of Population Health, University of Oxford, Oxford OX3 7DQ, UK; [3]Medical Research Council Population Health Research Unit (MRC PHRU) at the University of Oxford, Oxford OX3 7DQ, UK; [4]Department of Cellular Pathology, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford OX3 7DU, UK and [5]Sussex Cancer Centre, Royal Sussex County Hospital, Eastern Road, Brighton BN2 5BE, UK

Background: The majority of anal cancers (84–95%) are driven by infection with human papillomavirus (HPV). HPV-positive tumours show significantly better responses to chemo-radiotherapy when compared with HPV-negative tumours. HPV infection is linked to alterations in DNA damage response proteins, including MRE11. MRE11 is a potential predictive biomarker for response to radiotherapy in muscle-invasive bladder cancer and may hold predictive power in other cancers.

Methods: Using a previously reported cohort, we evaluated the levels of MRE11 in anal cancer and assessed its predictive value in this disease.

Results: We found no association between the level of MRE11 and relapse-free survival following chemo-radiotherapy.

Conclusions: MRE11 has no predictive value in the analysis of relapse-free survival after chemo-radiotherapy in anal cancer and does not add to the prognostic value of p16 and tumour-infiltrating lymphocyte scores. Further investigation into the role of DNA repair proteins in anal cancer is required.

Although anal cancer remains a relatively rare disease, its incidence is increasing (Shiels *et al*, 2015). Most cases (approximately 85%) are squamous cell carcinoma of the anus and anal canal (SCCA) with the remaining cases attributed to adenocarcinoma (10%) or other types (5%) (Nelson *et al*, 2013). Small (T1) tumours may be treated with surgical excision alone, but for most tumours, the standard of care is combination chemo-radiotherapy (Glynne-Jones *et al*, 2009), which achieves good responses with complete tumour regression in 80–90% of cases (Vinayan and Glynne-Jones, 2016). However, acute toxicities resulting from current treatment regimens are common and patients can suffer long-term effects, including bowel, bladder and sexual dysfunction, and more rarely lower limb venous thrombosis (Ghosn *et al*, 2015).

Infection with high-risk subtypes of human papillomavirus (HPV), predominately HPV16, are strongly associated with the development of anal cancer, with reported estimates of HPV positivity in anal cancer ranging from 84% to 95% (Baricevic *et al*, 2015). HPV status, either measured directly or by overexpression

*Correspondence: Professor AE Kiltie; E-mail: anne.kiltie@oncology.ox.ac.uk

Received 26 January 2017; revised 31 May 2017; accepted 1 June 2017; published online 22 June 2017

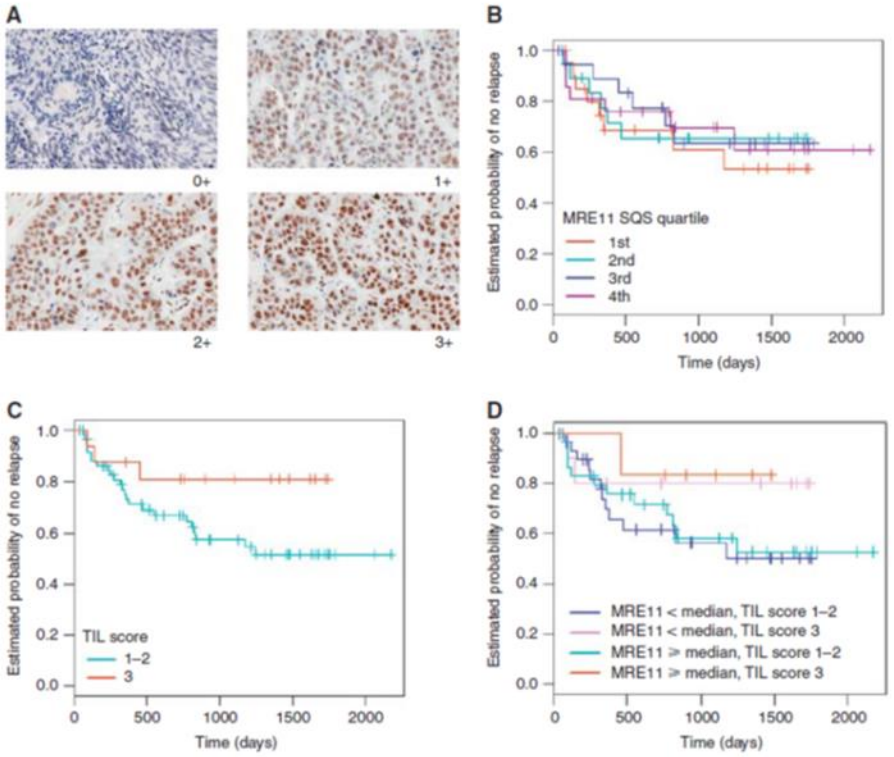


Figure 1. Results of MRE11 Immunohistochemistry in anal cancer specimens. (**A**) Representative staining intensities of MRE11 in anal cancer tissue. (**B**) Kaplan–Meier relapse-free survival curves for 82 anal cancer patients stratified by MRE11 semiquantitative score (SQS). MRE11 SQSs are grouped into four quartiles. (**C**) Kaplan–Meier relapse-free survival curves for 78 anal cancer patients stratified by tumour-infiltrating lymphocyte (TIL) score. TIL scores are grouped into low–moderate (1–2) or high (3). (**D**) Kaplan–Meier relapse-free survival curves for 78 anal cancer patients stratified by MRE11 SQS (greater than or equal to and less than the median score) and TIL scores (low–moderate (1–2) or high (3)).

of the surrogate marker p16, has a significant effect on patient outcome. Individuals identified as negative for HPV show significantly worse response to therapy compared with patients positive for HPV infection (Gilbert *et al*, 2013; Doll *et al*, 2014; Koerber *et al*, 2014; Serup-Hansen *et al*, 2014; Mai *et al*, 2015; Meulendijks *et al*, 2015), indicating that the current standard treatment is suboptimal for these patients. Further to this, the presence of tumour-infiltrating lymphocytes (TIL) has also been found to stratify p16-positive SCCA. Patients with high TIL scores show significantly better relapse-free survival compared with patients with low TIL scores (Gilbert *et al*, 2016).

The DNA damage response (DDR) is responsible for sensing, signalling and repair of DNA damage. It comprises a network of cellular pathways leading to the activation of a number of cellular mechanisms, including DNA damage repair, cell cycle checkpoints, apoptosis and chromatin remodelling. MRE11 is involved in the initiation of the DDR and functions within the MRN (MRE11-RAD50-NBS1) protein complex (Ciccia and Elledge, 2010). HPV infection is known to activate the DDR system for amplification of the viral genome and recruits DDR proteins to sites of viral replication (Moody and Laimins, 2009; Sakakibara *et al*, 2011; Reinson *et al*, 2013). Additionally, HPV-positive cells also show increased expression of DDR proteins, including MRE11 and other members of the MRN complex (Johnson *et al*, 2017).

MRE11 has been identified as a potential biomarker for outcome following radiotherapy in muscle-invasive bladder cancer. Patients who exhibited high levels of MRE11 showed improved cancer-specific survival following radiotherapy compared with those with low levels (Choudhury *et al*, 2010; Laurberg *et al*, 2012), but no association was seen following cystectomy. Owing to the aberrant activity of the DDR system caused by HPV infection and the high proportion of HPV-driven anal carcinomas, we wished to test the hypothesis that MRE11 might be a useful additional biomarker in anal cancer.

MATERIALS AND METHODS

With the appropriate ethical approval (11/LO/1032), clinical details and corresponding tumour blocks were retrieved from patients treated with radical chemo-radiotherapy for non-metastatic SCCA from 2004 to 2009 as previously described (Gilbert *et al*, 2013).

Immunohistochemistry was conducted on four-micron thick tissue sections using a Leica Bond-max autostainer (Leica Microsystems GmbH, Wetzlar, Germany) with a Bond Polymer Refine Detection kit (DS9800: Leica Microsystems Inc., Newcastle, UK) and assay-specific reagents. Epitope retrieval was conducted with low-pH buffer for 20 min. A protein blocking step with 10% BSA for

Table 1. Patient demographics for 82 patients whose samples were stained for MRE11

Characteristic	Number of individuals (%)
Sex	
Female	43 (52.4)
Male	39 (47.6)
T stage ($n=68$)	
T1	7 (8.5)
T2	26 (31.7)
T3	22 (26.8)
T4	13 (15.9)
NA	14 (17.1)
N stage ($n=71$)	
N0	40 (48.8)
N1	9 (11.0)
N2	19 (23.2)
N3	3 (3.6)
NA	11 (13.4)
p16 ($n=82$)	
p16 positive	72 (87.8)
p16 negative	10 (12.2)
p53 ($n=82$)	
Strong p53 staining	19 (23.2)
Negative-moderate p53 staining	63 (76.8)
TIL score ($n=78$)	
TIL 1 (low/absent)	18 (23.1)
TIL 2 (moderate)	44 (56.4)
TIL 3 (high)	16 (20.5)
NA	4 (5.1)
Relapse ($n=82$)	27

Abbreviations: NA = not available; TIL = tumour-infiltrating lymphocyte.

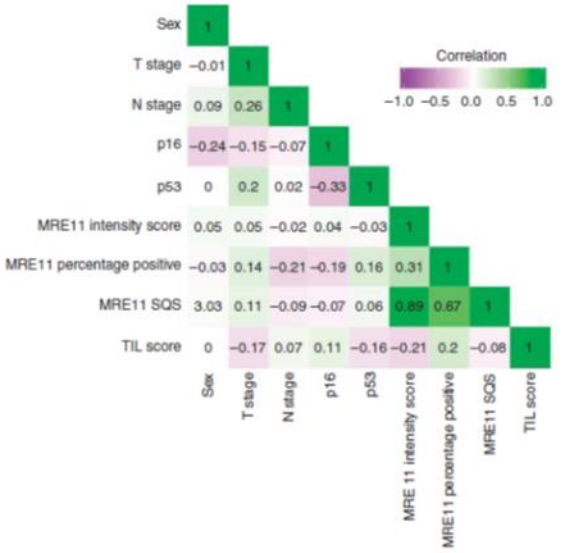


Figure 2. Correlation matrix for all variables used in the analysis. Negative values of Pearson's correlation coefficient, in purple, indicate a negative correlation between variables. Positive values, in green, indicate a positive correlation between variables. The intensity of the purple or green colour indicates the strength of the correlation between the variables.

30 min was applied before incubation with MRE11 primary antibody (ab214, Abcam, Cambridge, UK). Primary antibody was used at a 1:6000 dilution with samples incubated for 8 min. Leica polymer and postpolymer were applied for 8 min followed by DAB for 10 min and a haematoxylin counterstain for 1 min.

The Aperio ScanScope CS2 digital slide scanner (Leica Microsystems GmbH, Wetzlar, Germany) was used to image the stained sections at ×400 magnification, which were viewed with Aperio Image Scope viewing software (Leica Biosystems Imaging, Vista, CA, USA). Invasive cancerous areas of stained tissue were marked on H&E sections of corresponding tissue by a trained pathologist (EC). Up to 10 high-magnification images of these areas were captured. The intensity and percentage of cancer cells stained for MRE11 were scored in identified invasive cancerous areas by two independent scorers. The intensity score was determined according to a range of 0+ (negative), 1+ (weak), 2+ (moderate) or 3+ (strong staining) (Figure 1A). A consensus score for both intensity and percentage of cancer cells stained in each section was reached and semiquantitative scores (SQSs) were generated (intensity × percentage positive).

Associations between clinicopathological characteristics and relapse-free survival were assessed using multivariable Cox proportional hazards models. Relapse-free survival time was defined as the time from diagnosis of SCCA to the date of diagnosis of confirmed relapse. Individuals were censored at the date of last follow-up. The proportional hazards assumption was checked using scaled Schoenfeld residuals. Statistical analyses were conducted using the statistical software R (R Core Team, 2016).

RESULTS

Analysis was conducted on a subset of samples from a previously reported SCCA cohort (Gilbert *et al*, 2013). In total, attempts were made to stain tissue from 124 out of 138 individual patients. Loss of patients from the cohort arose due to lack of remaining tissue in block, absence of invasive pathology or presence of cauterisation.

MRE11 immunohistochemistry was then conducted on 124 patients. Forty-two patients were excluded from the analysis due to loss of evaluable tissue. The patient demographic data for the 82 patients' evaluable samples is shown in Table 1. This patient subset showed the same associations with relapse for p16 and p53 as previously reported (Gilbert *et al*, 2013). In univariate analysis, p16-positive cases showed significantly lower risk of relapse (HR = 0.16, (95% CI = 0.07–0.40, $P<0.001$)) compared with the p16-negative cases, whereas strong p53 staining was significantly associated with a higher risk of relapse (HR = 2.3, (95% CI = 1.05–5.22, $P=0.037$)) compared with negative-to-moderate p53 staining.

Univariate analysis of the MRE11 staining, assessed using SQS, showed no association between levels of MRE11 and risk of relapse (HR = 0.999, 95% CI = 0.994–1.004, $P=0.61$; Figure 1B). In multivariable Cox regression analysis with T stage, N stage, sex, p16 status, p53 status and MRE11 SQS as covariates, only p16 was found to be of prognostic value (HR = 0.093, 95% CI = 0.024–0.359, $P<0.001$), a result consistent with previously published data on this cohort (Gilbert *et al*, 2013). This result was also seen when multivariate Cox regression analysis was conducted using constituent MRE11 intensity and percentage of positive scores in place of MRE11 SQS (HR for p16: 0.039 (95% CI = 0.006–0.230, $P<0.001$), HR per unit increase of MRE11 log(percentage positive): 0.37 (95% CI = 0.065–2.080, $P=0.26$), HR for MRE11

intensity 2 *vs* 1: 1.98 (95% CI = 0.548–7.175, $P = 0.30$) and HR for MRE11 intensity 3 *vs* 1: 0.65 (95% CI = 0.128–3.287, $P = 0.60$)).

Analysis was attempted on p16-negative cases but the number of p16-negative individuals was too small for reliable statistical analysis. There was no substantial difference in the range of staining observed between p16-positive and p16-negative cases. TIL have recently been reported to add prognostic value over and above p16 immunohistochemistry in SCCA, in part using this cohort of patients (Gilbert *et al*, 2016). We therefore investigated MRE11 alongside TIL scores on this cohort. The addition of MRE11 SQS to TIL score had no significant prognostic value over and above TIL score alone (Figures 1C and D).

MRE11 levels, assessed by SQS, percentage positivity or intensity scoring, were not highly correlated with any of the variables used in the analysis (Figure 2).

DISCUSSION

In contrast to our previous findings in muscle-invasive bladder cancer, we did not identify MRE11 as having predictive value in assessing outcome in SCCA, despite both types of cancer being treated with definitive chemo-radiotherapy. Presumably this represents the heterogeneity that exists between not only different cancers but also different cancer subtypes (Burrell *et al*, 2013). We additionally assessed whether MRE11 had any role to play over and above previously described prognostic markers (p16, p53 and TIL (Jones *et al*, 2017)) but found no evidence of an association with outcome. It is possible that, despite the lack of correlation to outcome in p16/HPV-positive tissue, the assessment of MRE11 levels in p16/HPV-negative tissues might yield positive findings. However, owing to the rarity of the pathology, this would be difficult to investigate.

In SCCA, chemo-radiotherapy achieves a good response rate. However, the toxicities resulting from treatment can be significantly detrimental to patient's quality of life, and for some patients, particularly those with HPV-negative tumours, outcomes are poor. There is much evidence for alteration in the DDR during HPV transformation (Anacker and Moody, 2017), but this has not yet been fully examined in the context of SCCA and, in particular, its response to therapy. To improve treatment outcomes, further investigation into the molecular diversity of anal cancer is needed.

ACKNOWLEDGEMENTS

We thank Marcus Green for his expert technical assistance with block sectioning. This work was funded by CRUK Programme Grant C5255/A15935 (to AEK) and an MRC studentship to AKW (MR/K501256/1).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Anacker DC, Moody CA (2017) Modulation of the DNA damage response during the life cycle of human papillomaviruses. *Virus Res* **231**: 41–49.

Baricevic I, He X, Chakrabarty B, Oliver AW, Bailey C, Summers J, Hampson L, Hampson I, Gilbert DC, Renehan AG (2015) High-sensitivity human papilloma virus genotyping reveals near universal positivity in anal squamous cell carcinoma: different implications for vaccine prevention and prognosis. *Eur J Cancer* **51**: 776–785.

Burrell RA, McGranahan N, Bartek J, Swanton C (2013) The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**: 338–345.

Choudhury A, Nelson LD, Teo MTW, Chilka S, Bhattarai S, Johnston CF, Elliott F, Lowery J, Taylor CF, Churchman M, Bentley J, Knowles MA, Harnden P, Bristow RG, Bishop DT, Kiltie AE (2010) MRE11 expression is predictive of cause-specific survival following radical radiotherapy for muscle-invasive bladder cancer. *Cancer Res* **70**: 7017–7026.

Ciccia A, Elledge SJ (2010) The DNA damage response: making it safe to play with knives. *Mol Cell* **40**: 179–204.

Doll CM, Winter K, Ajani J, Klimowicz A, Crane C, Kachnic L, Okawara G, Lukka H, Berk L, Roof K, Becker M, Grisell D, Guha C, Magliocco AM (2014) The significance of p16 and p53 expression on clinical outcome in patients with anal cancer treated with chemoradiation therapy: an analysis of RTOG 98-11. *Int J Radiat Oncol* **90**: S32.

Ghosn M, Kourie HR, Abdayem P, Antoun J, Nasr D (2015) Anal cancer treatment: current status and future perspectives. *World J Gastroenterol* **21**: 2294–2302.

Gilbert DC, Serup-Hansen E, Linnemann D, Høgdall E, Bailey C, Summers J, Havsteen H, Thomas GJ (2016) Tumour-infiltrating lymphocyte scores effectively stratify outcomes over and above p16 post chemo-radiotherapy in anal cancer. *Br J Cancer* **114**: 134–137.

Gilbert DC, Williams A, Allan K, Stokoe J, Jackson T, Linsdall S, Bailey CM, Summers J (2013) p16INK4A, p53, EGFR expression and KRAS mutation status in squamous cell cancers of the anus: correlation with outcomes following chemo-radiotherapy. *Radiother Oncol* **109**: 146–151.

Glynne-Jones R, Northover J, Oliveira J. ESMO Guidelines Working Group O behalf of the EGW (2009) Anal cancer: ESMO clinical recommendations for diagnosis, treatment and follow-up. *Ann Oncol* **20**(Suppl 4): 57–60.

Johnson BA, Aloor HL, Moody CA (2017) The Rb binding domain of HPV31 E7 is required to maintain high levels of DNA repair factors in infected cells. *Virology* **500**: 22–34.

Jones CM, Goh V, Sebag-Montefiore D, Gilbert DC (2017) Biomarkers in anal cancer: from biological understanding to stratified treatment. *Br J Cancer* **116**: 156–162.

Koerber SA, Schoneweg C, Slynko A, Krug D, Haefner MF, Herfarth K, Debus J, Sterzing F, Von M, Doeberitz K, Prigge E-S, Reuschenbach M (2014) Influence of human papillomavirus and p16 INK4a on treatment outcome of patients with anal cancer. *Radiother Oncol* **113**: 331–336.

Laurberg JR, Brems-Eskildsen AS, Nordentoft I, Fristrup N, Schepeler T, Ulhøi BP, Agerbaek M, Hartmann A, Bertz S, Wittlinger M, Fietkau R, Rödel C, Borre M, Jensen JB, Ørntoft T, Dyrskjøt L (2012) Expression of TIP60 (tat-interactive protein) and MRE11 (meiotic recombination 11 homolog) predict treatment-specific outcome of localised invasive bladder cancer. *BJU Int* **110**: E1228–E1236.

Mai S, Welzel G, Ottstadt M, Lohr F, Severa S, Prigge ES, Wentzensen N, Trunk MJ, Wenz F, von Knebel-Doeberitz M, Reuschenbach M (2015) Prognostic relevance of HPV infection and p16 overexpression in squamous cell anal cancer. *Int J Radiat Oncol Biol Phys* **93**: 819–827.

Meulendijks D, Tomasoa NB, Dewit L, Smits PH, Bakker R, van Velthuysen ML, Rosenberg EH, Beijnen JH, Schellens JH, Cats A (2015) HPV-negative squamous cell carcinoma of the anal canal is unresponsive to standard treatment and frequently carries disruptive mutations in TP53. *Br J Cancer* **112**: 1358–1366.

Moody CA, Laimins LA (2009) Human papillomaviruses activate the ATM DNA damage pathway for viral genome amplification upon differentiation. *PLoS Pathog* **5**: e1000605.

Nelson RA, Levine AM, Bernstein L, Smith DD, Lai LL (2013) Changing patterns of anal canal carcinoma in the United States. *J Clin Oncol* **31**: 1569–1575.

R Core Team (2016) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria.

Reinson T, Toots M, Kadaja M, Pipitch R, Allik M, Ustav E, Ustav M (2013) Engagement of the ATR-dependent DNA damage response at the human papillomavirus 18 replication centers during the initial amplification. *J Virol* **87**: 951–964.

Sakakibara N, Mitra R, McBride AA (2011) The papillomavirus E1 helicase activates a cellular DNA damage response in viral replication foci. *J Virol* **85**: 8981–8995.

Serup-Hansen E, Linnemann D, Skovrider-Ruminski W, Hgødall E, Geertsen PF, Havsteen H (2014) Human papillomavirus genotyping and p16 expression as prognostic factors for patients with American Joint Committee on Cancer stages I to III carcinoma of the anal canal. *J Clin Oncol* **32**: 1812–1817.

Shiels MS, Kreimer AR, Coghill AE, Darragh TM, Devesa SS (2015) Anal cancer incidence in the United States, 1977-2011: distinct patterns by histology and behavior. *Cancer Epidemiol Biomarkers Prev* **24**: 1548–1556.

Vinayan A, Glynne-Jones R (2016) Anal cancer–what is the optimum chemoradiotherapy? *Best Pract Res Clin Gastroenterol* **30**: 641–653.

Results from the germline genotyping of 189 MIBC cancer patients for the rs1805363 SNP.

Patient samples identified as positive for the A minor allele are highlighted in pink.

Sample ID	Genotype	Sample ID	Genotype	Sample ID	Genotype	Sample ID	Genotype	Sample ID	Genotype	Sample ID	Genotype	Sample ID	Genotype
s1446	GG	s0236	GG	s0187	GG	s0409	GG	s0968	GG	s0692	GG	s0563	GG
s0113	**AG**	s0986	GG	**s0521**	**AG**	s1833	GG	s1161	GG	s0610	GG	**s0168**	**AG**
s0225	GG	**s2200**	**AG**	s1472	GG	s0181	GG	s0270	GG	s0124	GG	s0942	GG
s0066	GG	s0615	GG	s0829	GG	s0058	GG	s0392	GG	s1919	GG	s1474	GG
s0809	GG	**s0472**	**AG**	s0436	GG	s1182	GG	s0431	GG	s0054	GG	s0625	GG
s0136	**AG**	s0674	GG	s1225	GG	s0908	GG	s1460	GG	s0428	GG	**s1402**	**AG**
s0142	GG	s1140	GG	**s0230**	**AG**	s1362	GG	s0985	GG	s0314	GG	s0639	GG
s0427	GG	s0552	GG	s1116	GG	s0710	GG	s0335	GG	s0605	GG	s0537	GG
s0500	**AG**	s0582	GG	s0398	GG	s0023	GG	s1353	GG	s0141	GG	s0229	GG
s0929	GG	s0983	GG	s0594	GG	s1801	GG	s0815	GG	s0608	GG	s0673	GG
s0189	GG	s0573	GG	s0883	GG	s0570	GG	s2000	GG	s1395	GG	s0221	GG
s0242	GG	s0423	GG	s0694	GG	**s0637**	**AG**	s0144	GG	s0896	GG	s1834	GG
s0729	**AG**	s0578	GG	s0215	GG	s1184	GG	s1448	GG	s0044	GG	s1177	GG
s0237	**AG**	s0086	GG	s1459	GG	s0539	GG	s0243	GG	s0140	GG	s1485	GG
s0193	GG	s0660	GG	s0550	GG	s0624	GG	s0420	GG	s0020	GG	s0091	GG
s0139	GG	s0982	GG	s0873	GG	s0263	GG	s0063	GG	**s0643**	**AG**	s0226	GG
s0559	GG	s0731	GG	s1209	GG	s1186	GG	s0278	GG	s0590	GG	s0551	GG
s0396	GG	s0164	GG	s0400	GG	**s2209**	**AG**	s0309	GG	**s1434**	**AG**	s0128	GG
s1272	GG	s1401	GG	s1200	GG	s0407	GG	s0753	GG	s1239	GG	s0389	GG
s0566	GG	s0682	GG	s1908	GG	s0426	GG	s1354	GG	s0035	GG	**s0724**	**AG**
s0684	GG	s0916	GG	s0287	GG	s0876	GG	s0566	GG	s0418	GG	**s1514**	**AG**
s0167	GG	**s0890**	**AG**	s1222	GG	s0414	GG	s1194	GG	s0080	GG	s0735	GG
s0549	GG	s1486	GG	s1224	GG	s0672	GG	s1906	GG	s0425	GG	s0810	GG
s0632	GG	**s1214**	**AG**	s0051	GG	s0891	GG	s0572	GG	**s1829**	**AG**	s0304	GG
s0532	GG	s0564	GG	s0298	GG	s0171	GG	s0628	GG	s1821	GG		
s0874	GG	s0421	GG	s0569	GG	s1558	GG	s1806	GG	s0052	GG		
s1352	GG	**s0295**	**AG**	s0525	GG	s1349	GG	s0631	GG	**s1126**	**AG**		

Appendix H

The IHC stains analysed by RTO. IHC stains are listed in order they were placed into the game. Information on the predominant cellular localisation of the staining, number of tissue cores scored, total number of classifications and average number of classifications per core is presented for each IHC stain.

Marker	Localisation of antibody stain	Total number of cores analysed by RTO	Total number of classifications	Average number of classifications per core
MRE11	nuclear	831	910,008	1095
RAD50	nuclear	786	455,429	579
p21	nuclear	845	1,040,933	1232
53BP1	nuclear	841	201,390	239
p53	nuclear	849	265,334	313
CK5/6	membranous	814	86,469	106
CK20	membranous	806	78,613	98
Tip60	nuclear	850	90,850	107
MRE11 re-optimised	nuclear	524	49,857	95
MRE11 c-terminal	nuclear	376	102,417	272
Ki67	nuclear	775	68,637	89

Appendix I

Univariate Cox proportional hazard modelling for all of the IHC stains scored by RTO by percentiles. Hazard ratios, 95% confidence intervals and p values are display for the Cystectomy cohort, 1995-1999 RT cohort and 2002-2006 RT cohort. Significant p values are highlighted (*= p≤0.05, **= p≤0.01, ***=p≤0.001). The analysis compared patients with IHC scores ≤25th percentile to patients with scores within the 26-50th, 51-75th and >75th percentiles. Statistics performed by Dr Christiana Kartsonaki.

	Cystectomy cohort Hazard ratio, [95% confidence interval, p value]	**1995-1999 RT cohort** Hazard ratio, [95% confidence interval, p value]	**2002-2006 RT cohort** Hazard ratio, [95% confidence interval, p value]
MRE11 ≤25th vs 26-50th ≤25th vs 51-75th ≤25th vs >75th percentile	n=75 1.808 [0.721- 4.534, 0.207] 1.4124 [0.547- 3.649, 0.476] 0.5646 [0.1788- 1.783, 0.330]	n=67 1.041 [0.477- 2.273, 0.919] 0.550 [0.200- 1.516, 0.248] 0.141 [0.032- 0.620, **0.009****]	n=77 0.374 [0.125- 1.119, 0.079] 0.6814 [0.263- 1.764, 0.429] 0.283 [0.095- 0.847, **0.024***]
RAD50 ≤25th vs 26-50th ≤25th vs 51-75th ≤25th vs >75th percentile	n=71 1.2285 [0.577- 2.616, 0.594] 0.6223 [0.205- 1.892, 0.403] 3.7724 [0.809- 17.581, 0.091]	n=66 1.355 [0.569- 3.224, 0.492] 0.656 [0.253- 1.700, 0.385] 0.484 [0.167- 1.404, 0.182]	n=76 1.508 [0.390- 5.837, 0.552] 1.650 [0.459- 5.931, 0.443] 0.998 [0.264- 3.767, 0.998]
p21 ≤25th vs 26-50th 5th vs 51-75th ≤25th vs >75th percentile	n=75 2.569 [1.105- 5.974, **0.028***] 1.313 [0.500- 3.449, 0.581] 1.8584[0.706- 4.889, 0.209]	n=67 2.129 [0.936- 4.844, 0.072] 0.555 [0.186- 1.659, 0.292] 0.633 [0.170- 2.355, 0.495]	n=76 1.056 [0.366- 3.046, 0.919] 0.985 [0.331- 2.935, 0.979] 1.835 [0.331- 2.935, 0.251]
53BP1 ≤25th vs 26-50th ≤25th vs 51-75th ≤25th vs >75th percentile	n=74 0.697 [0.326- 1.489, 0.351] 0.703 [0.277- 1.787, 0.460] 5.061 [1.080- 23.717, **0.040***]	n= 67 1.347 [0.505- 3.595, 0.552] 0.840 [0.330- 2.140, 0.714] 0.480 [0.174- 1.328, 0.158]	n=77 0.613 [0.222- 1.694, 0.345] 0.951 [0.366- 2.471, 0.918] 0.4203 [0.137- 1.291, 0.130]
p53 ≤25th vs 26-50th ≤25th vs 51-75th ≤25th vs >75th percentile	n=73 6.165 [2.014- 18.871, **0.001*****] 1.428 [0.403- 5.063, 0.581] 2.635 [0.837- 8.301, 0.098]	n= 67 0.623 [0.226- 1.718, 0.361] 1.221 [0.518- 2.878, 0.648] 0.674 [0.244- 1.858, 0.445]	n=77 6.709 [1.899- 23.707, **0.003****] 2.6069 [0.690- 9.848, 0.158] 4.0597 [1.007- 16.366, 0.049]

NA*: estimates not available due to insufficient data

CK5/6 ≤25th vs 26-50th ≤25th vs 51-75th ≤25th vs >75th percentile	n=75 1.906 [0.411- 8.827, 0.410] 1.187 [0.265- 5.309, 0.823] 0.349 [0.320- 6.289, 0.646]	n=67 1.066 [0.484- 2.351, 0.874] 0.589 [0.169-2.054, 0.406] 0.780 [0.280- 2.168, 0.633]	n=77 0.762 [0.263- 2.206, 0.617] 0.556 [0.169- 1.824, 0.332] 0.8194 [0.297- 2.260, 0.700]
CK20 ≤25th vs 26-50th ≤25th vs 51-75th ≤25th vs >75th percentile	n=75 1.209 [0.542- 2.695, 0.643] 0.4941 [0.185- 1.319, 0.159] 0.4997 [0.187- 1.333, 0.166]	n=65 0.521 [0.203- 1.339, 0.176] 0.291 [0.114- 0.746, **0.010****] 0.169 [0.049- 0.581, **0.005***]	n=77 0.982 [0.344- 2.805, 0.973] 0.921 [0.333- 2.543, 0.873] 0.686 [0.241- 1.958, 0.482]
Tip60 ≤25th vs 26-50th ≤25th vs 51-75th ≤25th vs >75th percentile	n=76 0.512 [0.212- 1.233, 0.135] 0.756 [0.298- 1.918, 0.556] 0.679 [0.268- 1.723, 0.415]	n=68 1.609 [0.652- 3.967, 0.302] 0.803 [0.305- 2.111, 0.656] 1.538 [0.604- 3.915, 0.367]	n=75 0.852 [0.286- 2.537, 0.773] 0.593 [0.208- 1.694, 0.330] 0.705 [0.255- 1.951, 0.501]
MRE11 re-optimised ≤25th vs 26-50th ≤25th vs 51-75th ≤25th vs >75th percentile	n=58 1.354 [0.524- 3.501, 0.532] 1.092 [0.304- 3.923, 0.892] 1.731 [0.382- 7.839, 0.476]	n=62 0.965 [0.379- 2.455, 0.939] 0.643 [0.225- 1.840, 0.410] 0.411 [0.120- 1.407, 0.157]	n=74 2.114 [0.236- 18.944, 0.503] 1.034 [0.107- 9.959, 0.977] 1.293 [0.162- 10.352, 0.809]
MRE11 c terminal ≤25th vs 26-50th ≤25th vs 51-75th ≤25th vs >75th percentile	n=23 0.494 [0.138- 1.775, 0.280] NA*	n=62 0.714 [0.268- 1.904, 0.501] 0.498 [0.180- 1.377, 0.179] 1.258 [0.450- 3.520, 0.662]	n=33 NA*
Ki67 ≤25th vs 26-50th ≤25th vs 51-75th ≤25th vs >75th percentile	n= 74 0.293 [0.105- 0.819, **0.019***] 0.557 [0.247- 1.257, 0.159] 0.572 [0.229- 1.430, 0.232]	n=65 0.503 [0.175- 1.441, 0.200] 0.489 [0.154- 1.556, 0.226] 1.277 [0.508- 3.209, 0.603]	n=75 0.464 [0.189- 1.139, 0.094] 0.258 [0.074- 0.896, 0.033] 0.292 [0.084- 1.008, 0.052]

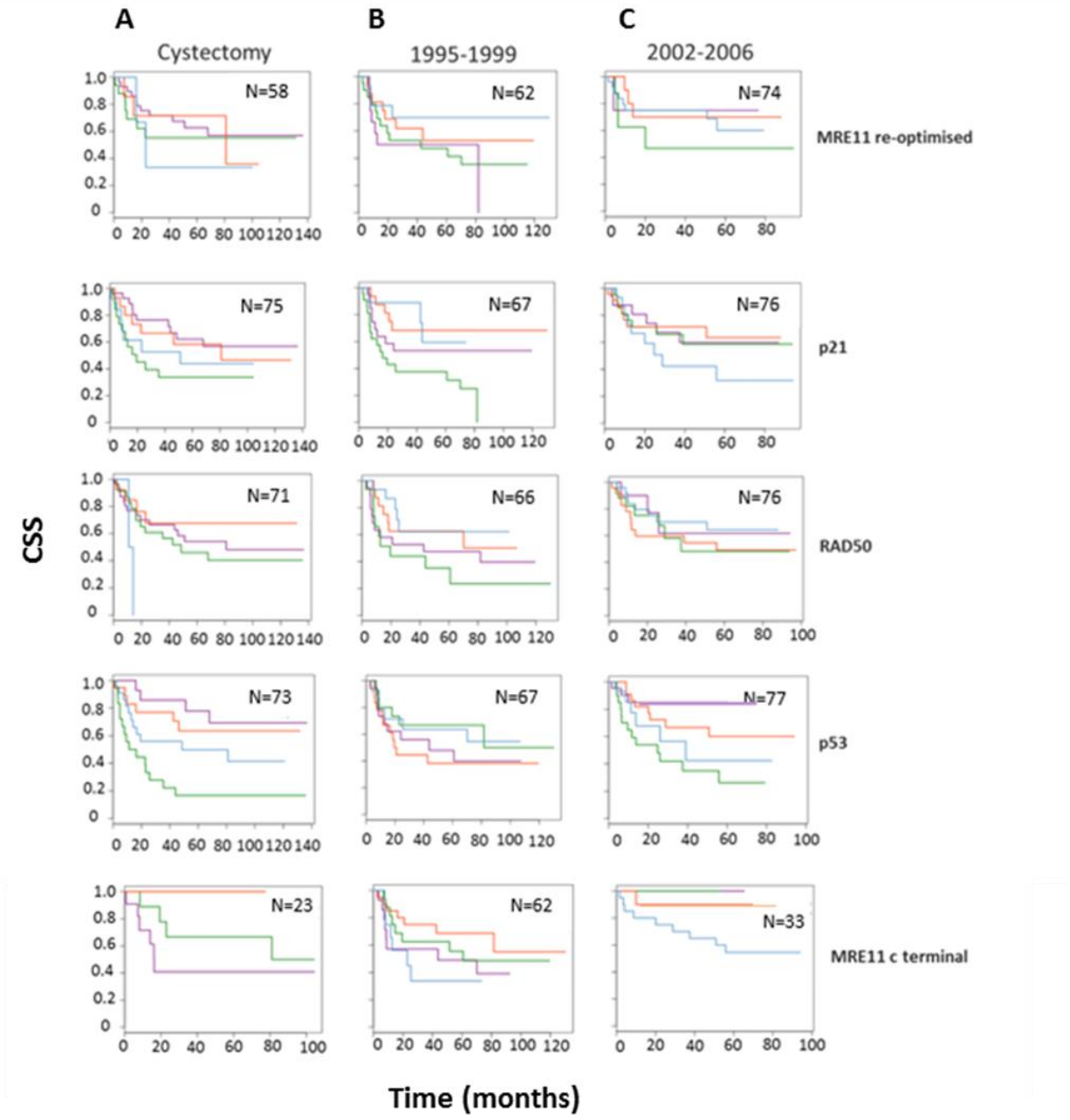
NA*: estimates not available due to insufficient data

Appendix J

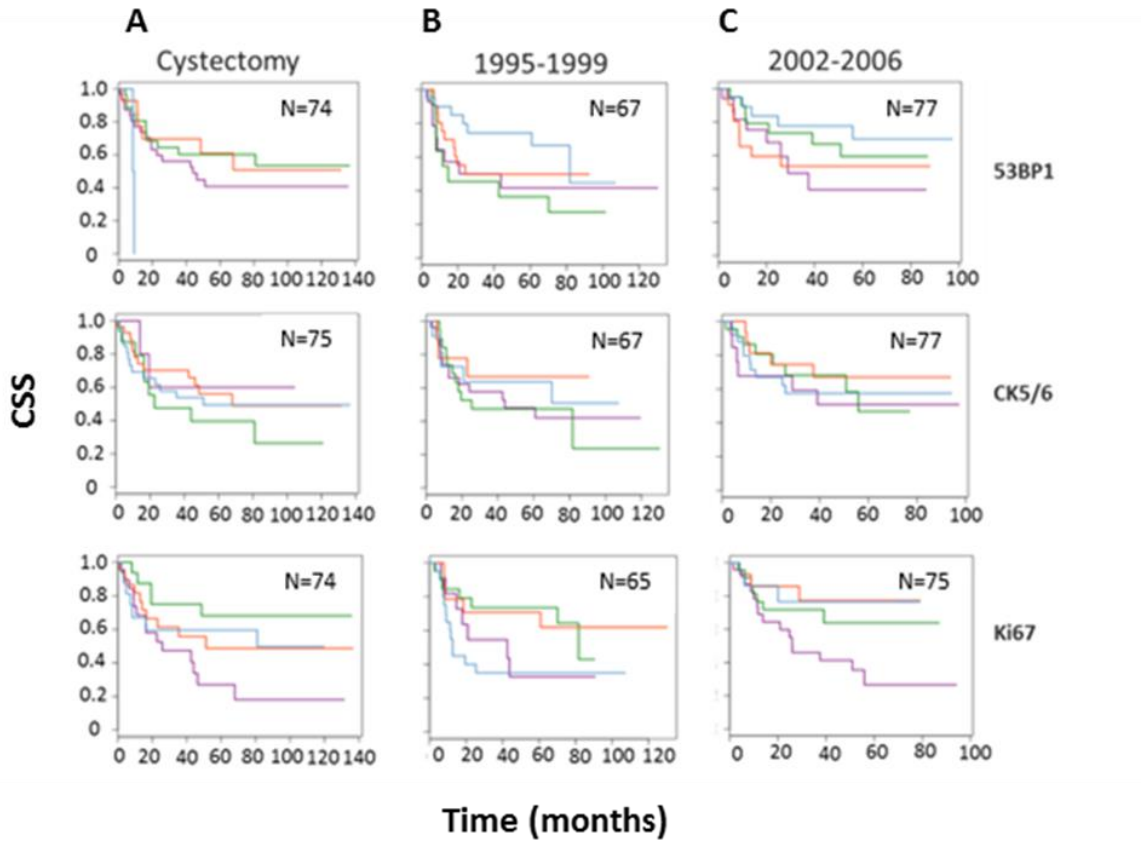
Univariate Cox proportional hazard modelling for all of the IHC stains scored by RTO by H-score unit increase. Hazard ratios, 95% confidence intervals and p values are display for the Cystectomy cohort, 1995-1999 RT cohort and 2002-2006 RT cohort. Significant p values are highlighted(*= p≤0.05, **= p≤0.01, ***=p≤0.001). The numerical unit increase in H-score was used as the variable in this analysis. Statistics performed by Dr Christiana Kartsonaki.

	Cystectomy cohort Hazard ratio, [95% confidence interval, p value]	**1995-1999 cohort** Hazard ratio, [95% confidence interval, p value]	**2002-2006 cohort** Hazard ratio, [95% confidence interval, p value]
MRE11 H-score (per unit increase)	n=75 0.997 [0.991- 1.004, 0.407]	n=67 0.991 [0.986- 0.997, **0.004****]	n=77 0.993 [0.987-1.00, 0.060]
RAD50 H-score (per unit increase)	n=71 1.003 [0.995-1.012, 0.437]	n=66 0.995 [0.989-1.002, 0.154]	n=76 0.998 [0.991-1.006, 0.662]
p21 H-score (per unit increase)	n=75 1.003 [0.994-1.012, 0.474]	n=67 0.990 [0.976-1.003, 0.133]	n=76 1.006 [0.997-1.015, 0.186]
53BP1 H-score (per unit increase)	n=74 0.997 [0.988-1.007, 0.570]	n= 67 0.995 [0.988-1.001, 0.097]	n=77 0.995 [0.988-1.002, 0.153]
p53 H-score (per unit increase)	n=73 0.999 [0.996-1.003, 0.662]	n= 67 1.000 [0.995-1.004, 0.912]	n=77 1.002 [0.997-1.007, 0.512]
CK5/6 H-score (per unit increase)	n=75 1.002 [0.997-1.007, 0.407]	n=67 1.001 [0.995-1.008, 0.746]	n=77 0.999 [0.992-1.006, 0.759]
CK20 H-score (per unit increase)	n=75 0.996 [0.993-1.000, **0.050 ***]	n=65 0.992 [0.988-0.997, **0.002****]	n=77 0.998 [0.994-1.001, 0.198]
Tip60 H-score (per unit increase)	n=76 0.997 [0.988-1.006, 0.505]	n=68 0.999 [0.990-1.009, 0.888]	n=75 1.000 [0.989-1.010, 0.949]
MRE11 re-optimised H-score (per unit increase)	n=58 1.002 [0.995-1.009, 0.648]	n=62 0.995 [0.989-1.001, 0.113]	n=74 1.000 [0.994-1.007, 0.909]
MRE11 c terminal H-score (per unit increase)	n=23 0.997 [0.986-1.009, 0.671]	n=62 0.999 [0.993-1.005, 0.761]	n=33 1.012 [0.995-1.029, 0.177]
Ki67 H-score (per unit increase)	n= 74 -0.994 [0.985-1.003, 0.210]	n=65 1.004 [0.997-1.012, 0.246]	n=75 0.991 [0.980-1.002, 0.113]

Appendix K



Kaplan-Meier CSS survival curves for the re-optimised MRE11, p21, RAD50, p53 and MRE11 C terminal IHC stains, scored by RTO. Plots within column A, B and C display the results for the cystectomy, 1995-1999 RT and 2002-2006 RT cohorts, respectively. The ≤25th (purple), 26-50th (green), 51-75th (red) and >75th (blue) percentiles of the IHC scoring for each stain were used to stratify patients within the curves. The survival statistics for each of the stained cohorts in this figure are contained in Appendix H. Statistics performed by Dr Christiana Kartsonaki.



Kaplan-Meier CSS survival curves for 53BP1, CK5/6 and KI67 HC stains, scored by RTO. Plots within column A, B and C display the results for the cystectomy, 1995-1999 RT and 2002-2006 RT cohorts, respectively. The ≤25th (purple), 26-50th (green), 51-75th (red) and >75th (blue) percentiles of the IHC scoring for each stain were used to stratify patients within the curves. The survival statistics for each of the stained cohorts in this figure are contained in Appendix H. Statistics performed by Dr Christiana Kartsonaki.

References

Abbas T, Dutta A (2009) p21 in cancer: intricate networks and multiple activities. *Nat Rev Cancer* 9: 400–414, doi:10.1038/nrc2657.

Abrahao-Machado LF, Scapulatempo-Neto C (2016) HER2 testing in gastric cancer: An update. *World J Gastroenterol* 22: 4619–4625, doi:10.3748/wjg.v22.i19.4619.

Advanced Bladder Cancer Meta-analysis Collaboration (2003) Neoadjuvant chemotherapy in invasive bladder cancer: a systematic review and meta-analysis. *Lancet (London, England)* 361: 1927–1934.

Aeffner F, Wilson K, Martin NT, *et al.* (2017) The Gold Standard Paradox in Digital Image Analysis: Manual Versus Automated Scoring as Ground Truth. *Arch Pathol Lab Med* 141: 1267–1275, doi:10.5858/arpa.2016-0386-RA.

Agerbaek M, Alsner J, Marcussen N, *et al.* (2006) Focal S100A4 Protein Expression Is an Independent Predictor of Development of Metastatic Disease in Cystectomized Bladder Cancer Patients. *Eur Urol* 50: 777–785, doi:10.1016/j.eururo.2006.02.027.

Ahnesorg P, Smith P, Jackson SP (2006) XLF Interacts with the XRCC4-DNA Ligase IV Complex to Promote DNA Nonhomologous End-Joining. *Cell* 124: 301–313, doi:10.1016/j.cell.2005.12.031.

Albain KS, Barlow WE, Shak S, *et al.* (2010) Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol* 11: 55–65, doi:10.1016/S1470-2045(09)70314-6.

Alfred Witjes J, Lebret T, Compérat EM, *et al.* (2017) Updated 2016 EAU Guidelines on Muscle-invasive and Metastatic Bladder Cancer. *Eur Urol* 71: 462–475, doi:10.1016/j.eururo.2016.06.020.

Allott EH, Geradts J, Sun X, *et al.* (2016) Intratumoral heterogeneity as a source of discordance in breast cancer biomarker classification. *Breast Cancer Res* 18: 68, doi:10.1186/s13058-016-0725-1.

Alymani NA, Smith MD, Williams DJ, Petty RD (2010) Predictive biomarkers for personalised anti-cancer drug use: Discovery to clinical implementation. *Eur J Cancer* 46: 869–879, doi:10.1016/j.ejca.2010.01.001.

Amado RG, Wolf M, Peeters M, *et al.* (2008) Wild-Type *KRAS* Is Required for Panitumumab Efficacy in Patients With Metastatic Colorectal Cancer. *J Clin Oncol* 26: 1626–1634, doi:10.1200/JCO.2007.14.7116.

Amin MB (2009) Histological variants of urothelial carcinoma: diagnostic, therapeutic and prognostic implications. *Mod Pathol* 22: S96–S118, doi:10.1038/modpathol.2009.26.

Anacker DC, Gautam D, Gillespie KA, *et al.* (2014) Productive replication of human papillomavirus 31 requires DNA repair factor Nbs1. *J Virol* 88: 8528–8544, doi:10.1128/JVI.00517-14.

Anacker DC, Moody CA (2016) Modulation of the DNA damage response during the life cycle of human papillomaviruses. *Virus Res* doi:10.1016/j.virusres.2016.11.006.

Anand R, Ranjha L, Cannavo E, Cejka P (2016) Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Resection. *Mol Cell* 64: 940–950, doi:10.1016/J.MOLCEL.2016.10.017.

Antoni S, Ferlay J, Soerjomataram I, *et al.* (2017) Bladder Cancer Incidence and Mortality: A Global Overview and Recent Trends. *Eur Urol* 71: 96–108, doi:10.1016/j.eururo.2016.06.010.

Arnaudeau C, Lundin C, Helleday T (2001) DNA double-strand breaks associated with replication forks are predominantly repaired by homologous recombination involving an exchange mechanism in mammalian cells11Edited by J. Karn. *J Mol Biol* 307: 1235–1245, doi:10.1006/jmbi.2001.4564.

de Azambuja E, Cardoso F, de Castro G, *et al.* (2007) Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. *Br J Cancer* 96: 1504–1513, doi:10.1038/sj.bjc.6603756.

Aziz A, May M, Burger M, *et al.* (2014) Prediction of 90-day Mortality After Radical Cystectomy for Bladder Cancer in a Prospective European Multicenter Cohort. *Eur Urol* 66: 156–163, doi:10.1016/j.eururo.2013.12.018.

Babjuk M, Böhle A, Burger M, *er al.* (2017) EAU Guidelines on Non–Muscle-invasive Urothelial Carcinoma of the Bladder: Update 2016. *Eur Urol* 71: 447–461, doi:10.1016/j.eururo.2016.05.041.

Barbagallo GM V., Paratore S, Caltabiano R, *et al.* (2014) Long-term therapy with temozolomide is a feasible option for newly diagnosed glioblastoma: a single-institution experience with as many as 101 temozolomide cycles. *Neurosurg Focus* 37: E4, doi:10.3171/2014.9.FOCUS14502.

Baricevic I, He X, Chakrabarty B, Oliver AW, *et al.* (2015) High-sensitivity human papilloma virus genotyping reveals near universal positivity in anal squamous cell carcinoma: Different implications for vaccine prevention and prognosis. *Eur J Cancer* 51: 776–785, doi:10.1016/j.ejca.2015.01.058.

Bartkova J, Tommiska J, Oplustilova L, *et al.* (2008) Aberrations of the MRE11–RAD50–NBS1 DNA damage sensor complex in human breast cancer: MRE11 as a candidate familial cancer-predisposing gene. *Mol Oncol* 2: 296–316, doi:10.1016/J.MOLONC.2008.09.007.

Bedford MT, Clarke SG (2009) Protein arginine methylation in mammals: who, what, and why. *Mol Cell* 33: 1–13, doi:10.1016/j.molcel.2008.12.013.

Reprinted from *Mol cell*, 33. Bedford MT, Clarke SG. Protein arginine methylation in mammals: who, what, and why. 1-13 (2009) with permission from Elsevier. doi:10.1016/j.molcel.2008.12.013.

Bentzen AG, Balteskard L, Wanderås EH, *et al.* (2013) Impaired health-related quality of life after chemoradiotherapy for anal cancer: Late effects in a national cohort of 128 survivors. *Acta Oncol (Madr)* 52: 736–744, doi:10.3109/0284186X.2013.770599.

Berger I, Martini T, Wehrberger C, *et al.* (2014) Perioperative complications and 90-day mortality of radical cystectomy in the elderly (75+): a retrospective, multicentre study. *Urol Int* 93: 296–302, doi:10.1159/000357127.

Berliner JL, Fay AM, Practice Issues Subcommittee of the National Society of Genetic Counselors' Familial Cancer Risk Counseling Special Interest Group (2007) Risk assessment and genetic counseling for hereditary breast and ovarian cancer: recommendations of the National Society of Genetic Counselors. *J Genet Couns* 16: 241–260, doi:10.1007/s10897-007-9090-7.

Bétermier M, Bertrand P, Lopez BS, Hendrickson E, Ramsden D (2014) Is Non-Homologous End-Joining Really an Inherently Error-Prone Process? *PLoS Genet* 10: e1004086, doi:10.1371/journal.pgen.1004086.

Bi L, Huang H, Fan X, *et al.* (2014) Extended vs non-extended pelvic lymph node dissection and their influence on recurrence-free survival in patients undergoing radical cystectomy for bladder cancer: a systematic review and meta-analysis of comparative studies. *BJU Int* 113: E39–E48, doi:10.1111/bju.12371.

Binabaj MM, Bahrami A, ShahidSales S, *et al.* (2018) The prognostic value of MGMT promoter methylation in glioblastoma: A meta-analysis of clinical trials. *J Cell Physiol* 233: 378–386, doi:10.1002/jcp.25896.

Boisvert F-M, Déry U, Masson J-Y, Richard S (2005a) Arginine methylation of MRE11 by PRMT1 is required for DNA damage checkpoint control. *Genes Dev* 19: 671–676, doi:10.1101/gad.1279805.

Boisvert F-M, Hendzel MJ, Masson J-Y, Richard S (2005b) Methylation of MRE11 Regulates its Nuclear Compartmentalization. *Cell Cycle* 47: 981–989.

Booth CM, Siemens DR, Li G, *et al.* (2014) Curative Therapy for Bladder Cancer in Routine Clinical Practice: A Population-based Outcomes Study. *Clin Oncol* 26: 506–514, doi:10.1016/j.clon.2014.05.007.

Boulet G, Horvath C, Vanden Broeck D *et al.* (2007) Human papillomavirus: E6 and E7 oncogenes. *Int J Biochem Cell Biol* 39: 2006–2011, doi:10.1016/j.biocel.2007.07.004.

Boyer SN, Wazer DE, Band V (1996) E7 Protein of Human Papilloma Virus-16 Induces Degradation of Retinoblastoma Protein through the Ubiquitin-Proteasome Pathway. *Cancer Res* 56:.

Brait M, Begum S, Carvalho AL, Dasgupta S, Vettore AL, Czerniak B, Caballero OL, Westra WH, Sidransky D, Hoque MO (2008) Aberrant promoter methylation of multiple genes during pathogenesis of bladder cancer. *Cancer Epidemiol Biomarkers Prev* 17: 2786–2794, doi:10.1158/1055-9965.EPI-08-0192.

Brandt S, Samartzis EP, Zimmermann A-K, *et al.* (2017) Lack of MRE11-RAD50-NBS1 (MRN) complex detection occurs frequently in low-grade epithelial ovarian cancer. *BMC Cancer* 17: 44, doi:10.1186/s12885-016-3026-2.

Brehm A, Nielsen SJ, Miska EA, *et al.* (1999) The E7 oncoprotein associates with Mi2 and histone deacetylase activity to promote cell growth. *EMBO J* 18: 2449–2458.

Bremang M, Cuomo A, Agresta AM, *et al.* (2013) Mass spectrometry-based identification and characterisation of lysine and arginine methylation in the human proteome. *Mol Biosyst* 9: 2231, doi:10.1039/c3mb00009e.

Brouwer I, Sitters G, Candelli A, *et al.* (2016) Sliding sleeves of XRCC4–XLF bridge DNA and connect fragments of broken DNA. *Nature* 535: 566–569, doi:10.1038/nature18643.

Bruins HM, Veskimae E, Hernandez V, *et al.*(2014) The Impact of the Extent of Lymphadenectomy on Oncologic Outcomes in Patients Undergoing Radical Cystectomy for Bladder Cancer: A Systematic Review. *Eur Urol* 66: 1065–1077, doi:10.1016/j.eururo.2014.05.031.

Bryant HE, Schultz N, Thomas HD, *et al* (2005) Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* 434: 913–917, doi:10.1038/nature03443.

Buis J, Wu Y, Deng Y, *et al.* (2008) MRE11 nuclease activity has essential roles in DNA repair and genomic stability distinct from ATM activation. *Cell* 135: 85–96, doi:10.1016/j.cell.2008.08.015.

Burger M, Catto JWF, Dalbagni G, *et al.* (2013) Epidemiology and Risk Factors of Urothelial Bladder Cancer. *Eur Urol* 63: 234–241, doi:10.1016/J.EURURO.2012.07.033.

Burrell RA, McGranahan N, Bartek J, Swanton C (2013) The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* 501: 338–345, doi:10.1038/nature12625.

Caffo O, Fellin G, Graffer U, Luciani L (1996) Assessment of quality of life after cystectomy or conservative therapy for patients with infiltrating bladder carcinoma: A survey by a self-administered questionnaire. *Cancer* 78: 1089–1097, doi:10.1002/(SICI)1097-0142(19960901)78:5<1089::AID-CNCR20>3.0.CO;2-Y.

Camp RL, Neumeister V, Rimm DL (2008) A decade of tissue microarrays: progress in the discovery and validation of cancer biomarkers. *J Clin Oncol* 26: 5630–5637, doi:10.1200/JCO.2008.17.3567.

Campisi J (2013) Aging, cellular senescence, and cancer. *Annu Rev Physiol* 75: 685–705, doi:10.1146/annurev-physiol-030212-183653.

Cancer Research UK, a. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bladder-cancer#heading-Zero (accessed: 09/03/2017).

Cancer Research UK, b. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/anal-cancer#heading-Zero (accessed: 13/01/2017).

Cancer Research UK (2015) http://www.cancerresearchuk.org/about-cancer/bladder-cancer/types-stages-grades/stages (accessed: 27/03/2017).

Cancer Research UK (2017) http://www.cancerresearchuk.org/about-cancer/anal-cancer/stages-types/tnm-staging (accessed: 27/03/2017).

Candido Dos Reis FJ, Lynn S, Ali HR, *et al.* (2015) Crowdsourcing the General Public for Large Scale Molecular Pathology Studies in Cancer. *EBioMedicine* 2: 681–689, doi:10.1016/j.ebiom.2015.05.009.

Carson CT, Schwartz RA, Stracker TH, *et al.* (2003) The Mre11 complex is required for ATM activation and the G2/M checkpoint. *EMBO J* 22: 6610–6620, doi:10.1093/emboj/cdg630.

Carvajal-Hausdorf DE, Schalper KA, Neumeister VM, Rimm DL (2015) Quantitative measurement of cancer tissue biomarkers in the lab and in the clinic. *Lab Invest* 95: 385–396, doi:10.1038/labinvest.2014.157.

Di Cataldo S, Ficarra E, Macii E (2012) Computer-aided techniques for chromogenic immunohistochemistry: Status and directions. *Comput Biol Med* 42: 1012–1025, doi:10.1016/j.compbiomed.2012.08.004.

Cerosaletti K, Concannon P (2004) Independent roles for nibrin and Mre11-Rad50 in the activation and function of Atm. *J Biol Chem* 279: 38813–38819, doi:10.1074/jbc.M404294200.

Chahal R, Sundaram SK, Iddenden R, *et al.* (2003) A study of the morbidity, mortality and long-term survival following radical cystectomy and radical radiotherapy in the treatment of invasive bladder cancer in Yorkshire. *Eur Urol* 43: 246–257.

Chang HHY, Watanabe G, Gerodimos CA, *et al.* (2016) Different DNA End Configurations Dictate Which NHEJ Components Are Most Important for Joining Efficiency. *J Biol Chem* 291: 24377–24389, doi:10.1074/jbc.M116.752329.

Chaudhry P, Srinivasan R, Patel FD (2009) Utility of Gene Promoter Methylation in Prediction of Response to Platinum-Based Chemotherapy in Epithelial Ovarian Cancer (EOC). *Cancer Invest* 27: 877–884, doi:10.1080/07357900902849699.

Chen L, Gilkes DM, Pan Y, *et al.* (2005) ATM and Chk2-dependent phosphorylation of MDMX contribute to p53 activation after DNA damage. *EMBO J* 24: 3411–3422, doi:10.1038/sj.emboj.7600812.

Choi W, Czerniak B, Ochoa A, *et al.* (2014a) Intrinsic basal and luminal subtypes of muscle-invasive bladder cancer. *Nat Rev Urol* 11: 400–410, doi:10.1038/nrurol.2014.129.

Choi W, Porten S, Kim S, *et al.* (2014b) Identification of distinct basal and luminal subtypes of muscle-invasive bladder cancer with different sensitivities to frontline chemotherapy. *Cancer Cell* 25: 152–165, doi:10.1016/j.ccr.2014.01.009.

Choudhury A, Elliott F, Iles MM, *et al.* (2008) Analysis of variants in DNA damage signalling genes in bladder cancer. *BMC Med Genet* 9: 69, doi:10.1186/1471-2350-9-69.

Choudhury A, Nelson LD, Teo MTW, *et al.* (2010) MRE11 expression is predictive of cause-specific survival following radical radiotherapy for muscle-invasive bladder cancer. *Cancer Res* 70: 7017–7026, doi:10.1158/0008-5472.CAN-10-1202.

Ciccia A, Constantinou A, West SC (2003) Identification and Characterization of the Human Mus81-Eme1 Endonuclease. *J Biol Chem* 278: 25172–25178, doi:10.1074/jbc.M302882200.

Cooper S, Khatib F, Treuille A, *et al.* (2010) Predicting protein structures with a multiplayer online game. doi:10.1038/nature09304.

Copson ER, Maishman TC, Tapper WJ, *et al.* (2018) Germline BRCA mutation and outcome in young-onset breast cancer (POSH): a prospective cohort study. *Lancet Oncol* 0: doi:10.1016/S1470-2045(17)30891-4.

Corn PG, Smith BD, Ruckdeschel ES, *et al.* (2000) E-cadherin expression is silenced by 5' CpG island methylation in acute leukemia. *Clin Cancer Res* 6: 4243–4248, doi:10.1158/1078-0432.ccr-04-0175.

Cortez D (2015) Preventing replication fork collapse to maintain genome integrity. *DNA Repair (Amst)* 32: 149–157, doi:10.1016/j.dnarep.2015.04.026.

Costa RLB, Gradishar WJ (2017) Triple-Negative Breast Cancer: Current Practice and Future Directions. *J Oncol Pract* 13: 301–303, doi:10.1200/JOP.2017.023333.

Curtis V (2014) Online citizen science games: Opportunities for the biological sciences. *Appl Transl Genomics* 3: 90–94, doi:10.1016/j.atg.2014.07.001.

Curtis V (2015) Motivation to Participate in an Online Citizen Science Game: A Study of Foldit. *Sci Commun* 37: 723–746, doi:10.1177/1075547015609322.

Van Cutsem E, Cervantes A, Adam R, *et al.* (2016) ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann Oncol* 27: 1386–1422, doi:10.1093/annonc/mdw235.

D’Amours D, Jackson SP (2002) The mre11 complex: at the crossroads of dna repair and checkpoint signalling. *Nat Rev Mol Cell Biol* 3: 317–327, doi:10.1038/nrm805.

Dadhania V, Zhang M, Zhang L, *et al.* (2016) Meta-Analysis of the Luminal and Basal Subtypes of Bladder Cancer and the Identification of Signature Immunohistochemical Markers for Clinical Use. *EBioMedicine* 12: 105–117, doi:10.1016/j.ebiom.2016.08.036.

Damiola F, Pertesi M, Oliver J, *et al.* (2014) Rare key functional domain missense substitutions in MRE11A, RAD50, and NBNcontribute to breast cancer susceptibility: results from a Breast Cancer Family Registry case-control mutation-screening study. *Breast Cancer Res* 16: R58, doi:10.1186/bcr3669.
https://creativecommons.org/licenses/by/2.0/legalcode

Damrauer JS, Hoadley KA, Chism DD, *et al.* (2014) Intrinsic subtypes of high-grade bladder cancer reflect the hallmarks of breast cancer biology. *Proc Natl Acad Sci* 111: 3110–3115, doi:10.1073/pnas.1318376111.

Darnell GA, Schroder WA, Antalis TM, *et al.* (2007) Human papillomavirus E7 requires the protease calpain to degrade the retinoblastoma protein. *J Biol Chem* 282: 37492–37500, doi:10.1074/jbc.M706860200.

Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. *Genes Dev* 25: 1010–1022, doi:10.1101/gad.2037511.

Déry U, Coulombe Y, Rodrigue A, *et al.* (2008) A glycine-arginine domain in control of the human MRE11 DNA repair protein. *Mol Cell Biol* 28: 3058–3069, doi:10.1128/MCB.02025-07.

Deshpande RA, Lee J-H, Arora S, *et al.* (2016) Nbs1 Converts the Human Mre11/Rad50 Nuclease Complex into an Endo/Exonuclease Machine Specific for Protein-DNA Adducts. *Mol Cell* 64: 593–606, doi:10.1016/j.molcel.2016.10.010.

Dhar NB, Klein EA, Reuther AM, *et al.* (2008) Outcome After Radical Cystectomy With Limited or Extended Pelvic Lymph Node Dissection. *J Urol* 179: 873–878, doi:10.1016/j.juro.2007.10.076.

Diamandis EP (2012) The failure of protein cancer biomarkers to reach the clinic: why, and what can be done to address the problem? *BMC Med* 10: 87, doi:10.1186/1741-7015-10-87.

DiCarlo JJ, Zoccolan D, Rust NC (2012) How does the brain solve visual object recognition? *Neuron* 73: 415–434, doi:10.1016/j.neuron.2012.01.010.

Divrik RT, Yildirim Ü, Zorlu F, Özen H (2006) The Effect of Repeat Transurethral Resection on Recurrence and Progression Rates in Patients With T1 Tumors of the Bladder Who Received Intravesical Mitomycin: A Prospective, Randomized Clinical Trial. *J Urol* 175: 1641–1644, doi:10.1016/S0022-5347(05)01002-5.

Djaladat H, Katebian B, Bazargani ST, *et al.* (2016) 90-Day complication rate in patients undergoing radical cystectomy with enhanced recovery protocol: a prospective cohort study. *World J Urol* doi:10.1007/s00345-016-1950-z.

Dmitrieva NI, Malide D, Burg MB (2011) Mre11 is expressed in mammalian mitochondria where it binds to mitochondrial DNA. *Am J Physiol Regul Integr Comp Physiol* 301: R632-40, doi:10.1152/ajpregu.00853.2010.

Doll CM, Winter K, Ajani J, *et al.* (2014) The Significance of p16 and p53 Expression on Clinical Outcome in Patients with Anal Cancer Treated with Chemoradiation Therapy: An Analysis of RTOG 98-11. *Int J Radiat Oncol* 90: S32, doi:10.1016/j.ijrobp.2014.05.140.

Domchek SM, Friebel TM, Singer CF, *et al.* (2010) Association of risk-reducing surgery in BRCA1 or BRCA2 mutation carriers with cancer risk and mortality. *JAMA* 304: 967–975, doi:10.1001/jama.2010.1237.

Douillard J-Y, Siena S, Cassidy J, *et al.* (2010) Randomized, phase III trial of panitumumab with infusional fluorouracil, leucovorin, and oxaliplatin (FOLFOX4) versus FOLFOX4 alone as first-line treatment in patients with previously untreated metastatic colorectal cancer: the PRIME study. *J Clin Oncol* 28: 4697–4705, doi:10.1200/JCO.2009.27.4860.

Drabovich AP, Martínez-Morillo E, Diamandis EP (2015) Toward an integrated pipeline for protein biomarker development. *Biochim Biophys Acta - Proteins Proteomics* 1854: 677–686, doi:10.1016/j.bbapap.2014.09.006.

Dreyer G (2012) Screening for gynaecologic cancers in genetically predisposed women. *Best Pract Res Clin Obstet Gynaecol* 26: 267–282, doi:10.1016/j.bpobgyn.2011.11.002.

Drucker E, Krapfenbauer K (2013) Pitfalls and limitations in translation from biomarker discovery to clinical utility in predictive and personalised medicine. *EPMA J* 4: 7, doi:10.1186/1878-5085-4-7.

Dudley WN, Wickham R, Coombs N (2016) An Introduction to Survival Statistics: Kaplan-Meier Analysis. *J Adv Pract Oncol* 7: 91–100.

Duensing S, Münger K (2002) The Human Papillomavirus Type 16 E6 and E7 Oncoproteins Independently Induce Numerical and Structural Chromosome Instability. *Cancer Res* 62:.

Duffy MJ, Harbeck N, Nap M, *et al.* (2017) Clinical use of biomarkers in breast cancer: Updated guidelines from the European Group on Tumor Markers (EGTM). *Eur J Cancer* 75: 284–298, doi:10.1016/j.ejca.2017.01.017.

Dunstan RW, Wharton KA, Quigley C, Lowe A (2011) The Use of Immunohistochemistry for Biomarker Assessment--Can It Compete with Other Technologies? *Toxicol Pathol* 39: 988–1002, doi:10.1177/0192623311419163.

Dupré A, Boyer-Chatenet L, Gautier J (2006) Two-step activation of ATM by DNA and the Mre11–Rad50–Nbs1 complex. *Nat Struct Mol Biol* 13: 451–457, doi:10.1038/nsmb1090.

Eckel-Passow JE, Lohse CM, Sheinin Y, *et al.* (2010) Tissue microarrays: one size does not fit all. *Diagn Pathol* 5: 48, doi:10.1186/1746-1596-5-48.

Efstathiou JA, Spiegel DY, Shipley WU, *et al.* (2012) Long-Term Outcomes of Selective Bladder Preservation by Combined-Modality Therapy for Invasive Bladder Cancer: The MGH Experience. *Eur Urol* 61: 705–711, doi:10.1016/j.eururo.2011.11.010.

el-Deiry WS, Tokino T, Velculescu VE, *et al.* (1993) WAF1, a potential mediator of p53 tumor suppression. *Cell* 75: 817–825.

Elbendary AA, Cirisano FD, Evans AC, *et al.* (1996) Relationship between p21 expression and mutation of the p53 tumor suppressor gene in normal and malignant ovarian epithelial cells. *Clin Cancer Res* 2: 1571–1575.

Eriksson D, Stigbrand T (2010) Radiation-induced cell death mechanisms. *Tumor Biol* 31: 363–372, doi:10.1007/s13277-010-0042-8.

Esteller M, Corn PG, Urena JM, *et al.* (1998) Inactivation of glutathione S-transferase P1 gene by promoter hypermethylation in human neoplasia. *Cancer Res* 58: 4515–4518.

Euser AM, Zoccali C, Jager KJ, Dekker FW (2009) Cohort studies: prospective versus retrospective. *Nephron Clin Pract* 113: c214-7, doi:10.1159/000235241.

Evans T, Matulonis U (2017) PARP inhibitors in ovarian cancer: evidence, experience and clinical potential. *Ther Adv Med Oncol* 9: 253–267, doi:10.1177/1758834016687254.

EVE Online: Project Discovery https://www.eveonline.com/discovery/ (accessed: 21/02/2017).

Everhard S, Tost J, El Abdalaoui H, *et al.* (2009) Identification of regions correlating MGMT promoter methylation and gene expression in glioblastomas. *Neuro Oncol* 11: 348–356, doi:10.1215/15228517-2009-001.

Fajkovic H, Halpern JA, Cha EK, *et al.*(2011) Impact of gender on bladder cancer incidence, staging, and prognosis. *World J Urol* 29: 457–463, doi:10.1007/s00345-011-0709-9.

Farmer H, McCabe N, Lord CJ, *et al.* (2005) Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* 434: 917–921, doi:10.1038/nature03445.

Fehrmann F, Laimins LA (2003) Human papillomaviruses: targeting differentiating epithelial cells for malignant transformation. *Oncogene* 22: 5201–5207, doi:10.1038/sj.onc.1206554.

Fell VL, Schild-Poulter C (2015) The Ku heterodimer: Function in DNA repair and beyond. *Mutat Res Mutat Res* 763: 15–29, doi:10.1016/j.mrrev.2014.06.002.

Forker LJ, Choudhury A, Kiltie AE (2015) Biomarkers of Tumour Radiosensitivity and Predicting Benefit from Radiotherapy Statement of Search Strategies Used and Sources of Information. *Clin Oncol* 27: 561–569, doi:10.1016/j.clon.2015.06.002.

Fraga MF, Ballestar E, Villar-Garea A, *et al.* (2005) Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat Genet* 37: 391–400, doi:10.1038/ng1531.

Friedrich T, Leong S, Lieu CH (2016) Beyond RAS and BRAF: a target rich disease that is ripe for picking. *J Gastrointest Oncol* 7: 705–712, doi:10.21037/jgo.2016.06.11.

Fukuda T, Sumiyoshi T, Takahashi M, *et al.* (2001) Alterations of the double-strand break repair gene MRE11 in cancer. *Cancer Res* 61: 23–26.

Funk JO, Waga S, Harry JB, *et al.* (1997) Inhibition of CDK activity and PCNA-dependent DNA replication by p21 is blocked by interaction with the HPV-16 E7 oncoprotein. *Genes Dev* 11: 2090–2100.

Gakis G, Efstathiou J, Lerner SP, *et al.* (2013) ICUD-EAU International Consultation on Bladder Cancer 2012: Radical Cystectomy and Bladder Preservation for Muscle-Invasive Urothelial Carcinoma of the Bladder. *Eur Urol* 63: 45–57, doi:10.1016/j.eururo.2012.08.009.

Ganguly N, Parihar SP (2009) Human papillomavirus E6 and E7 oncoproteins as risk factors for tumorigenesis. *J Biosci* 34: 113–123.

Gannon HS, Woda BA, Jones SN (2012) ATM Phosphorylation of Mdm2 Ser394 Regulates the Amplitude and Duration of the DNA Damage Response in Mice. *Cancer Cell* 21: 668–679, doi:10.1016/j.ccr.2012.04.011.

Garcia V, Phelps SEL, Gray S, Neale MJ (2011) Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. *Nature* 479: 241–244, doi:10.1038/nature10515.

Garrison LP, Babigumira JB, Masaquel A, *et al.* (2015) The Lifetime Economic Burden of Inaccurate HER2 Testing: Estimating the Costs of False-Positive and False-Negative HER2 Test Results in US Patients with Early-Stage Breast Cancer. *Value Heal* 18: 541–546, doi:10.1016/J.JVAL.2015.01.012.

Gerlinger M, Catto JW, Orntoft TF, *et al.* (2015) Intratumour Heterogeneity in Urologic Cancers: From Molecular Evidence to Clinical Implications. *Eur Urol* 67: 729–737, doi:10.1016/j.eururo.2014.04.014.

Geybels MS, Zhao S, Wong C-J, *et al.* (2015) Epigenomic profiling of DNA methylation in paired prostate cancer versus adjacent benign tissue. *Prostate* 75: 1941–1950, doi:10.1002/pros.23093.

Ghosn M, Kourie HR, Abdayem P, *et al.* (2015) Anal cancer treatment: Current status and future perspectives. *World J Gastroenterol* 21: 2294–2302, doi:10.3748/wjg.v21.i8.2294.

Giannakeas V, Narod SA (2017) The expected benefit of preventive mastectomy on breast cancer incidence and mortality in BRCA mutation carriers, by age at mastectomy. *Breast Cancer Res Treat* doi:10.1007/s10549-017-4476-1.

Giannini G, Ristori E, Cerignoli F, *et al.* (2002) Human MRE11 is inactivated in mismatch repair-deficient cancers. *EMBO Rep* 3: 248–254, doi:10.1093/embo-reports/kvf044.

Giannopoulou L, Chebouti I, Pavlakis K, *et al.* (2017) RASSF1A promoter methylation in high-grade serous ovarian cancer: A direct comparison study in primary tumors, adjacent morphologically tumor cell-free tissues and paired circulating tumor DNA. *Oncotarget* doi:10.18632/oncotarget.15249.

Gilbert DC, Williams A, Allan K, *et al.* (2013) P16INK4A, p53, EGFR expression and KRAS mutation status in squamous cell cancers of the anus: Correlation with outcomes following chemo-radiotherapy. *Radiother Oncol* 109: 146–151, doi:10.1016/j.radonc.2013.08.002.

Gillett CE, Springall RJ, Barnes DM, Hanby AM (2000) Multiple tissue core arrays in histopathology research: a validation study. *J Pathol* 192: 549–553, doi:10.1002/1096-9896(2000)9999:9999<::AID-PATH721>3.0.CO;2-0.

Gluz O, Nitz UA, Christgen M, *et al.* (2016) West German Study Group Phase III PlanB Trial: First Prospective Outcome Data for the 21-Gene Recurrence Score Assay and Concordance of Prognostic Markers by Central and Local Pathology Assessment. *J Clin Oncol* 34: 2341–2349, doi:10.1200/JCO.2015.63.5383.

Glynne-Jones R, Nilsson PJ, Aschele C, *et al.* (2014) Anal cancer: ESMO-ESSO-ESTRO clinical practice guidelines for diagnosis, treatment and follow-up. *Radiother Oncol* 111: 330–339, doi:10.1016/j.radonc.2014.04.013.

Glynne-Jones R, Northover J, Oliveira J, *et al.* (2009) Anal cancer: ESMO clinical recommendations for diagnosis, treatment and follow-up. *Ann Oncol* 20 Suppl 4: 57–60, doi:10.1093/annonc/mdp129.

Goebell PJ, Knowles MA (2010) Bladder cancer or bladder cancers? Genetically distinct malignant conditions of the urothelium. *Urol Oncol Semin Orig Investig* 28: 409–428, doi:10.1016/j.urolonc.2010.04.003.

Gofrit ON, Yutkin V, Shapiro A, *et al.* (2016) The Response of Variant Histology Bladder Cancer to Intravesical Immunotherapy Compared to Conventional Cancer. *Front Oncol* 6: 43, doi:10.3389/fonc.2016.00043.

Gonzalez-Angulo AM, Litton JK, Broglio KR, *et al.* (2009) High risk of recurrence for patients with breast cancer who have human epidermal growth factor receptor 2-positive, node-negative tumors 1 cm or smaller. *J Clin Oncol* 27: 5700–5706, doi:10.1200/JCO.2009.23.2025.

Goossens N, Nakagawa S, Sun X, Hoshida Y (2015) Cancer biomarker discovery and validation. *Transl Cancer Res* 4: 256–269, doi:10.3978/j.issn.2218-676X.2015.06.04.

Gown AM (2008) Current issues in ER and HER2 testing by IHC in breast cancer. *Mod Pathol* 21: S8–S15, doi:10.1038/modpathol.2008.34.

Groth P, Orta ML, Elvers I, *et al.* (2012) Homologous recombination repairs secondary replication induced DNA double-strand breaks after ionizing radiation. *Nucleic Acids Res* 40: 6585–6594, doi:10.1093/nar/gks315.

Gui Y, Guo G, Huang Y, *et al.* (2011) Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat Genet* 43: 875–878, doi:10.1038/ng.907.

Gunn A, Stark JM (2012) I-SceI-Based Assays to Examine Distinct Repair Outcomes of Mammalian Chromosomal Double Strand Breaks. In Methods in Molecular Biology (Clifton, N.J.), pp. 379–391.

Gurcan MN, Boucheron LE, Can A, *et al.* (2009) Histopathological image analysis: a review. *IEEE Rev Biomed Eng* 2: 147–171, doi:10.1109/RBME.2009.2034865.

Hager M, Kolbitsch C, Tiefenthaler W, *et al.* (2007) Tissue microarrays from renal cell tumors: Exclusion criteria and rate of exclusion. *Scand J Urol Nephrol* 41: 485–489, doi:10.1080/00365590701520552.

Harper J V., Anderson JA, O'Neill P (2010) Radiation induced DNA DSBs: Contribution from stalled replication forks? *DNA Repair (Amst)* 9: 907–913, doi:10.1016/j.dnarep.2010.06.002.

Harris SL, Levine AJ (2005) The p53 pathway: positive and negative feedback loops. *Oncogene* 24: 2899–2908, doi:10.1038/sj.onc.1208615.

Hartmann MD, Mendler CT, Bassler J, *et al.* (2016) α/β coiled coils. *Elife* 5: doi:10.7554/eLife.11861.

Hartsuiker E, Neale MJ, Carr AM (2009) Distinct requirements for the Rad32(Mre11) nuclease and Ctp1(CtIP) in the removal of covalently bound topoisomerase I and II from DNA. *Mol Cell* 33: 117–123, doi:10.1016/j.molcel.2008.11.021.

Hashimoto Y, Chaudhuri AR, Lopes M, Costanzo V (2010) Rad51 protects nascent DNA from Mre11-dependent degradation and promotes continuous DNA synthesis. *Nat Struct Mol Biol* 17: 1305–1311, doi:10.1038/nsmb.1927.

Hashimoto Y, Puddu F, Costanzo V (2011) RAD51- and MRE11-dependent reassembly of uncoupled CMG helicase complex at collapsed replication forks. *Nat Struct Mol Biol* 19: 17–24, doi:10.1038/nsmb.2177.

Hashine K, Ide T, Nakashima T, *et al.* (2016) Results of second transurethral resection for high-grade T1 bladder cancer. *Urol Ann* 8: 10, doi:10.4103/0974-7796.163798.

He G, Siddik ZH, Huang Z, *et al.* (2005) Induction of p21 by p53 following DNA damage inhibits both Cdk4 and Cdk2 activities. *Oncogene* 24: 2929–2943, doi:10.1038/sj.onc.1208474.

Hegi ME, Diserens A-C, Gorlia T, *et al.* (2005) *MGMT* Gene Silencing and Benefit from Temozolomide in Glioblastoma. *N Engl J Med* 352: 997–1003, doi:10.1056/NEJMoa043331.

Heikkinen K, Karppinen S-M, Soini Y, *et al.* (2003) Mutation screening of Mre11 complex genes: indication of RAD50 involvement in breast and ovarian cancer susceptibility. *J Med Genet* 40: e131, doi:10.1136/JMG.40.12.E131.

Helleday T (2011) The underlying mechanism for the PARP and BRCA synthetic lethality: Clearing up the misunderstandings. *Mol Oncol* 5: 387–393, doi:10.1016/J.MOLONC.2011.07.001.

Hennequin C, Quero L, Favaudon V (2008) Déterminants et facteurs prédictifs pour la radiosensibilité tumorale. *Cancer/Radiothérapie* 12: 3–13, doi:10.1016/j.canrad.2007.11.007.

Henningsohn L, Wijkström H, Dickman PW, *et al.* (2002) Distressful symptoms after radical radiotherapy for urinary bladder cancer. *Radiother Oncol* 62: 215–225, doi:10.1016/S0167-8140(01)00455-8.

Henson DE (2005) Back to the drawing board on immunohistochemistry and predictive factors. *J Natl Cancer Inst* 97: 1796–1797, doi:10.1093/jnci/dji449.

Hernández B, Parnell A, Pennington SR (2014) Why have so few proteomic biomarkers 'survived' validation? (Sample size and independent validation considerations). *Proteomics* 14: 1587–1592, doi:10.1002/pmic.201300377.

HERR HW, Grimm M, Ackermann R, *et al.* (1999) The value of a second transurethral resection in evaluating patients with bladder tumors. *J Urol* 162: 74–76, doi:10.1097/00005392-199907000-00018.

Hirao A, Kong YY, Matsuoka S, *et al.* (2000) DNA damage-induced activation of p53 by the checkpoint kinase Chk2. *Science* 287: 1824–1827.

Hirsch FR, Varella-Garcia M, Cappuzzo F (2009) Predictive value of EGFR and HER2 overexpression in advanced non-small-cell lung cancer. *Oncogene* 28: S32–S37, doi:10.1038/onc.2009.199.

Hoa NN, Kobayashi J, Omura M, *et al.* (2015) BRCA1 and CtIP Are Both Required to Recruit Dna2 at Double-Strand Breaks in Homologous Recombination. *PLoS One* 10: e0124495, doi:10.1371/journal.pone.0124495.

Hoffman RA, Kothari S, Wang MD (2014) Comparison of normalization algorithms for cross-batch color segmentation of histopathological images. *Conf Proc . Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf* 2014: 194–197, doi:10.1109/EMBC.2014.6943562.

Hon GC, Hawkins RD, Caballero OL, *et al.* (2012) Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res* 22: 246–258, doi:10.1101/gr.125872.111.

Hopfner K-P, Craig L, Moncalian G, *et al.* (2002) The Rad50 zinc-hook is a structure joining Mre11 complexes in DNA recombination and repair. *Nature* 418: 562–566, doi:10.1038/nature00922.

Hoskin PJ, Rojas AM, Bentzen SM, Saunders MI (2010) Radiotherapy with concurrent carbogen and nicotinamide in bladder carcinoma. *J Clin Oncol* 28: 4912–4918, doi:10.1200/JCO.2010.28.4950.

Howat WJ, Lewis A, Jones P, *et al.* (2014) Antibody validation of immunohistochemistry for biomarker discovery: Recommendations of a consortium of academic and pharmaceutical based histopathology researchers. *Methods* 70: 34–38, doi:10.1016/j.ymeth.2014.01.018.

Hsieh M-C, Sung M-T, Chiang P-H, *et al.* (2015) The Prognostic Impact of Histopathological Variants in Patients with Advanced Urothelial Carcinoma. *PLoS One* 10: e0129268, doi:10.1371/journal.pone.0129268.

Huddart RA, Hall E, Lewis R, Birtle A (2010) Life and death of spare (selective bladder preservation against radical excision): reflections on why the spare trial closed. *BJU Int* 106: 753–755, doi:10.1111/j.1464-410X.2010.09537.x.

Iacovides I, Jennett C, Cornish-Trestrail C, Cox AL (2013) Do games attract or sustain engagement in citizen science? In CHI ’13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA ’13, (New York, New York, USA: ACM Press), p. 1101.

Iakovlev V V, Pintilie M, Morrison A, *et al.* (2007) Effect of distributional heterogeneity on the analysis of tumor hypoxia based on carbonic anhydrase IX. *Lab Investig* 87: 1206–1217, doi:10.1038/labinvest.3700680.

International Collaboration of Trialists, Medical Research Council Advanced Bladder Cancer Working Party (now the National Cancer Research Institute Bladder Cancer Clinical Studies Group), European Organisation for Research and Treatment of Cancer Genito-Urinary Tract Cancer Group, *et al.* (2011) International Phase III Trial Assessing Neoadjuvant Cisplatin, Methotrexate, and Vinblastine Chemotherapy for Muscle-Invasive Bladder Cancer: Long-Term Results of the BA06 30894 Trial. *J Clin Oncol* 29: 2171–2177, doi:10.1200/JCO.2010.32.3139.

Jackson SP, Bartek J (2009) The DNA-damage response in human biology and disease. *Nature* 461: 1071–1078, doi:10.1038/nature08467.

de Jager M, Dronkert ML, Modesti M, *et al.* (2001a) DNA-binding and strand-annealing activities of human Mre11: implications for its roles in DNA double-strand break repair pathways. *Nucleic Acids Res* 29: 1317–1325.

de Jager M, van Noort J, van Gent DC, Dekker C, Kanaar R, Wyman C (2001b) Human Rad50/Mre11 is a flexible complex that can tether DNA ends. *Mol Cell* 8: 1129–1135.

de Jager M, Trujillo KM, Sung P, *et al.* (2004) Differential Arrangements of Conserved Building Blocks among Homologs of the Rad50/Mre11 DNA Repair Protein Complex. *J Mol Biol* 339: 937–949, doi:10.1016/j.jmb.2004.04.014.

James ND, Hussain SA, Hall E, *et al.* (2012) Radiotherapy with or without Chemotherapy in Muscle-Invasive Bladder Cancer. *N Engl J Med* 366: 1477–1488, doi:10.1056/NEJMoa1106106.

Janowczyk A, Madabhushi A (2016) Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *J Pathol Inform* 7: 29, doi:10.4103/2153-3539.186902.

Jeong JH, Kang SS, Park KK, *et al.* (2010) p53-independent induction of G1 arrest and p21WAF1/CIP1 expression by ascofuranone, an isoprenoid antibiotic, through downregulation of c-Myc. *Mol Cancer Ther* 9: 2102–2113, doi:10.1158/1535-7163.MCT-09-1159.

Jiao L, Zhu J, Hassan MM, *et al.* (2007) K-ras mutation and p16 and preproenkephalin promoter hypermethylation in plasma DNA of pancreatic cancer patients: in relation to cigarette smoking. *Pancreas* 34: 55–62, doi:10.1097/01.mpa.0000246665.68869.d4.

Jones CM, Goh V, Sebag-Montefiore D, Gilbert DC (2017) Biomarkers in anal cancer: from biological understanding to stratified treatment. *Br J Cancer* 116: 156–162, doi:10.1038/bjc.2016.398.

Jones DL, Alani RM, Münger K (1997) The human papillomavirus E7 oncoprotein can uncouple cellular differentiation and proliferation in human keratinocytes by abrogating p21Cip1-mediated inhibition of cdk2. *Genes Dev* 11: 2101–2111, doi:10.1101/GAD.11.16.2101.

Kaelin WG (2005) The Concept of Synthetic Lethality in the Context of Anticancer Therapy. *Nat Rev Cancer* 5: 689–698, doi:10.1038/nrc1691.

Kamat AM, Hahn NM, Efstathiou JA, *et al.* (2016) Bladder cancer. *Lancet* 388: 2796–2810, doi:10.1016/S0140-6736(16)30512-8.

Kamiya N, Suzuki H, Suyama T, *et al.* (2016) Clinical outcomes of second transurethral resection in non-muscle invasive high-grade bladder cancer: a retrospective, multi-institutional, collaborative study. *Int J Clin Oncol* doi:10.1007/s10147-016-1048-z.

Kampf C, Olsson I, Ryberg U, *et al* (2012) Production of tissue microarrays, immunohistochemistry staining and digitalization within the human protein atlas. *J Vis Exp* doi:10.3791/3620.

Kandimalla R, van Tilborg AA, Zwarthoff EC (2013) DNA methylation-based biomarkers in bladder cancer. *Nat Rev Urol* 10: 327–335, doi:10.1038/nrurol.2013.89.

Karapetis CS, Khambata-Ford S, Jonker DJ, *et al.* (2008) *K-ras* Mutations and Benefit from Cetuximab in Advanced Colorectal Cancer. *N Engl J Med* 359: 1757–1765, doi:10.1056/NEJMoa0804385.

Karen KA, Hoey PJ, Young CSH, Hearing P (2009) Temporal regulation of the Mre11-Rad50-Nbs1 complex during adenovirus infection. *J Virol* 83: 4565–4573, doi:10.1128/JVI.00042-09.

Karsli-Ceppioglu S, Dagdemir A, Judes G, *et al.* (2017) The Epigenetic Landscape of Promoter Genome-wide Analysis in Breast Cancer. *Sci Rep* 7: 6597, doi:10.1038/s41598-017-06790-z.

Kaufman PA, Bloom KJ, Burris H, *et al.* (2014) Assessing the discordance rate between local and central HER2 testing in women with locally determined HER2-negative breast cancer. *Cancer* 120: 2657–2664, doi:10.1002/cncr.28710.

Kaur G, Dufour JM (2012) Cell lines: Valuable tools or useless artifacts. *Spermatogenesis* 2: 1–5, doi:10.4161/spmg.19885.

Kawrykow A, Roumanis G, Kam A, *et al.* (2012) Phylo: A Citizen Science Approach for Improving Multiple Sequence Alignment. *PLoS One* 7: e31362, doi:10.1371/journal.pone.0031362.

Khanna. KK, Keating KE, Kozlov S, *et al.* (1998) ATM associates with and phosphorylates p53: mapping the region of interaction. *Nat Genet* 20: 398–400, doi:10.1038/3882.

Khatib F, Cooper S, Tyka MD, *et al.* (2011a) Algorithm discovery by protein folding game players. *Proc Natl Acad Sci U S A* 108: 18949–18953, doi:10.1073/pnas.1115898108.

Khatib F, DiMaio F, Cooper S, *et al.* (2011b) Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nat Struct Mol Biol* 18: 1175–1177, doi:10.1038/nsmb.2119.

Khosravi R, Maya R, Gottlieb T, *et al.* (1999) Rapid ATM-dependent phosphorylation of MDM2 precedes p53 accumulation in response to DNA damage. *Proc Natl Acad Sci U S A* 96: 14973–14977, doi:10.1073/PNAS.96.26.14973.

Khouja MH, Baekelandt M, Sarab A, *et al.* (2010) Limitations of tissue microarrays compared with whole tissue sections in survival analysis. *Oncol Lett* 1: 827–831, doi:10.3892/ol_00000145.

Kisiel JB, Raimondo M, Taylor WR, *et al.* (2015) New DNA Methylation Markers for Pancreatic Cancer: Discovery, Tissue Validation, and Pilot Testing in Pancreatic Juice. *Clin Cancer Res* 21: 4473–4481, doi:10.1158/1078-0432.CCR-14-2469.

Kiss B, Burkhard FC, Thalmann GN (2016) Open radical cystectomy: still the gold standard for muscle invasive bladder cancer. *World J Urol* 34: 33–39, doi:10.1007/s00345-015-1729-7.

Kiss B, Wyatt AW, Douglas J, *et al.* (2017) Her2 alterations in muscle-invasive bladder cancer: Patient selection beyond protein expression for targeted therapy. *Sci Rep* 7: 42713, doi:10.1038/srep42713.

Reprinted by permission from Springer Nature: Nat Rev Cancer Knowles MA, Hurst CD (2014) Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. Nat Rev Cancer 15: 25–41, doi:10.1038/nrc3817

Knowles MA, Hurst CD (2014) Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. *Nat Rev Cancer* 15: 25–41, doi:10.1038/nrc3817.

Kochańczyk T, Nowakowski M, Wojewska D, *et al.* (2016) Metal-coupled folding as the driving force for the extreme stability of Rad50 zinc hook dimer assembly. *Sci Rep* 6: 36346, doi:10.1038/srep36346.

Kodama M, Otsubo C, Hirota T, *et al.* (2010) Requirement of ATM for rapid p53 phosphorylation at Ser46 without Ser/Thr-Gln sequences. *Mol Cell Biol* 30: 1620–1633, doi:10.1128/MCB.00810-09.

Koerber SA, Schoneweg C, Slynko A, *et al.* (2014) HPV in anal cancer Influence of human papillomavirus and p16 INK4a on treatment outcome of patients with anal cancer. doi:10.1016/j.radonc.2014.11.013.

Kolinjivadi AM, Sannino V, De Antoni A, *et al.* (2017) Smarcal1-Mediated Fork Reversal Triggers Mre11-Dependent Degradation of Nascent DNA in the Absence of Brca2 and Stable Rad51 Nucleofilaments. *Mol Cell* 67: 867–881.e7, doi:10.1016/j.molcel.2017.07.001.

Kondo T, Kobayashi J, Saitoh T, *et al.* (2013) DNA damage sensor MRE11 recognizes cytosolic double-stranded DNA and induces type I interferon by regulating STING trafficking. *Proc Natl Acad Sci U S A* 110: 2969–2974, doi:10.1073/pnas.1222694110.

Kothari S, Phan JH, Stokes TH, *et al.* (2014) Removing batch effects from histopathological images for enhanced cancer diagnosis. *IEEE J Biomed Heal informatics* 18: 765–772, doi:10.1109/JBHI.2013.2276766.

Kotwal S, Choudhury A, Johnston C, *et al.* (2008) Similar treatment outcomes for radical cystectomy and radical radiotherapy in invasive bladder cancer treated at a United Kingdom specialist treatment center. *Int J Radiat Oncol Biol Phys* 70: 456–463, doi:10.1016/j.ijrobp.2007.06.030.

Kourou K, Exarchos TP, Exarchos KP, *et al.* (2015) Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J* 13: 8–17, doi:10.1016/J.CSBJ.2014.11.005.

Krajewski W, Zdrojowy R, Tupikowski K, *et al.* (2016) How to lower postoperative complications after radical cystectomy - a review. *Cent Eur J Urol* 69: 370–376, doi:10.5173/ceju.2016.880.

Krasnow RE, Drumm M, Roberts HJ, *et al.* (2017) Clinical Outcomes of Patients with Histologic Variants of Urothelial Cancer Treated with Trimodality Bladder-sparing Therapy. *Eur Urol* 72: 54–60, doi:10.1016/j.eururo.2016.12.002.

Kretschmer A, Grimm T, Buchner A, *et al.* (2016a) Prospective evaluation of health-related quality of life after radical cystectomy: focus on peri- and postoperative complications. *World J Urol* doi:10.1007/s00345-016-1992-2.

Kretschmer A, Grimm T, Buchner A, *et al.* (2016b) Prognostic features for quality of life after radical cystectomy and orthotopic neobladder. *Int Braz J Urol* 42: 1109–1120, doi:10.1590/S1677-5538.IBJU.2015.0491.

Krogh BO, Llorente B, Lam A, Symington LS (2005) Mutations in Mre11 phosphoesterase motif I that impair Saccharomyces cerevisiae Mre11-Rad50-Xrs2 complex stability in addition to nuclease activity. *Genetics* 171: 1561–1570, doi:10.1534/genetics.105.049478.

Kumar R, Srivastava R, Srivastava S (2015) Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features. *J Med Eng* 2015: 457906, doi:10.1155/2015/457906.

Laé M, Couturier J, Oudard S, *et al.* (2010) Assessing HER2 gene amplification as a potential target for therapy in invasive urothelial bladder cancer with a standardized methodology: results in 1005 patients. *Ann Oncol Off J Eur Soc Med Oncol* 21: 815–819, doi:10.1093/annonc/mdp488.

Lafrance-Vanasse J, Williams GJ, Tainer JA (2015) Envisioning the dynamics and flexibility of Mre11-Rad50-Nbs1 complex to decipher its roles in DNA replication and

repair. *Prog Biophys Mol Biol* 117: 182–193, doi:10.1016/j.pbiomolbio.2014.12.004. https://creativecommons.org/licenses/by/2.0/legalcode

Lammens K, Bemeleit DJ, Möckel C, *et al.* (2011) The Mre11:Rad50 structure shows an ATP-dependent molecular clamp in DNA double-strand break repair. *Cell* 145: 54–66, doi:10.1016/j.cell.2011.02.038.

Reprinted from *Cell* 145. Lammens K, Bemeleit DJ, Möckel C, *et al.* (2011) The Mre11:Rad50 structure shows an ATP-dependent molecular clamp in DNA double-strand break repair. 54–66 (2011) with permission from Elsevier. doi:10.1016/j.cell.2011.02.038

Larcher A, Sun M, Schiffmann J, *et al.* (2015) Peri-operative mortality and long-term survival after partial versus radical cystectomy for muscle invasive bladder cancer. *Bladder* 2: 14, doi:10.14440/bladder.2015.55.

Laurberg JR, Brems-Eskildsen AS, Nordentoft I, *et al.* (2012) Expression of TIP60 (tat-interactive protein) and MRE11 (meiotic recombination 11 homolog) predict treatment-specific outcome of localised invasive bladder cancer. *BJU Int* 110: E1228–E1236, doi:10.1111/j.1464-410X.2012.11564.x.

Lavallée LT, Schramm D, Witiuk K, *et al.* (2014) Peri-Operative Morbidity Associated with Radical Cystectomy in a Multicenter Database of Community and Academic Hospitals. *PLoS One* 9: e111281, doi:10.1371/journal.pone.0111281.

Lavin MF (2007) ATM and the Mre11 complex combine to recognize and signal DNA double-strand breaks. *Oncogene* 26: 7749–7758, doi:10.1038/sj.onc.1210880.

Lavin MF, Gueven N (2006) The complexity of p53 stabilization and activation. *Cell Death Differ* 13: 941–950, doi:10.1038/sj.cdd.4401925.

Lawson J, Robinson-Vyas RJ, McQuillan JP, *et al.* (2017) Crowdsourcing for translational research: analysis of biomarker expression using cancer microarrays. *Br J Cancer* 116: 237–245, doi:10.1038/bjc.2016.404.

Lee J-H, Lim D-S (2006) Dual role of Nbs1 in the ataxia telangiectasia mutated-dependent DNA damage response. *FEBS J* 273: 1630–1636, doi:10.1111/j.1742-4658.2006.05191.x.

Lee J-H, Mand MR, Deshpande RA, *et al.* (2013) Ataxia telangiectasia-mutated (ATM) kinase activity is regulated by ATP-driven conformational changes in the Mre11/Rad50/Nbs1 (MRN) complex. *J Biol Chem* 288: 12840–12851, doi:10.1074/jbc.M113.460378.

Lee J-H, Paull TT (2004) Direct activation of the ATM protein kinase by the Mre11/Rad50/Nbs1 complex. *Science* 304: 93–96, doi:10.1126/science.1091496.

Lee J-H, Paull TT (2005) ATM Activation by DNA Double-Strand Breaks Through the Mre11-Rad50-Nbs1 Complex. *Science (80-)* 308: 551–554, doi:10.1126/science.1108297.

Lee J, Kladwang W, Lee M, *et al.* (2014) RNA design rules from a massive open laboratory. *Proc Natl Acad Sci U S A* 111: 2122–2127, doi:10.1073/pnas.1313039111.

Leow JJ, Martin-Doyle W, Rajagopal PS, *et al.* (2014) Adjuvant Chemotherapy for Invasive Bladder Cancer: A 2013 Updated Systematic Review and Meta-Analysis of Randomized Trials. *Eur Urol* 66: 42–54, doi:10.1016/j.eururo.2013.08.033.

Levenson RM, Krupinski EA, Navarro VM, *et al.* (2015) Pigeons (Columba livia) as Trainable Observers of Pathology and Radiology Breast Cancer Images. *PLoS One* 10: e0141357, doi:10.1371/journal.pone.0141357.

Li X, Xu H, Xu C, *et al.* (2013) The yin-yang of DNA damage response: roles in tumorigenesis and cellular senescence. *Int J Mol Sci* 14: 2431–2448, doi:10.3390/ijms14022431.

Li Y, Dowbenko D, Lasky LA (2002) AKT/PKB Phosphorylation of p21 $^{Cip/WAF1}$ Enhances Protein Stability of p21 $^{Cip/WAF1}$ and Promotes Cell Survival. *J Biol Chem* 277: 11352–11361, doi:10.1074/jbc.M109062200.

Liao S, Tammaro M, Yan H (2016) The structure of ends determines the pathway choice and Mre11 nuclease dependency of DNA double-strand break repair. *Nucleic Acids Res* 44: 5689–5701, doi:10.1093/nar/gkw274.

Liedberg F (2010) Early Complications and Morbidity of Radical Cystectomy. *Eur Urol Suppl* 9: 25–30, doi:10.1016/j.eursup.2010.01.007.

Lim HS, Kim JS, Park YB, *et al.* (2011) Crystal structure of the Mre11-Rad50-ATPγS complex: understanding the interplay between Mre11 and Rad50. *Genes Dev* 25: 1091–1104, doi:10.1101/gad.2037811.

Limbo O, Chahwan C, Yamada Y, *et al.* (2007) Ctp1 Is a Cell-Cycle-Regulated Protein that Functions with Mre11 Complex to Control Double-Strand Break Repair by Homologous Recombination. *Mol Cell* 28: 134–146, doi:10.1016/j.molcel.2007.09.009.

Linderoth J, Ehinger M, Åkerman M, *et al.* (2007) Tissue microarray is inappropriate for analysis of BCL6 expression in diffuse large B-cell lymphoma. *Eur J Haematol* 79: 146–149, doi:10.1111/j.1600-0609.2007.00892.x.

Lindgren D, Sjödahl G, Lauss M, *et al.* (2012) Integrated Genomic and Gene Expression Profiling Identifies Two Major Genomic Circuits in Urothelial Carcinoma. *PLoS One* 7: e38863, doi:10.1371/journal.pone.0038863.

Liu T, Huang J (2014) Quality control of homologous recombination. *Cell Mol Life Sci* 71: 3779–3797, doi:10.1007/s00018-014-1649-5.

Liu Y, Shevchenko A, Shevchenko A, Berk AJ (2005) Adenovirus Exploits the Cellular Aggresome Response To Accelerate Inactivation of the MRN Complex. *J Virol* 79: 14004–14016, doi:10.1128/JVI.79.22.14004-14016.2005.

Liu Y, Sung S, Kim Y, *et al.* (2016) ATP-dependent DNA binding, unwinding, and resection by the Mre11/Rad50 complex. *EMBO J* 35: 743–758, doi:10.15252/embj.201592462.

Lobachev KS, Gordenin DA, Resnick MA (2002) The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. *Cell* 108: 183–193.

Loignon M, Fetni R, Gordon AJ, Drobetsky EA (1997) A p53-independent pathway for induction of p21waf1cip1 and concomitant G1 arrest in UV-irradiated human skin fibroblasts. *Cancer Res* 57: 3390–3394.

Longworth MS, Wilson R, Laimins LA (2005) HPV31 E7 facilitates replication by activating E2F2 transcription through its interaction with HDACs. *EMBO J* 24: 1821–1830, doi:10.1038/sj.emboj.7600651.

Loree JM, Kopetz S (2017) Recent developments in the treatment of metastatic colorectal cancer. *Ther Adv Med Oncol* 9: 551–564, doi:10.1177/1758834017714997.

Lorenzo Y, Costa S, Collins AR, Azqueta A (2013) The comet assay, DNA damage, DNA repair and cytotoxicity: hedgehogs are not always dead. *Mutagenesis* 28: 427–432, doi:10.1093/mutage/get018.

Luo Y, Hurwitz J, Massagué J (1995) Cell-cycle inhibition by independent CDK and PCNA binding domains in p21Cip1. *Nature* 375: 159–161, doi:10.1038/375159a0.

Macleod KF, Sherry N, Hannon G, *et al.* (1995) p53-dependent and independent expression of p21 during cell growth, differentiation, and DNA damage. *Genes Dev* 9: 935–944.

Madabhushi A, Lee G (2016) Image analysis and machine learning in digital pathology: Challenges and opportunities. *Med Image Anal* 33: 170–175, doi:10.1016/j.media.2016.06.037.

Magliocco AM, Moughan J, Simko J, *et al.* (2017) The impact of MRE11 in nuclear to cytoplasmic ratio on outcomes in muscle invasive bladder cancer an analysis of NRG/RTOG 8802, 8903, 9506, 9706, 9906, and 0233. *J Clin Oncol* 35: 343–343, doi:10.1200/JCO.2017.35.6_suppl.343.

Mai S, Welzel G, Ottstadt M, *et al.* (2015) Prognostic Relevance of HPV Infection and p16 Overexpression in Squamous Cell Anal Cancer. *Int J Radiat Oncol* 93: 819–827, doi:10.1016/j.ijrobp.2015.08.004.

Mak KS, Smith AB, Eidelman A, *et al.* (2016) Quality of Life in Long-term Survivors of Muscle-Invasive Bladder Cancer. *Int J Radiat Oncol* 96: 1028–1036, doi:10.1016/j.ijrobp.2016.08.023.

Mak KS, Smith AB, Eidelman A, *et al.* (2015) Quality of Life in Long-term Survivors of Muscle-Invasive Bladder Cancer. *Int J Radiat Oncol* 93: S23, doi:10.1016/j.ijrobp.2015.07.2206.

Manoharan M, Ayyathurai R, Soloway MS (2009) Radical cystectomy for urothelial carcinoma of the bladder: an analysis of perioperative and survival outcome. *BJU Int* 104: 1227–1232, doi:10.1111/j.1464-410X.2009.08625.x.

De Marco F (2013) Oxidative stress and HPV carcinogenesis. *Viruses* 5: 708–731, doi:10.3390/v5020708.

Martin JW, Carballido EM, Ahmed A, *et al* (2016) Squamous cell carcinoma of the urinary bladder: Systematic review of clinical characteristics and therapeutic approaches. *Arab J Urol* 14: 183–191, doi:10.1016/j.aju.2016.07.001.

Marullo R, Werner E, Zhang H, *et al.* (2015) HPV16 E6 and E7 proteins induce a chronic oxidative stress response via NOX2 that causes genomic instability and increased susceptibility to DNA damage in head and neck cancer cells. *Carcinogenesis* 36: 1397–1406, doi:10.1093/carcin/bgv126.

de Matos LL, Trufelli DC, de Matos MGL, da Silva Pinhal MA (2010) Immunohistochemistry as an important tool in biomarkers detection and clinical practice. Biomark Insights 2010: 9–20, doi:10.4137/BMI.S2185.

Matsuoka S, Huang M, Elledge SJ (1998) Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. *Science* 282: 1893–1897, doi:10.1126/SCIENCE.282.5395.1893.

Maya R, Balass M, Kim ST, *et al* (2001) ATM-dependent phosphorylation of Mdm2 on serine 395: role in p53 activation by DNA damage. *Genes Dev* 15: 1067–1077, doi:10.1101/gad.886901.

McHugh ML (2012) Interrater reliability: the kappa statistic. *Biochem medica* 22: 276–282.

Mehta S, Shelling A, Muthukaruppan A, *et al.* (2010) Predictive and prognostic molecular markers for cancer medicine. *Ther Adv Med Oncol* 2: 125–148, doi:10.1177/1758834009360519.

Menendez D, Inga A, Resnick MA (2009) The expanding universe of p53 targets. *Nat Rev Cancer* 9: 724–737, doi:10.1038/nrc2730.

Meulendijks D, Tomasoa NB, Dewit L, *et al.* (2015) HPV-negative squamous cell carcinoma of the anal canal is unresponsive to standard treatment and frequently carries disruptive mutations in TP53. *Br J Cancer* 112: 1358–1366, doi:10.1038/bjc.2015.20.

Meyer D, Fu BXH, Heyer W-D (2015) DNA polymerases δ and λ cooperate in repairing double-strand breaks by microhomology-mediated end-joining in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* 112: E6907-16, doi:10.1073/pnas.1507833112.

Millikan R, Dinney C, Swanson D, *et al.* (2001) Integrated therapy for locally advanced bladder cancer: final report of a randomized trial of cystectomy plus adjuvant M-VAC versus cystectomy with both preoperative and postoperative M-VAC. *J Clin Oncol* 19: 4005–4013, doi:10.1200/JCO.2001.19.20.4005.

Miquel C, Jacob S, Grandjouan S, *et al.* (2007) Frequent alteration of DNA damage signalling and repair pathways in human colorectal cancers with microsatellite instability. *Oncogene* 26: 5919–5926, doi:10.1038/sj.onc.1210419.

Mockenhaupt S, Makeyev E V (2015) Non-coding functions of alternative pre-mRNA splicing in development. *Semin Cell Dev Biol* 47–48: 32–39, doi:10.1016/j.semcdb.2015.10.018.

Momand J, Zambetti GP, Olson DC, *et al.* (1992) The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. *Cell* 69: 1237–1245.

Moody CA, Laimins LA, Hausen H zur, *et al.* (2009) Human papillomaviruses activate the ATM DNA damage pathway for viral genome amplification upon differentiation. PLoS Pathog 5: e1000605, doi:10.1371/journal.ppat.1000605.

Morgan MA, Lawrence TS (2015) Molecular Pathways: Overcoming radiation resistance by targeting DNA damage response pathways. *Clin Cancer Res* 21: 2898–2904, doi:10.1158/1078-0432.CCR-13-3229.

Morrison CD, Liu P, Woloszynska-Read A, *et al.* (2014) Whole-genome sequencing identifies genomic heterogeneity at a nucleotide and chromosomal level in bladder cancer. *Proc Natl Acad Sci U S A* 111: E672-81, doi:10.1073/pnas.1313580111.

Mukaka MM (2012) Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Med J* 24: 69–71.

Nag S, Qin J, Srivenugopal KS, *et al.* (2013) The MDM2-p53 pathway revisited. *J Biomed Res* 27: 254–271, doi:10.7555/JBR.27.20130030.

Nakano K, Vousden KH (2001) PUMA, a novel proapoptotic gene, is induced by p53. *Mol Cell* 7: 683–694.

Nelson RA, Levine AM, Bernstein L, *et al.* (2013) Changing patterns of anal canal carcinoma in the United States. *J Clin Oncol* 31: 1569–1575, doi:10.1200/JCO.2012.45.2524.

NICE (2015) Bladder cancer: diagnosis and management. NICE guideline [NG2]. https://www.nice.org.uk/guidance/ng2 (accessed: 07/03/2017).

Nicholson J, Jevons S, Groselj B, *et al.* (2017) E3 ligase cIAP2 mediates downregulation of MRE11 and radiosensitization in response to HDAC inhibition in bladder cancer. *Cancer Res* canres.3232.2016, doi:10.1158/0008-5472.CAN-16-3232.

Nieminuszczy J, Schwab RA, Niedzwiedz W (2016) The DNA fibre technique – tracking helicases at work. Methods. 2016 Oct 1;108:92-8. doi: 10.1016/j.ymeth.2016.04.019

Nimonkar A V, Genschel J, Kinoshita E, *et al.* (2011) BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. *Genes Dev* 25: 350–362, doi:10.1101/gad.2003811.

Nordentoft I, Lamy P, Birkenkamp-Demtröder K, *et al.* (2014) Mutational Context and Diverse Clonal Development in Early and Late Bladder Cancer. *Cell Rep* 7: 1649–1663, doi:10.1016/j.celrep.2014.04.038.

Nørregaard Jensen O (2004) Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr Opin Chem Biol* 8: 33–41, doi:10.1016/j.cbpa.2003.12.009.

Nuovo G (2016) False-positive results in diagnostic immunohistochemistry are related to horseradish peroxidase conjugates in commercially available assays. *Ann Diagn Pathol* 25: 54–59, doi:10.1016/j.anndiagpath.2016.09.010.

O'Hurley G, Sjöstedt E, Rahman A, *et al.* (2014) Garbage in, garbage out: A critical evaluation of strategies used for validation of immunohistochemical biomarkers. *Mol Oncol* 8: 783–798, doi:10.1016/j.molonc.2014.03.008.

Oh ST, Kyo S, Laimins LA (2001) Telomerase activation by human papillomavirus type 16 E6 protein: induction of human telomerase reverse transcriptase expression through Myc and GC-rich Sp1 binding sites. *J Virol* 75: 5559–5566, doi:10.1128/JVI.75.12.5559-5566.2001.

Okoye JO, Nnatuanya IN (2015) Immunohistochemistry : A Revolutionary Technique in Laboratory Medicine. *Clin Med Diagnostics* 5: 60–69, doi:10.5923/j.cmd.20150504.02.

Olive PL, Banáth JP (2006) The comet assay: a method to measure DNA damage in individual cells. *Nat Protoc* 1: 23–29, doi:10.1038/nprot.2006.5.

Orth M, Lauber K, Niyazi M, *et al.* (2014) Current concepts in clinical radiation oncology. *Radiat Environ Biophys* 53: 1–29, doi:10.1007/s00411-013-0497-2.

Paik S, Shak S, Tang G, *et al.* (2004) A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. *N Engl J Med* 351: 2817–2826, doi:10.1056/NEJMoa041588.

Paik S, Tang G, Shak S, *et al.* (2006) Gene Expression and Benefit of Chemotherapy in Women With Node-Negative, Estrogen Receptor–Positive Breast Cancer. *J Clin Oncol* 24: 3726–3734, doi:10.1200/JCO.2005.04.7985.

Park YB, Chae J, Kim YC, *et al.* (2011) Crystal structure of human Mre11: understanding tumorigenic mutations. *Structure* 19: 1591–1602, doi:10.1016/j.str.2011.09.010.

Park YB, Hohl M, Padjasek M, *et al.* (2017) Eukaryotic Rad50 functions as a rod-shaped dimer. *Nat Struct Mol Biol* 24: 248–257, doi:10.1038/nsmb.3369.

Paull TT, Gellert M (1999) Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by the Mre11/Rad50 complex. *Genes Dev* 13: 1276–1288.

Paz MF, Fraga MF, Avila S, *et al.* (2003) A systematic profile of DNA methylation in human cancer cell lines. *Cancer Res* 63: 1114–1121.

Perez EA, Cortés J, Gonzalez-Angulo AM, Bartlett JMS (2014) HER2 testing: Current status and future directions. *Cancer Treat Rev* 40: 276–284, doi:10.1016/j.ctrv.2013.09.001.

Permuth-Wey J, Boulware D, Valkov N, *et al.* (2009) Sampling Strategies for Tissue Microarrays to Evaluate Biomarkers in Ovarian Cancer. *Cancer Epidemiol Biomarkers Prev* 18: 28–34, doi:10.1158/1055-9965.EPI-08-0713.

Petermann E, Orta ML, Issaeva N, *et al.* (2010) Hydroxyurea-Stalled Replication Forks Become Progressively Inactivated and Require Two Different RAD51-Mediated Pathways for Restart and Repair. *Mol Cell* 37: 492–502, doi:10.1016/j.molcel.2010.01.021.

Petkov VI, Miller DP, Howlader N, *et al.* (2016) Breast-cancer-specific mortality in patients treated based on the 21-gene assay: a SEER population-based study. *NPJ breast cancer* 2: 16017, doi:10.1038/npjbcancer.2016.17.

Petrelli F, Viale G, Cabiddu M, Barni S (2015) Prognostic value of different cut-off levels of Ki-67 in breast cancer: a systematic review and meta-analysis of 64,196 patients. *Breast Cancer Res Treat* 153: 477–491, doi:10.1007/s10549-015-3559-0.

Pham N-A, Morrison A, Schwock J, *et al.* (2007) Quantitative image analysis of immunohistochemical stains using a CMYK color model. *Diagn Pathol* 2: 8, doi:10.1186/1746-1596-2-8.

Ponciano L, Brasileiro F, Simpson R, Smith A (2014) Volunteers' Engagement in Human Computation for Astronomy Projects. *Comput Sci Eng* 16: 52–59, doi:10.1109/MCSE.2014.4.

Ponec RM, Peddi P, Callahan RD (2017) PARP Inhibitors in Breast Cancer: Latest Evidence. *Curr Breast Cancer Rep* 9: 188–194, doi:10.1007/s12609-017-0251-x.

Povirk LF, F. L (2012) Processing of damaged DNA ends for double-strand break repair in mammalian cells. *ISRN Mol Biol* 2012: doi:10.5402/2012/345805.

Prichard JW (2014) Overview of Automated Immunohistochemistry. *Arch Pathol Lab Med* 138: 1578–1582, doi:10.5858/arpa.2014-0083-RA.

Pryor JM, Waters CA, Aza A, *et al.* (2015) Essential role for polymerase specialization in cellular nonhomologous end joining. *Proc Natl Acad Sci U S A* 112: E4537-45, doi:10.1073/pnas.1505805112.

R Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rai R, Zheng H, He H, *et al.* (2010) The function of classical and alternative non-homologous end-joining pathways in the fusion of dysfunctional telomeres. *EMBO J* 29: 2598–2610, doi:10.1038/emboj.2010.142.

Rakha EA, Pinder SE, Bartlett JMS, *et al.* (2015) Updated UK Recommendations for HER2 assessment in breast cancer. *J Clin Pathol* 68: 93–99, doi:10.1136/jclinpath-2014-202571.

Rallapalli G, Fraxinus Players F, Saunders DG, *et al.* (2015) Lessons from Fraxinus, a crowd-sourced citizen science game in genomics. *Elife* 4: e07460, doi:10.7554/eLife.07460.

Ramani VAC, Maddineni SB, Grey BR, Clarke NW (2010) Differential Complication Rates Following Radical Cystectomy in the Irradiated and Nonirradiated Pelvis. *Eur Urol* 57: 1058–1063, doi:10.1016/j.eururo.2009.12.002.

Rass U, Compton SA, Matos J, *et al* (2010) Mechanism of Holliday junction resolution by the human GEN1 protein. *Genes Dev* 24: 1559–1569, doi:10.1101/gad.585310.

Ray Chaudhuri A, Hashimoto Y, Herrador R, *et al.* (2012) Topoisomerase I poisoning results in PARP-mediated replication fork reversal. *Nat Struct Mol Biol* 19: 417–423, doi:10.1038/nsmb.2258.

Regal JA, Festerling TA, Buis JM, Ferguson DO (2013) Disease-associated MRE11 mutants impact ATM/ATR DNA damage signaling by distinct mechanisms. *Hum Mol Genet* 22: 5146–5159, doi:10.1093/hmg/ddt368.

Reinson T, Toots M, Kadaja M, *et al.* (2013) Engagement of the ATR-dependent DNA damage response at the human papillomavirus 18 replication centers during the initial amplification. *J Virol* 87: 951–964, doi:10.1128/JVI.01943-12.

Richman SD, Fairley J, Butler R, Deans ZC (2017) How close are we to standardised extended RAS gene mutation testing? The UK NEQAS evaluation. *J Clin Pathol* 70: 58–62, doi:10.1136/jclinpath-2016-203822.

Robertson S, Azizpour H, Smith K, Hartman J (2017) Digital image analysis in breast pathology—from image processing techniques to artificial intelligence. *Transl Res* doi:10.1016/j.trsl.2017.10.010.

Robson M, Im S-A, Senkus E, *et al.* (2017) Olaparib for Metastatic Breast Cancer in Patients with a Germline *BRCA* Mutation. *N Engl J Med* 377: 523–533, doi:10.1056/NEJMoa1706450.

Rodriguez-Brenes IA, Wodarz D, Komarova NL (2016) Quantifying replicative senescence as a tumor suppressor pathway and a target for cancer therapy. *Sci Rep* 5: 17660, doi:10.1038/srep17660.

Rodríguez-Paredes M, Esteller M (2011) Cancer epigenetics reaches mainstream oncology. *Nat Med* 17: 330–339, doi:10.1038/nm.2305.

Rodriguez MS, Desterro JM, Lain S,*et al.* (2000) Multiple C-terminal lysine residues target p53 for ubiquitin-proteasome-mediated degradation. *Mol Cell Biol* 20: 8458–8467.

Rojowska A, Lammens K, Seifert FU, *et al.* (2014) Structure of the Rad50 DNA double-strand break repair protein in complex with DNA. *EMBO J* 33: 2847–2859, doi:10.15252/embj.201488889.

Roth B, Thalmann GN (2015) Standard cystectomy fits all: truth or myth? *Transl Androl Urol* 4: 254–260, doi:10.3978/j.issn.2223-4683.2015.04.08.

Roy S, Andres SN, Vergnes A, *et al.* (2012) XRCC4's interaction with XLF is required for coding (but not signal) end joining. *Nucleic Acids Res* 40: 1684–1694, doi:10.1093/nar/gkr1315.

Roy S, de Melo AJ, Xu Y, *et al.* (2015) XRCC4/XLF Interaction Is Variably Required for DNA Repair and Is Not Required for Ligase IV Stimulation. *Mol Cell Biol* 35: 3017–3028, doi:10.1128/MCB.01503-14.

Sacho EJ, Maizels N (2011) DNA Repair Factor MRE11/RAD50 Cleaves 3'-Phosphotyrosyl Bonds and Resects DNA to Repair Damage Caused by Topoisomerase 1 Poisons. *J Biol Chem* 286: 44945–44951, doi:10.1074/jbc.M111.299347.

Saito S, Goodarzi AA, Higashimoto Y, *et al.* (2002) ATM mediates phosphorylation at multiple p53 sites, including Ser(46), in response to ionizing radiation. *J Biol Chem* 277: 12491–12494, doi:10.1074/jbc.C200093200.

Sakakibara N, Mitra R, McBride AA (2011) The papillomavirus E1 helicase activates a cellular DNA damage response in viral replication foci. *J Virol* 85: 8981–8995, doi:10.1128/JVI.00541-11.

Salati SA, Al Kadi A (2012) Anal cancer - a review. *Int J Health Sci (Qassim)* 6: 206–230.

Saleh-Gohari N, Bryant HE, Schultz N, *et al.* (2005) Spontaneous homologous recombination is induced by collapsed replication forks that are caused by endogenous DNA single-strand breaks. *Mol Cell Biol* 25: 7158–7169, doi:10.1128/MCB.25.16.7158-7169.2005.

San Filippo J, Chi P, Sehorn MG, *et al.* (2006) Recombination mediator and Rad51 targeting activities of a human BRCA2 polypeptide. *J Biol Chem* 281: 11649–11657, doi:10.1074/jbc.M601249200.

Sandoval J, Esteller M (2012) Cancer epigenomics: beyond genomics. *Curr Opin Genet Dev* 22: 50–55, doi:10.1016/j.gde.2012.02.008.

Sartori AA, Lukas C, Coates J, *et al.* (2007) Human CtIP promotes DNA end resection. *Nature* 450: 509–514, doi:10.1038/nature06337.

Satyanarayana A, Hilton MB, Kaldis P (2008) p21 Inhibits Cdk1 in the absence of Cdk2 to maintain the G1/S phase DNA damage checkpoint. *Mol Biol Cell* 19: 65–77, doi:10.1091/mbc.E07-06-0525.

Adapted by permission from [Springer Nature]: [Nature] [Nature Reviews Drug Discovery] Tissue microarrays in drug discovery. Sauter G, Simon R, Hillan K (2003) *2*:962-972, doi:10.1038/nrd1254

Schiffmann J, Gandaglia G, Larcher A, *et al.* (2014) Contemporary 90-day mortality rates after radical cystectomy in the elderly. *Eur J Surg Oncol* 40: 1738–1745, doi:10.1016/j.ejso.2014.10.004.

Schiller CB, Lammens K, Guerini I, *et al.* (2012) Structure of Mre11-Nbs1 complex yields insights into ataxia-telangiectasia-like disease mutations and DNA damage signaling. *Nat Struct Mol Biol* 19: 693–700, doi:10.1038/nsmb.2323.

Schlacher K, Christ N, Siaud N, *et al.* (2011) Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell* 145: 529–542, doi:10.1016/j.cell.2011.03.041.

Scholzen T, Gerdes J (2000) The Ki-67 protein: From the known and the unknown. *J Cell Physiol* 182: 311–322, doi:10.1002/(SICI)1097-4652(200003)182:3<311::AID-JCP1>3.0.CO;2-9.

Seferina SC, Nap M, van den Berkmortel F, *et al.* (2013) Reliability of receptor assessment on core needle biopsy in breast cancer patients. *Tumor Biol* 34: 987–994, doi:10.1007/s13277-012-0635-5.

Seifert FU, Lammens K, Stoehr G, *et al.* (2016) Structural mechanism of ATP-dependent DNA binding and DNA end bridging by eukaryotic Rad50. *EMBO J* 35: 759–772, doi:10.15252/embj.201592934.

Seisen T, Compérat E, Léon P, Roupret M (2014) Impact of histological variants on the outcomes of nonmuscle invasive bladder cancer after transurethral resection. *Curr Opin Urol* 24: 524–531, doi:10.1097/MOU.0000000000000086.

Seligson DB, Horvath S, Shi T, *et al.*(2005) Global histone modification patterns predict risk of prostate cancer recurrence. *Nature* 435: 1262–1266, doi:10.1038/nature03672.

Seo AN, Kwak Y, Kim D-W, *et al.* (2014) HER2 status in colorectal cancer: its clinical significance and the relationship between HER2 gene amplification and expression. *PLoS One* 9: e98528, doi:10.1371/journal.pone.0098528.

Serup-Hansen E, Linnemann D, Skovrider-Ruminski W, *et al.* (2014) Human papillomavirus genotyping and p16 expression as prognostic factors for patients with American Joint Committee on Cancer stages I to III carcinoma of the anal canal. *J Clin Oncol* 32: 1812–1817, doi:10.1200/JCO.2013.52.3464.

Shibata A, Moiani D, Arvai AS, *et al.* (2014) DNA double-strand break repair pathway choice is directed by distinct MRE11 nuclease activities. *Mol Cell* 53: 7–18, doi:10.1016/j.molcel.2013.11.003.

Shiels MS, Kreimer AR, Coghill AE, *et al.* (2015) Anal Cancer Incidence in the United States, 1977-2011: Distinct Patterns by Histology and Behavior. *Cancer Epidemiol Biomarkers Prev* 24: 1548–1556, doi:10.1158/1055-9965.EPI-15-0044.

Shin M-K, Balsitis S, Brake T, Lambert PF (2009) Human papillomavirus E7 oncoprotein overrides the tumor suppressor activity of p21Cip1 in cervical carcinogenesis. *Cancer Res* 69: 5656–5663, doi:10.1158/0008-5472.CAN-08-3711.

Shipley WU, Winter KA, Kaufman DS, *et al.* (1999) Phase III Trial of Neoadjuvant Chemotherapy in Patients With Invasive Bladder Cancer Treated With Selective Bladder Preservation by Combined Radiation Therapy and Chemotherapy: Initial Results of Radiation Therapy Oncology Group 89-03. *J Urol* 162: 623–624, doi:10.1016/S0022-5347(05)68645-4.

Short JM, Liu Y, Chen S, *et al.* (2016) High-resolution structure of the presynaptic RAD51 filament on single-stranded DNA by electron cryo-microscopy. *Nucleic Acids Res* 106: gkw783, doi:10.1093/nar/gkw783.

Shridhar R, Shibata D, Chan E, Thomas CR (2015) Anal cancer: Current standards in care and recent changes in practice. *CA Cancer J Clin* 65: 139–162, doi:10.3322/caac.21259.

Simon R, Altman DG (1994) Statistical aspects of prognostic factor studies in oncology. *Br J Cancer* 69: 979–985.

Skvortsov S, Debbage P, Cho WC, *et al.* (2014) Putative biomarkers and therapeutic targets associated with radiation resistance. *Expert Rev Proteomics* 11: 207–214, doi:10.1586/14789450.2014.893194.

Slamon DJ, Clark GM, Wong SG, *et al.* (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 235: 177–182.

SNPeadia (2015) http://www.snpedia.com/index.php/Rs1805363 (accessed: 30/01/2017).

So S, Davis AJ, Chen DJ (2009) Autophosphorylation at serine 1981 stabilizes ATM at DNA damage sites. *J Cell Biol* 187: 977–990, doi:10.1083/jcb.200906064.

Sogo JM, Lopes M, Foiani M (2002) Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. *Science* 297: 599–602, doi:10.1126/science.1074023.

Sood A, Kachroo N, Abdollah F, *et al.* (2017) An Evaluation of the Timing of Surgical Complications Following Radical Cystectomy: Data From the American College of Surgeons National Surgical Quality Improvement Program. *Urology* doi:10.1016/j.urology.2017.01.036.

Spagnolo L, Rivera-Calzada A, Pearl LH, Llorca O (2006) Three-Dimensional Structure of the Human DNA-PKcs/Ku70/Ku80 Complex Assembled on DNA and Its Implications for DNA DSB Repair. *Mol Cell* 22: 511–519, doi:10.1016/j.molcel.2006.04.013.

Sparano JA, Gray RJ, Makower DF, *et al.* (2015) Prospective Validation of a 21-Gene Expression Assay in Breast Cancer. *N Engl J Med* 373: 2005–2014, doi:10.1056/NEJMoa1510764.

de Stanchina E, Querido E, Narita M, *et al.* (2004) PML is a direct p53 target that modulates p53 effector functions. *Mol Cell* 13: 523–535.

Stracker TH, Carson CT, Weitzman MD (2002) Adenovirus oncoproteins inactivate the Mre11–Rad50–NBS1 DNA repair complex. *Nature* 418: 348–352, doi:10.1038/nature00863.

Stracker TH, Lee D V, Carson CT, *et al.* (2005) Serotype-Specific Reorganization of the Mre11 Complex by Adenoviral E4orf3 Proteins. *J Virol* 79: 6664–6673, doi:10.1128/JVI.79.11.6664-6673.2005.

Strand S, Orntoft T, Sorensen K (2014) Prognostic DNA Methylation Markers for Prostate Cancer. *Int J Mol Sci* 15: 16544–16576, doi:10.3390/ijms150916544.

Strande NT, Waters CA, Ramsden DA, *et al.* (2012) Resolution of complex ends by Nonhomologous end joining - better to be lucky than good? *Genome Integr* 3: 10, doi:10.1186/2041-9414-3-10.

Stratton MR, Campbell PJ, Futreal PA (2009) The cancer genome. *Nature* 458: 719–724, doi:10.1038/nature07943.

Strom CE, Johansson F, Uhlen M, *et al.* (2011) Poly (ADP-ribose) polymerase (PARP) is not involved in base excision repair but PARP inhibition traps a single-strand intermediate. *Nucleic Acids Res* 39: 3166–3175, doi:10.1093/nar/gkq1241.

Stuart-Harris R, Caldas C, Pinder SE, Pharoah P (2008) Proliferation markers and survival in early breast cancer: A systematic review and meta-analysis of 85 studies in 32,825 patients. *The Breast* 17: 323–334, doi:10.1016/j.breast.2008.02.002.

Stupp R, Hegi ME, Mason WP, *et al.* (2009) Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol* 10: 459–466, doi:10.1016/S1470-2045(09)70025-7.

Sturgeon C, Hill R, Hortin GL, Thompson D (2010) Taking a new biomarker into routine use--a perspective from the routine clinical biochemistry laboratory. *Proteomics Clin Appl* 4: 892–903, doi:10.1002/prca.201000073.

Sugiyama T, Kowalczykowski SC (2002) Rad52 Protein Associates with Replication Protein A (RPA)-Single-stranded DNA to Accelerate Rad51-mediated Displacement of RPA and Presynaptic Complex Formation. *J Biol Chem* 277: 31663–31672, doi:10.1074/jbc.M203494200.

Tadi SK, Sebastian R, Dahal S, *et al.* (2016) Microhomology-mediated end joining is the principal mediator of double-strand break repair during mitochondrial DNA lesions. *Mol Biol Cell* 27: 223–235, doi:10.1091/mbc.E15-05-0260.

Tanaka T, Suh KS, Lo AM, De Luca LM (2007) p21 $^{WAF1/CIP1}$ Is a Common Transcriptional Target of Retinoid Receptors. *J Biol Chem* 282: 29987–29997, doi:10.1074/jbc.M701700200.

Taylor AMR, Groom A, Byrd PJ (2004) Ataxia-telangiectasia-like disorder (ATLD)—its clinical presentation and molecular basis. *DNA Repair (Amst)* 3: 1219–1225, doi:10.1016/j.dnarep.2004.04.009.

Taylor CF, Platt FM, Hurst CD, *et al.* (2014) Frequent inactivating mutations of STAG2 in bladder cancer are associated with low tumour grade and stage and inversely related to chromosomal copy number changes. *Hum Mol Genet* 23: 1964–1974, doi:10.1093/hmg/ddt589.

Teo MTW, Dyrskjøt L, Nsengimana J, *et al.* (2014) Next-generation sequencing identifies germline MRE11A variants as markers of radiotherapy outcomes in muscle-invasive bladder cancer. *Ann Oncol Off J Eur Soc Med Oncol* 25: 877–883, doi:10.1093/annonc/mdu014. https://creativecommons.org/licenses/by-nc/4.0/legalcode.

Thariani R, Veenstra DL, Carlson JJ, *et al.* (2012) Paying for personalized care: Cancer biomarkers and comparative effectiveness. *Mol Oncol* 6: 260–266, doi:10.1016/j.molonc.2012.02.006.

Theodosiou Z, Kasampalidis IN, Livanos G, *et al.* (2007) Automated analysis of FISH and immunohistochemistry images: A review. *Cytom Part A* 71A: 439–450, doi:10.1002/cyto.a.20409.

Thomas M, Pim D, Banks L (1999) The role of the E6-p53 interaction in the molecular pathogenesis of HPV. *Oncogene* 18: 7690–7700, doi:10.1038/sj.onc.1202953.

Thomsen MBH, Nordentoft I, Lamy P, *et al.* (2017) Comprehensive multiregional analysis of molecular heterogeneity in bladder cancer. *Sci Rep* 7: 11702, doi:10.1038/s41598-017-11291-0.

Tilki D, Brausi M, Colombo R, *et al.* (2013) Lymphadenectomy for Bladder Cancer at the Time of Radical Cystectomy. *Eur Urol* 64: 266–276, doi:10.1016/j.eururo.2013.04.036.

Trenz K, Smith E, Smith S, Costanzo V (2006) ATM and ATR promote Mre11 dependent restart of collapsed replication forks and prevent accumulation of DNA breaks. *EMBO J* 25: 1764–1774, doi:10.1038/sj.emboj.7601045.

Tretyakova NY, Groehler A, Ji S, Ji S (2015) DNA-Protein Cross-Links: Formation, Structural Identities, and Biological Outcomes. *Acc Chem Res* 48: 1631–1644, doi:10.1021/acs.accounts.5b00056.

True LD (2008) Quality control in molecular immunohistochemistry. *Histochem Cell Biol* 130: 473–480, doi:10.1007/s00418-008-0481-0.

Truong LN, Li Y, Shi LZ, *et al.* (2013) Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proc Natl Acad Sci U S A* 110: 7720–7725, doi:10.1073/pnas.1213431110.

Uchisaka N, Takahashi N, Sato M, *et al.* (2009) Two Brothers with Ataxia-Telangiectasia-like Disorder with Lung Adenocarcinoma. *J Pediatr* 155: 435–438, doi:10.1016/j.jpeds.2009.02.037.

Uziel T, Lerenthal Y, Moyal L, *et al.* (2003) Requirement of the MRN complex for ATM activation by DNA damage. *EMBO J* 22: 5612–5621, doi:10.1093/emboj/cdg541.

Vale CL (2005a) Adjuvant Chemotherapy in Invasive Bladder Cancer: A Systematic Review and Meta-Analysis of Individual Patient Data. *Eur Urol* 48: 189–201, doi:10.1016/j.eururo.2005.04.005.

Vale CL (2005b) Neoadjuvant Chemotherapy in Invasive Bladder Cancer: Update of a Systematic Review and Meta-Analysis of Individual Patient Data. *Eur Urol* 48: 202–206, doi:10.1016/j.eururo.2005.04.006.

Veeratterapillay R, Heer R, Johnson MI, *et al.* (2016) High-Risk Non-Muscle-Invasive Bladder Cancer-Therapy Options During Intravesical BCG Shortage. *Curr Urol Rep* 17: 68, doi:10.1007/s11934-016-0625-z.

Villarreal DD, Lee K, Deem A, *et al.* (2012) Microhomology Directs Diverse DNA Break Repair Pathways and Chromosomal Translocations. *PLoS Genet* 8: e1003026, doi:10.1371/journal.pgen.1003026.

Vinayan A, Glynne-Jones R (2016) Anal cancer – What is the optimum chemoradiotherapy? *Best Pract Res Clin Gastroenterol* 30: 641–653, doi:10.1016/j.bpg.2016.06.005.

Voduc D, Kenney C, Nielsen TO (2008) Tissue microarrays in clinical oncology. *Semin Radiat Oncol* 18: 89–97, doi:10.1016/j.semradonc.2007.10.006.

Walker JR, Corpina RA, Goldberg J (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* 412: 607–614, doi:10.1038/35088000.

Wang B, Xiao Z, Ko HL, Ren E-C (2010) The p53 response element and transcriptional repression. *Cell Cycle* 9: 870–879, doi:10.4161/cc.9.5.10825.

Wang F, Flanagan J, Su N, *et al.* (2012) RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J Mol Diagn* 14: 22–29, doi:10.1016/j.jmoldx.2011.08.002.

Wang H, Liu S, Zhang P, *et al.* (2009) S-Phase Cells Are More Sensitive to High-Linear Energy Transfer Radiation. *Int J Radiat Oncol* 74: 1236–1241, doi:10.1016/j.ijrobp.2008.12.089.

Warford A, Howat W, McCafferty J (2004) Expression profiling by high-throughput immunohistochemistry. *J Immunol Methods* 290: 81–92, doi:10.1016/j.jim.2004.04.010.

Waring P, Tie J, Maru D, Karapetis CS (2016) RAS Mutations as Predictive Biomarkers in Clinical Management of Metastatic Colorectal Cancer. *Clin Colorectal Cancer* 15: 95–103, doi:10.1016/j.clcc.2015.10.006.

Weber AM, Ryan AJ (2015) ATM and ATR as therapeutic targets in cancer. *Pharmacol Ther* 149: 124–138, doi:10.1016/J.PHARMTHERA.2014.12.001.

Weinstein JN, Akbani R, Broom BM, *et al.* (2014) Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* 507: 315–322, doi:10.1038/nature12965.

Wen Q, Scorah J, Phear G, *et al.* (2008) A mutant allele of MRE11 found in mismatch repair-deficient tumor cells suppresses the cellular response to DNA replication fork stress in a dominant negative manner. *Mol Biol Cell* 19: 1693–1705, doi:10.1091/mbc.E07-09-0975.

Wick W, Platten M, Meisner C, *et al.* (2012) Temozolomide chemotherapy alone versus radiotherapy alone for malignant astrocytoma in the elderly: the NOA-08 randomised, phase 3 trial. *Lancet Oncol* 13: 707–715, doi:10.1016/S1470-2045(12)70164-X.

Williams GJ, Lees-Miller SP, Tainer JA (2010a) Mre11-Rad50-Nbs1 conformations and the control of sensing, signaling, and effector responses at DNA double-strand breaks. *DNA Repair (Amst)* 9: 1299–1306, doi:10.1016/j.dnarep.2010.10.001.

Williams JR, Zhang Y, Zhou H, Gridley DS, Koch CJ, Dicello JF, Slater JM, Little JB (2010b) Tumor response to radiotherapy is dependent on genotype-associated mechanisms in vitro and in vivo. *Radiat Oncol* 5: 71, doi:10.1186/1748-717X-5-71.

Williams RS, Moncalian G, Williams JS, *et al.* (2008) Mre11 dimers coordinate DNA end bridging and nuclease processing in double-strand-break repair. *Cell* 135: 97–109, doi:10.1016/j.cell.2008.08.017.

Williams VM, Filippova M, Filippov V, *et al.* (2014) Human papillomavirus type 16 E6* induces oxidative stress and DNA damage. *J Virol* 88: 6751–6761, doi:10.1128/JVI.03355-13.

Wong NACS, Gonzalez D, Salto-Tellez M, *et al.* (2014) RAS testing of colorectal carcinoma—a guidance document from the Association of Clinical Pathologists Molecular Pathology and Diagnostics Group. *J Clin Pathol* 67: 751–757, doi:10.1136/jclinpath-2014-202467.

Wosnitzer MS, Hruby GW, Murphy AM, *et al.* (2012) A comparison of the outcomes of neoadjuvant and adjuvant chemotherapy for clinical T2-T4aN0-N2M0 bladder cancer. *Cancer* 118: 358–364, doi:10.1002/cncr.26278.

Wu VS, Kanaya N, Lo C, *et al.* (2015) From bench to bedside: What do we know about hormone receptor-positive and human epidermal growth factor receptor 2-positive breast cancer? *J Steroid Biochem Mol Biol* 153: 45–53, doi:10.1016/j.jsbmb.2015.05.005.

Wu X-R (2005) Urothelial tumorigenesis: a tale of divergent pathways. *Nat Rev Cancer* 5: 713–725, doi:10.1038/nrc1697.

Xia B, Sheng Q, Nakanishi K, *et al.* (2006) Control of BRCA2 Cellular and Clinical Functions by a Nuclear Partner, PALB2. *Mol Cell* 22: 719–729, doi:10.1016/j.molcel.2006.05.022.

Xiao Y, Weaver DT (1997) Conditional gene targeted deletion by Cre recombinase demonstrates the requirement for the double-strand break repair Mre11 protein in murine embryonic stem cells. *Nucleic Acids Res* 25: 2985–2991.

Xie M, Park D, You S, *et al.* (2015) Bcl2 inhibits recruitment of Mre11 complex to DNA double-strand breaks in response to high-linear energy transfer radiation. *Nucleic Acids Res* 43: 960–972, doi:10.1093/nar/gku1358.

Xu H, Wang Z, Jin S, *et al.* (2014) Dux4 induces cell cycle arrest at G1 phase through upregulation of p21 expression. *Biochem Biophys Res Commun* 446: 235–240, doi:10.1016/j.bbrc.2014.02.105.

Xylinas E, Hassler MR, Zhuang D, *et al.* (2016) An Epigenomic Approach to Improving Response to Neoadjuvant Cisplatin Chemotherapy in Bladder Cancer. *Biomolecules* 6: doi:10.3390/biom6030037.

Yang LS, Shan BL, Shan LL, *et al.* (2016) A systematic review and meta-analysis of quality of life outcomes after radical cystectomy for bladder cancer. *Surg Oncol* 25: 281–297, doi:10.1016/j.suronc.2016.05.027.

Yang Y, Liu Z, Wang F, *et al.* (2015) FANCD2 and REV1 cooperate in the protection of nascent DNA strands in response to replication stress. *Nucleic Acids Res* 43: 8325–8339, doi:10.1093/nar/gkv737.

Yin M, Joshi M, Meijer RP, *et al.* (2016) Neoadjuvant Chemotherapy for Muscle-Invasive Bladder Cancer: A Systematic Review and Two-Step Meta-Analysis. *Oncologist* 21: 708–715, doi:10.1634/theoncologist.2015-0440.

Ying S, Hamdy FC, Helleday T (2012) Mre11-Dependent Degradation of Stalled DNA Replication Forks Is Prevented by BRCA2 and PARP1. *Cancer Res* 72: 2814–2821, doi:10.1158/0008-5472.CAN-11-3417.

Yoo S, Dynan WS (1999) Geometry of a complex formed by double strand break repair proteins at a single DNA end: recruitment of DNA-PKcs induces inward translocation of Ku protein. *Nucleic Acids Res* 27: 4679–4686.

You Z, Chahwan C, Bailis J, *et al.* (2005) ATM Activation and Its Recruitment to Damaged DNA Require Binding to the C Terminus of Nbs1. *Mol Cell Biol* 25: 5363–5379, doi:10.1128/MCB.25.13.5363-5379.2005.

Yu J, Zhang L, Hwang PM, *et al.* (2001) PUMA induces the rapid apoptosis of colorectal cancer cells. *Mol Cell* 7: 673–682.

Yu Z, Vogel G, Coulombe Y, *et al.* (2012) The MRE11 GAR motif regulates DNA double-strand break processing and ATR activation. *Cell Res* 22: 305–320, doi:10.1038/cr.2011.128.

Yuan C-H, Filippova M, Duerksen-Hughes P (2012) Modulation of apoptotic pathways by human papillomaviruses (HPV): mechanisms and implications for therapy. *Viruses* 4: 3831–3850, doi:10.3390/v4123831.

Yue X, Zhao Y, Xu Y, *et al.* (2017) Mutant p53 in Cancer: Accumulation, Gain-of-Function, and Therapy. *J Mol Biol* 429: 1595–1606, doi:10.1016/j.jmb.2017.03.030.

Yue Y, Astvatsaturyan K, Cui X, *et al.* (2016) Stratification of Prognosis of Triple-Negative Breast Cancer Patients Using Combinatorial Biomarkers. *PLoS One* 11: e0149661, doi:10.1371/journal.pone.0149661.

Yun MH, Hiom K (2009) CtIP-BRCA1 modulates the choice of DNA double-strand-break repair pathway throughout the cell cycle. *Nature* 459: 460–463, doi:10.1038/nature07955.

Zahran MH, El-Hefnawy AS, Zidan EM, *et al.* (2014) Health-related quality of life after radical cystectomy and neobladder reconstruction in women: Impact of voiding and continence status. *Int J Urol* 21: 887–892, doi:10.1111/iju.12452.

Zelic R, Fiano V, Grasso C, *et al.* (2015) Global DNA hypomethylation in prostate cancer development and progression: a systematic review. *Prostate Cancer Prostatic Dis* 18: 1–12, doi:10.1038/pcan.2014.45.

Zhang K, Wang X, Zhou B, Zhang L (2013) The prognostic value of MGMT promoter methylation in Glioblastoma multiforme: a meta-analysis. *Fam Cancer* 12: 449–458, doi:10.1007/s10689-013-9607-1.

Zhang X, Zhang M, Hou Y, *et al.* (2016) Single-cell analyses of transcriptional heterogeneity in squamous cell carcinoma of urinary bladder. *Oncotarget* 7: 66069–66076, doi:10.18632/oncotarget.11803.

Zhang Z, Chan DW (2010) The road from discovery to clinical diagnostics: lessons learned from the first FDA-cleared in vitro diagnostic multivariate index assay of proteomic biomarkers. *Cancer Epidemiol Biomarkers Prev* 19: 2995–2999, doi:10.1158/1055-9965.EPI-10-0580.

Zietman AL, Sacco D, Skowronski U, *et al.* (2003) Organ Conservation in Invasive Bladder Cancer by Transurethral Resection, Chemotherapy and Radiation: Results of a Urodynamic and Quality of Life Study on Long-term Survivors. *J Urol* 170: 1772–1776, doi:10.1097/01.ju.0000093721.23249.c3.