DNA repair in bladder cancer predisposition and radiotherapy treatment response

by

Mark Teo Teong Wey

Submitted in accordance with the

requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Medicine

December 2012

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Chapter 8.3 was formed from the publication, “The role of microRNA-binding site polymorphisms in DNA repair genes as risk factors for bladder cancer and breast cancer and their impact on radiotherapy outcomes” by Teo MT, Landi D, Taylor C, Elliott F, Vaslin L, Cox DG, Hall J, Landi S, Bishop DT, Kiltie AE (2012) *Carcinogenesis*, 33(3), 581-6.

The candidate undertook the sample preparation, genotype calling, statistical analysis (with support from Elliott F) and interpretation for the bladder cancer cohort, coordination between collaborators and the preparation of the manuscript for publication. Landi D and Landi S performed the bioinformatics analysis, Taylor C investigated the *ATM* SNPs, Vaslin L, Cox DG and Hall J analysed the breast cancer cohort, and Bishop DT and Kiltie AE provided supervision and along with Hall J, critical review of the final manuscript.

Mark Teo Teong Wey

1. Acknowledgements

I would first like to thank my supervisors, Dr Anne Kiltie and Professor Tim Bishop for their support, advice and guidance throughout all stages of this doctoral thesis. I have had significant support for the more technical aspects of this study as follows:

Dr Juliette Randerson-Moore and Dr Mark Harland for their general everyday support, teaching and help for most of the genetics aspects of this thesis especially in the preparation of master DNA sample plates. Ms Helen Snowden for her hard work with the library preparation and genotyping. Dr Claire Taylor for her invaluable training in high resolution melting mutation scanning and conventional Sanger sequencing, plus the genotyping of the *ATM* SNPs. Dr Joanne Morgan for training in long PCR amplification and, with Professor Graham Taylor, for performing the next-generation sequencing on the Illumina platform. Dr Phillip Chambers, Dr Gina Scott and Dr Jonathan Laye for their advice and support in developing an *MRE11A* isoform expression assay, and to Mr Christien Buchwald for the sequencing of the bladder cancer cell lines.

Dr Ian Carr for his support in the use of his custom-designed Illuminator software, and Dr Aengus Stewart and Dr Mike Mitchell for training in bioinformatics functional predictions. Dr Lee Hazlewood, Dr Alastair Droop and Mr John Alexander for the development of custom software tools and for local application of the Syzygy software. Ms Faye Elliott for her general statistical advice, training in STATA use and in database management. Professor Jenny Barrett and Dr Jeremie Nsengimana for their statistical expertise and in undertaking the collapsed rare variant analyses.

Dr Filomena Estevez and Dr Mike Shires for all their help and expertise in histopathology with tissue sectioning, and antibody and immunohistochemistry protocol optimisation. Dr Selina Bhattarai for bladder cancer pathology teaching, and the delineation of bladder cancer sections. Dr Boling Qiao for training in cell culture and Western blotting. Professor Maggie Knowles, Dr Carolyn Hurst, Dr Fiona Platt and Dr Joanne Brown for bladder tumour sample management and provision of primary bladder tumour and bladder cancer cell line samples. Mr Sei Chung Sak recruited bladder cancer cases and controls from 2002-2005, Mr Sanjeev Kotwal collected cases from 2005-2008, and Dr Ananya Choudhury collected controls from 2004-2007.

I would like to thank the collaborators involved in the work in this thesis. In the DNA repair gene microRNA-binding site SNP study, Dr Stefano Landi and Dr Debora Landi in the bioinformatics prediction of potential microRNA binding sites, and Prof Janet Hall and Dr David Cox for their work on their breast cancer cohort, and advice on the final haplotype analysis and contribution in the final published manuscript. In the validation of germline *MRE11A* variants, Professor Torben Orntoft and Dr Lars Dyrskjot Anderson for their contribution of a cystectomy treated cohort. The work in this thesis was funded by Cancer Research UK and Yorkshire Cancer Research.

Finally, I would like to thank my wife for all her support and patience throughout my fellowship and putting up with the long hours I have spent in the laboratory and in writing this thesis (and apologise for bringing work on holiday).

2. Abstract

The genetic contribution to bladder cancer risk remains undetermined, while the role of radiotherapy versus surgery in muscle-invasive bladder cancer (MIBC) treatment is hotly debated with the need for predictive biomarkers of treatment response. DNA repair pathways are involved in repairing DNA damage from carcinogens thus preventing carcinogenesis, but also form one of the 5 R's of radiobiology for determining cancer response to radiotherapy.

The aims of this project were:

1) To study the contribution of germline DNA repair gene variants, specifically rare variants (RV) and 3'-untranslated region (3'UTR) single nucleotide polymorphisms (SNP), on bladder cancer risk.
2) To investigate the predictive value of germline DNA repair gene variants and tumour DNA repair protein expression on radiotherapy outcomes in MIBC.

RVs can only be identified by sequencing so a developmental multiplexed next-generation sequencing (NGS) project was undertaken, identifying two approaches, with the choice of method based on balancing costs and labour versus accuracy and data needed. Using these methods, candidate RVs were identified in the DNA repair genes, *MUTYH* and *XPC*, with *XPC* RVs being associated with an increased bladder cancer risk (P=0.008) independent of previously identified GWAS SNPs. Putatively functional DNA repair gene 3'UTR SNPs, *PARP1* rs8679 and *RAD51* rs7180135, were found to increase bladder cancer risk (P=0.05) and predict improved survival following radiotherapy (P=0.01) respectively. Multiplexed NGS of *MRE11A* identified rs1805363 to be predictive of survival following radiotherapy (P=0.001) but not surgery (P=0.89), and to affect *MRE11A* isoform expression. Tumour DNA repair

protein expression of CtIP, MUTYH and XPC were not found to predict survival following radiotherapy.

This study demonstrated the contribution of DNA repair gene variants in bladder cancer risk and predicting radiotherapy response. These findings could contribute to the goal of personalised medicine for targeted primary prevention, early diagnosis and treatment individualisation.

Table of contents

Table of figures

Table of tables

3. Glossary

∆G	Gibbs binding free energy
Ago	Argonaute
BER	base excision repair
bp	base pairs
BSA	bovine serum albumin
CDCV	Common Disease – Common Variant hypothesis
CDRV	Common Disease – Rare Variant hypothesis
CI	confidence interval
CSS	cause specific survival
CT	computed tomography
CTV	clinical target volume
DSB	double strand breaks
dsDNA	double stranded DNA
EORTC	European Organization for Research and Treatment of Cancer
FFPE	formalin-fixed paraffin-embedded
FPRP	False Positive Reporting Probability
Gb	gigabases
GGR	global genome repair
GWAS	genome wide association study
H&E	haematoxylin and eosin
HR	hazard ratio
HRM	high resolution melting
HRR	homologous recombination repair
IFC	Integrated Fluidic Circuit
IR	ionising radiation
LBCS	Leeds Bladder Cancer Study
LD	linkage disequilibrium
LIMM	Leeds Institute of Molecular Medicine
LOD	logarithm of odds
LTHT	Leeds Teaching Hospitals NHS Trust
MAF	minor allele frequency
MAP	*MUTYH*-associated polyposis
Mb	megabases
MIBC	muscle-invasive bladder cancer

miRNA	micro-RNA
MMEJ	microhomology-mediated end joining
MRI	magnetic resonance imaging
MRN	MRE11-RAD50-NBS1 complex
NCI	National Cancer Institute
ncRNA	non-coding RNA
NER	nucleotide excision repair
NGS	next-generation sequencing
NHEJ	non-homologous end joining
OR	odds ratio
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pre-miRNA	precursor-microRNA
pri-miRNA	primary-microRNA
psi	pound-force per square inch
RCT	randomised control trial
RISC	RNA-induced silencing complex
RT	radiotherapy
RTOG	Radiation Therapy Oncology Group
RV	rare variant
SBC	streptavidin-biotin complex
SIR	standardised incidence ratio
SJIO	St James's Institute of Oncology
SJUH	St James's University Hospital
SNP	single nucleotide polymorphism
SNV	single nucleotide variants
SQS	semi-quantitative score
TBE	Tris/ Borate/ EDTA
TBS	Tris-buffered Saline
TCR	transcription-coupled repair
TURBT	transurethral resection of bladder tumour
UC	urothelial carcinoma
UCSC	University of California Santa Cruz
VUS	variants of unknown significance
WTCC	Wellcome Trust Case-Control Consortium

4. Glossary of Next-Generation Sequencing Terminology

Bar-coding	The tagging of DNA fragments in each sequencing library with a unique six oligonucleotide sequence. Used together with multiplexing.
Coverage (Depth of)	The average number of times each nucleotide base has been sequenced.
Indexing	Same as “Bar-coding”.
Multiplexing	The pooling of multiple different DNA samples together.
Paired-end Reads	The sequencing of DNA fragments from the 5’- to 3’-end followed by sequencing from the 3’- to 5’-end.
Read Depth	Same as “Coverage”
Read Length	The number of contiguous nucleotide bases sequenced for each DNA fragment. On the Illumina platform, usually user-defined by the number of DNA sequencing-by-synthesis cycles performed (ie. the number of cycles labelled nucleotides are added and imaged).
Sequence Alignment	The lining up of DNA fragment sequences against a reference sequence and to each other to determine the DNA sample’s sequence.
Sequencing Depth	Same as “Coverage”.
Single-end Reads	The sequencing of DNA fragments from the 5’- to 3’-end only.
Target Size	The total length (in nucleotide base pairs) of the DNA sample being sequenced.
Variant Calling	The identification of specific sequence variation in the DNA sample compared to the reference sequence used for sequence alignment.

5. Introduction

5.1. Bladder cancer

5.1.1. Epidemiology

Bladder cancer is the seventh most common cancer in the United Kingdom (UK) and the fourth commonest cancer among men, with approximately 11000 new cases diagnosed per year (Cancer Research UK 2010). It is also the eighth highest cause of cancer death in the UK killing approximately 5000 people per year and is the sixth most common cause of male cancer death. In the United States, bladder cancer has the fifth highest healthcare budgetary costs amongst all cancers and the highest monitoring and treatment costs per patient from diagnosis to death (Botteman *et al.* 2003). Bladder cancer is predominantly seen in men, with a male to female ratio of 5:2, and the elderly, with 80 percent of cases occurring in people over the age of 65 years.

In the developed world, tobacco smoking and occupational exposure to industrial chemicals, such as from the dye, printing and rubber industries, are major pre-disposing risk factors for bladder cancer (Glashan and Cartwright 1981; Morrison *et al.* 1984; Kogevinas *et al.* 1998). The population attributable risk (the additional disease risk in a population due to an exposure) of bladder cancer from cigarette smoking is about 50% with current smokers having quadruple the bladder cancer risk of never smokers (Freedman *et al.* 2011). In the UK, occupational carcinogen exposure was estimated to account for 5.3% (7.1% in men, 1.9% in women) of the population bladder cancer attributable risk and 245 attributable deaths per year (Brown *et al.* 2012). However, in Africa especially Egypt, chronic bladder inflammation secondary to endemic schistosomiasis parasitic infections is the main mechanism of carcinogenesis leading to high bladder cancer rates (Mostafa *et al.* 1999).

Genetic factors also play a role in bladder cancer pre-disposition as suggested by racial differences and the association of family history with bladder cancer risk. Caucasian Americans have a higher incidence of bladder cancer compared to African Americans with a greater dose-response effect with increasing cigarette consumption on bladder cancer risk (Schairer *et al.* 1988; Harris *et al.* 1990). Kantor *et al* reported that a family history of a first-degree relative with bladder cancer resulted in an increased bladder cancer risk (relative risk = 1.45) in the United States (Kantor *et al.* 1985), while in Sweden, Plna *et al* observed familial risks of 1.35 to 2.29 (Plna and Hemminki 2001). Similar findings were then reported by Aben *et* al in a Dutch population (odds ratio (OR) 1.8 [95% confidence interval (CI) 1.3 – 2.7]) (Aben *et al.* 2002) and by the Spanish Bladder Cancer Study (OR 1.32 [95%CI 1.01 - 1.50]) who also reported an increased risk of developing early onset disease (Murta-Nascimento *et al.* 2007).

5.1.2. Histopathology and Carcinogenesis

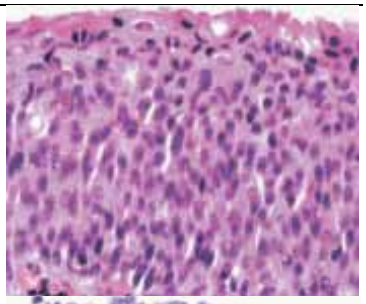
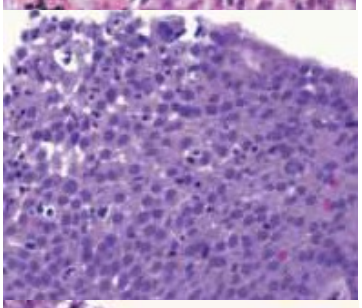
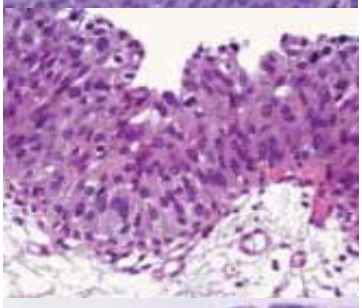
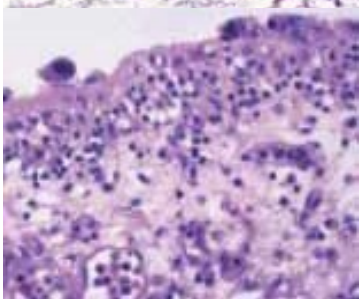
5.1.2.1. Histopathology

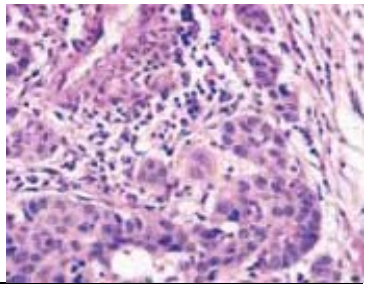
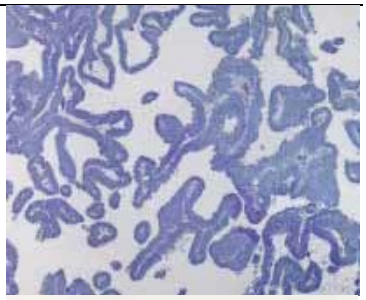
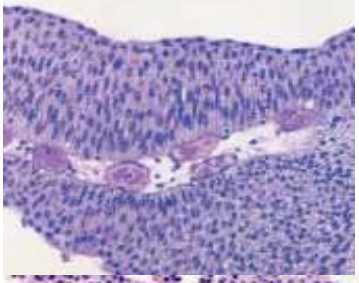
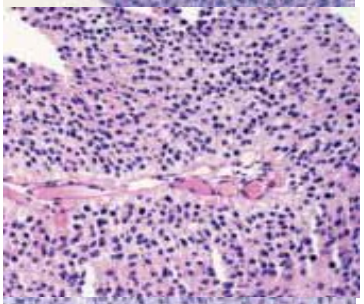
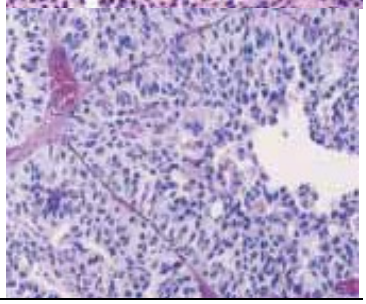
In the Western world, urothelial carcinoma (UC) accounts for more than 90% of bladder cancer and, as the name denotes, originates from the bladder urothelial epithelium. Macroscopically, UC tumours present as flat or papillary lesions while microscopic grading is based on cytological and architectural atypia, nuclear pleomorphisms and mitoses (Stevens and Lowe 1995; Grignon 2009). The 2004 WHO/ 1998 ISUP morphologic classification is typically used for histological grading and is summarised in Table 1.

Due to the entire bladder urothelium being exposed to causative carcinogens in the urine, there is often a “field-change” throughout the urothelium with 25% of cases showing dysplasia around primary tumours, and 60% of cases developing

recurrences distant to the primary tumours with multi-focal or multiple co-existing UC tumours also seen (Harris and Neal 1992; Stevens and Lowe 1995). These co-existing multi-focal tumours were also found to be predominantly of oligoclonal origin, with each tumour having distinctly different genetic alterations (Jones *et al.* 2005). There is frequent divergent differentiation in UCs with a variable proportion of cancer cells showing adenomatous, squamous, or small cell differentiation with the latter two possibly indicative of a worse prognosis (Lopez-Beltran *et al.* 2004).

Table 1: 2004 WHO/ 1998 ISUP morphologic classification defining the microscopic morphologic appearances of bladder lesions (adapted from (Lopez-Beltran *et al.* 2004; Grignon 2009; Babjuk *et al.* 2011)). Images from (Lopez-Beltran *et al.* 2004)

Flat lesions	
Hyperplasia (flat lesion without atypia or papillary aspects)	
Reactive atypia (flat lesion with atypia) Atypia of unknown significance	
Dysplasia (Intra-Urothelial Neoplasia, low-grade)	
Carcinoma *in situ* (Intra-Urothelial Neoplasia, high-grade)	

Invasive carcinoma	
Papillary lesions	
Urothelial papilloma (benign lesion)	
Papillary urothelial neoplasm of low malignant potential	
Papillary urothelial carcinoma, low-grade	
Papillary urothelial carcinoma, high-grade	

The remaining approximately 10% of bladder cancers consist of pure squamous cell carcinomas (<5%), typically associated with chronic inflammation from calculi or schistosomiasis, adenocarcinomas (< 2%) usually originating from urachal remnants,

and rarer histological types such as small cell carcinoma, lymphoma and sarcoma. This thesis will predominantly focus on UCs of the bladder.

5.1.2.2. Molecular genetics of bladder carcinogenesis

Carcinogenesis is "the evolution of an invasive cancer cell from a normal cell" (Martin 1998). The multistep model of carcinogenesis involves sequential mutations resulting in down-regulation of tumour suppressor genes and up-regulation of oncogenes, genes that are involved in inhibiting or contributing to cancer development and progression respectively. These sequential mutations lead to initiation of a tumour, promotion of tumour growth and progression to invasive carcinoma (Okey *et al.* 2005). Multiple somatic mutations have been identified in bladder tumours affecting growth factor receptors, signalling pathways and cell-cycle control genes. Chromosomal abnormalities are seen in over 60% of bladder UCs with loss of heterozygosity reported in 2q (long arm of chromosome 2) (58%), 5q (6-50%), 8p (short arm of chromosome 8) (18-83%), 9p (33-82%), 9q (43-90%), 10q (39-45%), 11p (9-72%), 18q (36-51%) and the Y chromosome (11%) as well as amplifications at 1q (37-54%), 5p (24-25%), 8q (37-54%), and 17q (29-49%) (Lopez-Beltran *et al.* 2004; Knowles 2006). Identification of all the genes responsible for tumour survival and growth in these large chromosomal regions remains work in progress.

However, distinct patterns of somatic mutations have been identified in papillary and flat/ dysplastic UC lesions with allelic loss of chromosome 9 being more common in the former while *TP53* mutations and loss of chromosome 17 are more common in the latter (Spruck *et al.* 1994; Knowles 2006). Two separate bladder tumorigenesis pathways have thus been proposed for the development of these two phenotypically different UC tumour subtypes (Figure 1) (Castillo-Martin *et al.* 2010; Goebell and Knowles 2010).

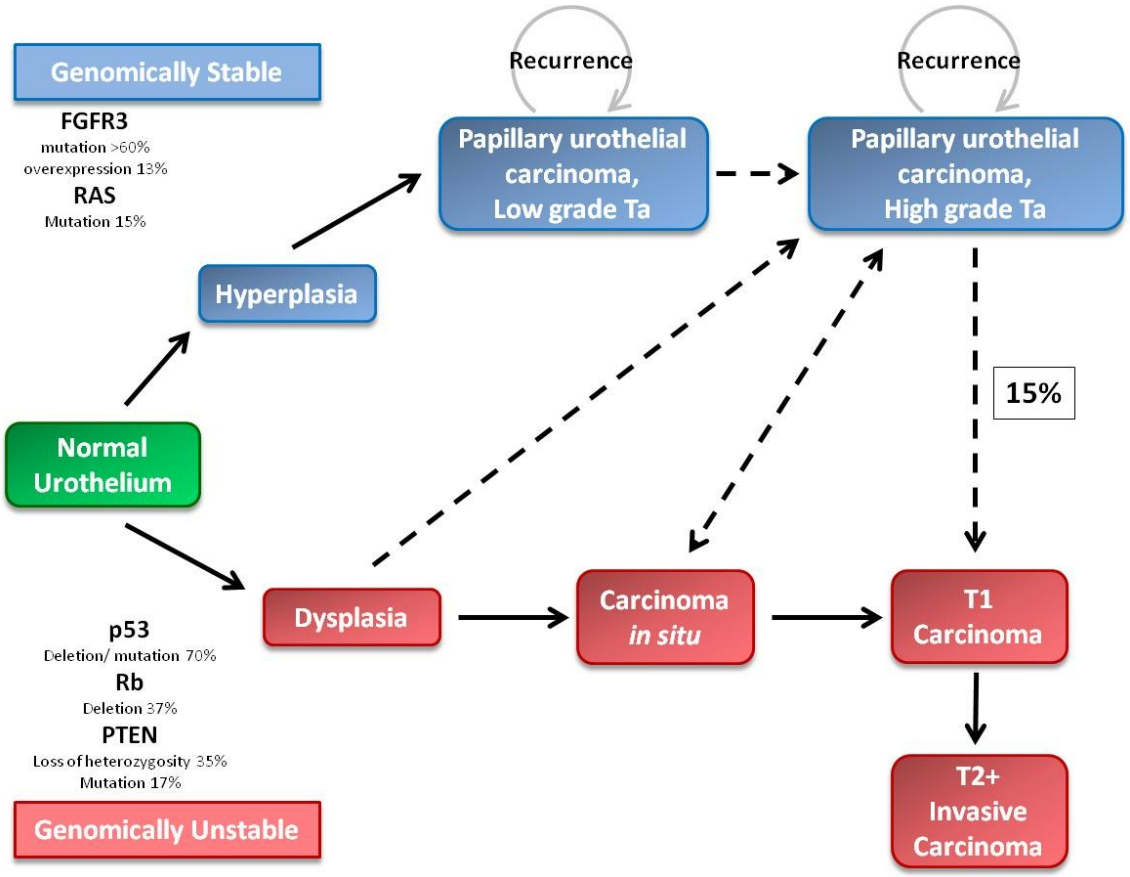


Figure 1: Pathways of bladder carcinogenesis (adapted from (Goebell and Knowles 2010)). The genomically stable papillary tumour pathway (blue) involving *FGFR3* and *RAS* oncogene mutations, progressing to frequently recurring low grade papillary tumours, while the genomically unstable invasive carcinoma pathway (red) with mutations in *TP53*, *RB* and *PTEN* tumour suppressor genes, progressing to invasive carcinoma. The broken arrows indicate potential alternative pathways of progression.

The papillary tumour pathway, as previously stated, is typically associated with loss-of-heterozygosity in chromosome 9 and commonly early activating mutations in oncogenes such as *FGFR3*, *RAS* and in the PI3-kinase pathway (Knowles 2006; Castillo-Martin *et al.* 2010). FGFR3, fibroblast growth factor receptor 3, is a cell surface receptor thought to activate the Ras-MAPK signalling pathway. Activating mutations or overexpression of FGFR3 has been reported in 80% of superficial papillary tumours compared to 50% of invasive cancers (Goebell and Knowles 2010) and is not seen in papillary tumours carrying *Ras* mutations (Jebar *et al.* 2005), perhaps indicating distinct processes of papillary tumorigenesis. Despite frequent recurrences, papillary tumours remain genetically stable.

The invasive carcinoma pathway, however, is associated instead with tumour initiating deletions or mutations of tumour suppressor genes resulting in inactivation of *TP53, RB1,* and *PTEN* (Knowles 2006; McConkey *et al.* 2010). Mutations in *TP53*, a transcription factor involved in cell cycle arrest and apoptosis, are seen in 70% of invasive bladder carcinomas, but less than 20% of papillary tumours, with *TP53* and *FGFR3* mutations being almost mutually exclusive (Bakkar *et al.* 2003; Knowles 2006). *CDKN2A* mutations are frequently seen in both superficial and invasive tumours (Goebell and Knowles 2010). Invasive tumours are genetically unstable, frequently accumulating further genetic aberrations and mutations, thus progressing to more advanced disease.

5.1.3. Diagnosis and Staging

5.1.3.1. Clinical Presentation

Bladder cancer presents with frank haematuria with or without dysuria in 85% of cases (Lopez-Beltran *et al.* 2004); other patients present with symptoms of cystitis or anaemia (Blandy 1998). Tumours at the bladder neck may present with urgency or frequency while tumours obstructing the ureteric orifice may cause hydronephrosis with resulting flank pain and renal dysfunction. First presentation with extensive pelvic or metastatic disease is uncommon in bladder cancer.

5.1.3.2. Diagnosis and staging

The current gold standard for diagnosis of bladder cancer is by cystoscopy and transurethral resection of the bladder tumour (TURBT- see section 5.1.4.1, page 36) to obtain a biopsy specimen (British Association of Urological Surgeons and British Uro-oncology Group 2007). Further investigations and management will be dependent on the histological findings, multi-focal nature of the disease and previous

history and treatment for bladder cancer. WHO tumour staging is shown in Figure 2 while the full 2010 TNM (tumour, nodes, metastasis) classification in Table 2.

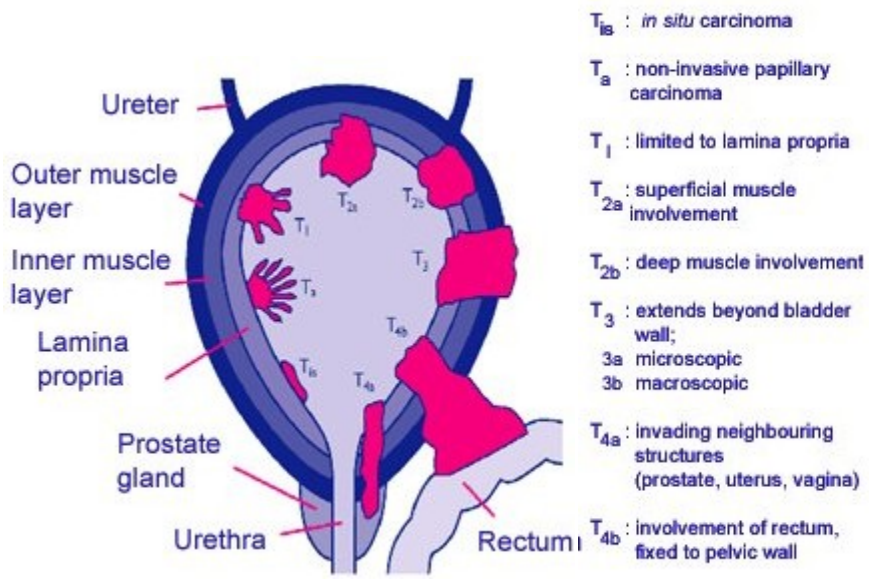


Figure 2 Bladder cancer T-staging (Cancer Research UK 2010)

High risk (large, multifocal or treatment-resistant tumours) or high grade superficial disease may require further resection to confirm the extent of disease, random biopsies to exclude multifocal disease or intravenous urography/ ultrasonography to detect synchronous upper urinary tract tumours (British Association of Urological Surgeons and British Uro-oncology Group 2007). About 80% of bladder cancers present as superficial papillary tumours with the remainder presenting as invasive disease. Fifteen percent of superficial disease will progress to become muscle invasive disease (Castillo-Martin *et al.* 2010). In view of this, regular follow-up cystoscopy is performed for all patients with superficial bladder cancers for the early detection of disease recurrence or progression (Babjuk *et al.* 2011).

Muscle-invasive disease consists of WHO T-stage of 2 or greater tumours (Table 2), with invasion into the detrusor muscle, the smooth muscle layer of the bladder wall, and is associated with access to the lymphatic drainage system, and hence the potential for tumour spread to regional lymph nodes and more distant spread (Blandy

1998). Staging for muscle-invasive disease requires an overall assessment of patient fitness for treatment and magnetic resonance imaging (MRI) of the pelvis and computed tomography (CT) of the thorax, abdomen and pelvis to assess the extent of bladder wall invasion, local and distant spread (British Association of Urological Surgeons and British Uro-oncology Group 2007). Routine blood tests performed are a full blood count for anaemia, biochemistry for renal dysfunction, and alkaline phosphatase for potential bone involvement.

Table 2: 2010 TNM classification for bladder cancer (adapted from (Edge *et al.* 2010)).

2010 Tumour Node Metastases Classification of Urinary Bladder Cancer

T - Primary Tumour		
Tx	Primary tumour cannot be assessed.	
T0	No evidence of primary tumour.	***Superficial Disease***
Ta	Noninvasive papillary carcinoma.	
Tis	Carcinoma in situ: "flat tumour."	
T1	Tumour invades subepithelial connective tissue.	
T2	Tumour invades muscularis propria.	***Muscle-Invasive Disease (Non-metastatic)***
T2a	Tumour invades superficial muscularis propria (inner half).	
T2b	Tumour invades deep muscularis propria (outer half).	
T3	Tumour invades perivesical tissue.	
T3a	Microscopically.	
T3b	Macroscopically (extravesical mass).	
T4	Tumour invades any of the following: prostatic stroma, seminal vesicles, uterus, vagina, pelvic wall, abdominal wall.	
T4a	Tumour invades prostatic stroma, uterus, vagina.	
T4b	Tumour invades pelvic wall, abdominal wall.	
N - Lymph Nodes		
NX	Lymph nodes cannot be assessed.	
N0	No lymph node metastasis.	
N1	Single regional lymph node metastasis in the true pelvis (hypogastric, obturator, external iliac or presacral lymph node).	***Advanced Disease (Metastatic)***
N2	Multiple regional lymph node metastases in the true pelvis (hypogastric, obturator, external iliac or presacral lymph node).	
N3	Lymph node metastases to the common iliac lymph nodes.	
M - Distant Metastases		
M1	Distant metastasis.	
M0	No distant metastasis.	

5.1.4. Treatment of superficial bladder cancer

5.1.4.1. Transurethral Resection of Bladder Cancer (TURBT)

TURBT is the primary method for diagnosis and treatment of all superficial bladder tumours (British Association of Urological Surgeons and British Uro-oncology Group 2007; Babjuk *et al.* 2011). Under direct visualisation by flexible cystoscopy, the bladder tumour is excised down to the detrusor muscle layer of the bladder wall with a surrounding resection margin. Complete resection with detrusor muscle seen in the resection sample is associated with reduced recurrence rates (Mariappan *et al.* 2010).

Table 3: Non-muscle invasive bladder cancer recurrence risk factor weighting, and the probability of recurrence by risk group and recurrence score (adapted from (Sylvester *et al.* 2006; Babjuk *et al.* 2011)).

Risk Factor	Recurrence weighting score
Number of tumours	
Single	0
2-7	3
> 8	6
Tumour diameter	
< 3 cm	0
> 3 cm	3
Prior recurrence rate	
Primary	0
< 1 recurrence/year	2
> 1 recurrence/year	4
Category	
Ta	0
T1	1
Concurrent CIS	
No	0
Yes	1
Grade (WHO 1973)	
G1	0
G2	1
G3	2
Total score	**0-17**

Risk group	Recurrence score	Probability of recurrence at 1 year	
		%	*(95% CI)*
Low risk	0	15	(10-19)
Intermediate risk	1-4	24	(21-26)
	5-9	38	(35-41)
High risk	10-17	61	(55-67)

5.1.4.2. Intra-vesical Mitomycin C chemotherapy

Over 24% of intermediate-risk superficial tumours recur following TURBT (Table 3) (Millan-Rodriguez *et al.* 2000; Babjuk *et al.* 2011). In this group of patients, adjuvant intravesical chemotherapy with a single instillation of Mitomycin C is recommended with a 24% relative risk reduction in recurrence rates (British Association of Urological Surgeons and British Uro-oncology Group 2007; Babjuk *et al.* 2011).

5.1.4.3. Local Bacillus Calmette-Guerin (BCG) injection

BCG immunotherapy is recommended for high risk superficial tumours (Table 3) as over 60% of cases recur and are at high risk of progression to muscle-invasive disease (British Association of Urological Surgeons and British Uro-oncology Group 2007; Babjuk *et al.* 2011). Maintenance BCG treatment with 27 doses over 3 years is associated with improved recurrence-free and overall survival (Lamm *et al.* 2000).

5.1.5. Treatment of muscle invasive bladder cancer (MIBC)

Curative treatments for MIBC in the United Kingdom consist of either radical cystectomy or radical RT with / without neoadjuvant chemotherapy (Figure 3) (British Association of Urological Surgeons and British Uro-oncology Group 2007).

5.1.5.1. Neoadjuvant, concurrent and adjuvant chemotherapy

Neoadjuvant cisplatin-based chemotherapy prior to curative surgery or radiotherapy to downstage disease and eradicate micrometastases is recommended for all non-metastatic MIBC. This is based on a Cochrane review meta-analysis of 11 trials finding a significant benefit in disease-specific and overall survival at 5 years (absolute improvement of 9% and 5% respectively) with platinum-based combination chemotherapy (Vale and Advanced Bladder Cancer Meta-analysis Collaboration 2004; Vale and ABC Meta-analysis Group 2005).

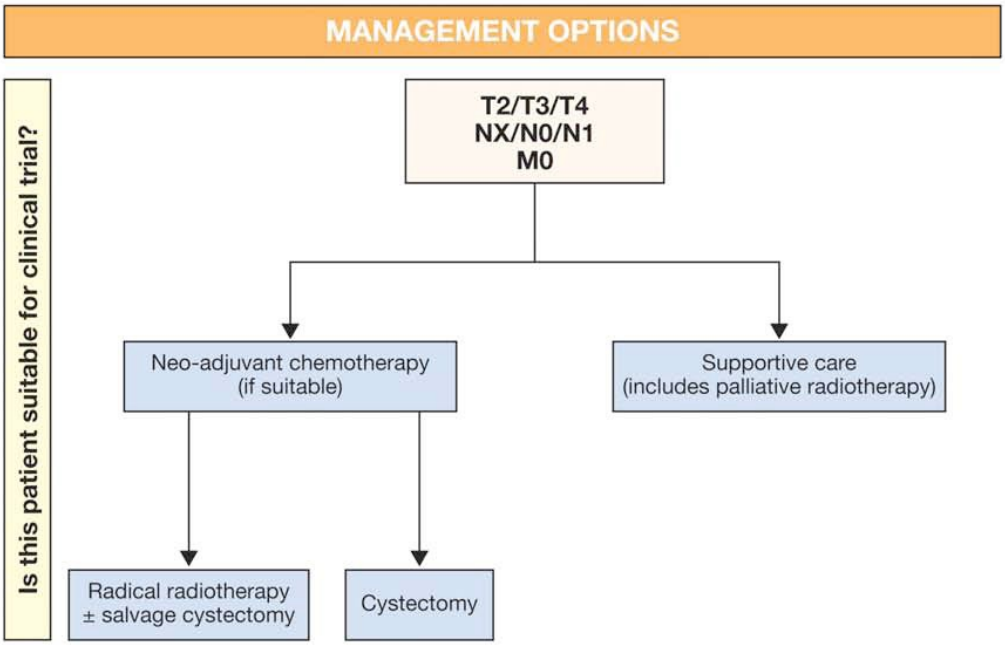


Figure 3: Treatment algorithm for non-metastatic MIBC (British Association of Urological Surgeons and British Uro-oncology Group 2007)

The optimum regimen of concurrent chemo-radiotherapy remains unclear. Until recently, there has been only one published small randomised trial (99 patients only) comparing radiotherapy with or without concurrent cisplatin chemotherapy, which found a significant improvement in control of pelvic disease but not survival although its power was very limited (Coppin *et al.* 1996). In a single centre study looking at four consecutive patient cohorts with gradual introduction of different platinum-based chemotherapy regimens between 1982 and 2006, the authors found improvement in 5-year overall survival in the concurrent chemotherapy cohorts though this could be confounded by advancements in patient care and RT techniques (Ott *et al.* 2009). A recent Phase II trial using concurrent gemcitabine with hypofractionated RT has generated promising results with 3-year overall survival rates of 75% compared to 46% in previously reported RT-only results (Hoskin *et al.* 2010; Choudhury *et al.* 2011). The recently published BC2001 phase III randomised control trial (RCT) comparing radical radiotherapy versus radiotherapy with synchronous Mitomycin C and 5-fluorouracil has reported a significant improvement in loco-regional control,

with no increase in treatment morbidity, but no significant difference in cancer-specific survival or overall survival although this may have been an issue of statistical power with subjective separation of the survival curves (James *et al.* 2012).

The BCON phase III RCT (Hoskin *et al.* 2010) explored the addition of concurrent carbogen (a normobaric 98% oxygen/ 2% carbon dioxide gas mixture), and nicotinamide as radiosensitisers to RT treatment. These radiosensitisers were to overcome chronic diffusion-limited tumour hypoxia by enhancing oxygen delivery, and to stabilise oxygen delivery by reducing intermittent tumour blood flow closure thus overcoming acute perfusion-limited tumour hypoxia respectively (Chaplin *et al.* 1991). In this trial, use of concurrent carbogen and nicotinamide with RT resulted in an absolute 11% improvement in 3-year relapse-free survival and 13% improvement in 3-year overall survival with no increase in radiotherapy toxicity.

The role of adjuvant chemotherapy in advanced localised disease (T3 and T4) remains unclear due to small study sizes and failure of recruitment in trials (Calabro and Sternberg 2009). A meta-analysis of six trials using cisplatin-based chemotherapy found a significant 25% relative risk reduction of death with chemotherapy (P=0.02). However, this meta-analysis was limited by small numbers, poor treatment arm compliance, and trials closing early (Vale and ABC Meta-analysis Group 2005).

5.1.5.2. Radical radiotherapy

Radical RT is a curative treatment modality for non-metastatic MIBC allowing the preservation of the bladder thus maintenance of urinary continence in most cases. In the UK, besides patient choice, factors favouring radiotherapy treatment are related to fitness for surgical treatment, successful technical delivery and safety of radiotherapy, concurrent co-morbidities, bladder function and the presence of ureteric

obstruction and hydronephrosis (British Association of Urological Surgeons and British Uro-oncology Group 2007).

Maximal TURBT is preferred for maximum tumour debulking prior to RT (Efstathiou *et al.* 2012; Hindson *et al.* 2012). A planning CT scan is performed with the patient supine. The clinical target volume (CTV) is delineated encompassing the entire bladder and a 0.5 cm margin around the tumour bearing wall and regions clinically or radiologically suspect for extra-vesicular extension. A further 1.5 to 2 cm margin is added to the CTV to generate the planning target volume (PTV) (James and Zarkar 2006). The PTV receives at least 95% of the total prescribed radiotherapy dose of 64 Gy in 32 fractions using conventional fractionation or a hypofractionated schedule of 52.5 - 55Gy in 20 fractions (James and Zarkar 2006; Hindson *et al.* 2012).

Acute radiotherapy toxicity, which occurs during treatment and for several weeks afterwards, includes lethargy, radiation dermatitis, diarrhoea, tenesmus, proctitis and cystitis (British Association of Urological Surgeons and British Uro-oncology Group 2007), with the majority of patients developing bowel and/or urinary symptoms, although reported severe toxicity rates (Radiation Therapy Oncology Group (RTOG)/ European Organization for Research and Treatment of Cancer (EORTC) radiation toxicity score ≥ 3) are low (3% and 5% respectively) (Sengelov and von der Maase 1999; Majewski and Tarnawski 2009). Late radiotherapy side-effects, which develop three or more months post-RT treatment, include bladder fibrosis and shrinkage, haematuria from bladder telangiectasia, impotence, radiation proctitis, and, more uncommonly, ileus or fistula formation and, rarely, secondary malignancies. Of the 60% of male patients with sexual dysfunction post-RT, 75% of these reported that this was secondary to other factors and not just RT *per se* (Fokdal *et al.* 2004). Five-year risk of severe late radiation bladder or bowel toxicity is between 3-8% (Sengelov and von der Maase 1999; Majewski and Tarnawski 2009).

5.1.5.3. Radical and salvage cystectomy

Radical cystectomy is the curative surgical option for MIBC; it involves the excision of the whole bladder, with any macroscopic tumour extension beyond the bladder wall, the adjacent distal ureters and pelvic lymph nodes (Blandy 1998; British Association of Urological Surgeons and British Uro-oncology Group 2007; Stenzl *et al.* 2011). In men, the prostate gland is also resected, with over 65% of patients found to have an unsuspected prostate adenocarcinoma or urothelial cancer (Stenzl *et al.* 2011). Salvage cystectomy involves the same surgical technique but is performed in the setting of local treatment failure or relapse following radical radiotherapy. Post-cystectomy, there are three main options for urinary diversion: 1) via an abdominal stoma by direct diversion of the ureters, through an ileal conduit or a continent abdominal pouch, 2) by drainage into the large bowel via a rectosigmoid diversion, and, 3) via the urethra by construction of an orthotopic neobladder using a section of bowel (Blandy 1998; Stenzl *et al.* 2011).

Common early surgical complications are prolonged ileus after surgery, anastamotic leakage or suture breakdown of urinary diversions and reconstructed bladders (with associated risk of peritonitis), thromboembolic events, and respiratory infections (Blandy 1998). About 50% of cystectomy patients develop a surgical complication, with the rate being even higher in salvage cystectomy cases (Stenzl *et al.* 2011; Eswara *et al.* 2012). Long-term morbidity is mainly related to complications of the urinary diversion and includes urinary tract infections and incontinence, but also includes male impotence and erectile dysfunction (Stenzl *et al.* 2011).

5.1.5.4. Radical cystectomy versus radical radiotherapy

There are no definitive international guidelines on the curative treatment modality of choice for MIBC with guidelines and practices varying between countries and centres (British Association of Urological Surgeons and British Uro-oncology Group 2007;

Stenzl *et al.* 2011; Efstathiou *et al.* 2012; Hindson *et al.* 2012). The gold standard in the United States of America (USA) and Europe remains radical cystectomy while in the UK and a few centres in Europe and at the Massachusetts General Hospital, the preferred treatment option is radical radiotherapy (RT) with or without concurrent chemotherapy (James and Hussain 2005; Rodel *et al.* 2005; Efstathiou *et al.* 2012). There has been only one randomised Phase III clinical trial comparing radical cystectomy versus radical RT but this trial was closed early due to poor accrual, partly due to those patients declining randomisation preferring bladder preservation, and hence RT (Huddart *et al.* 2010). A Cochrane review covering only three trials (all prior to 1995, with two trials before 1980) comparing pre-operative RT and radical cystectomy versus radical RT and salvage cystectomy found a small survival benefit with surgery (Shelley *et al.* 2008). However, all three trials were relatively small, and both surgical and RT techniques have advanced significantly since then, with orthotopic bladder reconstruction in surgery and, for RT, CT planning, improved dose delivery, conformal and image-guided adaptive RT (Pos and Remeijer 2010; Lalondrelle *et al.* 2011; Thariat *et al.* 2012). In an audit in Leeds looking at MIBC patients treated by cystectomy or RT between 1995 and 2000, there was no significant difference seen in 5-year disease specific survival (53.4% versus 56.8% respectively) (Kotwal *et al.* 2008). A recent study of the Surveillance, Epidemiology, and End Results database of MIBC patients between 1988 and 2006 in the USA also found no significant difference in overall survival between radical cystectomy or RT after adjusting for confounding factors (Kozak *et al.* 2012).

5.2. Genetics

Understanding the genetics of inherited bladder cancer risk is important to improve knowledge of the underlying mechanisms of bladder carcinogenesis which in turn could lead to the discovery of new therapeutic targets. Knowledge of the genetic determinants of bladder cancer inherited risk could also potentially allow the prospect of estimating individual cancer risk, thus providing targeted lifestyle advice for primary prevention or for targeted screening.

5.2.1. Common Disease - Common Variant Hypothesis

Single nucleotide polymorphisms (SNP) are common single base DNA variations that have a minor allele frequency (MAF) of more than one percent. The "common disease - common variant" (CDCV) hypothesis proposes that a few common SNPs sited at a few specific disease loci are predominantly responsible for the genetically attributable risk of common diseases such as cancer (Reich and Lander 2001).

Previously, the acetylator activity status of N-acetyltransferase, a gene involved in detoxifying aromatic hydrocarbons such as those from tobacco smoke, was demonstrated to be associated with bladder cancer risk (Cartwright *et al.* 1982), with acetylator status being known to be related to SNPs in this gene (Evans and White 1964). Candidate SNP studies by the Spanish Bladder Cancer Study proceeded to demonstrate that SNPs resulting in deletion of the glutathione-S-transferase M1 (*GSTM1*) gene (also involved in detoxifying aromatic hydrocarbons) and slow N-acetyltransferase acetylator (*NAT2)* status resulted in an overall increased risk (OR 4.21 [95%CI 1.26-14.14] and OR 4.76 [95%CI 1.25-18.09]) of bladder cancer (Garcia-Closas *et al.* 2005).

Genome wide association studies (GWAS) explore the CDCV hypothesis further by unbiased high-throughput genotyping of hundreds of thousands of SNPs spread throughout the genome in large case-control studies. Geographically close SNPs tend to be in “linkage”, that is they tend to be inherited together, and “tag” SNPs are SNPs which provide representative information of the presence of surrounding linked SNPs. Thus by targeting tag SNPs, the GWAS approach is able to detect the effects of SNPs not genotyped and, with sufficiently dense SNP spacing, is able to provide information on the majority of SNPs in the whole genome (Iles 2008). Multiple cancer GWAS have been conducted with many positive cancer disease-susceptibility loci discovered (Figure 4) (Cooper and Shendure 2011). However, not all associations were replicated (Lohmueller *et al.* 2003) and the majority of successfully replicated SNPs had low penetrance (ie. only a small proportion of carriers go on to develop cancer) with small effect sizes (mean OR between 1.2 and 1.5) with only a few having ORs greater than two (Bodmer and Bonilla 2008; Iles 2008; Schork *et al.* 2009; Carvajal-Carmona 2010).

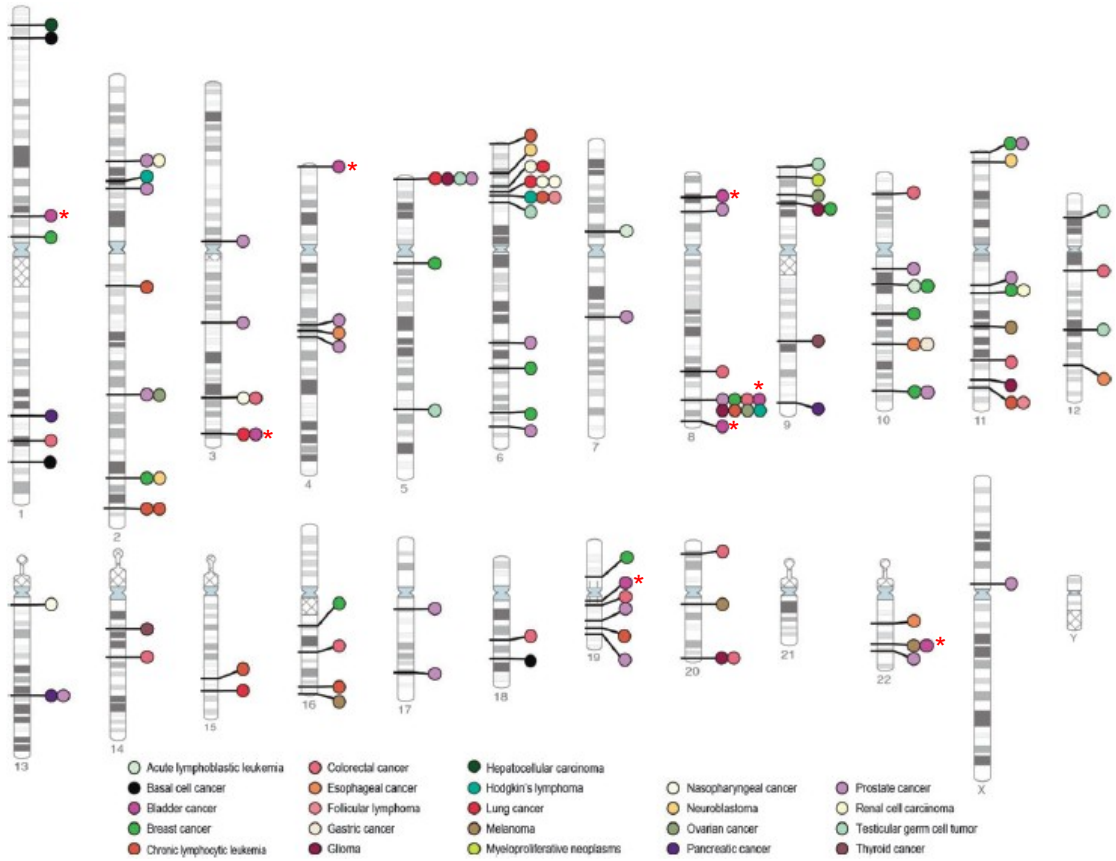


Figure 4: Cancer susceptibility loci hits identified by respective cancer site GWAS studies up to 2010 (red asterisk for GWAS bladder cancer loci) (Hindorff *et al.* 2011).

5.2.1.1. Genome wide association studies in bladder cancer

The three initial bladder cancer GWAS completed in Europe and the USA identified eight disease loci, with replication of disease loci on chromosomes 3q28 (closest gene *TP63)*, 4p16.3 (*FGFR3*), 8q24.21 (*MYC)* and 8q24.3 (*PSCA*) in all three studies, with allelic ORs of about 1.2 (Kiemeney *et al.* 2008; Wu *et al.* 2009; Rothman *et al.* 2010; Fu *et al.* 2012). The roles of these genes in bladder carcinogenesis is biologically feasible: *FGFR3* mutations are commonly seen in superficial tumours as mentioned earlier (Goebell and Knowles 2010), PSCA is overexpressed with increasing grade of bladder tumour, with PSCA mRNA expression levels being predictive of superficial tumour recurrence, the transcription factor MYC is involved in cell proliferation and is associated with other malignancies, and the tumour suppressor TP63 is involved in cell cycle arrest and apoptosis (Kiltie 2010). Wu *et al* demonstrated that the different alleles of the GWAS SNP hit *PSCA* rs2294008

resulted in alterations in the transcriptional promoter activity of PSCA, while resequencing the *PSCA* gene identifed seven common *PSCA* SNPs all in strong linkage disequilibrium (LD) with rs2294008 and having similar effects on bladder cancer risk (Wu *et al.* 2009).

Pathway based analysis of the National Cancer Institute (NCI) GWAS identified signals in three main cellular processes – vesicle biogenesis and budding involved in intra- and extra-cellular signalling, mitosis, and metabolic detoxification including aromatic amine metabolism (Menashe *et al.* 2012). As mentioned earlier, aromatic amine metabolism is implicated in bladder cancer risk from candidate gene SNP studies of *NAT2* and *GSTM1* genes (Garcia-Closas *et al.* 2005). By imputation and targeted Sanger sequencing of the NCI GWAS hit rs11892031 within intron 1 of the *UGT1A* gene (coding for an enzyme involved in aromatic amine detoxification) located on chromosome 2q37.1, an uncommon synonymous coding *UGT1A* SNP, rs17863783 (MAF 2.5%), was identified. This SNP explained the genetic association seen with rs11892031, and was found to be protective for bladder cancer risk (allelic OR 0.55) and associated with increased mRNA expression of the *UGT1A* functional splicing isoform (Tang *et al.* 2012). Further imputation of newly discovered SNPs from the 1000 Genomes project (The 1000 Genomes Project Consortium 2010) in the European GWAS has further identified an additional locus in *SLC14A*, a urea transporter gene involved in urine production (Rafnar *et al.* 2011).

However despite these efforts in bladder cancer and in other diseases, only about five percent of inheritable risk in complex diseases has been explained by GWAS and the CDCV hypothesis (Schork *et al.* 2009; Bodmer and Tomlinson 2010; Cirulli and Goldstein 2010; Gorlov *et al.* 2011). An alternative hypothesis was thus needed to explain this “missing heritability” (Manolio *et al.* 2009).

5.2.2. Common Disease - Rare Variant Hypothesis

The "common disease-multiple rare variants" (CDRV) hypothesis proposes that a significant proportion of inherited cancer susceptibility is due to the summation of several different rare gene variants each having a significant effect on disease risk (Bodmer and Bonilla 2008). Rare variants are base variations with a MAF between 0.1 and 1% (Table 4) (Cirulli and Goldstein 2010) but they make up to 50% of variants in the dbSNP database (http://www.ncbi.nlm.nih.gov/snp) (Gorlov *et al.* 2011).

Table 4: Comparison of common versus rare disease variant characteristics (adapted from (Bodmer and Bonilla 2008).

Common disease variants	Rare disease variants
Discovery by population association, case-control studies, using genome-wide markers (WGA)	Discovery by DNA resequencing of candidate genes, preferably in early onset cases with one or more relatives affected
Mostly MAF > 5%	MAF > 0.1% to 1% Higher than rare familial mutations, lower than polymorphisms. Often population specific
Explained by LD with functional variant	Not detected by WGA
OR mostly between 1.2 and 1.5	OR mostly ≥ 2
Need large studies with control for ethnic heterogeneity to achieve statistical significance and minimise false positives	Assess significance by increased frequencies in cases vs. controls and by functional analysis of variant
Make substantial contribution to population attributable risk	Summation of effects of several variants make significant contribution to population attributable risk
Low penetrance makes prophylactic intervention unlikely	Penetrance often high enough to justify prophylactic interventions
Hard to find functionally relevant variant	Variants identified are functionally relevant
Contribution to disease aetiology questionable	Make a contribution to understanding disease aetiology
May suggest candidates for rare variant search	Effect may be modified by common variants

They represent recent germline genetic mutations arising within the last 100 generations (Figure 5). Thus, the rapid population growth over the last 2000 years has resulted in the abundance of rare variants detected with many variants being population-specific (Coventry *et al.* 2010; Ju *et al.* 2011). Due to genetic selection

having insufficient time to influence these recent variations, rare variants are predicted to have a larger effect on disease risk and to be more likely to have a functional effect than common SNPs (Coventry *et al.* 2010; Gorlov *et al.* 2011).

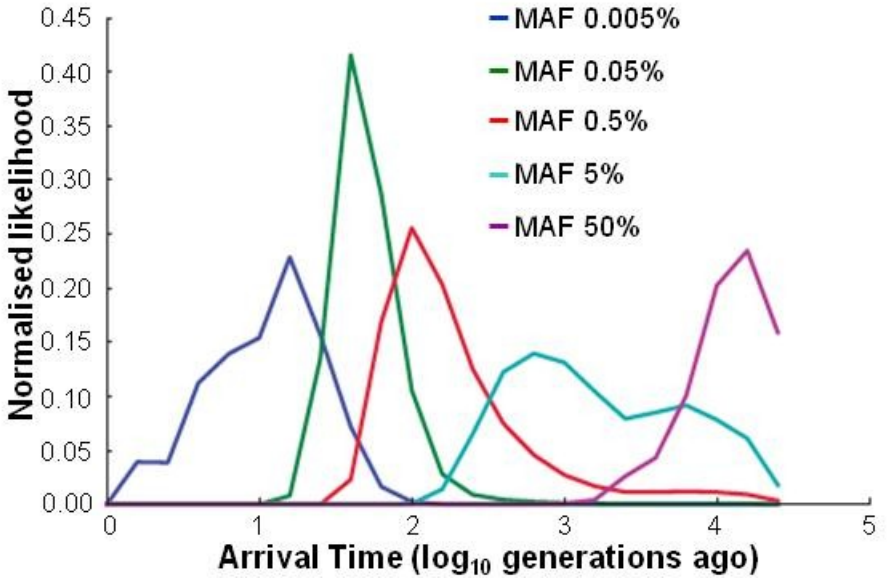


Figure 5: Estimated distribution of variants by current population MAFs relative to when the variant first arose in log 10 number of generations ago (adapted from (Coventry *et al.* 2010).

Rare variant alleles are predicted to have dominant effects due to loss of heterozygosity or gain of function, with each rare variant allele at least doubling disease risk (Gibson 2011). For a disease with low population incidence, it has been predicted that many individuals may carry a few different disease risk rare variants without developing disease (Figure 6). It is also proposed that a potentially large proportion of GWAS common SNP hits may be secondary to “synthetic associations” with causal rare variants within several megabases (Mb) of the GWAS significant SNP (Dickson *et al.* 2010).

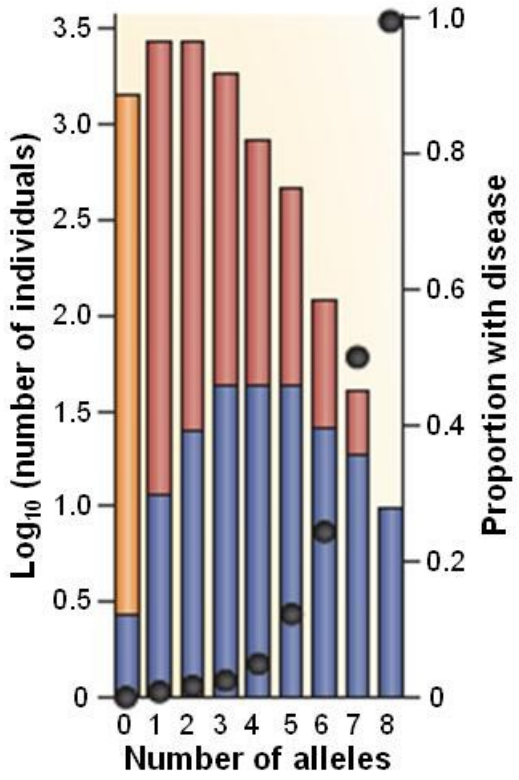


Figure 6: Predicted distribution of number of rare variant risk alleles carried in a case–control population. Distribution of cases (blue) and controls (red) assume a rare variant allele frequency of one percent, a background population disease risk of 0.2% and 100 rare variant disease loci each conferring a genotype relative risk of 2.2 (Gibson 2011).

5.2.2.1. Rare variants in common diseases

Highly penetrant rare germline variants are known to be associated with several rare familial disorders such as maturity onset diabetes of the young (Weedon and Frayling 2007), familial BRCA1/ BRCA2 related breast cancers (Easton *et al.* 2007) and MUTYH-associated polyposis (Al-Tassan *et al.* 2002). Fearnhead *et al* first demonstrated the role of rare variants in common diseases by identifying 13 rare variants with a combined OR of 2.2 involved in the inherited susceptibility of colorectal adenomas (Fearnhead *et al.* 2004). In hypertriglyceridaemia, targeted sequencing of four GWAS loci found a significantly increased rare variant carrier frequency in cases compared to controls (Johansen *et al.* 2010). In the WECARE population-based breast cancer study, a greater risk of developing a second, contralateral breast cancer was described in carriers of rare *BRCA1* and *BRCA2*

variants of unknown significance as well as an increased risk of breast cancer in their first degree relatives (Capanu *et al.* 2011). In colorectal cancer genetic susceptibility, carriers of rare variants in the cell cycle regulation gene, *Cyclin D1*, had an increased risk of developing multiple adenomas and colorectal cancer while *MUTYH* rare variants have been associated with early onset colorectal cancer with all identified rare variants in these two studies reported having per-allele odds ratios greater than two (Farrington *et al.* 2005; Bonilla *et al.* 2011).

GWAS are not adequately powered to detect rare variants due to their low allele frequency, high allelic heterogeneity and being population specific (Iles 2008; Schork *et al.* 2009; Bodmer and Tomlinson 2010). Thus, reliable rare variant discovery can only be achieved by direct DNA resequencing of the whole genome, whole exome or candidate genes (Bodmer and Tomlinson 2010; Cirulli and Goldstein 2010). Most strategies proposed for the study of rare variants involve a two-stage approach of sequencing a study sub-set or population enriched for rare variants, followed by confirmation in a large case-control study using standard genotyping techniques (Cirulli and Goldstein 2010; Kim *et al.* 2010; Yang and Thomas 2011).

However, standard conventional Sanger sequencing is too costly and labour intensive to use for extensive DNA resequencing for rare variants but the advent of next-generation sequencing technologies has now made such studies feasible.

5.2.3. DNA sequencing

5.2.3.1. Conventional Sanger sequencing

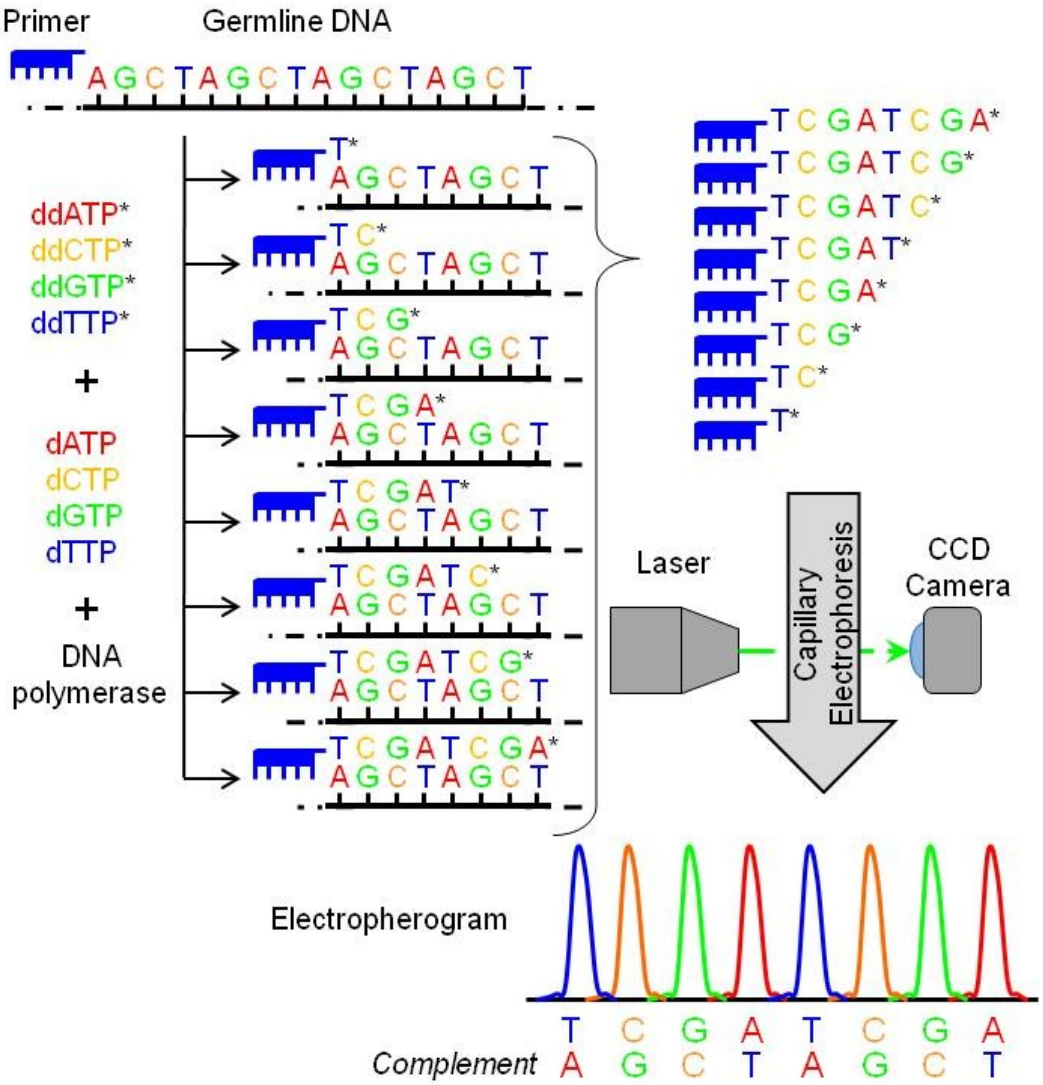


Figure 1: Conventional Sanger sequencing. Using a primer upstream of the DNA template, standard deoxynucleotides (dNTP), DNA polymerase and chain-terminating dideoxynucleotides (ddNTP) labelled with a different fluorescent dye for each nucleotide are added for DNA strand elongation resulting in DNA fragments of varied sizes which, based on size, travel at different speeds in an electric field during capillary electrophoresis. Laser excitation of the fluorescent dye and capture of this fluorescence generates an electropherogram with each fluorescent peak denoting the respective ddNTP with the complement being the sequence of the original DNA template.

Conventional sequencing is performed using capillary-based Sanger sequencing, involving the incorporation of chain-terminating fluorescently-labelled dideoxynucleotides during DNA strand elongation. As each complementary fragment is terminated at different points, capillary electrophoresis separates each fragment by size while laser excitation of the fluorescent label generates an electropherogram for

interpretation of the DNA sequence (Figure 7). Conventional sequencing is capable of sequencing up to about 1 kilobase at a cost of $500 per megabase (Shendure and Ji 2008; Tucker *et al.* 2009).

5.2.3.2. Next-generation sequencing

There are currently three commercially available second generation next-generation sequencing (NGS) platforms: Illumina Genome Analyzer's polymerase-based "sequencing by synthesis" platform, Applied Biosystems SOLiD Sequencer ligation-based platform and Roche GS-FLX 454 Genome Sequencer pyrosequencing-based platform (Table 5). Using these platforms, it is now possible to sequence a whole genome within a week.

Table 5: Comparison between different next-generation sequencing platforms (adapted from (Tucker *et al.* 2009). * Cost quoted is as reported in 2009.

Platform	Amplification	Approach	Read Length (bp)	Run Time and Throughput	Raw Accuracy	Cost (US$/Mb)*
Illumina Genome Analyzer	polymerase-based sequencing by synthesis	bridge PCR	75-150	17 gigabases (Gb) in 7 days	98.50%	6.00
ABI SOLiD Sequencer	ligation-based	emulsion PCR	50	10–15 Gb in 3-7 days	99.94%	5.80
Roche 454 Sequencer	pyrosequencing	emulsion PCR	400	400–600 Mb in 10 hr	99%	84.40

Focusing on the Illumina platform used in this thesis, the DNA sequence of interest is fragmented and ligated with oligonucleotide adaptors to form sequencing libraries. Fragments are hybridised to complementary adaptors coated on the flow cell surface. Bridge amplification of hybridised fragments result in the formation of clusters of each fragment sequence. Fluorescently-labelled nucleotides with reversible chain-terminating inhibitors are added to the 5'-end and the surface imaged to determine the first base. The terminator and label are enzymatically removed and the next base

added and imaged and so on for currently up to a read length of 150 bases (Figure 8).

As the cost for whole genome sequencing of a single sample to a high read depth remains relatively high, sequencing a large cohort would prove excessively expensive. However, by sequencing only a target-enriched region of interest, it is possible to sequence pools of multiple samples, termed multiplexing, while maintaining high read depths and improving cost-effectiveness. Several studies have thus recently been exploring the feasibility and the various strategies of multiplexing with regards to target size, amplification processes and the ideal pooling strategies (Out *et al.* 2009; Harris *et al.* 2010; Kim *et al.* 2010; Harakalova *et al.* 2011; Lee *et al.* 2011).

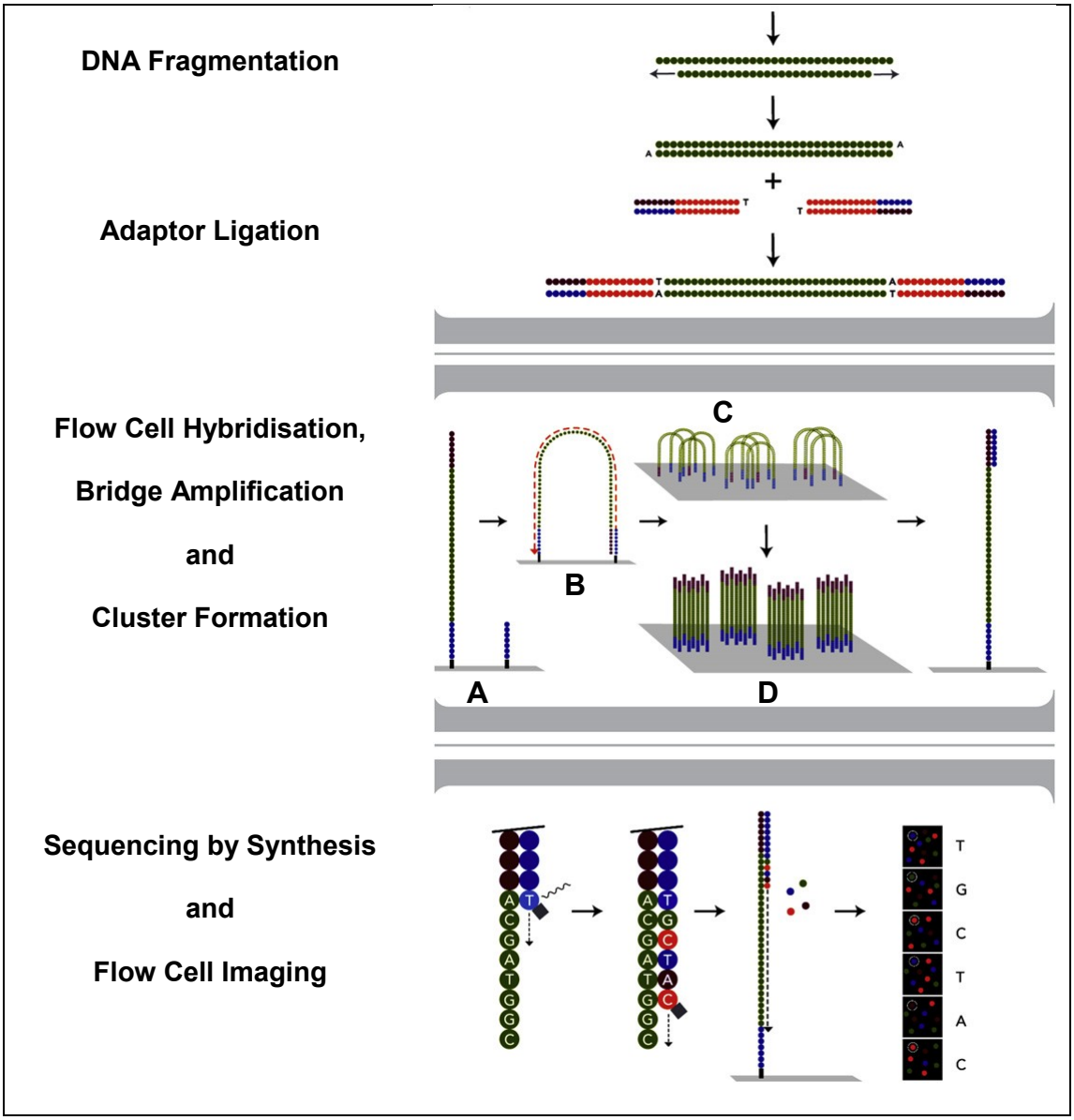


Figure 8: Illumina Genome Analyzer sequencing workflow (Tucker *et al.* 2009). The DNA sequence of interest is randomly fragmented into 200 base pair (bp) fragments followed by ligation to oligonucleotide adaptors at both the 3'- and 5'-ends. Fragments are then hybridised to a flow cell coated with complementary oligonucleotides to the adaptors (A) resulting in the formation of "bridges" between hybridised fragments and complementary oligonucleotides and amplification from the 3' to the 5'-end (B). By repeated bridge amplification of hybridised fragments and newly synthesised fragments (C), discrete clusters of each fragment are formed (D). Sequencing by synthesis is then performed from the 5'-end using fluorescently labelled nucleotides with reversible chain-terminating inhibitors and imaging of each consecutive base with each cycle. The number of cycles is equivalent to the read length.

5.2.4. MicroRNAs

5.2.4.1. Biology and post-transcriptional regulation

Non-coding RNAs (ncRNA), which include microRNAs (miRNAs), small nucleolar RNAs, Piwi protein-interacting RNAs and large intergenic non-coding RNAs, are a recently discovered group of endogenous RNAs involved in epigenetic, ribosomal and post-transcriptional regulation (Esteller 2011). miRNAs are the most extensively studied of these. They are small ncRNAs of approximately 22 to 27 nucleotides, accounting for 1% of all genes, and are believed to regulate more than 60% of coding genes (Brennecke *et al.* 2005; Friedman *et al.* 2009; Esteller 2011). miRNA genes are either independent genes or are located within introns of coding genes known as mirtrons. They are transcribed by RNA polymerase II into primary-miRNA (pri-miRNA) which is processed by the Drosha complex into precursor-miRNA (pre-miRNA) which is then exported out of the nucleus by Exportin-5 (Figure 9) (Krol *et al.* 2010; Esteller 2011; Pasquinelli 2012). The loop region is excised by Dicer and the miRNA duplex loaded onto the catalytic RNA endonuclease protein Argonaute (Ago) to form an RNA-induced silencing complex (RISC).

RISC regulates post-transcriptional gene regulation by pairing with complementary binding sites predominantly in the 3'untranslated region (3'UTR) of messenger RNA (mRNA), with perfect complementation resulting in mRNA degradation while imperfect pairing causes inhibition of mRNA translation (Zeng *et al.* 2003; Liu *et al.* 2008). High complementation of the second to eighth nucleotide from the 5' end of the miRNA known as the miRNA "seed" site, is required for functional miRNA-mRNA pairing (Brennecke *et al.* 2005; Chi *et al.* 2012; Pasquinelli 2012). Multiple bio-informatics algorithms have thus been developed using these "seed pairing rules" for predicting miRNA targets (Lewis *et al.* 2003; John *et al.* 2004; Kiriakidou *et al.* 2004; Krek *et al.* 2005; Rusinov *et al.* 2005; Griffiths-Jones *et al.* 2006). However, functional non-canonical miRNA binding sites have recently been described: "centered pairing"

with highly contiguous pairing between miRNA nucleotides four to fifteen (Shin *et al.* 2010); 3'-end pairing (Lal *et al.* 2009; Pasquinelli 2012); and "pivot-pairing" involving the formation of a transitional nucleation state of complementary seed pairing between nucleotides two to six followed by a bulge in the mRNA to allow downstream base-pairing (Chi *et al.* 2012).

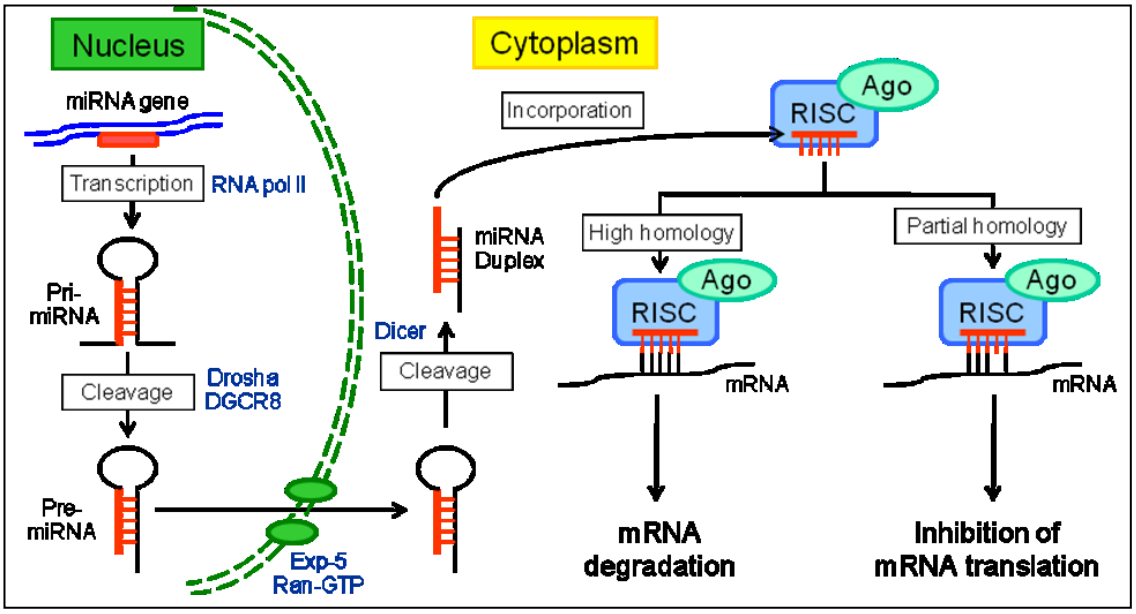


Figure 9: MicroRNA biogenesis and its role in mRNA post-transcriptional regulation (adapted from (Esquela-Kerscher and Slack 2006). miRNA genes are transcribed by RNA polymerase II into pri-miRNAs which in turn are processed by DGCR8 double-stranded RNA binding protein and Drosha RNAase II enzyme into a ~70 nucleotide pre-miRNA stem-loop hairpin structure. The pre-miRNA is exported out of the nucleus by Exportin-5 and Ran-GTP and the loop excised by Dicer to form a miRNA duplex. The mature miRNA is incorporated into the RISC complex with Argonaute and, depending on perfect or imperfect complementation with target mRNA, results in mRNA degradation or inhibition of mRNA translation.

5.2.4.2. MicroRNAs in carcinogenesis and cancer susceptibility

miRNAs are commonly dysregulated in cancers with differential expression between normal tissues, tumours, and across cancer types (Esteller 2011). Fifty percent of miRNA genes are localised to cancer-associated "fragile sites" of the genome with miRNA dysregulation caused by genetic mutation of miRNA or miRNA biogenesis genes, by epigenetic changes such as CpG island hypermethylation of miRNA or mirtron associated gene promoters, or by gene copy number variations (Esquela-Kerscher and Slack 2006; Zhang *et al.* 2006; Krol *et al.* 2010; Esteller 2011). miRNAs

can act both as oncogenic miRNAs (or "oncomirs"), or as tumour suppressors. miRNAs commonly associated with malignancy are miR-15, miR-16-1, miR-17-92, miR-21, miR-125 and miR-145 and the let-7 miRNA family (Esquela-Kerscher and Slack 2006; Hammond 2007; Esteller 2011).

In bladder cancer, genomic miRNA profiling has demonstrated differences in miRNA expression not only between normal urothelium and bladder cancer but differences between low-grade, high grade and invasive tumours (Catto *et al.* 2009; Dyrskjot *et al.* 2009; Catto *et al.* 2011). The differential expression of several miRNAs in low grade tumours is also associated with risk of progression to a high grade phenotype (Catto *et al.* 2009; Dyrskjot *et al.* 2009). Transfection of various miRNAs downregulated in bladder tumours, such as miR-125, miR-129 and miR-145, into bladder cancer cell lines have resulted in cell growth and colony formation inhibition, apoptosis and cancer cell death, highlighting the tumour suppressor roles of these miRNAs (Dyrskjot *et al.* 2009; Ichimi *et al.* 2009; Ostenfeld *et al.* 2010; Huang *et al.* 2011).

SNPs in miRNA and miRNA biogenesis genes are associated with cancer predisposition by affecting miRNA-mRNA binding, miRNA maturation or RISC stability (Horikawa *et al.* 2008; Krol *et al.* 2010; Liu *et al.* 2010; Yang *et al.* 2010). In bladder cancer, homozygote carriers of the miRNA processing gene *GEMIN3* SNP rs197414 rare allele had a significantly increased risk of developing bladder cancer (Yang *et al.* 2008). SNPs within miRNA-target binding sites may also alter miRNA-mRNA binding and Yu *et al* reported a negative selection against SNPs within these predicted 3'UTR miRNA "seed" sites compared to other SNPs within the 3'UTR, potentially due to their deleterious effects (Yu *et al.* 2007). Bioinformatics predictions for these 3'UTR SNPs affecting miRNA-mRNA binding have since been successful in

identifying breast cancer and colorectal cancer susceptibility alleles (Landi *et al.* 2008; Nicoloso *et al.* 2010).

5.3. DNA repair and radiation biology

Maintaining cellular genetic integrity is crucial in preventing somatic mutations and carcinogenesis. Cellular DNA is constantly assailed by both intrinsic agents, such as free radicals generated from metabolic processes, and extrinsic agents such as chemical carcinogens, carcinogens derived from tobacco smoking and ionising radiation. Following DNA damage, DNA repair, cell cycle arrest and apoptotic pathways are activated rapidly with the purpose of repairing these DNA lesions or if not possible, inducing cell death. These pathways are thus important in determining carcinogenesis in normal tissues and cancer cell death or survival following radiotherapy treatment.

5.3.1. DNA Repair Pathways

There are five main DNA repair pathways: mismatch repair (MMR), base excision repair (BER), nucleotide excision repair (NER), and the two double strand break (DSB) repair pathways – homologous recombination and non-homologous end joining.

5.3.1.1. Mismatch Repair

Errors in DNA replication due to "DNA polymerase slippage" can result in mismatched nucleotides or DNA insertion-deletion loops in regions of repetitive DNA (Bristow and Harrington 2005; Hall and Giaccia 2006; Jalal *et al.* 2011). These errors are recognised, excised and repaired by MMR. Mutations in the MMR gene families, *MSH*, *MLH* and *PSM*, result in insertions or deletions of short DNA repeats termed microsatellite instability and is associated classically with hereditary non-polyposis colorectal cancer but also seen in a variety of other malignancies. MMR deficiency is also predictive of sensitivity to topoisomerases and resistance to cisplatin and

fluorouracil chemotherapy (Ribic *et al.* 2003; Bristow and Harrington 2005; Bertagnolli *et al.* 2009).

5.3.1.2. Base excision repair

BER is required to repair DNA base lesions, including methylated bases from industrial carcinogens or chemotherapeutics, and oxidative base damage from reactive oxygen species, such as that from smoking-related carcinogens and from ionising radiation, and also DNA SSBs (Hegde *et al.* 2008). Different DNA glycosylases recognise and excise specific base lesions with a level of cross-activity between glycosylases and target base lesions (David *et al.* 2007). The most common oxidative base lesion is 7,8-dihydro-8-oxyguanine (8-OxoG) which is particularly deleterious, as it is able to mimic thymine (T) in base pairing with adenine (A) during DNA replication, resulting in a guanine (G) to T transversions if unrepaired (Figure 10). OGG1 detects and excises 8-oxoG resulting in an apurinic site, which in turn is excised by APE1 leading to repair by either the short-patch or long-patch BER. Short-patch BER involves the removal of only the damaged nucleotide while long-patch BER involves the excision of 2 to 13 (Evans *et al.* 2000). Both pathways generate an SSB and results in poly-ADP-ribosylation of the SSB by PARP1 and recruitment of the downstream repair proteins DNA polymerase β, LIG3 and XRCC1 (Bristow and Harrington 2005; Hall and Giaccia 2006; David *et al.* 2007; Hegde *et al.* 2008; Jalal *et al.* 2011).

MUTYH, like OGG1, screens DNA for 8-oxoG but then locates and excises the mispaired A instead, thus restoring the 8-oxoG:C substrate for OGG1 repair and preventing the G to T transversion (Figure 10). It is a unique glycosylase with no "backup" glycosylase recognising the same lesion (David *et al.* 2007; Kundu *et al.* 2009). In MUTYH-associated polyposis (MAP), carriage of *MUTYH* rare missense variants result in increased somatic G to T transversions in the *APC* and *KRAS*

genes and accounts for approximately 1% of all colorectal cancers (Al-Tassan *et al.* 2002; Lipton *et al.* 2003; Fleischmann *et al.* 2004). MAP patients have also been noted to have an increased risk in extra-colonic tumours (including bladder cancer) (Vogt *et al.* 2009). Colorectal cancer epidemiology studies of two rare *MUTYH* variants Y179C (rs34612342) and G396D (rs36053993), have found a 100% penetrance of developing colorectal cancer by the age of 65 in individuals with biallelic *MUTYH* defects, as well as an increased risk in heterozygous carriers (Farrington *et al.* 2005; Theodoratou *et al.* 2010). Increased rates of breast, gastric and endometrial cancers have also been observed in heterozygote carriers of *MUTYH* variants (Win *et al.* 2011; Rennert *et al.* 2012).

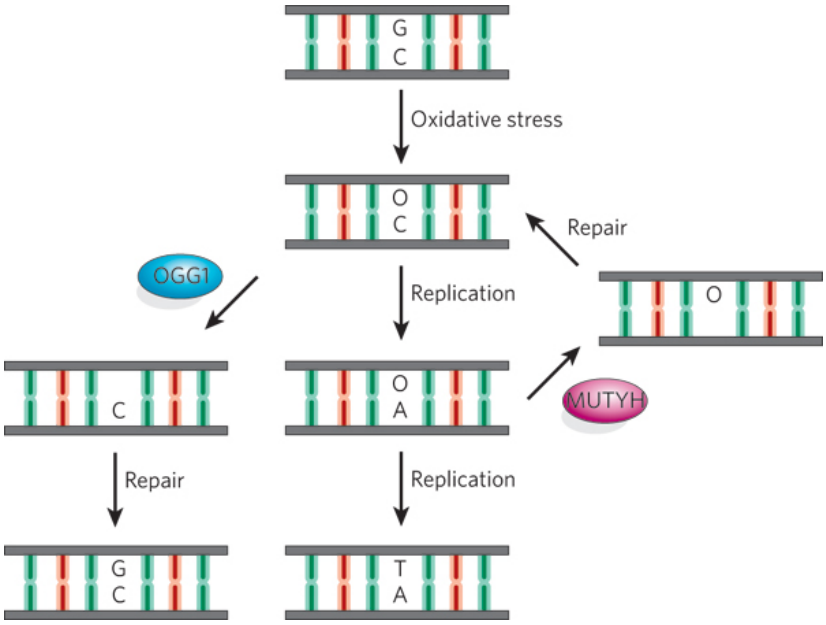


Figure 10: Base excision repair of 8-oxoG by MUTYH and OGG1 (David *et al.* 2007). Oxidative stress results in the formation of 8-oxoG which is detected and excised by OGG1 and repaired (left pathway). If unrepaired prior to replication, 8-oxoG mispairs with A and results in a G:C to T:A mutation (middle pathway). MUTYH detects 8-oxoG:A mispairing and excises A (right pathway), restoring the 8-oxoG:C substrate for OGG1 repair.

5.3.1.3. Nucleotide excision repair

NER targets DNA helix-distorting lesions such as ultraviolet-induced photolesions and bulky DNA adducts from a variety of compounds including aromatic carcinogens

from industrial chemicals and cigarette smoke, and chemotherapy agents. It consists of two sub-pathways: global genome repair (GGR) and transcription-coupled repair (TCR) (Figure 11). Following lesion recognition, GGR and TCR activate a common pathway with unwinding of the DNA helix, excision of the lesion with 24 to 34 surrounding nucleotides, gap filling by DNA polymerase and finally ligation of the re-synthesised DNA (Sugasawa 2008; Bergink *et al.* 2012).

XPC is crucial in GGR for the surveillance of the whole genome for DNA distorting lesions by searching for non-hydrogen bonded nucleotides within the genome, binding to them and recruiting downstream repair factors (Camenisch *et al.* 2009). XPC exists as a heterotrimer complex with HR23B and Centrin 2, which are essential for stabilising XPC by regulating ubiquitinylation and enhancing stable XPC binding to DNA (Sugasawa 2008; Sugasawa 2010; Bergink *et al.* 2012). There is also evidence that XPC modulates DSB repair (Despras *et al.* 2007; Zhang *et al.* 2009), acts as a co-factor for OGG1 in BER (D'Errico *et al.* 2006), regulates gene transcription (Le May *et al.* 2010), and activates a p-53 independent apoptosis pathway (Wang *et al.* 2012).

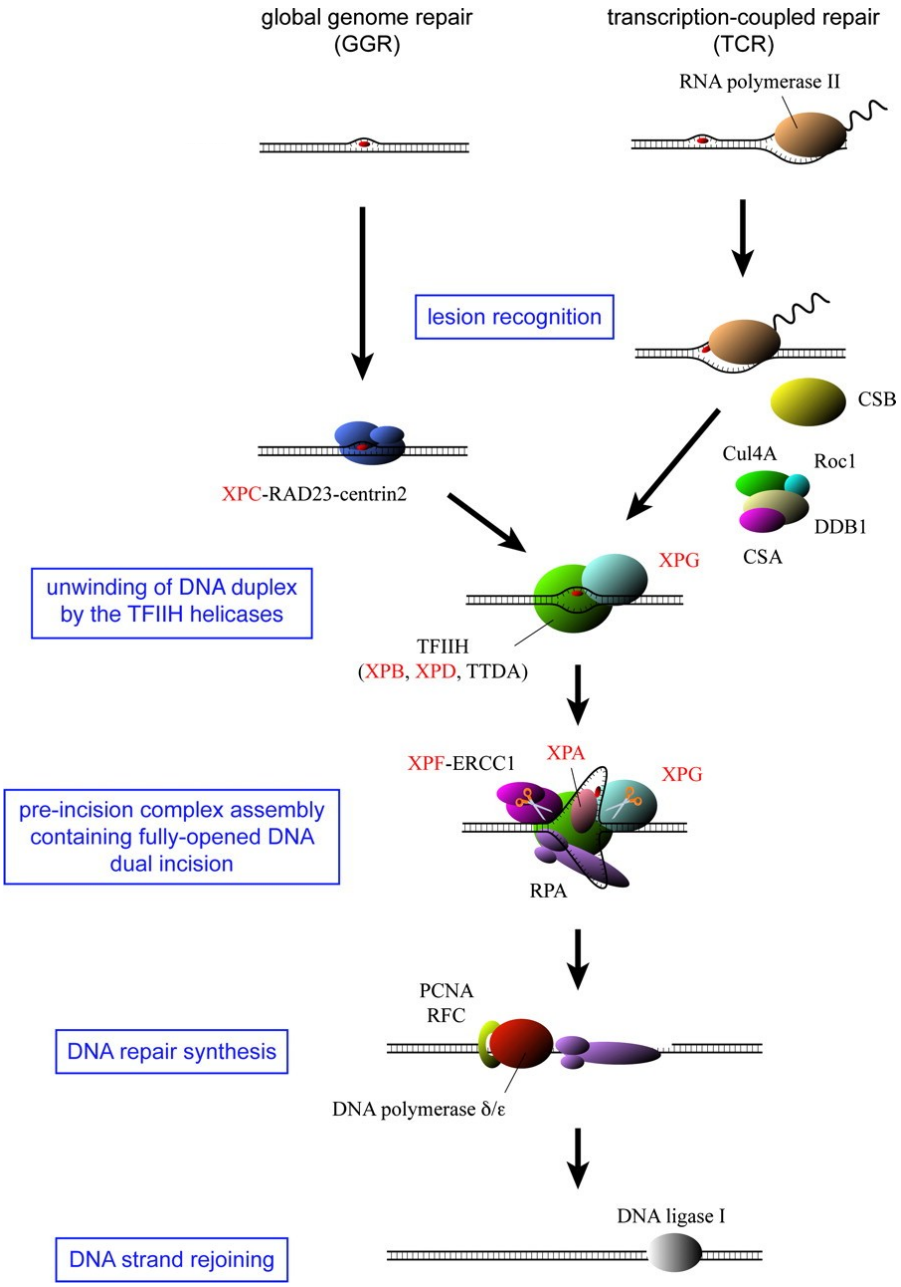


Figure 11: Nucleotide excision repair pathway (adapted from (Sugasawa 2008). NER is initiated by the detection of DNA helix-distorting lesions in global genomic repair by the XPC-HR23B(RAD23)-Centrin2 or in the case of ultraviolet-induced lesions or in transcription coupled repair by the detection of a translocation blockage during transcription by RNA polymerase II and recruitment of the Cockayne syndrome proteins, CSA and CSB. The TFIIH complex is recruited by both XPC and CSA/ CSB resulting in initiation of a common repair pathway with DNA helix unwinding and excision of the damaged DNA at two incision sites by the incision complex generating a 24 to 34 base gap. The gap is filled by DNA synthesis using the undamaged DNA strand as a template and finally rejoined by DNA ligase I.

5.3.1.4. DSB repair

Depending on the cell cycle phase, either of two DSB repair pathways are activated: homologous recombination repair (HR) in cells in S/G_2 phase, or the classical or alternative non-homologous end joining (NHEJ) pathways in cells in the G_0/G_1 phase (Figure 12) (Williams *et al.* 2007; Iijima *et al.* 2008). Phosphorylation of CtIP by ATM and CDK has been reported to influence the cell-cycle dependent choice of pathway activated (Iijima *et al.* 2008; Yun and Hiom 2009; Kass and Jasin 2010).

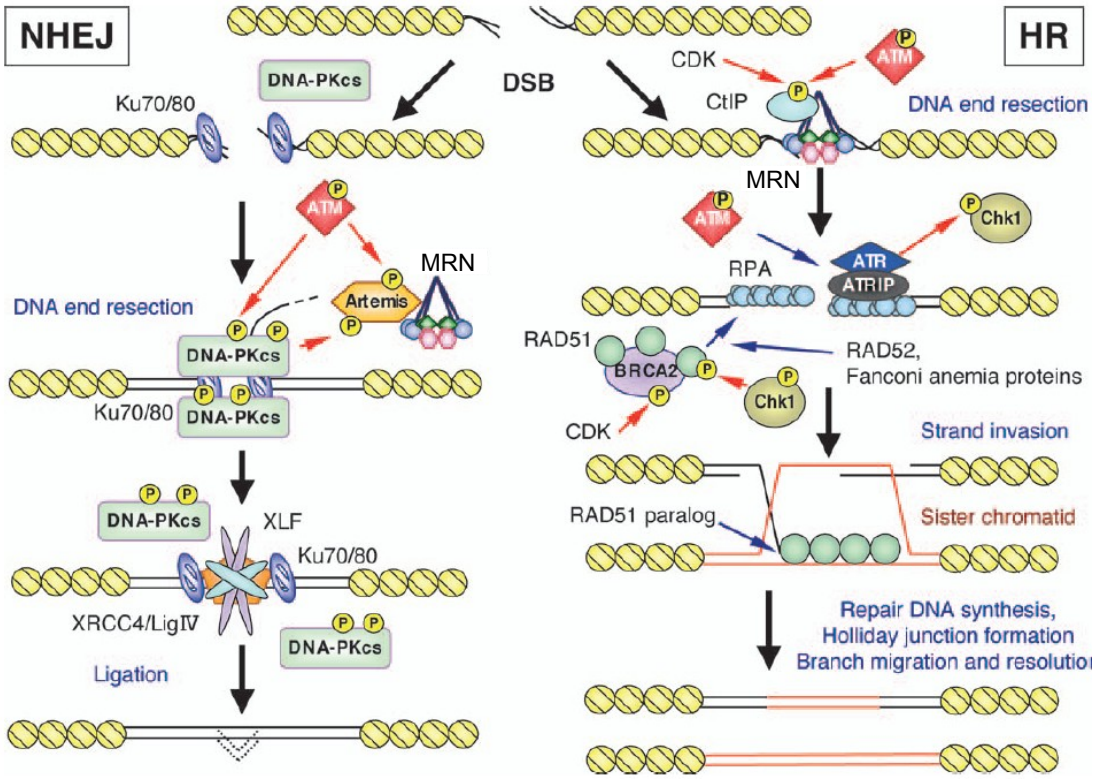


Figure 12: DSB repair pathways – NHEJ and HR (Iijima *et al.* 2008). In NHEJ, the Ku70/80 heterodimer binds to the DSB ends and recruits DNA-PKcs, which in turn phosphorylates and recruits the MRE11-RAD50-NBS1 (MRN) complex or Artemis for DSB end-processing. The DSB is then ligated by the XRCC4/LIGIV heterodimer with potential loss or gain of a few nucleotides. In HR, MRN detects and binds to the DSBs, initiates end-processing facilitated by CtIP and amplifies the repair signal by phosphorylating ATM. RAD51 and BRCA2, regulated by CDK, Chk1, RAD52 and Fanconi anaemia proteins, substitutes RPA at the processed DNA ends, initiating sister chromatid strand invasion and Holliday junction formation. The damaged DNA is then repaired by DNA synthesis using the sister chromatid template for error-free repair.

5.3.1.5. Homologous recombination

HR is a high fidelity error-free repair process for DNA DSBs due to the availability of a sister chromatid template for repair (Figure 12). The MRE11-RAD50-NBS1 (MRN) complex detects DSBs and binds to the DNA ends forming a scaffold to tether the broken ends, and recruits and phosphorylates ATM (Figure 13). This in turn phosphorylates H2AX, resulting in a cascade of phosphorylation events amplifying the repair signal leading to end resection and processing by CtIP and MRE11, histone modification and chromatin remodeling, activation of cell cycle checkpoints and initiation of HR (Lavin 2007; Sartori *et al.* 2007; Williams *et al.* 2007; Iijima *et al.* 2008; Zha *et al.* 2009; Eid *et al.* 2010). RAD51 and BRCA2 are then recruited and promote sister chromatid strand invasion and formation of Holliday junctions. DNA repair synthesis using the sister template is completed by DNA polymerases followed by ligase sealing of the strand breaks, Holliday junction disengagement and resolution of repair (Figure 12) (Bristow and Harrington 2005; Hall and Giaccia 2006; Kass and Jasin 2010). PARP1 has also been implicated in promoting MRN recruitment and binding to DSBs (Haince *et al.* 2008).

5.3.1.6. Non-homologous end-joining

NHEJ is an error-prone DSB repair pathway that does not require sequence homology. Classical NHEJ requires recognition of DSBs by the DNA-dependent protein kinase complex (DNA-PK) consisting of Ku70, Ku86 and DNA-PK_{cs}. Ku competes competitively with MRN and PARP1 to bind to DSB ends thus inhibiting DNA end resection, HR and microhomology-mediated end joining (MMEJ) (see below) (Fattah *et al.* 2010; Kass and Jasin 2010; Cheng *et al.* 2011). DNA-PK activates Artemis for end-processing and gap-filling usually resulting in short nucleotide insertions or deletions, followed by bridging and ligation of DSB ends by XRCC4 and LIG4.

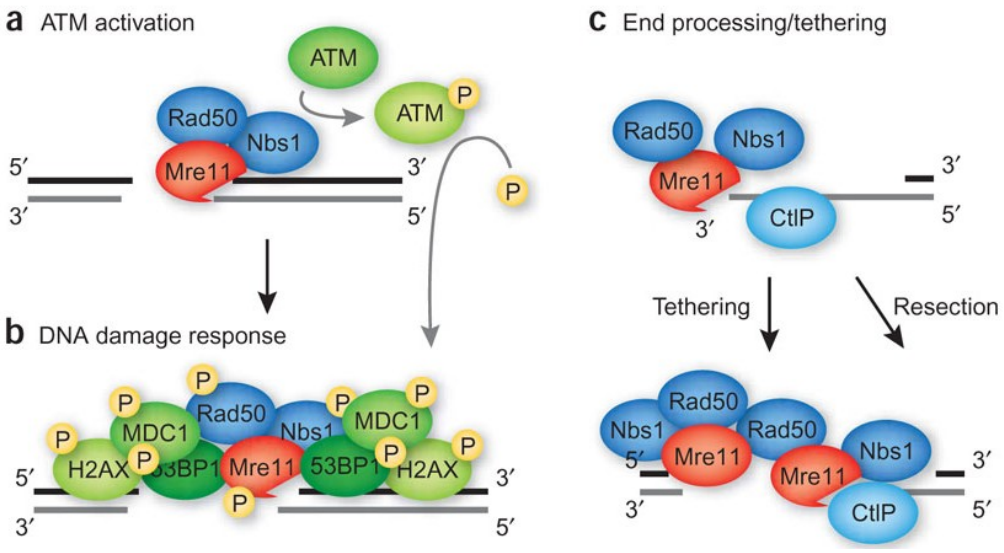


Figure 13: The MRE11-RAD50-NBS1 (MRN) complex and its functions in DSB repair signalling (Zha *et al.* 2009). (a) The MRN complex binds to DNA DSB ends and phosphorylates ATM thus, (b) activating the DNA damage response with downstream ATM phosphorylation targets, and (c) as well as binding to the DSB ends, the MRN complexes tethers the two broken ends together and initiates end-resection and processing with CtIP.

The alternative or back-up NHEJ pathway also known as MMEJ, is less efficient and more error-prone than classical NHEJ, and is predominantly seen in the event of defective classical NHEJ. MMEJ requires short homologous sequences (microhomologies) near the DSB break which are paired for DNA repair resulting in large deletions. MRN and PARP1 detect the lesion and undertake end-resection to the sites of microhomology with the DSB ends finally joined by XRCC1 and LIG3 (Audebert *et al.* 2004; Fattah *et al.* 2010; Cheng *et al.* 2011; Della-Maria *et al.* 2011). MMEJ has been implicated in chromosomal translocations and has a greater role in DSB repair in cancer cells compared to normal cells (Della-Maria *et al.* 2011). High-grade and muscle-invasive bladder tumours have been found to preferentially repair microhomology related DSBs by MMEJ over the classical pathway (Bentley *et al.* 2004; Bentley *et al.* 2009).

5.3.2. DNA Repair Variants and Cancer Predisposition

DNA damage repair mechanisms are crucial for the prevention of the fixation of somatic mutations caused by carcinogens and hence tumorigenesis (Wu *et al.* 2006). DNA repair gene SNPs may impair DNA repair capacity and, from multiple case-control studies, GWAS and meta-analyses, have been found to be associated with cancer susceptibility in most cancer sites (Berwick and Vineis 2000; Goode *et al.* 2002; Hung *et al.* 2005; Landi *et al.* 2006; Shen *et al.* 2010; Ricceri *et al.* 2011; Yu *et al.* 2011). These studies have focused predominantly on candidate common coding SNPs and, as with other common SNP studies, effects sizes have been small. Studies of DNA repair rare variants have focused on candidate genes implicated in familial cancer disorders such as in *BRCA1* and *MUTYH*, as previously mentioned (Theodoratou *et al.* 2010; Capanu *et al.* 2011).

In bladder cancer, epidemiological studies have found associations between coding single nucleotide polymorphisms (SNP) in BER, NER and DSB repair genes and bladder cancer susceptibility (Garcia-Closas *et al.* 2006; Figueroa *et al.* 2007; Figueroa *et al.* 2007; Kiltie 2009). In a pooled analysis by the International Consortium on Bladder Cancer, three DNA repair gene coding SNPs: *ERCC2* D312N (rs1799793); *NBS1* E185Q (rs1805794) and *XPC* A499V (rs2228000), were confirmed to significantly increase bladder cancer risk, though their effect sizes were small (OR ~1.10) (Stern *et al.* 2009). Analysis of tagging SNPs of all known NER genes identified SNPs in six genes modifying bladder cancer risk with SNP-SNP and SNP-smoking interactions (Xing *et al.* 2012). Dr Kiltie's group investigating the role of rare variants in bladder cancer predisposition, have recently identified four new rare *XPC* risk variants (OR=3.1, 95%CI 1.0-9.8, p=0.048) that alter gene regulation and *in vivo* function (Qiao *et al.* 2011).

5.3.3. Ionising Radiation and DNA

Ionising radiation (IR), such as radiotherapy, damages DNA directly or indirectly by ionizing water and generating free radicals, resulting in base damage, single-strand breaks (SSB) and double-strand breaks (DSB) (Table 6), usually with a cluster of DNA lesions within a few nanometers of each other (Steel 2002). DNA DSBs are the most lethal form of ionising radiation-induced DNA damage and it is the failure to repair these DSBs that results in cancer cell death.

Table 6: Lesions per Gray of ionising radiation (Steel 2002)

	Number of lesions per Gy
DNA crosslinks	2000
Base damage	1000
Single-strand breaks	1000
Double-strands breaks	40

5.3.3.1. DNA Repair in Radiobiology

The five R's of radiation biology describe the principles determining the response of cancer and normal cells to ionising radiation. They are: 1) Repair, the cellular capability to repair radiation induced DNA damage, 2) Redistribution or reassortment, the movement of cells through the cell cycle phases from radioresistant (S) to more radiosensitive (late G_2 / M) phases, 3) Repopulation, regrowth due to cell division, 4) Reoxygenation, the oxygen availability in cells which is crucial for the fixation of indirect DNA damage from ionising radiation, and, 5) Radiosensitivity, the intrinsic radiosensitivity of different cell types (Steel 2002). DNA repair, forming one of the 5 R's, influences RT cancer cell kill by its role in repairing both lethal DSBs and sublethal/ potentially lethal DNA damage - lesions that could potentially generate DSBs from stalled replication forks. The repair of these latter DNA lesions is also important in acute normal tissue radiation toxicity, with one of the rationales for RT

treatment fractionation being to allow sufficient time for normal tissue cells to repair these lesions prior to the delivery of the next RT fraction.

5.3.4. Biomarkers of Bladder Cancer and Radiotherapy Outcomes

5.3.4.1. Predictive and prognostic markers in bladder cancer

With adoption of the concept of personalised cancer medicine, predictive and prognostic markers are vital for treatment individualisation (La Thangue and Kerr 2011; Mirnezami *et al.* 2011). Predictive markers indicate the likelihood of cancer response to a specific treatment, while prognostic markers indicate the likely disease outcomes for an individual. **Clinical** prognostic markers in bladder cancer are TNM staging, tumour grade and multifocality, hydronephrosis and concurrent carcinoma *in situ*, and have been used to identify individuals who may gain from increased surveillance, treatment intensification or adjuvant treatment (Lopez-Beltran *et al.* 2004). Several clinical nomograms for MIBC post-radical cystectomy have been generated, though these are not yet in routine clinical use (Karakiewicz *et al.* 2006; Shariat *et al.* 2006; Nuhn *et al.* 2012).

A **molecular** biomarker is a biochemical "characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" (Biomarker Definitions Working Group 2001). Several predictive and prognostic biomarkers are already in routine clinical use such as HER2 and oestrogen receptor status in breast cancer, *BCR-ABL* translocations in chronic myeloid leukaemia, and *BRAF* mutations in melanoma (La Thangue and Kerr 2011).

Biomarker studies in bladder cancer have predominantly employed immunohistochemistry methods (Matsushita *et al.* 2011). FGFR3, p53 and Ki-67 (a marker of cell proliferation) expression, have been associated with tumour grade,

progression, recurrence and cancer survival (Hernandez *et al.* 2006; Mhawech-Fauceglia *et al.* 2006; Margulis *et al.* 2009; Goebell *et al.* 2010; van Rhijn *et al.* 2012). Though *TP53* mutations have been reported to be predictive of cisplatin-based chemosensitivity (Nishiyama *et al.* 2008), a recent phase III RCT failed to detect any prognostic or predictive value of TP53 protein expression in MIBC patients receiving adjuvant chemotherapy (Stadler *et al.* 2011). Immunohistochemistry studies of proteins involved in cell cycle regulation, apoptotic signalling, growth factor receptors and angiogenesis as potential biomarkers have so far been inconclusive (Lopez-Beltran *et al.* 2004; Proctor *et al.* 2010; Matsushita *et al.* 2011). Genetic studies have also identified gene expression signatures, epigenetic histone methylation marks and miRNAs as potential prognostic biomarkers (Catto *et al.* 2009; Dyrskjot *et al.* 2009; Schneider *et al.* 2011; Riester *et al.* 2012). However, none of the above markers has been successfully implemented clinically to date.

5.3.4.2. Biomarkers of radiotherapy

Radiotherapy biomarker research has so far aimed to identify predictors of normal tissue and tumour radiosensitivity, which may identify potential drug targets or allow individualisation of radiotherapy dosimetry, dose-fractionation or concurrent chemotherapy or radiosensitiser treatments (Alsner *et al.* 2008; Mukherjee *et al.* 2010; Sak and Stuschke 2010; Toustrup *et al.* 2012). High tumour EGFR expression or *EGFR* mutation, and a hypoxic gene expression profile have been associated with cancer radioresistance and worse outcomes in a variety of cancer sites (Mukherjee *et al.* 2010; Klopp and Eifel 2012; Toustrup *et al.* 2012). Positive human papilloma virus infection status in HPV-related squamous cell carcinomas are both prognostic and predictive of good radiotherapy response (Begg 2012; Klopp and Eifel 2012) while γ-H2AX assays of normal patient lymphocytes during radiotherapy have had moderate success in predicting normal tissue radiation toxicity (Sak and Stuschke 2010).

In MIBC, high tumour MRE11 expression in a test and validation cohort was found to be predictive of improved cancer specific survival (CSS) following radical RT but not radical cystectomy (Choudhury *et al.* 2010). This result has recently been independently validated (Laurberg *et al.* 2012). High MRN expression has been correlated with reduced local recurrence following post-operative radiotherapy for early breast cancer while overexpression of NBS1 in head and neck squamous cell carcinomas was associated with radiosensitisation (Rhee *et al.* 2007; Soderlund *et al.* 2007). Further validation is currently underway, testing MRE11 as a clinical biomarker to identify patients best managed with bladder preserving treatment. Other MIBC studies have also indicated associations of APE1 and XRCC1 expression and ERCC1 mRNA levels with RT and chemo-RT response (Sak *et al.* 2005; Kawashima *et al.* 2011).

The new field of radiogenetics aims to identify the genetic determinants of individual clinical radiosensitivity (Burnet *et al.* 2006; Alsner *et al.* 2008). Multiple studies have demonstrated links between DNA repair gene SNPs with acute and late radiation normal tissue toxicity (Pugh *et al.* 2009; Parliament and Murray 2010) and the multi-centre RAPPER radiogenomics study is currently underway to investigate the role of common SNPs in radiation toxicity using GWAS techniques (Burnet *et al.* 2006; Barnett *et al.* 2012). The role of germline SNPs in predicting radiotherapy outcomes has been less extensively investigated. Candidate SNP studies have found associations of DNA repair gene SNPs with pathological response or survival in several cancer sites following radiotherapy treatment, predominantly in combination with concurrent chemotherapy (Parliament and Murray 2010; Yin *et al.* 2011; Yoon *et al.* 2011). In MIBC, Sakano *et al* (2006) identified coding *ERCC2* and *XRCC1* SNPs as potential prognostic markers following platinum-based chemoradiotherapy (Sakano *et al.* 2006). However, all these small sized, positive studies remain focused

on selected candidate common coding SNPs which as previously discussed are less likely to have clinically significant functional effects.

6. Aims and Objectives

The two main aims of this project were:

1) *To study the contribution of DNA repair gene variants on bladder cancer risk, with the ultimate goal of improving the understanding of inherited risk in bladder cancer.*

   The following objectives were set to achieve this aim:

   i. To develop a protocol for multiplexed NGS to allow the cost-effective targeted sequencing of multiple samples for the identification of rare germline variants, exploring issues of target amplification, optimum sample pooling, rare variant calling and selection, and sensitivity and specificity of the method.

   ii. To test the CDRV hypothesis in two candidate DNA repair genes, *XPC* and *MUTYH*, by undertaking a pilot multiplexed NGS study to identify rare germline variants in *XPC* and *MUTYH* that may influence bladder cancer susceptibility in an enriched case-control population followed by confirmation of candidate, putatively-functional variants in a larger case-control study.

   iii. To investigate in a large case-control study the association of DNA repair gene 3'UTR SNPs predicted to affect miRNA binding *in silico* on bladder cancer risk. These SNPs would also be investigated for associations with clinical outcomes in MIBC treated with radical RT.

2) *To investigate DNA repair as a potential marker of RT response and outcomes in MIBC to allow future treatment individualisation, aid patient choice of the most effective treatment, and potentially improve overall bladder cancer survival.*

The following objectives were set to achieve this aim:

i. To use novel bar-coded multiplexed next-generation sequencing technology to identify germline *MRE11A* variants associated with RT outcomes in MIBC patients treated with radical RT.

ii. To examine tumour protein expression of several DNA repair proteins involved in DSB processing, NER and BER as potential predictive biomarkers of RT response in MIBC.

7. Materials and Methods

7.1. Materials

Reagents and kits are listed in Appendix A.

7.2. Study Population

7.2.1. Ethical approval

Local ethical approval was obtained from Leeds (East) Local Research Ethical Committee (studies 02/060, 02/192 and 04/Q1206/62) (Appendix B).

7.2.2. Leeds Bladder Cancer Study (LBCS) Population

As previously described (Choudhury *et al.* 2008), between August 2002 and October 2009, bladder cancer cases (N = 853) were recruited from the Pyrah Department of Urology, St James's University Hospital, Leeds (SJUH). All recruited cases had urothelial carcinomas except for two cases that also had synchronous squamous cell carcinomas. Hospital-based controls (N = 546) were recruited specifically for this study from the ophthalmology, and ear, nose and throat (ENT) outpatient departments, SJUH. Community controls (N = 227) were previously recruited for a colorectal cancer case-control study undertaken by the Leeds Cancer Research UK Centre (Barrett *et al.* 2003). Controls had no history of cancer or haematuria, with attempts made to frequency match for age and gender. As both recruited control groups were cancer-free, of similar age and gender frequencies, and from the same georgraphical region and ethnic distribution (over 97.8% Caucasian), it was deemed unlikely that the community recruited control group would incur any genetic biases. Each participant was provided with a patient information leaflet, consent form and a questionnaire covering smoking, and occupational and family history, which were then discussed and completed (Appendix B). Five millilitres of blood was collected in

an EDTA tube for DNA extraction. DNA extraction from peripheral blood leucocytes was performed by the Yorkshire Regional Genetics Service, SJUH, using a salting-out protocol involving isolation of cell nuclei by centrifugation, lysis of nuclei and use of a high concentration sodium chloride solution to precipitate out proteins. The DNA supernatant was then purified by ethanol precipitation and eluted in Tris-EDTA buffer. Sample concentrations were measured by spectrophotometry (Nanodrop, Thermo Scientific, USA) and stored at -20°C.

7.2.3. Muscle Invasive Bladder Cancer (MIBC) Study Population

7.2.3.1. Germline DNA

Between August 2002 and October 2009, 201 bladder cancer patients referred for radical RT were recruited in the Pyrah Department of Urology, SJUH. Consent was obtained for the study at their first oncology outpatient's appointment and a blood sample for germline DNA extraction (as described above) was collected prior to initiating any treatment. All patients had MIBC (WHO T-stage 2 or greater) on pre-treatment transurethral biopsy specimens and no evidence of metastases on radiological staging. Patients were treated using three-dimensional conformal external beam radiotherapy (52.5 – 55 Gy in 20 fractions over four weeks), initially at Cookridge Hospital and latterly at St James's Institute of Oncology (SJIO), Leeds, UK (Kotwal *et al.* 2008). Clinical outcomes and follow-up data were collected from clinic letters and electronic medical notes prospectively. Radiation Therapy Oncology Group (RTOG)/European Organisation for Research and Treatment of Cancer (EORTC) scores (Cox *et al.* 1995) for late bladder and bowel radiation toxicity were collected at each follow-up visit using a set proforma (Appendix B).

7.2.3.2. Formalin Fixed Paraffin Embedded (FFPE) TURBT Blocks

This previously described population (Choudhury *et al.* 2010) consisted of non-metastatic MIBC patients treated by radical RT between 2002-2005 for MIBC at

Cookridge Hospital and SJIO, Leeds, UK. Clinical FFPE tumour blocks taken at pre-treatment TURBT were available for 88 individuals. Patient treatment details were the same as above with clinical data also collected prospectively.

7.3. Histopathology

7.3.1. Cell lines and culture

The RT112 bladder urothelial carcinoma cell line was a generous gift from Prof M Knowles; the Daudi Burkitt lymphoma cell line was generously provided by Dr G Doody. The HeLa cervical squamous cell carcinoma cell line and GM15983 XPC SV40-transformed fibroblasts were obtained from American Type Culture Collection (LGC Standards, UK). RT112 and Daudi cells were cultured in RPMI 1640 (Sigma-Aldrich, UK) growth medium, and GM15983 and HeLa cell lines in MEM Eagle growth medium (Sigma-Aldrich, UK). Both media were supplemented with 10% v/v foetal bovine serum (Sigma-Aldrich, UK) and 1% 2 mM L-glutamine (Sigma-Aldrich, UK) and all cell lines were incubated in a 37°C 5% CO_2 humidified atmosphere.

7.3.2. FFPE cell line pellets

Over 1 x 10^7 cells in solution were needed to generate a cell pellet. Daudi cell lines were used directly. For adherent cell lines (RT112, HeLa and GM15983), the growth medium was discarded and the cells washed with 10 ml phosphate buffered saline (PBS): 0.1% EDTA for 10 minutes. The wash was discarded and the cells were then incubated with 0.25% Trypsin-EDTA solution at 37°C for five minutes. Cells were suspended in growth medium and 20 µl of suspended cells pipetted onto a haematocytometer to determine cell concentration. The appropriate volume of suspended cells were transferred into a conical tube and centrifuged at 2000 rpm at 4°C for four minutes. The supernatant was removed without disturbing the pellet and the cells resuspended and washed with PBS, and recentrifuged at 2000 rpm at 4°C for four minutes. The PBS was then discarded (without disturbing the pellet) and the

cell pellet fixed in 5 ml 4% Formalin in PBS pH 9 overnight. The next day, the formalin was removed and 5 ml 70% ethanol added to the fixed cell pellet with the cell pellet. The cell pellet was then brought to the Histopathology laboratory for paraffin embedding.

7.3.3. Cell protein lysates and protein concentration quantification

Cells were grown to 80-90% confluence for adherent cells in a 50cm^2 cell culture flask, incubated with 0.25% Trypsin-EDTA solution as described above, suspended in growth medium, and transferred to a 50 ml centrifuge tube. The cell suspension was centrifuged at 450 g for five minutes. The medium was then carefully aspirated without disturbing the cell pellet and discarded. The cell pellet was then washed three times by resuspending in ice cold PBS, centrifuging at 450 g for five minutes and discarding as much of the supernatant as possible. The cell pellet was then resuspended in 1 ml PBS, transferred to an Eppendorf tube, centrifuged again at 13000 rpm for one minute, and the supernatant discarded. The cell pellet was then incubated on ice with a mixture of 100 µl RIPA buffer (Sigma-Aldrich, UK), 1 µl phosphatase inhibitor (Roche, Switzerland) and 1 µl proteinase inhibitor (Roche, Switzerland) for 20 minutes, then centrifuged at 13000 rpm for one minute. The cell lysate was then transferred to a fresh Eppendorf tube for quantification.

Cell lysate protein concentration was determined using the Bradford protein assay. By diluting bovine serum albumin (BSA) (Sigma-Aldrich, UK) in sterile water, 50 µl protein standards containing 0 µg, 2 µg, 4 µg, 6 µg, 8 µg and 10 µg of BSA were made. One millilitre of Bradford solution (Sigma-Aldrich, UK) was added and the mixture transferred into a transparent cuvette. The absorption spectrum was measured for each protein standard at 595 nm using a spectrophotometer (Jenway, UK) to generate a standard curve. In triplicate, one microlitre of the cell lysate sample was diluted in 49 µl sterile water, 1 ml Bradford solution was added, and the mixture

transferred to a cuvette. The absorption spectrum was then measured for each triplicate sample and the mean absorption spectrum of the three samples used to determine the sample protein concentration using the generated standard curve. Cell lysate samples were stored at -20°C.

7.3.4. Western Blotting

All primary antibodies were tested by western blotting to confirm target specificity on cell lysates from Daudi, HeLa and RT112 cell lines. A 10% acrylamide gel was made using the recipe below.

Resolving Gel

Reagent	*Volume*
2M Tris pH 8.8	1.87 ml
10% sodium dodecyl sulfate	100 µl
Protogel	3.33 ml
Sterile distilled water	4.7 ml
TEMED	6 µl
10% ammonium persulphate	30 µl

Stacking gel

Reagent	*Volume*
2M Tris pH 6.8	
10% sodium dodecyl sulfate	100 µl
Protogel	1.33 ml
Sterile distilled water	7.94 ml
TEMED	6 µl
10% ammonium persulphate	30 µl

50 µg of cell lysate protein sample was mixed with 2x sodium dodecyl sulphate and 20% mercaptoethanol loading buffer, heated to 95°C for five minutes and placed on ice. Samples were then loaded onto the acrylamide gel alongside 10 µl of a dual colour protein marker (Bio-Rad Laboratories, UK) and run at 130 V in Tris-Glycine-SDS running buffer (Bio-Rad Laboratories, UK) for two hours. Proteins were then transferred onto a nitrocellulose supported membrane (Amersham Pharmacia Biotech, UK) at 100 V for one hour in ice-cold Tris-Glycine transfer buffer (Bio-Rad Laboratories, UK). Protein transfer was checked by staining with Ponceau Red (Sigma-Aldrich, UK). Membrane blocking was performed by incubation in 50% Odyssey Blocking Buffer (Licor, USA) in PBS for one hour at room temperature. The membrane was then incubated with the primary antibody at the manufacturer recommended dilution overnight at 4°C, washed four times with PBS-0.01% Tween

20 (Sigma-Aldrich, UK), followed by incubation with the appropriate fluorescent secondary antibody, Alexa Fluor 680 anti-rabbit (Molecular Probes, UK) or IR Dye 800 Anti-mouse (Rockland Inc, USA), for one hour at room temperature. The membrane was washed again four times with PBS-0.01% Tween 20 then imaged on an Odyssey Infrared Imaging System (Licor, USA).

7.3.5. Haemotoxylin and Eosin (H&E) Staining

For each patient, a 4 μm FFPE tissue section was stained with haematoxylin and eosin (H&E). Sections were deparaffinised in three changes of xylene for five minutes each, then rehydrated in two changes of absolute alcohol for two minutes each then 90% alcohol for two minutes and 70% alcohol for two minutes followed by washing in tap water. The sections were then stained in haematoxylin (Sigma-Aldrich, UK) for three minutes, Scott's tap water (Sigma-Aldrich, UK) for one minute and eosin (Sigma-Aldrich, UK) for three minutes. Finally, the sections were dehydrated in three changes of absolute alcohol and two changes of Xylene for one minute each and mounted with a coverslip using DPX mountant (Sigma-Aldrich, UK). Slides were reviewed by a consultant uropathologist and areas of muscle-invasive urothelial carcinoma were outlined.

7.3.6. Immunohistochemistry (IHC)

As per Table 7, immunohistochemistry was undertaken using a standard streptavidin-biotin complex (SBC) method or the MenaPath X-Cell Plus HRP-Polymer Detection Kit (Menarini Diagnostics, UK). FFPE tissue section slides were deparaffinised, rehydrated and washed as described above. Endogenous peroxidases were blocked using 3% hydrogen peroxide (Sigma-Aldrich, UK) for 20 minutes, followed by tissue section antigen retrieval by pressure-cooking at 15 psi for two minutes in the appropriate buffer (see Table 7). Slides were then stained vertically using a Shandon coverplate assembly (ThermoFisher Scientific, UK). For antibodies using the SBC

method, endogenous avidin binding sites, biotin and biotin receptors were blocked using an Avidin/ Biotin Blocking kit (Vector Laboratories, USA) with an initial avidin block for 15 minutes, two washes with Tris-buffered saline (TBS) followed by a biotin block for 15 minutes. This step was not needed for the MenaPath kit. The slides were then washed twice with TBS and incubated for 30 minutes with 10% normal goat serum (for SBC protocol) or casein (for the Menapath kit) diluted in TBS to block non-specific antibody binding. Slides were then incubated with the optimised concentration for the respective primary antibody for 60 minutes at room temperature or overnight at 4°C (Table 7).

Table 7: Primary antibody details, IHC conditions and positive control tissue used.

Primary Antibody	*Antigen Retrieval Buffer*	*IHC kit/ protocol*	*Dilution*	*Incubation time*	*Positive control tissue*
Anti-CtIP Mouse monoclonal Supplied by R Baer (Yu and Baer 2000)	10 mM citrate buffer pH 6.0	SBC	1:50	Overnight at 4°C	Daudi cell pellet
Anti-XPC Rabbit polyclonal Sigma C-terminus X1129	10 mM citrate buffer pH 6.0	SBC	1:8000	60 minutes at room temperature	Daudi cell pellet
Anti-MUTYH Mouse monoclonal Abcam ab55551	10 mM Tris-1 mM EDTA buffer pH 8.0 + 0.05% Triton X-100	Menapath	1:200	60 minutes at room temperature	HeLa cell pellet

For the SBC method, sections were incubated in biotinylated secondary antibody for 30 minutes, washed twice with TBS, incubated with streptavidin peroxidase (Dakocytomation, Denmark) for a further 30 minutes and washed again twice with TBS. For the Menapath kit, sections were incubated with the Universal Probe for 30 minutes, washed twice with TBS-0.1% Triton X-100 (Promega, USA), incubated with the HRP-Polymer for a further 30 minutes and washed again twice with TBS-0.1% Triton X-100. For both protocols, bound antibodies were visualised by incubation with diaminobenzidine (Dakocytomation, Denmark) for 10 minutes then washed in tap water. Sections were then counterstained in haematoxylin (Sigma-Aldrich, UK) for 15 seconds and Scott's tap water (Sigma-Aldrich, UK) for one minute. Finally, the

sections were dehydrated in three changes of absolute alcohol and two changes of Xylene for one minute each and mounted with a coverslip using DPX-Mountant (Sigma-Aldrich, UK).

Primary antibodies were tested and concentrations optimised against sections from a panel of tissue (normal breast, skin, placenta, tonsil and papillary bladder tumour) as well as cell line pellets (HeLa, RT112 and Daudi cell lines) to identify the ideal positive control. The final primary antibody dilution was chosen by two observers so that on a scale of 0 - 3 (Figure 14) the nuclear staining in the positive control scored 2. Negative controls consisted of the positive control tissue with the primary antibody omitted.

For the anti-XPC antibody, due to the presence of a second unspecified band on Western blotting, a cell pellet made from the XPC non-expressing GM15983 cell line was used as a separate negative control tissue. The optimum antibody concentration was the highest concentration with no staining in the GM15983 cell pellet but positive staining in the Daudi cell pellet positive control. A GM15983 negative control was used for all anti-XPC IHC staining runs.

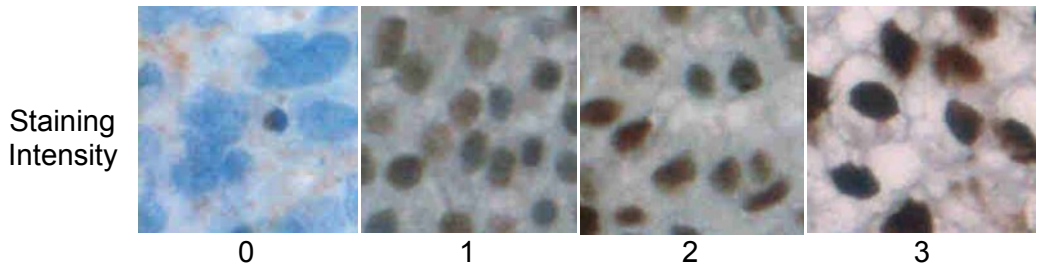


Figure 14: Representative sections immunostained for MRE11 (Abcam mouse monoclonal ab214, UK) demonstrating the different staining intensities.

7.3.7. Microscopy and Photography and Scoring

Using the H&E-stained slides as a reference, areas of muscle-invasive tumour were demarcated on the IHC stained slides. Digital images were taken of 10 random fields

from within these areas using an Olympus BX50 microscope and c-3030 camera (Olympus UK, UK) at 600x magnification. A “quickscore” method was used for semi-quantitative assessment of protein expression (Reiner *et al.* 1990; Detre *et al.* 1995). For each field, the proportion of bladder cancer cells staining positive was scored from 1 to 4 (1 = 0 – 10%, 2 = 11 – 30%, 3 = 31 – 70%, 4 = 71 – 100%) as well as the nuclear staining intensity which was scored 0 to 3 as per Figure 14. The nuclear staining intensity was scored independently by two observers, and discordant scores reviewed together and a consensus reached. A semi-quantitative score (SQS) was used to assess overall tumour DNA repair protein expression based on the product of the median score of the proportion of positive staining cells and the modal intensity. For MUTYH, due to its expression in mitochondria (Ohtsubo *et al.* 2000), bladder cancer cytoplasmic staining was also scored as present (1) or absent (0) as previously described by Gao *et al* (2004) (Gao *et al.* 2004).

7.4. Genetics

7.4.1. DNA master plate design and alliquoting

	1	2	3	4	5	6	7	8	9	10	11	12
A	Empty for orientation/ identification	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Empty for Assay Controls
B	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Empty for Assay Controls
C	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Empty for Assay Controls
D	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Empty for Assay Controls
E	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Empty for Assay Controls
F	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Empty for Assay Controls
G	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Sample	Empty for Assay Controls
H	Sample	Sample	Sample	Sample	Sample	Sample	Replicate/ Control sample	Replicate/ Control sample	Replicate/ Control sample	Replicate/ Control sample	Replicate/ Control sample	Empty for Assay Controls

Figure 15: Blueprint for 96-well master plates for both the LBCS and MIBC study germline DNA samples. The empty well for orientation/ identification (red) is changed for each master plate to aid plate identification.

Ninety six-well master DNA plates were designed for germline DNA samples from both the LBCS and MIBC study populations using the blueprint shown in Figure 15.

For alliquoting, stock DNA tubes were arranged in 96 tube racks based on the corresponding 96-well plate design. Two different identifiers were checked by two people for each sample to ensure accurate layout. Samples were then alliquoted using an automated robotic liquid handling workstation (Corbett Robotics, Australia) and diluted with molecular grade water to produce 100 µl of 50 ng/ µl germline DNA for each sample.

7.4.2. Reference sequences

The wildtype genomic sequences and gene annotations for the candidate genes were obtained from the University of California Santa Cruz (UCSC) genome browser (http://genome.ucsc.edu): *MUTYH* (NCBI Accession NG_008189 human genome build 18 accessed October 2008), *XPC* (NCBI Accession NT_022517 human genome build 18 accessed August 2008 and NCBI Accession NG_011763.1 human

genome build 19 accessed March 2011), and *MRE11A* (NCBI Accession ng_007261 human genome build 19 accessed January 2010). Reference protein sequences were obtained from the NCBI protein database (http://www.ncbi.nlm.noh.gov/protein) for *MUTYH* isoform 1 (NP_036354.1 accessed February 2009), *XPC* isoform 1 (NP_004619.3 accessed February 2009) and *MRE11A* isoform 1 (NP_005582.1 accessed June 2010).

7.4.3. Primer design

PCR primers were designed using Primer3 (http://frodo.wi.mit.edu/primer3/) using the default settings with some alterations depending on primer function (Table 8). The respective wildtype reference genomic sequence was used for primer design with surrounding common SNPs annotated.

Table 8: Primer3 settings for primer design according to primer purpose.

Settings	Conventional Sequencing Primers	High Resolution Melting Primers	Long PCR Primers
Product Size	100-800 bases	150-300 bases	Up to 15 kb
Primer Size	18-25 bases Optimum: 20 bases	18-25 bases Optimum: 20 bases	22-27 bases Optimum: 25 bases
Primer Melting Temperature	52-63°C Optimum: 55°C	57-65°C Optimum: 63°C	55-65°C Optimum: 60°C
Maximum melting temperature difference	100°C (Default)	1°C	100°C (Default)
CG Clamp	0	0	1-2
Other	Variant of interest not within 20 bases of at least one of the primers	Variant of interest not within 20 bases of at least one of the primers No other common SNPs within the fragment	Primer not located within 200 bases 5'- or 3' of an exon. Primer not sited on a common SNP or within a known DNA repeat region.

7.4.4. Polymerase Chain Reaction (PCR)

7.4.4.1. Standard PCR

Standard PCR was carried out using the HotStarTaq master mix kit (Qiagen, UK) in a 10 µl reaction mix using 40 pmol of each primer, 10 ng of genomic DNA, 5 µl of 2x HotStarTaq master mix and made up to 10 µl with molecular grade water. Reaction conditions were – initial denaturation: 95°C for 15 minutes, 36 cycles of [denature: 95°C for 30 seconds; anneal: 50-65°C for 30 seconds; extend: 72°C for 30 seconds], followed by a final extension of 72°C for 10 minutes. The optimum annealing temperature for each primer pair set was determined by amplifying control germline DNA samples using an annealing temperature gradient followed by gel electrophoresis of PCR products.

7.4.4.2. Long Polymerase Chain Reaction (PCR) Amplification

For the initial pilot NGS study, *MUTYH* was divided into two amplicons and *XPC* into nine amplicons. *MRE11A* was divided into 12 amplicons, while for the resequencing of *XPC*, the gene was divided into four amplicons. Several proof-reading capable long PCR kits, ABgene Extensor Long PCR Master Mix (ThermoFisher Scientific, UK), Bio-X-act (BioLine, UK), Invitrogen SequelPrep (Life Technologies, UK) and Fermentas Phusion Flash (ThermoFisher Scientific, UK), were tested for each primer set using control germline DNA samples on an annealing temperature gradient for successful PCR amplification with PCR products imaged following gel electrophoresis. Standard 10 µl reaction mixes and thermocycling conditions are shown in Table 9 and Table 10.

Table 9: Standard 10 µl long PCR reaction mixes for the different long PCR kits used.

Reagent	ABgene Extensor Long PCR Master Mix kit	Bioline Bio-X-act kit	Invitrogen SequelPrep kit	Fermentas Phusion Flash kit
DNA sample (25 ng/ µl)	2 µl	2 µl	2 µl	2 µl
Long PCR kit reagent 1	5 µl 2x master mix 1/2	5 µl 2x master mix	1 µl 10x buffer	5 µl 2x master mix
Long PCR kit reagent 2			0.5-1 µl 10x Enhancer A/B	
Long PCR kit reagent 3			0.18 µl Long Polymerase	
Forward Primer (100 µM)	0.05 µl	0.05 µl	0.05 µl	0.05 µl
Reverse Primer (100 µM)	0.05 µl	0.05 µl	0.05 µl	0.05 µl
DMSO			0.2 µl	
Molecular grade water	Made up to 10 µl	Made up to 10 µl	Made up to 10 µl	Made up to 10 µl

Table 10: Standard thermocycling conditions for the different long PCR kits used.

Thermocycling Step	ABgene Extensor Long PCR Master Mix kit	Bioline Bio-X-act kit	Invitrogen SequelPrep kit	Fermentas Phusion Flash kit
Initial Denaturation	92°C for 2 minutes	95°C for 3 minutes	94°C for 2 minutes	98°C for 10 seconds
Phase 1:	15 cycles	35 cycles	10 cycles	35 cycles
Denaturation	92°C for 10 seconds	95°C for 30 seconds	94°C for 10 seconds	98°C for 3 seconds
Annealing	50-68°C for 30 seconds	50-68°C for 45 seconds	50-68°C for 30 seconds	55-72°C for 5 seconds
Extension	68°C for 1 minute/ kb	72°C for 1 minute/ kb	68°C for 1 minute/ kb	72°C for 30 seconds/ kb
Phase 2:	20 cycles		25 cycles	
Denaturation	92°C for 10 seconds		94°C for 10 seconds	
Annealing	50-68°C for 30 seconds		50-68°C for 30 seconds	
Extension	68°C for 1 minute/ kb + 10 seconds per cycle		68°C for 1 minute/ kb + 20 seconds per cycle	
Final Extension	68°C for 7 minutes	72°C for 10 minutes	72°C for 5 minutes	72°C for 1 minute

PCR amplification conditions were further optimised to maximise PCR yield by adjusting the duration of thermocycling steps and/or the number of thermocycling cycles accordingly. Details for each candidate gene amplicons are shown below in Table 11 to Table 13 with full long PCR conditions in Appendix A. All candidate gene amplicons were amplified individually for each study sample.

Table 11: *XPC* and *MUTYH* long PCR amplicons and primers for the pilot multiplexed NGS study.

Gene	Amplicon	Exons	Forward PCR Primer	Reverse PCR Primer	Amplicon Size (bp)
XPC	1	1	TGAATAAATAATTGT TGCGCTCAC	GGAGGAAAGCATA GATTTAAAGAGG	1265
	2	2 - 3	GAGGTTAGCTGACA TTTAAGATCTGG	TACTCACTCACACT GCCACCTAAG	3261
	3.1	4 - 5.1	TCTAAAGACTGTCT GGGTTGTGTG	TCTAATCCTTTCCC CTACATAAAGC	1994
	3.2	5.2 - 6	AAGCAATTTGTATTT TGAAGCTTTG	TAGCACATCTGACC AAAAACTACG	1316
	4.1	7 - 8	TAGCACTAGACTGT TTCCAGAGTCC	GACAGTATCATGTC TGTCTCATTCC	3434
	4.2	9	AGTGAATTGCCTAC TGAGAAATGAG	AGAGACAGGGTTTC ATCGTGTTAG	1255
	5	10	AATCATTTGACTGTT CCATGTGC	GTGTCCTGTGAATA GTCTCTGTGG	1269
	6.1	11 - 14	ACAGGCTCTACTGA TGGACAGTTAC	AGACAGAAGACTGA GGTGTCCTAAC	2111
	6.2	15 - 16	AAAAGCCCTTCATC TGACTTTATG	GTCTGTTTCTCAAC AAGAGGTTCC	1987
MUTYH	1	1 - 4	GTATGAGCCTGGGT AAATTACTTCC	TTCAGTAAATAAAA CATCAAAAGTTCC	3049
	2	5 - 9	TAACAGTTAGGGAG CAGTGAAAATC	ATTAGAGCTGACCT GAGACGTAAAC	6583

Table 12: *MRE11A* long PCR amplicons and primers.

Gene	Amplicon	Exons	Forward PCR Primer	Reverse PCR Primer	Amplicon Size (bp)
MRE11A	1	1 - 3	CAGACGTTCTAATT CAAAAGTTCTAC	AGCTTTGACAGTCA CCTTACATTTC	3960
	2	4	TAACAGTTTGCCTC TTCCATTAGAG	TGTCAACAAGAGTC CATAAACAAAG	1250
	3	5 - 7	TCAAACTAATTGAA ATGAATTGTCG	TCTACCTTTTTGTTG AATTGAGGAG	3903
	4	8 - 9	CACTTGTCTGTGTC CAATAAGTTTG	TCCAGTATTCCTCT CTCTCAATGAC	1701
	5	10	AGGAAAGCCTTATT GAAACATGAG	AGACTGTGAGTGGT GCTATAAGAGG	2233
	6	11	CTTTCTCTTGAGGC ACATCTCTAAG	GAGGCCTCAGAAAT AACAATGC	2068
	7	12 - 13	TATGAACGAGAAAT ACTGACTGTGG	AGAGAAGATACTAT CCATGGGGAAC	2067
	8	14	ACGATAATATATGG CACATTGAAGG	GGTTGCTAACTTGT AGATTCAAACC	1394
	9	15 - 16	ATAGCTCACTGCAG CCTGTATCTC	CTCTCTCTGTTGCT AGGGTAGAGTC	2448
	10	17 - 18	TCTCAATAAGCTGG GAAATAATGAG	GGAGAGAACACAG TCTTCCATTAAG	2223
	11	19	TTTTCAGAAGTGGA AGATGGTAATG	AGGCACATTACAAA GAAGAAGACAC	1369
	12	20	AAAATTACAGCTGT TTCTGAGTTGG	TTGAGATTGAGTTT CACTCTTGTTG	3620

Table 13: Long PCR amplicons and primers for NGS resequencing of *XPC*.

Gene	Amplicon	Exons	Forward PCR Primer	Reverse PCR Primer	Amplicon Size (bp)
XPC	1	1 - 3	TAACCTAACTCAAC TGGTCCCTATG	GTAAACAGCCTTCT TTGGTAACTTG	9800
	2	4 - 7	GCTACTGTATACAA GCAACGTAGGG	TGTTTTATTCAGACA GAGCCTTACC	4649
	3	8 - 10	ACCCATAACCCAGT ATATGGATGC	TGTGGGGTGTGTTT TATACTGC	4394
	4	11 - 16	CTTTACCCCATCCT GATAGTCTGC	GTCTGTTTCTCAAC AAGAGGTTCC	8079

7.4.5. Gel Electrophoresis

Successful PCR amplification was confirmed by running PCR products on a 1% agarose gel. The gel was made by heating and dissolving 1 g of molecular grade agarose in 100 ml of Tris/Borate/EDTA (TBE) buffer (Bio-Rad Laboratories, UK). One microlitre of ethidium bromide was dissolved in the liquid agarose gel then the gel poured into a mould with the appropriate sized comb *in situ* and the gel left to set. Four microlitres of each PCR product was mixed with 1 µl sterile water and 1 µl 6x loading dye (Norgen Biotek, Canada). The set agarose gel was submerged in TBE buffer in a gel electrophoresis tank. PCR products were then loaded into each well alongside an appropriate sized DNA ladder: <1 kb ladder (New England Biolabs, USA), <10 kb ladder (Norgen Biotek, Canada), and >10 kb ladder (Invitrogen, UK). The samples were electrophoresed in a 100-150 V electric potential for 60-90 minutes and the PCR product bands visualised using a gel documentation ultraviolet imager (Bio-Rad Laboratories, UK).

7.4.6. PCR product purification

PCR products were purified to remove excess nucleotides, primers, enzymes and salts prior to downstream application.

7.4.6.1. Column purification

Column purification was performed using the standard QIAquick PCR Purification Kit (Qiagen, UK) protocols with the DNA eluted in 30 µl of 10 mM Tris-Cl (pH 8.5) buffer.

7.4.6.2. Exonuclease – Shrimp Alkaline Phosphotase

This PCR product purification process was used only for clean up prior to conventional dye terminator sequencing as the exonuclease – shrimp alkaline phosphatase reagents were found to interfere with DNA fragmentation during NGS library preparation. One microlitre of ExoSAP-IT (Affymetrix, USA), or 1 µl of

exonuclease 1 and 2 µl of shrimp alkaline phosphatase were added to 2.5 µl of PCR product to digest excess primers, degrade excess nucleotides and dephosphorylate enzymes. The mixture was incubated at 37°C for 30 minutes followed by enzyme inactivation at 80°C for 15 minutes. The treated sample was then diluted with 4 µl molecular grade water.

7.4.6.3. Polymer-based purification

The microCLEAN DNA clean-up reagent (Web Scientific, UK) was used for 96-well plate-based PCR product purification. Equal volumes of the microCLEAN reagent and PCR product were mixed in a 96-well plate and left at room temperature for five minutes. The plate was centrifuged at 4000 rpm for 40 minutes then centrifuged again upside down on tissue paper at 200 rpm for 30 seconds. Finally, the DNA was eluted in 30 µl of molecular grade water for five minutes and transferred to a fresh microcentrifuge tube.

7.4.7. DNA quantification

The concentrations of each purified amplicon were determined using a DNA standard curve assay. The Invitrogen Quant-IT BR PicoGreen Kit (Life Technologies, UK) was used. Two microlitres of each sample and lambda DNA standards (supplied at 0 ng/ µl, 5 ng/ µl, 10 ng/ µl, 20 ng/ µl, 40 ng/ µl, 60 ng/ µl, 80 ng/ µl and 100 ng/ µl) were mixed with 1 µl of Picogreen dsDNA reagent (a 502/523 nm fluorescent double-stranded DNA (dsDNA) binding compound) and 199 µl of 20x Tris-EDTA buffer on a black 96-well assay plate. The fluorescence intensity was then measured for each well using a spectrofluorometer (excitation 480 nm, emission 520 nm) (BMG LABTECH, Germany). A DNA standard curve was plotted using the fluorescence measurements of the lambda DNA standards and from this plot, the sample concentrations calculated.

7.4.8. Genotyping

7.4.8.1. High-throughput Taqman SNP genotyping

Genomic DNA samples were sent for genotyping using the Taqman SNP Genotyping Assay (Applied Biosystems, USA) by Gen-Probe Life Sciences (West Lothian, Scotland) or by the Leeds Cancer Research UK Genomics Facility. For each sample, five nanograms of germline DNA was dried on a 384-well plate and resuspended in 2 µl of reaction mix consisting of 1x Taqman Universal PCR Master Mix (Applied Biosystems, UK), which incorporates a Taq polymerase, and 1x SNP Genotyping Assay (Applied Biosystems, UK) containing the target specific forward and reverse primers, and two allele specific probes labelled with a VIC or FAM fluorescent dye linked to a fluorescence quenching molecule. All primers and probes used are listed in Appendix A. Blind duplicates from five percent of samples were included to ensure concordance of genotyping calls.

The sample mixtures were PCR amplified using the following thermocycler conditions: enzyme activation 95°C for 10 minute, 40 cycles of [denature: 92°C for 15 seconds; anneal and extend: 60°C for one minute]. During PCR amplification and DNA elongation, on reaching a perfectly complementary probe bound to the SNP of interest, the 5'-nuclease activity of Taq polymerase cleaves the probe separating the fluorescent dye from the quencher. Imperfectly bound probes are displaced but not degraded by the Taq polymerase. Fluorescence measurements were carried out using an Applied Biosystems 7900 real-time PCR machine (Applied Biosystems, UK) and a fluorescence plot of each sample measurement generated. SDS 2.2 software (Applied Biosystems, UK) was used for allelic discrimination and genotype calling of each sample.

In the study of DNA repair gene 3'UTR SNPs, as the *ATM* SNPs rs1137918 and rs227091 failed Taqman genotyping due to being located within Alu repeats, Taqman

genotyping was performed on a nested PCR product. PCR primers located in unique sequence flanking the Alu repeat were designed using Primer3 as previously described. Primers for rs1137918 were AAGGAACATCTCTGCTTTCACTC and GAGTGATCTTACTAGGAAAAATCCAAA (398 bp product) and for rs227091 were TGCACACAAGCCCATTCTTA and AGCTGGGGGACAGAGAAATG (842 bp product). PCR was carried out using the HotStarTaq master mix (Qiagen, UK) in a 10 µl reaction mix as previously described with a 58°C annealing temperature. Two microlitres of diluted 1:10 PCR product was then dried on a 384-well plate and Taqman SNP genotyping performed as described above. Conventional DNA sequencing was used to confirm the genotype. For sequencing, the primers GGAGTTTCGCTCTTGTCACC (rs1137918) and GTGCAGTGGCATGATCTCAG (rs227091) were used. There was complete concordance of genotype by sequencing and Taqman in all samples tested.

7.4.8.2. Fluidigm SNP genotyping

The Fluidigm 96.96 Dynamic Array Integrated Fluidic Circuit (IFC) system with custom Fluidigm SNPtype Assays (Fluidigm, USA) were used for SNP genotyping of genomic germline DNA samples. Each DNA sample was first pre-amplified in a 5 µl reaction mix of 1.25 µl germline DNA (at 10 ng/ µl to 90 ng/ µl), 2.5 µl of Qiagen 2x Muliplex PCR Master Mix (Qiagen, UK), 0.5 µl of 10x SNPtype Specific Target Amplification primer pool (consisting of 500 nM of each SNPtype Assay STA primer and SNPtype Assay LSP Primer combined from all 96 SNP assays) (Fluidigm, USA) and 0.75 µl of molecular grade water, and using the following thermal cycling conditions: initial denaturation: 95°C for fifteen minutes and 14 cycles of [denature: 95°C for 15 seconds; anneal and extend: 60°C for four minutes].

Pre-amplified samples were then diluted 1:100 in Tris-EDTA buffer. A sample mix was then made by combining 2.5 µl of each diluted pre-amplified sample with 3.5 µl of a master mix containing 3 µl Biotium 2x Fast Probe Master Mix (Biotium, USA), 0.3 µl SNPtype 20x Sample Loading Reagent (Fluidigm, USA), 0.1 µl SNPtype Reagent (Fluidigm, USA), 0.036 µl 50x ROX Reference Dye (Invitrogen, UK) and 0.064 µl of molecular grade water. A 10x assay mix was made for each assay by mixing 2.5 µl 2x Assay Loading Reagent (Fluidigm, USA), 1.5 µl molecular grade water and 1 µl of SNPtype Assay Mix (consisting of 7.5 µM SNPtype Assay ASP1/ASP2 and 20 µM SNPtype Assay LSP Primer for each SNP assay) (Fluidigm, USA).

Using the Fluidigm IFC Controller HX (Fluidigm, USA), the Fluidigm 96.96 Dynamic Array IFC chip was loaded with 4 µl of each 10x assay mix and 5 µl of each sample mix in the corresponding assay and sample inlets respectively. The IFC chip was then loaded onto the Biomark HD system (Fluidigm, USA) for PCR amplification and fluorescence image capture. Allelic discrimination and genotype calling of each sample was performed using the Fluidigm Genotyping Analysis software (Fluidigm, USA).

7.4.8.3. *High resolution melting mutation scanning*

High resolution melting (HRM) mutation scanning in genotyping works by measuring the differences in the melting temperature of dsDNA due to DNA sequence variation (Taylor 2009). Forward and reverse primers were designed using Primer3 as described earlier. PCR amplification was performed using 5 µl 2x HotShot Diamond PCR Master Mix (Clent Life Science, UK), 1 µl LCGreen PLUS dye (Idaho Technology, USA), 40 pmol of each primer, and 1 ng of genomic DNA made up to a 10 µl reaction with molecular grade water and with 1 µl DMSO added as required. Thermal cycling conditions were: initial denaturation: 95°C for five minutes, 45 cycles of [denature: 95°C for 30 seconds; anneal: 50-65°C for 30 seconds; extend: 72°C for

15 seconds], followed by a final denaturation step of 95°C for 30 seconds. The PCR annealing temperature was optimized for each primer pair using a control genomic DNA sample on a temperature gradient. HRM primers for each variant tested are listed in Table 14.

The LCGreen PLUS dsDNA intercalating fluorescent dye only fluoresces when bound to dsDNA. Thus, using the LightScanner (Idaho Technology, USA), the fluorescence of each sample was measured as each PCR sample was gradually heated over a temperature gradient of 70°C to 95°C with the fluorescence intensity falling off rapidly with dsDNA denaturation (Figure 16). Analysis of these different melting curve profiles on the LightScanner allowed identification of any samples containing variant alleles. These variant calls were confirmed by conventional DNA sequencing, using the above PCR products directly in the cycle sequencing reaction following Exonuclease – Shrimp Alkaline Phosphatase PCR purification.

Table 14: HRM primer sequences and PCR conditions for each variant tested.

Gene	HG19 reference position/ dbSNP rs number	Variant base change	Forward Primer	Reverse Primer	Amplicon Size (bp)	Annealing Temp (°C)	DMSO
MUTYH	Chr1: 45795058	T>A	GAGTCGGGGAAAGGGAGAGA	GTCACTGGGCTGCACTGTTG	222	60	+
MUTYH	Chr1: 45805788	C>T	GCCTGAGTCGTCTGTGGGTA	GGCCAGCACAGGCCAATAGGCAAT	184	60	+
MUTYH	rs1140199	G>A	GGTCCGGAGTTCAAAACC	AGGCTGAAGTGCAAAATGG	192	60	-
MUTYH	rs3219495	G>A	CGTGTGTATCAGGGCCAACA	TCTACCCTGCACCCCACAAT	202	57	+
MUTYH	rs3219497	C>T	GAGTCGGGGAAAGGGAGAGA	GTCACTGGGCTGCACTGTTG	222	60	+
XPC	Chr3: 14186418	C>T	GGGAAACAGTCCCAGCTCCT	TCCAGCCTCCATGAAAACAAA	161	57	-
XPC	Chr3: 14186922	A>C	TTCATCTGTCCGACAAGTTCACTC	TCCCAGCAGATGACCTGTACTTC	156	57	-
XPC	Chr3: 14187269	T>G	AAGGCAAACTGAGGCAGCAT	GGTGAGTGGGCTTTGGTAGC	150	56	+
XPC	Chr3: 14189405	T>G	TCCCCACAGGACTGATGGATAC	AGTGTTGCTTCCAGCTTCTG	152	58	-
XPC	Chr3: 14190068	A>C	CTGCCCAGCATGATGCCTAT	TTTCCATCCCCATCTCTGGA	206	55	+
XPC	Chr3: 14190169	T>G	CTGCCCAGCATGATGCCTAT	TTTCCATCCCCATCTCTGGA	206	55	+
XPC	Chr3: 14190324	T>G	TCCTCTGGTGCAGATGGTGA	CTGGTCCTGAGCCCTTCTGA	180	57	+
XPC	Chr3: 14199728	T>G	AGCGATGGTGAGAAGGCAGA	TCACTGTCAATGCCCACCAC	182	60	-
XPC	Chr3: 14199759	T>G	AGCGATGGTGAGAAGGCAGA	TCACTGTCAATGCCCACCAC	182	60	-
XPC	Chr3: 14200112	T>G	TGCCAAAGGGAAGAGGAACA	CCAGCCTCATCACTCCCACT	180	60	-
XPC	Chr3: 14219968	del AGGG	ATCCAAGGCCAAGAGCAAGG	CGCAGCAACCTCCACCAG	151	58	+
XPC	Chr3: 14220193	C>T	CTACGTCGTCCGCCATGTTT	CCCTTCGTTGGAGGCCTAGT	211	58	-
XPC	rs3731069	G>T	TCAGAGGGCTGGAACCATGT	TGTTCCCTGGGCTCATAATCA	153	60	-
XPC	rs3731072	G>C	CTGCCTCCTAAGAAAGTTGA	TGTGCCCTCTACTGGACTAC	150	55	-

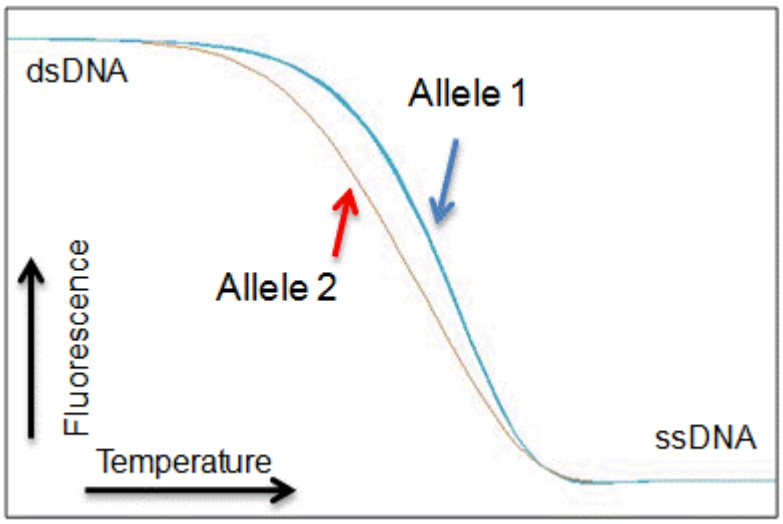


Figure 16: Temperature-shifted, fluorescence-normalised melting curve (adapted from (Taylor 2009). At lower temperatures, the fluorescence intensity remains high with the fluorescent dye bound to the dsDNA molecules. As the temperature increases, the dsDNA molecules denature to ssDNA releasing the fluorescent dye with a rapid fall in the fluorescence measured until complete denaturation of all dsDNA molecules. The different sequence of Allele 1 compared to Allele 2 results in the different melting curve profile observed.

7.4.9. Conventional DNA sequencing

A standard dye terminator chemistry protocol was used. In brief, forward and reverse primers were designed using Primer3 as previously described. Primers used are detailed in Appendix A. PCR amplification was carried out as described earlier in a 10 µl reaction mix of HotStarTaq master mix (Qiagen, UK) and 5 ng of genomic DNA. Successful PCR was confirmed by gel electrophoresis. PCR products were purified using the Exonuclease - Shrimp Alkaline Phosphatase (Exo-SAP) protocol. The cycle sequencing reaction was performed using standard BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, UK) protocols. A 10 µl reaction mix of 4 µl molecular grade water, 1 µl of purified PCR product, 0.5 µl of BigDye® Terminator v1.1 Ready Reaction Mix, 3.5 µl of 5x Sequencing buffer and 1 µl of either the forward or reverse PCR primer (1.6 µm). Two cycle sequencing reactions were performed for each sample PCR product: one with the forward and one with the reverse primer. Thermocycling conditions were: initial denaturation: 96°C for one

minute and 25 cycles of [denature: 96°C for 10 seconds; anneal: 55°C for five seconds; extend: 60°C for four minutes].

Sequencing extension products were purified using a standard ethanol precipitation protocol. DNA was precipitated by incubating each product sample with 1 µl of 3M, pH5.3 sodium acetate and 25 µl of 95% ethanol for 30 minutes at room temperature. Samples were then centrifuged at 2250g for 30 minutes followed by centrifuging inverted on tissue paper at 180g for one minute. The samples were then washed with 70 µl of 70% ethanol and centrifuged at 1650g at 4°C for 15 minutes, followed by centrifuging inverted on tissue paper at 180g for one minute. The invisible DNA pellet was then dried at 95°C for one minute.

The pellet was resuspended in 15 µl of Hi-Di formamide (Life Technologies, UK) and heated at 95°C for one minute then cooled on ice. Samples were loaded on a 3130xl Genetic Analyser (Applied Biosystems, UK) for capillary electrophoresis electrophoresis and sequence data collection. Data analysis was by visual inspection of generated electropherograms using Sequencing Analysis Software version 5.1 (Applied Biosystems, UK).

7.4.10. Next-generation sequencing

7.4.10.1. Long PCR product normalisation

The concentrations of each purified amplicon were measured using the Quant-IT BR PicoGreen Kit (Life Technologies, UK) as described earlier. For each individual, all their amplicons were normalised based on the relative size of each amplicon to produce one microgramme of the amplified candidate gene/genes containing equimolar quantities of each amplicon. Thus, the volume needed of each amplicon for each sample was calculated using the following formula:

$$\text{Volume needed of Amplicon X} = \frac{\left[\frac{\text{Size of Amplicon X (bp)}}{\text{Total Size of All Amplicons (bp)}} \times \text{Total DNA needed (ng)}\right]}{\text{Concentration of Amplicon X (ng/µl)}}$$

For each sample, all the amplicons were pooled and made up to 50 µl with 10 mM Tris-Cl (pH 8.5) buffer. For the pilot study sequencing *XPC* and *MUTYH*, the normalised samples were pooled immediately into pools of 20 case or 20 control samples prior to standard Illumina sequencing library preparation. For the indexed sequencing of *MRE11A* and *XPC*, each sample underwent library preparation individually.

7.4.10.2. Illumina sequencing library preparation and indexing

Illumina sequencing libraries were prepared using the standard LIMM Next Generation Sequencing Facility protocol. Samples were made up to 250 µl with molecular grade TE buffer, transferred to a shearing tube then fragmented to 150-200 bp using adaptive focused acoustics on a Covaris S2 Sonicator (KBiosciences, UK). Sheared samples were purified with fragment size selection using the MinElute PCR Purification Kit (Qiagen, UK) – the protocol was as per column purification above but instead a MinElute column (instead of a Qiaquick column) was used and the DNA was eluted in 10 µl (instead of 30 µl) of 10 mM Tris-Cl (ph8.5) buffer. Successful fragmentation was confirmed by running 1 µl of each purified sample on an Agilent 2100 Bioanalyzer (Agilent Technologies, USA) using the DNA 1000 Kit (Agilent Technologies, USA) for sizing and quantification of DNA fragments.

DNA fragment ends were then repaired using the End-It DNA End Repair kit (EpiCentre, USA). At room temperature, 9 µl of each sample was incubated with 5 µl

10x End-Repair buffer, 5 µl dNTP mix, 5 µl ATP, 1 µl End-Repair Enzyme mix and made up to a 50 µl reaction with molecular grade water, for 45 minutes. Each sample was then purified using the QIAquick PCR Purification Kit (Qiagen, UK) as described earlier but instead eluted in 34.45 µl of 10 mM Tris-Cl (ph8.5) buffer. A 3'-deoxyadenosine was then added by incubating each sample at 37°C for 30 minutes with a master mix containing 5 µl DNA Polymerase 10x buffer (Promega, USA), 10 µl 1 mM dATP and 0.55 µl of 9 Unit/ µl DNA Polymerase I Large (Klenow) Fragment Exonuclease Minus (Promega, USA) for each sample. Samples were then purified again using the MinElute PCR Purification Kit (Qiagen, UK) and eluted in 10 µl of 10 mM Tris-Cl (pH 8.5) buffer.

Either standard Illumina sequencing adaptor oligonucleotides for unindexed NGS runs or adaptors with a unique six nucleotide bar-code attached for indexed NGS runs, were then ligated to the DNA fragments for each sample using the LigaFast Rapid DNA Ligation System kit (Promega, USA). Ten microlitres of each sample was mixed with 15 µl 2x Rapid Ligation Buffer, 0.5 µl standard or indexed adaptors, 3 µl T4 DNA Ligase and 2 µl molecular grade water and incubated at room temperature for 15 minutes then 65°C for 20 minutes to inactivate the enzyme. The sequencing adaptor oligonucleotides and six nucleotide bar-codes are shown in Appendix A. Each bar-code has a minimum of two bases difference between indexes to reduce the risk of sequencing error resulting in mis-assignment of DNA fragment sequences.

Adaptor-ligated samples were then purified and size selected using the Agencourt AMPure XP magnetic bead system (Beckman Coulter, USA). The ligation reaction mix was mixed thoroughly with 27 µl of AMPure XP and incubated for five minutes at room temperature to allow DNA binding to the magnetic beads. The beads were then separated from solution by placing samples on a magnetic separator rack (Invitrogen, UK) for 10 minutes until the solution clears. While still on the magnetic rack, the

solution was discarded and the beads washed twice with 200 µl 70% ethanol. All the ethanol was aspirated and the beads dried at 37°C until all excess ethanol has evaporated. The beads were resuspended in 40 µl 10 mM Tris-Cl (pH 8.5) buffer to elute the DNA then separated from solution by placing back on a magnetic rack for 10 minutes. The eluted DNA solution was then transferred to a new microcentrifuge tube.

NGS libraries were then generated by amplifying the eluted DNA by PCR in a 50 µl reaction containing 2 µl purified DNA, 25 µl Fermentas Phusion High Fidelity Master Mix (ThermoFisher Scientific, UK) and 1 µl each of the Illumina library amplification forward and reverse primers (CAAGCAGAAGACGGCATACGAGATCGGTCTCGGC ATTCCTGCTGAACCGCTCTTCCGATCT and AATGATACGGCGACCACCGAGAT CTACACTCTTTCCCTACACGACGCTCTTCCGATCT respectively). Thermocycling conditions were: initial denaturation: 96°C for one minute, 12 cycles of [denature: 98°C for 30 seconds; anneal: 65°C for 30 seconds; extend: 72°C for 30 seconds], and final extension: 72°C for five minutes. PCR products were purified again using the Agencourt AMPure XP magnetic bead system (Beckman Coulter, USA) but instead using 90 µl of AMPure XP. Finally, the library quality was checked on an Agilent 2100 Bioanalyzer (Agilent Technologies, USA) using the DNA 1000 Kit (Agilent Technologies, USA) and library concentration measured using the Invitrogen Quant-IT BR PicoGreen Kit (Life Technologies, UK). For indexed NGS, equal quantities of each indexed library were pooled in groups of 20 to 25 samples per pool for a total of 1 µg of DNA library.

7.4.10.3. Illumina sequencing

Libraries were sent for sequencing at the LIMM Next-Generation Sequencing facility. Flow cell cluster generation and sequencing by synthesis using the Illumina GA II was performed following standard Illumina protocols with single end reads. The pilot multiplexed NGS study of *XPC* and *MUTYH* used 36 bp reads while the indexed NGS of *MRE11A* and *XPC* used 100 bp reads.

7.4.11. *MRE11A* mRNA isoform expression profiling

For RNA extraction, macrodissection was undertaken on five 10 µm FFPE tissue sections from each corresponding patient tumour blocks for regions of muscle-invasive bladder cancer. Tissue from the muscle-invasive tumour regions were placed in a microcentrifuge tube, deparaffinised in one millilitre of Xylene, vortexed, then centrifuged at 13200 rpm for two minutes. The supernatant was discarded and the tissue pellet washed with one millilitre of 100% ethanol, vortexed again, and centrifuged at 13200 rpm for two minutes. The supernatant was discarded and the pellet dried at 37°C for 10 minutes.

Tumour RNA was extracted using the AllPrep DNA/RNA FFPE Kit (Qiagen, UK) following standard protocols. The RNA sample was then quantified using a NanoDrop 1000 spectrophotometer (ThermoScientific, UK). Reverse transcription of RNA to cDNA was performed following standard protocols using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, UK). A 10 µl master mix of 2 µl 10X RT Buffer, 0.8 µl 25X 100 mM dNTP Mix, 2 µl 10X RT Random Primers, 1 µl MultiScribe Reverse Transcriptase, 1 µl RNase Inhibitor and 3.2 µl molecular grade water was mixed with 2 µg of RNA made up to 10 µl with molecular grade water. Thermal cycling conditions were: 25°C for 10 minutes, 37°C for two hours, and 85°C for five minutes. A negative control sample for later steps underwent the same

process and thermal cycling but the master mix contained no reverse transcriptase. The reverse transcription product was then diluted 1:5 with molecular grade water.

Primers were designed using Primer3 for cDNA sequence flanking *MRE11A* rs1805363 with forward and reverse primers sited in common sequence for both *MRE11A* isoforms in exon 1 and exon 2 respectively. The PCR primers used were AAACGGACGCCGTTCTCT and GGGACCAGGTTCTTCTCCAA (139 bp product for isoform 1 and 246 bp product for isoform 2). PCR amplification was performed on 2 µl cDNA with Fermentas Phusion Flash (ThermoFisher Scientific, UK) PCR kit following standard thermocycling protocols as mentioned earlier except with a 65°C annealing temperature, 10 second extension time, and for 40 cycles. The *GAPDH* housekeeper gene was also amplified in parallel as a normalization control in a 10 µl reaction mix containing 2 µl cDNA, 5 µl 2x Taqman Universal PCR Master Mix (Applied Biosystems, UK) and 0.5 µl Taqman GAPDH Control Reagents (Human) (Applied Biosystems, UK). Thermocycling conditions were: enzyme activation 50°C for 2 minutes, denaturation 95°C for 10 minutes, 40 cycles of [denature: 95°C for 15 seconds; anneal and extend: 60°C for one minute]. Both PCR products were electrophoresed in parallel on a 2% agarose gel with a 1 kb ladder (New England Biolabs, USA) in a 100V electric potential for two hours. Bands were visualised using a gel documentation ultraviolet imager (Bio-Rad Laboratories, UK) and images taken. Images were analysed using Quantity One 1-D Analysis Software (Bio-Rad Laboratories, UK) to measure for each band the intensity per mm^2.

The bands with corresponding sizes for the PCR products of isoform 1 and isoform 2 were excised with a sterile scalpel and placed into separate microcentrifuge tubes for conventional Sanger sequencing to confirm the identity of each band. The PCR products were then extracted using the Qiaquick Gel Extraction Kit (Qiagen, UK) as per manufacturer's instructions. One microlitre of each purified sample was then

used direct in the cycle sequencing reaction as previously described for the conventional Sanger sequencing of each sample.

7.5. Bioinformatics

7.5.1. Next-generation sequencing

7.5.1.1. Illumina Genome Analyser Pipeline image analysis and base calling

The standard Illumina Pipeline software (Illumina, USA) was used for interpretation of signal intensities for each cluster captured images by the Illumina GA II and base calling of each cluster for each cycle. Cluster read sequences were ouput as a _prb.txt file (Pipeline software version 1.1) for the initial six NGS runs for the pilot multiplexed NGS study of *XPC* and *MUTYH*, then as a _qseq.txt file for later runs due to an update in the Pipeline software (Pipeline software version 1.3+).

7.5.1.2. Sequence alignment and variant calling

NextGene version 1.10 (SoftGenetics, USA) was used for sequence alignment and variant calling for the pilot multiplexed NGS of *XPC* and *MUTYH* study. Reads were aligned using the _prb.txt and _qseq.txt files against reference sequences *MUTYH* (NG_008189) and *XPC* (NT_022517) from build 18 of the Human Genome assembly. NextGene alignment and mutation calling settings were: Condensation method: Elongation, Coverage Index: 2500, and Mutation Filter: 2.0%.

Illuminator (written by Ian Carr, LIMM (Carr *et al.* 2011)) was used for index sorting, sequence alignment and variant calling for all the indexed multiplexed NGS studies. Index sorting was first performed using the Illuminator Data Extractor by inputting the _qseq.txt files and a list of the six nucleotide index sequences used in the corresponding pool. This generated a .fasta file for each sample index containing the corresponding sample index read data. Sequence alignment was performed for each

.fasta file against either the reference sequence human genome build 19 *XPC* (NG_011763.1) or *MRE11A* (NG_007261). The default heterozygous cut-off of 20% was used for mutation calling.

Analysis using Syzygy (http://www.broadinstitute.org/software/syzygy/) was undertaken by the University of Leeds Biomedical Health Research Centre Bioinformatics Technology Group. In short, a binary sequence alignment .bam data file was generated using SAMtools (http://samtools.sourceforge.net/) for each pool sequenced from the corresponding _prb.txt and _qseq.txt files of the pilot multiplexed NGS of *XPC* and *MUTYH* study. Sequence alignment was performed using these .bam files against the reference sequences *MUTYH* (NG_008189) and *XPC* (NT_022517) and mutation calling undertaken using default settings.

7.5.2. Bioinformatics functional prediction

7.5.2.1. Coding variants

Non-synonymous coding variants and amino acid changes were determined using the UCSC genome browser and these were analysed for potential functional effects using Polyphen (http://coot.embl.de/PolyPhen/) and SIFT (http://sift.jcvi.org/), changes to protein secondary structure with NPS@ (http://pbil.ibcp.fr/), lost of conservation in active sites using Pfam (http://pfam.sa nger.ac.uk/) as well as looking for significant changes in amino acid chemical properties. Reference protein sequences used have been detailed earlier (Section 7.4.2).

7.5.2.2. Predicting miRNA binding sites

For 5'UTR and 3'UTR variants identified by next-generation sequencing, variants were checked if they were sited at predicted miRNA binding sites using miRbase (http://www.mirbase.org/) and miRanda (http://www.microrna.org/microrna/) and if so, the ΔG was calculated using RNAhybrid (http://bibiserv.techfak.uni-

bielefeld.de/rnahybrid/) (Kruger and Rehmsmeier 2006). UTRscan (http://itbtools.ba.itb.cnr.it/utrscan) (Mignone *et al.* 2005) was used to predict if a new miRNA binding site may have been generated by the variant.

For the identification of DNA repair gene miRNA binding site SNPs, twenty DNA repair genes involved in NER, BER, NHEJ, HR and DSB signalling and end-processing pathways were selected for investigation. In July 2009, in collaboration with Dr S Landi's group (Landi *et al.* 2008), the 3′UTR genomic sequence for all genes were obtained from the UCSC genome browser (http://genome.ucsc.edu) and putative miRNA-binding sites within these regions were then predicted using the miRBase (Griffiths-Jones *et al.* 2006), miRanda (John *et al.* 2004), PicTar (Krek *et al.* 2005), MicroInspector (Rusinov *et al.* 2005), Diana-MicroT (Kiriakidou *et al.* 2004) and TargetScanS (Lewis *et al.* 2003) algorithms. Using the dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP), BLAST (http://www.ncbi.nlm.-nih.gov/BLAST/Blast.cgi) and BLAST-SNP (http://www.ncbi.nlm.nih.gov/SNP/ snpblastByChr.ht ml) algorithms, common SNPs (MAF>0.10) sited within these putative miRNA-binding sites were identified. Using miRanda, the Gibbs binding free energy (ΔG, expressed in KJ/mol) for both common and variant alleles of each SNP was calculated with $\Delta\Delta$G being the difference in ΔG of the two alleles. As some SNPs may be predicted to reside in more than one miRNA binding site increasing the likelihood of a true binding site existing, the sum of all $|\Delta\Delta G|$s for each SNP ($|\Delta\Delta G_{tot}|$) was used to grade the impact of the SNP on miRNA binding. The eight SNPs with the highest $|\Delta\Delta G_{tot}|$ not previously studied were selected for genotyping.

7.5.2.3. Intronic variants

Intronic variants located in canonical GT donor and AG acceptor splice sites were identified manually. NetGene2 (http://www.cbs.dtu.dk/services/NetGene2/) and

ESEFinder 3.0 (http://rulai.cshl.edu/tools/ESE3/) were used to predict potential generation of new splice sites. Using chromatin immunoprecipitation sequencing data from the UCSC genome browser (http://genome.ucsc.edu), consistent epigenetic regulatory regions across all cell lines, such as transcription factor binding sites, histone modifications, and predicted chromatin state segmentation regions, were identified.

7.6. Statistical analysis

7.6.1. Statistical software

Statistical analysis was performed using STATA version 10 (StataCorp, USA), SPSS version 18 (IBM, USA) and GraphPad Prism version 5 (GraphPad Software, USA). Haploview (Broad Institute, USA) was used for generating linkage disequilibrium plots and haplotype block analysis.

7.6.2. Case-control epidemiological studies

Power calculations were calculated for the relevant study populations. Hardy-Weinberg equilibrium was assessed for the variant genotype frequency in the control samples at a significance threshold of P = 0.05 for common SNPs. A Cochrane-Armitage trend test for common SNPs and Fisher's exact test for trend for rare variants was used to test for association of each variant with bladder cancer risk estimating the odds ratio (OR) for each genotype with the respective 95% confidence interval (95%CI). A multivariate logistic regression analysis was also performed for each variant adjusting for gender, age, smoking status (ever versus never) and occupational dye exposure (ever versus never). Trend tests were used as the genetic inheritance model was unknown and this test was deemed to provide the best power for detecting an association (Li *et al.* 2009). Gene-environment and gene-gene interactions were assessed using the multiplicative model of interaction with P-values less than 0.05 considered to be significant.

In the DNA repair gene miRNA binding site SNP study, SNPs showing a significant risk effect were incorporated into a secondary analysis for the total number of unfavourable genotype groups carried, adjusting for the same factors listed above.

The rare variant collapsed analysis was performed using a simple single-direction collapsed variant analyses (Barrett and Nsengimana 2011; Konig *et al.* 2011). This method involves for each gene, the number of copies of each minor allele for all variants with a MAF less than 0.05 being counted for each individual and each gene (according to the number of rare variants carried per individual) being tested for association with bladder cancer risk tested using a Bonferroni-corrected significance level (P = 0.05 / No of genes tested*)*.

7.6.3. Clinical outcome and survival analyses

Survival power calculations were calculated for the relevant studies. D' and r^2 values of linkage disequilibrium were calculated for all common SNPs. Kaplan Meier curves were plotted for cancer-specific survival (CSS; deaths due to bladder cancer only with other deaths censored) for the relevant categories of protein expression levels or germline variant genotype status. CSS was used as bladder cancer patients treated by radiotherapy tend to be elderly thus more likely to die of other causes (Kotwal *et al.* 2008). A log-rank statistic or a survival-adjusted trend test was used to compare CSS times across categories of protein expression levels or germline variant genotype status respectively. Cox Proportional Hazard models were used to estimate hazard ratios (HRs) and 95% confidence intervals and multivariate analysis adjusting for age, gender, tumour and nodal stage, tumour grade, hydronephrosis and neoadjuvant or concurrent chemotherapy or radiosensitiser use. Ordered logistic regression was used to test for genotype associations with late bladder and bowel radiation toxicity.

8. Results

8.1. Developing multiplexed next-generation sequencing for detecting rare variants

8.1.1. Unindexed multiplexed next-generation sequencing

In 2008, technical experience in multiplexing, the sequencing of multiple samples in a single pool, for NGS internationally was very limited while indexing or bar-coding technology had yet to be developed. This project, which started in October 2008, thus aimed to develop protocols for multiplexed NGS for the cost-effective targeted sequencing of multiple pooled samples for the identification of rare germline variants.

8.1.1.1. Candidate Gene Amplification

Long range PCR amplicons were successfully generated for the target candidate genes *MUTYH* (two amplicons) and *XPC* (nine amplicons) (total 33 kb) (see Methods chapter 7.4.4.2). The youngest 280 cases and 280 controls of the LBCS were selected to enrich for genetic influences, and were alliquoted in eight 96-well PCR plates comprising four case plates and four control plates. Each of the *MUTYH* and *XPC* amplicons were amplified individually for each sample. For quality control purposes, each amplicon of twelve samples from Row A of each PCR reaction plate were checked by gel electrophoresis; a PCR efficiency of greater than 95% was achieved in nine out of eleven amplicons (Figure 17).

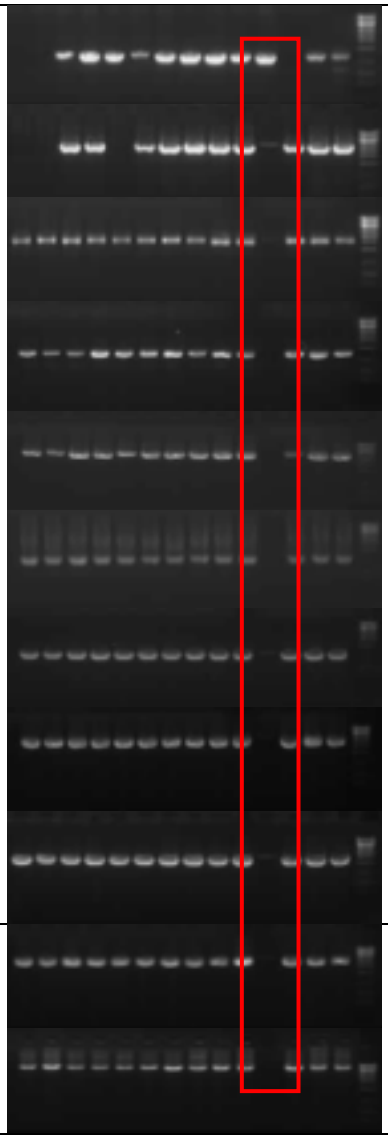
Gene	Amplicon	Exons	Gel Image of Row A Case Plate 3	PCR efficiency (%)
XPC	1	1		92
	2	2, 3		94
	3.1	4, 5.1		96
	3.2	5.2, 6		98
	4.1	7, 8		97
	4.2	9		98
	5	10		97
	6.1	11, 12, 13, 14		98
	6.2	15, 16		98
MUTYH	1	1,2,3,4		98
	2	5, 6, 7, 8, 9		98

Figure 17: PCR Efficiency of *MUTYH* and *XPC* amplicons. Sample gel electrophoresis image of Case plate 3 with persistent PCR failure for the same sample (red box) in all amplicons, except *XPC* amplicon 1, suggestive of variations in germline DNA quality or elution buffer conditions.

8.1.1.2. Optimum sample pooling

It was necessary to determine the optimum number of pooled samples per Illumina Genome Analyzer flow cell lane while maintaining sensitivity to detect a rare variant. Equimolar pooled amplicons from four previously conventionally-sequenced case samples of differing *XPC* heterozygous genotypes were thus NGS sequenced together at 1:1.2, 1:10, 1:20 and 1:50 dilutions respectively, as a pilot multiplexed

experiment. This pilot multiplexed NGS experiment was performed twice on two separate lanes on the same Illumina flow cell, either with PCR product samples purified prior to normalisation to generate equimolar pooled amplicons for each sample or with samples not purified prior to normalisation (Figure 18). This was done to establish whether PCR product purification was required prior to normalisation, as if not required, this would reduce the workload involved in individual sample processing. DNA concentration quantification using Picogreen® was performed prior to normalisation for each sample amplicon, and post-purification in the appropriate samples.

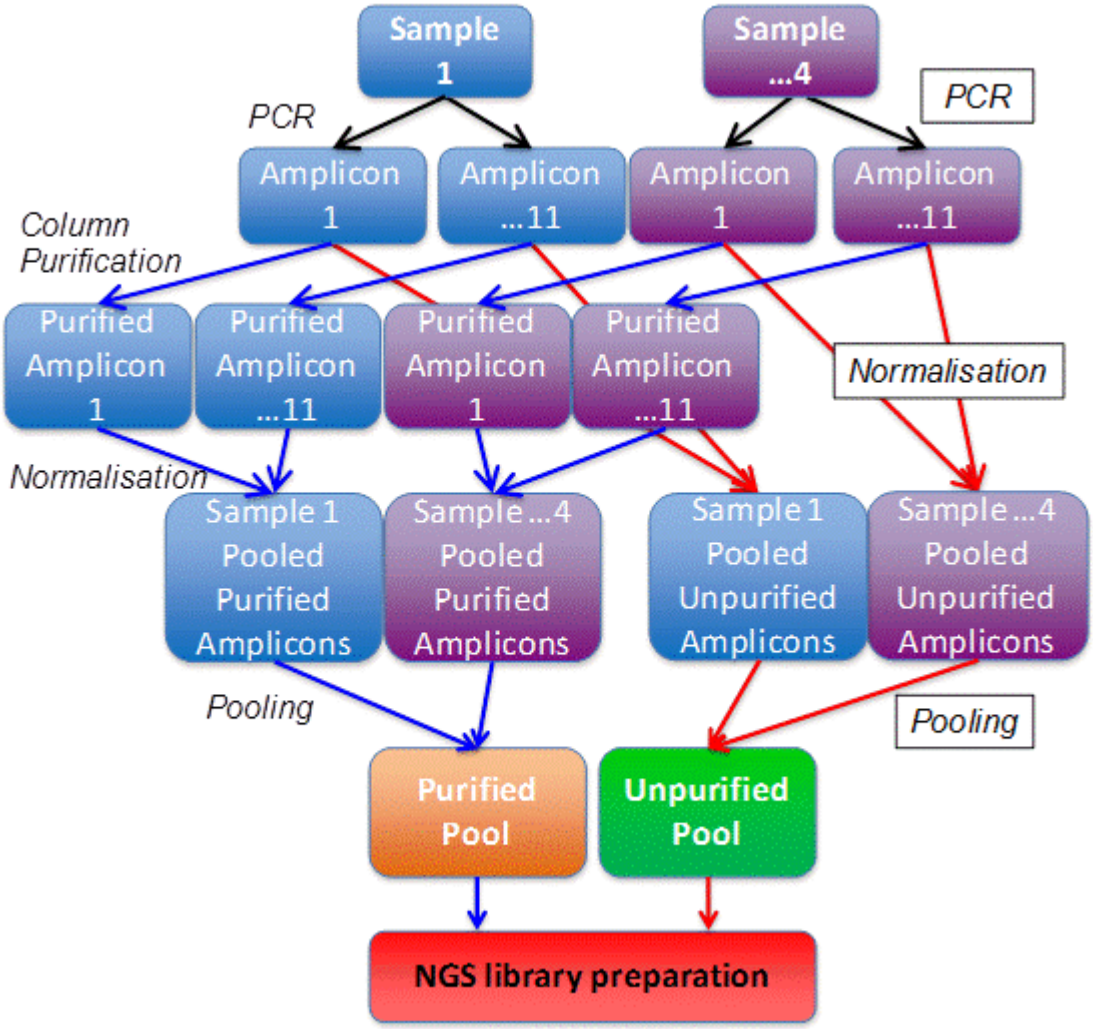


Figure 18: Flow chart of sample processing and pooling of purified (blue arrows) and unpurified (red arrows) PCR products for pilot multiplexed NGS.

From these two NGS experiments as shown in Figure 19, firstly, the sensitivity of detecting the respective variants by multiplexed NGS were equivalent if not better in the unpurified samples compared to the purified samples. Secondly, 1:20 was the highest dilution at which the variant allele was still detectable. Thus, for subsequent multiplexed NGS, 20 normalised samples were pooled per lane in 14 case and 14 control pools. For rare variants, in 20 samples, we would expect only one heterozygote sample or one variant allele (2.5%) of 40 alleles. Based on this and the pilot experiment above (Figure 19, Sample 3), a minimum threshold of 2% was set as a filter for the minimum percentage of reads containing a variant detectable by the SoftGenetics NextGene software.

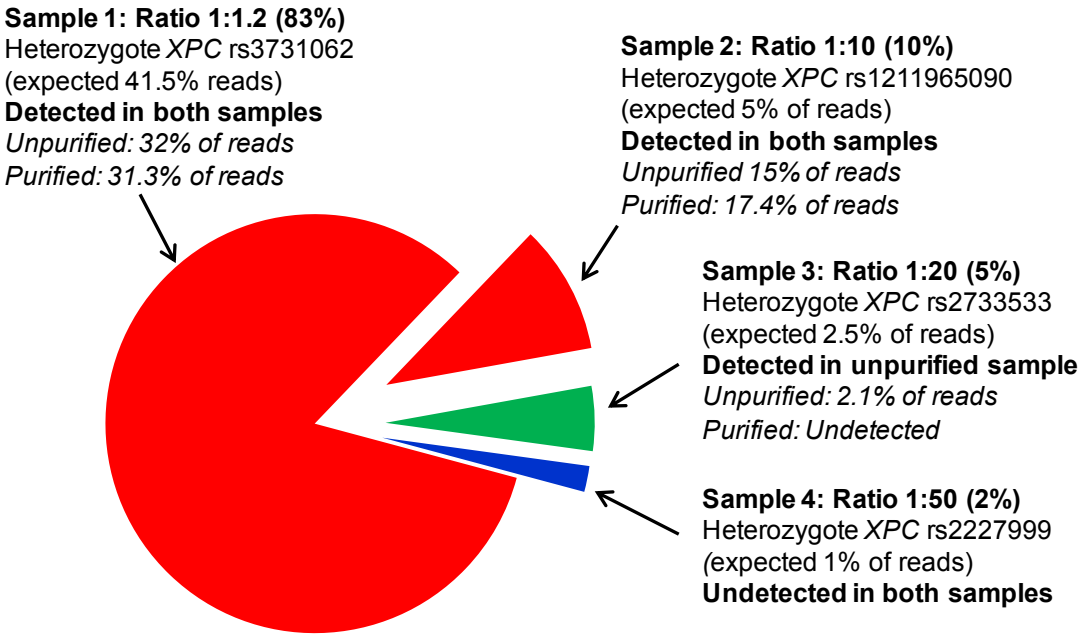
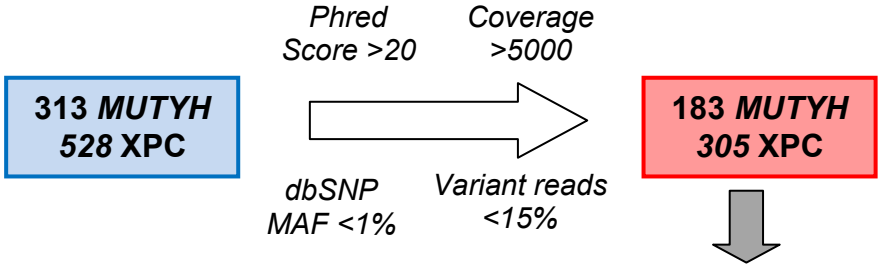


Figure 19: Relative proportions of each sample in the multiplexed NGS pool and the detection of previously identified variants (detection in both purified and unpurified samples (red), unpurified sample only (green) and undetected in both samples (blue)). The expected reads were the expected percentage of reads the SNP minor allele would be detected in taking into account the relative concentrations of that sample within the pool. The discrepancy in expected and actual read proportions for Sample 1 and Sample 2 was thought to be probably secondary to a systematic pipetting inaccuracy for Sample 2 during pooling resulting in a greater proportion of Sample 2.

8.1.1.3. Multiplexed NGS rare variant calling and candidate variant selection

All 14 case and 14 control pools were sequenced using single-end 36bp reads on the Illumina platform, this being the longest available read length at the time (now up to 150bp). Overall mean sequencing depth and base coverage per pool were 152x and 8969 reads respectively (*MUTYH*: 136x and 8629 reads, *XPC*: 163x and 9191 reads) with an overall mapped base count of 45 Gb per pool indicating good sequencing depth and coverage of the two genes. This compared favourably with the mapped base count seen in the 1000 Genomes Project for single end reads with the Illumina platform (12.8 Gb per sample) (The 1000 Genomes Project Consortium 2010).

A total of 313 *MUTYH* and 528 *XPC* variants were called following sequencing of all 28 pools. Further filtering criteria were applied: having a high base calling quality (Phred score greater than 20), good sequencing coverage (greater than 5000 reads), and likelihood of being a rare variant (seen in under 15% of reads within a pool or if reported in dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP), a reported MAF of under 1%) (Figure 20); with the aim of reducing the number of variants to a more practical number for bioinformatics analysis and candidate variant selection.



	MUTYH		XPC	
Gene Region	dbSNP	New	dbSNP	New
Coding	3	15	4	54
3'UTR	0	0	2	13
Intronic	8	157	29	203

Figure 20: Flow chart showing the filtering criteria and identified rare variants by gene region. The filtering criteria used to reduce the number of variants identified were good coverage, good base calling quality scores and those likely to be rare variants.

Filtered variants were then divided into coding, 3'UTR and intronic for bioinformatics functional analysis. Eighteen candidate variants, six intronic (3 *MUTYH*, 3 *XPC*), ten coding (4 *MUTYH*, 6 *XPC*) and two 3'UTR (both XPC), were selected for further investigation, based on their (i) being predicted to have a significant effect on protein function or alteration of miRNA binding energies, or (ii) being sited in a canonical splice site and (iii) the relative number of case to control pools in which the variant was detected.

8.1.1.4. Accuracy of rare variant calling

Preliminary high throughput Taqman genotyping was performed in 750 cases and 706 controls by Gen-Probe Life Sciences (West Lothian, Scotland). Unexpectedly, custom Taqman assays for all 11 *XPC* variants and one of the intronic *MUTYH* variants failed to discriminate. Therefore, high resolution melting (HRM) mutation scanning of individual samples from two NGS pools positive for each of these variants was performed to determine if this was due to custom Taqman assay failure or was due to false positive calls. Only the triallelic *XPC* rare variants rs3731072 G>C and G>T were found to be true variants (Figure 21) with confirmation by direct Sanger sequencing (Figure 22).

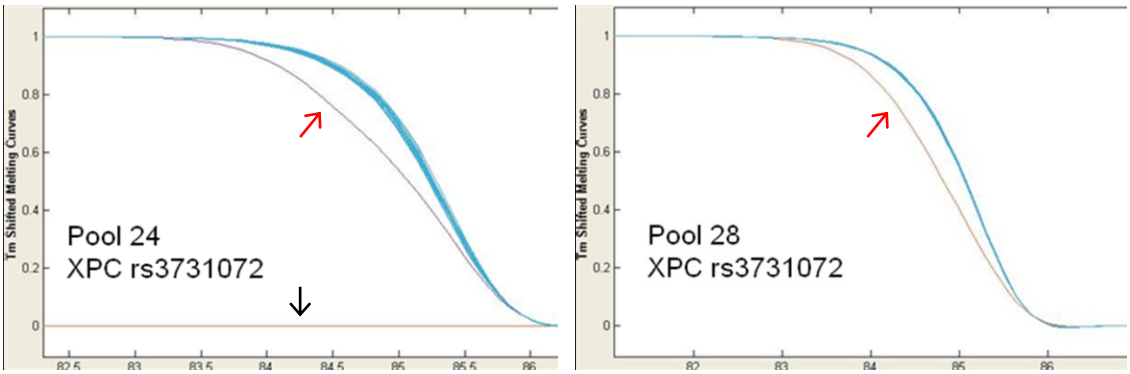


Figure 21: High resolution melting curves for *XPC* rs3731072 in Pool 24 and 28 demonstrating a different genotype in one sample of each pool (Red arrow: Pool 24 - purple and Pool 28 - brown line) compared with the other samples (blue lines). Brown line (black arrow) in Pool 24 represents a sample that failed to PCR amplify.

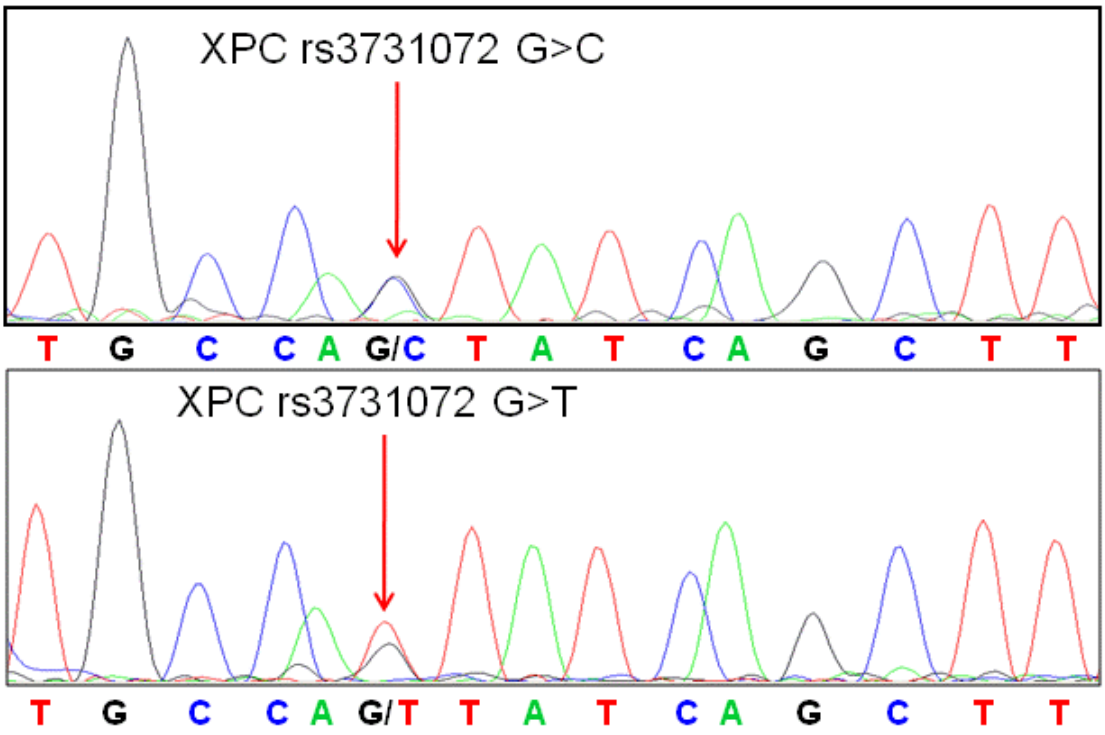
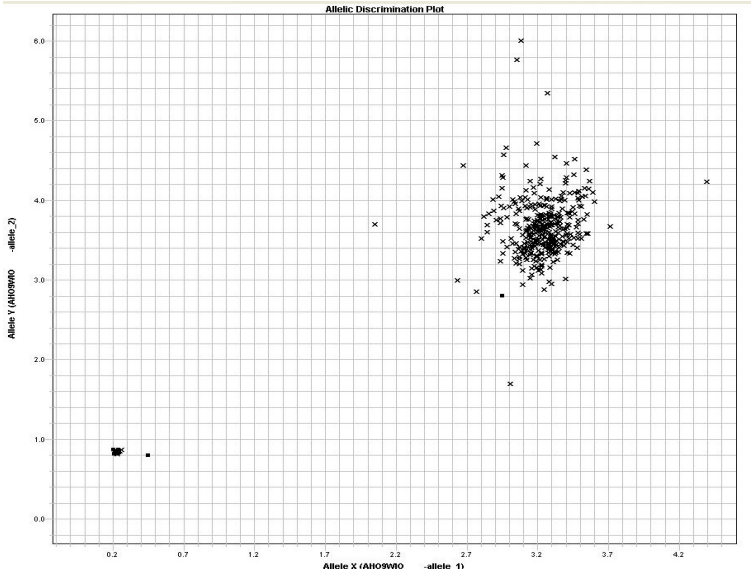


Figure 22: Conventional Sanger sequencing electropherogram confirming the triallelic *XPC* rare variants rs3731072 G>C and G>T.

Repeat genotyping using the same custom assay for rs3731072 G>C in several sequenced samples locally indicated that the failure to discriminate this rare allele in the original genotyping was due to technical failures with the genotyping itself rather than the assay (Figure 23).

a)



b)

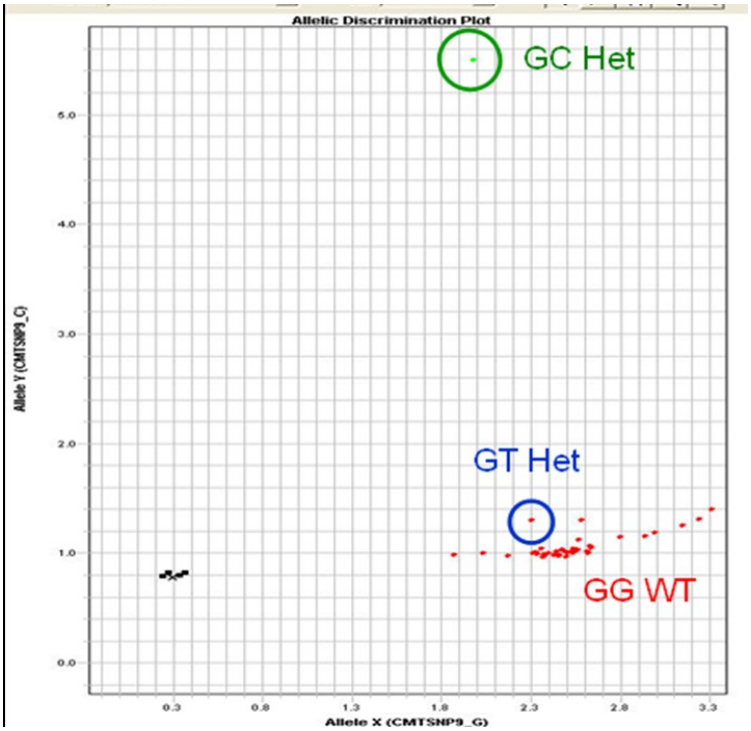


Figure 23: Allelic discrimination plots using custom Taqman assay AHGI9A3 for *XPC* variant rs3731072 G>C: a) Original genotyping in 750 cases and 706 controls showing a mono-cluster; b) Repeat genotyping in NGS pools 24 and 28 with the known sample heterozygous for the C-allele (green circle) clear discriminated and the T-allele (blue circle) heterozygote sample clustering with the wildtypes.

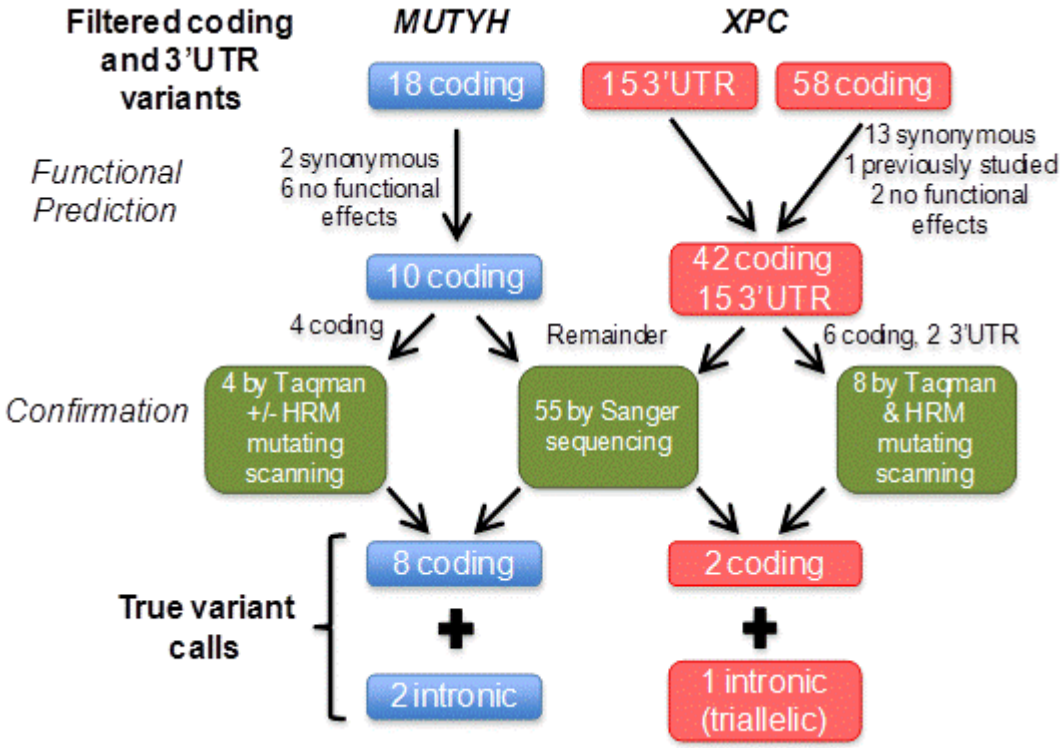


Figure 24: Flowchart of filtered NextGene coding and 3'UTR variants called that were selected for confirmation and the number of true variants identified.

All remaining 42 putatively functional coding variants (6 *MUTYH* and 36 *XPC*) and remaining 13 *XPC* 3'UTR variants called by NextGene underwent confirmatory conventional sequencing in their respective pools, confirming a further six coding variants (4 *MUTYH* and 2 *XPC*) only (Figure 24). The sensitivity of multiplexed NGS to detect the presence of a true rare variant allele was estimated by using the successful rare variant Taqman genotyping results as the gold standard test and the following formula:

$$\text{Sensitivity} = \frac{\text{No of true positive calls}}{(\text{No of true positive calls} + \text{No of false negative calls})}$$

A true positive call was hence defined as the detection of a rare variant in a multiplexed NGS pool containing a rare variant heterozygote / homozygote individual genotyped on Taqman, while a false negative call was the failure of detecting a rare variant in a multiplexed NGS pool containing a rare variant heterozygote /

homozygote individual genotyped on Taqman. The number of true positive and false positive calls was thus summed up for all variants successfully genotyped by Taqman at several different variant calling thresholds and the sensitivity calculated accordingly (Table 15). At a 2% variant calling threshold, there was a sensitivity of 73% for detecting a true rare variant with the sensitivity rising or falling as the variant calling threshold is increased or decreased. The true negative rate (thus specificity) could not be calculated without resequencing all samples by conventional sequencing.

Table 15: The sensitivity of multiplexed NGS to detect the presence of a true rare variant allele relative to Taqman genotyping results and the effect of varying the NGS variant calling thresholds.

	NGS variant calling threshold		
	1%	*2%*	*2.5%*
Rare variant allele detected	24	19	15
Rare variant allele undetected	2	7	11
Sensitivity	**92%**	**73%**	**58%**

However, on examining only the potentially functional coding variants selected for validation, this sensitivity is tempered by a high false positive rate (Table 16) of 80.8% at the 2% variant calling threshold only decreasing to 54.5% at the 2.5% variant calling threshold. As discussed later, these results actually compared favourably with the only contemporary published report at that time using unindexed multiplexed NGS that attempted to validate their variants, with that study having all false positive calls (Harakalova *et al.* 2011). However, assuming an equivalent false positive rate among the intronic variant calls, identifying further candidate intronic variants for further investigation would prove to be very labour intensive and non-cost effective due to the large number of variants and hence samples needed to be tested.

Table 16: False positive and true positive rates of potentially functional coding variants called by multiplexed NGS successfully validated.

	NGS variant calling threshold	
	2%	*2.5%*
Total coding variants called	52	22
No of True +ve coding variants	10	10
No of False +ve coding variants	42	12
False positive rate (%)	**80.8**	**54.5**
True positive rate (%)	**19.2**	**45.5**

8.1.2. Indexed multiplexed next-generation sequencing

After this initial work was undertaken on unindexed multiplexed NGS, NGS technology had advanced significantly, with improvements in reagents and amplification accuracy, higher flow cell adaptor density, read lengths and the development of paired-end sequencing for greater coverage. Indexing/ bar-coding technology was also developed, which involves the labelling of each individual multiplexed sample with a unique 6 nucleotide "bar-code" tag (Craig *et al.* 2008; Kenny *et al.* 2010). This allows the identification of the sample source of each read fragment during NGS and the analysis of each sample separately, thus enabling the use of a higher variant calling threshold for greater accuracy and accurate sample genotype determination.

8.1.2.1. Pilot indexed multiplexed next-generation sequencing of XPC and MUTYH

In order to assess the impact of these advances, *XPC* and *MUTYH* were resequenced in the samples from one pool, using single-end 90-base reads with indexing of all 20 samples. Using a default 30% variant calling threshold for each sample and filtering for low-base calling quality or coverage, the two confirmed *MUTYH* variants previously detected were detected again as well as five novel

variants (four intronic and one 3'UTR). These were, in fact, called in the original unindexed pool but were not selected for further investigation as they either had no predicted functional effects or were seen in most pools (hence, were not rare). All of these five variants were successfully confirmed by conventional sequencing.

Only one of the false positive intronic *MUTYH* "variant" rs1140199 A allele from the untagged NGS experiment was detected though in a small percentage of reads in six of the tagged samples (Table 17). Rs1140199 is sited at the end of a short homopolymer stretch of four A bases which may account for the false positive call. Without indexing, this low level of sequencing or alignment error would have reached the variant calling threshold but with indexing, was easily identified and excluded.

Table 17: Variant calling of the false positive *MUTYH* rs1140199 variant alleles in the original unindexed pool compared to the indexed pool samples.

		Original Pool	Indexed Samples					
			1	2	3	4	5	6
***MUTYH* rs1140199**	% variant reads	2.24	3.0	6.1	6.4	6.6	10.7	21.6
	Coverage	5653	1191	2458	882	897	979	527

This newer, improved method was thus used in subsequent NGS projects. As there were sufficient rare variant hits in *MUTYH*, only *XPC* was chosen for resequencing in a smaller subset of the bladder cancer case-control for rare variants especially in view of the low number of true positive rare variants identified compared to previous work done by the group (Qiao *et al.* 2011). *MRE11A* was also sequenced using indexed multiplexed NGS to look for genetic markers of radiotherapy response in a set of germline DNA from MIBC patients treated with radiotherapy (see Results chapter 8.4).

8.1.2.2. Target Amplification of MRE11A and XPC

Both *MRE11A* and *XPC* genes were amplified by long range PCR, *MRE11A* in 12 amplicons (total 28.2 kb) in 186 RT-treated MIBC cases, and *XPC*, using newer long range PCR kits, in just four amplicons (total 26.9 kb) in the youngest 100 bladder cancer cases and 100 controls. All PCR amplicons were purified, quantified and normalised to generate equimolar pooled amplicons for each sample.

8.1.2.3. Summary of indexed multiplexed NGS of MRE11A and XPC

Separate indexed NGS libraries were prepared for each sample prior to multiplexing in equimolar quantities of 20 to 24 indexed libraries per pool for *MRE11A* and 25 indexed libraries per pool for *XPC*. Each pool was single-end sequenced for 90 cycles (ie. 90 base read length) on a single flow-cell lane of an Illumina GAII. Illuminator, an in-house developed software (Carr *et al.* 2011), was used to sort indexed reads according to sample origin and sequence alignment of the reads from each sample against the reference wildtype genomic sequence. For *MRE11A*, a total of nine pools were sequenced with a median coverage of 2507x per sample (range 673 – 9891) and 95% of target candidate gene regions were sequenced at greater than 221x coverage per sample. *XPC* was resequenced in eight pools with a median coverage of 1719x per sample (range 727 – 3579) and 95% of target candidate gene regions sequenced to 201x coverage.

8.1.2.4. Accuracy of rare variant calling

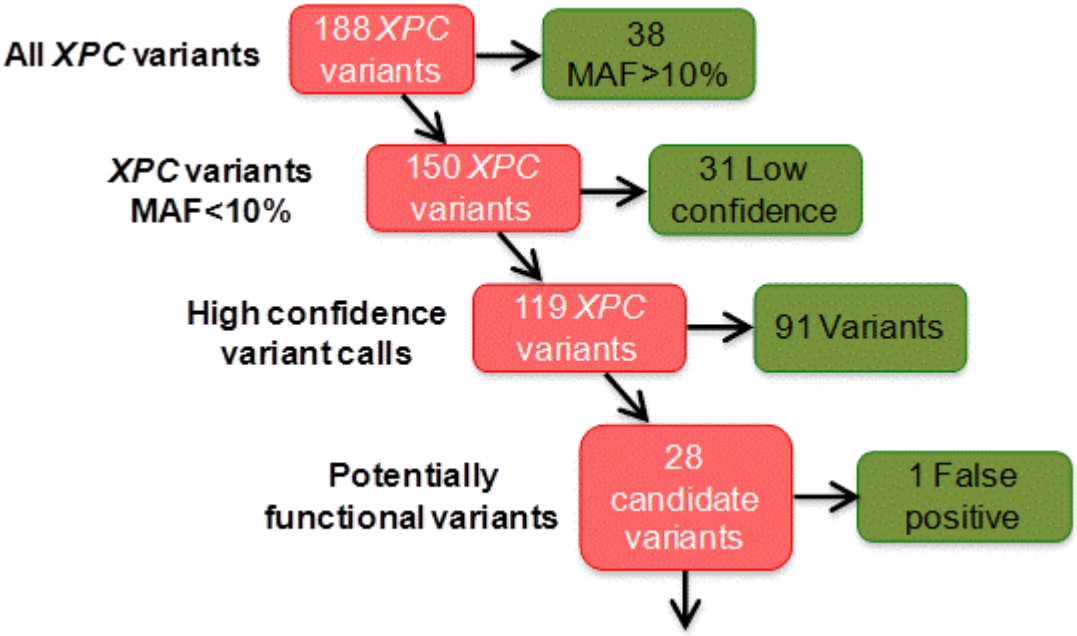
Illuminator was also used for variant calling, with a default 30% variant calling threshold and a minimum coverage of 200x. Poor confidence and high-confidence variant calls were identified based on visual inspection of sequence alignment and read coverage data. A total of 121 *MRE11A* variants were called, 85 variant calls were deemed high confidence variants and 36 poor confidence. Confirmatory

sequencing of all 121 variants successfully validated all 85 high confidence calls and only 3 of the 36 poor confidence calls (Table 18).

Table 18: Summary of confirmed *MRE11A* variants identified by MAF, presence in the dbSNP database and gene location.

	dbSNP	**New**
***MRE11A* Variant type (MAF)**	No of variants	No of variants
SNP (>0.01)	40	3
Intronic	*37*	*3*
3'UTR	*3*	*0*
Rare Variant (≤0.01)	14	31
Intronic	*10*	*25*
Coding	*3*	*1*
3'UTR	*1*	*5*

On resequencing *XPC*, a total of 188 variants were called, of which 150 variants had an overall MAF less than 10% in the 100 bladder cancer cases and 100 controls (Figure 25). 119 of these were high confidence calls and 31 low confidence calls. Twenty eight of the high confidence variants were then selected for confirmatory sequencing of the respective indexed sample, specifically all 3'UTR, 5'UTR and coding variants, and variants within 200bp of a splice site, with 27 being successfully validated. The false positive "variant" was sited within a SINE repeat region which could suggest an issue with sequence alignment.



	dbSNP	New
***XPC* Variant type**	No of variants	No of variants
5'UTR	0	1
Intronic	10	5
Coding	2	7
3'UTR	0	2

Figure 25: Flow chart of candidate *XPC* variant selection and the number of Sanger sequencing confirmed true variants by gene region.The 91 high confidence *XPC* variants excluded were intronic variants with minimal predicted functional effects so were not selected for further confirmation (see Appendix C for full details of all variants).

Based on the high confidence calls only, indexed multiplexed NGS had a true positive rare variant call rate of 98.6% for all *MRE11A* and *XPC* rare variants selected for confirmatory Sanger sequencing, with only one of 70 rare variants (42 *MRE11A* and 28 *XPC* rare variants) being a false positive call. This marked improvement has resulted in these methods being instigated as the standard multiplexing protocol for candidate gene resequencing at the Leeds Institute of Molecular Medicine's (LIMM) NGS facility.

8.1.3. Unindexed multiplexed next-generation sequencing revisited

In 2011, Rivas *et al* at the Broad Institute, USA, investigating GWAS loci in inflammatory bowel disease, published the first successful unindexed multiplexed NGS for disease susceptibility rare variants (Rivas *et al.* 2011). Using a custom-designed variant discovery software, Syzygy, which uses various error-modelling algorithms to account for sequencing errors and uneven variant read frequency, they identified 429 high-confidence variant calls, of which 137 were selected for validation, with 91.2% of these being true positive calls. Despite some differences in their protocols compared to this project (germline DNA pooling prior vs post PCR amplification respectively, 50 samples vs 20 samples per pool, 107.5 kb vs 33 kb target size and single-end 76bp vs single-end 36bp read length), the NGS variant calling issues of potentially uneven PCR amplification between individual samples and achieving adequate coverage taking into account the sequencing depth, target size and pooling size differences were relatively similar. Therefore, the original data from the unindexed multiplexed NGS of *MUTYH* and *XPC* was reanalysed using Syzygy.

8.1.3.1. Syzygy software and accuracy

The Syzygy software was installed by the LIMM Bioinformatics Group and the data from all 14 bladder cancer case pools and 14 control pools analysed. Syzygy called 398 variants and estimated the MAF for each variant based on the variant allele read frequency in case pools and in control pools. Using this MAF estimation, Syzygy performs a weighted association likelihood ratio test for each individual variant and generates a Chi^2-statistic which could be a potential tool for prioritising variants for further investigation.

Of note, three *MUTYH* rare variants found to be significantly associated with bladder cancer susceptibility in the preliminary Taqman genotyping, as well as the common *XPC* SNP rs2228000 previously associated with bladder cancer risk (Stern *et al.* 2009), were all observed to have a Chi^2-statistic value greater than two (Appendix C). Thus, a filter of a Chi^2-statistic value of greater than two was used to try aid candidate variant selection resulting in 15 *MUTYH* and 25 *XPC* variants (Figure 26), of which 11 had an estimated population MAF greater than 5%, while four *MUTYH* rare variants had previously been validated in the original work. Of the remaining 25 "rare" variants, four *MUTYH* and three *XPC* variants were selected, based on predicted functional effects and the Chi^2-statistic value, for validation by HRM mutation scanning and Sanger sequencing of the corresponding sample pools – all were true positive. These seven variants were subsequently selected for genotyping in the larger case-control study (see Results chapter 8.2).

On cross-referencing all Syzygy variant calls with all previously validated true positive variant calls and known false positive variant calls, Syzygy had a predicted true positive rare variant call rate of 80.6% and a sensitivity of 89.3% of detecting a true positive rare variant.

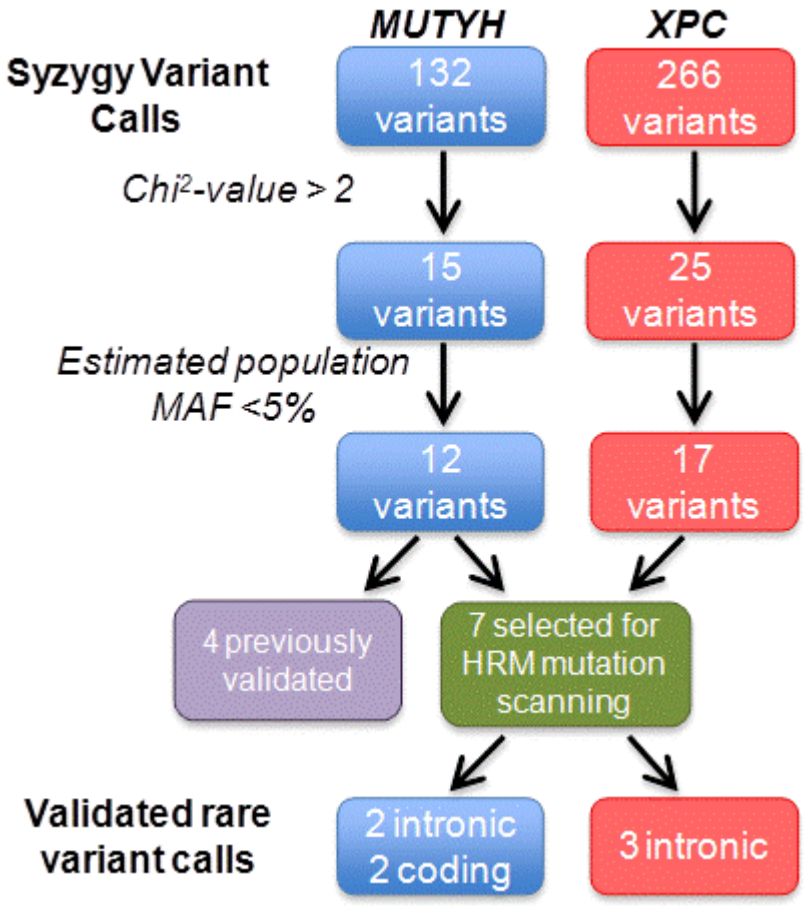


Figure 26: Flowchart of number of Syzygy variants called, and number of variants at each stage of filtering and selection for confirmation. Filtering criteria were Syzygy Chi^2-statistic value and estimated population MAF.

8.1.4. Discussion

This project was successful in developing two methods for the identification of rare variants in candidate genes using multiplexed NGS at the LIMM. Despite initial high false positive call rates with unindexed multiplexed NGS, the use of new custom-multiplexed NGS software markedly improved the true positive call rate and the sensitivity of detecting rare variants. This project also demonstrated the higher accuracy of rare variant detection using indexing technology and current NGS reagents in multiplexed NGS.

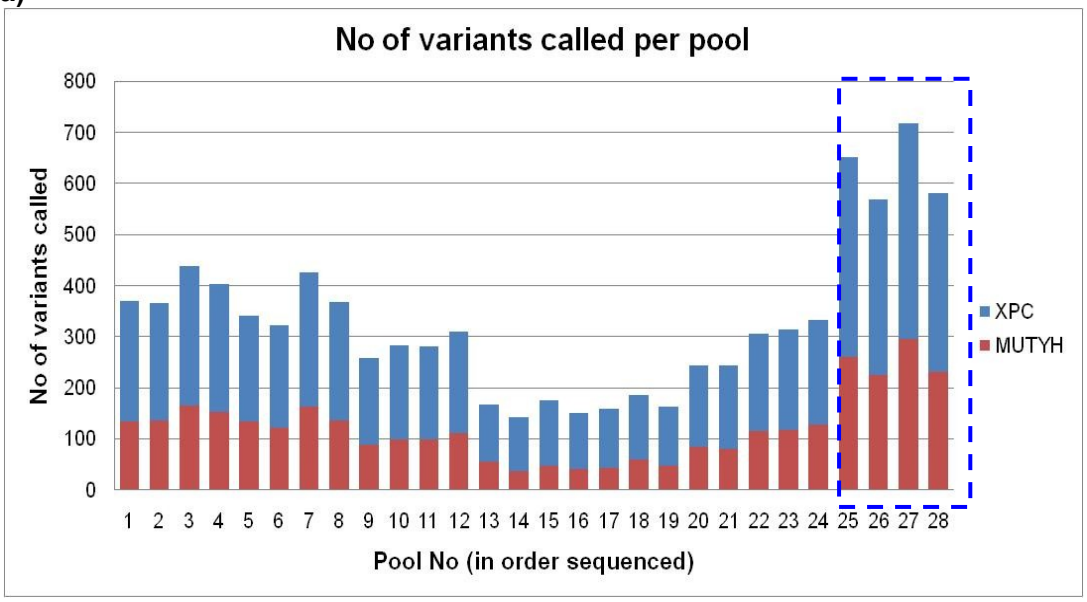
8.1.4.1. Unindexed multiplexed next-generation sequencing

In the original analysis using the NextGene software in 2009, despite achieving high sequencing depths and coverage, a significant proportion of candidate variants selected were false positive calls, with the majority in *XPC* (Table 16). In the 1000 Genomes Project undertaking whole genome sequencing of individual samples using contemporary NGS protocols and kits as this project, heterozygote call accuracy for SNPs with MAF 2-5% ranged from 83.08% to 99.97% for sequencing depths of 1x to 16x respectively with overall base call error rates of 1-3% (The 1000 Genomes Project Consortium 2010). Therefore, with multiplexing at a sequencing depth of 152x per pool, a sequencing depth of about 8x per sample would be expected assuming even amplification correlating with a call accuracy of 99.56%. However, the lower variant calling threshold of 2% used would encroach on the Illumina platform's average base call error rates of 1-1.5%, thus searching for variants with rarer MAFs likely resulted in the lower accuracy and higher false positive rates seen (Shendure and Ji 2008). This could also explain the relatively higher number of "variants" detected in this study (25.5 variants per kb sequenced in 560 individuals) compared to the 1000 Genomes Project (6.2 variants per kb in 179 individuals).

There was a relatively low number of true *XPC* rare variants identified compared to *MUTYH* (3 variants vs 10 variants) in the original analysis, which was surprising as in our group's work, conventional sequencing of only 33 case samples identified four novel rare variants (Qiao *et al.* 2011). There are several possible explanations for this. As individual amplicon PCR products were not purified prior to normalisation, unwanted oligonucleotides could distort the quantification of amplicon concentrations thus affecting accurate normalisation. As the *XPC* amplicons were also relatively smaller (1.5 – 3 kb) compared to *MUTYH* (3 – 6 kb), small volumes (less than 0.5 µL) were needed for some amplicons during normalisation. Any pipetting errors could thus result in significantly less DNA normalised and therefore the potential for any erroneous sequences to make up a relatively higher proportion of total DNA.

Over the course of sequencing all the pools, there was a drift in the number of variants called in the later pools sequenced, with a predominance of variant T>G and A>C transversions (Figure 27). These T>G and A>C transversions corresponded to the majority of the false positive variants. This could be due to known changes in Illumina NGS kit chemistry or from automatic NextGene software updates (Version 1.10 in Pools 1 to 19 and Version 1.12 in Pools 20 to 28) over this period. On comparing the read coverage of variants called between Pools 1 to 19 and Pools 20 to 28, there was a lower mean variant coverage in the latter pools (15182 reads vs 11294 reads, $p<0.001$). There was no significant difference between Pools 1 to 10 and Pools 11 to 20 (14805 reads vs 15848 reads, $p=0.20$). This difference in coverage could be due to problems with cluster amplification in the later pools or from altered sequence alignment with the software update. Unfortunately, it was not possible to perform a direct comparison between Illumina kit batches or the two NextGene versions as earlier kits and software versions were no longer available.

a)



b)

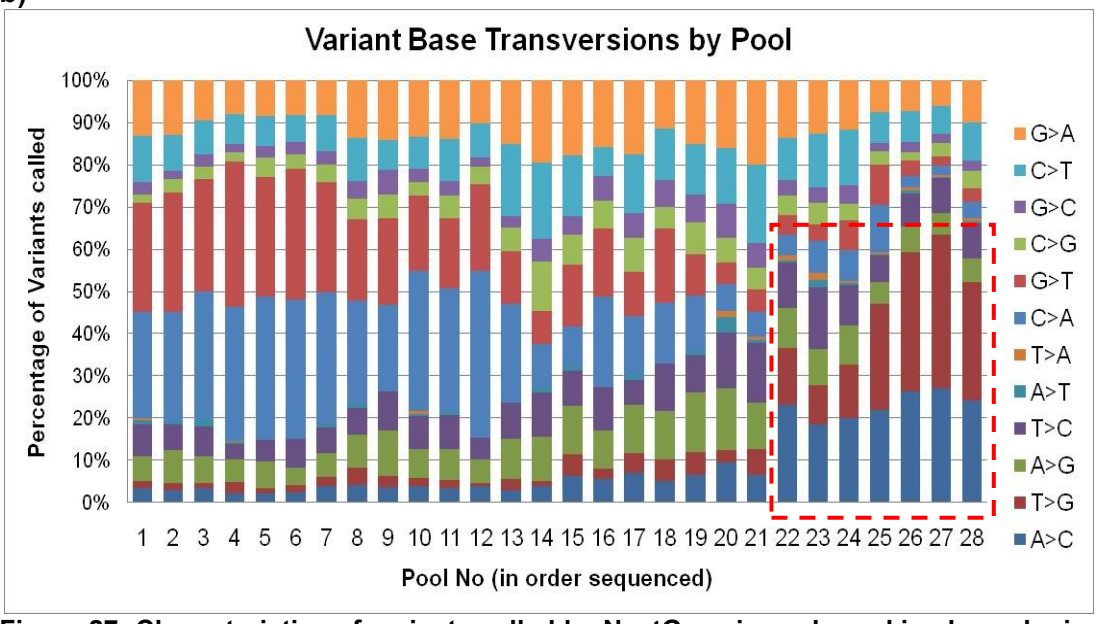


Figure 27: Characteristics of variants called by NextGene in each pool in chronological order of sequencing: a) Number of variants called in each pool, b) Percentage of each variant base transversion in each pool. Differences seen in later sequenced pools are highlighted in the blue and red squares.

Despite these issues, the number of true rare variants originally detected compared favourably with other published unindexed multiplexing studies at that time. Harakalova *et al* used a similar methodology with individual sample amplification prior to sample pooling in pools of 20 samples (Harakalova *et al.* 2011). This study sequenced 45 candidate genes (total 500 kb) in 100 patients but found that all twenty candidate variants selected were false positive. Out *et al* sequenced by NGS the

MUTYH gene in 88 previously Sanger-sequenced samples as a single pool, detecting two out of the five known rare variants, but they did not investigate for false positive calls (Out *et al.* 2009). This study pooled 287 genomic samples prior to target amplification but no rare variants were detected. Lupton *et al* also pooled samples prior to target amplification of the Nicastrin gene (3 kb) in 311 cases in one case pool and 360 controls in one control pool, but from the variants identified only investigated two known coding rare variants (Lupton *et al.* 2011).

The high false positive rate of rare variant calls observed by Harakalova *et al* and in the original NextGene analysis suggested that unindexed multiplexed NGS was not a promising strategy for rare variant detection. However, the success of Rivas *et al* and their development of the Syzygy software as mentioned earlier (Rivas *et al.* 2011), has led to a reassessment of this view. Particularly surprising was the marked superiority of its variant-calling algorithm compared to the original analysis (true positive rate 80.6% vs 45.5% at best respectively), even when using the same NGS raw sequencing data. By incorporating error-modelling and testing of consistency between forward and reverse strands, the Syzygy software was successful in excluding low-quality base calls and thus identify predominantly true-variant calls.

Recently, Kelleher III *et al* used a novel multiplexed NGS design with two different NGS platforms (Illumina and Helicos) to cross-validate variant calls to identify rare variants in 18 autism related genes (total 40 kb) (Kelleher *et al.* 2012). Of note, their Illumina pooling strategy was the same as this project (20 samples per pool) with similar false positive rates of greater than 50% (Figure 28). However, by combining both platforms for cross-validation, they were able to reduce the false positive rate below 10%. Thus, with the development of new custom software and alternative NGS study design approaches, unindexed multiplexed NGS remains a feasible option for rare variant detection.

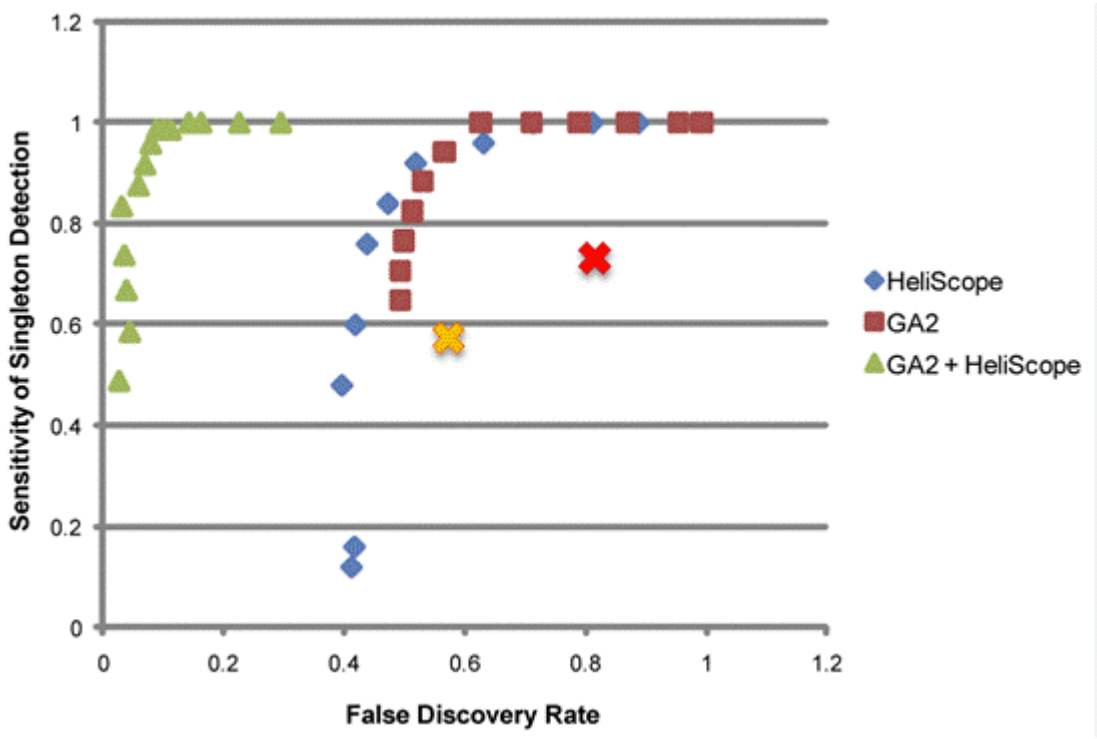


Figure 28: Receiver operator curve for sensitivity and false discovery rate of detecting a variant with unindexed multiplexed NGS as reported by Kelleher *et al* (Kelleher *et al.* 2012). Red squares are results with the Illumina GA2 platform, blue diamonds with the Helicos Heliscope platform and green triangles with a combination of both platforms. Crosses are the equivalent results on the Illumina GA2 platform reported in this study at the 2% (red cross) and 2.5% (orange cross) calling thresholds.

8.1.4.2. Indexed multiplexed next-generation sequencing

The use of oligonucleotide bar-codes for multiplexed NGS was first demonstrated by Craig *et al* (2008), reporting at best an 88.7% true positive rate and a 90.8% false negative rate on multiplexing 46 indexed PCR-enriched samples from previously genotyped individuals (Craig *et al.* 2008). In the last two years, there have been a flurry of indexed NGS studies validating this methodology on different NGS platforms in previously Sanger sequenced or genotyped samples, demonstrating its use in clinical genetic diagnostics and in the investigation of candidate disease-related gene variants (Kenny *et al.* 2010; Morgan *et al.* 2010; Nijman *et al.* 2010; Amstutz *et al.* 2011; Lotta *et al.* 2012; Pritchard *et al.* 2012; Rossetti *et al.* 2012). Despite using different target-enrichment techniques, target sizes, NGS platforms, sequencing coverage and numbers of multiplexed samples, results similar to those shown here

were reported in all these publications, with all studies having greater than 93% true positive rates. Most studies reported sensitivities of over 90% for detecting a true single nucleotide variant.

Two papers have reported the potential use of indexed multiplexed NGS in clinical genetic diagnostics (Morgan *et al.* 2010; Pritchard *et al.* 2012). Pritchard *et al* demonstrated the use of new target enrichment capture arrays for sequencing of all colorectal cancer associated genes for pathogenic mutations and variants of unknown significance (VUS), by using indexed multiplexing of 96 samples on the latest Illumina HiSeq NGS platform. A sensitivity, specificity and accuracy of 99.4% was seen on confirmation by Sanger sequencing(Pritchard *et al.* 2012). Locally, diagnostic testing using indexed multiplexed NGS for *BRCA1* and *BRCA2* mutations in breast cancer cases, and *TP53* mutations in Li-Fraumeni cases is now routine with a 100% sensitivity and zero false positive calls reported (Morgan *et al.* 2010).

8.1.4.3. Unindexed versus indexed multiplexed NGS

It is thus feasible to identify rare variants using both unindexed and indexed multiplexed NGS. As indexed multiplexed NGS was found to have close to a 100% true positive rate with the added benefit of providing information on sample genotype, this would seem to be the better NGS option of the two. However, indexing techniques require the labour intensive preparation of individual NGS libraries for each sample prior to pooling (ie. twenty libraries for twenty samples per pool) whereas for non-indexed libraries only one library is needed per pool (ie. one library for twenty samples per pool). Unindexed NGS, though, requires confirmation of a whole pool of samples for each rare variant detected rather than just two samples (one variant carrier and one wildtype) for indexed NGS. Fortunately, this is relatively

easy and cost-effective, with minimal added workload using HRM mutation screening as an initial sample screening step prior to Sanger sequencing as shown above.

Thus, the choice of unindexed versus indexed multiplexing will be dependent on the number of samples to be sequenced and the need for genotyping data. Unindexed multiplexing would be more suitable for case-control and rare variant studies where large numbers of samples require to be sequenced to maximise the likelihood of detecting disease susceptibility variants and where the individual sample genotype is not required, as this would be determined in a second stage by large scale genotyping. In contrast, indexed multiplexing would be more useful for small to medium sized cohort studies, such as biomarker studies as shown in Results section 8.4, where the individual sample genotype is needed for correlation with patient-specific outcome measures, and where analysis of all variants detected rather than just selected candidate variants is desirable.

8.1.4.4. Newer Technologies and Study Limitations

For detection of rare variants, the larger the sample size, the larger the number of rare variants that are likely to be detected up to the number of rare variants present in the sample set. The high workload of manual library preparation for indexing could thus potentially prove prohibitive for some research projects. Recently, several commercial automated library preparation kits and systems (Illumina Nextera, Agilent Bravo, Bechman Coulter SPRIworks and Fluidigm Access Array) have been developed, which are capable of automating the whole process or elements of the library preparation process for 10 to 96 samples per run. However, this reduction in hands-on time comes with a consumables cost at least ten-times higher than the manual process. Potentially, as costs fall, high-throughput automation could become widely available making large multiplexing NGS projects more accessible.

NGS technology and platforms continue to develop and improve rapidly with greater sequencing depth, more accurate software and newer chemistry. One could predict that these newer technologies would improve NGS accuracy, increase multiplexing or target sizes, and reduce costs and sequencing time. However, NGS remains a relatively new approach with much developmental work still in progress. Two studies have undertaken cross-NGS platform comparisons of the Applied Biosystems SOLiD, the Roche 454 GS-FLX and the Illumina GA2 platforms for standard single sample NGS (Harismendy *et al.* 2009; Suzuki *et al.* 2011). These revealed false positive single nucleotide variation call rates of approximately 3%, 7% and 6-16% for each of the platforms respectively, and false negative call rates of approximately 3%, 1% and 0% respectively. These calling errors have been attributed to similar systematic errors, typical of the short-read sequencing used by all three platforms and seen at regions of repetitive sequences, simple repeats, insertion-deletions within 30bp on the variant and homopolymer stretches (Harismendy *et al.* 2009). It thus remains unclear if the greater coverage of advancing NGS technology would be able to resolve these issues.

New high-throughput target-enrichment methods have also recently been developed using PCR-based assays, molecular inversion probes and hybrid-capture arrays with some of these systems capable of incorporating library preparation during the target-enrichment process though this is usually at the cost of reduced target size (Mamanova *et al.* 2010). These different methods differ in their optimum cost-efficacy (Figure 29) with, for example, PCR assays being better suited for sequencing of a small target size in a large number of samples, while hybrid-capture arrays are more suitable for larger targets. The availability of these high-throughput methods could potentially allow megabase sequencing for rare variants in large sample sets.

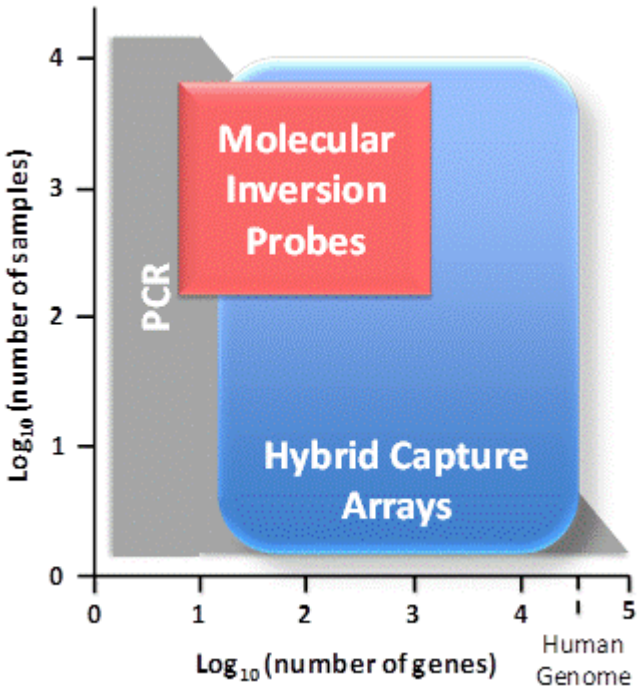


Figure 29: Optimum target enrichment strategy relative to number of samples sequenced and target size (adapted from (Mamanova *et al.* 2010).

One of the limitations of this study is the inability to accurately determine the false negative call rate which would require resequencing of the whole target gene by conventional Sanger sequencing as a "gold" standard. Potentially pathogenic or clinically significant variants could thus be missed. A difficulty also now recognised with NGS is the sheer bulk of data generated which requires handling of the large number of variants called (Ding *et al.* 2010). For epidemiology studies, the selection of candidate variants for further investigation and second stage genotyping remains very subjective, particularly due to the limited validation of bioinformatics functional prediction tools and correlation with actual clinical impact. Syzygy's Chi^2-value prediction provides a potential objective candidate variant selection tool. Finally, the rapid technological developments in NGS has made keeping pace difficult. The "novel" developments in this study, such as multiplexed sequencing and target enrichment by long PCR for NGS, are now commonplace or outmoded with the development of indexing technology and array-based enrichment. It is likely that these technologies too will be replaced, especially with the continued development of

new “third” generation sequencing platforms such as the Life Technologies Ion Torrent semiconductor-based system and the previously mentioned single-molecule sequencing Helicos Heliscope (Pareek *et al.* 2011).

8.1.4.5. Summary

This study has successfully developed local protocols for unindexed and indexed multiplexed NGS and demonstrated its efficacy in rare variant discovery. Both approaches have advantages and disadvantages, in terms of cost, accuracy and manual workload. The rapid advances in NGS technology, with higher read lengths and coverage, and newer automated or high-throughput systems for library preparation and target enrichment, will improve the cost-efficacy and accuracy of both approaches. This will allow even larger multiplexing projects to be conducted and potentially increase the success of identifying rare variants of clinical significance.

8.2. *MUTYH* and *XPC* rare variants in bladder cancer susceptibility

Using the *MUTYH* and *XPC* rare variants identified by multiplexed NGS, this study aimed to test the common disease – rare variant hypothesis by genotyping candidate RVs in a large bladder cancer case-control study population and investigate the associations between individual or multiple RVs and bladder cancer risk. The top SNP hits reported from the three bladder cancer GWAS (Kiemeney *et al.* 2008; Wu *et al.* 2009; Kiemeney *et al.* 2010; Rothman *et al.* 2010; Garcia-Closas *et al.* 2011; Rafnar *et al.* 2011; Tang *et al.* 2012) were also genotyped to determine if risk contributions of candidate RVs were independent of these SNPs.

8.2.1. Bladder cancer case-control study population

Germline DNA samples from 853 cases and 773 controls for the LBCS population were collected between 2002 and 2011 at St James University Hospital as described in the methods (Methods chapter 7.2.2). A further 1232 controls from the Wellcome Trust Case-Control Consortium (WTCC) (The Wellcome Trust Case Control Consortium 2007) were also included to try boost statistical power. This resulted in a final study population of 853 cases and 2005 controls. The statistical power of this study is shown in Table 19 for the LBCS study population on its own (for multivariate analyses) and for the whole combined LBCS and WTCC population (for univariate analyses) at several ORs and at several different MAFs (either single variant or collapsed) at a 5% significance level.

Table 19: Statistical power of LBCS and combined LBCS and WTCC populations at different ORs and MAFs at a 5% significance level.

P = 0.05	**LBCS (853 cases : 773 controls)**			**LBCS + WTCC (853 cases : 2005 controls)**		
Single variant/ Collapsed MAF	*OR 1.5*	*OR 2.0*	*OR 3.0*	*OR 1.5*	*OR 2.0*	*OR 3.0*
0.001[a]	3.7	3.3	9.6	3.7	8.8	25.2
0.01	15.2	46.4	92.3	26.2	69.0	98.7
0.05	64.7	99.1	100.0	83.1	99.9	100.0
0.1	89.2	100.0	100.0	97.5	100.0	100.0

[a] Statistical power estimations at small MAFs may be inaccurate.

The study population demographics are shown in Table 20. The control population were significantly younger (P < 0.001) with a higher proportion of women (P < 0.001) especially as there were no attempts at frequency matching for the WTCC population. Between cases and controls for the LBCS population on its own, there were no significant differences between the groups for age (Mean 72.1yr v 72.1yr respectively, P = 0.96) but there remained significantly more women among the controls compared to cases (37.1% v 29.4%, P = 0.001). Among the demographic variables, only smoking history was independently associated with an increased bladder cancer risk (OR 1.86, 95%CI 1.46 – 2.37, P <0.001) on multivariate analysis.

Table 20: Combined LBCS and WTCC case-control study population demographics and tumour characteristics.

Demographic variables		Case N (%)	Control N (%)	OR (95% CI)[b]	P-value[b]
Age	Mean	72.1	53.4		**<0.001**[c]
	(Range)	(21 – 99)	(15 – 98)		
Gender	Female	251 (29.4)	928 (46.4)		**<0.001**[d]
	Male	602 (70.6)	1073 (53.6)		
Ethnicity[a]	Non-caucasian	15 (1.8)	18 (2.3)	1	
	Caucasian	838 (98.2)	755 (97.7)	1.52 (0.75 – 3.09)	0.25
Smoking history[a]	Never	185 (21.7)	281 (36.4)	1	
	Ever	668 (78.3)	492 (63.6)	1.86 (1.46 – 2.37)	**<0.001**
Family history[a]	None	809 (95.3)	587 (96.9)	1	
	>1 family member	40 (4.7)	19 (3.1)	1.49 (0.85 – 2.62)	0.16
Number of occupational exposures[a]	0	623 (73.5)	606 (80.0)	1	
	1	183 (21.6)	126 (16.6)		
	2	31 (3.7)	22 (2.9)		
	3	8 (0.9)	3(0.4)	1.11 (0.93 – 1.33)	0.24
	4	1 (0.1)	1 (0.1)		
	6	2 (0.2)	0 (0)		
Bladder tumour characteristics[a]					
Tumour stage	Ta	360 (42.2)			
	CIS	15 (1.8)			
	T1	200 (23.4)			
	T2	134 (15.7)			
	T3	69 (8.1)			
	T4	13 (1.5)			
	Tx	62 (7.3)			
Pathological grade	G1	76 (8.9)			
	G2	324 (38.1)			
	G3	412 (48.4)			
	Unrecorded	39 (4.6)			
Nodal stage	N0	203 (23.8)			
	N1	17 (2.0)			
	N2	2 (0.2)			
	N3	1 (0.1)			
	Nx	630 (73.9)			

NB. Missing LBCS data for gender in 4 patients and occupational history in 20 patients.

[a] Data available only for the LBCS population.

[b] Multivariate logistic regression adjusting for all demographic variables.

[c] Two-tailed T-test.

[d] Pearson's Chi-squared test.

8.2.2. Candidate variants for genotyping

8.2.2.1. MUTYH and XPC rare variants identified by NGS

In summary from all the unindexed and indexed NGS of *MUTYH* and *XPC*, a total of 14 *MUTYH* and 34 *XPC* RVs were identified and confirmed by conventional sequencing or Taqman genotyping as follows from the different NGS methods development stage as described in the last chapter: i) 10 *MUTYH* and 4 *XPC* RVs from the initial unindexed multiplexed NGS of 280 cases and 280 controls analysed using NextGene; ii) 27 *XPC* RVs from the indexed multiplexed NGS of 100 cases and 100 controls for *XPC*; and iii) 4 *MUTYH* and 3 *XPC* RVs from the reanalysis of the unindexed multiplexed NGS of 280 cases and 280 controls using Syzygy.

Candidate RVs were then ranked and selected based on the number of the following criteria met: (i) their being predicted to have a significant effect on protein function or alteration of miRNA binding energies, (ii) their being sited in or within 100bp of a splice site, (iii) the relative number of case to control samples or pools in which the variant was detected, (iv) their having a Syzygy Chi^2-value greater than two (only for RVs identified from the Syzygy analysis as discussed in the last chapter), and (v) being able to design a suitable Fluidigm genotyping assay. All the *MUTYH* (four intronic and 10 coding) and 24 *XPC* (10 intronic, 11 coding, two 3'UTR and one 5'UTR) RVs were selected for genotyping in the full bladder case-control study and these are shown in Table 21 with the bioinformatics functional predictions. The full list of all variants identified by NGS is in Appendix C.

Table 21: *MUTYH* and *XPC* rare variants detected by NGS selected for genotyping and functional predictions.

Gene	Variant NCBI hg19 position/ dbSNP rs number	NGS run detected in	Case: Control (No of pools or No of samples or Estimated MAF)[a]	Gene Region	Syzygy Chi^2 value[b]	Coding			Predicted miRNA binding		Splicing/ Transcription factor binding site/ Epigenetics
						Protein Change	*Polyphen*	*SIFT*	*miRNA-binding site*	*ΔΔG (kJ/mol)*	
MUTYH	rs3219470 C>T	Unindexed multiplexed (NextGene)	3 : 0	Intron 1	2.31[b]	Non-coding					Predicted chromatin promoter region & H3K27Ac histone mark
MUTYH	Chr 1: 45803644 G>C	Unindexed multiplexed (NextGene)	1 : 4	Intron 1	2.39[b]	Non-coding					Possible H3K36Me3 histone mark
MUTYH	Chr 1: 45805788 G>A	Unindexed multiplexed (Syzygy)	0.006 : 0[a]	Intron 1 / 2	3.23	Non-coding					Exon 2 acceptor splice site in Isoform Beta 3
MUTYH	rs3219495 C>T	Unindexed multiplexed (Syzygy)	0 : 0.009 [a]	Intron 12	4.66	Non-coding					73bp 3' from exon 12 donor splice site
MUTYH	rs34612342 A>G	Unindexed multiplexed (NextGene)	2 : 0	Exon 7	2.16[b]	Y179C	Probably damaging	Affect protein function	Nil		Nil
MUTYH	rs149866955 G>A	Unindexed multiplexed (NextGene)	1 : 0	Exon 10	1.07[b]	R274Q	Possibly damaging	Affect protein function	Nil		Nil

MUTYH	rs138089183 C>T	Unindexed multiplexed (NextGene)	0 : 2	Exon 10	1.86[b]	R309C	Benign	Tolerated	Nil	Nil
MUTYH	rs149342980 G>A	Unindexed multiplexed (NextGene)	1 : 0	Exon 10	1.07[b]	R311K	Benign	Tolerated	Nil	Exon 10 GA donor splice site
MUTYH	Chr 1: 45797428 G>A	Unindexed multiplexed (NextGene)	0 : 1	Exon 12	1.86[b]	R364H	Possibly damaging	Affect protein function (low confidence)	Nil	Nil
MUTYH	rs36053993 G>A	Unindexed multiplexed (NextGene)	3 : 0	Exon 13	5.49[b]	G396D	Probably damaging	Affect protein function	Nil	Nil
MUTYH	rs140118273 C>T	Unindexed multiplexed (NextGene)	4 : 1	Exon 16	0.83[b]	S515F	Benign	Tolerated	Nil	Nil
MUTYH	Chr 1: 45795058 A>T	Unindexed multiplexed (Syzygy)	0.004 : 0 [a]	Exon 16	2.15	M511L	Benign	Tolerated	Nil	Nil
MUTYH	rs147923905 G>T	Unindexed multiplexed (NextGene)	1 : 0	Exon 16	1.07[b]	D530Y	Possibly damaging	Affect protein function	Nil	Nil
MUTYH	rs3219497 G>A	Unindexed multiplexed (Syzygy)	0 : 0.006 [a]	Exon 16	3.02	R534Q	Benign	Tolerated	Nil	Nil
XPC	rs3731072 G>T	Unindexed multiplexed (NextGene)	1 : 0	Intron 2	N/A	Non-coding				Predicted chromatin transcriptional elongation region

XPC	rs3731072 G>C	Unindexed multiplexed (NextGene)	2 : 1	Intron 2	0.41[b]	Non-coding	
XPC	Chr 3: 14220193 G>A	Unindexed multiplexed (Syzygy)	0.004 : 0 [a]	5'-promoter region	2.15	Non-coding	Predicted chromatin promoter region with multiple transcription factor binding sites; CpG methylation region
XPC	rs3731069 C>A	Unindexed multiplexed (Syzygy)	0 : 0.009 [a]	Intron 2	4.58	Non-coding	Predicted chromatin transcriptional elongation region
XPC	Chr 3: 14186418 G>A	Unindexed multiplexed (Syzygy)	0.01 : 0.002 [a]	3'-post-gene	2.99	Non-coding	Predicted chromatin enhancer region & H3K27Ac histone mark
XPC	rs147584831 C>G	Indexed multiplexed (*XPC* only)	0 : 4	Intron 1	N/A	Non-coding	Predicted chromatin promoter region with multiple transcription factor binding sites; CpG methylation region
XPC	rs77167750 A>G	Indexed multiplexed (XPC only)	1 : 3	Intron 10	3.2[b]	Non-coding	Nil
XPC	rs146126554 T>C	Indexed multiplexed (*XPC* only)	3 : 0	Intron 11	N/A	Non-coding	Nil
XPC	Chr 3: 14214734 G>A	Indexed multiplexed (*XPC* only)	1 : 0	Intron 1	N/A	Non-coding	Nil

XPC	rs3731078 A>G	Indexed multiplexed (*XPC* only)	4 : 8	Intron 2	N/A	Non-coding				H3K27Ac histone mark
XPC	Chr 3: 14187236 A>C	Indexed multiplexed (*XPC* only)	0 : 1	3'UTR	N/A	Non-coding			Nil	Nil
XPC	Chr 3: 14186649 A>G	Indexed multiplexed (*XPC* only)	0 : 1	3'UTR	N/A	Non-coding			Nil	3'-transcription end site
XPC	Chr 3: 14220118 T>G	Indexed multiplexed (*XPC* only)	1 : 0	5'UTR	N/A	Non-coding			Nil	Predicted chromatin promoter region with multiple transcription factor binding sites; CpG methylation region
XPC	Chr 3: 14220032 G>C	Indexed multiplexed (*XPC* only)	0 : 1	Exon 1	0.26[b]	G13R	Possibly damaging	Affect protein function	Nil	Nil
XPC	rs1870134 C>G	Indexed multiplexed (*XPC* only)	1 : 2	Exon 1	N/A	L16V	Benign	Affect protein function (low confidence)	Nil	Nil
XPC	Chr 3: 14214527 C>T	Indexed multiplexed (*XPC* only)	1 : 0	Exon 2	1.07[b]	L47F	Benign	Affect protein function (low confidence)	Nil	Nil
XPC	rs3731062 C>T	Indexed multiplexed (*XPC* only)	4 : 8	Exon 2	1.08[b]	L48F	Benign	Affect protein function (low confidence)	Nil	Nil
XPC	rs183478541 G>A	Indexed multiplexed (*XPC* only)	0 : 1	Exon 2	N/A	G64E	Possibly damaging	Affect protein function (low confidence)	Nil	Nil

XPC	Chr 3: 14206335 G>A	Indexed multiplexed (*XPC* only)	1 : 0	Exon 7	1.07[b]	R293Q	Possibly damaging	Tolerated	Nil	Nil
XPC	Chr 3: 14200154 A>G	Unindexed multiplexed (NextGene)	1 : 0	Exon 9	1.07[b]	E410G	Possibly damaging	Affect protein function	Nil	Nil
XPC	Chr 3: 14199980 G>C	Indexed multiplexed (*XPC* only)	0 : 1	Exon 9	0.92[b]	R468T	Possibly damaging	Affect protein function	hsa-miR-185-3p -6.5	Nil
XPC	Chr 3: 14199857 G>T	Indexed multiplexed (*XPC* only)	0 : 1	Exon 9	0.92[b]	R509I	Probably damaging	Tolerated	Nil	Nil
XPC	Chr 3: 14189426 T>C	Indexed multiplexed (*XPC* only)	1 : 0	Exon 14	0.28[b]	I832I	Synonymous		Nil	Nil
XPC	Chr 3: 14188798 G>A	Unindexed multiplexed (NextGene)	0 : 1	Exon 15	0.92[b]	G866R	Possibly damaging	Tolerated	Nil	Nil

[a] The ratio of detection in cases versus controls was dependent on which NGS methods development stage, as described in the last chapter, that the RV was detected in. The ratios are thus reported as follows: i) for unindexed multiplexed NGS (NextGene) detected RVs, the number of case pools versus control pools; ii) for indexed multiplexed NGS of *XPC* detected RVs, the number of case samples versus control samples; or iii) for unindexed multiplexed NGS (Syzygy) detected RVs, the estimated case MAF versus control MAF as calculated by Syzygy.

[b] For variants selected from the non-Syzygy analysed runs, the Syzygy Chi^2-value is displayed only out of interest if the variant was also detected by the Syzygy analysis.

8.2.2.2. MUTYH rare variants in bladder cancer risk

During initial preliminary Taqman genotyping of 750 cases and 706 controls in early 2010 as mentioned in the last chapter, three *MUTYH* RVs out of six *MUTYH* RVs genotyped looked suggestive of an influence on bladder cancer risk though only one (Chr 1: 45803644 G>C) reached statistical significance (Table 22). As *MUTYH* RVs had been previously associated with several different cancers (especially rs34612342 (Y179C) and rs36053993 (G396D) in colorectal cancer) (Vogt *et al.* 2009; Win *et al.* 2011; Castillejo *et al.* 2012; Rennert *et al.* 2012) but not investigated directly in bladder cancer, it was decided to also genotype all known *MUTYH* RVs that were potentially functional. Using the UCSC genome browser (accessed October 2011) with MAF details from the dbSNP database, there were 43 known variants in *MUTYH* with reported MAFs less than 0.01 or unknown. From these, 22 RVs were selected: 14 non-synonymous coding variants and eight intronic variants within predicted epigenetic or splice regions (Table 23).

Table 22: Preliminary univariate logistic regression analysis results following Taqman genotyping in 750 bladder cancer cases and 706 controls for select candidate *MUTYH* rare variants.

***MUTYH* Variant**[a]	**Case**			**Control**			**Per-allele OR (95%CI)**	**P-value**[c]
	VV[b]	*VW*[b]	*WW*[b]	*VV*[b]	*VW*[b]	*WW*[b]		
rs3219470 C>T	0	12	646	1	8	620	1.14 (0.51-2.56)	0.66
Chr 1: 45803644 G>C	0	1	740	0	7	690	0.13 (0.02-1.09)	**0.03**
rs34612342 A>G (Y179C)	0	4	738	0	0	701	Variant not seen in controls	0.13
rs138089183 C>T (R309C)	0	1	742	0	2	701	0.47 (0.04-5.22)	0.62
rs36053993 G>A (G396D)	0	11	726	0	4	692	2.62 (0.83-8.27)	0.09
rs140118273 C>T (S515F)	0	14	673	0	13	648	1.04 (0.48-2.25)	0.93

[a] *MUTYH* RVs subjectively suggestive of an effect on bladder cancer risk on comparing number of carriers among cases versus controls are in bold.
[b] VV – homozygote variant, VW – heterozygote variant, WW – homozygote wildtype
[c] Significant Fishers exact test P-values (P ≤ 0.05) are in bold.

Table 23: Candidate known *MUTYH* rare variants from the UCSC genome browser selected for genotyping in the bladder cancer case-control study population.

Gene	Variant dbSNP rs number	dbSNP MAF	Predicted functional effects
MUTYH	rs3219496 C>A	0.006	Missense : L529M
	rs3219494 G>A	0.003	Missense : G503E
	rs121908381 G>T	Unknown	Nonsense : E480X
	rs121908383 A>G	Unknown	Missense : Q414R
	rs121908382 C>T	Unknown	Missense : P405S
	rs35352891 C>T	0.005	Missense : A373V
	rs112422930 T>G	Unknown	Missense : L257R
	rs34126013 C>T	Unknown	Missense : R241W
	rs11545695 C>T	Unknown	Missense : A227V
	rs121908380 C>A	Unknown	Nonsense : Y104X
	rs1140507 T>C	Unknown	Missense : W103R
	rs75321043 G>A	Unknown	Missense : G25D
	rs3219484 G>A	0.001	Missense : V22M
	rs79777494 C>T	Unknown	Missense : P18L
	rs7522089 C>T	0.002	Intron 1/2: H3K36me3 histone mark
	rs3219471 G>A	0.009	Intron 1/2: H3K36me3 histone mark
	rs74688071 C>A	Unknown	Intron 1: H3K4Me2 histone mark
	rs115058761 C>T	0.008	Intron 1: H3K4Me3 histone mark
	rs115629902 G>A	0.008	Intron 1: Predicted chromatin promoter region with multiple histone marks
	rs3219469 C>G	0.006	Intron 1: Predicted chromatin promoter region with multiple histone marks and transcription factor binding sites
	rs77949009 C>T	Unknown	Intron 1: Predicted chromatin promoter region with multiple histone marks and transcription factor binding sites
	rs3219468 C>G	0.005	Intron 1: 3 bp 3' from Exon 1 donor splice site; decreased *MUTYH* expression reported in carriers (Plotz *et al.* 2012)

8.2.2.3. CDRV versus CDCV hypothesis

To test the relative contributions of the CDRV hypothesis versus the CDCV hypothesis on bladder cancer risk, up to the top 10 (if reported) most significant SNPs from each of the published papers from the bladder cancer GWAS studies (Kiemeney *et al.* 2008; Wu *et al.* 2009; Kiemeney *et al.* 2010; Rothman *et al.* 2010; Garcia-Closas *et al.* 2011; Rafnar *et al.* 2011; Tang *et al.* 2012) were selected to be genotyped alongside the above RVs in the LBCS study population resulting in a total of 36 SNPs (Table 24). The *XPC* SNP rs2228000 previously associated with an increased bladder cancer risk was also included (Stern *et al.* 2009). Inclusion of these SNPs would allow identification of indirect or synthetic associations of GWAS common SNP bladder cancer loci in linkage association with causal rare variants (which can potentially be over 9 Mb apart (Dickson *et al.* 2010)), and the independent risk contributions of candidate RVs and known bladder cancer risk SNPs.

Table 24: Most significant SNPs from the corresponding bladder cancer GWAS publications selected for genotyping in the bladder cancer case-control study.

Variant dbSNP rs number	Chromosome	Closest reported genes	Citation
rs17418689 A>G	2	*THSD7B*	(Kiemeney *et al.* 2008)
rs710521 T>C	3	*TP63, LEPREL1*	
rs10240737 A>G	7	*PLXNA4*	
rs9642880 G>T	8	*MYC*	
rs1092116 A>G	10	*ZMIZ1*	
rs233716 C>T	12	*PTPN11, MIR1302-1*	
rs233722 G>A	12	*PTPN11, MIR1302-1*	
rs12584999 G>A	13	*DDX6P2*	
rs12982672 G>A	19	*ARHGAP33*	
rs6610426 A>G	X	*ATP6AP2*	
rs2819049 A>G	10	*TCEB1P3*	(Wu *et al.* 2009)
rs11615848 G>T	12	*UBA52P7*	
rs4902033 C>A	14	*SLC38A6, TMEM30B*	
rs6100488 T>C	20	*PHACTR3*	
rs6969519 T>C	7	*MEOX2*	
rs11782640 C>T	8	*TRAPPC9*	
rs1584415 C>T	9	*CYLC2*	
rs16936133 G>A	9	*CYLC2*	
rs2294008 C>T	8	*PSCA*	
rs2228000 C>T	3	*XPC*	(Stern *et al.* 2009)
rs11892031 A>C	2	*UGT1A*	(Rothman *et al.* 2010)
rs1495741 G>A	8	*NAT2*	
rs8102137 T>C	19	*CCNE1*	
rs1014971 C>T	22	*APOBEC3, CBX6*	
rs17717312 G>A	10	*ABLIM1*	(Garcia-Closas *et al.* 2011)
rs10775480 T>C	18	*SLC14A1*	
rs10853535 C>T	18	*SLC14A1*	
rs11082469 A>G	18	*SLC14A1*	
rs11877720 A>G	18	*SLC14A1*	
rs7238033 T>C	18	*SLC14A1*	
rs16982241 G>A	19	*FUT2*	
rs12975781 C>T	19	*RASIP1, IZUMO1, FUT1, FGF21*	
rs12982115 A>G	19	*RASIP1, IZUMO1, FUT1, FGF21*	
rs1474680 A>G	20	*TGM6*	
rs1058396 G>A	18	*SLC14A1*	(Rafnar *et al.* 2011)
rs17674580 C>T	18	*SLC14A1*	
rs17863783 G>T	2	*UGT1A*	(Tang *et al.* 2012)

8.2.3. Rare variants and bladder cancer risk

For the 97 variants, genomic DNA samples with blind duplicates from 7.6% of samples were genotyped for 96 variants using custom Fluidigm SNPtype Assays on a Fluidigm 96.96 Dynamic Array. *XPC* Chr 3: 14188798 G>A (G866R) was genotyped using a custom Taqman genotyping assay. Fifteen assays failed to discriminate (Table 25). Of these, 11 *MUTYH* variants were identified from the UCSC Genome Browser so it is unclear if the failure to discriminate was due to assay failure or due to these rare variants not being present in the study population. The remaining two *MUTYH* and two *XPC* variants identified by NGS had been previously confirmed by conventional sequencing in a sub-population thus represented assay failure. Two of these variants were previously noted to be in a GC rich region or had another SNP in close proximity thus possibly accounting for the assay failure.

Table 25: List of variants that failed to discriminate on genotyping with details of source where the variant was identified and the genotyping assay type used.

Gene	Variant NCBI hg19 position/ dbSNP rs number	Variant Source	Assay Type
MUTYH	rs3219494 G>A	UCSC Genome Browser	Fluidigm
MUTYH	rs121908381 G>T	UCSC Genome Browser	Fluidigm
MUTYH	rs121908383 A>G	UCSC Genome Browser	Fluidigm
MUTYH	rs121908382 C>T	UCSC Genome Browser	Fluidigm
MUTYH	rs35352891 C>T	UCSC Genome Browser	Fluidigm
MUTYH	rs112422930 A>C	UCSC Genome Browser	Fluidigm
MUTYH	rs11545695 C>T	UCSC Genome Browser	Fluidigm
MUTYH	rs1140507 T>C	UCSC Genome Browser	Fluidigm
MUTYH	rs115058761 G>A	UCSC Genome Browser	Fluidigm
MUTYH	rs115629902 C>T	UCSC Genome Browser	Fluidigm
MUTYH	rs3219469 C>G	UCSC Genome Browser	Fluidigm
MUTYH	Chr 1: 45797428 G>A	Unindexed multiplexed NGS (NextGene)	Fluidigm
MUTYH	rs149866955 C>T	Unindexed multiplexed NGS (NextGene)	Fluidigm
XPC	Chr 3: 14199857 G>T	Indexed multiplexed NGS (XPC only)	Fluidigm
XPC	Chr 3: 14188798 G>A	Unindexed multiplexed NGS (NextGene)	Taqman

There was 99.9% concordance of replicates with non-concordant samples dropped from analysis. Excluding the above 15 “failed” assays, for the remaining 82 genotyping assays, the mean assay failure rate was 3.1% with 75 of 82 assays having an assay failure rate of less than 5%. Hardy-Weinberg equilibrium was maintained for all common SNPs in the controls. Due to the differences in age and gender between the case and control populations, all of the following single marker analyses were adjusted for age and gender to try correct for these differences. Further multivariate analyses adjusting for ethnicity, family history, smoking history and occupational exposure was also performed but only included individuals from the LBCS study population (where this data was collected). Full analyses results for all variants are in Appendix C.

8.2.3.1. GWAS SNPs and common MUTYH and XPC SNPs

Of 36 GWAS SNPs, 14 SNPs were significantly associated with bladder cancer predisposition with P-values less than 0.05 (Table 26), of which the minor allele of three SNPs were associated with increased bladder cancer risk and the remaining 11 being protective. The “SNP” with the strongest effect was on chromosome 2, rs17418689 A>G (Control MAF 0.002, age and gender adjusted OR 0.09 [95% CI 0.01 - 0.78], P_{trend} = 0.03), which interestingly was an uncommon SNP (MAF 0.06) with the minor allele being a risk allele (OR 1.17) in the original publication (Kiemeney *et al.* 2008) yet a protective allele and a rare variant in the LBCS and WTCC population. The most significant SNP was also on chromosome 2, rs11892031 A>C (age and gender adjusted OR 0.66 [0.52 - 0.84], P_{trend} < 0.001), previously reported to be secondary to the tagging of the functional uncommon SNP rs17863783 G>T (Tang *et al.* 2012). On Bonferroni correction for 36 GWAS SNPs, only rs11892031 remained significant at a Bonferroni corrected P-value of 0.001.

Table 26: Genotyping results for GWAS SNPs with significant associations with bladder cancer susceptibility on univariate or multivariate logistic regression. The age and gender adjusted analysis was for the combined LBCS and WTCC populations while the multivariate analysis only included the LBCS study population.

dbSNP rs number Chromosome (closest genes)	MAF Case	MAF Control	No of alleles	Case N (%)	Control N (%)	Age & gender adjusted OR (95% CI)[a]	P_{trend}[b]	Multivariate OR (95% CI)[c]	P_{trend}[b]
rs17418689	0.001	0.002	0	843 (99.9)	1941 (99.5)	1		1	
A>G			1	1 (0.1)	9 (0.5)	0.09 (0.01 - 0.78)	**0.03**	0.10 (0.01 - 0.87)	**0.04**
Chr 2 (*THSD7B*)									
rs11892031	0.07	0.10	0	730 (86.6)	1590 (81.6)	1		1	
A>C			1	107 (12.7)	342 (17.6)	0.62 (0.48 - 0.80)	**<0.001**	0.56 (0.42 - 0.74)	**<0.001**
Chr 2 (*UGT1A*)			2	6 (0.7)	16 (0.8)	0.88 (0.30 - 2.60)		1.27 (0.31 - 5.17)	
rs17863783	0.01	0.02	0	821 (97.6)	1859 (95.2)	1		1	
G>T			1	20 (2.4)	91 (4.7)	0.46 (0.27 - 0.79)	**<0.01**	0.48 (0.27 - 0.85)	**0.01**
Chr 2 (*UGT1A*)			2	0 (0)	2 (0.1)	Not seen in cases		Not seen in cases	
rs1495741	0.19	0.22	0	550 (66.3)	1185 (61)	1		1	
G>A			1	250 (30.2)	649 (33.4)	0.78 (0.64 - 0.96)	**<0.01**	0.80 (0.64 - 1.01)	**<0.01**
Chr 8 (*NAT2*)			2	29 (3.5)	109 (5.6)	0.56 (0.35 - 0.89)		0.51 (0.31 - 0.86)	
rs9642880	0.51	0.46	0	207 (25.1)	591 (31.4)	1		1	
G>T			1	401 (48.7)	868 (46.1)	1.38 (1.10 - 1.72)	**<0.01**	1.41 (1.09 - 1.82)	**<0.01**
Chr 8 (*MYC*)			2	216 (26.2)	424 (22.5)	1.50 (1.16 - 1.95)		1.50 (1.11 - 2.02)	
rs2294008	0.46	0.44	0	237 (28.2)	593 (30.6)	1		1	
C>T			1	430 (51.1)	982 (50.6)	1.10 (0.89 - 1.36)	**0.05**	1.00 (0.78 - 1.28)	0.06
Chr 8 (*PSCA*)			2	174 (20.7)	366 (18.9)	1.31 (1.00 - 1.72)		1.41 (1.02 - 1.94)	
rs11615848	0.07	0.10	0	722 (86)	1610 (82.3)	1		1	
G>T			1	115 (13.7)	320 (16.4)	0.70 (0.54 - 0.91)	**<0.01**	0.69 (0.52 - 0.92)	**<0.01**
Chr 12 (*UBA52P7*)			2	3 (0.4)	26 (1.3)	0.29 (0.08 - 1.02)		0.23 (0.06 - 0.86)	
rs233722	0.40	0.40	0	306 (36.6)	701 (35.9)	1		1	
G>A			1	401 (47.9)	929 (47.7)	1.00 (0.81 - 1.22)		0.90 (0.71 - 1.14)	
Chr 12 (*PTPN11, MIR1302-1*)			2	130 (15.5)	319 (16.4)	0.78 (0.59 - 1.02)	0.12	0.72 (0.53 - 0.99)	**0.05**

rs233716 C>T Chr 12 (*PTPN11, MIR1302-1*)	0.39	0.43	0 1 2	299 (36.9) 386 (47.6) 126 (15.5)	638 (34.2) 857 (46) 370 (19.8)	1 0.97 (0.78 - 1.19) 0.72 (0.55 - 0.95)	**0.04**	1 0.85 (0.67 - 1.09) 0.75 (0.54 - 1.04)	0.07
rs8102137 T>C Chr 19 (*CCNE1*)	0.37	0.33	0 1 2	330 (39.8) 386 (46.5) 114 (13.7)	857 (44.9) 828 (43.4) 222 (11.6)	1 1.18 (0.97 - 1.44) 1.32 (0.98 - 1.78)	**0.04**	1 1.27 (1.01 - 1.60) 1.36 (0.96 - 1.91)	**0.03**
rs12975781 C>T Chr 19 (*RASIP1, IZUMO1, FUT1, FGF21*)	0.38	0.39	0 1 2	329 (39.4) 379 (45.4) 126 (15.1)	709 (36.7) 923 (47.8) 300 (15.5)	1 0.81 (0.66 - 0.99) 0.82 (0.62 - 1.08)	0.07	1 0.73 (0.58 - 0.93) 0.80 (0.58 - 1.11)	**0.05**
rs1474680 A>G Chr 20 (*TGM6*)	0.33	0.35	0 1 2	392 (46.7) 349 (41.6) 98 (11.7)	828 (42.4) 881 (45.1) 243 (12.4)	1 0.76 (0.62 - 0.92) 0.79 (0.58 - 1.07)	**0.02**	1 0.75 (0.60 - 0.94) 0.75 (0.53 - 1.06)	**0.02**
rs6100488 T>C Chr 20 (*PHACTR3*)	0.18	0.18	0 1 2	564 (67.1) 249 (29.6) 28 (3.3)	1311 (67) 575 (29.4) 70 (3.6)	1 0.97 (0.79 - 1.19) 0.85 (0.51 - 1.39)	0.54	1 0.84 (0.67 - 1.06) 0.66 (0.38 - 1.12)	**0.05**
rs1014971 C>T Chr 22 (*APOBEC3, CBX6*)	0.35	0.37	0 1 2	345 (41.1) 394 (47) 100 (11.9)	764 (39.1) 945 (48.3) 246 (12.6)	1 0.81 (0.66 - 0.98) 0.80 (0.59 - 1.09)	**0.05**	1 0.76 (0.60 - 0.96) 0.70 (0.50 - 0.99)	**0.01**

[a] Logistic regression of the combined LBCS and WTCC populations adjusting for age and gender.
[b] P-values less than 0.05 are in bold.
[c] Multivariate logistic regression of only the LBCS study population adjusting for age, gender, ethnicity, family history, smoking history and occupational exposure.

On conditional analysis of all GWAS SNPs in the full study population, only seven SNPs continued to have an independent contribution to bladder cancer predisposition (Table 27). Interestingly, of the three chromosome 2 SNPs, only rs11892031 A>C remained significant but not rs17418689 A>G or rs17863783 G>T perhaps suggesting other functional variants being tagged by rs11892031 A>C not explained by rs17863783 G>T.

Table 27: GWAS SNPs with an independent contribution to bladder cancer susceptibility on conditional analysis of the combined LBCS and WTCC study populations.

dbSNP rs number	OR (95% CI)	P-value
rs11892031 A>C	0.73 (0.55 - 0.97)	0.03
rs1495741 G>A	0.81 (0.68 - 0.97)	0.02
rs9642880 G>T	1.25 (1.08 - 1.43)	0.002
rs233716 C>T	0.48 (0.26 - 0.91)	0.02
rs11615848 G>T	0.58 (0.45 - 0.76)	<0.001
rs8102137 T>C	1.18 (1.01 - 1.36)	0.03
rs1474680 A>G	0.82 (0.71 - 0.96)	0.01

Among the *MUTYH* variants, there were two *MUTYH* variants, both obtained from the UCSC genome browser, with observed control MAFs greater than 0.05, rs3219484 G>A (observed MAF 0.08) and rs74688071 G>T (observed MAF 0.17), with neither being associated with bladder cancer susceptibility (P_{trend} = 0.46 and 0.52 respectively). There were no *XPC* variants with a MAF greater than 0.05 in the study population, with common *XPC* SNP rs2228000 C>T (dbSNP MAF 0.24) unexpectedly having only an observed control MAF of 0.03. On univariate analysis, *XPC* SNP rs2228000 C>T had the most significant result of all variants genotyped (P_{trend} = 3.31 x 10^{-6}). However, it was not associated with bladder cancer risk following age and gender adjustment or on full multivariate analysis (OR 1.24 [95%CI 0.90 - 1.71], P_{trend} = 0.13; and OR 1.37 [0.96 - 1.95], P_{trend} = 0.54 respectively) which was surprising based on previous publications (Stern *et al.* 2009).

8.2.3.2. Individual MUTYH and XPC uncommon SNPs and rare variants

In the combined LBCS and WTCC population, 35.5% of the study population carried at least one minor allele of an uncommon SNP or RV (observed MAF < 0.05) in *MUTYH* or *XPC* (Table 28). 11.8% of the study population carried at least one minor rare variant allele (MAF < 0.01) in *MUTYH* or *XPC*.

Table 28: The number of minor alleles of variants with a control MAF under 0.05 or under 0.01 carried by individuals in the combined LBCS and WTCC population.

No of minor alleles carried	Control MAF < 0.05 N (%)	Control MAF < 0.01 N (%)
0	1842 (64.5)	2522 (88.2)
1	818 (28.6)	319 (11.2)
2	167 (5.8)	15 (0.5)
3 or more	31 (1.1)	2 (0.1)

On single variant analyses, only two *MUTYH* and *XPC* variants were found to be associated with increased bladder cancer risk either on univariate or multivariate analyses (P < 0.05), the *MUTYH* rare variant rs36053993 G>A (G396D) and *XPC* uncommon SNP, rs3731078 A>G (Table 29). As earlier mentioned, the former variant *MUTYH* rs36053993 G>A (G396D) has been extensively investigated and linked with colorectal cancer risk (Theodoratou *et al.* 2010). *XPC* rs3731078 A>G remained significant even after adjusting for rs2228000 (P = 0.02). These results are tempered by the low statistical power of the study population size and issues of multiple testing. However, the application of the Bonferroni correction here may prove too conservative, as this presumes complete independence of all RVs within a gene.

Table 29: Genotyping results for *MUTYH* and *XPC* variants with significant associations with bladder cancer susceptibility. Age and gender adjusted, and multivariate logistic regression analysis results are displayed.

Gene & dbSNP rs number	MAF Case	MAF Control	No of alleles	Case N (%)	Control N (%)	Age and gender adjusted OR (95% CI)[a]	P_{trend}[b]	Multivariate OR (95% CI)[c]	P_{trend}[b]
MUTYH rs36053993 G>A	0.009	0.006	0	827 (98.2)	1927 (98.7)	1		1	
			1	15 (1.8)	25 (1.3)	1.61 (0.76 - 3.40)	0.21	5.83 (1.31 - 25.90)	**0.02**
XPC rs3731078 A>G	0.04	0.03	0	786 (92.9)	1861 (95.2)	1		1	
			1	59 (7.0)	92 (4.7)	1.53 (1.04 - 2.26)	**0.02**	1.29 (0.82 - 2.04)	0.22
			2	1 (0.1)	2 (0.1)	2.94 (0.16 - 54.05)		Insufficient samples to calculate	

[a] Logistic regression of the combined LBCS and WTCC populations adjusting for age and gender.
[b] P-values less than 0.05 are in bold.
[c] Multivariate logistic regression of only the LBCS study population adjusting for age, gender, ethnicity, family history, smoking history and occupational exposure.

8.2.3.3. MUTYH and XPC collapsed analyses

This study undertook a simple single-direction collapsed variant analyses (performed by Dr J Nsengimana), as previously described (Barrett and Nsengimana 2011; Konig *et al.* 2011). In summary, this method collapses all uncommon SNPs (MAF < 0.05) and RVs (MAF < 0.01) within a gene (e.g. *MUTYH* or *XPC*) by generating a total count or gene score of all the minor alleles carried of these variants by an individual. It then tests for the association between this gene score with bladder cancer risk applying a Bonferroni-corrected significance level (P = 0.025 for two genes*)*. For the multivariate analyses, among the GWAS SNPs, only rs11892031 was included being the only GWAS SNP that remained significant on Bonferroni correction. rs2228000 was also corrected for, having the most significant result on univariate analyses. Age and gender, were adjusted for to account for the differences in case and control populations. No correction was performed for other population demographics as this would result in dropping the WTCC control population (which had no other demographic data available) and loss of statistical power.

The collapsed *XPC* variants were significantly associated with increased bladder cancer susceptibility on univariate analyses (Per-RV allele OR 1.31 [95%CI 1.07 – 1.60], P = 0.008) though they were just short of significance on multivariate analyses (1.29 [1.03 – 1.60], P = 0.026). Correction for all 36 GWAS SNPs and *XPC* rs2228000 revealed the collapsed *XPC* variants to be an independent genetic risk factor separate from these SNPs (P = 0.02). However, there was no significant association found for the collapsed *MUTYH* variants and bladder cancer risk on both univariate and multivariate analyses (0.94 [0.74 – 1.20], P = 0.62; and 0.92 [0.70 – 1.21], P = 0.57 respectively).

With regards to validation of the Syzygy calls as part of the unindexed multiplexed NGS methods development, there was no association found with bladder cancer risk (1.01 [0.72 – 1.44], P = 0.93) on collapsed analyses of all *MUTYH* and *XPC* variants called by Syzygy to have a Chi^2-value greater than two, unfortunately, invalidating the use of the Syzygy Chi^2-value as a tool for selecting candidate RVs detected by unindexed multiplexed NGS for further investigation. The *MUTYH* rs34612342 (Y179C) and rs36053993 (G396D) RVs have been extensively studied and implicated in colorectal cancer risk (Farrington *et al.* 2005; Tenesa *et al.* 2006; Theodoratou *et al.* 2010) and demonstrated to result in dysfunctional MUTYH activity (Ali *et al.* 2008; Kundu *et al.* 2009). It was hypothesised that these two functional RVs would affect bladder cancer risk. Preliminary Taqman genotyping was suggestive of an effect for these two RVs individually (Table 22, page 140) with collapsed analysis of these two RVS being significant (P=0.02, Table 23). In the complete Fluidigm genotyping of the LBCS and WTCC study populations, collapsed analysis of these two RVs did not reveal any association with bladder cancer predisposition (P = 0.37, Table 23). However, multivariate analyses in just the LBCS population adjusting for all covariates (age, gender, ethnicity, family history, smoking history and occupational exposure) and for GWAS SNP rs11892031 and *XPC* rs2228000, there was an observed increase in bladder cancer risk in carriers of these two RVs (P = 0.02).

Table 30: Collapsed analyses of *MUTYH* rs34612342 and rs36053993 from the preliminary Taqman genotyping, and the Fluidigm genotyping for the combined LBCS and WTCC population and the LBCS population only.

Collapsed analysis: rs34612342 + rs36053993	**Case**			**Control**			**OR (95%CI)**	**P-value[b]**
	VV[a]	*VW*[a]	*WW*[a]	*VV*[a]	*VW*[a]	*WW*[a]		
Preliminary Taqman: LBCS sub-population	0	15	728	0	4	699	3.60 (1.18-10.93)	**0.02**
Fluidigm genotyping: LBCS + WTCC population	0	16	823	0	29	1909	1.39 (0.68-2.87)	0.42[c]
Fluidigm genotyping: LBCS population only	0	16	823	0	6	759	5.85 (1.31-26.06)	**0.02**[d]

[a] VV – homozygote variant, VW – heterozygote variant, WW – homozygote wildtype.
[b] Fisher's exact test P-value. Significant P-values (P < 0.05) in bold.
[c] Adjusted for age, gender, GWAS SNP rs11892031 and *XPC* SNP rs2228000.
[d] Adjusted for age, gender, ethnicity, family history, smoking history, occupational exposure, GWAS SNP rs11892031 and XPC SNP rs2228000.

8.2.3.4. Gene-gene and gene-environment interactions and gene-phenotype associations

Tests for gene-gene interactions of collapsed *MUTYH* variants and collapsed *XPC* variants found no multiplicative effects of carrying variants in both genes on bladder cancer risk in the combined LBCS and WTCC populations (OR 1.64 [95% CI 0.89 – 3.03], P = 0.12). There was also no evidence of interactions between either smoking or occupational exposure with collapsed variants in *MUTYH* (P = 0.40 and 0.97, respectively) and *XPC* (P = 0.30 and 0.77, respectively) on bladder cancer risk in the LBCS population.

Among the bladder cases, variants were tested for correlations with bladder tumour phenotype (superficial tumours, Ta and T1, versus invasive tumours, Tis and T2+) and histological grade (low grade (G<2) versus high grade (G3)). Once again, no associations were found for the collapsed *MUTYH* and *XPC* variants with either tumour phenotype or histological grade (*MUTYH*: P = 0.72 and P = 0.57, respectively; *XPC*: P = 0.15 and P = 0.65, respectively). Of the GWAS SNPs, only four SNPs had associations with either bladder tumour phenotype or histological grade though none of these would have reached significance on Bonferroni correction (P = 0.001) (Table 30).

Table 31: Associations between bladder tumour phenotype and histological grade with GWAS SNPs previously associated with bladder cancer susceptibility.

	Non-invasive versus invasive phenotype		**Low versus High Histological Grade**	
GWAS SNP	*Per-allele OR (95% CI)*	*P-value*[a]	*Per-allele OR (95% CI)*	*P-value*[a]
rs16982241 G>A	1.21 (0.88 - 1.65)	0.24	1.41 (1.05 - 1.60)	**0.02**
rs17717312 G>A	0.65 (0.43 - 0.99)	**0.05**	0.80 (0.57 - 1.14)	0.22
rs6610426 A>G	0.74 (0.56 - 0.97)	**0.03**	0.84 (0.68 - 1.05)	0.12
rs6969519 T>C	1.69 (1.16 - 2.47)	**0.006**	1.32 (0.91 - 1.91)	0.15

[a] P-values less than 0.05 are in bold.

8.2.4. Discussion

8.2.4.1. CDRV hypothesis in bladder cancer susceptibility

This is the first study exploring the CDRV hypothesis in bladder cancer genetic predisposition using an NGS RV discovery stage. This identified *XPC* variants to be associated with increased bladder cancer risk on univariate analysis though was not significant on multivariate analysis with a Bonferroni correction. This result partly supports previous work by Dr Kiltie's group on RVs in bladder cancer (Qiao *et al.* 2011), which found four *XPC* rare variants associated with increased bladder cancer risk. Multiple studies have found associations between common *XPC* coding SNPs and multiple cancer sites (Doherty *et al.* 2011; Jiao *et al.* 2011; Gil *et al.* 2012; Yang *et al.* 2012). Bladder cancer case-control SNP studies had predominantly investigated the rs2228000 (A499V variant) and rs2228001 (K939Q variant) (Sak *et al.* 2005; Sak *et al.* 2006; Fontana *et al.* 2008; Liu *et al.* 2012) but a pooled analysis found only rs2228000 weakly associated with bladder cancer risk (Stern *et al.* 2009).The current study adds to this literature by demonstrating the independence of the effects of these *XPC* RVs from *XPC* rs2228000 and previously discovered GWAS SNPs based on the CDCV hypothesis.

However, the risk effect observed was far smaller than predicted by the CDRV hypothesis with a combined OR of only about 1.30 observed, though this matches some of the observations by Rivas *et al* of RVs in inflammatory bowel disease with ORs of 1.40 to 1.60 (Rivas *et al.* 2011). These results, thus, could suggest that the relative contribution of the CDRV hypothesis to overall inherited disease susceptibility may not be as large as proposed. Looking at individual *XPC* markers, only *XPC* rs3731078 A>G was associated with an increased bladder cancer risk though on taking into account multiple testing, this could very well be a false positive result.

Validation and functional studies will be needed to clarify the relevance of this uncommon SNP on bladder cancer risk.

In this study, no contribution was observed on collapsed analysis of *MUTYH* RVs to bladder cancer risk. Previous *MUTYH* RV studies have predominantly focused on the rs36053993 (G396D) and rs34612342 (Y179C) variants in MAP and colorectal cancer (Tenesa *et al.* 2006; Theodoratou *et al.* 2010), with reports of other solid cancers in MAP patients: duodenal, testicular germ cell, ovarian, bladder, skin, lung, and thyroid cancers (Vogt *et al.* 2009; Pervaiz *et al.* 2010; von der Thusen *et al.* 2011; Castillejo *et al.* 2012). A breast cancer case-control study reported an increased breast cancer risk in heterozygote carriers of the G396D and Y179C variants (Rennert *et al.* 2012), while Win *et al* observed in heterozygote carriers of 12 previously identified *MUTYH* missense and nonsense RVs an increased colorectal cancer (standardised incidence ratios (SIR) 2.04), gastric cancer (SIR 3.24), liver cancer (SIR 3.09) and endometrial cancer (SIR 2.33) compared to the general population (Win *et al.* 2011). The MUTYH G396D and Y179C protein variants have been demonstrated to have defective DNA binding and glycosylase functional activity. These studies highlighted the strong evidence for *MUTYH* RVs in population cancer susceptibility thus it was surprising for the lack of effect seen in the current study.

The 1000 Genomes Exon Pilot Project undertook exome sequencing of 800 individuals from seven populations identifying 12758 variants of which 73.7% had MAFs of under 1% with significantly low inter-population sharing of RVs (Marth *et al.* 2011). They also reported that a significantly higher proportion of RVs (MAF<1%) resulted in missense or nonsense amino acid changes, and were predicted to have a functional impact (Figure 30). These findings were replicated by the NHLBI Exome Sequencing Project exome sequencing 2440 individuals with over 80% of all single

nucleotide variants (SNV) and 95.7% of putatively functional coding SNVs having a MAF under 0.5%, plus they also successfully modelled the growth of RVs with recent population growth, and the negative selection of coding RVs (Tennessen *et al.* 2012). They also demonstrated that, on average, each individual carried up to 580 putatively functional coding SNVs of which 35 were nonsense variants with each individual being homozygous for at least one of these nonsense variants. Nelson *et al* sequenced 202 candidate drug target genes in 14002 individuals, finding 95% of SNVs detected being rare, with over half of these RVs being predicted to be functionally deleterious (Nelson *et al.* 2012). These three studies, thus, underline the abundance and the potential pathogenicity of RVs in the general population.

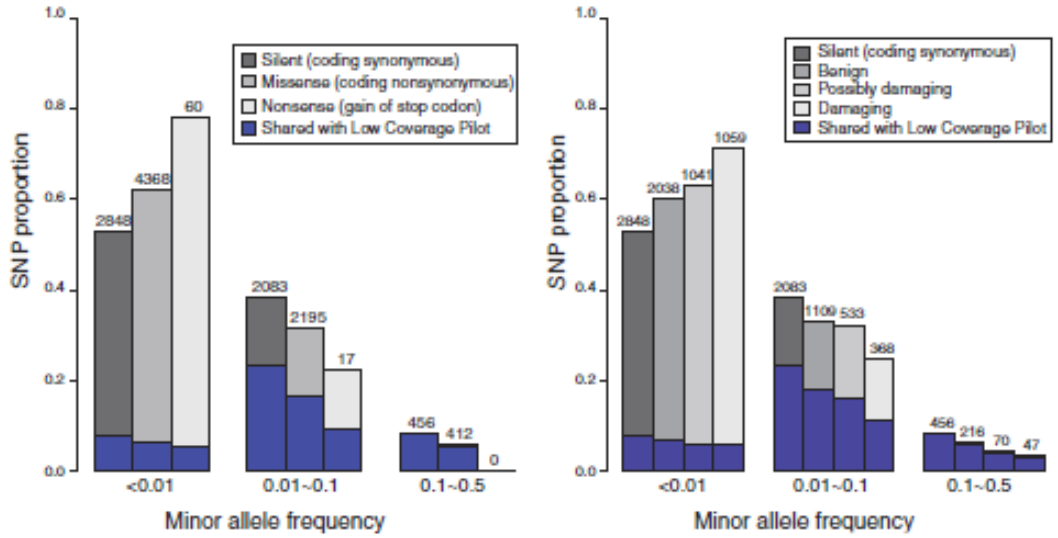


Figure 30: Distribution of variants identified by the 1000 Genomes Exon Pilot Project by allele frequency, amino acid change (left) and predicted functional effects by SIFT and PolyPhen-2 (right) (Marth *et al.* 2011).

Over the last year, several studies have been published investigating predominantly coding RVs in common disease predisposition. These studies in inflammatory bowel disease (Rivas *et al.* 2011), asthma (Torgerson *et al.* 2012), deep vein thrombosis (Lotta *et al.* 2012), multiple sclerosis (Ramagopalan *et al.* 2011) and degenerative disorders (Raychaudhuri *et al.* 2011; Jin *et al.* 2012) identified RVs by conventional Sanger sequencing, multiplexed NGS or whole exome sequencing, with significant

moderate to large risk (individual rare variant OR 1.44 to greater than 10) or protective (individual rare variant OR 0.29 – 0.72) effects seen on genotyping. Some of these studies also proceeded to demonstrate the independence of significant RVs effects from known common GWAS disease variants or common haplotype risk groups (Raychaudhuri *et al.* 2011; Rivas *et al.* 2011), and the biological effects of these variants on gene or protein function (Rivas *et al.* 2011). Only one RV study by Heinzen *et al* on idiopathic generalised epilepsy has reported a negative result for individual RV association but this could be secondary to insufficient statistical power, the use of the over-conservative Bonferonni correction and not accounting for the overall effects of multiple RVs (Heinzen *et al.* 2012).

In cancer, RV studies have focused on specific candidate genes or known candidate variants. Breast cancer studies have investigated DSB repair and HRR genes due to their relationship with the *BRCA1/2* genes, identifying *XRCC2* (Park *et al.* 2012), *PALB2* (Tischkowitz *et al.* 2012) and potentially *ATM* (Goldgar *et al.* 2011) (not significant but study was underpowered) missense coding RVs associated with increased breast cancer risk. Lefevre *et al* examined 31 known RVs from candidate genes previously associated with colorectal, gastric, breast and prostate cancer for associations with multiple adenomatous polyposis and early-onset colorectal cancer, and found four individually significant risk RVs and pooled analysis of all RVs with MAF under 0.5% to be associated with increased disease risk (OR 3.14, P-value = 0.02) (Lefevre *et al.* 2012). In melanoma, a rare missense coding variant (E318K; MAF 0.008) in the *MITF* transcription factor gene was confirmed as a susceptibility variant (overall OR 2.19) in two large case-control study populations, and was shown to be associated with a family history of melanoma, early disease onset and multiple melanomas, and found to increase the transcriptional activity of MITF target genes (Yokoyama *et al.* 2011). These studies thus highlight the role of the CDRV hypothesis in cancer genetic susceptibility.

8.2.4.2. Study limitations and future work

This study suffers from several limitations inherent for most RV studies. The biggest limitation is the issue of achieving adequate statistical power to detect an association. Due to the rarity of RVs, single RV marker analysis would be severely underpowered using current case-control study sizes with reported powers under 5% for a sample size of 1000 cases and 1000 controls (Li and Leal 2008; Bansal *et al.* 2010; Witte 2012), and clearly demonstrated for this study in Table 19. The CDRV hypothesis and the assumption for collapsed RV analysis is that, "in aggregate, (RVs) may be common enough to account for variation in common traits", thus by collapsing these RVs together a greater power could be achieved (Ladouceur *et al.* 2012). As shown in Figure 31, over 100,000 samples would be needed to detect an RV with a MAF of 0.001 (black line) for an OR of 2.0, but if 25 of such RVs were collapsed, this would in effect shift the sample size needed to the equivalent of a variant with MAF 0.025 (blue line) reducing the study size by over ten-fold.

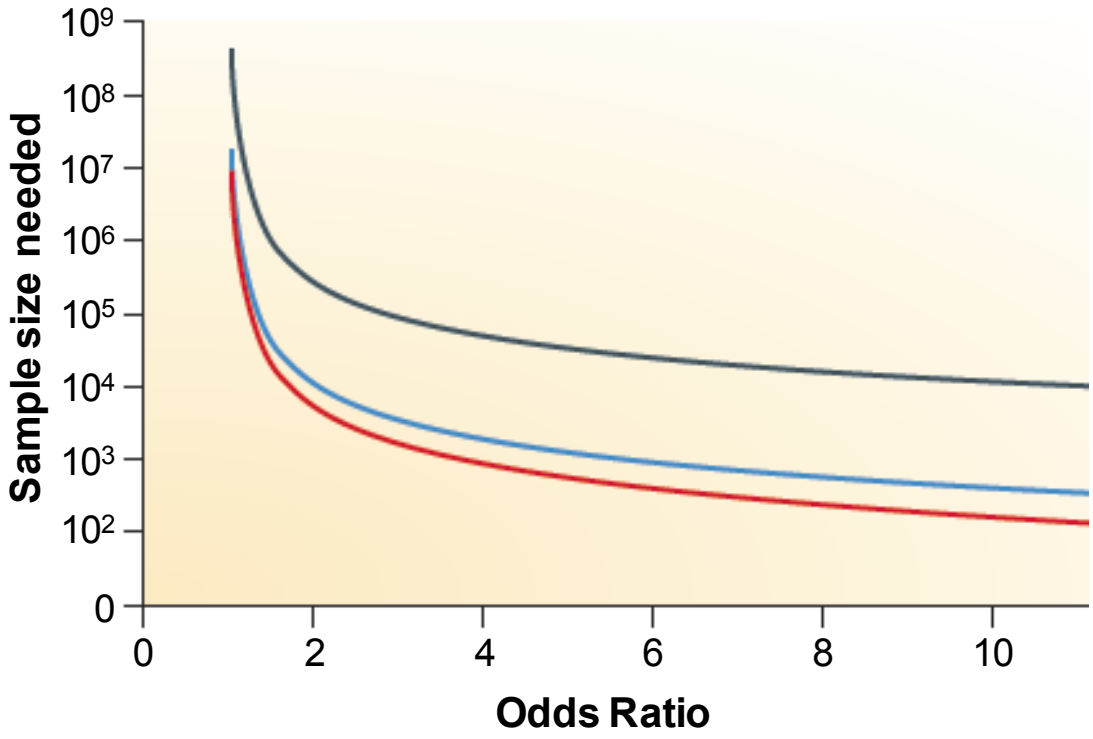


Figure 31: The case-control study sample size needed to achieve 80% statistical power for a single variant with a MAF of 0.001 (black line), 0.025 (blue line) and 0.05 (red line) at different odds ratios of effect (adapted from (Bansal *et al.* 2010).

There are four main statistical approaches to collapsed RV analysis (Bansal *et al.* 2010; Dering *et al.* 2011; Ladouceur *et al.* 2012): collapsing all RVs across a region/ gene (Li and Leal 2008), *a priori* weighted collapsing of RVs based on MAF or biological effects (Madsen and Browning 2009; Flanagan *et al.* 2010; Price *et al.* 2010), RV variance distribution across cases and controls (Liu and Leal 2010; Ionita-Laza *et al.* 2011; Wu *et al.* 2011), and data-adaptive summation (Han and Pan 2010). Comparisons of these different collapsing analytical approaches reported the greatest statistical power using the permutation-based variance distribution and data-adaptive approaches, but even then the power remained low with maximum reported powers of just 30% for a study of 500 cases and 500 controls (Dering *et al.* 2011; Dering *et al.* 2011; Ladouceur *et al.* 2012). This study used the first approach, as for single gene analysis, this approach was simulated to have marginally higher statistical power compared to a similar approach to the data-adaptive summation method that accounts for the direction of effect on disease risk for each variant (Barrett and Nsengimana 2011; Konig *et al.* 2011). A permutation-based data-adaptive sum test, as described by Han and Pan (Han and Pan 2010), was also tested producing similar results thus was not shown. In view of the low power even with these collapsing approaches, it is thus likely that the current study on *MUTYH* and *XPC* was not powered to detect all true causal RVs. The Genetic Analysis Workshop 17 has highlighted these issues with permutation-based approaches as well as issues of P-value estimation and the applicability of these methods to all scenarios, hence the need for further development of these statistical tools for RV association studies (Dering *et al.* 2011).

On sub-analysis, an effect was observed for *MUTYH* rs36053993 (G396D) singly and for *MUTYH* rs36053993 (G396D) and *MUTYH* rs34612342 (Y179C) RVs combined in the preliminary genotyping results (Table 22) and in the final analysis on multivariate analysis, with both analyses involving only the LBCS cohort. This could

suggest that the WTCC cohort was not an adequately representative control group. This cohort was made up of blood donors recruited from the UK Blood Services (The Wellcome Trust Case Control Consortium 2007) thus would have no past history of cancer. However, they were markedly younger (Mean age: 41 years [Range: 15 – 65 years]) than the LBCS cases (72.1 [21 – 99]) and LBCS controls (72.1 [27 – 98]). This age difference could be important as 80% of bladder cancer cases occurs over the age of 65 years (Cancer Research UK 2010), the maximum age in the WTCC cohort, making it probable that a sizeable proportion of individuals in the WTCC controls could go on to develop bladder cancer. It was presumed that statistically adjusting for age and gender would correct for these differences so it remains unclear how this population age difference could hide the influence of RVs in bladder cancer risk.

For the NGS of *MUTYH* and *XPC*, the whole gene was not sequenced, being limited by feasibility of successful long PCR amplification, with 12 kb and 6 kb of intronic sequence not targeted during the unindexed and indexed NGS runs respectively. There are higher frequencies of SNVs in intronic regions (~70 variants per kilobase) compared to coding and UTR regions (~50 variants per kilobase) (Nelson *et al.* 2012), which could be indicative of reduced negative selection of intronic variants but could also mean potentially missing any pathogenic RVs within these regions. This study also only focused on rare SNVs, but with some evidence of rare germline copy number variations also contributing to disease risk, potentially important genetic risk determinants could have been missed (Lalani *et al.* 2012; Pylkas *et al.* 2012).

Selection of candidate RVs for genotyping was influenced by bioinformatics functional predictions of deleterious gene effects. Prediction tools for coding variations are based on the principles of biochemical categorisation of amino acids and evolutionary preservation (Ng and Henikoff 2006). These tools have been more

extensively developed and more readily available than prediction tools for non-coding variations especially intronic variants. Due to this, a higher proportion of candidate RVs selected for genotyping were coding variations (35 coding versus 25 non-coding). However, there is reported to be only a 1% consensus between different bioinformatics prediction tools for non-synonymous coding variants (Tennessen *et al.* 2012), while there are also potentially five times more evolutionarily conserved non-coding positions in the human genome compared to coding positions, with previous identification of pathogenic non-coding RVs (Lettice *et al.* 2003; Cooper and Shendure 2011; Torgerson *et al.* 2012). The limited number of non-coding prediction tools, understanding of gene regulatory and splicing elements, and "benchmarking" of all coding and non-coding bioinformatics prediction tools with experimental and clinical data of true deleterious and benign sequence variations (Cooper and Shendure 2011; Mechanic *et al.* 2012), could thus result in truly deleterious variants being excluded from further investigation in this study. Alternatively, to overcome this limitation would require impartial genotyping of all identified RVs in the full bladder cancer case-control study but this approach at present remains highly costly. Hopefully, as NGS costs fall and high throughput library preparation protocols develop, a single stage indexed multiplexed NGS study combining both RV identification and genotyping would soon be feasible for large case-control studies.

Significant RVs in this study will require validation in other bladder cancer case-control populations of European descent. Due to the population specific origin of RVs, the *XPC* gene will need to be sequenced in other ethnic populations to determine if RVs in these candidate genes are involved in bladder cancer susceptibility across all populations (Mechanic *et al.* 2012). Though current studies on common SNP gene-gene and gene-environment interactions have only found moderate effects on complex disease risk (Milne *et al.* 2010; Ciampa *et al.* 2011), evaluation of RV gene-environment and gene- gene interactions have not yet been

assessed (Cordell 2009; Thomas 2010; Mechanic *et al.* 2012). However, in this study, no gene RV interactions were found with smoking and occupational exposure or between *MUTYH* and *XPC* RVs. However, these risk estimates are limited by the retrospective recall of participants of the degree of exposure or the use of arbitrary binary (ie. exposed/ not exposed) measures, and it is recognised that the collection of accurate measures of environmental carcinogen exposure can prove to be difficult (Mechanic *et al.* 2012).

Functional studies on the biological effects of candidate RVs on gene/ protein function could aid statistical analysis by allowing *a priori* weighting, and exclusion of functionally neutral RVs to reduce multiple testing and potentially improve statistical power. Functional assays for *XPC* coding and 3'UTR variants have previously been developed by Dr Kiltie's group (Qiao *et al.* 2011). Determining the functional effects of *XPC* rs3731078 A>G would be informative of the "truth" of its observed effect increasing bladder cancer risk. It is an intronic SNP with bioinformatics analysis indicating that it is located at a histone acetylation mark thus potentially affecting chromatin packing and gene transcriptional activity (Grunstein 1997). Experiments looking at the relative *XPC* expression between wildtypes and carriers of the rs3731078 G allele could thus be useful.

As previously developed for *XPC* coding variants, attempts were made to develop an *in vivo* functional assay of RFP-tagged MUTYH recruitment to areas of oxidative laser DNA damage in RT112 bladder cancer cell lines. However, this assay proved insufficiently sensitive, requiring cell-cycle synchronisation and continuous live confocal imaging for several hours as the cells moved through DNA replication (S-phase) when MUTYH would be active (data not shown). The functional effects of MUTYH Y179C and G396D protein variants on MUTYH DNA binding and mispaired

adenine removal have been previously reported using the *in vitro* DNA binding activity and MUTYH glycosylase assays (Ali *et al.* 2008; Kundu *et al.* 2009).

In summary, this study investigated the CDRV hypothesis of rare variants in bladder cancer susceptibility specifically investigating two candidate genes, *MUTYH* and *XPC*, demonstrating an increased bladder cancer risk with *XPC* variants only. This effect on bladder cancer risk was independent of SNPs previously identified in GWASs investigating the CDCV hypothesis. Though the effect observed was smaller than expected, these findings do support previous studies of the contribution of the CDRV hypothesis in complex diseases (Goldgar *et al.* 2011; Ramagopalan *et al.* 2011; Raychaudhuri *et al.* 2011; Rivas *et al.* 2011; Stacey *et al.* 2011; Yokoyama *et al.* 2011; Jin *et al.* 2012; Lefevre *et al.* 2012; Lotta *et al.* 2012; Park *et al.* 2012; Tischkowitz *et al.* 2012; Torgerson *et al.* 2012), and will contribute to the greater understanding of bladder cancer genetic inheritability. These results will require further validation and integration with other bladder cancer "omics" studies to improve knowledge of biological mechanisms and future patient cancer phenotyping, risk-stratisfication and treatment individualisation.

8.3. DNA repair gene microRNA-binding site SNPs in bladder cancer risk and radiotherapy outcomes

In recent years, there has been increasing interest in the role of miRNAs on cancer risk and treatment outcomes. To test the hypothesis that germline SNPs in the 3'UTR of DNA repair genes could affect miRNA-mRNA binding and post-transcriptional regulation, thus influencing DNA repair capacity and bladder cancer risk and RT outcomes, SNPs residing in the 3'UTR of twenty candidate DNA repair genes were assessed for their effects on miRNA-binding *in silico*. Potentially functional SNPs were then assessed for their influence on bladder cancer susceptibility in a case-control study of 752 bladder cancer cases and 704 controls, and for association with survival following radiotherapy treatment in 199 MIBC patients treated by RT.

8.3.1. Bladder cancer case-control study of DNA repair gene miRNA-binding site SNPs

8.3.1.1. Case-control study population demographics

Germline DNAs for the bladder cancer case-control study were collected between 2002 and 2006 at St James University Hospital as described earlier (7.2.2) and the demographics of this population is shown in Table 31. There were no significant differences in the matched traits of age, gender or ethnicity between the two groups. As previously reported (Choudhury *et al.* 2008), smoking history (ever versus never smoked) and the number of lifetime occupational carcinogen exposures were significantly associated with increased bladder cancer risk (OR 1.78 [95%CI 1.41-2.26], P<0.001 and Per-exposure OR 1.26 [95%CI 1.05-1.52], P_{trend}=0.02 respectively). In the cases, 65.4% had superficial non-invasive tumours, 25.4% MIBC and 9.2% had unknown T-staging. 42.8% of cases had high pathological grade

disease. With 752 cases and 704 controls, this case-control study had a statistical power of at least 80% for detecting an OR of 1.50 for a MAF of 0.2 at a 1% significance level.

Table 32: Case-control study population demographics and tumour characteristics.

Characteristic		Case (N)	Control (N)	OR (95% CI)	P-value
Age	Mean (Range)	73.2 (30.1 - 100.7)	73.4 (28.1 - 99.7)		0.71[a]
Gender	Male	528	467		0.08[b]
	Female	220	237		
Ethnicity	Caucasian	740	689		0.34[b]
	Non-Caucasian	11	15		
Smoking History	Never	161	226	1	**<0.001**[b]
	Ever	586	461	1.78 (1.41 - 2.26)	
Number of occupational exposures	0	556	543	1	**0.02**[c]
	1	160	123		
	2	23	14		
	3	6	3	1.26 (1.05 - 1.52)[c]	
	4	0	1		
	6	2	0		
Bladder tumour characteristics					
Tumour stage	Ta	316			
	CIS	14			
	T1	162			
	T2	128			
	T3	59			
	T4	4			
	Tx	69			
Pathological grade	G1	87			
	G2	306			
	G3	314			
	Unrecorded	45			

NB. Missing data for gender in 4 patients, occupational history in 25 patients, smoking history in 22 patients and ethnicity data for 1 patient.
[a] Two-tailed T-test
[b] Pearson's Chi-squared test
[c] Per-occupational exposure OR and Chi-squared test for trend

8.3.1.2. Bioinformatics prediction for miRNA-binding sites

Twenty DNA repair genes involved in the BER, NER, NHEJ, HR and DSB signalling pathways were selected for investigation (Table 32). Bio-informatics prediction of putative miRNA-binding sites were performed by collaborators (D Landi and S Landi) at the Dipartimento di Biologia, University of Pisa, Italy (see Methods chapter 7.5.2.2 for full details). In July 2009, putative miRNA-binding sites within the 3′UTR genomic sequences for all 20 genes were predicted using several online bioinformatics algorithms. Common SNPs (MAF > 0.10 in Caucasians) within these putative miRNA-binding sites were identified and the Gibbs free energy (ΔG) for both the wildtype and variant alleles for each identified SNP determined. The difference in ΔG for the two alleles (Wildtype allele ΔG – Variant allele ΔG) was calculated as $\Delta\Delta$G. Therefore, negative $\Delta\Delta$G values indicate weaker miRNA-mRNA binding in the presence of the variant allele than the wildtype allele (ie. the wildtype allele ΔG is more negative) while positive $\Delta\Delta$G values mean stronger binding in the presence of the variant allele (ie. the variant allele ΔG is more negative). The sum of all |$\Delta\Delta$G|s greater than 3kJ/mol for each SNP (|$\Delta\Delta$Gtot|) was used to score the impact of that SNP on miRNA binding. The absolute value of $\Delta\Delta$G (|$\Delta\Delta$G|) was used to avoid positive and negative $\Delta\Delta$G values negating each other.

Table 33: List of 20 candidate DNA repair genes investigated for putative miRNA-binding site SNPs within the gene 3'UTR.

Gene symbol	Gene Name
ATM	Ataxia telangiectasia mutated
ATRX (RAD54)	Alpha thalassemia/mental retardation syndrome X-linked (RAD54 homolog, S. cerevisiae)
BRCA1	Breast cancer susceptibility protein 1
BRCA2	Breast cancer susceptibility protein 2
CETN2	Centrin 2
FEN-1	Flap structure-specific endonuclease 1
LIG1	DNA ligase I
LIG3	DNA ligase III
LIG4	DNA ligase IV
MRE11	Meiotic recombination 11 homolog A (S. cerevisiae)
NBS1	Nijmegen breakage syndrome protein 1
PARP1	Poly (ADP-ribose) polymerase family, member 1
PARP2	Poly (ADP-ribose) polymerase family, member 2
PCNA	Proliferating cell nuclear antigen
DNA-PK	DNA-dependent protein kinase catalytic subunit
RAD23B (HR23B)	RAD23 homolog B (S. cerevisiae)
RAD50	RAD50 homolog (S. cerevisiae)
RAD51	RAD51 homolog (RecA homolog, E. coli) (S. cerevisiae)
XPC	Xeroderma pigmentosum complementation group C
XRCC1	X-ray repair complementing defective repair in Chinese hamster cells 1 (H. sapiens)

Of the 20 genes, five genes, *ATRX*, *CETN2*, *DNA-PK, RAD50* and *XRCC1,* had no SNPs located within their 3'UTR, while the 3'UTR SNPs in *LIG1* and *PCNA* were not sited within a predicted miRNA-binding site. Of the remainder, there were 21 common SNPs located within a predicted miRNA-binding site in 13 genes (Appendix D). Ten of these SNPs had a $|\Delta\Delta Gtot|$ greater than 5 kJ/mol. Of these 10 SNPs, two had been previously genotyped in the bladder cancer case-control study, namely *NBS1* rs1063054 (Choudhury *et al.* 2008) and *XPC* rs2229090 (Sak *et al.* 2006) and showed no association. *XPC* rs2229090 was also subsequently imputed from Hapmap phase 2 variants in the Nijmegen and Icelandic bladder cancer study populations (B Kiemeney, personal communication), with no association found with bladder cancer risk (combined OR 0.99 [95%CI 0.89-1.09], P=0.80) despite this

variant having the highest predicted |ΔΔGtot| of 113.27 for differential miRNA binding. The remaining eight candidate SNPs and corresponding predicted miRNA-binding sites with a |ΔΔG| greater than 5 kJ/moL are listed in Table 33.

8.3.1.3. Candidate SNP genotyping and bladder cancer risk

Genomic DNA samples from the bladder cancer case-control series were sent for high-throughput Taqman genotyping to Gen-Probe Life Sciences in West Lothian, Scotland. Blind duplicates from five percent of samples were included to ensure concordance of genotyping calls. Custom Taqman assays for *ATM* SNPs rs1137918 and rs227091 failed genotyping and these SNPs were found to be located within Alu repeats. Capillary sequencing for *ATM* rs1137918 in 20 genomic control samples (undertaken by Dr C Taylor) failed to identify this common variant allele (dbSNP MAF = 0.22 reported by Sequenom in 92 individuals from a Caucasian population) making it very unlikely that this SNP truly exists.

ATM rs227091 was successfully genotyped by Taqman genotyping of a nested PCR product for each sample (undertaken by Dr C Taylor), with full concordance in replicates and an assay failure rate of 6.25%. The remaining six candidate SNPs were genotyped with greater than 98% concordance in replicates and an assay failure rate of less than 5%. Non-concordant replicates were due to exhaustion of these DNA samples and so were dropped from analysis. Hardy-Weinberg equilibrium for all SNPs was maintained in the controls.

Table 34: Candidate DNA repair gene miRNA-binding site SNPs selected for genotyping in the bladder cancer case-control study population. Predicted miRNA-binding with ΔΔG greater than 3 kJ/mol are listed.

Gene, dbSNP ID & Allele Substitution	MAF	Predicted miRNA binding	ΔΔG[a]	$\|\Delta\Delta G_{tot}\|$[b]
LIG3 rs4796030 C>A	0.47	miR-612	-5.36	42.17
		miR-423-3p	-5.23	
		miR-346	-5.17	
		miR-221	-4.49	
		miR-888*	-4.36	
		miR-512-5p	-3.90	
		miR-615-3p	-3.70	
		miR-222	-3.44	
		miR-525-3p	-3.30	
		miR-508-5p	-3.22	
ATM rs227091 T>C	0.44	miR-217	-5.41	38.91
		miR-338-3p	-5.37	
		miR-199b-5p	-4.79	
		miR-199a-5p	-4.70	
		miR-24	-3.83	
		miR-593*	-3.49	
		miR-134	-3.26	
		miR-196a	3.34	
		miR-939	4.72	
BRCA1 rs12516 C>T	0.36	miR-874	-9.92	25.76
		miR-324-3p	-8.83	
		miR-623	-7.01	
BRCA1 rs8176318 G>T	0.35	miR-328	-4.10	21.82
		miR-565	-3.95	
		miR-149	-3.71	
		miR-146b-3p	-3.49	
		miR-345	-3.30	
		miR-892b	-3.27	
PARP1 rs8679 T>C	0.17	miR-145	-5.71	17.42
		miR-105	-4.91	
		miR-630	-3.62	
		miR-302a	3.18	
ATM rs1137918 A>G	0.22	miR-615-5p	-5.14	12.80
		miR-193a-5p	-3.69	
		miR-939	3.97	
NBN (NBS1) rs2735383 G>C	0.32	miR-499	-4.36	8.70
		miR-508-3p	-4.34	
RAD51 rs7180135 A>G	0.47	miR-197	6.87	6.87

[a]The ΔΔG was calculated by deducting the variant ΔG from the wildtype ΔG. A negative ΔΔG indicates decreased binding in the variant compared to the wildtype while a positive ΔΔG indicates increased binding in the variant.
[b]$|\Delta\Delta G_{tot}|$ is the total of the absolute values of ΔΔG (|ΔΔG|).

Multivariate logistic regression of the seven SNPs for association with bladder cancer risk adjusting for gender, age, smoking status and occupational exposures showed that only *PARP1* rs8679 (P_{trend} = 0.05) was associated with increased bladder cancer risk (Table 34). Collaborators (L Vaslin, DG Cox and J Hall) at Institut Curie, Orsay, France, also genotyped *PARP1* rs8679 in 257 breast cancer cases and 512 controls but did not identify an association with increased breast cancer risk (P_{trend} = 0.45) though this population was underpowered (24% power at 5% significance level) to detect an equivalent effect size. However, CC homozygotes were noted to have a greater risk of developing breast cancer compared with TT (adjusted OR 1.90 [95% CI 1.05-3.43], P=0.03, data not shown).

Visual inspection of Table 34 suggested that the dominant genotype association model may better describe the gene effects of *PARP1* rs8679 rather than the additive genotype association model used here with the ORs for CC homozygotes and CT heterozygotes being similar. Under a dominant model, *PARP1* rs8679 (CT + CC genotypes) as well as *LIG3* rs4796030 (AC + AA genotypes) were associated with increased bladder cancer risk (OR 1.28 [95%CI 1.02-1.62], P=0.03 and OR 1.26 [95%CI 1.00-1.58], P=0.05 respectively). The combined effects of *LIG3* rs4796030 and *PARP1* rs8679 was assessed by collapsing unfavourable genotypes (AC + AA and CT + CC genotypes respectively) for analysis on bladder cancer risk (Table 35). This demonstrated an additive genotype-dose effect with the number of unfavourable SNP genotype groups carried (P_{trend}=0.002). These results, however, have to be interpreted with care in view of the potential application of the incorrect genetic model and the use of multiple comparisons.

Table 35: Multivariate logistic regression analysis of candidate DNA repair gene 3'UTR SNPs on bladder cancer risk.

Candidate SNP	Observed MAF (Reported MAF)	Genotype	Case	Control	OR[a] (95% CI)	P_{trend}[b,c]
LIG3 rs4796030 C>A	0.46 (0.47)	CC	204	226	1	0.15
		AC	369	321	1.28 (1.00 – 1.63)	
		AA	160	148	1.21 (0.90 – 1.64)	
		AC + AA	529	469	1.26 (1.00 – 1.58)	**0.05**[d]
ATM rs227091 T>C	0.45 (0.44)	TT	201	209	1	0.70
		CT	320	341	0.92 (0.69 – 1.21)	
		CC	146	148	0.93 (0.69 – 1.27)	
		CT + CC	466	488	0.92 (0.71 – 1.23)	0.55[d]
BRCA1 rs12516 C>T	0.33 (0.36)	CC	342	292	1	0.43
		CT	297	318	0.80 (0.64 – 1.01)	
		TT	79	68	1.04 (0.72 – 1.50)	
		CT + TT	376	386	0.84 (0.68 – 1.05)	0.12[d]
PARP1 rs8679 T>C	0.21 (0.21)	TT	412	424	1	**0.05**
		CT	266	214	1.29 (1.02 – 1.62)	
		CC	45	37	1.23 (0.77 – 1.95)	
		CT + CC	311	251	1.28 (1.02 – 1.62)	**0.03**[d]
BRCA1 rs8176318 G>T	0.32 (0.35)	GG	349	296	1	0.29
		GT	303	321	0.80 (0.64 – 1.00)	
		TT	78	70	0.98 (0.68 – 1.41)	
		GT + TT	381	391	0.83 (0.67 – 1.03)	0.09[d]
NBS1 rs2735383 G>C	0.32 (0.32)	GG	305	313	1	0.27
		CG	337	308	1.09 (0.87 – 1.36)	
		CC	69	59	1.22 (0.83 – 1.81)	
		CG + CC	406	367	1.11 (0.89 – 1.38)	0.35[d]
RAD51 rs7180135 A>G	0.44 (0.47)	AA	224	237	1	0.97
		AG	369	327	1.06 (0.80 – 1.41)	
		GG	145	135	0.90 (0.66 – 1.21)	
		AG + GG	514	462	0.99 (0.76 – 1.30)	0.97[d]

[a] Adjusted for age, gender, smoking status and occupational exposure.
[b] Statistically significant p-values ($P \leq 0.05$) are in bold.
[c] Additive model of genetic inheritance.
[d] Dominant model of genetic inheritance.

Table 36: Multivariate logistic regression analysis of number of unfavourable LIG3 rs4796030 and PARP1 rs8679 SNP genotypes on bladder cancer risk.

No. of unfavourable genotypes carried[a]	Case	Control	OR[b] (95% CI)	P_{trend}[c]
0	115	145	1	
1	375	352	1.33 (0.99-1.78)	**0.002**
2	220	169	1.65 (1.20-2.29)	

[a]Unfavourable genotypes defined as *LIG3* rs4796030 AC + AA and *PARP1* rs8679 CT + CC.
[b] Adjusted for age, gender, smoking status and occupational exposure.
[c] Statistically significant p-values (P ≤ 0.05) are in bold.

8.3.2. DNA repair gene miRNA-binding site SNPs and MIBC radiotherapy outcomes

8.3.2.1. Radiotherapy cohort demographics

The MIBC radiotherapy cohort consisted of 199 muscle-invasive bladder cancer cases treated with 3D-conformal external beam radiotherapy (52.5 – 55 Gy in 20 fractions over four weeks) between August 2002 and October 2009 at Cookridge Hospital and St James Institute of Oncology, Leeds, UK. Median follow-up time was 67.6 months (range: 20.0 – 97.1 months). For 100 cancer-specific events (assuming an overall 50% 5-year CSS, this study had a statistical power of 80% of detecting a hazard ratio (HR) of 2.00 in rare homozygotes and heterozygotes combined for a MAF of 0.2 at a 5% significance level.

As shown in Table 36, MIBC patients treated by RT are predominantly male and tend to be an elderly population (median 78.5 years old), as would be expected in bladder cancer but also partly due to fit (hence, typically younger) patients being offered the option of radical cystectomy. Hydronephrosis rates in this population were much higher than those reported in recent published MIBC RT trials but this could be due to poor renal function being an exclusion criteria in these trials (Hoskin *et al.* 2010; James *et al.* 2012). However, these rates were similar to previous retrospective RT series (Scrimger *et al.* 2001; Tran *et al.* 2009). As neo-adjuvant chemotherapy has

only been introduced recently to standard guidelines, only a small proportion of the population received this treatment while concurrent chemotherapy or radiosensitisers were only used as part of clinical trials. Eleven percent of the cohort had local persistent disease or relapse requiring salvage cystectomy and the 5-year CSS for the whole population was 51.6%. Cox multivariate analysis of CSS identified only tumour (P=0.04) and nodal stage (P=0.05) as independent poor prognostic factors. Neo-adjuvant chemotherapy, concurrent chemotherapy or radiosensitiser use was associated with better 5-year CSS outcomes compared to without (60.6% vs 51.9% respectively, P=0.04).

Table 37: Radiotherapy cohort clinical characteristics and Cox multivariate survival analysis of the contribution of each variable to cancer-specific survival.

Clinical Characteristics	Study population No of patients (%)	Cancer-specific Survival	
		HR (95% CI)	P-value[c]
Age (years)			
Median (range)	78.5 (51 – 92.5)	1.02 (0.98-1.05)[d]	0.37
Gender			
Female	53 (26.6)	1	
Male	146 (73.4)	1.22 (0.70-2.12)	0.49
Tumour stage			
T2	112 (56.3)	1	
T3	76 (38.2)	1.54 (1.02-2.35)	**0.04**
T4	11 (5.5)		
Nodal stage			
N0	192 (96.5)	1	
N1	4 (2.0)	2.23 (1.00-5.00)	**0.05**
N2	2 (1.0)		
Nx	1 (0.5)		
Histological grade			
<G3	14 (7.0)	1	
G3	178 (89.5)	0.93 (0.61-1.41)	0.73
Gx	7 (3.5)		
Hydronephrosis			
No	147 (73.9)	1	
Yes	52 (26.1)	1.22 (0.68-2.20)	0.51
Neoadjuvant/ concurrent therapy			
None	168 (84.5)	1	
Neoadjuvant chemotherapy	15 (7.5)[a]	0.39 (0.16-0.95)[e]	**0.04** [e]
Concurrent chemotherapy/ radiosensitizer	16 (8.0)[b]		
Salvage chemotherapy			
Not Received	189 (95.0)		
Received	10 (5.0)		
Salvage cystectomy			
Not Received	177 (88.9)		
Received	22 (11.1)		

[a] All received platinum based combination chemotherapy.
[b] 11 patients received concurrent gemcitabine (100 mg/m^2) weekly ×4 and 5 patients received concurrent carbogen and nicotinamide as part of a phase II clinical trial and as part of the BCON phase III clinical trial respectively.
[c] Statistically significant p-values (P ≤ 0.05) are in bold.
[d] Hazard ratio per-year increase in age.
[e] Hazard ratio calculated for receiving any form of chemotherapy or radiosensitizer.

8.3.2.2. Candidate SNP genotyping and association with clinical outcomes

Candidate DNA repair gene 3'UTR SNPs were genotyped in the RT cohort and correlated with RT clinical outcomes. Carriers of at least one *RAD51* rs7180135 minor allele (AG + GG) were found to have significantly better 5-year CSS rates compared to common homozygotes (AA) ((Figure 32), 56.2% vs 35.1% respectively, log-rank P=0.02). This remained true even after adjusting for covariates on multivariate analysis (HR 0.52 [95%CI 0.31-0.87], P=0.01) (Table 37). A dominant model was used, as like the case-control study, this model best described the genetic effects observed (AG: HR 0.51 [95%CI 0.29-0.89]; GG: HR 0.55 [95%CI 0.28-1.05]). There were no associations found with late radiotherapy bladder or bowel toxicity.

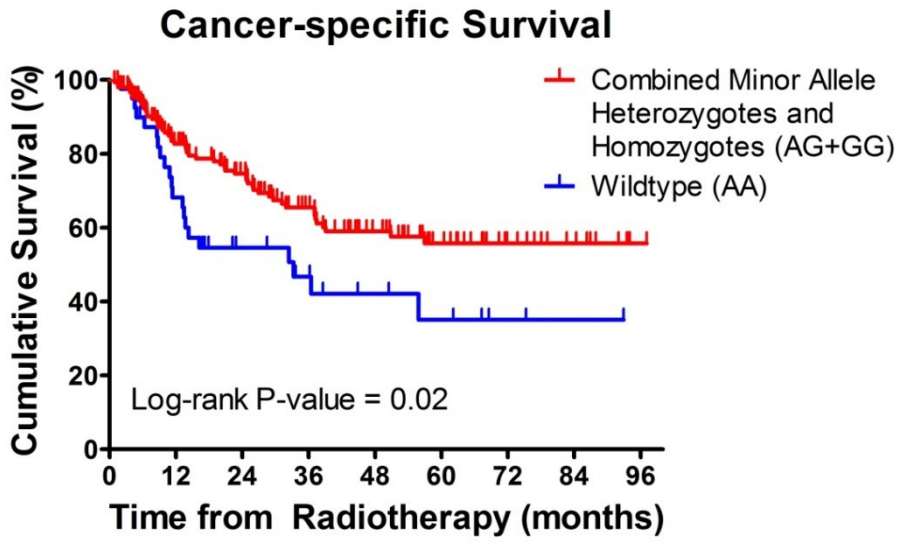


Figure 32: Kaplan-Meier graph of cancer-specific survival in combined heterozygotes and rare homozygotes of *RAD51* rs7180135 (AG+GG) versus wildtype (AA).

Table 38: Cox multivariate analysis of cancer-specific survival and logistic regression for late radiation normal tissue toxicity following radical radiotherapy and associations with candidate SNPs using a dominant model.

Candidate SNP	Genotype		Cancer specific survival		Late Bladder Toxicity (RTOG 2+)		Late Bowel Toxicity (RTOG 2+)	
		N	HR (95% CI)[a]	P-value[b]	OR (95%CI)	P-value[b]	OR (95%CI)	P-value[b]
LIG3 rs4796030 C>A	CC	56	1		1		1	
	AC + AA	129	1.32 (0.77-2.26)	0.31	1.35 (0.68-2.70)	0.39	1.98 (0.64-6.15)	0.24
ATM rs227091 T>C	TT	52	1		1		1	
	CT + CC	133	1.34 (0.71-2.54)	0.37	0.73 (0.35-1.54)	0.41	0.82 (0.28-2.38)	0.71
BRCA1 rs12516 C>T	CC	85	1		1		1	
	CT + TT	101	0.84 (0.53-1.34)	0.47	0.87 (0.47-1.60)	0.65	1.45 (0.58-3.63)	0.43
PARP1 rs8679 T>C	TT	117	1		1		1	
	CT + CC	71	1.14 (0.71-1.86)	0.57	0.54 (0.28-1.04)	0.06	0.74 (0.28-1.90)	0.53
BRCA1 rs8176318 G>T	GG	87	1		1		1	
	GT + TT	93	0.90 (0.56-1.44)	0.66	0.92 (0.50-1.70)	0.80	1.90 (0.74-4.91)	0.18
NBS1 rs2735383 G>C	GG	81	1		1		1	
	CG + CC	101	0.83 (0.52-1.34)	0.44	1.60 (0.84-3.00)	0.15	1.35 (0.54-3.39)	0.52
RAD51 rs7180135 A>G	AA	50	1		1		1	
	AG + GG	139	0.51 (0.30-0.86)	**0.01**	1.08 (0.50-2.30)	0.85	1.76 (0.49-6.28)	0.38

[a] Adjusted for age, gender, tumour and nodal stage, tumour grade, hydronephrosis and neoadjuvant or concurrent chemotherapy or radiosensitiser.
[b] Statistically significant p-values (P ≤ 0.05) are in bold.

8.3.3. Discussion

8.3.3.1. DNA repair gene miRNA-binding site 3'UTR SNPs and bladder cancer susceptibility

This was the first published study demonstrating an effect of DNA repair gene 3'UTR SNPs within putative miRNA-binding sites on bladder cancer susceptibility (Teo *et al.* 2012). *PARP1* rs8679 and possibly *LIG3* rs4796030 were found to be associated with increased bladder cancer risk, with an additive effect seen with carriage of both SNP minor alleles. Further validation of these findings in an independent bladder cancer case-control population will be needed.

LIG3 is a BER protein and is stabilised by XRCC1. It is recruited through interactions with PARP1 and XRCC1 to DNA SSBs and ligates the SSBs generated during short patch BER. It may also act as a "back-up" ligase for long-patch BER (Petermann *et al.* 2006). PARP1, another BER protein, carries out the poly(ADP-ribosyl)ation of histones, topoisomerases and DNA repair signalling proteins following DNA damage, resulting in chromatin unwinding and recruitment of DNA repair proteins (Megnin-Chanet *et al.* 2010). PARP1 is also involved in the error-prone alternative NHEJ pathway, MMEJ, together with LIG3 and XRCC1 (Audebert *et al.* 2004). The $\Delta\Delta$Gs for the miRNAs predicted to bind to *LIG3* rs4796030 and *PARP1* rs8679 (Table 33) were predominantly negative, indicating decreased miRNA-mRNA binding in the variant alleles compared to the wildtype alleles. Thus, the variant allele would be expected to be associated with increased LIG3 and PARP1 expression. This result is counter-intuitive as it would be expected that higher levels of DNA repair enzymes would be associated with increased repair and reduced cancer risk. However, DNA repair enzyme over-expression in tumours may provide a selective advantage for tumour cell growth with over-expression having been associated with a greater risk of metastasis (Sarasin and Kauffmann 2008). The additive effect seen in our study of

carrying both *LIG3* rs4796030 and *PARP1* rs8679 on bladder cancer risk suggests that either an enhanced BER pathway or MMEJ repair pathway predisposes individuals to bladder cancer. PARP1 competes with Ku (classical NHEJ) for binding to DSB ends and initiating the MMEJ pathway (Wang *et al.* 2006). Increased PARP1 and LIG3 expression due to *LIG3* rs4796030 and *PARP1* rs8679 could thus possibly promote DSB repair via the MMEJ over the classical NHEJ pathway resulting in more error-prone repair and mutagenesis, with the greatest influence seen in carriers of both SNPs.

From the published literature of the miRNAs predicted to bind to *LIG3* rs4796030, a pre-miRNA SNP of miR-423 has previously been associated with increased bladder cancer risk (Yang *et al.* 2008) while miR-221 has been reported to be upregulated in bladder cancer cell lines with miR-221 silencing leading to apoptosis (Lu *et al.* 2010). For the *PARP1* rs8679 SNP, miR-145 binding was predicted to have the greatest $\Delta\Delta G$ (-5.71kJ/mol) (Table 33), indicating decreased miR-145 binding to the variant *PARP1* mRNA 3'UTR and thus increased PARP1 expression. Previous studies have reported down-regulation of miR-145 expression in bladder cancer (Catto *et al.* 2009; Dyrskjot *et al.* 2009), while transfection of miR-145 in bladder cancer cell lines had a tumour suppressor effect, reducing cell proliferation and inducing apoptotic pathways (Ostenfeld *et al.* 2010).

Candidate gene association studies have identified cancer risk associations with 3'UTR SNPs which were then investigated for possible miRNA-binding sites (Liang *et al.* 2010; Pharoah *et al.* 2011). In bladder cancer, DSB repair *MRE11* rs2155209 3'UTR SNP was previously found to increase bladder cancer risk (Choudhury *et al.* 2008). This variant was predicted to be within the binding sites for miR-584 and miR-744 but the minor allele is predicted to alter the $\Delta\Delta G$ by less than 1 kJ/mol for both

miRNAs (Table 38 and Appendix D), thus the mechanism underlying this association remains to be determined.

Table 39: Predicted miRNA bindings sites and predicted ΔG values for *MRE11A* rs2155209.

Gene, dbSNP ID & Allele Substitution	Minor Allele Frequency	Predicted miRNA binding	Wildtype Allele ΔG	Variant Allele ΔG	ΔΔG	|ΔΔG|
MRE11A rs2155209 A>G	0.29	miR-744	-20.99	-20.02	-0.97	0.97
		miR-584	-18.6	-18.35	-0.25	0.25

Other studies also used *in silico* methods to try identify SNPs located within predicted miRNA-binding sites, then tested for associations with cancer susceptibility (Chin *et al.* 2008; Landi *et al.* 2008; Nicoloso *et al.* 2010). Using the same bioinformatics analysis methods, Landi *et al* reported an association between the *CD86* 3'UTR SNP rs17281995, predicted to affect the binding of four miRNAs, with increased colorectal cancer risk (Landi *et al.* 2008) while Naccarati *et al* identified two NER 3'UTR SNPs, *GTF2H1* rs4596 and *RPA2* rs7356, to be protective and to predispose to colorectal cancer respectively (Naccarati *et al.* 2012). Chin *et al* demonstrated that *KRAS* 3'UTR SNP rs712 was sited within a *let-7* binding site, resulted in increased luciferase activity compared to wild-type in a *KRAS* 3'UTR reporter assay, and was associated with increased lung cancer risk (Chin *et al.* 2008). Nicoloso *et al* identified two predicted miRNA-binding site SNPs within the coding regions of *BRCA1* and *TGFR1* that affected post-transcriptional miRNA regulation (Nicoloso *et al.* 2010), thus highlighting the possibility of miRNA-binding sites beyond the 3'UTR.

8.3.3.2. DNA repair gene miRNA-binding site SNPs as markers of MIBC radiotherapy outcomes

As ionising radiation from radiotherapy causes cancer cell death by inducing DNA damage, the candidate SNPs were also investigated as potential predictive markers of radiotherapy treatment outcomes in muscle invasive bladder cancer. Carriers of the *RAD51* rs7180135 G minor allele were found to have an improved 5-year CSS with an absolute survival advantage at 5 years of 21.1% indicating a potential role as a predictive marker of RT treatment outcomes in MIBC.

However, local recurrence would be a better measure for a local treatment such as radiotherapy. An accurate measure of local recurrence would be influenced by response at first check cystoscopy three months post-radiotherapy and frequency of follow-up cystoscopy. However, a high number of patients did not have their first cystoscopy performed or their findings documented, plus the frequency of follow-up cystoscopy was variable in the referring urology units in the region. Thus, analysis was performed on CSS instead but this measure would include deaths from metastatic disease. *RAD51* rs7180135 could thus potentially be just a measure of a poorer prognostic phenotype and not a predictive marker of RT. This SNP will thus need to be validated in an independent RT treated as well as a surgically treated cohort to clarify its clinical value.

It is hypothesised that miR-197 binds more strongly to the *RAD51* rs7180135 G allele ($\Delta\Delta$G 6.87 kJ/mol) resulting in a reduction of RAD51 expression. RAD51 is involved in DSB homologous recombination repair, so in tumours, alteration of RAD51 expression could potentiate tumour radiosensitisation. In breast cancer, low tumour *RAD51* mRNA expression has been associated with lower local recurrence and improved survival following adjuvant chemotherapy and radiotherapy treatment (Le

Scodan *et al.* 2010). Differential expression of miR-197 has been reported in different cancer types with up-regulation seen in squamous cell carcinoma of the tongue and male breast carcinoma, and down-regulation in gastric cancer (Wong *et al.* 2008; Lehmann *et al.* 2010; Li *et al.* 2011). Plasma miR-197 levels have also been proposed as a potential diagnostic biomarker, being elevated in lung cancer (Zheng *et al.* 2011). In human colon cancer cell lines, miR-197 expression levels were found to be down-regulated following oxaliplatin and 5-fluorouracil treatment (Zhou *et al.* 2010). In response to DNA damage, the effect of miR-197 down-regulation in cancer cells could thus be to increase DNA repair by increasing RAD51 expression. The current results suggest that the *RAD51* rs7180135 genotype and its possible effects on miR-197-mRNA-binding may predict tumour radiosensitivity and radiotherapy outcomes in bladder cancer. If successfully validated, this might be used clinically as a predictive marker of radiotherapy outcome.

8.3.3.3. Study strengths and limitations

This was a large study able to detect moderate to large effects of the candidate common SNPs on bladder cancer risk and RT outcomes in MIBC. Further validation is needed in an independent bladder cancer case-control population and RT-treated MIBC population. This study may potentially have missed other miRNA-binding site SNPs within the candidate genes. The bioinformatics software used for predicting miRNA-binding sites were based purely on canonical “seed pairing rules” hence non-canonical miRNA-binding sites would have been missed. Also by only examining SNPs within the 3’UTR, potential miRNA-binding sites within the 5’UTR or coding regions would not have been detected.

This study examined seven polymorphisms for associations with risk of bladder cancer and cause-specific mortality following RT using multiple genetic models, thus

runs the risk of false positive results due to multiple comparisons. If a simple Bonferroni correction were applied for the number of SNPs genotyped, the significance threshold would be 0.007 thus none of the associations observed would remain statistically significant. However, with the complex correlations between these candidate genes, such as gene-gene interactions and linkage disequilibrium, influencing these multiple comparisons, it is difficult to determine the true number of tests carried out for simple Bonferroni correction. All the SNPs were chosen based on their prior probability of functional significance, due to their predicted influence on miRNA-binding. However, the magnitude of this prior probability would be difficult to estimate for use in correction methods such as the false positive reporting probability (FPRP) (Wacholder *et al.* 2004). Due to the subjectivity of the assumptions necessary, no formal multiple comparison corrections were used for the p-values reported, hence I acknowledge the possibility of false-positive discovery.

In summary, this project examined the 3'UTR of 20 candidate DNA repair genes using *in silico* methods to identify SNPs that potentially affect miRNA-binding. Associations between *PARP1* rs8679 with increased bladder cancer risk, and *RAD51* rs7180135 with improved cancer-specific survival following radiotherapy treatment in muscle-invasive bladder cancer were observed. Whilst the first of its kind in bladder cancer, this project also corroborates the findings seen in other cancer sites of the association of miRNA-binding site SNPs with cancer susceptibility (Chin *et al.* 2008; Landi *et al.* 2008; Liang *et al.* 2010; Nicoloso *et al.* 2010). These findings will need to be successfully validated in other bladder cancer series and the functional effects of these SNPs elicited. This will provide further insight into the biology of bladder carcinogenesis, potential predictive biomarker of RT outcomes, and potential novel targets for new cancer treatments.

8.4. Germline *MRE11A* variants as a biomarker of radiotherapy outcomes in muscle-invasive bladder cancer

In MIBC, Dr Kiltie's group had previously shown tumour MRE11 protein expression to be predictive of survival following RT but not surgery (Choudhury *et al.* 2010), and that a common germline *MRE11A* SNP was associated with an increased risk of developing bladder cancer (Choudhury *et al.* 2008). *MRE11A* SNPs had also been shown to be associated with altered DSB repair response (Ricceri *et al.* 2011). As tumours initially inherit the host genotype (Stewart 2010), it was hypothesised that germline SNPs and rare variants in *MRE11A* may potentially be predictive of both tumour response and toxicity following radical radiotherapy in MIBC. To test this hypothesis, indexed multiplexed NGS was undertaken to identify germline *MRE11A* variants in MIBC patients treated by radical RT and to assess their association with survival and late radiation normal tissue toxicity. Validation studies were then undertaken in a surgically-treated MIBC cohort.

8.4.1. Radiotherapy treated MIBC cohort demographics

The characteristics and RT treatment details of the full MIBC radiotherapy cohort were described previously in Section 7.2.3.1. However, long PCR amplification failed in 15 samples so these patients were excluded with 70 cancer specific events in the remaining 186 patients. Patient clinical demographics of just the 186 included patients are displayed in Table 39. Therefore, this study was able to detect, with a statistical power of 80% at a 5% significance level, a HR of 2.20 for a MAF (or collapsed rare variant allele frequency) of 0.2, or a HR of 2.50 for a MAF of 0.1.

Table 40: Clinical demographics of patients from the MIBC radiotherapy cohort with successful long PCR amplification of *MRE11A*.

Variable	Study population (N=186) No of patients (%)
Age (years)	
Median (range)	79 (55 - 93)
Gender	
Male	139 (74.7)
Female	47 (25.3)
Tumour stage	
T2	118 (63.5)
T3	51 (27.5)
T4	9 (4.9)
Tx	8 (4.4)
Nodal stage	
N0	179 (96.3)
N1	4 (2.2)
N2	2 (1.1)
Nx	1 (0.6)
Histological Grade	
High grade	167 (89.8)
Low grade	14 (7.6)
Unknown	5 (2.7)
Hydronephrosis	
No	135 (72.6)
Yes	51 (27.5)
Neoadjuvant/ concurrent therapy	
None	160 (86.0)
Neoadjuvant chemotherapy	11 (5.9)[a]
Concurrent chemotherapy/ radiosensitiser	15 (8.1)[b]
Salvage chemotherapy	
Not Received	177 (95.2)
Received	9 (4.9)
Salvage Cystectomy	
Not Received	167 (89.8)
Received	19 (10.3)

[a] All received platinum based combination chemotherapy.
[b] 10 patients received concurrent gemcitabine (100 mg/m^2) weekly ×4 as part of a phase II clinical trial and 5 patients received concurrent carbogen and nicotinamide as part of the BCON phase III clinical trial.

8.4.2. Germline *MRE11A* variants from indexed multiplexed NGS

A total of nine pools of 19 to 24 indexed samples per pool were sequenced on the Illumina GAII platform using 100bp single end reads. There was a median coverage of 2507x per sample (range 673 – 9891) and 95% of target candidate gene regions were sequenced at a coverage greater than 221x per sample.

8.4.2.1. Variants identified

In total, 121 variants were called. Alignment and coverage data were reviewed for variants in regions of poor coverage or with poor alignment consensus between forward and reverse strands (Figure 33), resulting in 85 variants being deemed high confidence calls and 36 variants low confidence calls.

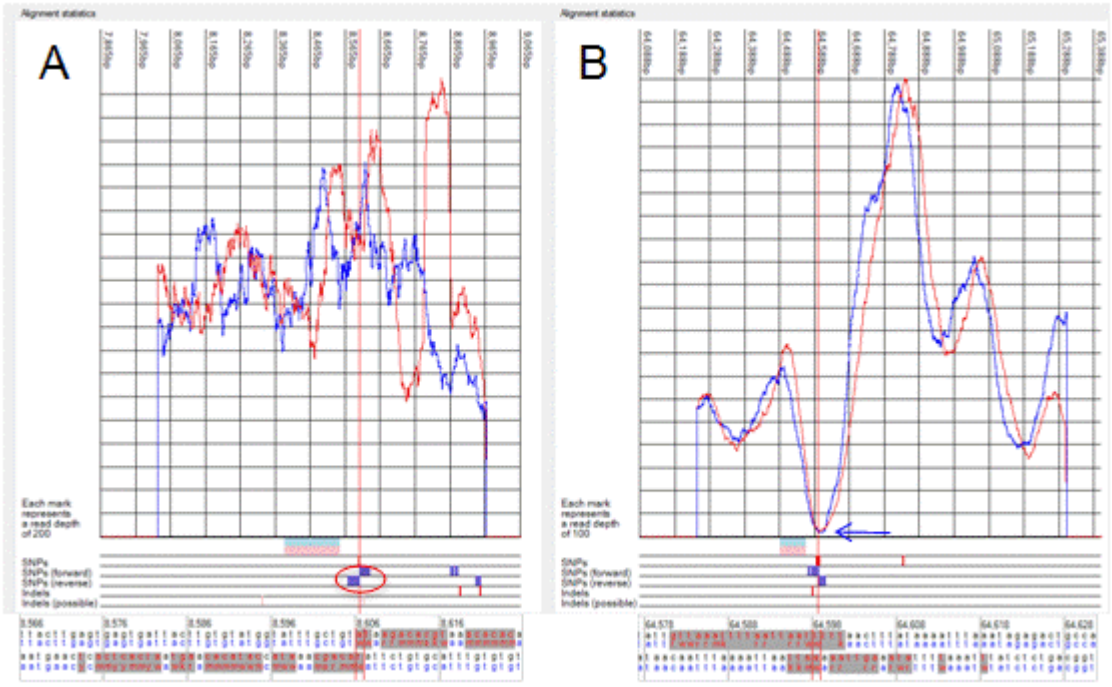


Figure 33: Illuminator alignment and coverage data demonstrating regions of poor confidence calling: A) a variant sited in a region with poor consensus between forward and reverse strands (red circle); B) a variant sited in a region of poor coverage (blue arrow).

One variant carrier sample and one wildtype sample were conventionally Sanger sequenced for all variants to validate the NGS variant calls, with all high confidence

variants but just three low confidence variants confirmed. A summary of confirmed variants by MAF and gene region is shown in Table 40 with equal proportions of SNPs and rare variants identified. The full list of confirmed variants is in Appendix C. Thirty common SNPs were in linkage disequilibrium with R-squared values greater than 0.80 (Figure 35) with nine common SNP haplotype groups observed (Figure 34).

Table 41: Summary of confirmed *MRE11A* variants identified. Variants are divided by MAF, novel variants or variants within the dbSNP database and location within the gene.

	dbSNP	New
Variant type (MAF)	No of variants	No of variants
SNP (>0.01)	40	3
Intronic	*37*	*3*
3'-UTR	*3*	*0*
Rare Variant (≤0.01)	14	31
Intronic	*10*	*25*
Coding	*3*	*1*
3'-UTR	*1*	*5*

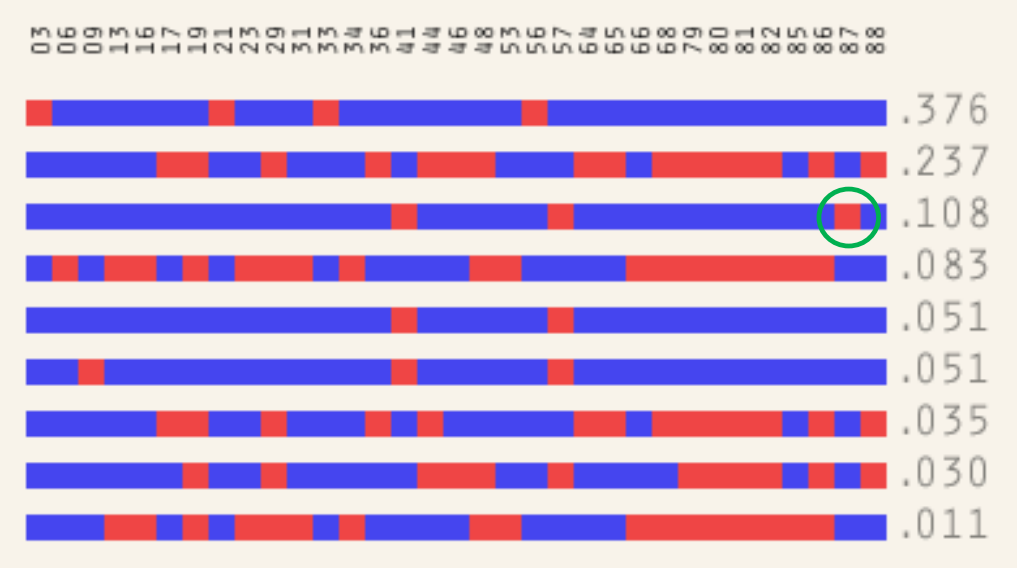


Figure 34: *MRE11A* haplotypes in the study population with observed frequencies shown on the right. Each red square represents carriage of a corresponding SNP minor allele (Haploview ID at the top (see Figure 35)). The rs1805363 A minor allele is circled in green.

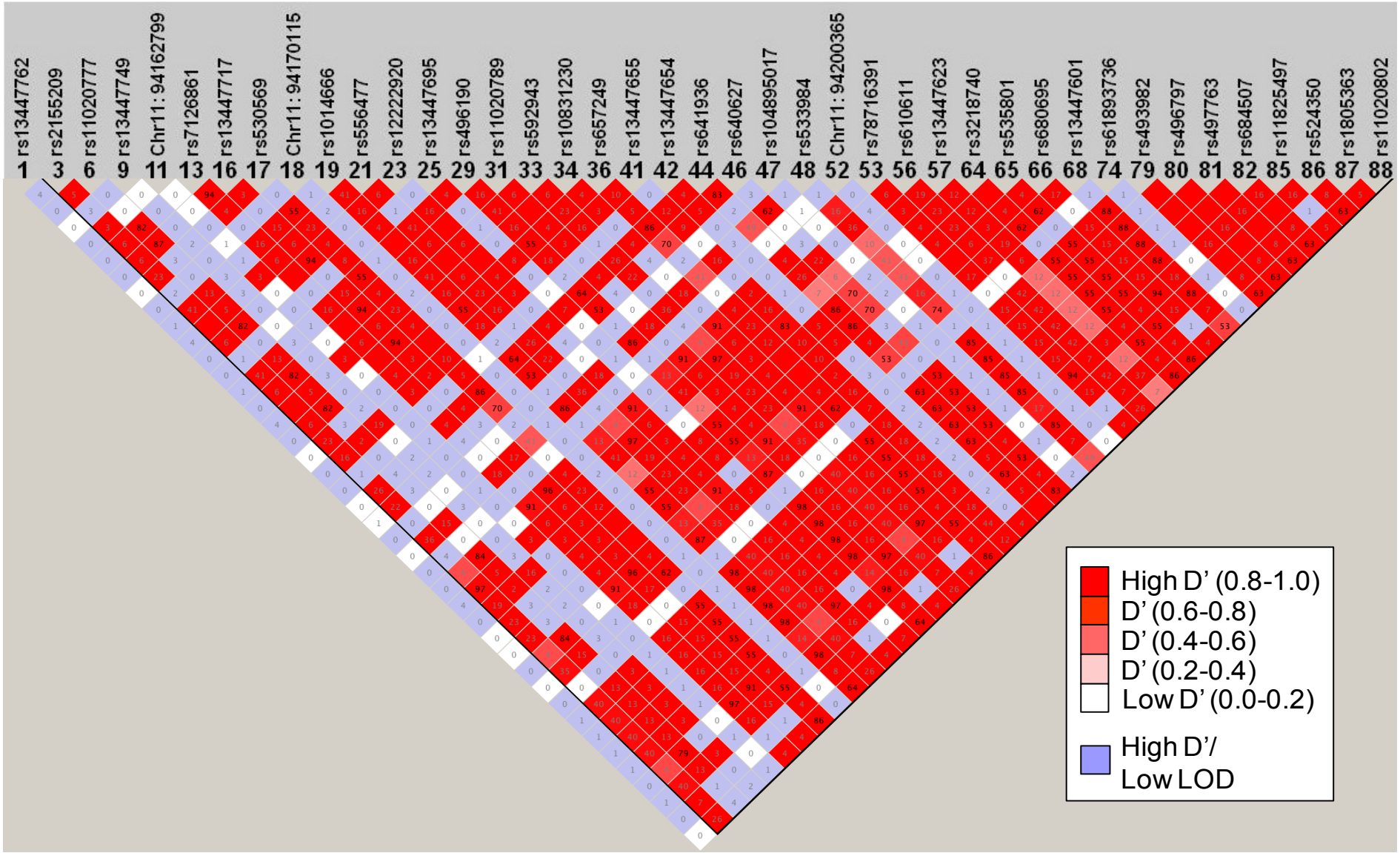


Figure 35: Linkage disequilibrium plot for all common *MRE11A* SNPs (MAF>0.01) generated using Haploview (http://www.broadinstitute.org/) – R-squared values are shown in each square while the heat map (see key) displays the D'-values. In blue are SNPs with a high D' but low logarithm of odds (LOD) thus a low likelihood of a true linkage. Haploview ID and corresponding SNP details are at the top.

8.4.2.2. *Association with cancer-specific survival*

On multivariate Cox proportional hazards analysis, none of the covariates were significantly associated with CSS (Table 41). Individual variant analysis of all *MRE11A* variants detected identified only two common SNPs to be significantly associated with CSS with a $P_{trend} \leq 0.05$, rs1805363 ($P_{trend} = 0.001$) and rs13447623 ($P_{trend} = 0.05$) (Table 41 and in Appendix C). However, the rs1805363 variant A allele was in linkage disequilibrium with the rs13447623 variant G allele (D' = 1.00, r^2 = 0.38, SNP ID 41 in Figure 34). By adjusting for the rs1805363 genotype status, conditional analysis demonstrated that the contribution of rs13447623 on survival was completely dependent on the rs1805363 genotype status (Table 41). Several rare variants, with only one allele seen in the whole study population, attained P-values less than 0.05 individually (Appendix C) but had very wide 95% confidence intervals, hence they were interpreted with caution and used only as part of the collapsed analyses discussed later.

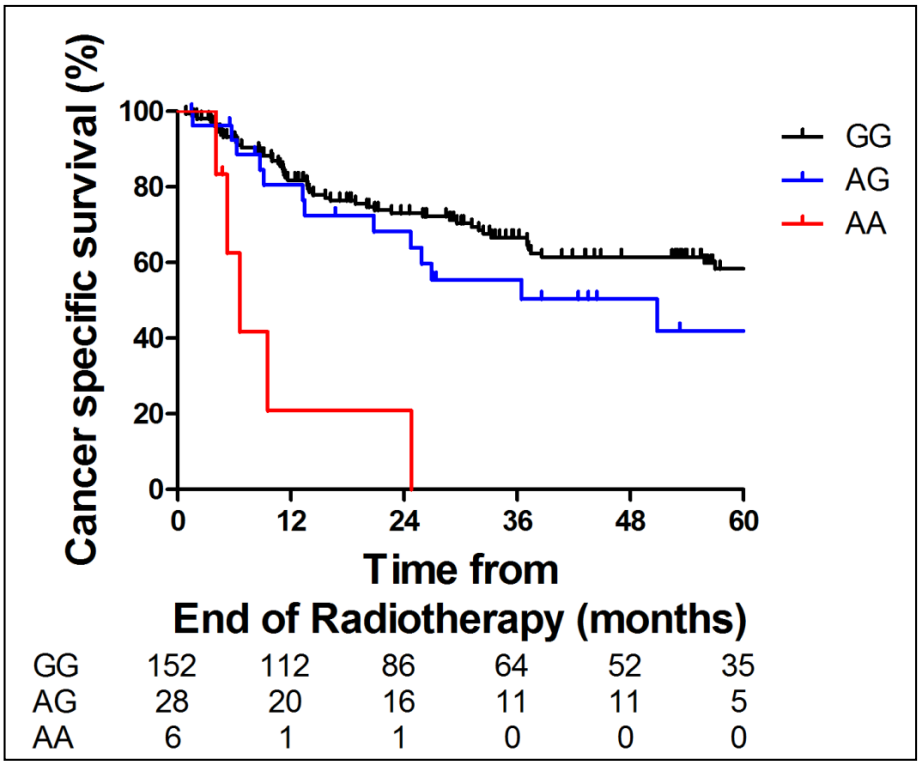


Figure 36: Kaplan Meier curve of cancer-specific survival for *MRE11A* rs1805363 genotypes in the MIBC radiotherapy cohort.

The rs1805363 A minor allele was associated with worse CSS following radiotherapy treatment (Per-allele Hazard Ratio (HR) 2.10, 95% Confidence Interval (CI) 1.34 – 3.28, P_{trend} = 0.001) and a gene-dosage effect was seen with increasing number of A minor alleles carried with a 5-year CSS of 58.3% in wildtypes, 42.0% in heterozygotes and 0% in minor allele homozygotes (Figure 36). This suggests rs1805363 as a potential prognostic marker of bladder cancer survival following radical radiotherapy treatment.

Table 42: Cox proportional hazards multivariate analysis of cancer specific survival in covariates, rs1805363, rs13447623 and 3'UTR rare variants.

Covariates			**HR (95% CI)**	**P-value[a]**
Age at Diagnosis			1.01 (0.98 - 1.04)	0.61
Gender			0.72 (0.39 - 1.31)	0.28
T Stage			1.50 (0.98 - 2.30)	0.06
N Stage			1.96 (0.88 - 4.37)	0.10
Histological Grade			1.12 (0.50 - 2.48)	0.79
Hydronephrosis			1.21 (0.66 - 2.22)	0.53
Neoadjuvant/ Concurrent Chemotherapy			0.52 (0.22 - 1.20)	0.12
Variant (Observed MAF)	**Genotype**	**N**		
rs1805363 G>A (0.11)	GG	152	1	**0.001**
	AG	28	1.49 (0.80 - 2.78)	
	AA	6	8.00 (2.93 - 21.90)	
rs13447623 A>G (0.24)	AA	110	1	**0.05**
	AG	62	0.86 (0.50 - 1.49)	
	GG	14	3.68 (1.78 - 7.60)	
Conditional Analysis				
rs1805363 G>A			2.19 (1.12 - 4.30)	**0.02**
rs13447623 A>G			0.95 (0.54 - 1.66)	0.86
Rare Variants Collapsed Analysis				
No 3'UTR rare variants		179	1	
1+ 3'UTR rare variants		7	4.04 (1.42 - 11.51)	**0.009**

[a] Significant p-values (<0.05) are In bold.

Collapsed analyses were performed by Dr J Nsengimana. Simple collapsed analysis of rare variants a uni-directional analyses looking at the presence or absence of rare variant carriage on survival revealed that carriage of at least one rare variant in the 3'UTR of MRE11 was significantly associated with worse CSS (P = 0.009; Table 41) in carriers compared to non-carriers (5-year CSS: 42.9 vs 54.8% respectively; Figure 37). However, using the data-adaptive sum test as described by Han and Pan (Han and Pan 2010), for a permutation-based collapsed analysis of the uncommon and rare variants (MAF<0.05) in this study taking into account the observed direction of effect for each different variant on survival, did not reveal any significant associations with CSS for all uncommon and rare variants (P = 0.45) or just uncommon and rare exonic variants (P = 0.59). Analysis of the nine common SNP haplotypes shown in Figure 34, did not reveal any haplotype groups associated with CSS (all P>0.05) unexplained by the rs1805363 A minor allele.

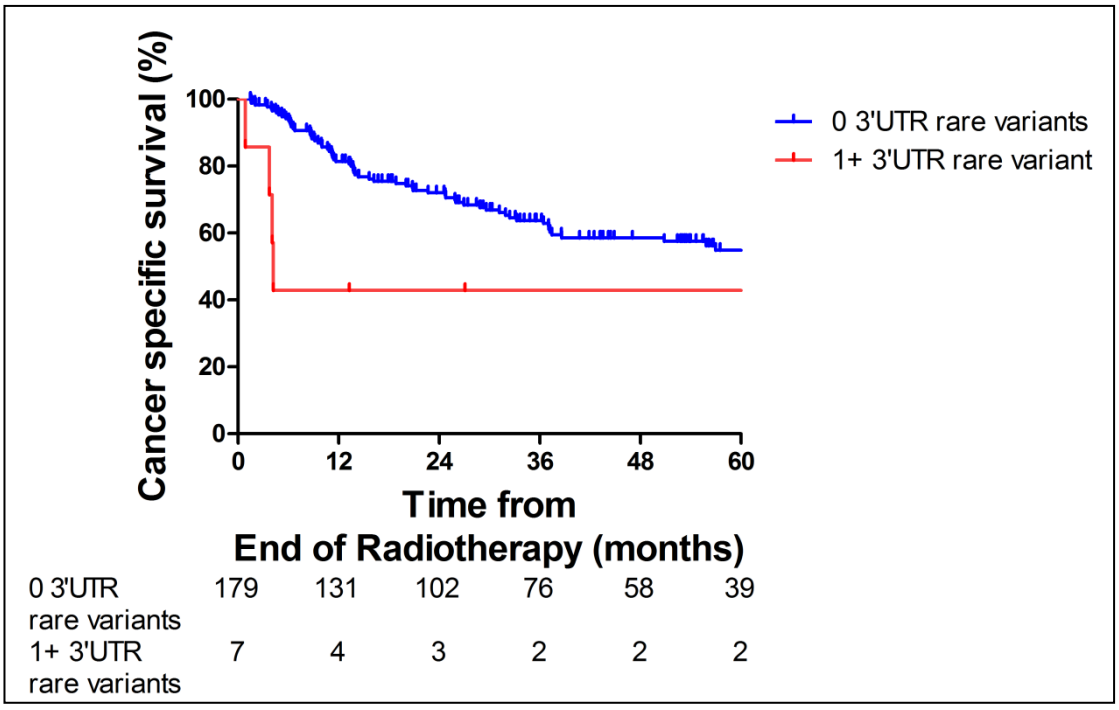


Figure 37: Kaplan Meier curve of cancer-specific survival for carriers of at least one 3'UTR *MRE11A* rare variant allele in the MIBC radiotherapy cohort.

8.4.2.3. Association with late radiation bladder and rectal toxicity

Neither rs1805363 nor the 3'UTR rare variants were associated with developing late radiation bladder (P = 0.09 and 0.33 respectively) or rectal (P = 0.17 for rs1805363; no events for 3'UTR rare variants) toxicity. However, rs13447623 and the common 3'UTR SNP rs2155209, previously found to be associated with increased bladder cancer risk (Choudhury *et al.* 2008), were associated with the likelihood of developing RTOG grade 2 or greater late bladder toxicity and rectal toxicity (Table 42 and in Appendix C for full list). Collapsed analysis did not identify any associations of rare variants with bladder or rectal toxicity (P = 0.71 and P = 0.50 respectively) while there were no associations in the nine common SNP haplotypes unexplained by rs13447623 and rs2155209.

Table 43: Ordered logistic regression of late radiation bladder and rectal toxicity with genotype of rs13447623 and rs2155209.

			Late Bladder Toxicity		Late Rectal Toxicity	
Variant (Observed MAF)	**Genotype**	**N**	**OR (95%CI)**	**P-value**	**OR (95%CI)**	**P-value**
Rs13447623 A>G (0.24)	AA	110	1		1	
	AG	62	3.15 (1.61 - 6.18)	0.003	2.42 (0.94 - 6.22)	0.04
	GG	14	2.69 (0.85 - 8.50)		3.06 (0.72 - 13.01)	
Rs2155209 A>G (0.38)	AA	70	1		1	
	AG	90	0.64 (0.33 - 1.24)	0.02	0.67 (0.28 - 1.63)	0.04
	GG	26	0.27 (0.08 - 0.88)		N/A[a]	

[a] Insufficient events to calculate.

8.4.2.4. Bioinformatics functional predictions

Bioinformatics functional predictions for all *MRE11A* SNPs and variants identified are listed in Appendix C. Three intronic common SNPs including rs1805363 were postulated to have potential effects being sited within a predicted binding site or close proximity to a known splice site (Table 43). Of the four coding rare variants, three

were non-synonymous with only rs61749249 predicted to affect protein function. Two of the coding rare variants and one 3'UTR rare variant were predicted to be sited within a miRNA-binding site with a |ΔΔG| greater than 4kJ/mol.

Table 44: *MRE11A* common SNPs and rare variants with predicted functional effects.

Variant (Observed MAF) [Gene region]	**Coding**			**Predicted miRNA binding**		**Splicing/ Transcription factor binding site**
	Protein Change	*Polyphen*	*SIFT*	*miRNA-binding site*	*ΔΔG (kJ/ mol)*	
Common SNPs						
rs11020802 G>T (0.30) [5'-upstream]		Non-coding		Non-coding		ELF/BRCA1 transcription factor binding site
rs1805363 G>A (0.11) [Intron 1 (Isoform 1) / 5'UTR (Isoform 2)]		Non-coding		Nil		5 bases 3' from Exon 1 GC donor splice site
rs535801 G>A (0.27) [Intron 5]		Non-coding		Non-coding		6 bases 5' from Exon 6 AG acceptor splice site
Rare Variants						
rs3218740 C>T (0.003) [Exon 6]	D142D (Synonymous)			Nil		Nil
rs115244417 C>G (0.003) [Exon 9]	S334R	Benign	Tolerated	Nil		Nil
Chr 11: 94197302 A>G (0.003) [Exon 11]	H401R	Benign	Tolerated	hsa-miR-892a	-4.4	Nil
rs61749249 C>A (0.008) [Exon 13]	A492D	Probably damaging	Affect protein function	hsa-miR-3659	6.6	Nil
rs104895004 G>T (0.003) [Exon 20 3'UTR]		Non-coding		hsa-miR-338-5p	4.4	Nil

8.4.3. Validation of rs1805363

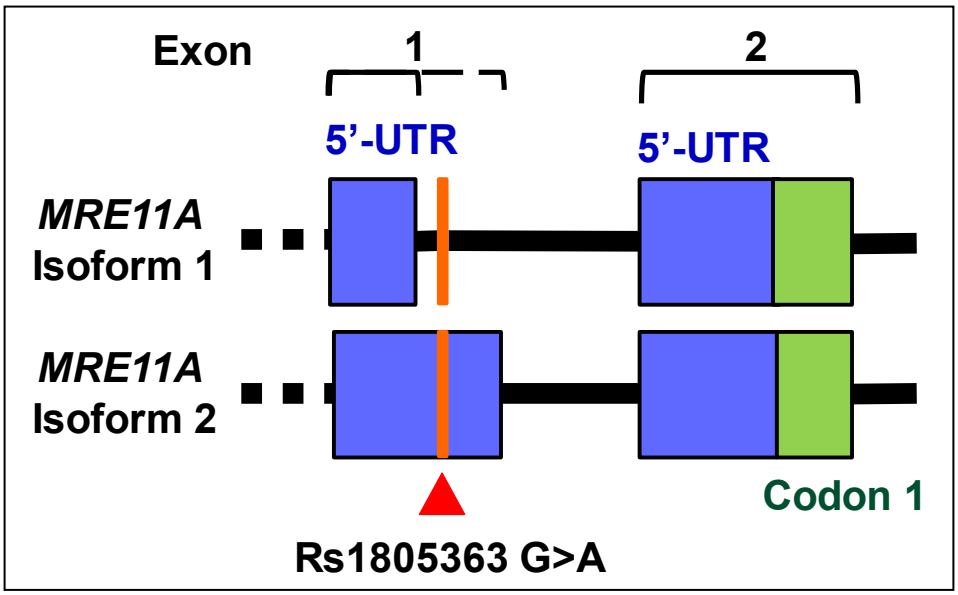


Figure 38: *MRE11A* isoform 1 and isoform 2 alternative splicing and the relative position of common SNP rs1805363.

Rs1805363 is an intronic SNP in *MRE11A* isoform 1, five bases 3' from the Exon 1 AG donor splice site, or is located within the 5'UTR in *MRE11A* isoform 2 (Figure 38). It could thus potentially influence the splicing and/ or the expression of one isoform over the other. However, there were no predicted effects on exonic splice enhancer sites (Figure 39) or on miRNA-binding and post-transcriptional regulation (Table 43), but the accuracy of these prediction tools remains uncertain.

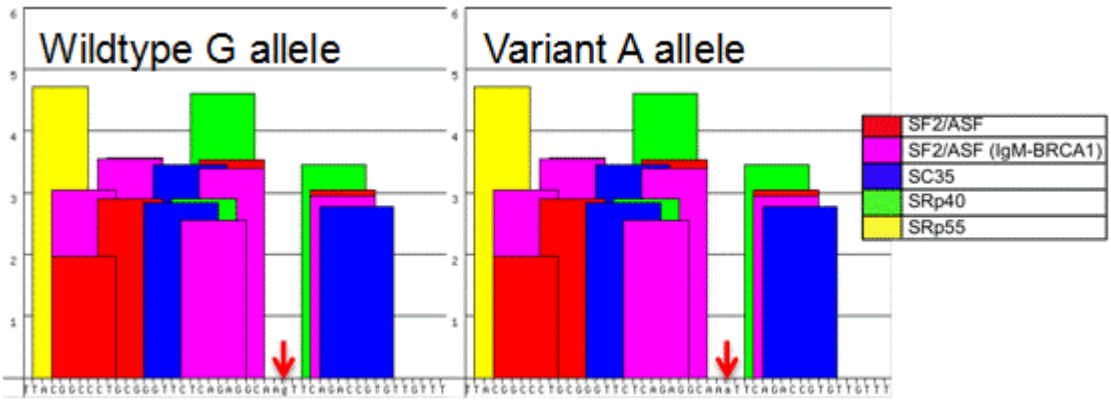


Figure 39: Predicted exonic splice enhancer splice sites in the region surrounding rs1805363 (red arrows) using the ESEFinder 3.0 (http://rulai.cshl.edu/tools/ESE3/) prediction tool for the wildtype G allele and the variant A allele.

8.4.3.1. MRE11A mRNA isoform profiling

To investigate the influence of rs1805363 on isoform expression, Dr Kiltie's group (performed by Mr C Buchwald) sequenced genomic DNA from eight bladder cancer and bladder-derived cell lines showing all were homozygous for the rs1805363 wildtype G allele (Figure 40).

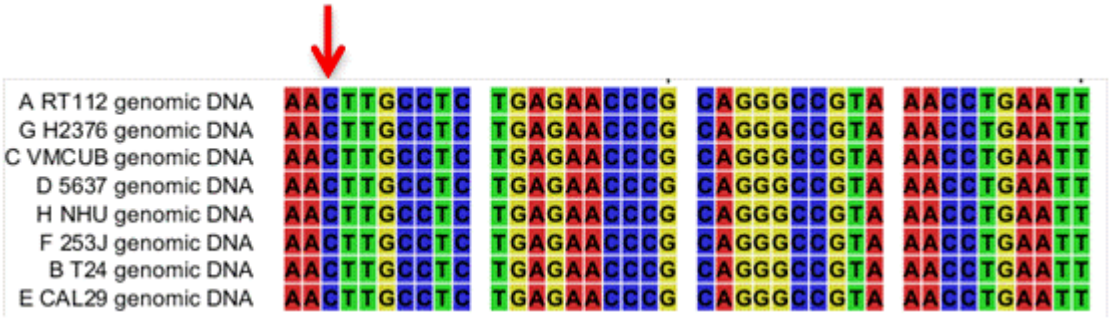


Figure 40: Sequence data for genomic DNA from eight bladder cancer and bladder-derived cell lines with all eight cell lines being wildtype for rs1805363 (red arrow).

Of the 34 carriers of the rs1805363 minor A allele in the MIBC radiotherapy cohort, only one heterozygous individual had matched RNA already extracted from their bladder tumour (now referred to as tumour A) available (generously provided by Prof M Knowles). RNA was thus extracted from macro-dissected muscle-invasive regions from the FFPE bladder tumour blocks of another heterozygous individual (tumour B) and a homozygous minor A allele individual (tumour C), and cDNA generated by reverse transcription for all three tumours. cDNA for six of the above bladder cancer cell lines were provided by Prof M Knowles as wildtype controls. PCR primers sited in common sequence for both *MRE11A* isoforms and spanning exons 1 and 2 were used to produce two PCR products of different sizes corresponding to isoform 1 (139 bp) and isoform 2 (246 bp) for cDNA from the six bladder cancer cell lines and tumours A to C. By gel electrophoresis, the band intensities of isoform 1 and isoform 2 relative to a housekeeper gene control, *GAPDH*, was measured for each sample and the relative expression of each isoform calculated (Figure 41A).

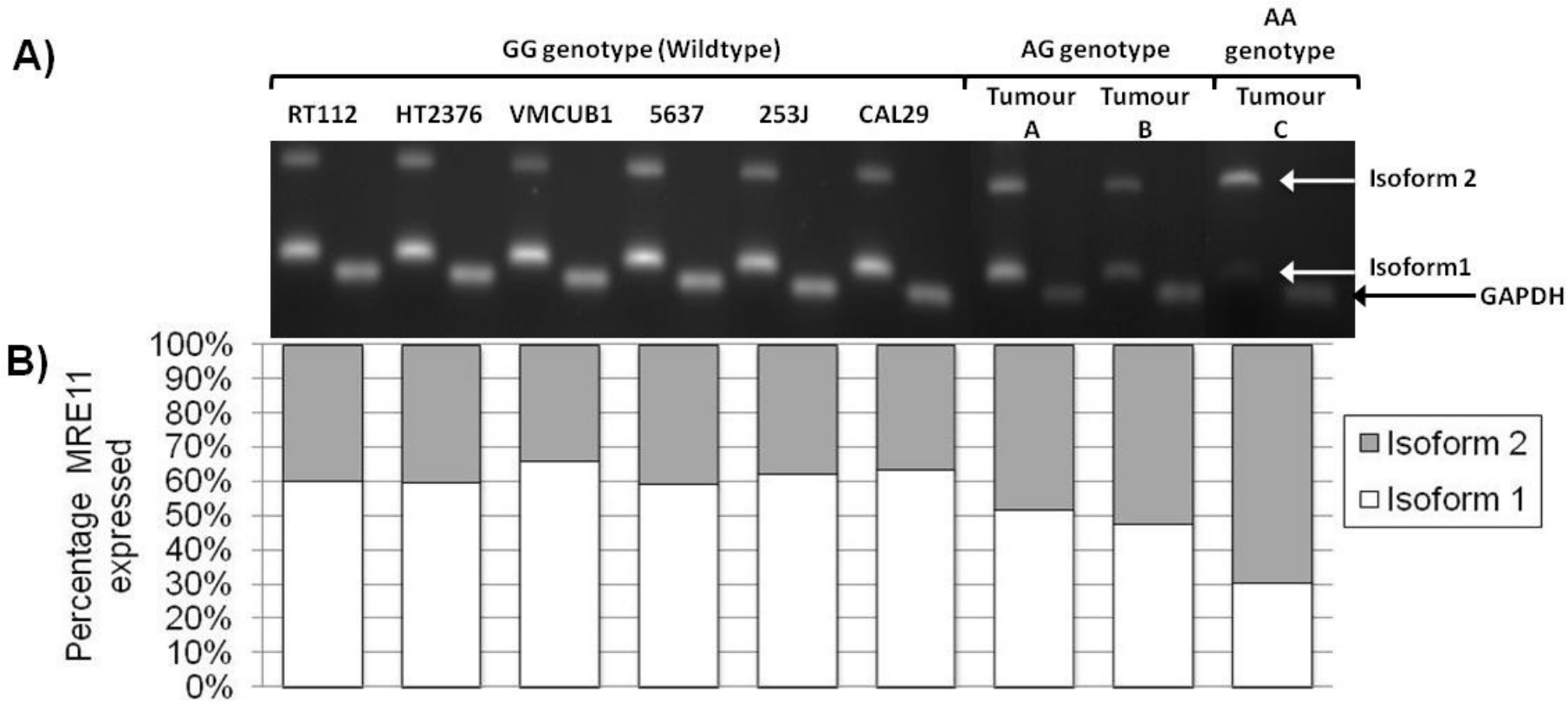


Figure 41: *MRE11A* isoform expression in bladder cancer cell lines and primary bladder tumours with different genotypes for rs1805363: A) Gel electrophoresis bands for *MRE11A* isoform 1 and 2, and *GAPDH* PCR amplified from cDNA from six bladder cancer cell lines (all wildtype GG genotype for rs1805363), primary bladder tumours A and B (heterozygote AG genotype for rs1805363), and primary bladder tumour C (homozygote variant AA genotype for rs1805363). B) Percentage expression of *MRE11A* for each isoform relative to overall *MRE11A* expression for each sample.

As shown in Figure 41B, the percentage of *MRE11A* isoform 1 relative to overall *MRE11A* (sum of isoforms 1 and 2) falls with each rs1805363 minor A allele carried. The average percentage of isoform 1 was 61.7% in wildtypes, 49.6% in heterozygotes and 30.3% in minor allele homozygotes. The two isoform gel bands were excised and sequenced to confirm they corresponded with isoforms 1 and 2 (Figure 42). Overall *MRE11A* expression normalised to *GAPDH* expression varied between cell lines and tumours (Table 44). No obvious difference in overall *MRE11A* expression was observed between genotypes but more samples would need to be analysed for any firm conclusions to be made.

Table 45: Overall *MRE11A* expression normalised to *GAPDH* expression in bladder cancer cell lines and primary bladder tumours A, B and C by *MRE11A* rs1805363 genotype.

Cell Line/ Tumour	Overall *MRE11A* expression normalised to *GAPDH* expression
GG genotype	
RT112	1.78
HT2376	1.82
VMCUB1	1.77
5637	2.13
253J	1.89
CAL29	2.45
Mean	*1.97*
AG genotype	
Tumour A	3.91
Tumour B	1.79
Mean	*2.50*
AA genotype	
Tumour C	2.46

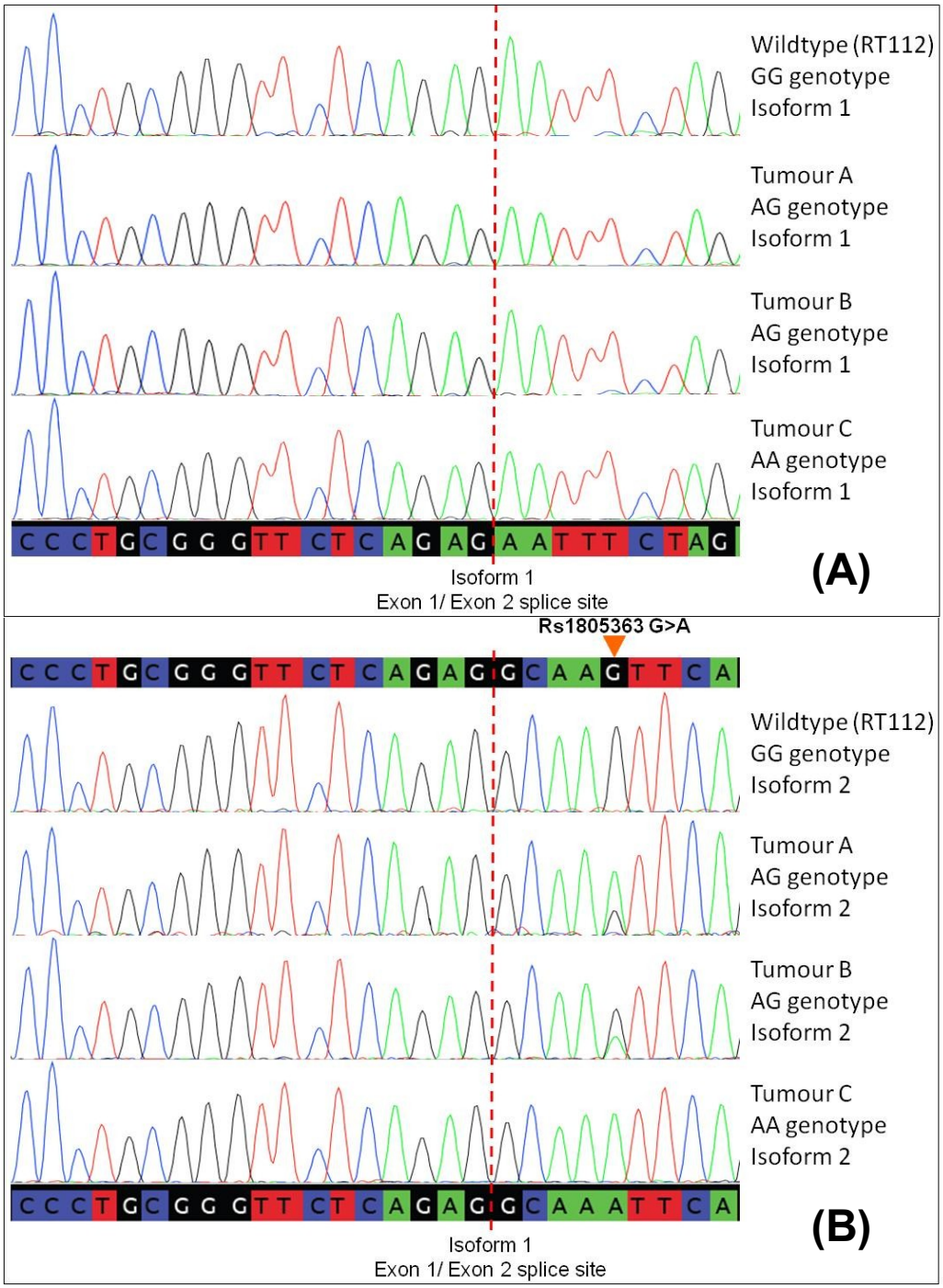


Figure 42: Sequencing electropherograms for PCR products corresponding to isoform 1 and isoform 2 for RT112 bladder cancer cell line, and primary bladder tumours A, B and C: A) Isoform 1/ 139 bp band, and B) isoform 2/ 246 bp band.

8.4.3.2. Radical cystectomy treated MIBC cohort

To determine the prognostic versus predictive role of rs1805363, a cohort of 256 MIBC patients treated by radical cystectomy from collaborators (Dr L Dyrskjøt and Prof T Ørntoft) at Aarhus University Hospital, Denmark, between 1992 and 2008 were genotyped for rs1805363. The clinical demographics of this population are shown in Table 44. This cohort was notably younger which would be expected as younger (thus healthier) patients would be more likely to receive surgery (Choudhury *et al.* 2010). Treatment details of this cohort were previously described (Jensen *et al.* 2011) but in summary, all patients had a radical cystectomy with two-thirds having a limited lymph node dissection to the obturator fossa and the remaining third of patients having an extended lymph node dissection to the level of the inferior mesenteric artery. None of these patients received neo-adjuvant or adjuvant chemotherapy.

Table 46: Clinical demographics of patients from the MIBC cystectomy cohort.

Variable	Study population (N=256) No of patients (%)
Age (years)	
Median (range)	65 (34 - 85)
Gender	
Male	187 (73.0)
Female	69 (27.0)
Tumour stage[a]	
T2	94 (36.7)
T3	112 (47.6)
T4	40 (15.7)
Nodal stage[a]	
N0	184 (71.9)
N1	32 (12.5)
N2	27 (10.5)
N3	12 (4.7)
Nx	1 (0.4)
Histological Grade[a]	
High grade	251 (98.0)
Low grade	5 (2.0)

[a] Bladder cancer classification as per the seventh edition of the TNM classification (Sobin *et al.* 2010).

Survival data was obtained from the Danish Central Personal Registry with cancer-specific deaths defined as deaths in patients with known bladder cancer recurrence following surgery. All other deaths were categorised as deaths from other causes. Median follow-up time was 63.7 months (range: 2.0 – 220.6 months) with 95 cancer-specific events. Assuming the same MAF for rs1805363, this cohort had a statistical power of 80.5% at a 5% significance level for detecting a HR of 2.10.

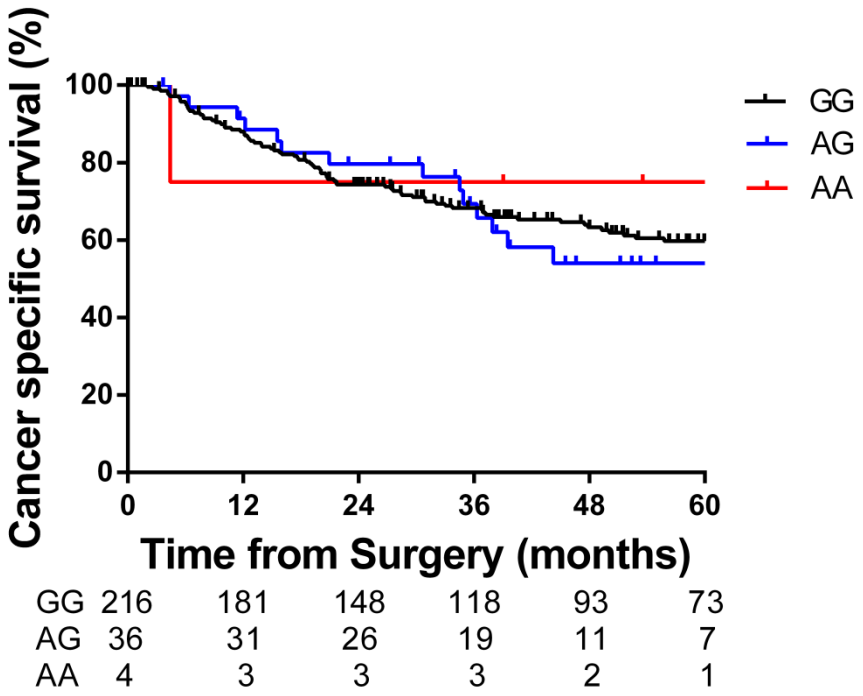


Figure 43: Kaplan Meier curve of cancer-specific survival for *MRE11A* rs1805363 genotypes in the Aarhus MIBC cystectomy cohort.

Genotyping for rs1805363 in this cohort found no association with CSS (P_{trend} = 0.89) with a 5-year CSS of 59.7% in wildtypes, 54.0% in heterozygotes and 75.0% in minor allele homozygotes (Figure 43). Multivariate analysis, adjusting for age, gender, tumour stage, nodal stage and histological grade, did not reveal any significant associations of the rs1805363 genotypes with CSS (Table 45).

Table 47: Cox proportional hazards multivariate analysis of cancer specific survival for rs1805363 in the cystectomy treated cohort adjusting for age, gender, tumour stage, nodal stage and histological grade.

Variant (Observed MAF)	Genotype	N	HR (95% CI)	P-value
rs1805363 G>A (0.09)	GG	216	1	0.88
	AG	36	1.21 (0.15 – 9.49)	
	AA	4	0.96 (0.13 – 7.03)	

8.4.4. Discussion

The *MRE11A* gene has a key role in DSB detection and repair signalling, with the expression of tumour MRE11 protein previously associated with survival following radiotherapy in 180 MIBC patients (Choudhury *et al.* 2010). The *MRE11A* gene was thus investigated for potential underlying genetic variants associated with both tumour response and radiation toxicity. Previous candidate gene studies of common coding DNA repair SNPs have found associations with pathological response or survival in several cancer sites following radiotherapy treatment, predominantly in combination with chemotherapy (Parliament and Murray 2010; Yin *et al.* 2011; Yoon *et al.* 2011; Wibom *et al.* 2012). Although a candidate gene approach was also used here, the use of NGS technology allows the in depth investigation of all *MRE11* variants in the study population including novel rare ones, rather than studying only known variants or tag SNPs.

In this pilot study, for the first time germline *MRE11* SNPs and rare variants were found to be associated with survival and late radiation normal tissue toxicity following radical radiotherapy in MIBC. The common SNP rs1805363 and *MRE11A* 3'UTR rare variant carrier status were identified as markers of poor radiotherapy outcomes and rs13447623 and rs2155209 as potential markers for increased late radiation normal tissue toxicity, thus potentially identifying an MIBC patient population not best served by bladder conserving treatment. Investigation of rs1805363 in a cystectomy treated MIBC cohort findings did not reveal any association with prognosis supporting the hypothesis of rs1805363 as a predictive inherited genetic marker of radiotherapy outcomes in MIBC.

Ricceri *et al* previously identified eight intronic *MRE11A* SNPs significantly associated with H2AX dephosphorylation at three hours, thus suggestive of increased completion of DSB repair (Taneja *et al.* 2004), following 2 Gy irradiation of

peripheral blood mononuclear cells from 118 healthy individuals (Ricceri *et al.* 2011). Though rs1805363 failed genotyping in their work, two of their significant *MRE11A* SNPs were detected by NGS in the MIBC radiotherapy cohort but were not associated with survival or toxicity. However, from HapMap data (http://hapmap.ncbi.nlm.nih.gov/), rs1805363 and rs13447623 are in moderate to complete *D'* linkage (*D'* > 0.6) with the top three *MRE11A* SNPs in Ricceri et al's work.

There are two expressed isoforms of *MRE11A* transcripts: isoform 1 (4772 bases: shorter 5'UTR and exon 16 expressed) and isoform 2 (4668 bases: longer 5'UTR and exon 16 not expressed). However, there are no published data available on the relative expression or the functional differences of these two isoforms. As rs1805363 was situated only five bases from the 3' terminal of exon 1 in isoform 1 of *MRE11A*, it was postulated that this SNP may influence the *MRE11A* isoform expressed. PCR amplification of isoforms 1 and 2 in cDNA from wildtype bladder cancer cell lines and primary bladder tumours from heterozygote and minor allele homozygotes demonstrated a clear functional effect of rs1805363 on isoform expression with a reduction in isoform 1 expression with each minor allele carried. However, what remains unclear is why this relative change in isoform expression affects survival with several possible hypotheses raised: (1) Does isoform 2 result in more efficient DSB repair thus promoting cancer cell survival? (2) Alternatively, is isoform 1 the crucial factor for initiating the DSB repair cascade and cell death pathways thus reduction in isoform 1 results in the failure of cancer cells to initiate apoptosis after radiotherapy? (3) The MRN complex exists as a heterotetramer with two MRE11 molecules (Lavin 2007). Are both isoforms 1 and 2 required in the MRN complex for complex stability or DSB detection with a critical balance needed in relative expression (hence the smaller effect on survival in heterozygotes)? These will require further investigation to

understand the role of these isoforms in DSB repair following radiotherapy treatment and to identify any potential therapeutic targets.

This was the first study demonstrating an association of germline rare variants with radiotherapy outcomes. Borgmann *et al* previously sequenced the cDNA of several DSB repair genes for germline mutations associated with late normal tissue radiation toxicity in head and neck cancer but failed to identify any *MRE11* coding variants (Borgmann *et al.* 2002). In contrast, we identified three known and one novel coding rare variants in our study, three of which are non-synonymous (Table 43). There are only nine *MRE11A* non-synonymous coding variants reported in the dbSNP database (http://www.ncbi.nlm.nih.gov/snp/), all with MAFs under 1%, indicating a high negative selection pressure against coding variants. Coding *MRE11A* mutations, after all, cause the rare Ataxia-Telangiectasia Like Disorder, associated with genomic instability and hypersensitivity to ionising radiation (Delia *et al.* 2004). Collapsed analysis of the coding rare variants in this study did not reveal any significant association with radiotherapy outcomes but this could be due to only five individuals carrying these rare variants. One known and five novel rare variants in the 3'UTR of *MRE11A* were also identified and were associated with survival following radiotherapy (Table 41). However, only one of these was predicted to affect miRNA binding (Table 43). If confirmed to affect miRNA binding, this would be in keeping with the previously demonstrated association with MIBC survival following RT treatment of another DNA repair gene 3'UTR SNP, *RAD51* rs7180135, which was predicted to be within a miRNA binding site (Teo *et al.* 2012).

This study demonstrated the first use of NGS technology for the in-depth investigation of germline common and rare variants within a candidate gene for potential biomarkers of radiotherapy outcomes. Currently, only three other studies, all investigating haematological malignancies, have applied NGS technology to

investigate for potential prognostic biomarkers specifically, leukaemic transformation in myelodysplasia (Jadersten *et al.* 2011) and chronic myelomonocytic leukemia (Kohlmann *et al.* 2010; Smith *et al.* 2010). With rising read lengths, paired end sequencing technology and higher throughput sample preparation, it would be possible for larger targets and larger multiplexing sequencing reactions, hence, larger discovery studies thus accelerating the discovery of new predictive and prognostic markers (Harakalova *et al.* 2011; Rivas *et al.* 2011). With diagnostic laboratories using this technology for clinical genetic testing (Morgan *et al.* 2010; Ozcelik *et al.* 2012), the widespread use of next generation sequencing to aid clinical decision making is imminent (Human Genomics Strategy Group 2012).

The main limitation of this study was the relatively small sample sizes and the inability to validate the findings in a separate RT treated MIBC study population. Muscle-invasive bladder cancer is an uncommon cancer, thus achieving the numbers seen in more common tumour sites is difficult, as demonstrated by two of the three most recent multi-centre phase III radiotherapy clinical trials on bladder cancer struggling to recruit and closing early (Huddart *et al.* 2010; James *et al.* 2012). However, there remains a need to validate if rs1805363 is a true predictive marker of radiotherapy response in a second independent radiotherapy treated MIBC cohort. Unfortunately, none of the recent or on-going phase III radiotherapy clinical trials on bladder cancer collected germline DNA samples for translational research (Huddart *et al.* 2009; Hoskin *et al.* 2010). As all cancers would be expected to mount a DNA repair response following radiotherapy, it was hypothesised that rs1805363, if a true predictive marker of radiotherapy response, would potentially be also predictive of survival in other radiotherapy treated cancers.

Wibom *et al* investigated 1458 DNA repair gene tag SNPs for associations with survival in a Swedish and Danish cohort of glioblastoma cases (Wibom *et al.* 2012). Rs1805363 was not genotyped in this study but of note, eight of the 11 *MRE11A* *S*NPs genotyped were significant in their predominantly radiotherapy treated test cohort. However, these findings were not validated in their validation cohort though no treatment details were available in this cohort. A collaboration has thus been set up with Professor Mellin's group to validate rs1805363 as a marker of radiotherapy response in the radiotherapy treated GBM cases of the test cohort. The results of this collaboration are eagerly awaited.

The low number of patients in the MIBC radiotherapy cohort receiving neo-adjuvant or concurrent chemotherapy allowed the specific assessment of the identified variants on RT response. However, whether the current standard of neo-adjuvant chemotherapy treatment and the likely introduction of concurrent chemotherapy and radiosensitisers would affect the predictive value of the identified *MRE11A* variants by involvement of other repair pathways thus bypassing the dependence on MRE11 remains unclear.

In summary, this study, using NGS technology, demonstrated germline *MRE11A* SNPs and rare variants as potential markers of radiotherapy outcomes and toxicity in muscle-invasive bladder cancer. Of note, carriage of the *MRE11A* rs1805363 minor A allele was found to be predictive of poor radiotherapy outcomes but not surgery in MIBC patients, and was shown to affect relative *MRE11A* isoform expression. Further validation of rs1805363 is urgently needed for its translation into a clinical predictive tool for personalised medicine.

8.5. Tumour DNA repair protein expression as a marker of radiotherapy response

Dr Kiltie's group had previously studied MIBC tumour expression of several DNA repair proteins involved in BER and DSB repair for associations with bladder cancer specific survival following RT treatment demonstrating several associations with survival (Sak *et al.* 2005; Choudhury *et al.* 2010). The current study thus investigated several other candidate DNA repair proteins as potential markers of MIBC RT outcomes, specifically CtIP, MUTYH and XPC.

These candidate DNA repair proteins were selected due to their potential direct or indirect influence on DSB repair. CtIP is closely involved with MRE11 in DSB end resection and processing (Sartori *et al.* 2007; Zha *et al.* 2009) and is associated with BRCA1 in regulating DSB repair pathway choice (Yu and Baer 2000; Yun and Hiom 2009). Ionising radiation forms clusters of oxidative base damage surrounding SSBs or DSBs which slows DSB repair as well as potentially generating further DSBs during repair (Gulston *et al.* 2004; Harrison *et al.* 2006; Eccles *et al.* 2011). MUTYH was thus investigated due to its involvement in the repair of the most common oxidative base lesion, 8-oxoG. Besides its role in GGR-NER, XPC has also been indicated to modulate DSB repair (Despras *et al.* 2007; Zhang *et al.* 2009).

8.5.1. Study population demographics

The study population consisted of non-metastatic MIBC patients treated by radical RT between 2002 and 2005 in Leeds. Clinical FFPE tumour blocks taken at pre-treatment TURBT were available for 88 individuals for CtIP. However, six blocks had too little tissue left, and so were dropped from further investigation for MUTYH and XPC. In the cohort of 88 individuals with 34 CSS events, assuming 75% of

individuals had high expression and 25% had low expression, this study had a power of 86.5% to detect a HR of 2.5 at the 5% significance level. For the reduced cohort size of 82 individuals with 33 CSS events, with the same division of protein expression and same significance level, the study had a power of 84.4% to detect a HR of 2.5. The study demographics are shown in Table 46.

Table 48: Clinical demographics of bladder cancer patients treated by radical RT between 2002 and 2005 with available pre-treatment TURBT FFPE tumour blocks.

Variable		Study population (N=88) No of patients (%)
Age (years)	Median (range)	78.2 (55.7 - 89.9)
Gender	Male	67 (76.1)
	Female	21 (23.9)
Tumour stage	T1	2 (2.3)
	T2	48 (54.5)
	T3	29 (33)
	T4	9 (10.2)
Nodal stage	N0	84 (95.5)
	N1	2 (2.3)
	N2	2 (2.3)
Histological grade	G3	76 (86.4)
	<G3	9 (10.2)
	Gx	3 (3.4)
Hydronephrosis	No	58 (65.9)
	Yes	30 (34.1)
Neoadjuvant/ concurrent therapy	None	73 (83)
	Neoadjuvant chemotherapy	8 (9.1)[a]
	Concurrent chemotherapy/ radiosensitiser	7 (8)[b]
Salvage chemotherapy	Not received	83 (94.3)
	Received	5 (5.7)
Salvage cystectomy	Not received	83 (94.3)
	Received	5 (5.7)

[a] All received platinum based combination chemotherapy.
[b] 6 patients received concurrent gemcitabine (100 mg/m^2) weekly ×4 as part of a phase II clinical trial and one patient received concurrent carbogen and nicotinamide as part of the BCON phase III clinical trial.

8.5.2. Antibody specificity validation and optimisation

Antibody specificity was validated using cell line extracts by Western blotting (Figure 44). Corresponding FFPE cell line pellets were then made and used as positive controls and IHC protocols optimised using these positive controls (see Methods section 7.3).

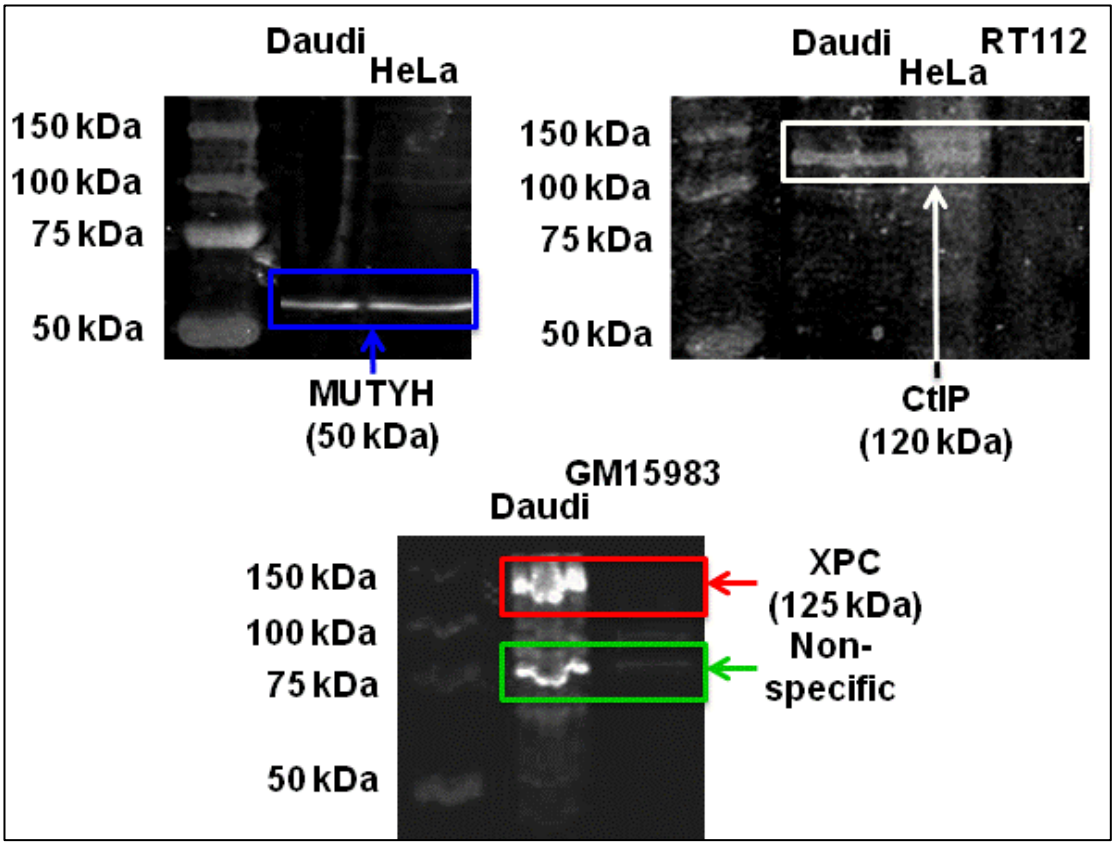


Figure 44: Western blots of cell line extracts for MUTYH (top left, blue), CtIP (top right, white) and XPC (bottom, red). Non-specific band (green) noted with anti-XPC antibody. XPC western blot was performed by Dr J Bentley and Ms M McCarthy.

As shown in Figure 44, the anti-XPC antibody had a second unspecified band on Western blotting so a cell pellet of the known null-XPC expressing GM15983 cell line was used as a separate negative control tissue for IHC protocol optimisation. Stained positive control cell pellets with the respective antibodies are shown in Figure 45.

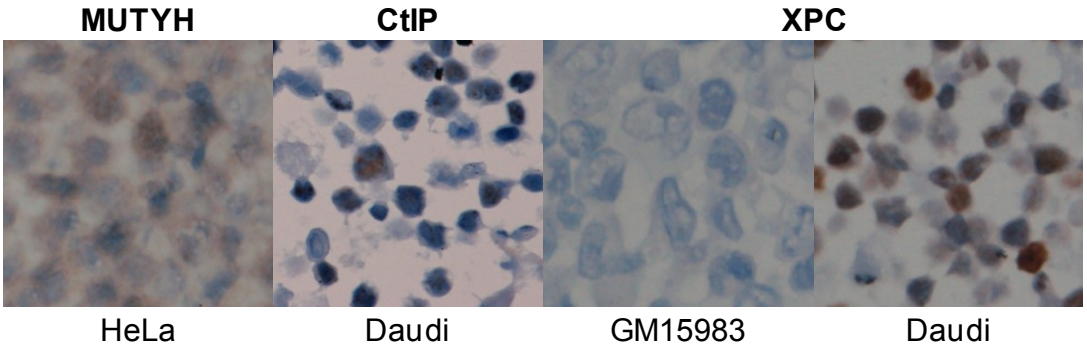


Figure 45: Stained positive control (and negative control in the case of GM15983) cell line pellets for MUTYH, CtIP and XPC.

8.5.3. Bladder tumour CtIP, MUTYH and XPC expression

Tumour DNA repair protein expression was graded using a multiplicative SQS (see Methods chapter 7.3.7). Nuclear CtIP expression varied across the cohort with a median SQS of 3 (range: 0 – 12). Of interest, among CtIP expressing malignant cells within the same bladder tumour, some bladder cancer cells had distinct nuclear foci of CtIP (Figure 46A) while others had a diffuse expression throughout the nucleus (Figure 46B). The significance of these foci are unclear but previous associations have been reported of nuclear CtIP foci with S-phase replication forks (Gu and Chen 2009) and Akt signalling (Xu *et al.* 2007). In contrast, most tumours had low nuclear MUTYH expression with 74.4% of tumours having an SQS of 0 or 1 (range: 0 – 10.5) but just over half (51.2%) of tumours had cytoplasmic MUTYH expression (Figure 46C and D). Tumour nuclear XPC expression was relatively evenly spread with a median SQS of 6 (range: 1 – 12). No associations were found between CtIP, MUTYH and XPC expression levels with tumour stage or histological grade (P = 0.54 and 0.22; 0.13 and 0.67; and 0.89 and 0.86 respectively).

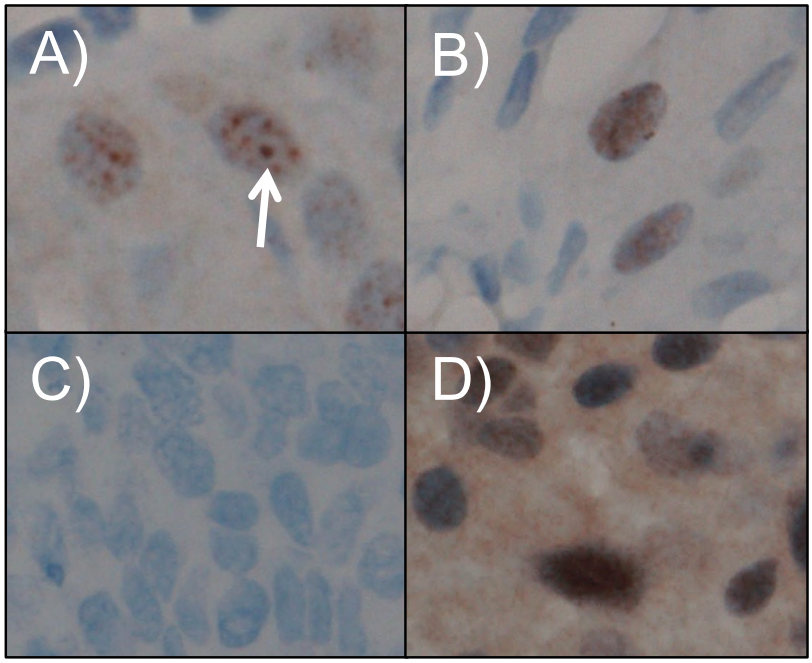


Figure 46: Representative images of variation in nuclear CtIP staining and MUTYH cytoplasmic staining: A) Bladder tumour nuclear CtIP expression with the presence of distinct foci (white arrow), B) Diffuse nuclear CtIP expression in malignant cells from the same bladder tumour, C) Absent MUTYH cytoplasmic expression, and (D) Positive MUTYH cytoplasmic expression.

8.5.4. Association with cancer-specific survival

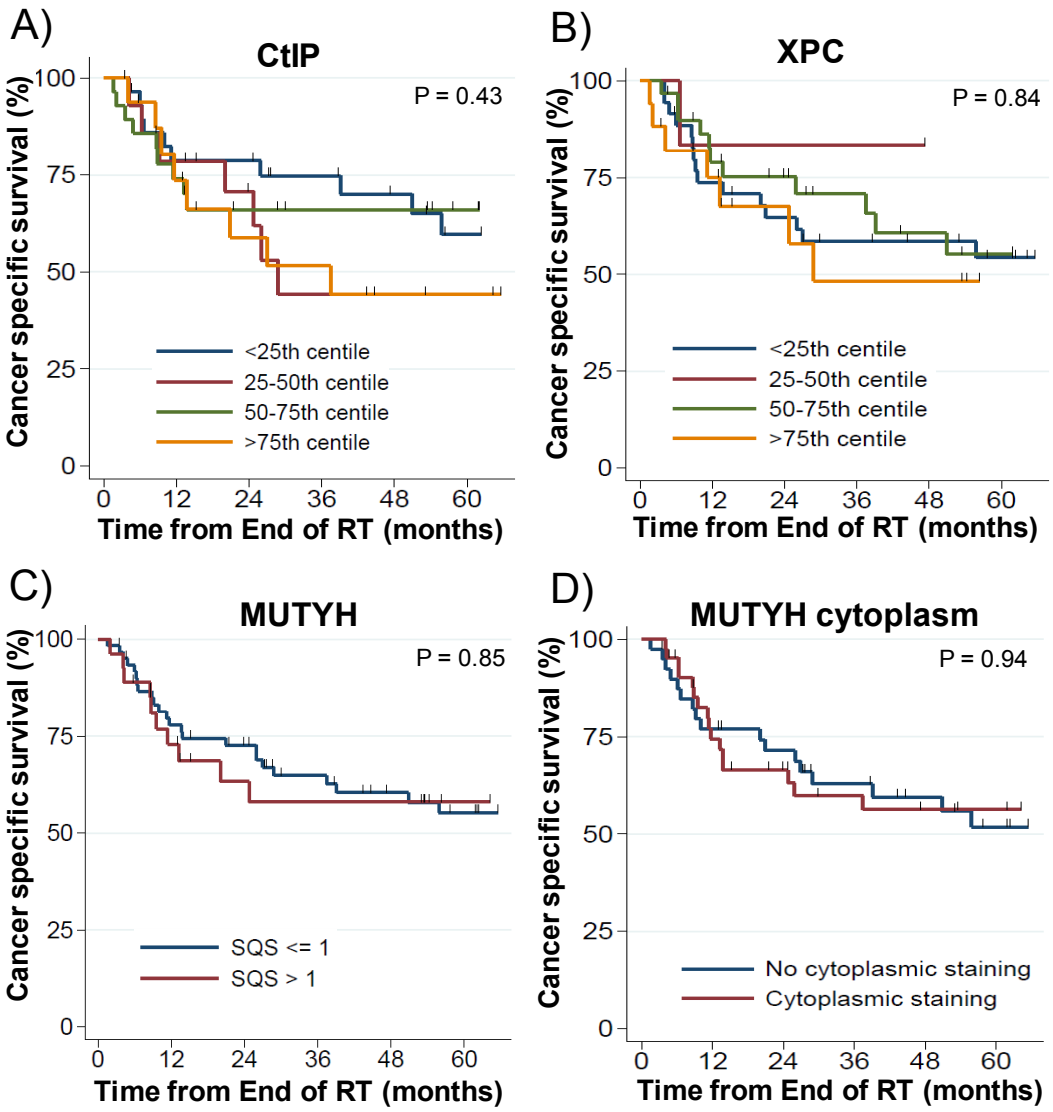


Figure 47: Kaplan-Meier survival curves of CSS for CtIP, XPC and MUTYH: A) SQS quartiles of CtIP expression, B) SQS quartiles of XPC expression, C) SQS of one or less versus greater than one for nuclear MUTYH expression, and D) absent versus present cytoplasmic MUTYH expression.

Nuclear CtIP and XPC expression SQS were divided into quartiles and analysed for any trend of increasing protein expression with CSS (Figure 47A and B) but no significant associations were found (P = 0.43 and 0.84 respectively). For nuclear MUTYH expression, “low” expressing tumours with SQS of less than or equal to one, were compared with “high expressing” tumours with SQS greater than one. Once again, no significant associations were found with CSS (P = 0.85, Figure 47C) nor

were there any associations with the presence or absence of cytoplasmic expression (P = 0.94, Figure 47D).

8.5.5. Discussion

This study investigating three candidate DNA repair proteins involved in DSB repair, BER and NER found no significant associations between the three candidate proteins, CtIP, MUTYH and XPC, with cancer-specific survival following radical radiotherapy treatment for MIBC.

CtIP's involvement in DSB end resection and processing with MRE11, previously shown to be predictive of CSS in MIBC following RT (Choudhury *et al.* 2010; Laurberg *et al.* 2012) highlighted CtIP as a potential candidate for investigation. On top of this, CtIP's also regulates DSB repair HRR or NHEJ pathway choice (Yu and Baer 2000; Yun and Hiom 2009), and maintains S-phase/ G2-phase cell cycle checkpoints following DNA damage (Kousholt *et al.* 2012). This negative result may indicate that MRE11's position upstream of CtIP in the DSB repair pathway and its regulation of CtIP phosphorylation by CDK2 (Buis *et al.* 2012), supersedes any influence CtIP levels have on DSB repair and cancer cell survival. Only one study has looked at CtIP expression in primary human tumours with clinical outcomes, observing an association with response to endocrine treatment in breast cancer (Wu *et al.* 2007). However, in that study, CtiP expression levels were based on cytoplasmic staining of CtIP which is peculiar as CtIP is supposed to be only localised within the nucleus (Yu and Baer 2000).

What remains unclear is the significance of the observed diffuse CtIP nuclear staining versus CtIP nuclear foci. The CtIP SQS scoring used did not differentiate between the two forms of staining in scoring if a malignant cell was positive or negative for CtIP expression. The formation of CtIP foci following ionising radiation in

cancer cell lines could suggest that scoring of these foci, either by the number of foci per cell or cells with CtIP foci, could potentially be a better measure of CtIP activity (Yuan and Chen 2009).

MUTYH, as previously mentioned, is involved in maintaining genomic stability by detecting 8-oxoG following oxidative damage and excising any mispaired A during replication which results in an abasic site. The repair of this abasic site results in the formation of a SSB (David *et al.* 2007; Hegde *et al.* 2008), which in regions of clustered DNA damage as in following IR, the presence of a SSB on the sister strand could result in the formation of additional lethal DSBs (Gulston *et al.* 2004; Harrison *et al.* 2006; Eccles *et al.* 2011). Expression of both nuclear and mitochondrial MUTYH was assessed in this study. High nuclear MUTYH expression has previously been associated with better prognosis in gastric cancer (Shinmura *et al.* 2011). Mitochondrial MUTYH in cancer prognosis has not been previously researched but mitochondrial genetic instability has been linked with improved survival in Duke's C colorectal cancer (Tsai *et al.* 2009). Mitochondrial dysfunction, secondary to mitochondrial genetic instability, results in impaired p53 function and IR-induced cell death (Kim *et al.* 2006; Compton *et al.* 2011).

However, this report did not find any correlation of nuclear or mitochondrial MUTYH with radiotherapy outcomes in MIBC. MUTYH would only be active during DNA replication (ie. S-phase) thus any additional DSBs formed could be compensated by the activity of HRR. Also studies in *Escherichia coli* found that clustered 8-oxoG lesions with or without uracil mispairing did not result in the formation of DSBs (Malyarchuk *et al.* 2004), which could account for the result seen in this study.

Besides its activity in NER, XPC has been implicated in the repair of DSBs and oxidative base damage (D'Errico *et al.* 2006; Despras *et al.* 2007; Zhang *et al.* 2009).

Despite this, no association was found in this study with prognosis following RT treatment. Several other studies have examined XPC expression in cancer outcomes though not in relation to radiotherapy. Yang *et al* in a study of 66 bladder tumours, with over a third being superficial non-invasive disease, reported associations of null XPC expression with worse bladder cancer pathological grade and prognosis (Yang *et al.* 2010). However, the XPC expression seen in the MIBC RT cohort (predominantly high grade disease) did not correlate with Yang *et al*'s findings with only four tumours having less than 10% of malignant cells expressing XPC. In breast cancer, one study reported lower XPC expression in malignant cells compared to normal breast tissue, but no correlation was seen with survival on multivariate analysis (Bai *et al.* 2012).

Studies investigating XPC and chemosensitivity (predominantly platinum-based chemotherapy) have been more fruitful with several SNP studies reporting associations with survival in breast, ovarian and lung cancer (Dong *et al.* 2012; Fleming *et al.* 2012; Yang *et al.* 2012). Lai *et al* reported decreased cisplatin chemotherapy response and survival in lung adenocarcinomas with high tumour XPC expression (Lai *et al.* 2011). Only 8 individuals in the MIBC RT cohort received platinum-based chemotherapy but with current practice recommending neoadjuvant cisplatin-based chemotherapy (Vale and ABC Meta-analysis Group 2005) and interest in the use of concurrent cisplatin chemoradiotherapy (Efstathiou *et al.* 2012), XPC expression as a possible prognostic/ predictive marker of the added benefit of chemotherapy in these groups may be of future clinical significance.

Studies of tumour expression of multiple DNA repair proteins have been investigated for markers of radiotherapy or chemoradiotherapy outcomes in cervical carcinoma, glioblastoma, oesophageal cancer and squamous cell carcinomas of the head and neck, with associations seen with DNA-PK, ERCC1, ERCC4, Ku80 and PARP1

(Kase *et al.* 2011; Liang *et al.* 2011; Moeller *et al.* 2011; Sun *et al.* 2011; Vaezi *et al.* 2011; Alexander *et al.* 2012). In MIBC, as previously mentioned, Dr Kiltie's group had initially found that tumours with a "high" percentage of positive nuclear staining malignant cells for APE1 and XRCC1 was associated with CSS following RT (Sak *et al.* 2005). High MRE11 expression was then identified and validated as a predictive biomarker of better CSS following RT in two radiotherapy cohorts but not following cystectomy treatment (Choudhury *et al.* 2010). This finding was further validated by an independent group who also identified high expression of TIP60 acetyltransferase, which is involved in multiple signalling pathways including DSB repair, histone acetylation and chromatin modelling, as predictive of improved CSS following cystectomy but not following RT (Sun *et al.* 2010; Laurberg *et al.* 2012). These biomarkers are currently under development.

One of the strengths of this study was that immunohistochemistry protocols and scoring methods are widely used clinically and in research thus are well validated. However, one of the main limitations was the difficulties in identifying adequately robust antibodies for immunohistochemistry. Immunohistochemistry antibody choice was based, as recommended, on the specificity seen on Western blotting on a complex protein sample such as cell line lysates, with minimal non-specific cross-reaction bands seen (Kurien *et al.* 2011; Signore and Reeder 2012). However, antibodies for Western blotting are designed to target short linear amino acid sequences under reducing conditions. These antibodies may thus not bind to protein in its tertiary conformation as in during immunohistochemistry. In turn, antibodies with non-specific binding on Western blotting may have specific binding to only the target protein on immunohistochemistry. For this study, what was thus unclear, for example, was whether the weak MUTYH staining throughout the cohort truly representative of low MUTYH expression or just representative of differences in antibody activity

between application types. Without fully profiled control FFPE tissue, it was very difficult to be confident of the specificity of an antibody for immunohistochemistry.

Another difficulty encountered was the reliability of manufacturer datasheets for the selection of suitable antibodies for validation. During preliminary validation for an OGG1 antibody, the most cited anti-OGG1 antibody used for immunohistochemistry (Li *et al.* 2001; Fukae *et al.* 2005; Sheehan *et al.* 2005) was tested by Western blotting using the manufacturer recommended positive control cell lysate. This revealed a very weak target band and two much stronger non-specific bands (Figure 48, Lane A) not seen on the manufacturer's datasheet (Figure 48, Lane B) yet present on the supplier's internal quality assurance blot (Figure 48, Lane C). The significance of these non-specific bands on immunohistochemistry was unknown but raised questions on the reliability of previously published work plus highlights the costs and time required to identify and validate viable antibodies for a project with in this case, OGG1 being dropped from further investigation.

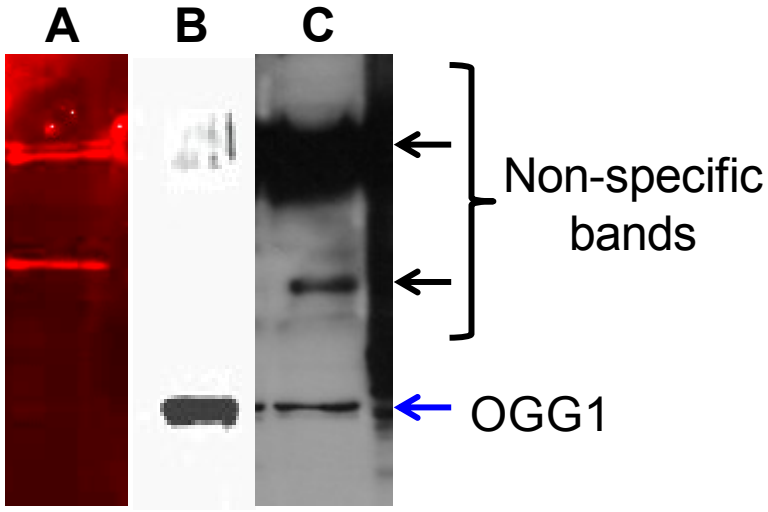


Figure 48: Western blots using rabbit polyclonal anti-OGG1 antibody on manufacturer recommended cell line lysate. Lane A: Experimental western blot; Lane B: Manufacturer's supplied datasheet western blot image; Lane C: Supplier's internal quality assurance western blot image on request.

It is likely that the majority of clinically relevant predictive tools will consist of a panel of biomarkers, each with moderate predictive power, rather than a single biomarker with strong predictive power (Shariat *et al.* 2010). However, this study investigated only three candidate proteins in a small study size with power to detect only biomarkers with relatively large effects. It is thus possible that smaller yet clinically relevant effects could have been missed. Pathway-based strategies or larger panels could thus potentially yield greater success (Choudhury *et al.* 2010; Moeller *et al.* 2011; Alexander *et al.* 2012). Due to the heterogeneity inherent within bladder tumours (Jones *et al.* 2005), the high-throughput advantages of using tissue microarrays for MIBC is limited as selected tumour cores may not be representative of the whole tumour hence the use of tissue sections in this study. This potentially limits the project's sample size, statistical power and number of candidate markers studied based on available time and resources. Due to the subjective nature of differentiating malignant cells from surrounding stroma, inflammatory or immune cells, the use of current automated scoring software was also not feasible to help boost throughput and study power.

In summary, MIBC tumour expression of three candidate DNA repair proteins, CtIP, MUTYH and XPC were investigated for potential markers of CSS following RT treatment with no significant associations found. Further research potentially studying larger target protein panels and sample sizes would be useful for the identification of new potential biomarkers.

9. General discussion and future work

Genetic instability is one of the ten hallmarks of cancer and is probably the most important (Hanahan and Weinberg 2011). It triggers the mutational changes required for the acquisition and progression of the other nine hallmarks. Cytotoxic agents and radiotherapy treatment work in the therapeutic window created by the differential capability to repair DNA damage generated by these treatment modalities in genetically unstable cancer cells compared to genetically stable normal cells, thus balancing cancer cell kill versus damage to normal tissues. The maintenance of genomic stability by the DNA repair pathways is thus integral to prevention of carcinogenesis (Hoeijmakers 2001; Bristow and Harrington 2005) as well as determining treatment response to radiotherapy (Steel 2002; Hall and Giaccia 2006). This thesis investigated the role of germline genetic variants in DNA repair genes on bladder cancer inherited risk, and DNA repair germline genetic variants and tumour protein expression in bladder cancer survival following RT treatment.

The CDCV hypothesis had been the premise of all GWAS studies on the genetic inheritability of bladder cancer (Kiemeney *et al.* 2008; Wu *et al.* 2009; Kiemeney *et al.* 2010; Rothman *et al.* 2010; Garcia-Closas *et al.* 2011; Rafnar *et al.* 2011), while common DNA repair gene SNPs had been implicated in bladder cancer risk (Stern *et al.* 2009). However, the contribution of rare variants and the CDRV hypothesis had not been explored in bladder cancer but is now feasible with NGS technology. The NGS methods development study demonstrated the capability of what is already becoming out-dated technology for the cost-effective multiplexed sequencing of candidate genes for RVs, showing two approaches involving unindexed pooling or bar-coding technology, with the choice of approach based on balancing between costs and manual laboratory time versus accuracy and the need for genotype data respectively. However, unexpectedly the issue discovered was not one of detecting

RVs but of how to prioritise candidate RVs for further investigation from the hundreds detected. With non-coding variants being the most frequent RVs encountered, the limited bioinformatics functional prediction tools available made prioritisation of these variants very arbitrary. Unfortunately, the Syzygy software generated Chi^2-value for detected RVs proved not to be an effective predictor of disease susceptibility RVs. Other equivalent NGS studies of common diseases, have either just described the discovered RVs (Kelleher *et al.* 2012; Lotta *et al.* 2012), gone on to genotype just novel variants (Doyle *et al.* 2012), or to pay the costs to perform large-scale genotyping of most/ all detected RVs with mixed results (Rivas *et al.* 2011; Heinzen *et al.* 2012).

What also remains unclear is whether the current recommended two-stage design of RV, namely RV discovery followed by genotyping in a large case-control study, represents the best approach for future disease risk RV discovery. This protocol has been based on trying to maximise cost-efficacy for NGS use (Kang *et al.* 2012) and statistical power (Yang and Thomas 2011). With the rapid development in newer, higher throughput and automated technology, such as “third” generation sequencing platforms (Pareek *et al.* 2011), larger scale target-enrichment methods (Mamanova *et al.* 2010) and automated library preparation protocols, undertaking a large scale whole genome sequencing studies does not seem far away. This technological surge will require the development of newer protocols, study designs and analytical tools for the handling and filtering of the immense data load for what is clinically relevant. With falling sequencing costs, a two stage sequencing study could soon be feasible with whole genome sequencing of an enriched population followed by multiplexed sequencing of candidate genes with an over-/under-representation of RVs in the enriched population in a larger case-control study.

Investigation of the CDRV hypothesis in bladder cancer genetic susceptibility specifically looking at two DNA repair genes, *MUTYH* and *XPC*, yielded mixed results. Collapsed analysis of *XPC* RVs found an association with increased bladder cancer risk as previously found by the group (Qiao *et al.* 2011), but the effect seen was far smaller than predicted by the CDRV hypothesis. However, this effect size emulated some of the results found in other RV studies (Rivas *et al.* 2011). Surprisingly, *MUTYH* was not associated with bladder cancer risk despite multiple associations of *MUTYH* RVs with colorectal cancer and other cancer types (Vogt *et al.* 2009; Pervaiz *et al.* 2010; Theodoratou *et al.* 2010; Win *et al.* 2011; Castillejo *et al.* 2012; Rennert *et al.* 2012). Like all RV association studies, this study was severely limited by having insufficient samples to achieve statistical power with a power of only 69% for an OR of 2.0 at a RV MAF of 0.01, thus the negative result does not exclude the involvement of *MUTYH* RVs in bladder cancer risk.

miRNAs mediate post-transcriptional regulation and are observed to be dysregulated in bladder cancer (Catto *et al.* 2009; Catto *et al.* 2011) with SNPs in miRNA-related and biogenesis genes increasing bladder cancer risk (Yang *et al.* 2008). What had not yet been investigated was how SNPs in DNA repair gene miRNA-binding sites could affect bladder cancer susceptibility by altering miRNA-binding and DNA repair gene expression. Examination of putatively functional 3'UTR SNPs in DNA repair genes found the BER genes, *PARP1* rs8679 and *LIG3* rs4796030, to be associated with increased bladder cancer risk, with an additive effect seen with carriage of both SNPs. However, functional predictions of miRNA binding propose an increase in expression of these two genes in carriers of these SNPs which would be expected to increase DNA repair thus reduce mutation and cancer risk, the reverse effect on bladder cancer risk to what was seen. It was thus hypothesised that increased PARP1 and LIG3 expression due to *LIG3* rs4796030 and *PARP1* rs8679 could possibly promote DSB repair via the MMEJ over the classical NHEJ pathway

resulting in more error-prone repair and mutagenesis. Overexpression of PARP1 had been shown to promote MMEJ over classical NHEJ (Wang *et al.* 2006), while increased MMEJ repair in peripheral blood lymphocytes had been observed in women with breast cancer or with a familial risk of bladder cancer potentially implicating MMEJ repair in cancer susceptibility (Keimling *et al.* 2012). Since this study, SNPs in miRNA-binding sites in NER genes have been implicated in colorectal cancer risk (Naccarati *et al.* 2012) thus emphasising the influence of miRNAs on DNA repair gene activity and overall cancer risk.

These case-control studies reinforce the contribution of DNA repair gene variants in bladder cancer risk despite none being detected by the bladder cancer GWAS studies. However, further work is needed to validate the results reported here. Due to the population specific nature of RVs, validation of RVs in other ethnically similar cohorts would be required. To this purpose, preliminary results of this study were recently presented at the International Bladder Cancer Consortium meeting. Experimental studies assessing the functional consequences of RVs would be highly useful not just to corroborate significant associations but as a method to prioritise candidate variants for genotyping and to provide an *a priori* weighting in statistical analyses thus potentially boosting statistical power (Cooper and Shendure 2011; Zhu *et al.* 2011). In a similar way, functional assays will be needed to assess the effects of rs8679 and rs4796030 on PARP1 and LIG3 expression respectively, as well as any influence on choice of MMEJ over NHEJ repair of DSBs.

DNA repair is one of the 5 R's of radiobiology for determining cancer response to radiotherapy (Steel 2002). In MIBC, with there being no clear evidence of superiority of cystectomy over radiotherapy or vice-versa, predictive biomarkers of radiotherapy responders are needed to aid clinicians in personalising treatment for each patient, thus reducing morbidity of ineffective or over-invasive treatment and decreasing

delay (and risk of metastasis) to receiving optimum treatment. The study of putatively functional DNA repair gene 3'UTR SNPs identified *RAD51* rs7180135 as a marker of good prognosis following radiotherapy treatment with functional predictions of this SNP resulting in stronger miR-197 binding thus decreased RAD51 expression. This would fit with other studies demonstrating increased radiosensitivity of tumour cell lines following RAD51 knockdown (Short *et al.* 2011) but also highlighted miR-197 as a possible therapeutic target for radiosensitisation.

Following the success in identifying tumour MRE11 protein expression (Choudhury *et al.* 2010) and previously unpublished work on *MRE11A* SNPs as predictive markers of RT outcomes by Dr Kiltie's group, multiplexed NGS of *MRE11A* was undertaken to interrogate this candidate gene in detail for SNPs and RVs associated with RT outcomes. This proved to be a successful endeavour identifying *MRE11A* rs1805363 G>A as being predictive of CSS following RT but not cystectomy, and functional studies went on to demonstrate its effect on *MRE11A* mRNA isoform expression. These results support the small number of studies demonstrating associations of inherited DNA repair gene SNPs with treatment response (Parliament and Murray 2010) thus highlighting the potential of developing a panel of genetic markers for the prediction of radiotherapy response in the future.

Unfortunately, investigation of the tumour protein expression of three DNA repair proteins, CtIP, MUTYH and XPC, did not reveal any correlation of tumour protein expression with radiotherapy outcomes unlike previous work by Dr Kiltie's group in this area (Sak *et al.* 2005; Choudhury *et al.* 2010). These proteins have never been investigated before for associations with radiotherapy outcomes in any cancer. Though not predictive of radiotherapy outcomes, XPC expression was previously associated with increased cisplatin chemotherapy resistance in adenocarcinomas of the lung (Lai *et al.* 2011). It could thus still have a role in determining patients who

would benefit from cisplatin as part of neo-adjuvant chemotherapy or concurrent chemo-radiotherapy regimes.

Future work is needed to validate *MRE11A* rs1805363 and *RAD51* rs7180135 as markers of radiotherapy outcomes. Due to the unavailability of other cohorts of radiotherapy treated MIBC, collaborations are under way exploring the role of the *MRE11A* SNP rs18053636 as a potential universal marker of radiotherapy response due to the central role of MRE11 in DSB repair. Ultimately, this SNP will need to be validated as part of a prospective clinical trial to assess its true applicability in the clinical environment.

A greater understanding of the inherited risk determinants of bladder cancer will allow the risk stratification of individuals for the targeted screening of high risk sub-populations for earlier diagnosis or the monitoring of only patients at high risk of recurrence thus reducing the morbidity of repeated invasive cystosopic procedures with the associated high costs to stretched healthcare budgets. Improved prediction of RT treatment outcomes in MIBC could spare individuals from either having to endure the morbidity of four or more weeks of ineffective RT treatment, or the morbidity of having a urostomy following a cystectomy when bladder function could have been preserved by RT treatment. This individualisation of MIBC treatment could also result in an overall improvement in bladder cancer survival. This study has shown the genetic contributions of DNA repair gene SNPs that influence post-transcriptional regulation and RVs on bladder cancer risk as well as identifying potentially new clinical genetic markers for predicting RT outcomes in MIBC.

10. References

Aben, K. K., J. A. Witjes, M. P. Schoenberg, C. Hulsbergen-van de Kaa, A. L. Verbeek and L. A. Kiemeney (2002). "Familial aggregation of urothelial cell carcinoma." Int J Cancer **98**(2): 274-278.

Al-Tassan, N., N. H. Chmiel, J. Maynard, N. Fleming, A. L. Livingston, G. T. Williams*, et al.* (2002). "Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors." Nat Genet **30**(2): 227-232.

Alexander, B. M., X. Z. Wang, A. Niemierko, D. T. Weaver, R. H. Mak, K. S. Roof*, et al.* (2012). "DNA repair biomarkers predict response to neoadjuvant chemoradiotherapy in esophageal cancer." Int J Radiat Oncol Biol Phys **83**(1): 164-171.

Ali, M., H. Kim, S. Cleary, C. Cupples, S. Gallinger and R. Bristow (2008). "Characterization of mutant MUTYH proteins associated with familial colorectal cancer." Gastroenterology **135**(2): 499-507.

Alsner, J., C. N. Andreassen and J. Overgaard (2008). "Genetic markers for prediction of normal tissue toxicity after radiotherapy." Semin Radiat Oncol **18**(2): 126-135.

Amstutz, U., G. Andrey-Zurcher, D. Suciu, R. Jaggi, J. Haberle and C. R. Largiader (2011). "Sequence capture and next-generation resequencing of multiple tagged nucleic acid samples for mutation screening of urea cycle disorders." Clin Chem **57**(1): 102-111.

Audebert, M., B. Salles and P. Calsou (2004). "Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining." J Biol Chem **279**(53): 55117-55126.

Babjuk, M., W. Oosterlinck, R. Sylvester, E. Kaasinen, A. Böhle, J. Palou and M. Rouprêt (2011). Guidelines on non-muscle invasive bladder cancer (TaT1 and CIS), European Association of Urology.

Bai, X., F. Jin, Y. Fu, Z. Yu, L. Zhao, J. Ren*, et al.* (2012). "Clinicopathological significance and prognostic value of Xeroderma pigmentosum complementary group C (XPC) expression in sporadic breast cancer patients." Med Oncol **29**(3): 1543-1553.

Bakkar, A. A., H. Wallerand, F. Radvanyi, J. B. Lahaye, S. Pissard, L. Lecerf*, et al.* (2003). "FGFR3 and TP53 gene mutations define two distinct pathways in urothelial cell carcinoma of the bladder." Cancer Res **63**(23): 8108-8112.

Bansal, V., O. Libiger, A. Torkamani and N. J. Schork (2010). "Statistical analysis strategies for association studies involving rare variants." Nat Rev Genet **11**(11): 773-785.

Barnett, G. C., C. E. Coles, R. M. Elliott, C. Baynes, C. Luccarini, D. Conroy*, et al.* (2012). "Independent validation of genes and polymorphisms reported to be associated with radiation toxicity: a prospective analysis study." Lancet Oncol **13**(1): 65-77.

Barrett, J. H. and J. Nsengimana (2011). "Two-stage analyses of sequence variants in association with quantitative traits." BMC Proc **5 Suppl 9**: S53.

Barrett, J. H., G. Smith, R. Waxman, N. Gooderham, T. Lightfoot, R. C. Garner*, et al.* (2003). "Investigation of interaction between N-acetyltransferase 2 and heterocyclic amines as potential risk factors for colorectal cancer." Carcinogenesis **24**(2): 275-282.

Begg, A. C. (2012). "Predicting recurrence after radiotherapy in head and neck cancer." Semin Radiat Oncol **22**(2): 108-118.

Bentley, J., C. P. Diggle, P. Harnden, M. A. Knowles and A. E. Kiltie (2004). "DNA double strand break repair in human bladder cancer is error prone and involves microhomology-associated end-joining." Nucleic Acids Res **32**(17): 5249-5259.

Bentley, J., C. L'Hote, F. Platt, C. D. Hurst, J. Lowery, C. Taylor*, et al.* (2009). "Papillary and muscle invasive bladder tumors with distinct genomic stability profiles have different DNA repair fidelity and KU DNA-binding activities." Genes Chromosomes Cancer **48**(4): 310-321.

Bergink, S., W. Toussaint, M. S. Luijsterburg, C. Dinant, S. Alekseev, J. H. Hoeijmakers*, et al.* (2012). "Recognition of DNA damage by XPC coincides with disruption of the XPC-RAD23 complex." J Cell Biol **196**(6): 681-688.

Bertagnolli, M. M., D. Niedzwiecki, C. C. Compton, H. P. Hahn, M. Hall, B. Damas*, et al.* (2009). "Microsatellite instability predicts improved response to adjuvant therapy with irinotecan, fluorouracil, and leucovorin in stage III colon cancer: Cancer and Leukemia Group B Protocol 89803." J Clin Oncol **27**(11): 1814-1821.

Berwick, M. and P. Vineis (2000). "Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review." J Natl Cancer Inst **92**(11): 874-897.

Biomarker Definitions Working Group (2001). "Biomarkers and surrogate endpoints: preferred definitions and conceptual framework." Clin Pharmacol Ther **69**(3): 89-95.

Blandy, J. (1998). Bladder Cancer. Lecture Notes on Urology. Oxford, Blackwell Publishing**:** 161-175.

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

Bodmer, W. and I. Tomlinson (2010). "Rare genetic variants and the risk of cancer." Curr Opin Genet Dev **20**(3): 262-267.

Bonilla, C., J. H. Lefevre, B. Winney, E. Johnstone, S. Tonks, C. Colas*, et al.* (2011). "Cyclin D1 rare variants in UK multiple adenoma and early-onset colorectal cancer patients." J Hum Genet **56**(1): 58-63.

Borgmann, K., B. Roper, R. El-Awady, S. Brackrock, M. Bigalke, T. Dork*, et al.* (2002). "Indicators of late normal tissue response after radiotherapy for head and neck cancer: fibroblasts, lymphocytes, genetics, DNA repair, and chromosome aberrations." Radiother Oncol **64**(2): 141-152.

Botteman, M. F., C. L. Pashos, A. Redaelli, B. Laskin and R. Hauser (2003). "The health economics of bladder cancer: a comprehensive review of the published literature." Pharmacoeconomics **21**(18): 1315-1330.

Brennecke, J., A. Stark, R. B. Russell and S. M. Cohen (2005). "Principles of microRNA-target recognition." PLoS Biol **3**(3): e85.

Bristow, R. and L. Harrington (2005). Genomic Stability and DNA Repair. The Basic Science of Oncology. I. F. Tannock, R. P. Hill, R. Bristow and L. Harrington. Toronto, McGraw-Hill Medical Publishing**:** 77-99.

British Association of Urological Surgeons and British Uro-oncology Group (2007). "MDT (Multi-disciplinary Team) Guidance for Managing Bladder Cancer."

Brown, T., R. Slack and L. Rushton (2012). "Occupational cancer in Britain." British journal of cancer **107 Suppl 1**: S76-84.

Buis, J., T. Stoneham, E. Spehalski and D. O. Ferguson (2012). "Mre11 regulates CtIP-dependent double-strand break repair by interaction with CDK2." Nat Struct Mol Biol **19**(2): 246-252.

Burnet, N. G., R. M. Elliott, A. Dunning and C. M. West (2006). "Radiosensitivity, radiogenomics and RAPPER." Clin Oncol **18**(7): 525-528.

Calabro, F. and C. N. Sternberg (2009). "Neoadjuvant and adjuvant chemotherapy in muscle-invasive bladder cancer." Eur Urol **55**(2): 348-358.

Camenisch, U., D. Trautlein, F. C. Clement, J. Fei, A. Leitenstorfer, E. Ferrando-May and H. Naegeli (2009). "Two-stage dynamic DNA quality check by xeroderma pigmentosum group C protein." Embo J **28**(16): 2387-2399.

Cancer Research UK (2010). "Bladder cancer statistics - Key Facts."

Capanu, M., P. Concannon, R. W. Haile, L. Bernstein, K. E. Malone, C. F. Lynch*, et al.* (2011). "Assessment of rare BRCA1 and BRCA2 variants of unknown significance using hierarchical modeling." Genet Epidemiol **35**(5): 389-397.

Carr, I. M., J. E. Morgan, C. P. Diggle, E. Sheridan, A. F. Markham, C. V. Logan*, et al.* (2011). "Illuminator, a desktop program for mutation detection using short-read clonal sequencing." Genomics **98**(4): 302-309.

Cartwright, R. A., R. W. Glashan, H. J. Rogers, R. A. Ahmad, D. Barham-Hall, E. Higgins and M. A. Kahn (1982). "Role of N-acetyltransferase phenotypes in bladder carcinogenesis: a pharmacogenetic epidemiological approach to bladder cancer." Lancet **2**(8303): 842-845.

Carvajal-Carmona, L. G. (2010). "Challenges in the identification and use of rare disease-associated predisposition variants." Curr Opin Genet Dev.

Castillejo, A., A. B. Sanchez-Heras, R. Jover, M. I. Castillejo, C. Guarinos, S. Oltra*, et al.* (2012). "Recurrent Testicular Germ Cell Tumors in a Family With MYH-Associated Polyposis." J Clin Oncol **30**(23): e216-217.

Castillo-Martin, M., J. Domingo-Domenech, O. Karni-Schmidt, T. Matos and C. Cordon-Cardo (2010). "Molecular pathways of urothelial development and bladder tumorigenesis." Urol Oncol **28**(4): 401-408.

Catto, J. W., A. Alcaraz, A. S. Bjartell, R. De Vere White, C. P. Evans, S. Fussel*, et al.* (2011). "MicroRNA in Prostate, Bladder, and Kidney Cancer: A Systematic Review." Eur Urol.

Catto, J. W., S. Miah, H. C. Owen, H. Bryant, K. Myers, E. Dudziec*, et al.* (2009). "Distinct microRNA alterations characterize high- and low-grade bladder cancer." Cancer Res **69**(21): 8472-8481.

Chaplin, D. J., C. E. Peters, M. R. Horsman and M. J. Trotter (1991). "Drug induced perturbations in tumor blood flow: therapeutic potential and possible limitations." Radiother Oncol **20 Suppl 1**: 93-101.

Cheng, Q., N. Barboule, P. Frit, D. Gomez, O. Bombarde, B. Couderc*, et al.* (2011). "Ku counteracts mobilization of PARP1 and MRN in chromatin damaged with DNA double-strand breaks." Nucleic Acids Res **39**(22): 9605-9619.

Chi, S. W., G. J. Hannon and R. B. Darnell (2012). "An alternative mode of microRNA target recognition." Nat Struct Mol Biol **19**(3): 321-327.

Chin, L. J., E. Ratner, S. Leng, R. Zhai, S. Nallur, I. Babar*, et al.* (2008). "A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk." Cancer Res **68**(20): 8535-8540.

Choudhury, A., F. Elliott, M. M. Iles, M. Churchman, R. G. Bristow, D. T. Bishop and A. E. Kiltie (2008). "Analysis of variants in DNA damage signalling genes in bladder cancer." BMC Med Genet **9**: 69.

Choudhury, A., L. D. Nelson, M. T. Teo, S. Chilka, S. Bhattarai, C. F. Johnston*, et al.* (2010). "MRE11 expression is predictive of cause-specific survival following radical radiotherapy for muscle-invasive bladder cancer." Cancer Res **70**(18): 7017-7026.

Choudhury, A., L. D. Nelson, M. T. Teo, S. Chilka, S. Bhattarai, C. F. Johnston*, et al.* (2010). "MRE11 expression is predictive of cause-specific survival following radical radiotherapy for muscle-invasive bladder cancer." Cancer Res **70**(18): 7017-7026.

Choudhury, A., R. Swindell, J. P. Logue, P. A. Elliott, J. E. Livsey, M. Wise*, et al.* (2011). "Phase II Study of Conformal Hypofractionated Radiotherapy With Concurrent Gemcitabine in Muscle-Invasive Bladder Cancer." J Clin Oncol **29**(6): 733-738.

Ciampa, J., M. Yeager, L. Amundadottir, K. Jacobs, P. Kraft, C. Chung*, et al.* (2011). "Large-scale exploration of gene-gene interactions in prostate cancer using a multistage genome-wide association study." Cancer Res **71**(9): 3287-3295.

Cirulli, E. T. and D. B. Goldstein (2010). "Uncovering the roles of rare variants in common disease through whole-genome sequencing." Nat Rev Genet **11**(6): 415-425.

Compton, S., C. Kim, N. B. Griner, P. Potluri, I. E. Scheffler, S. Sen*, et al.* (2011). "Mitochondrial dysfunction impairs tumor suppressor p53 expression/function." J Biol Chem **286**(23): 20297-20312.

Cooper, G. M. and J. Shendure (2011). "Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data." Nat Rev Genet **12**(9): 628-640.

Coppin, C. M., M. K. Gospodarowicz, K. James, I. F. Tannock, B. Zee, J. Carson*, et al.* (1996). "Improved local control of invasive bladder cancer by concurrent cisplatin and preoperative or definitive radiation. The National Cancer Institute of Canada Clinical Trials Group." J Clin Oncol **14**(11): 2901-2907.

Cordell, H. J. (2009). "Detecting gene-gene interactions that underlie human diseases." Nat Rev Genet **10**(6): 392-404.

Coventry, A., L. M. Bull-Otterson, X. Liu, A. G. Clark, T. J. Maxwell, J. Crosby*, et al.* (2010). "Deep resequencing reveals excess rare recent variants consistent with explosive population growth." Nat Commun **1**: 131.

Cox, J. D., J. Stetz and T. F. Pajak (1995). "Toxicity criteria of the Radiation Therapy Oncology Group (RTOG) and the European Organization for Research and Treatment of Cancer (EORTC)." Int J Radiat Oncol Biol Phys **31**(5): 1341-1346.

Craig, D. W., J. V. Pearson, S. Szelinger, A. Sekar, M. Redman, J. J. Corneveaux*, et al.* (2008). "Identification of genetic variants using bar-coded multiplexed sequencing." Nat Methods **5**(10): 887-893.

D'Errico, M., E. Parlanti, M. Teson, B. M. de Jesus, P. Degan, A. Calcagnile*, et al.* (2006). "New functions of XPC in the protection of human skin cells from oxidative damage." Embo J **25**(18): 4305-4315.

David, S. S., V. L. O'Shea and S. Kundu (2007). "Base-excision repair of oxidative DNA damage." Nature **447**(7147): 941-950.

Delia, D., M. Piane, G. Buscemi, C. Savio, S. Palmeri, P. Lulli*, et al.* (2004). "MRE11 mutations and impaired ATM-dependent responses in an Italian family with ataxia-telangiectasia-like disorder." Hum Mol Genet **13**(18): 2155-2163.

Della-Maria, J., Y. Zhou, M. S. Tsai, J. Kuhnlein, J. P. Carney, T. T. Paull and A. E. Tomkinson (2011). "Human Mre11/human Rad50/Nbs1 and DNA ligase IIIalpha/XRCC1 protein complexes act together in an alternative nonhomologous end joining pathway." J Biol Chem **286**(39): 33845-33853.

Dering, C., C. Hemmelmann, E. Pugh and A. Ziegler (2011). "Statistical analysis of rare sequence variants: an overview of collapsing methods." Genet Epidemiol **35 Suppl 1**: S12-17.

Dering, C., A. Ziegler, I. R. Konig and C. Hemmelmann (2011). "Comparison of collapsing methods for the statistical analysis of rare variants." BMC Proc **5 Suppl 9**: S115.

Despras, E., P. Pfeiffer, B. Salles, P. Calsou, S. Kuhfittig-Kulle, J. F. Angulo and D. S. Biard (2007). "Long-term XPC silencing reduces DNA double-strand break repair." Cancer Res **67**(6): 2526-2534.

Detre, S., G. Saclani Jotti and M. Dowsett (1995). "A "quickscore" method for immunohistochemical semiquantitation: validation for oestrogen receptor in breast carcinomas." J Clin Pathol **48**(9): 876-878.

Dickson, S. P., K. Wang, I. Krantz, H. Hakonarson and D. B. Goldstein (2010). "Rare variants create synthetic genome-wide associations." PLoS Biol **8**(1): e1000294.

Ding, L., M. C. Wendl, D. C. Koboldt and E. R. Mardis (2010). "Analysis of next-generation genomic data in cancer: accomplishments and challenges." Hum Mol Genet **19**(R2): R188-196.

Doherty, J. A., N. S. Weiss, S. Fish, W. Fan, M. M. Loomis, L. C. Sakoda*, et al.* (2011). "Polymorphisms in nucleotide excision repair genes and endometrial cancer risk." Cancer Epidemiol Biomarkers Prev **20**(9): 1873-1882.

Dong, J., Z. Hu, Y. Shu, S. Pan, W. Chen, Y. Wang*, et al.* (2012). "Potentially functional polymorphisms in DNA repair genes and non-small-cell lung cancer survival: a pathway-based analysis." Mol Carcinog **51**(7): 546-552.

Doyle, G. A., A. T. Lai, A. D. Chou, M. J. Wang, X. Gai, E. F. Rappaport and W. H. Berrettini (2012). "Re-sequencing of ankyrin 3 exon 48 and case-control association analysis of rare variants in bipolar disorder type I." Bipolar Disord.

Dyrskjot, L., M. S. Ostenfeld, J. B. Bramsen, A. N. Silahtaroglu, P. Lamy, R. Ramanathan*, et al.* (2009). "Genomic profiling of microRNAs in bladder cancer: miR-129 is associated with poor outcome and promotes cell death in vitro." Cancer Res **69**(11): 4851-4860.

Easton, D. F., A. M. Deffenbaugh, D. Pruss, C. Frye, R. J. Wenstrup, K. Allen-Brady*, et al.* (2007). "A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes." Am J Hum Genet **81**(5): 873-883.

Eccles, L. J., P. O'Neill and M. E. Lomax (2011). "Delayed repair of radiation induced clustered DNA damage: friend or foe?" Mutat Res **711**(1-2): 134-141.

Edge, S. B., D. R. Byrd, C. C. Compton, A. G. Fritz, F. L. Greene and A. E. Trotti (2010). "Urinary bladder." AJCC Cancer Staging Manual 7th edition: 497-505.

Efstathiou, J. A., D. Y. Spiegel, W. U. Shipley, N. M. Heney, D. S. Kaufman, A. Niemierko*, et al.* (2012). "Long-Term Outcomes of Selective Bladder Preservation by Combined-Modality Therapy for Invasive Bladder Cancer: The MGH Experience." Eur Urol **61**(4): 705-711.

Eid, W., M. Steger, M. El-Shemerly, L. P. Ferretti, J. Pena-Diaz, C. Konig*, et al.* (2010). "DNA end resection by CtIP and exonuclease 1 prevents genomic instability." EMBO Rep **11**(12): 962-968.

Esquela-Kerscher, A. and F. J. Slack (2006). "Oncomirs - microRNAs with a role in cancer." Nat Rev Cancer **6**(4): 259-269.

Esteller, M. (2011). "Non-coding RNAs in human disease." Nat Rev Genet **12**(12): 861-874.

Eswara, J. R., J. A. Efstathiou, N. M. Heney, J. Paly, D. S. Kaufman, W. S. McDougal*, et al.* (2012). "Complications and long-term results of salvage cystectomy after failed bladder sparing therapy for muscle invasive bladder cancer." J Urol **187**(2): 463-468.

Evans, A. R., M. Limp-Foster and M. R. Kelley (2000). "Going APE over ref-1." Mutat Res **461**(2): 83-108.

Evans, D. A. and T. A. White (1964). "Human Acetylation Polymorphism." J Lab Clin Med **63**: 394-403.

Farrington, S. M., A. Tenesa, R. Barnetson, A. Wiltshire, J. Prendergast, M. Porteous*, et al.* (2005). "Germline susceptibility to colorectal cancer due to base-excision repair gene defects." Am J Hum Genet **77**(1): 112-119.

Fattah, F., E. H. Lee, N. Weisensel, Y. Wang, N. Lichter and E. A. Hendrickson (2010). "Ku regulates the non-homologous end joining pathway choice of DNA double-strand break repair in human somatic cells." PLoS Genet **6**(2): e1000855.

Fearnhead, N. S., J. L. Wilding, B. Winney, S. Tonks, S. Bartlett, D. C. Bicknell*, et al.* (2004). "Multiple rare variants in different genes account for multifactorial inherited susceptibility to colorectal adenomas." Proc Natl Acad Sci U S A **101**(45): 15992-15997.

Figueroa, J. D., N. Malats, F. X. Real, D. Silverman, M. Kogevinas, S. Chanock*, et al.* (2007). "Genetic variation in the base excision repair pathway and bladder cancer risk." Hum Genet **121**(2): 233-242.

Figueroa, J. D., N. Malats, N. Rothman, F. X. Real, D. Silverman, M. Kogevinas*, et al.* (2007). "Evaluation of genetic variation in the double-strand break repair pathway and bladder cancer risk." Carcinogenesis **28**(8): 1788-1793.

Flanagan, S. E., A. M. Patch and S. Ellard (2010). "Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations." Genet Test Mol Biomarkers **14**(4): 533-537.

Fleischmann, C., J. Peto, J. Cheadle, B. Shah, J. Sampson and R. S. Houlston (2004). "Comprehensive analysis of the contribution of germline MYH variation to early-onset colorectal cancer." Int J Cancer **109**(4): 554-558.

Fleming, N. D., H. Agadjanian, H. Nassanian, C. W. Miller, S. Orsulic, B. Y. Karlan and C. S. Walsh (2012). "Xeroderma pigmentosum complementation group C single-nucleotide polymorphisms in the nucleotide excision repair pathway correlate with prolonged progression-free survival in advanced ovarian cancer." Cancer **118**(3): 689-697.

Fokdal, L., M. Hoyer, P. Meldgaard and H. von der Maase (2004). "Long-term bladder, colorectal, and sexual functions after radical radiotherapy for urinary bladder cancer." Radiother Oncol **72**(2): 139-145.

Fontana, L., R. Bosviel, L. Delort, L. Guy, N. Chalabi, F. Kwiatkowski*, et al.* (2008). "DNA repair gene ERCC2, XPC, XRCC1, XRCC3 polymorphisms and associations with bladder cancer risk in a French cohort." Anticancer Res **28**(3B): 1853-1856.

Freedman, N. D., D. T. Silverman, A. R. Hollenbeck, A. Schatzkin and C. C. Abnet (2011). "Association Between Smoking and Risk of Bladder Cancer Among Men and Women." JAMA **306**(7): 737-745.

Friedman, R. C., K. K. Farh, C. B. Burge and D. P. Bartel (2009). "Most mammalian mRNAs are conserved targets of microRNAs." Genome Res **19**(1): 92-105.

Fu, Y. P., I. Kohaar, N. Rothman, J. Earl, J. D. Figueroa, Y. Ye*, et al.* (2012). "Common genetic variants in the PSCA gene influence gene expression and bladder cancer risk." Proc Natl Acad Sci U S A.

Fukae, J., M. Takanashi, S. Kubo, K. Nishioka, Y. Nakabeppu, H. Mori*, et al.* (2005). "Expression of 8-oxoguanine DNA glycosylase (OGG1) in Parkinson's disease and related neurodegenerative disorders." Acta Neuropathol **109**(3): 256-262.

Gao, D., C. Wei, L. Chen, J. Huang, S. Yang and A. M. Diehl (2004). "Oxidative DNA damage and DNA repair enzyme expression are inversely related in murine models of fatty liver disease." Am J Physiol Gastrointest Liver Physiol **287**(5): G1070-1077.

Garcia-Closas, M., N. Malats, F. X. Real, R. Welch, M. Kogevinas, N. Chatterjee*, et al.* (2006). "Genetic variation in the nucleotide excision repair pathway and bladder cancer risk." Cancer Epidemiol Biomarkers Prev **15**(3): 536-542.

Garcia-Closas, M., N. Malats, D. Silverman, M. Dosemeci, M. Kogevinas, D. W. Hein*, et al.* (2005). "NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses." Lancet **366**(9486): 649-659.

Garcia-Closas, M., Y. Ye, N. Rothman, J. D. Figueroa, N. Malats, C. P. Dinney*, et al.* (2011). "A genome-wide association study of bladder cancer identifies a new susceptibility locus within SLC14A1, a urea transporter gene on chromosome 18q12.3." Hum Mol Genet **20**(21): 4282-4289.

Gibson, G. (2011). "Rare and common variants: twenty arguments." Nat Rev Genet **13**(2): 135-145.

Gil, J., D. Ramsey, A. Stembalska, P. Karpinski, K. A. Pesz, I. Laczmanska*, et al.* (2012). "The C/A polymorphism in intron 11 of the XPC gene plays a crucial role in the modulation of an individual's susceptibility to sporadic colorectal cancer." Mol Biol Rep **39**(1): 527-534.

Glashan, R. W. and R. A. Cartwright (1981). "Occupational bladder cancer and cigarette smoking in West Yorkshire." Br J Urol **53**(6): 602-604.
Goebell, P. J., S. G. Groshen and B. J. Schmitz-Drager (2010). "p53 immunohistochemistry in bladder cancer--a new approach to an old question." Urol Oncol **28**(4): 377-388.
Goebell, P. J. and M. A. Knowles (2010). "Bladder cancer or bladder cancers? Genetically distinct malignant conditions of the urothelium." Urol Oncol **28**(4): 409-428.
Goldgar, D. E., S. Healey, J. G. Dowty, L. Da Silva, X. Chen, A. B. Spurdle*, et al.* (2011). "Rare variants in the ATM gene and risk of breast cancer." Breast Cancer Res **13**(4): R73.
Goode, E. L., C. M. Ulrich and J. D. Potter (2002). "Polymorphisms in DNA repair genes and associations with cancer risk." Cancer Epidemiol Biomarkers Prev **11**(12): 1513-1530.
Gorlov, I., O. Gorlova, M. Frazier, M. Spitz and C. Amos (2011). "Evolutionary evidence of the effect of rare variants on disease etiology." Clin Genet **79**(3): 199-206.
Griffiths-Jones, S., R. J. Grocock, S. van Dongen, A. Bateman and A. J. Enright (2006). "miRBase: microRNA sequences, targets and gene nomenclature." Nucleic Acids Res **34**(Database issue): D140-144.
Grignon, D. J. (2009). "The current classification of urothelial neoplasms." Mod Pathol **22 Suppl 2**: S60-69.
Grunstein, M. (1997). "Histone acetylation in chromatin structure and transcription." Nature **389**(6649): 349-352.
Gu, B. and P. L. Chen (2009). "Expression of PCNA-binding domain of CtIP, a motif required for CtIP localization at DNA replication foci, causes DNA damage and activation of DNA damage checkpoint." Cell Cycle **8**(9): 1409-1420.
Gulston, M., C. de Lara, T. Jenner, E. Davis and P. O'Neill (2004). "Processing of clustered DNA damage generates additional double-strand breaks in mammalian cells post-irradiation." Nucleic Acids Res **32**(4): 1602-1609.
Haince, J. F., D. McDonald, A. Rodrigue, U. Dery, J. Y. Masson, M. J. Hendzel and G. G. Poirier (2008). "PARP1-dependent kinetics of recruitment of MRE11 and NBS1 proteins to multiple DNA damage sites." J Biol Chem **283**(2): 1197-1208.
Hall, E. J. and A. J. Giaccia (2006). Repair of Radiation Damage and the Dose-Rate Effect. Radiobiology for the Radiologist. Philadelphia, Lippincott Williams & Wilkins**:** 60-64.
Hammond, S. M. (2007). "MicroRNAs as tumor suppressors." Nat Genet **39**(5): 582-583.
Han, F. and W. Pan (2010). "A data-adaptive sum test for disease association with multiple common or rare variants." Hum Hered **70**(1): 42-54.
Hanahan, D. and R. A. Weinberg (2011). "Hallmarks of cancer: the next generation." Cell **144**(5): 646-674.
Harakalova, M., I. J. Nijman, J. Medic, M. Mokry, I. Renkens, J. D. Blankensteijn*, et al.* (2011). "Genomic DNA Pooling Strategy for Next-Generation Sequencing-Based Rare Variant Discovery in Abdominal Aortic Aneurysm Regions of Interest-Challenges and Limitations." J Cardiovasc Transl Res **4**(3): 271-280.
Harismendy, O., P. C. Ng, R. L. Strausberg, X. Wang, T. B. Stockwell, K. Y. Beeson*, et al.* (2009). "Evaluation of next generation sequencing platforms for population targeted sequencing studies." Genome Biol **10**(3): R32.
Harris, A. L. and D. E. Neal (1992). "Bladder cancer--field versus clonal origin." N Engl J Med **326**(11): 759-761.
Harris, J. K., J. W. Sahl, T. A. Castoe, B. D. Wagner, D. D. Pollock and J. R. Spear (2010). "Comparison of normalization methods for construction of large,

multiplex amplicon pools for next-generation sequencing." Appl Environ Microbiol **76**(12): 3863-3868.
Harris, R. E., J. Y. Chen-Backlund and E. L. Wynder (1990). "Cancer of the urinary bladder in blacks and whites. A case-control study." Cancer **66**(12): 2673-2680.
Harrison, L., K. L. Brame, L. E. Geltz and A. M. Landry (2006). "Closely opposed apurinic/apyrimidinic sites are converted to double strand breaks in Escherichia coli even in the absence of exonuclease III, endonuclease IV, nucleotide excision repair and AP lyase cleavage." DNA repair **5**(3): 324-335.
Hegde, M. L., T. K. Hazra and S. Mitra (2008). "Early steps in the DNA base excision/single-strand interruption repair pathway in mammalian cells." Cell Res **18**(1): 27-47.
Heinzen, E. L., C. Depondt, G. L. Cavalleri, E. K. Ruzzo, N. M. Walley, A. C. Need*, et al.* (2012). "Exome sequencing followed by large-scale genotyping fails to identify single rare variants of large effect in idiopathic generalized epilepsy." Am J Hum Genet **91**(2): 293-302.
Hernandez, S., E. Lopez-Knowles, J. Lloreta, M. Kogevinas, A. Amoros, A. Tardon*, et al.* (2006). "Prospective study of FGFR3 mutations as a prognostic factor in nonmuscle invasive urothelial bladder carcinomas." J Clin Oncol **24**(22): 3664-3671.
Hindorff, L. A., E. M. Gillanders and T. A. Manolio (2011). "Genetic architecture of cancer and other complex diseases: lessons learned and future directions." Carcinogenesis **32**(7): 945-954.
Hindson, B. R., S. L. Turner, J. L. Millar, F. Foroudi, N. K. Gogna, M. Skala*, et al.* (2012). "Australian & New Zealand Faculty of Radiation Oncology Genito-Urinary Group: 2011 consensus guidelines for curative radiotherapy for urothelial carcinoma of the bladder." J Med Imaging Radiat Oncol **56**(1): 18-30.
Hoeijmakers, J. H. (2001). "Genome maintenance mechanisms for preventing cancer." Nature **411**(6835): 366-374.
Horikawa, Y., C. G. Wood, H. Yang, H. Zhao, Y. Ye, J. Gu*, et al.* (2008). "Single nucleotide polymorphisms of microRNA machinery genes modify the risk of renal cell carcinoma." Clin Cancer Res **14**(23): 7956-7962.
Hoskin, P. J., A. M. Rojas, S. M. Bentzen and M. I. Saunders (2010). "Radiotherapy with concurrent carbogen and nicotinamide in bladder carcinoma." J Clin Oncol **28**(33): 4912-4918.
Hoskin, P. J., A. M. Rojas, S. M. Bentzen and M. I. Saunders (2010). "Radiotherapy with concurrent carbogen and nicotinamide in bladder carcinoma." J Clin Oncol **28**(33): 4912-4918.
Huang, L., J. Luo, Q. Cai, Q. Pan, H. Zeng, Z. Guo*, et al.* (2011). "MicroRNA-125b suppresses the development of bladder cancer by targeting E2F3." Int J Cancer **128**(8): 1758-1769.
Huddart, R., E. Hall, R. Lewis, A. Birtle and SPARE Trial Management Group (2010). "Life and death of spare (selective bladder preservation against radical excision): reflections on why the spare trial closed." BJU Int **106**(6): 753-755.
Huddart, R. A., N. D. James, F. Adab, I. Syndikus, P. Jenkins, C. Rawlings*, et al.* (2009). "BC2001: A multicenter phase III randomized trial of standard versus reduced volume radiotherapy for muscle invasive bladder cancer (ISCRTN:68324339)." J Clin Oncol **27**(15s (suppl; abstr 5022)).
Human Genomics Strategy Group (2012). Building on our inheritance. Genomic technology in healthcare. D. o. Health. United Kingdom.
Hung, R. J., J. Hall, P. Brennan and P. Boffetta (2005). "Genetic polymorphisms in the base excision repair pathway and cancer risk: a HuGE review." Am J Epidemiol **162**(10): 925-942.

Ichimi, T., H. Enokida, Y. Okuno, R. Kunimoto, T. Chiyomaru, K. Kawamoto*, et al.* (2009). "Identification of novel microRNA targets based on microRNA signatures in bladder cancer." Int J Cancer **125**(2): 345-352.

Iijima, K., M. Ohara, R. Seki and H. Tauchi (2008). "Dancing on damaged chromatin: functions of ATM and the RAD50/MRE11/NBS1 complex in cellular responses to DNA damage." J Radiat Res (Tokyo) **49**(5): 451-464.

Iles, M. M. (2008). "What can genome-wide association studies tell us about the genetics of common disease?" PLoS Genet **4**(2): e33.

Ionita-Laza, I., J. D. Buxbaum, N. M. Laird and C. Lange (2011). "A new testing strategy to identify rare variants with either risk or protective effect on disease." PLoS Genet **7**(2): e1001289.

Jadersten, M., L. Saft, A. Smith, A. Kulasekararaj, S. Pomplun, G. Gohring*, et al.* (2011). "TP53 mutations in low-risk myelodysplastic syndromes with del(5q) predict disease progression." J Clin Oncol **29**(15): 1971-1979.

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

James, N. and S. A. Hussain (2005). "Management of muscle invasive bladder cancer--British approaches to organ conservation." Semin Radiat Oncol **15**(1): 19-27.

James, N. and A. Zarkar (2006). Bladder cancer. Radiotherapy in Practice: External Beam Therapy. P. J. Hoskin. Oxford, Oxford University Press**:** 213-221.

James, N. D., S. A. Hussain, E. Hall, P. Jenkins, J. Tremlett, C. Rawlings*, et al.* (2012). "Radiotherapy with or without chemotherapy in muscle-invasive bladder cancer." N Engl J Med **366**(16): 1477-1488.

Jebar, A. H., C. D. Hurst, D. C. Tomlinson, C. Johnston, C. F. Taylor and M. A. Knowles (2005). "FGFR3 and Ras gene mutations are mutually exclusive genetic events in urothelial cell carcinoma." Oncogene **24**(33): 5218-5225.

Jensen, J. B., P. P. Munksgaard, C. M. Sorensen, N. Fristrup, K. Birkenkamp-Demtroder, B. P. Ulhoi*, et al.* (2011). "High expression of karyopherin-alpha2 defines poor prognosis in non-muscle-invasive bladder cancer and in patients with invasive bladder cancer undergoing radical cystectomy." Eur Urol **59**(5): 841-848.

Jiao, X., J. Ren, H. Chen, J. Ma, S. Rao, K. Huang*, et al.* (2011). "Ala499Val (C>T) and Lys939Gln (A>C) polymorphisms of the XPC gene: their correlation with the risk of primary gallbladder adenocarcinoma--a case-control study in China." Carcinogenesis **32**(4): 496-501.

Jin, S. C., P. Pastor, B. Cooper, S. Cervantes, B. A. Benitez, C. Razquin*, et al.* (2012). "Pooled-DNA sequencing identifies novel causative variants in PSEN1, GRN and MAPT in a clinical early-onset and familial Alzheimer's disease Ibero-American cohort." Alzheimers Res Ther **4**(4): 34.

Johansen, C. T., J. Wang, M. B. Lanktree, H. Cao, A. D. McIntyre, M. R. Ban*, et al.* (2010). "Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia." Nat Genet **42**(8): 684-687.

John, B., A. J. Enright, A. Aravin, T. Tuschl, C. Sander and D. S. Marks (2004). "Human MicroRNA targets." PLoS Biol **2**(11): e363.

Jones, T. D., M. Wang, J. N. Eble, G. T. MacLennan, A. Lopez-Beltran, S. Zhang*, et al.* (2005). "Molecular evidence supporting field effect in urothelial carcinogenesis." Clin Cancer Res **11**(18): 6512-6519.

Ju, Y. S., J. I. Kim, S. Kim, D. Hong, H. Park, J. Y. Shin*, et al.* (2011). "Extensive genomic and transcriptional diversity identified through massively parallel DNA and RNA sequencing of eighteen Korean individuals." Nat Genet **43**(8): 745-752.

Kang, G., D. Lin, H. Hakonarson and J. Chen (2012). "Two-stage extreme phenotype sequencing design for discovering and testing common and rare genetic variants: efficiency and power." Hum Hered **73**(3): 139-147.

Kantor, A. F., P. Hartge, R. N. Hoover and J. F. Fraumeni, Jr. (1985). "Familial and environmental interactions in bladder cancer risk." Int J Cancer **35**(6): 703-706.

Karakiewicz, P. I., S. F. Shariat, G. S. Palapattu, A. E. Gilad, Y. Lotan, C. G. Rogers*, et al.* (2006). "Nomogram for predicting disease recurrence after radical cystectomy for transitional cell carcinoma of the bladder." J Urol **176**(4 Pt 1): 1354-1361; discussion 1361-1352.

Kase, M., M. Vardja, A. Lipping, T. Asser and J. Jaal (2011). "Impact of PARP-1 and DNA-PK expression on survival in patients with glioblastoma multiforme." Radiother Oncol **101**(1): 127-131.

Kass, E. M. and M. Jasin (2010). "Collaboration and competition between DNA double-strand break repair pathways." FEBS Lett **584**(17): 3703-3708.

Kawashima, A., M. Nakayama, Y. Kakuta, T. Abe, K. Hatano, M. Mukai*, et al.* (2011). "Excision repair cross-complementing group 1 may predict the efficacy of chemoradiation therapy for muscle-invasive bladder cancer." Clin Cancer Res **17**(8): 2561-2569.

Keimling, M., M. Deniz, D. Varga, A. Stahl, H. Schrezenmeier, R. Kreienberg*, et al.* (2012). "The power of DNA double-strand break (DSB) repair testing to predict breast cancer susceptibility." Faseb J **26**(5): 2094-2104.

Kelleher, R. J., 3rd, U. Geigenmuller, H. Hovhannisyan, E. Trautman, R. Pinard, B. Rathmell*, et al.* (2012). "High-throughput sequencing of mGluR signaling pathway genes reveals enrichment of rare variants in autism." PLoS One **7**(4): e35003.

Kenny, E. M., P. Cormican, W. P. Gilks, A. S. Gates, C. T. O'Dushlaine, C. Pinto*, et al.* (2010). "Multiplex Target Enrichment Using DNA Indexing for Ultra-High Throughput SNP Detection." DNA Res **18**(1): 31-38.

Kiemeney, L. A., P. Sulem, S. Besenbacher, S. H. Vermeulen, A. Sigurdsson, G. Thorleifsson*, et al.* (2010). "A sequence variant at 4p16.3 confers susceptibility to urinary bladder cancer." Nat Genet **42**(5): 415-419.

Kiemeney, L. A., S. Thorlacius, P. Sulem, F. Geller, K. K. Aben, S. N. Stacey*, et al.* (2008). "Sequence variant on 8q24 confers susceptibility to urinary bladder cancer." Nat Genet **40**(11): 1307-1312.

Kiltie, A. E. (2009). "Molecular epidemiology of DNA repair genes in bladder cancer." Methods Mol Biol **472**: 281-306.

Kiltie, A. E. (2010). "Common predisposition alleles for moderately common cancers: bladder cancer." Curr Opin Genet Dev **20**(3): 218-224.

Kim, G. J., K. Chandrasekaran and W. F. Morgan (2006). "Mitochondrial dysfunction, persistently elevated levels of reactive oxygen species and radiation-induced genomic instability: a review." Mutagenesis **21**(6): 361-367.

Kim, S. Y., Y. Li, Y. Guo, R. Li, J. Holmkvist, T. Hansen*, et al.* (2010). "Design of association studies with pooled or un-pooled next-generation sequencing data." Genet Epidemiol **34**(5): 479-491.

Kiriakidou, M., P. T. Nelson, A. Kouranov, P. Fitziev, C. Bouyioukos, Z. Mourelatos and A. Hatzigeorgiou (2004). "A combined computational-experimental approach predicts human microRNA targets." Genes Dev **18**(10): 1165-1178.

Klopp, A. H. and P. J. Eifel (2012). "Biological predictors of cervical cancer response to radiation therapy." Semin Radiat Oncol **22**(2): 143-150.

Knowles, M. A. (2006). "Molecular subtypes of bladder cancer: Jekyll and Hyde or chalk and cheese?" Carcinogenesis **27**(3): 361-373.

Kogevinas, M., M. Sala, P. Boffetta, N. Kazerouni, H. Kromhout and S. Hoar-Zahm (1998). "Cancer risk in the rubber industry: a review of the recent epidemiological evidence." Occup Environ Med **55**(1): 1-12.

Kohlmann, A., V. Grossmann, H. U. Klein, S. Schindela, T. Weiss, B. Kazak*, et al.* (2010). "Next-generation sequencing technology reveals a characteristic pattern of molecular mutations in 72.8% of chronic myelomonocytic leukemia by detecting frequent alterations in TET2, CBL, RAS, and RUNX1." J Clin Oncol **28**(24): 3858-3865.

Konig, I. R., J. Nsengimana, C. Papachristou, M. A. Simonson, K. Wang and J. A. Weisburd (2011). "Multiple testing in high-throughput sequence data: experiences from Group 8 of Genetic Analysis Workshop 17." Genet Epidemiol **35 Suppl 1**: S61-66.

Kotwal, S., A. Choudhury, C. Johnston, A. B. Paul, P. Whelan and A. E. Kiltie (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**(2): 456-463.

Kousholt, A. N., K. Fugger, S. Hoffmann, B. D. Larsen, T. Menzel, A. A. Sartori and C. S. Sorensen (2012). "CtIP-dependent DNA resection is required for DNA damage checkpoint maintenance but not initiation." J Cell Biol **197**(7): 869-876.

Kozak, K. R., M. Hamidi, M. Manning and J. S. Moody (2012). "Bladder Preservation for Localized Muscle-Invasive Bladder Cancer: The Survival Impact of Local Utilization Rates of Definitive Radiotherapy." Int J Radiat Oncol Biol Phys.

Krek, A., D. Grun, M. N. Poy, R. Wolf, L. Rosenberg, E. J. Epstein*, et al.* (2005). "Combinatorial microRNA target predictions." Nat Genet **37**(5): 495-500.

Krol, J., I. Loedige and W. Filipowicz (2010). "The widespread regulation of microRNA biogenesis, function and decay." Nat Rev Genet **11**(9): 597-610.

Kruger, J. and M. Rehmsmeier (2006). "RNAhybrid: microRNA target prediction easy, fast and flexible." Nucleic Acids Res **34**(Web Server issue): W451-454.

Kundu, S., M. K. Brinkmeyer, A. L. Livingston and S. S. David (2009). "Adenine removal activity and bacterial complementation with the human MutY homologue (MUTYH) and Y165C, G382D, P391L and Q324R variants associated with colorectal cancer." DNA Repair (Amst) **8**(12): 1400-1410.

Kurien, B. T., Y. Dorri, S. Dillon, A. Dsouza and R. H. Scofield (2011). "An overview of Western blotting for determining antibody specificities for immunohistochemistry." Methods Mol Biol **717**: 55-67.

La Thangue, N. B. and D. J. Kerr (2011). "Predictive biomarkers: a paradigm shift towards personalized cancer medicine." Nat Rev Clin Oncol **8**(10): 587-596.

Ladouceur, M., Z. Dastani, Y. S. Aulchenko, C. M. Greenwood and J. B. Richards (2012). "The empirical power of rare variant association methods: results from sanger sequencing in 1,998 individuals." PLoS Genet **8**(2): e1002496.

Lai, T. C., K. C. Chow, H. Y. Fang, H. C. Cho, C. Y. Chen, T. Y. Lin*, et al.* (2011). "Expression of xeroderma pigmentosum complementation group C protein predicts cisplatin resistance in lung adenocarcinoma patients." Oncol Rep **25**(5): 1243-1251.

Lal, A., F. Navarro, C. A. Maher, L. E. Maliszewski, N. Yan, E. O'Day*, et al.* (2009). "miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3'UTR microRNA recognition elements." Mol Cell **35**(5): 610-625.

Lalani, S. R., C. Shaw, X. Wang, A. Patel, L. W. Patterson, K. Kolodziejska*, et al.* (2012). "Rare DNA copy number variants in cardiovascular malformations with extracardiac abnormalities." Eur J Hum Genet.

Lalondrelle, S., R. Huddart, K. Warren-Oseni, V. N. Hansen, H. McNair, K. Thomas*, et al.* (2011). "Adaptive-predictive organ localization using cone-beam computed tomography for improved accuracy in external beam radiotherapy for bladder cancer." Int J Radiat Oncol Biol Phys **79**(3): 705-712.

Lamm, D. L., B. A. Blumenstein, J. D. Crissman, J. E. Montie, J. E. Gottesman, B. A. Lowe*, et al.* (2000). "Maintenance bacillus Calmette-Guerin immunotherapy

for recurrent TA, T1 and carcinoma in situ transitional cell carcinoma of the bladder: a randomized Southwest Oncology Group Study." J Urol **163**(4): 1124-1129.

Landi, D., F. Gemignani, A. Naccarati, B. Pardini, P. Vodicka, L. Vodickova*, et al.* (2008). "Polymorphisms within micro-RNA-binding sites and risk of sporadic colorectal cancer." Carcinogenesis **29**(3): 579-584.

Landi, S., F. Gemignani, F. Canzian, V. Gaborieau, R. Barale, D. Landi*, et al.* (2006). "DNA repair and cell cycle control genes and the risk of young-onset lung cancer." Cancer Res **66**(22): 11062-11069.

Laurberg, J. R., A. S. Brems-Eskildsen, I. Nordentoft, N. Fristrup, T. Schepeler, B. P. Ulhoi*, 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 **[E-publication ahead of print]**.

Laurberg, J. R., A. S. Brems-Eskildsen, I. Nordentoft, N. Fristrup, T. Schepeler, B. P. Ulhoi*, 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.

Lavin, M. F. (2007). "ATM and the Mre11 complex combine to recognize and signal DNA double-strand breaks." Oncogene **26**(56): 7749-7758.

Le May, N., D. Mota-Fernandes, R. Velez-Cruz, I. Iltis, D. Biard and J. M. Egly (2010). "NER factors are recruited to active promoters and facilitate chromatin modification for transcription in the absence of exogenous genotoxic attack." Mol Cell **38**(1): 54-66.

Le Scodan, R., G. Cizeron-Clairac, E. Fourme, D. Meseure, S. Vacher, F. Spyratos*, et al.* (2010). "DNA repair gene expression and risk of locoregional relapse in breast cancer patients." Int J Radiat Oncol Biol Phys **78**(2): 328-336.

Lee, J. S., M. Choi, X. Yan, R. P. Lifton and H. Zhao (2011). "On optimal pooling designs to identify rare variants through massive resequencing." Genet Epidemiol.

Lefevre, J. H., C. Bonilla, C. Colas, B. Winney, E. Johnstone, S. Tonks*, et al.* (2012). "Role of rare variants in undetermined multiple adenomatous polyposis and early-onset colorectal cancer." J Hum Genet.

Lehmann, U., T. Streichert, B. Otto, C. Albat, B. Hasemeier, H. Christgen*, et al.* (2010). "Identification of differentially expressed microRNAs in human male breast cancer." BMC Cancer **10**: 109.

Lettice, L. A., S. J. Heaney, L. A. Purdie, L. Li, P. de Beer, B. A. Oostra*, et al.* (2003). "A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly." Hum Mol Genet **12**(14): 1725-1735.

Lewis, B. P., I. H. Shih, M. W. Jones-Rhoades, D. P. Bartel and C. B. Burge (2003). "Prediction of mammalian microRNA targets." Cell **115**(7): 787-798.

Li, B. and S. M. Leal (2008). "Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data." Am J Hum Genet **83**(3): 311-321.

Li, D., W. Zhang, J. Zhu, P. Chang, A. Sahin, E. Singletary*, et al.* (2001). "Oxidative DNA damage and 8-hydroxy-2-deoxyguanosine DNA glycosylase/apurinic lyase in human breast cancer." Mol Carcinog **31**(4): 214-223.

Li, Q., G. Zheng, X. Liang and K. Yu (2009). "Robust tests for single-marker analysis in case-control genetic association studies." Ann Hum Genet **73**(2): 245-252.

Li, X., Y. Zhang, H. Zhang, X. Liu, T. Gong, M. Li*, et al.* (2011). "miRNA-223 promotes gastric cancer invasion and metastasis by targeting tumor suppressor EPB41L3." Mol Cancer Res **9**(7): 824-833.

Liang, D., L. Meyer, D. W. Chang, J. Lin, X. Pu, Y. Ye*, et al.* (2010). "Genetic variants in MicroRNA biosynthesis pathways and binding sites modify ovarian cancer risk, survival, and treatment response." Cancer Res **70**(23): 9765-9776.

Liang, Z. L., E. K. Song, Y. B. Ko, N. R. Lee, H. Y. Yhim, H. T. Noh*, et al.* (2011). "Excision repair cross-complementation group 1 expression predicts response and survival in locally advanced cervical carcinoma patients treated with concurrent chemoradiotherapy." Histopathology **59**(3): 564-567.

Lipton, L., S. E. Halford, V. Johnson, M. R. Novelli, A. Jones, C. Cummings*, et al.* (2003). "Carcinogenesis in MYH-associated polyposis follows a distinct genetic pathway." Cancer Res **63**(22): 7595-7599.

Liu, D. J. and S. M. Leal (2010). "A novel adaptive method for the analysis of next-generation sequencing data to detect complex trait associations with rare variants due to gene main effects and interactions." PLoS Genet **6**(10): e1001156.

Liu, Y., H. Wang, T. Lin, Q. Wei, Y. Zhi, F. Yuan*, et al.* (2012). "Interactions between cigarette smoking and XPC-PAT genetic polymorphism enhance bladder cancer risk." Oncol Rep **28**(1): 337-345.

Liu, Z., G. Li, S. Wei, J. Niu, A. K. El-Naggar, E. M. Sturgis and Q. Wei (2010). "Genetic variants in selected pre-microRNA genes and the risk of squamous cell carcinoma of the head and neck." Cancer **116**(20): 4753-4760.

Liu, Z., A. Sall and D. Yang (2008). "MicroRNA: An emerging therapeutic target and intervention tool." Int J Mol Sci **9**(6): 978-999.

Lohmueller, K. E., C. L. Pearce, M. Pike, E. S. Lander and J. N. Hirschhorn (2003). "Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease." Nat Genet **33**(2): 177-182.

Lopez-Beltran, A., G. Sauter, T. Gasser, A. Hartmann, B. J. Schmitz-Dräger, B. Helpap*, et al.* (2004). Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs. Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs. J. N. Ebele, G. Sauter, J. I. Epstein and I. A. Sesterhenn. Lyon, IARCC Press**:** 29-34.

Lotta, L. A., M. Wang, J. Yu, I. Martinelli, F. Yu, S. M. Passamonti*, et al.* (2012). "Identification of genetic risk variants for deep vein thrombosis by multiplexed next-generation sequencing of 186 hemostatic/pro-inflammatory genes." BMC Med Genomics **5**: 7.

Lu, Q., C. Lu, G. P. Zhou, W. Zhang, H. Xiao and X. R. Wang (2010). "MicroRNA-221 silencing predisposed human bladder cancer cells to undergo apoptosis induced by TRAIL." Urol Oncol **28**(6): 635-641.

Lupton, M. K., P. Proitsi, M. Danillidou, M. Tsolaki, G. Hamilton, R. Wroe*, et al.* (2011). "Deep sequencing of the nicastrin gene in pooled DNA, the identification of genetic variants that affect risk of Alzheimer's disease." PLoS One **6**(2): e17298.

Madsen, B. E. and S. R. Browning (2009). "A groupwise association test for rare mutations using a weighted sum statistic." PLoS Genet **5**(2): e1000384.

Majewski, W. and R. Tarnawski (2009). "Acute and late toxicity in radical radiotherapy for bladder cancer." Clin Oncol **21**(8): 598-609.

Malyarchuk, S., K. L. Brame, R. Youngblood, R. Shi and L. Harrison (2004). "Two clustered 8-oxo-7,8-dihydroguanine (8-oxodG) lesions increase the point mutation frequency of 8-oxodG, but do not result in double strand breaks or deletions in Escherichia coli." Nucleic Acids Res **32**(19): 5721-5731.

Mamanova, L., A. J. Coffey, C. E. Scott, I. Kozarewa, E. H. Turner, A. Kumar*, et al.* (2010). "Target-enrichment strategies for next-generation sequencing." Nat Methods **7**(2): 111-118.

Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter*, et al.* (2009). "Finding the missing heritability of complex diseases." Nature **461**(7265): 747-753.

Margulis, V., Y. Lotan, P. I. Karakiewicz, Y. Fradet, R. Ashfaq, U. Capitanio*, et al.* (2009). "Multi-institutional validation of the predictive value of Ki-67 labeling

index in patients with urinary bladder cancer." J Natl Cancer Inst **101**(2): 114-119.
Mariappan, P., A. Zachou and K. M. Grigor (2010). "Detrusor muscle in the first, apparently complete transurethral resection of bladder tumour specimen is a surrogate marker of resection quality, predicts risk of early recurrence, and is dependent on operator experience." Eur Urol **57**(5): 843-849.
Marth, G. T., F. Yu, A. R. Indap, K. Garimella, S. Gravel, W. F. Leong*, et al.* (2011). "The functional spectrum of low-frequency coding variation." Genome Biol **12**(9): R84.
Martin, E. A. (1998). Oxford Concise Colour Medical Dictionary. Oxford, Oxford University Press**:** 101.
Matsushita, K., E. K. Cha, K. Matsumoto, S. Baba, T. F. Chromecki, H. Fajkovic*, et al.* (2011). "Immunohistochemical biomarkers for bladder cancer prognosis." Int J Urol **18**(9): 616-629.
McConkey, D. J., S. Lee, W. Choi, M. Tran, T. Majewski, A. Siefker-Radtke*, et al.* (2010). "Molecular genetics of bladder cancer: Emerging mechanisms of tumor initiation and progression." Urol Oncol **28**(4): 429-440.
Mechanic, L. E., H. S. Chen, C. I. Amos, N. Chatterjee, N. J. Cox, R. L. Divi*, et al.* (2012). "Next generation analytic tools for large scale genetic epidemiology studies of complex diseases." Genet Epidemiol **36**(1): 22-35.
Megnin-Chanet, F., M. A. Bollet and J. Hall (2010). "Targeting poly(ADP-ribose) polymerase activity for cancer therapy." Cell Mol Life Sci **67**(21): 3649-3662.
Menashe, I., J. D. Figueroa, M. Garcia-Closas, N. Chatterjee, N. Malats, A. Picornell*, et al.* (2012). "Large-scale pathway-based analysis of bladder cancer genome-wide association data from five studies of European background." PLoS One **7**(1): e29396.
Mhawech-Fauceglia, P., R. T. Cheney, G. Fischer, A. Beck and F. R. Herrmann (2006). "FGFR3 and p53 protein expressions in patients with pTa and pT1 urothelial bladder cancer." Eur J Surg Oncol **32**(2): 231-237.
Mignone, F., G. Grillo, F. Licciulli, M. Iacono, S. Liuni, P. J. Kersey*, et al.* (2005). "UTRdb and UTRsite: a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs." Nucleic Acids Res **33**(Database issue): D141-146.
Millan-Rodriguez, F., G. Chechile-Toniolo, J. Salvador-Bayarri, J. Palou, F. Algaba and J. Vicente-Rodriguez (2000). "Primary superficial bladder cancer risk groups according to progression, mortality and recurrence." J Urol **164**(3 Pt 1): 680-684.
Milne, R. L., M. M. Gaudet, A. B. Spurdle, P. A. Fasching, F. J. Couch, J. Benitez*, et al.* (2010). "Assessing interactions between the associations of common genetic susceptibility variants, reproductive history and body mass index with breast cancer risk in the breast cancer association consortium: a combined case-control study." Breast Cancer Res **12**(6): R110.
Mirnezami, R., J. Nicholson and A. Darzi (2011). "Preparing for Precision Medicine." N Engl J Med **366**(6): 489-491.
Moeller, B. J., J. S. Yordy, M. D. Williams, U. Giri, U. Raju, D. P. Molkentine*, et al.* (2011). "DNA repair biomarker profiling of head and neck cancer: Ku80 expression predicts locoregional failure and death following radiotherapy." Clin Cancer Res **17**(7): 2035-2043.
Morgan, J. E., I. M. Carr, E. Sheridan, C. E. Chu, B. Hayward, N. Camm*, et al.* (2010). "Genetic diagnosis of familial breast cancer using clonal sequencing." Hum Mutat **31**(4): 484-491.
Morgan, J. E., I. M. Carr, E. Sheridan, C. E. Chu, B. Hayward, N. Camm*, et al.* (2010). "Genetic diagnosis of familial breast cancer using clonal sequencing." Hum Mutat **31**(4): 484-491.

Morrison, A. S., J. E. Buring, W. G. Verhoek, K. Aoki, I. Leck, Y. Ohno and K. Obata (1984). "An international study of smoking and bladder cancer." J Urol **131**(4): 650-654.

Mostafa, M. H., S. A. Sheweita and P. J. O'Connor (1999). "Relationship between schistosomiasis and bladder cancer." Clin Microbiol Rev **12**(1): 97-111.

Mukherjee, B., H. Choy, C. Nirodi and S. Burma (2010). "Targeting nonhomologous end-joining through epidermal growth factor receptor inhibition: rationale and strategies for radiosensitization." Semin Radiat Oncol **20**(4): 250-257.

Murta-Nascimento, C., D. T. Silverman, M. Kogevinas, M. Garcia-Closas, N. Rothman, A. Tardon*, et al.* (2007). "Risk of bladder cancer associated with family history of cancer: do low-penetrance polymorphisms account for the increase in risk?" Cancer Epidemiol Biomarkers Prev **16**(8): 1595-1600.

Naccarati, A., B. Pardini, S. Landi, D. Landi, J. Slyskova, J. Novotny*, et al.* (2012). "Polymorphisms in miRNA binding sites of nucleotide excision repair genes and colorectal cancer risk." Carcinogenesis.

Naccarati, A., B. Pardini, L. Stefano, D. Landi, J. Slyskova, J. Novotny*, et al.* (2012). "Polymorphisms in miRNA-binding sites of nucleotide excision repair genes and colorectal cancer risk." Carcinogenesis **33**(7): 1346-1351.

Nelson, M. R., D. Wegmann, M. G. Ehm, D. Kessner, P. St Jean, C. Verzilli*, et al.* (2012). "An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people." Science **337**(6090): 100-104.

Ng, P. C. and S. Henikoff (2006). "Predicting the effects of amino acid substitutions on protein function." Annu Rev Genomics Hum Genet **7**: 61-80.

Nicoloso, M. S., H. Sun, R. Spizzo, H. Kim, P. Wickramasinghe, M. Shimizu*, et al.* (2010). "Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility." Cancer Res **70**(7): 2789-2798.

Nijman, I. J., M. Mokry, R. van Boxtel, P. Toonen, E. de Bruijn and E. Cuppen (2010). "Mutation discovery by targeted genomic enrichment of multiplexed barcoded samples." Nat Methods **7**(11): 913-915.

Nishiyama, H., J. Watanabe and O. Ogawa (2008). "p53 and chemosensitivity in bladder cancer." Int J Clin Oncol **13**(4): 282-286.

Nuhn, P., M. May, M. Sun, H. M. Fritsche, S. Brookman-May, A. Buchner*, et al.* (2012). "External validation of postoperative nomograms for prediction of all-cause mortality, cancer-specific mortality, and recurrence in patients with urothelial carcinoma of the bladder." Eur Urol **61**(1): 58-64.

Ohtsubo, T., K. Nishioka, Y. Imaiso, S. Iwai, H. Shimokawa, H. Oda*, et al.* (2000). "Identification of human MutY homolog (hMYH) as a repair enzyme for 2-hydroxyadenine in DNA and detection of multiple forms of hMYH located in nuclei and mitochondria." Nucleic Acids Res **28**(6): 1355-1364.

Okey, A., P. A. Harper, D. M. Grant and R. P. Hill (2005). Chemical and Radiation Carcinogenesis. The Basic Science of Oncology. I. F. Tannock, R. P. Hill, R. Bristow and L. Harrington. Toronto, McGraw-Hill Medical Publishing**:** 25-48.

Ostenfeld, M. S., J. B. Bramsen, P. Lamy, S. B. Villadsen, N. Fristrup, K. D. Sorensen*, et al.* (2010). "miR-145 induces caspase-dependent and -independent cell death in urothelial cancer cell lines with targeting of an expression signature present in Ta bladder tumors." Oncogene **29**(7): 1073-1084.

Ott, O. J., C. Rodel, C. Weiss, M. Wittlinger, F. St Krause, J. Dunst*, et al.* (2009). "Radiochemotherapy for bladder cancer." Clin Oncol **21**(7): 557-565.

Out, A. A., I. J. van Minderhout, J. J. Goeman, Y. Ariyurek, S. Ossowski, K. Schneeberger*, et al.* (2009). "Deep sequencing to reveal new variants in pooled DNA samples." Hum Mutat **30**(12): 1703-1712.

Ozcelik, H., X. Shi, M. C. Chang, E. Tram, M. Vlasschaert, N. Di Nicola*, et al.* (2012). "Long-Range PCR and Next-Generation Sequencing of BRCA1 and BRCA2 in Breast Cancer." J Mol Diagn **[Epub ahead of print]**.

Pareek, C. S., R. Smoczynski and A. Tretyn (2011). "Sequencing technologies and genome sequencing." J Appl Genet **52**(4): 413-435.
Park, D. J., F. Lesueur, T. Nguyen-Dumont, M. Pertesi, F. Odefrey, F. Hammet*, et al.* (2012). "Rare mutations in XRCC2 increase the risk of breast cancer." Am J Hum Genet **90**(4): 734-739.
Parliament, M. B. and D. Murray (2010). "Single nucleotide polymorphisms of DNA repair genes as predictors of radioresponse." Semin Radiat Oncol **20**(4): 232-240.
Pasquinelli, A. E. (2012). "MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship." Nat Rev Genet **13**(4): 271-282.
Pervaiz, M. A., A. Eppolito and K. Schmidt (2010). "Papillary thyroid cancer in a patient with MUTYH-associated polyposis (MAP)." Fam Cancer **9**(4): 595-597.
Petermann, E., C. Keil and S. L. Oei (2006). "Roles of DNA ligase III and XRCC1 in regulating the switch between short patch and long patch BER." DNA Repair (Amst) **5**(5): 544-555.
Pharoah, P. D., R. T. Palmieri, S. J. Ramus, S. A. Gayther, I. L. Andrulis, H. Anton-Culver*, et al.* (2011). "The role of KRAS rs61764370 in invasive epithelial ovarian cancer: implications for clinical testing." Clin Cancer Res **17**(11): 3742-3750.
Plna, K. and K. Hemminki (2001). "Familial bladder cancer in the National Swedish Family Cancer Database." J Urol **166**(6): 2129-2133.
Plotz, G., M. Casper, J. Raedle, I. Hinrichsen, V. Heckel, A. Brieger*, et al.* (2012). "MUTYH gene expression and alternative splicing in controls and polyposis patients." Human mutation **33**(7): 1067-1074.
Pos, F. and P. Remeijer (2010). "Adaptive management of bladder cancer radiotherapy." Semin Radiat Oncol **20**(2): 116-120.
Price, A. L., G. V. Kryukov, P. I. de Bakker, S. M. Purcell, J. Staples, L. J. Wei and S. R. Sunyaev (2010). "Pooled association tests for rare variants in exon-resequencing studies." Am J Hum Genet **86**(6): 832-838.
Pritchard, C. C., C. Smith, S. J. Salipante, M. K. Lee, A. M. Thornton, A. S. Nord*, et al.* (2012). "ColoSeq Provides Comprehensive Lynch and Polyposis Syndrome Mutational Analysis Using Massively Parallel Sequencing." J Mol Diagn **14**(4): 357-366.
Proctor, I., K. Stoeber and G. H. Williams (2010). "Biomarkers in bladder cancer." Histopathology **57**(1): 1-13.
Pugh, T. J., M. Keyes, L. Barclay, A. Delaney, M. Krzywinski, D. Thomas*, et al.* (2009). "Sequence variant discovery in DNA repair genes from radiosensitive and radiotolerant prostate brachytherapy patients." Clin Cancer Res **15**(15): 5008-5016.
Pylkas, K., M. Vuorela, M. Otsukka, A. Kallioniemi, A. Jukkola-Vuorinen and R. Winqvist (2012). "Rare Copy Number Variants Observed in Hereditary Breast Cancer Cases Disrupt Genes in Estrogen Signaling and TP53 Tumor Suppression Network." PLoS Genet **8**(6): e1002734.
Qiao, B., A. H. Ansari, G. B. Scott, S. C. Sak, P. A. Chambers, F. Elliott*, et al.* (2011). "In vitro functional effects of XPC gene rare variants from bladder cancer patients." Carcinogenesis **32**(4): 516-521.
Qiao, B., A. H. Ansari, G. B. Scott, S. C. Sak, P. A. Chambers, F. Elliott*, et al.* (2011). "In vitro functional effects of XPC gene rare variants from bladder cancer patients." Carcinogenesis **32**(4): 516-521.
Rafnar, T., S. H. Vermeulen, P. Sulem, G. Thorleifsson, K. K. Aben, J. A. Witjes*, et al.* (2011). "European genome-wide association study identifies SLC14A1 as a new urinary bladder cancer susceptibility gene." Hum Mol Genet **20**(21): 4268-4281.

Ramagopalan, S. V., D. A. Dyment, M. Z. Cader, K. M. Morrison, G. Disanto, J. M. Morahan*, et al.* (2011). "Rare variants in the CYP27B1 gene are associated with multiple sclerosis." Ann Neurol **70**(6): 881-886.

Raychaudhuri, S., O. Iartchouk, K. Chin, P. L. Tan, A. K. Tai, S. Ripke*, et al.* (2011). "A rare penetrant mutation in CFH confers high risk of age-related macular degeneration." Nat Genet **43**(12): 1232-1236.

Reich, D. E. and E. S. Lander (2001). "On the allelic spectrum of human disease." Trends Genet **17**(9): 502-510.

Reiner, A., B. Neumeister, J. Spona, G. Reiner, M. Schemper and R. Jakesz (1990). "Immunocytochemical localization of estrogen and progesterone receptor and prognosis in human primary breast cancer." Cancer Res **50**(21): 7057-7061.

Rennert, G., F. Lejbkowicz, I. Cohen, M. Pinchev, H. S. Rennert and O. Barnett-Griness (2012). "MutYH mutation carriers have increased breast cancer risk." Cancer **118**(8): 1989-1993.

Rhee, J. G., D. Li, M. Suntharalingam, C. Guo, B. W. O'Malley, Jr. and J. P. Carney (2007). "Radiosensitization of head/neck squamous cell carcinoma by adenovirus-mediated expression of the Nbs1 protein." Int J Radiat Oncol Biol Phys **67**(1): 273-278.

Ribic, C. M., D. J. Sargent, M. J. Moore, S. N. Thibodeau, A. J. French, R. M. Goldberg*, et al.* (2003). "Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer." N Engl J Med **349**(3): 247-257.

Ricceri, F., G. Matullo and P. Vineis (2011). "Is there evidence of involvement of DNA repair polymorphisms in human cancer?" Mutat Res.

Ricceri, F., P. Porcedda, A. Allione, V. Turinetto, S. Polidoro, S. Guarrera*, et al.* (2011). "Involvement of MRE11A and XPA gene polymorphisms in the modulation of DNA double-strand break repair activity: a genotype-phenotype correlation study." DNA Repair (Amst) **10**(10): 1044-1050.

Riester, M., J. M. Taylor, A. Feifer, T. Koppie, J. E. Rosenberg, R. J. Downey*, et al.* (2012). "Combination of a novel gene expression signature with a clinical nomogram improves the prediction of survival in high-risk bladder cancer." Clin Cancer Res **18**(5): 1323-1333.

Rivas, M. A., M. Beaudoin, A. Gardet, C. Stevens, Y. Sharma, C. K. Zhang*, et al.* (2011). "Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease." Nat Genet **43**(11): 1066-1073.

Rodel, C., C. Weiss and R. Sauer (2005). "Organ preservation by combined modality treatment in bladder cancer: the European perspective." Semin Radiat Oncol **15**(1): 28-35.

Rossetti, S., K. Hopp, R. A. Sikkink, J. L. Sundsbak, Y. K. Lee, V. Kubly*, et al.* (2012). "Identification of gene mutations in autosomal dominant polycystic kidney disease through targeted resequencing." J Am Soc Nephrol **23**(5): 915-933.

Rothman, N., M. Garcia-Closas, N. Chatterjee, N. Malats, X. Wu, J. D. Figueroa*, et al.* (2010). "A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci." Nat Genet **42**(11): 978-984.

Rusinov, V., V. Baev, I. N. Minkov and M. Tabler (2005). "MicroInspector: a web tool for detection of miRNA binding sites in an RNA sequence." Nucleic Acids Res **33**(Web Server issue): W696-700.

Sak, A. and M. Stuschke (2010). "Use of gammaH2AX and other biomarkers of double-strand breaks during radiotherapy." Semin Radiat Oncol **20**(4): 223-231.

Sak, S. C., J. H. Barrett, A. B. Paul, D. T. Bishop and A. E. Kiltie (2005). "The polyAT, intronic IVS11-6 and Lys939Gln XPC polymorphisms are not associated with transitional cell carcinoma of the bladder." Br J Cancer **92**(12): 2262-2265.

Sak, S. C., J. H. Barrett, A. B. Paul, D. T. Bishop and A. E. Kiltie (2006). "Comprehensive analysis of 22 XPC polymorphisms and bladder cancer risk." Cancer Epidemiol Biomarkers Prev **15**(12): 2537-2541.

Sak, S. C., P. Harnden, C. F. Johnston, A. B. Paul and A. E. Kiltie (2005). "APE1 and XRCC1 protein expression levels predict cancer-specific survival following radical radiotherapy in bladder cancer." Clin Cancer Res **11**(17): 6205-6211.

Sakano, S., T. Wada, H. Matsumoto, S. Sugiyama, R. Inoue, S. Eguchi*, et al.* (2006). "Single nucleotide polymorphisms in DNA repair genes might be prognostic factors in muscle-invasive bladder cancer patients treated with chemoradiotherapy." Br J Cancer **95**(5): 561-570.

Sarasin, A. and A. Kauffmann (2008). "Overexpression of DNA repair genes is associated with metastasis: a new hypothesis." Mutat Res **659**(1-2): 49-55.

Sartori, A. A., C. Lukas, J. Coates, M. Mistrik, S. Fu, J. Bartek*, et al.* (2007). "Human CtIP promotes DNA end resection." Nature **450**(7169): 509-514.

Schairer, C., P. Hartge, R. N. Hoover and D. T. Silverman (1988). "Racial differences in bladder cancer risk: a case-control study." Am J Epidemiol **128**(5): 1027-1037.

Schneider, A. C., L. C. Heukamp, S. Rogenhofer, G. Fechner, P. J. Bastian, A. von Ruecker*, et al.* (2011). "Global histone H4K20 trimethylation predicts cancer-specific survival in patients with muscle-invasive bladder cancer." BJU Int **108**(8 Pt 2): E290-296.

Schork, N. J., S. S. Murray, K. A. Frazer and E. J. Topol (2009). "Common vs. rare allele hypotheses for complex diseases." Curr Opin Genet Dev **19**(3): 212-219.

Scrimger, R. A., A. D. Murtha, M. B. Parliament, P. M. Venner, J. Hanson, G. Houle and M. Chetner (2001). "Muscle-invasive transitional cell carcinoma of the urinary bladder: a population-based study of patterns of care and prognostic factors." Int J Radiat Oncol Biol Phys **51**(1): 23-30.

Sengelov, L. and H. von der Maase (1999). "Radiotherapy in bladder cancer." Radiother Oncol **52**(1): 1-14.

Shariat, S. F., P. I. Karakiewicz, G. S. Palapattu, G. E. Amiel, Y. Lotan, C. G. Rogers*, et al.* (2006). "Nomograms provide improved accuracy for predicting survival after radical cystectomy." Clin Cancer Res **12**(22): 6663-6676.

Shariat, S. F., Y. Lotan, A. Vickers, P. I. Karakiewicz, B. J. Schmitz-Drager, P. J. Goebell and N. Malats (2010). "Statistical consideration for clinical biomarker research in bladder cancer." Urol Oncol **28**(4): 389-400.

Sheehan, A. M., D. K. McGregor, A. Patel, V. Shidham, C. Y. Fan and C. C. Chang (2005). "Expression of human 8-oxoguanine DNA glycosylase (hOGG1) in follicular lymphoma." Mod Pathol **18**(11): 1512-1518.

Shelley, M., J. Barber, T. Wilt and M. Mason (2008). "Surgery versus radiotherapy for muscle invasive bladder cancer (Review)." Cochrane Database Syst Rev(4).

Shen, M., I. Menashe, L. M. Morton, Y. Zhang, B. Armstrong, S. S. Wang*, et al.* (2010). "Polymorphisms in DNA repair genes and risk of non-Hodgkin lymphoma in a pooled analysis of three studies." Br J Haematol **151**(3): 239-244.

Shendure, J. and H. Ji (2008). "Next-generation DNA sequencing." Nat Biotechnol **26**(10): 1135-1145.

Shin, C., J. W. Nam, K. K. Farh, H. R. Chiang, A. Shkumatava and D. P. Bartel (2010). "Expanding the microRNA targeting code: functional sites with centered pairing." Mol Cell **38**(6): 789-802.

Shinmura, K., M. Goto, M. Suzuki, H. Tao, H. Yamada, H. Igarashi*, et al.* (2011). "Reduced expression of MUTYH with suppressive activity against mutations caused by 8-hydroxyguanine is a novel predictor of a poor prognosis in human gastric cancer." J Pathol **225**(3): 414-423.

Short, S. C., S. Giampieri, M. Worku, M. Alcaide-German, G. Sioftanos, S. Bourne*, et al.* (2011). "Rad51 inhibition is an effective means of targeting DNA repair in glioma models and CD133+ tumor-derived cells." Neuro Oncol **13**(5): 487-499.

Signore, M. and K. A. Reeder (2012). "Antibody validation by Western blotting." Methods Mol Biol **823**: 139-155.

Smith, A. E., A. M. Mohamedali, A. Kulasekararaj, Z. Lim, J. Gaken, N. C. Lea*, et al.* (2010). "Next-generation sequencing of the TET2 gene in 355 MDS and CMML patients reveals low-abundance mutant clones with early origins, but indicates no definite prognostic value." Blood **116**(19): 3923-3932.

Sobin, L. H., M. K. Gospodarowicz and C. Wittekind (2010). TNM Classification of Malignant Tumours, 7th Edition. New York, Wiley-Blackwell.

Soderlund, K., O. Stal, L. Skoog, L. E. Rutqvist, B. Nordenskjold and M. S. Askmalm (2007). "Intact Mre11/Rad50/Nbs1 complex predicts good response to radiotherapy in early breast cancer." Int J Radiat Oncol Biol Phys **68**(1): 50-58.

Spruck, C. H., 3rd, P. F. Ohneseit, M. Gonzalez-Zulueta, D. Esrig, N. Miyao, Y. C. Tsai*, et al.* (1994). "Two molecular pathways to transitional cell carcinoma of the bladder." Cancer Res **54**(3): 784-788.

Stacey, S. N., P. Sulem, A. Jonasdottir, G. Masson, J. Gudmundsson, D. F. Gudbjartsson*, et al.* (2011). "A germline variant in the TP53 polyadenylation signal confers cancer susceptibility." Nat Genet **43**(11): 1098-1103.

Stadler, W. M., S. P. Lerner, S. Groshen, J. P. Stein, S. R. Shi, D. Raghavan*, et al.* (2011). "Phase III study of molecularly targeted adjuvant therapy in locally advanced urothelial cancer of the bladder based on p53 status." J Clin Oncol **29**(25): 3443-3449.

Steel, G. (2002). "Chapter 8: DNA damage and cell killing." Basic Clinical Radiobiology 3rd Edition.

Stenzl, A., N. C. Cowan, M. De Santis, M. A. Kuczyk, A. S. Merseburger, M. J. Ribal*, et al.* (2011). "Treatment of muscle-invasive and metastatic bladder cancer: update of the EAU guidelines." Eur Urol **59**(6): 1009-1018.

Stern, M. C., J. Lin, J. D. Figueroa, K. T. Kelsey, A. E. Kiltie, J. M. Yuan*, et al.* (2009). "Polymorphisms in DNA repair genes, smoking, and bladder cancer risk: findings from the international consortium of bladder cancer." Cancer Res **69**(17): 6857-6864.

Stevens, A. and J. Lowe (1995). Tumours of the Lower Urinary Tract. Pathology, Mosby**:** 346-347.

Stewart, D. J. (2010). "Tumor and host factors that may limit efficacy of chemotherapy in non-small cell and small cell lung cancer." Crit Rev Oncol Hematol **75**(3): 173-234.

Sugasawa, K. (2008). "Xeroderma pigmentosum genes: functions inside and outside DNA repair." Carcinogenesis **29**(3): 455-465.

Sugasawa, K. (2010). "Regulation of damage recognition in mammalian global genomic nucleotide excision repair." Mutat Res **685**(1-2): 29-37.

Sun, J. M., M. J. Ahn, M. J. Park, H. Y. Lee, J. S. Ahn, S. Lee*, et al.* (2011). "Expression of excision repair cross-complementation group 1 as predictive marker for nasopharyngeal cancer treated with concurrent chemoradiotherapy." Int J Radiat Oncol Biol Phys **80**(3): 655-660.

Sun, Y., X. Jiang and B. D. Price (2010). "Tip60: connecting chromatin to DNA damage signaling." Cell Cycle **9**(5): 930-936.

Suzuki, S., N. Ono, C. Furusawa, B. W. Ying and T. Yomo (2011). "Comparison of sequence reads obtained from three next-generation sequencing platforms." PLoS One **6**(5): e19534.

Sylvester, R. J., A. P. van der Meijden, W. Oosterlinck, J. A. Witjes, C. Bouffioux, L. Denis*, et al.* (2006). "Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials." Eur Urol **49**(3): 466-465; discussion 475-467.

Taneja, N., M. Davis, J. S. Choy, M. A. Beckett, R. Singh, S. J. Kron and R. R. Weichselbaum (2004). "Histone H2AX phosphorylation as a predictor of radiosensitivity and target for radiotherapy." J Biol Chem **279**(3): 2273-2280.

Tang, W., Y. P. Fu, J. D. Figueroa, N. Malats, M. Garcia-Closas, N. Chatterjee*, et al.* (2012). "Mapping of the UGT1A locus identifies an uncommon coding variant that affects mRNA expression and protects from bladder cancer." Hum Mol Genet **21**(8): 1918-1930.

Taylor, C. F. (2009). "Mutation scanning using high-resolution melting." Biochem Soc Trans **37**(Pt 2): 433-437.

Tenesa, A., H. Campbell, R. Barnetson, M. Porteous, M. Dunlop and S. M. Farrington (2006). "Association of MUTYH and colorectal cancer." Br J Cancer **95**(2): 239-242.

Tennessen, J. A., A. W. Bigham, T. D. O'Connor, W. Fu, E. E. Kenny, S. Gravel*, et al.* (2012). "Evolution and functional impact of rare coding variation from deep sequencing of human exomes." Science **337**(6090): 64-69.

Teo, M. T., D. Landi, C. F. Taylor, F. Elliott, L. Vaslin, D. G. Cox*, et al.* (2012). "The role of microRNA-binding site polymorphisms in DNA repair genes as risk factors for bladder cancer and breast cancer and their impact on radiotherapy outcomes." Carcinogenesis **33**(3): 581-586.

Thariat, J., S. Aluwini, Q. Pan, M. Caullery, P. Y. Marcy, M. Housset and J. L. Lagrange (2012). "Image-guided radiation therapy for muscle-invasive bladder cancer." Nat Rev Urol **9**(1): 23-29.

The 1000 Genomes Project Consortium (2010). "A map of human genome variation from population-scale sequencing." Nature **467**(7319): 1061-1073.

The Wellcome Trust Case Control Consortium (2007). "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls." Nature **447**(7145): 661-678.

Theodoratou, E., H. Campbell, A. Tenesa, R. Houlston, E. Webb, S. Lubbe*, et al.* (2010). "A large-scale meta-analysis to refine colorectal cancer risk estimates associated with MUTYH variants." Br J Cancer **103**(12): 1875-1884.

Thomas, D. (2010). "Gene--environment-wide association studies: emerging approaches." Nat Rev Genet **11**(4): 259-272.

Tischkowitz, M., M. Capanu, N. Sabbaghian, L. Li, X. Liang, M. P. Vallee*, et al.* (2012). "Rare germline mutations in PALB2 and breast cancer risk: a population-based study." Human mutation **33**(4): 674-680.

Torgerson, D. G., D. Capurso, R. A. Mathias, P. E. Graves, R. D. Hernandez, T. H. Beaty*, et al.* (2012). "Resequencing candidate genes implicates rare variants in asthma susceptibility." Am J Hum Genet **90**(2): 273-281.

Toustrup, K., B. S. Sorensen, J. Alsner and J. Overgaard (2012). "Hypoxia gene expression signatures as prognostic and predictive markers in head and neck radiotherapy." Semin Radiat Oncol **22**(2): 119-127.

Tran, E., L. Souhami, S. Tanguay and R. Rajan (2009). "Bladder conservation treatment in the elderly population: results and prognostic factors of muscle-invasive bladder cancer." Am J Clin Oncol **32**(4): 333-337.

Tsai, M. H., W. H. Fang, S. W. Lin, S. J. Yen, S. J. Chou and Y. C. Yang (2009). "Mitochondrial genomic instability in colorectal cancer: no correlation to nuclear microsatellite instability and allelic deletion of hMSH2, hMLH1, and p53 genes, but prediction of better survival for Dukes' stage C disease." Ann Surg Oncol **16**(10): 2918-2925.

Tucker, T., M. Marra and J. M. Friedman (2009). "Massively parallel sequencing: the next big thing in genetic medicine." Am J Hum Genet **85**(2): 142-154.
Vaezi, A., X. Wang, S. Buch, W. Gooding, L. Wang, R. R. Seethala*, et al.* (2011). "XPF expression correlates with clinical outcome in squamous cell carcinoma of the head and neck." Clin Cancer Res **17**(16): 5513-5522.
Vale, C. and ABC Meta-analysis Group (2005). "Adjuvant chemotherapy in invasive bladder cancer: a systematic review and meta-analysis of individual patient data Advanced Bladder Cancer (ABC) Meta-analysis Collaboration." Eur Urol **48**(2): 189-199; discussion 199-201.
Vale, C. and ABC Meta-analysis Group (2005). "Neoadjuvant chemotherapy in invasive bladder cancer: update of a systematic review and meta-analysis of individual patient data advanced bladder cancer (ABC) meta-analysis collaboration." Eur Urol **48**(2): 202-205; discussion 205-206.
Vale, C. and Advanced Bladder Cancer Meta-analysis Collaboration (2004). "Neoadjuvant chemotherapy for invasive bladder cancer." Cochrane Database Syst Rev(1).
van Rhijn, B. W., L. Liu, A. N. Vis, P. J. Bostrom, T. C. Zuiverloon, N. E. Fleshner*, et al.* (2012). "Prognostic value of molecular markers, sub-stage and European Organisation for the Research and Treatment of Cancer risk scores in primary T1 bladder cancer." BJU Int.
Vogt, S., N. Jones, D. Christian, C. Engel, M. Nielsen, A. Kaufmann*, et al.* (2009). "Expanded extracolonic tumor spectrum in MUTYH-associated polyposis." Gastroenterology **137**(6): 1976-1985 e1971-1910.
von der Thusen, J. H., M. D. van de Wetering, A. M. Westermann, D. A. Heideman and E. Thunnissen (2011). "Bronchioloalveolar adenocarcinoma and pulmonary langerhans cell histiocytosis in a patient with MUTYH-associated polyposis." J Clin Oncol **29**(8): e188-190.
Wacholder, S., S. Chanock, M. Garcia-Closas, L. El ghormli and N. Rothman (2004). "Assessing the Probability That a Positive Report is False: An Approach for Molecular Epidemiology Studies." J Natl Cancer Inst **96**(6): 434-442.
Wang, M., W. Wu, B. Rosidi, L. Zhang, H. Wang and G. Iliakis (2006). "PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways." Nucleic Acids Res **34**(21): 6170-6182.
Wang, Q. E., C. Han, B. Zhang, K. Sabapathy and A. A. Wani (2012). "Nucleotide excision repair factor XPC enhances DNA damage-induced apoptosis by downregulating the antiapoptotic short isoform of caspase-2." Cancer Res **72**(3): 666-675.
Weedon, M. N. and T. M. Frayling (2007). "Insights on pathogenesis of type 2 diabetes from MODY genetics." Curr Diab Rep **7**(2): 131-138.
Wibom, C., S. Sjostrom, R. Henriksson, T. Brannstrom, H. Broholm, P. Ryden*, et al.* (2012). "DNA-repair gene variants are associated with glioblastoma survival." Acta Oncol **51**(3): 325-332.
Williams, R. S., J. S. Williams and J. A. Tainer (2007). "Mre11-Rad50-Nbs1 is a keystone complex connecting DNA repair machinery, double-strand break signaling, and the chromatin template." Biochem Cell Biol **85**(4): 509-520.
Win, A. K., S. P. Cleary, J. G. Dowty, J. A. Baron, J. P. Young, D. D. Buchanan*, et al.* (2011). "Cancer risks for monoallelic MUTYH mutation carriers with a family history of colorectal cancer." Int J Cancer **129**(9): 2256-2262.
Witte, J. S. (2012). "Rare genetic variants and treatment response: sample size and analysis issues." Stat Med **(Epub ahead of print)**.
Wong, T. S., X. B. Liu, B. Y. Wong, R. W. Ng, A. P. Yuen and W. I. Wei (2008). "Mature miR-184 as Potential Oncogenic microRNA of Squamous Cell Carcinoma of Tongue." Clin Cancer Res **14**(9): 2588-2592.

Wu, M., D. R. Soler, M. C. Abba, M. I. Nunez, R. Baer, C. Hatzis*, et al.* (2007). "CtIP silencing as a novel mechanism of tamoxifen resistance in breast cancer." Mol Cancer Res **5**(12): 1285-1295.

Wu, M. C., S. Lee, T. Cai, Y. Li, M. Boehnke and X. Lin (2011). "Rare-variant association testing for sequencing data with the sequence kernel association test." Am J Hum Genet **89**(1): 82-93.

Wu, X., J. Gu, H. B. Grossman, C. I. Amos, C. Etzel, M. Huang*, et al.* (2006). "Bladder cancer predisposition: a multigenic approach to DNA-repair and cell-cycle-control genes." Am J Hum Genet **78**(3): 464-479.

Wu, X., Y. Ye, L. A. Kiemeney, P. Sulem, T. Rafnar, G. Matullo*, et al.* (2009). "Genetic variation in the prostate stem cell antigen gene PSCA confers susceptibility to urinary bladder cancer." Nat Genet **41**(9): 991-995.

Xing, J., C. P. Dinney, S. Shete, M. Huang, M. A. Hildebrandt, Z. Chen and J. Gu (2012). "Comprehensive pathway-based interrogation of genetic variations in the nucleotide excision DNA repair pathway and risk of bladder cancer." Cancer **118**(1): 205-215.

Xu, J., S. Lv, Y. Qin, F. Shu, Y. Xu, J. Chen*, et al.* (2007). "TRB3 interacts with CtIP and is overexpressed in certain cancers." Biochim Biophys Acta **1770**(2): 273-278.

Yang, F. and D. C. Thomas (2011). "Two-stage design of sequencing studies for testing association with rare variants." Hum Hered **71**(4): 209-220.

Yang, H., C. P. Dinney, Y. Ye, Y. Zhu, H. B. Grossman and X. Wu (2008). "Evaluation of genetic variants in microRNA-related genes and risk of bladder cancer." Cancer Res **68**(7): 2530-2537.

Yang, J., Z. Xu, J. Li, R. Zhang, G. Zhang, H. Ji*, et al.* (2010). "XPC epigenetic silence coupled with p53 alteration has a significant impact on bladder cancer outcome." J Urol **184**(1): 336-343.

Yang, R., B. Schlehe, K. Hemminki, C. Sutter, P. Bugert, B. Wappenschmidt*, et al.* (2010). "A genetic variant in the pre-miR-27a oncogene is associated with a reduced familial breast cancer risk." Breast Cancer Res Treat **121**(3): 693-702.

Yang, X., D. Liu, H. Wu, H. Kang, H. Pang, D. Huang*, et al.* (2012). "Association of XPC polymorphisms with susceptibility and clinical outcome to chemotherapy in breast cancer patients." Cancer Sci **[E-publication ahead of print]**.

Yin, M., Z. Liao, Y. J. Huang, Z. Liu, X. Yuan, D. Gomez*, et al.* (2011). "Polymorphisms of homologous recombination genes and clinical outcomes of non-small cell lung cancer patients treated with definitive radiotherapy." PLoS One **6**(5): e20055.

Yokoyama, S., S. L. Woods, G. M. Boyle, L. G. Aoude, S. MacGregor, V. Zismann*, et al.* (2011). "A novel recurrent mutation in MITF predisposes to familial and sporadic melanoma." Nature **480**(7375): 99-103.

Yoon, H. H., P. J. Catalano, K. M. Murphy, T. C. Skaar, S. Philips, M. Powell*, et al.* (2011). "Genetic variation in DNA-repair pathways and response to radiochemotherapy in esophageal adenocarcinoma: a retrospective cohort study of the Eastern Cooperative Oncology Group." BMC Cancer **11**(1): 176.

Yu, H., H. Zhao, L. E. Wang, Y. Han, W. V. Chen, C. I. Amos*, et al.* (2011). "An analysis of single nucleotide polymorphisms of 125 DNA repair genes in the Texas genome-wide association study of lung cancer with a replication for the XRCC4 SNPs." DNA Repair (Amst) **10**(4): 398-407.

Yu, X. and R. Baer (2000). "Nuclear localization and cell cycle-specific expression of CtIP, a protein that associates with the BRCA1 tumor suppressor." J Biol Chem **275**(24): 18541-18549.

Yu, Z., Z. Li, N. Jolicoeur, L. Zhang, Y. Fortin, E. Wang*, et al.* (2007). "Aberrant allele frequencies of the SNPs located in microRNA target sites are potentially associated with human cancers." Nucleic Acids Res **35**(13): 4535-4541.

Yuan, J. and J. Chen (2009). "N terminus of CtIP is critical for homologous recombination-mediated double-strand break repair." J Biol Chem **284**(46): 31746-31752.

Yun, M. H. and K. Hiom (2009). "CtIP-BRCA1 modulates the choice of DNA double-strand-break repair pathway throughout the cell cycle." Nature **459**(7245): 460-463.

Zeng, Y., R. Yi and B. R. Cullen (2003). "MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms." Proc Natl Acad Sci U S A **100**(17): 9779-9784.

Zha, S., C. Boboila and F. W. Alt (2009). "Mre11: roles in DNA repair beyond homologous recombination." Nat Struct Mol Biol **16**(8): 798-800.

Zhang, L., J. Huang, N. Yang, J. Greshock, M. S. Megraw, A. Giannakakis*, et al.* (2006). "microRNAs exhibit high frequency genomic alterations in human cancer." Proc Natl Acad Sci U S A **103**(24): 9136-9141.

Zhang, Y., L. H. Rohde and H. Wu (2009). "Involvement of nucleotide excision and mismatch repair mechanisms in double strand break repair." Current Genomics **10**(4): 250-258.

Zheng, D., S. Haddadin, Y. Wang, L. Q. Gu, M. C. Perry, C. E. Freter and M. X. Wang (2011). "Plasma microRNAs as novel biomarkers for early detection of lung cancer." Int J Clin Exp Pathol **4**(6): 575-586.

Zhou, J., Y. Zhou, B. Yin, W. Hao, L. Zhao, W. Ju and C. Bai (2010). "5-Fluorouracil and oxaliplatin modify the expression profiles of microRNAs in human colon cancer cells in vitro." Oncol Rep **23**(1): 121-128.

Zhu, C., X. Li and J. Yu (2011). "Integrating Rare-Variant Testing, Function Prediction, and Gene Network in Composite Resequencing-Based Genome-Wide Association Studies (CR-GWAS)." G3 (Bethesda) **1**(3): 233-243.

11. APPENDICES

11.1. Appendix A

11.1.1. Reagents

Compound	Supplier/ Manufacturer
2-mercaptoethanol	Sigma-Aldrich, UK
Absolute ethanol	Sigma-Aldrich, UK
Ammonium persulphate	Sigma-Aldrich, UK
Bovine serum albumin	Sigma-Aldrich, UK
Bradford solution	Sigma-Aldrich, UK
Citric acid	Sigma-Aldrich, UK
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich, UK
DPX mountant	Sigma-Aldrich, UK
Ethylenediaminetetraacetic acid (EDTA)	Sigma-Aldrich, UK
Eosin	Sigma-Aldrich, UK
Ethidium bromide	Sigma-Aldrich, UK
Formalin	Sigma-Aldrich, UK
Hi-Di formamide	Life Technologies, UK
Hydrochloric acid, concentrated (38.5 – 38.0%)	Sigma-Aldrich, UK
Hydrogen peroxide	Sigma-Aldrich, UK
L-glutamine	Sigma-Aldrich, UK
Meyers Haematoxylin	Sigma-Aldrich, UK
Molecular grade agarose	Promega, USA
Paraformaldehyde	Sigma-Aldrich, UK
Phosphate buffered saline, tablets	Sigma-Aldrich, UK
Phosphatase inhibitor	Roche, Switzerland
Ponceau red	Sigma-Aldrich, UK
Proteinase inhibitor	Roche, Switzerland
Protogel	National Diagnostics, UK
RIPA buffer	Sigma-Aldrich, UK
Scott's tap water	Sigma-Aldrich, UK
Sodium chloride	Sigma-Aldrich, UK
Sodium dodecyl sulphate (SDS)	Sigma-Aldrich, UK
Sodium hydroxide	Sigma-Aldrich, UK
TEMED	Sigma-Aldrich, UK
Tris base	Sigma-Aldrich, UK

Compound	Supplier/ Manufacturer
Tris/Borate/EDTA (TBE) buffer, 10 x solution (89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.3)	Bio-Rad Laboratories, UK
Tris-Chloride buffer (Buffer EB) (10 mM Tris-Cl pH 8.5)	Qiagen, UK
Tris-EDTA buffer, molecular grade (10 mM Tris pH 8, 1 mM EDTA)	Sigma-Aldrich, UK
Tris-Glycine transfer buffer, 10 x solution (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3)	Bio-Rad Laboratories, UK
Triton-X-100	Sigma-Aldrich, UK
Trypsin 0.25% - EDTA solution, sterile	Sigma-Aldrich, UK
Tween 20	Sigma-Aldrich, UK
Xylene	Sigma-Aldrich, UK

11.1.2. Buffers

Citric acid antigen retrieval buffer, 1 x solution:

4.2 g Citric acid

Made up to 2 l with water

Adjusted to pH 6 with sodium hydroxide

Phosphate buffered saline, 1 x solution:

5 PBS tablets made up to 1 l with water

Tris-buffered saline (TBS) pH 7.5, 10 x solution:

12.11 g Tris-base

87.66 g Sodium chloride

Made up to 1 l with water

Adjusted to pH 7.5 with concentrated hydrochloric acid

Tris-EDTA + 0.05% Triton-X-100 antigen retrieval buffer, 1 x solution:

2.42 g Tris-base

0.74 g EDTA

1 ml Tween 20

1 ml Triton-X-100

Made up to 2 l with water

Adjusted to pH 8 with concentrated hydrochloric acid

Tris-Glycine-SDS running buffer, 5 x solution:

15.1 g Tris base

94 g Glycine

50 ml 10% SDS

Made up to 1 l with sterile water

11.1.3. Cell Culture Media

RPMI-1640 medium Catalogue No R0883 (Sigma-Aldrich, UK)

Minimum Essential Medium Eagle medium Catalogue No M2279 (Sigma-Aldrich, UK)

Media were supplemented with 10% fetal bovine serum (Sigma-Aldrich, UK) and 1% L-glutamine (Sigma-Aldrich, UK).

11.1.4. PCR Kits

PCR kit	Supplier/ Manufacturer	Contents
ABgene Extensor Long PCR Master Mix Cat No: AB-0792	ThermoFisher Scientific, UK	2 x Extensor Long PCR master mix, Buffer 1/2
Bio-X-act Cat No: BIO-25024	BioLine, UK	2 x Bio-X-Act Long mix 50 mM $MgCl_2$ solution
Fermentas Phusion Flash High-Fidelity PCR Master Mix Cat No: F-548	ThermoFisher Scientific, UK	2 x Phusion Flash PCR master mix
Fermentas Phusion High Fidelity Master Mix Cat No: F-530	ThermoFisher Scientific, UK	Phusion DNA Polymerase (100 U) 5 x Phusion HF Buffer 5 x Phusion GC Buffer 50 mM $MgCl_2$ solution
HotShot Diamond PCR Master Mix Cat No: HS002	Clent Life Science, UK	2 x HotShot Diamond PCR master mix
HotStarTaq Master Mix kit Cat No: 203443	Qiagen, UK	2 x HotStarTaq master mix Distilled water

PCR kit	Supplier/ Manufacturer	Contents
Invitrogen SequelPrep Long PCR Kit Cat No: A10498	Life Technologies, UK	Sequel Prep Long Polymerase 5 U / µl Sequel Prep 10 x Reaction Buffer Sequel Prep 10 x Enhancer A Sequel Prep 10 x Enhancer B DMSO
LCGreen Plus+ Dye Cat No: BCHM-ASY-0005	Idaho Technology, USA	LCGreen dye
Qiagen Muliplex PCR Master Mix Cat No: 206143	Qiagen, UK	2x Muliplex PCR Master Mix Q Solution Buffer Distilled water
Biotium Fast Probe Master Mix Cat No: 31005	Biotium, USA	2x Fast Probe Master Mix

11.1.4.1. Miscellaneous kits

Item Name	Supplier/ Manufacturer	Cat No
QIAquick PCR Purification Kit	Qiagen, UK	28106
Dako REA Detection System, Peroxidase DAB+, Rabbit/ Mouse	Dakocytomation, Denmark	K500111
MenaPath X-Cell Plus HRP-Polymer Detection Kit	Menarini Diagnostics, UK	MP-XCP-U100
Invitrogen Quant-IT BR PicoGreen Kit	Life Technologies, UK	P7589
BigDye® Terminator v1.1 Cycle Sequencing Kit	Applied Biosystems, UK	4337449
MinElute PCR Purification Kit	Qiagen, UK	28006
DNA 1000 Kit	Agilent Technologies, USA	5067-1504
End-It DNA End Repair kit	EpiCentre, USA	ER81050
DNA Polymerase I Large (Klenow) Fragment Exonuclease Minus	Promega, USA	M2181
LigaFast Rapid DNA Ligation System	Promega, USA	M8225

Item Name	Supplier/ Manufacturer	Cat No
AllPrep DNA/RNA FFPE Kit	Qiagen, UK	80234
High Capacity cDNA Reverse Transcription Kit	Applied Biosystems, UK	4374966
Taqman GAPDH Control Reagents (Human)	Applied Biosystems, UK	402869
Invitrogen ROX Reference Dye	Life Technologies, UK	12223-012
QIAquick Gel Extraction Kit	Qiagen, UK	28704

11.1.5. Primers & PCR conditions

DNA oligonucleotides were purchased from Eurogentec, Belgium dissolved in molecular grade water at 100 µM concentration and stored at 4°C.

11.1.5.1. Long PCR

Table 49: *MUTYH* long PCR primers and conditions for the pilot multiplexed NGS study.

Gene	Amplicon	Exons	Forward Primer	Reverse Primer	Amplicon Size (bp)	Long PCR Kit	Annealing Temp (°C)	Extension Time (min)	Total no of cycles
MUTYH	1	1 - 4	GTATGAGCCTGG GTAAATTACTTCC	TTCAGTAAATAAAA CATCAAAAGTTCC	3049	SequelPrep Enhancer B	65	3	35
	2	5 - 9	TAACAGTTAGGGA GCAGTGAAAATC	ATTAGAGCTGACC TGAGACGTAAAC	6583	SequelPrep Enhancer B	68	8	40

Table 50: *XPC* long PCR primers and conditions for the pilot multiplexed NGS study.

Gene	Amplicon	Exons	Forward Primer	Reverse Primer	Amplicon Size (bp)	Long PCR Kit	Annealing Temp (°C)	Extension Time (min)	Total no of cycles
XPC	1	1	TGAATAAATAATTG TTGCGCTCAC	GGAGGAAAGCATA GATTTAAAGAGG	1265	Extensor Long PCR Master Mix 2	54	4	35
	2	2 - 3	GAGGTTAGCTGAC ATTTAAGATCTGG	TACTCACTCACAC TGCCACCTAAG	3261	Extensor Long PCR Master Mix 2	56	4	35
	3.1	4 - 5.1	TCTAAAGACTGTC TGGGTTGTGTG	TCTAATCCTTTCC CCTACATAAAGC	1994	Bio-X-act	61	4	35
	3.2	5.2 - 6	AAGCAATTTGTATT TTGAAGCTTTG	TAGCACATCTGAC CAAAAACTACG	1316	Extensor Long PCR Master Mix 1	59	2	35
	4.1	7 - 8	TAGCACTAGACTG TTTCCAGAGTCC	GACAGTATCATGT CTGTCTCATTCC	3434	Extensor Long PCR Master Mix 1	65	4	35
	4.2	9	AGTGAATTGCCTA CTGAGAAATGAG	AGAGACAGGGTTT CATCGTGTTAG	1255	SequelPrep Enhancer A	65	2	35
	5	10	AATCATTTGACTG TTCCATGTGC	GTGTCCTGTGAAT AGTCTCTGTGG	1269	Extensor Long PCR Master Mix 1	62	2	35
	6.1	11 - 14	ACAGGCTCTACTG ATGGACAGTTAC	AGACAGAAGACTG AGGTGTCCTAAC	2111	Extensor Long PCR Master Mix 1	65	3	35
	6.2	15 - 16	AAAAGCCCTTCAT CTGACTTTATG	GTCTGTTTCTCAA CAAGAGGTTCC	1987	Extensor Long PCR Master Mix 1	61	2	35

Table 51: *MRE11A* long PCR primers and conditions.

Gene	Amplicon	Exons	Forward Primer	Reverse Primer	Amplicon Size (bp)	Long PCR Kit	Annealing Temp (°C)	Extension Time (min)	Total no of cycles
MRE11A	1	1 - 3	CAGACGTTCTAATT CAAAAGTTCTAC	AGCTTTGACAGTCAC CTTACATTTC	3960	SequelPrep Enhancer A	55	4	35
	2	4	TAACAGTTTGCCTC TTCCATTAGAG	TGTCAACAAGAGTCC ATAAACAAAG	1250	Extensor Long PCR Master Mix 1	62	2	35
	3	5 - 7	TCAAACTAATTGAA ATGAATTGTCG	TCTACCTTTTTGTTG AATTGAGGAG	3903	SequelPrep Enhancer B	55	4	40
	4	8 - 9	CACTTGTCTGTGTC CAATAAGTTTG	TCCAGTATTCCTCTC TCTCAATGAC	1701	SequelPrep Enhancer B	59	2	35
	5	10	AGGAAAGCCTTATT GAAACATGAG	AGACTGTGAGTGGT GCTATAAGAGG	2233	SequelPrep Enhancer A	55	3	35
	6	11	CTTTCTCTTGAGGC ACATCTCTAAG	GAGGCCTCAGAAAT AACAATGC	2068	Extensor Long PCR Master Mix 1	68	3	35
	7	12 - 13	TATGAACGAGAAAT ACTGACTGTGG	AGAGAAGATACTATC CATGGGGAAC	2067	Extensor Long PCR Master Mix 1	59	2	35
	8	14	ACGATAATATATGG CACATTGAAGG	GGTTGCTAACTTGTA GATTCAAACC	1394	Extensor Long PCR Master Mix 1	62	2	40
	9	15 - 16	ATAGCTCACTGCAG CCTGTATCTC	CTCTCTCTGTTGCTA GGGTAGAGTC	2448	SequelPrep Enhancer B	68	3	35
	10	17 - 18	TCTCAATAAGCTGG GAAATAATGAG	GGAGAGAACACAGT CTTCCATTAAG	2223	Extensor Long PCR Master Mix 1	59	3	35
	11	19	TTTTCAGAAGTGGA AGATGGTAATG	AGGCACATTACAAAG AAGAAGACAC	1369	SequelPrep Enhancer B	59	2	35
	12	20	AAAATTACAGCTGT TTCTGAGTTGG	TTGAGATTGAGTTTC ACTCTTGTTG	3620	SequelPrep Enhancer B	62	4	40

Table 52: Long PCR primers and conditions for NGS resequencing of *XPC*.

Gene	Amplicon	Exons	Forward Primer	Reverse Primer	Amplicon Size (bp)	Long PCR Kit	Annealing Temp (°C)	Extension Time (min)	Total no of cycles
XPC	1	1 - 3	TAACCTAACTCAAC TGGTCCCTATG	GTAAACAGCCTTC TTTGGTAACTTG	9800	Phusion Flash	68	5	40[1]
	2	4 - 7	GCTACTGTATACAA GCAACGTAGGG	TGTTTTATTCAGAC AGAGCCTTACC	4649	Phusion Flash	62	2	35
	3	8 - 10	ACCCATAACCCAGT ATATGGATGC	TGTGGGGTGTGTT TTATACTGC	4394	Phusion Flash	62	2	35
	4	11 - 16	CTTTACCCCATCCT GATAGTCTGC	GTCTGTTTCTCAA CAAGAGGTTCC	8079	Phusion Flash	68	4	40[1]

[1] Phase 1 Denaturation and Annealing thermocycling steps were extended to 10 seconds each.

11.1.5.2. Conventional DNA sequencing primers

Table 53: Conventional sequencing primers designed for specific *MRE11A* and *XPC* variants.

Gene	NCBI hg19 position/ dbSNP rs number	Variant base change	Forward Primer	Reverse Primer	Amplicon Size (bp)	Annealing Temp (°C)
XPC	rs111768011	C>G	GAGGCCTGGTGGAGGTTG	AACTGCCTTCTCCGAGTTCA	214	57
	rs59774494	A>G	TTGGCCTCCTTTCTCTTCCT	TAGTATCCGCGGGGTATCTG	188	57
	Chr 3: 14214734	G>A	CCTGGCAGGAGGTTAGCTG	AGCAACCAGCTCTCAGCTTC	165	57
	rs3731078	A>G	TGTCCAGGCACAGCTATGTT	TGGAATTTTTAAAACCCCTCCT	209	57
	rs3731123	C>G	GCATGTTCCTGTTCTTCTACGA	CAGAGCCCGGAGAATCAGTA	228	57
	Chr 3: 14201378	T>A				
	rs3731137	C>T	AGAGAAAAGTCGCAGGGTGA	ATCTGTGGCTTGTGGCAACT	228	57
	rs77167750	A>G	TTCTTCCTAGGCAAGGGTTG	GTTTTTCCCCTGCCATTCTT	153	59
	Chr 3: 14192475	T>C	CTTCCCAGCCTGAATCCAT	AGGACAGTGGAGGCAAGAAA	300	59
	rs79619898	T>G				
	rs115373819	C>T	TGTCCCTGAAGTGGTGACAG	CGTGGCCTCCATCACATATT	155	57
	rs181225959	G>A	GAGGTTGAGGCGAGTGCAT	CCTCTGCCTCCTGGGTTT	178	59
	rs7638104	C>T	ATGATGGGTGAGGCTGGATA	GATGTCAACCCTCGAACCTG	230	63
	Chr 3: 14189586	C>T				
	Chr 3: 14189426	T>C	AGATGTGGCCCACTGTCTTC	GGCAGGAGCCTGCTGTATT	356	63
	Chr 3: 14186649	A>G	GCTAAAATCAGGAAAGAAAAGGA	AGCCTGCGCAAGACAGTAAG	242	57
MRE11A	rs11020802	G>T	CACGGGCTCAGGAGTTCTTA	CTCGCGACACTTCATGGATA	257	57
	rs1805363	G>A	GGACTTGAAGCATCTACGTT	ATGCGATTCCTAAATTACCC	224	57
	Chr 11: 94226459	C>T	CCCAAACCTGGCTATTTTCA	ATTTTCCTTTCCCGTGCTCT	314	57
	Chr 11: 94226398	A>T				

MRE11A	rs496797	G>A	CTGCAGATGCACTGTAAGCA	TCAAATTACTGCAAGACTCCAATC	216	57
	Chr 11: 94225305	deletion A	CATGCCTGTAATCCCAGCTA	TCCAAGAAGGTAGGTCAGATACTT	337	53
	Chr 11: 94225127	G>A				
	rs11404578	insertion T	TTCAGATTAATTTTGAGGCTTAAATG	TGAGAAGGCTAAACCCAGAGA	316	55
	rs13447588	T>C				
	rs13447590	G>C				
	rs61893736	C>T	CAAATTGAAACAATTCTGTGTTGG	GTAACCCAGGCCCAAAAGTT	220	57
	rs10485020	T>A	TTGAACTAGTGGTTAGGAAAACTGT	GCAGAAGCTTTGACAGTCACC	597	57
	Chr 11: 94223936	C>T				
	Chr 11: 94223855	T>C				
	Chr 11: 94218991	A>G	TGATCAGTCAGTCAACTTTGGTTT	CACTGCAACCTCTGCTTCTG	250	57
	rs3218740	C>T	CAGGTGATACGATATTCATGCAG	CGCAATCTTTGTGCTTCCTT	178	57
	Chr 11: 94211508	A>G	CAAAACTGCACTTTGATATCTTGG	TATTTCACTTCCTGCTCTTTCACTT	174	57
	Chr 11: 94210491	T>C	GCAGTGTTGTCCACCCTTTT	TGCAGAAATAAATATTAAAAGGGATG	244	57
	Chr 11: 94210210	C>A	GAATTTTACCATAGAAAACTCACATTT	GGGTCTTGCTATGTTGTCCAG	498	57
	rs13447619	G>A				
	Chr 11: 94209866	G>T				
	rs13447623	A>G	TTTCTCTGAAAAAGTTTGAGC	AATGGGATAAAGGAATGGAT	278	57
	rs115244417	C>G	AAAGGGAGGAAGATGAATATG	CCTTACAGGCTTCATGAGAA	275	55
	Chr 11: 94200953	T>C	TTTTGAGATTGAAGAAATGCTTG	CCGATGGTGATTGCTCTTCT	190	55
	Chr 11: 94200365	C>T	CCAGTGCCCTGATTTCACTT	TCTTTTCCACTGGTTATGGTGA	333	57
	Chr 11: 94199802	G>A	TTCATGAAAATTTGAGAAGTACTGG	TGGACAAAGGTTCCAGTGAA	572	57
	Chr 11: 94199571	A>G				
	Chr 11: 94199359	T>C				
	rs104895017	A>G	TCAGATGGTTTGCTTGGTAGG	CGCTAGGAAACAACAATTTGC	270	57

MRE11A	rs640627	G>A	TCAGATGGTTTGCTTGGTAGG	CGCTAGGAAACAACAATTTGC	270	57
	Chr 11: 94197302	A>G	GAACCTTTCAGTGTTCTTCG	ATTCCCACTGTCAATTTGTT	242	57
	rs641936	T>C				
	Chr 11: 94197239	T>C				
	rs13447654	A>G	CCCGGTGTGTGATGTTCC	TGCCCGTCAGAGAAATACAA	560	63
	Chr 11: 94196707	A>T				
	Chr 11: 94196702	A>C				
	Chr 11: 94196674	T>C				
	Chr 11: 94196475	T>C				
	Chr 11: 94195930	G>A	TTGTTGTCATTGCTTTTGGTG	AAGCAATGGGGAAAGGATTC	223	57
	Chr 11: 94193845	insertion CTC	GCTGCACAGCAGGGAAAA	AAGTTTCACTTAAGAATGTCCCTGA	145	57
	Chr 11: 94193208	A>C	GTTTTGGCATGGCTGTCTTC	TGGCAAAGAGGAACAAAGGT	278	57
	Chr 11: 94193178	C>T				
	rs61749249	C>A	ATGCCATTGAGGAATTAGTG	ACCAACCATATGCAAGACTC	297	57
	Chr 11: 94189265	G>T	AAGATGATGAAGTCCGTGAGG	AGAGACAGGGTCTCACTCTGTTG	362	57
	Chr 11: 94189161	G>A				
	rs13447695	A>C	CATCGAGAGGAGGGTCTCAA	CGTGACAAACACAGTGAAGC	280	57
	rs13447696	C>G				
	Chr 11: 94179824	G>A	AAAGTCAGTTTAAGAAAGAAAACTGC	TGGTGAAACCCCACCTCTAC	515	57
	Chr 11: 94179497	A>G				
	Chr 11: 94170115	A>G	GCTGGAAGGAGCTTTACCAA	TCTATAACCATTGAACTCACCTTCA	180	57
	rs61893706	G>A	TGTGTTCCCTTAGGTGCTCA	ATCACCTGGCAAGGAAACAA	235	57
	rs113009211	G>C	TTGATGCAGCATTAGTACAAGGA	TGGTGGTGCACGTCTGTAAT	226	57
	rs104895013	G>T	CCCCGTTTTACAGATGAGGA	CATCAAGTGTGCCTTTCTGG	205	61
	Chr 11: 94162799	T>C	CAGGGAGTACATGTGCAGGTT	GGAGGCCACTATCCTAAGCA	238	65

MRE11A	Chr 11: 94153102	G>A	TTTCCTGAGCATAACTCCAT	AACCCAGAACCTCTAGGAAA	279	57
	rs13447749	G>A	AAAACAATTTATGCGAACCT	CAATTTTTAACCCGTTTCCT	381	55
	Chr 11: 94152721	A>C				
	Chr 11: 94152660	C>T				
	rs11020777	T>C				
	Chr 11: 94151932	T>C	GCTGCCTCATATAGCACTTT	GCTCTTCCCTGACTTAACCT	249	55
	rs104895004	G>T	GACATAGCAGTTACAGAGGGTGA	CATTCCTATACCAACAGGTCTGAA	226	55
	rs2155209	A>G	TCCCTCTACAGGTCAAGAAA	GAATGGATTTTCCTCTTTATGA	386	55
	Chr 11: 94150560	G>A				
	rs13447762	T>G	CCAGCTTGATAAATTTGCCTAT	AGCGATTCTCCTGCCTCAG	287	55
	Chr 11: 94219326	deletion T	ATCACCTTGTGTGGCCTGA	TGTTTTCCTTGAGGGCTTATTT	315	55
	Chr 11: 94219035	A>T	TGATCAGTCAGTCAACTTTGGTTT	CACTGCAACCTCTGCTTCTG	250	57
	Chr 11: 94219035	deletion A				
	Chr 11: 94218775	deletion A	CTACTCGGGAAGCTGAGGTG	AAGTCAAAAGACAAGCCATTGA	332	57
	Chr 11: 94218692	A>T				
	Chr 11: 94218691	A>T				
	Chr 11: 94218690	A>T				
	Chr 11: 94218689	A>T				
	Chr 11: 94218688	T>G				
	Chr 11: 94218677	insertion GAGATTTTTT				
	Chr 11: 94212913	insertion A	AGTGGTCATATGCCAATGTAGA	TCATTTCCAAAATTCCAACAAA	254	55
	Chr 11: 94212912	insertion C				
	Chr 11: 94210968	deletion C	CCTCCACTTTCTAACCTTCAGTG	AAACACTGGCAAAATTGAAGG	241	55
	Chr 11: 94209742	deletion A	GCATGGTGGCTTATGCTTGT	TTGTTTTCTTACTTCGGCTTAAAA	271	55
	Chr 11: 94209722	insertion A				

MRE11A	Chr 11: 94189682	insertion T	CCCATAGGGTATTAGAATTTTTCC	GGGGCTACAATTAAGAAATGC	225	51
	Chr 11: 94189682	deletion T	CCCATAGGGTATTAGAATTTTTCC	GGGGCTACAATTAAGAAATGC	225	51
	Chr 11: 94170173	deletion T	GCTGGAAGGAGCTTTACCAA	TCTATAACCATTGAACTCACCTTCA	180	57
	Chr 11: 94163043	T>A	TGAATTATGTAGTTCATTGTGTTTGTT	TACCAAACCTCAGCATCACG	291	51
	Chr 11: 94162988	deletion T				
	Chr 11: 94162335	A>G	GGGTCAAATCGTAGCTCTGTT	CAAAGCAGCCAACAAACAAA	244	55
	Chr 11: 94151564	insertion T	TTAGGTGGGTCTGGGTGAGA	AAAAAGGTTTCCCTGTCACG	261	59
	Chr 11: 94151561	T>A				
	Chr 11: 94151548	insertion A				
	Chr 11: 94151548	deletion A				
	Chr 11: 94151547	T>A				

Table 54: Conventional sequencing primers for exons of *MUTYH* and *XPC*.

Gene	Exon	Forward Primer	Right Primer	Amplicon Size (bp)	Annealing Temp (°C)
MUTYH	6 - 7	TACCACCTTCAC CCTTGACC	GTTCCTACCCTC CTGCCATC	274	61
	8 - 9	CCAGGAGTCTTG GGTGTCTT	AGCAGAGCTCCT TTGCAGAC	400	61
	10 - 11	CAAAGGAGCTCT GCTTCACA	AGAACTGGAATG GGGCTTCT	394	61
	12	AAAGCCCTCTTG GCTTGAGT	CCGATTCCCTCC ATTCTCTC	297	60
	16	CCCCTCCCCCAA CTACAAG	CGAAACCAGTCT GAGCAACA	298	60
XPC	1	GTATGGGGTGGA GCTTCCTT	AGCTACGCAGGA GCTTGGAT	484	65
	2	CCTTCCACCCCT CACCTTAT	AATCTTCCATGGA CCCCAGT	341	65
	3	TGTTGATGGAGG AAGTGAGG	TGCAATTAGTGAT CTGACTCCA	359	62
	4	CATGCCTCACTT CCTCCTTC	CTCAGTCCTGGT CCCCTACA	328	65
	5.1	AGGAAATAGCTG GCTTGCAG	AAATAAAGCCTC GGTGAGCA	319	65
	5.2	TTTGGCAGCAAA AATTCCTC	ATGCCCACCACC TGATACAT	355	59
	6	GCTGGGGAAAGT AGGACAGA	TGTCGGTAACAC ACCTGGAA	345	59
	7	CTGGCTGTTTCC AGCTTTTC	GCTCGAAAGAAC CCACACTC	326	65
	8.1	GGGGACATCTTG ATGTATTGG	GCCTCATCACTC CCACTCTC	450	63
	8.2	GACAAGCAGGAG AAGGCAAC	ACCATCGCTGCA CATTTTCT	333	63
	8.3	GACCCAAGCTTG CCAGTG	GCTGGGCATATA TAAGGTGCTC	478	59
	9	GCTCCACCATCT GTTGTCAG	AAGGCTGCTAAT CCCATGC	327	64
	10	GTCCACGTTCAA GGCTGTTT	GGGAGGCTCATC ATCACTTC	354	63
	11	TTCTGAGGGTTC ACCAGGTA	GGGTGAGCAAGT CAGCATTT	320	60
	12	AGCATCAGAAGG GCTCAGG	AGCTTTCCATCC CCATCTCT	319	63
	13	AGATGTGGCCCA CTGTCTTC	GGCAGGAGCCTG CTGTATT	356	63
	14	CTTGGTGTGAAG GAGAGGCTA	ACTTGAGGATGG GGCAGAA	320	65
	15.1	ATTACTGACCCTC GCCTGTG	GTGCATGCTGCC TCAGTTT	397	63
	15.2	CACTACAGGCCC CACACCT	ATGGTCCTAGGT CCGCAAC	422	63
	15.3	AATGCGCTGATC GTTTCTT	AGAGCCAAATCT TTAGATAAATGC	420	61

11.1.6. Taqman SNP Genotyping Assays

Table 55: Applied Biosystems Taqman SNP Genotyping Primers and Probes for *MUTYH* rare variants genotyped.

Assay	Assay ID	Gene	Variant	Forward Primer	Reverse Primer	VIC Reporter 1	FAM Reporter 2
Custom	AHJ93TS	*MUTYH*	rs3219470 C>T	GGGTGATTGAATTT GAAAGAGGTTCCT	CCTCGGCCTCCCAAA GTG	CGGTGGCTCAC GCC	CGGTGGTTCA CGCC
Custom	AHLI1Z0		rs1140199 G>A	CTGGCCAACATGGT GAAACC	GCGCCCGCACCATAC	CCGTCTCTACT AAAAGTACAA	CGTCTCTACT AAAAATACAA
Custom	AHD1CYN		c.IVS1+2246 G>C	CCTTGACCACAGTC TTTAAAATTGCA	AGCATAAAAAGCAAAA ATTAGACCAAGAGG	CCCACCTTGGC ACTC	CCCACCTTCG CACTC
Custom	AHMRZ58		rs34612342 A>G	CCACAGGAGGTGAA TCAACTCT	CCTTCCGAGCTCCCT CCT	CCTGGGCTACT ATTCT	CTGGGCTGCT ATTCT
Custom	AHFAA4V		c.1099 C>T	CCTGTGGAGAGCCT GTGC	CCCTTCCCCAGTAGG CTTAC	TCTCTGGCGTG CCCG	CTCTGGCATG CCCG
Pre-designed	C_27860252 _10		rs36053993 G>A				
Custom	AHGI9A3		c.1718 C>T	CAGGGTTCCAAAAG GTCCCA	GCTGTGTGCATCAGT GGAGAT	CACGGAGAGGA CACC	CACGGAAAGG ACACC

Table 56: Applied Biosystems Taqman SNP Genotyping Primers and Probes for *XPC* rare variants genotyped.

Assay	Assay ID	Gene	Variant	Forward Primer	Reverse Primer	VIC Reporter 1	FAM Reporter 2
Custom	AHN0YCG	*XPC*	c.172 delAGGG	AGAAATCCAAGGCCAA GAGCAA	GGCGTCTCCCGCGA A	CTCTCACCCTC CTCCTC	CGCTCTCACCT CCTC
Custom	AHO9WIO		rs3731072 G>C	GGAGAAGAGGAAGAA CAGAGGGATA	GGACTAGAGGCTGC TGGAAAAG	TCACTGCCAGT ATCAG	TCACTGCCACT ATCAG
Custom	AHHR7HB		c.IVS14+2 T>G	CAGGCAGTCATTGAAA GGAAGGA	GCTTCTGCTGTCCC TCAGT	CATATGCGCTT ACCTCCT	ATGCGCTTCCC TCCT
Custom	AHI05NJ		c.1271 T>G	CGACGTCCGCATGGC	TCATCACTCCCACTC TCCTCTTTAT	TGGAGGCCACC CGC	TGGAGGCCCCC CGC
Custom	AHJ93TR		c.1624 T>G	AGTGGCTAGAGGTGTT CTGTGA	CACACCGTGCACAC AGTCTA	CACATACCCAC TTTTC	CACATACCCCC TTTTC
Custom	AHLI1ZZ		c.1655 T>G	TGTTCTGTGAGCAGGA GGAAAAG	GGCGTACTTGTAAC AGGTCAGA	CTGGCCCACCA CACC	CTGGCCCCCCA CACC
Custom	AHMRZ57		c.2240 T>G	ACTGGCAGACAGAGG AGTATCAG	CCTTCTGATGCTGC CCTTACC	TTCCCGTCCAC GGCCA	TTCCCGTCCCC GGCCA
Custom	AHN0YCF		c.2313 T>G	CTGCCCAGCATGATGC CTAT	CGGTGTAGATTGGG CAGGTT	CAGCTGGACAC AGCCA	CAGCTGGACCC AGCCA
Custom	AHO9WIN		c.2414 A>C	CTGTGTCCAGGCCATC ACT	CCTGTGTTTAGCCT CCATCGAA	CGGCTACTCCC ATCCCGT	CGGCTACTCCC CTCCCGT
Custom	AHQIUOV		c.*172 T>G	GGAGACGAGGCCAAG CT	AGCAAAAAGCTTTG AAGGCTTCAC	CTGCAGCACCT CCT	CTGCAGCCCCT CCT
Custom	AHRRSU3		c.*519 A>C	CCACTTACCTTCCCCT GAGTCA	ATGACCTGTACTTCT CTGCTCTCT	TGCCCTCACTG CCTC	TGCCCTCCCTG CCTC

Table 57: Applied Biosystems Taqman SNP Genotyping Primers and Probes for DNA repair 3'UTR SNPs genotyped.

Assay	Assay ID	Gene	Variant	Forward Primer	Reverse Primer	VIC Reporter 1	FAM Reporter 2
Custom	AHHR7HC	*ATM*	rs1137918 A>G	GGTGAAACCCTGTCT CTACTAAAAATACA	CCTCCCAGGTTCAA GAGATTCTC	AGTAGCTGGGA TTACAG	AGTAGCTGGGA CTACAG
Custom	AHI05NK	*ATM*	rs227091 T>C	CCTCCTGGGTTCAAG CAATTCTC	TGGCACACGCCTGT AGTC	AGCTACTCGGG AGGCT	AGCTACTCAGG AGGCT
Pre-designed	C_3178688_10	*BRCA1*	rs8176318 G>T				
Pre-designed	C_29356_10	*BRCA1*	rs12516 C>T				
Pre-designed	C_11881871_10	*LIG3*	rs4796030 C>A				
Pre-designed	C_2440679_10	*NBS1*	rs2735383 G>C				
Pre-designed	C_9632806_10	*PARP1*	rs8679 T>C				
Pre-designed	C_29187507_10	*RAD51*	rs7180135 A>G				

Table 58: Applied Biosystems Taqman SNP Genotyping Primers and Probes for *MRE11A* rs1805363.

Assay	Assay ID	Gene	Variant	Forward Primer	Reverse Primer	VIC Reporter 1	FAM Reporter 2
Pre-designed	C_11474841_10	*MRE11A*	rs1805363 G>A				

11.1.7. Illumina Sequencing Adaptors and Indexes

Table 59: Illumina sequencing adaptor oligonucleotide sequences and index sequences (in brackets).

Index	5'adaptor	3'adaptor
Unindexed	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	AGATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
1	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(CAACCT)	(GGTTGA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
2	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(AACCAT)	(TGGTTA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
3	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(AAGGAT)	(TCCTTA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
4	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(AATTAT)	(TAATTA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
5	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(ACACAT)	(TGTGTA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
6	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(GCATGT)	(CATGCA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
7	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(TCGATT)	(ATCGAA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
8	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(CGATCT)	(GATCGA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
9	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(AGCTAT)	(TAGCTA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
10	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(GGTTGT)	(CAACCA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
11	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(TGCATT)	(ATGCAA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
12	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(GCCGGT)	(CCGGCA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
13	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(GATCGT)	(CGATCA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
14	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(CCTTCT)	(GAAGGA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
15	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(TACGTT)	(ACGTAA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
16	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(TCAGTT)	(ACTGAA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
17	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(CAGTCT)	(GACTGA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
18	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(CGTACT)	(GTACGA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
19	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(TATATT)	(ATATAA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
20	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(AGTCAT)	(TGACTA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
21	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(GAAGGT)	(CCTTCA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
22	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(GACTGT)	(CAGTCA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG

Index	5'adaptor	3'adaptor
23	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(CATGCT)	(GCATGA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
24	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(TCCTTT)	(AAGGAA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
25	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(GCGCGT)	(CGCGCA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
26	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(CCGGCT)	(GCCGGA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
27	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(ACGTAT)	(TACGTA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
28	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(TCTCTT)	(AGAGAA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
29	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(ACTGAT)	(TCAGTA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
30	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(GGAAGT)	(CTTCCA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
31	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(TAATTT)	(AATTAA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
32	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(CACACT)	(GTGTGA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
33	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(GAGAGT)	(CTCTCA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
34	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(CCAACT)	(GTTGGA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
35	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(ACCAAT)	(TTGGTA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
36	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(GCTAGT)	(CTAGCA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
37	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(TGACTT)	(AGTCAA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
38	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(AGAGAT)	(TCTCTA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
39	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(GGCCGT)	(CGGCCA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
40	ACACTCTTTCCCTACACGACGCTCTTCCGATCT(CGCGCT)	(GCGCGA)GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG

11.2. Appendix B

11.2.1. Ethical Approval Letters

11.2.1.1. Leeds (East) Local Research Ethical Committee 02/192

11.2.1.2. Leeds (East) Local Research Ethical Committee studies 02/060

11.2.1.3. Leeds (East) Local Research Ethical Committee 04/Q1206/62

Academic Unit of Oncology and Haematological Oncology

CANCER RESEARCH UK

Division of Cancer Medicine Research
Cancer Research UK Clinical Centre
St James' University Hospital
Beckett Street
Leeds LS9 7TF
United Kingdom

F 0113 242 9886
T 0113 20 67227
e-mail: anne.kiltie@cancer.org.uk

AEK/sg

19th June 2002

Dr P Dear
Chairman
Local Research Ethics Committee
St James's University Hospital
LEEDS

Dear Dr Dear

Clinical Research Ethics Committee
Proposed project: Molecular epidemiology of DNA repair polymorphisms and bladder tumours.
AE Kiltie, A Paul, DT Bishop, M Knowles, P Whelan, A Joyce, I Eardley, S Lloyd, S Prescott, J Cartledge.
[also update to: Project number 99/156: A molecular analysis of transitional cell carcinoma]

We would be grateful if you could consider the enclosed Local Research Ethics Committee application.

We wish to perform an epidemiological study of DNA repair polymorphisms and bladder tumours. This will entail seeking consent from all incident cases of bladder tumours presenting to St James's University Hospital Urology Department for a blood sample, and patients will be requested to fill in a questionnaire regarding smoking, occupation and family history, which will be checked by a doctor or nurse. We already have a suitable group of 300 population based controls from a previous colorectal cancer study, and wish to recruit a further 150 hospital based controls in the first instance.

This study is closely related to **Project number 99/156: A molecular analysis of transitional cell carcinoma**, and the patient group is taken from the same population. As the consent form for project 99/156 is now outdated due to recent changes regarding consent for tissue collection and data storage, Professor Knowles has asked that the information sheet and consent form be combined for both projects and updated, and a copy is enclosed with our LREC submission.

We look forward to hearing from you soon.

Yours sincerely

Dr Anne Kiltie
Senior Lecturer/Honorary Consultant Clinical Oncologist

Mr Alan Paul
Consultant Urologist

Professor Tim Bishop
Professor of Genetic Epidemiology

Professor Margaret Knowles
Professor of Experimental Cancer Research

Encl.

The Leeds Teaching Hospitals NHS
NHS Trust
Local Research Ethics Committee

Room 8.7, Clinical Sciences Building, St James's University Hospital
Beckett Street, Leeds LS9 7TF
e-mail: comdhfo@stjames.leeds.ac.uk

Dr A Kiltie
Senior Lecturer/Honorary Consultant
Clinical Oncologist
Cancer Research UK Clinical Centre
St James's University Hospital

Enquiries to: Ann Prothero
(Ethics Secretary)

Direct Line/Ext: 0113 (20) 65652

10 July 2002

Dear Dr Kiltie

Project No 02/192: Molecular epidemiology of DNA repair polymorphisms and bladder tumours

Thank you for your letter of 19 June enclosing details of the above study and enclosing a combined information sheet and consent form for this study and a closely related study (Project No 99/156) which has been previously approved by the Ethics Committee. I am happy to take Chairman's action and approve this study on the basis that it involves an additional blood sample and the use of surplus tissue with consent and a questionnaire. I am happy with the updated information sheet and consent form.

The Committee would be very interested in receiving a report of your findings at some future date.

Yours sincerely

Ann Prothero

Dr P R F Dear
Chairman
Leeds Health Authority/St James's & Seacroft University Hospitals
Clinical Research (Ethics) Committee

The Leeds Teaching Hospitals NHS
NHS Trust
Local Research Ethics Committee

Room 8.7, Clinical Sciences Building, St James's University Hospital
Beckett Street, Leeds LS9 7TF
e-mail: comdhfo@stjames.leeds.ac.uk

Dr A Kiltie
Senior Lecturer/Honorary Consultant
Clinical Oncologist
Cancer Research UK Clinical Centre
St James's University Hospital

Enquiries to: Ann Prothero
(Ethics Secretary)

Direct Line/Ext: 0113 (20) 65652

29 October 2002

Dear Dr Kiltie

Project No 99/156: A molecular genetic analysis of transitional cell carcinoma

Project No 02/192: Molecular epidemiology of DNA repair polymorphisms and bladder tumours

Thank you for your letter of 14 October concerning amendments to the protocol for Project No 02/192 and enclosing revised information sheets and consent forms for the different patient groups. I am happy to take Chairman's action and approve these amendments on behalf of the Ethics Committee.

Yours sincerely

Ann Prothero

cc **Dr P R F Dear**
Chairman
Leeds Health Authority/St James's & Seacroft University Hospitals
Clinical Research (Ethics) Committee

16 [illegible]2003

The Leeds Teaching Hospitals NHS

NHS Trust

Local Research Ethics Committee

Room 5.1, Clinical Sciences Building, St James's University Hospital
Beckett Street, Leeds LS9 7TF
e-mail: a.d.prothero@leeds.ac.uk

Dr A Kiltie
Senior Lecturer/Honorary Consultant Clinical Oncologist
Cancer Research UK Clinical Centre
St James's University Hospital

Enquiries to: Ann Prothero
(Ethics Secretary)

Direct Line/Ext: 0113 (20) 65652

11 December 2003

Dear Dr A Kiltie

Project No 02/192: Molecular epidemiology of DNA repair polymorphisms and bladder tumours.

Thank you for your letter of 1 December concerning the recruitment of an additional 500 cases and 500 controls into the above study. I am happy to take Chairman's action and approve this amendment on behalf of the Ethics Committee

Yours sincerely

Ann Prothero

cc **Dr P R F Dear**
Chairman
Leeds (East) Research Ethics Committee

22 June 2005

Dr PRF Dear DM FRCP FRCPCH
Chairman
Clinical Research (Ethics) Committee
Room 5.2, Clinical Sciences Building
St James's University Hospital
Beckett Street
LEEDS LS9 7TF

Dear Dr Dear,

Project ref: 02/192 Molecular epidemiology of DNA repair polymorphisms and bladder tumours.

I respectfully submit a substantial amendment on the above.

In collaboration with Professor Margaret Knowles, we wish to extend the above project to study the TSC1 gene at the same time as ATM, H2AX, MRE11, NBS1 and RAD50 are genotyped later this year as approved in amendment 4.

In order to avoid repeated amendments we are also requesting that similar bladder cancer related genes can also be studied in the future but we would understand if you wished us to resubmit amendments related to these as required.

We hope that this meets with your approval.

Yours sincerely,

Dr Anne Kiltie
Senior Lecturer/Honorary Consultant Clinical Oncologist.

Leeds (East) Research Ethics Committee
Room 5.2, Clinical Sciences Building
St James's University Hospital
Beckett Street
Leeds
LS9 7TF

Telephone: 0113 2065652
Facsimile: 0113 2066772

21 July 2005

Dr Anne Kiltie
Senior Lecturer/Honorary Consultant Clinical Oncologist
Cancer Research UK Clinical Centre
St James's University Hospital

Dear Dr Kiltie

Study title: Molecular epidemiology of DNA repair polymorphisms and bladder tumours
REC reference: 02/192

Amendment number: *5*
Amendment date: 22 June 2005

The above amendment was reviewed at the meeting of the Sub-Committee of the Research Ethics Committee held on 19 July 2005.

Ethical opinion

The members of the Committee present gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

The members wondered whether, as this study has now been amended on a number of occasions, the study title was still relevant or whether it should be now be revised to make it less specific. I would be grateful for your comments on this.

Approved documents

The documents reviewed and approved at the meeting were:

Notice of substantial amendment (amendment 5) dated 22 June 2005

Membership of the Committee

The members of the Ethics Committee who were present at the meeting are listed on the attached sheet.

Management approval

All investigators and research collaborators in the NHS should notify the R&D Department for the relevant NHS care organisation of this amendment and check whether it affects local management approval of the research.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

REC reference: 02/292	**Please quote this number on all correspondence**

Yours sincerely,

Ann Prothero
Committee Administrator

E-mail: ann.prothero@leedsth.nhs.uk

List of members present at the meeting:

Ms Caroline Bedford, Pharmacist

Dr Carol Chu, Consultant in Clinical Genetics

NHS
National Research Ethics Service
Leeds (East) Research Ethics Committee
Room 5.2, Clinical Sciences Building
St James's University Hospital
Beckett Street
Leeds
LS9 7TF

Telephone: 0113 2065637

18 April 2007

Dr Anne Kiltie
Senior Lecturer/Honorary Consultant Clinical Oncologist
Cancer Research UK Clinical Centre
Leeds Institute of Molecular Medicine
St James's University Hospital

Dear Dr Kiltie

Study title: **Molecular epidemiology of DNA repair polymorphisms and bladder tumours.**
REC reference: **02/192**

Amendment number: 7
Amendment date: 05/04/2007

The above amendment was reviewed at the meeting of the Sub-Committee of the Research Ethics Committee held on

Ethical opinion

The members of the Committee present gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

Approved documents

The documents reviewed and approved at the meeting were:

Notice of substantial amendment form dated 05/04/2007.
Patient information sheet for controls version 2 dated 05/04/2007.
Consent form for controls version 2 dated 05/04/2007.
Protocol version 2 dated 05/04/2007.

Membership of the Committee

The members of the Ethics Committee who were present at the meeting are listed on the attached sheet.

Research governance approval

All investigators and research collaborators in the NHS should notify the R&D Department for the relevant NHS care organisation of this amendment and check whether it affects research governance approval of the research.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

REC reference number: 02/192	**Please quote this number on all correspondence**

Yours sincerely

Mrs Elaine Hazell
Assistant Co-ordinator

E-mail: Elaine.hazell@leedsth.nhs.uk

Copy to: R&D Department, LTHT

The Leeds Teaching Hospitals NHS
NHS Trust

Local Research Ethics Committee

Room 8.7, Clinical Sciences Building, St James's University Hospital
Beckett Street, Leeds LS9 7TF
e-mail: comdhfo@stjames.leeds.ac.uk

Dr A Kiltie
Senior Lecturer/
Honorary Consultant Clinical Oncologist
Cancer Research UK Clinical Centre
St James's University Hospital

Enquiries to: Ann Prothero
(Ethics Secretary)

Direct Line/Ext: 0113 (20) 65652

27 February 2002

Dear Dr Kiltie

Project No 02/060: Outcomes following radical treatment of bladder cancer: biological factors predictive of response to radiotherapy

Thank you for your letter of 13 February enclosing details of your proposed study. I am happy to take Chairman's action and approve your study on the basis that it involves the use of archived pathological material. I am satisfied with your justification for reviewing the notes of the group of patients involved without their consent or contacting the relatives of those who have died.

We would be very interested in receiving a report of your findings at some future date.

Yours sincerely

Ann Prothero

cc **Dr P R F Dear**
Chairman
Leeds Health Authority/St James's & Seacroft University Hospitals
Clinical Research (Ethics) Committee

NHS

National Research Ethics Service

Leeds (East) Research Ethics Committee
Room 5.2, Clinical Sciences Building
St James's University Hospital
Beckett Street
Leeds
LS9 7TF

Tel: 0113 2065652
Fax: 0113 2066772

24 July 2009

Dr Anne E Kiltie
Section of Experimental Oncology
Leeds Institute of Molecular Medicine
St James's University Hospital
Beckett Street
Leeds
LS9 7TF

Dear Dr Kiltie

Study title:	**Outcomes following radical treatment of bladder cancer: Biological factors predictive of response to radiotherapy**
REC reference:	**02/060**
Amendment number:	**2**
Amendment date:	**14 July 2009**

The above amendment was reviewed at the meeting of the Sub-Committee held on 22 July 2009.

Ethical opinion

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

Approved documents

The documents reviewed and approved at the meeting were:

Document	Version	Date
Notice of Substantial Amendment (non-CTIMPs)		14 July 2009
Covering Letter		12 July 2009

Membership of the Committee

The members of the Committee who took part in the review are listed on the attached sheet.

R&D approval

All investigators and research collaborators in the NHS should notify the R&D office for the relevant NHS care organisation of this amendment and check whether it affects R&D approval of the research.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

02/060:	**Please quote this number on all correspondence**

Yours sincerely

Miss Amy Beckitt
Committee Assistant Co-ordinator
E-mail: Amy.Beckitt@leedsth.nhs.uk

Enclosures: List of names and professions of members who took part in the review

Copy to: R&D office for Leeds Teaching Hospitals NHS Trust

15 JUL 2004

NHS

Dr Anne E Kiltie
Senior Lecturer/Honorary Consultant Clinical Oncologist
Cancer Research UK Clinical Centre in Leeds
Cancer Research Building
St James's University Hospital
Beckett St, Leeds
LS9 7TF

Leeds (East) Research Ethics Committee
Room 5.2, Clinical Sciences Building
St James's University Hospital
Beckett Street
Leeds
LS9 7TF

Enquiries to Ann Prothero
Ethics Secretary
Direct Line 0113 (20) 65652
Email: a.d.prothero@leeds.ac.uk

Dear Dr Kiltie,

14 July 2004

Full title of study: DNA damage response to double-strand breaks in bladder cancer: influence on response to radiotherapy and genomic instability
REC reference number: 04/Q1206/62

The Research Ethics Committee reviewed the above application at the meeting held on 06 July 2004.

Ethical opinion

The members of the Committee present gave a favourable ethical opinion to the above research on the basis described in the application form, protocol and supporting documentation.

Conditions of approval

The favourable opinion is given provided that you comply with the conditions set out in the attached document. You are advised to study the conditions carefully.

Approved documents

The documents reviewed and approved at the meeting were:

Application dated 18/06/2004
Investigator CV dated 17/06/2004
Protocol cohort A dated 17/06/2004
Protocol cohort B dated 17/06/2004
Protocol, fresh tissue cohort dated 17/06/2004
Covering Letter dated 16/06/2004
Statistician Comments dated 14/06/2004
Approval letter for project 99/156 dated 22/04/2002
Approval letter for projects 99/156 and 02/192 dated 25/06/2003
Approval letter for project 02/060 dated 27/02/2002

Management approval

You should arrange for all relevant host organisations to be notified that the research will be taking place, and provide a copy of the REC application, the protocol and this letter.

All researchers and research collaborators who will be participating in the research must obtain management approval from the relevant host organisation before commencing any research procedures. Where a substantive contract is not held with the host organisation, it may be necessary for an honorary contract to be issued before approval for the research can be given.

Membership of the Committee

The members of the Ethics Committee who were present at the meeting are listed on the attached sheet.

Notification of other bodies

We shall notify the research sponsor, The Leeds Teaching Hospitals N.H.S. Trust, that the study has a favourable ethical opinion.

Statement of compliance *(from 1 May 2004)*

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

REC reference number: 04/Q1206/62 **Please quote this number on all correspondence**

Yours sincerely,

Ann Prothero

Gr **Dr P R F Dear**
Chairman

Cc: Research and Development Department, LTHT

Enclosures *List of names and professions of members who were present at the meeting and those who submitted written comments*

Standard approval conditions SL-AC2

NHS

National Research Ethics Service
Leeds (East) Research Ethics Committee

Room 5.2, Clinical Sciences Building
St James's University Hospital
Beckett Street
Leeds
LS9 7TF

Tel: 0113 3926788
Fax: 0113 3926788

25 January 2010

Dr Anne E Kiltie
Cancer Research UK Clinical Centre in Leeds
Senior Lecturer/Honorary Consultant Clinical Oncologist
Cancer Research UK Clinical Centre in Leeds
Level 5, JIF Building
St James's University Hospital
Beckett St, Leeds
LS9 7TF

Dear Dr Kiltie

Study title:	**DNA damage response to double-strand breaks in bladder cancer: influence on response to radiotherapy and genomic instability**
REC reference:	**04/Q1206/62**
Amendment number:	**6**
Amendment date:	**30 December 2009**

The above amendment was reviewed at the meeting of the Sub-Committee held on 19 January 2010 by the Sub-Committee in correspondence.

Ethical opinion

The members of the Committee taking part in the review gave a favourable ethical opinion of the amendment on the basis described in the notice of amendment form and supporting documentation.

Approved documents

The documents reviewed and approved at the meeting were:

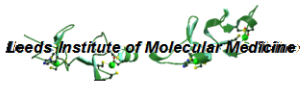
Document	Version	Date
Notice of Substantial Amendment (non-CTIMPs)		30 December 2009

Membership of the Committee

The members of the Committee who took part in the review are listed on the attached sheet.

11.2.2. Study Forms

11.2.2.1. LBCS Case Study Information Letter



UNIVERSITY OF LEEDS
Section of Oncology and Clinical Research
Cancer Research UK Clinical Centre
Leeds Institute of Molecular Medicine
Cancer Research Building
St James's University Hospital
Beckett Street
Leeds LS9 7TF

PATIENT INFORMATION SHEET (PRE-OPERATIVE PATIENTS):

A MOLECULAR GENETIC ANALYSIS OF UROTHELIAL CELL CARCINOMA (PROJECT 99/156)

EPIDEMIOLOGY OF DNA REPAIR AND BLADDER TUMOURS (PROJECT 02/192)

Your doctor has decided that an operation on your bladder or ureter (tube leading to the bladder from the kidney) is needed. During this operation, tissue may be removed for diagnosis and treatment. We would like to invite you to take part in two closely related RESEARCH studies.

You are under no obligation to take part in these studies and if you prefer not to do so this decision will be accepted without question, and your treatment will not be affected in any way. The findings will not be disclosed to you at a later date, will not affect the treatment you receive and will not affect you or your family.

The first study is looking at molecular changes found in tumour tissues and also in normal cells of patients who have bladder tumours or tumours in the ureter. This information may help us in the future to develop new methods of diagnosis and treatment and to decide which is the best method of treatment for individual patients.

This study involves providing the following samples for research purposes:

1. A blood sample (15 ml or 3 teaspoonfuls). This may be taken either at the time of your surgery or afterwards in the ward or clinic.
2. Urine samples. A sample will be collected at the time of surgery. You may be asked to provide a sample at the time of your visit to the pre-assessment clinic and on some occasions when you visit the clinic in the future you may be asked to provide a urine specimen.

Only a small amount of the tissue removed at operation is examined in detail by the pathology department to help with diagnosis and the remainder is disposed of. It is part of this surplus tumour tissue taken at this or future operations that we plan to use. Cells from the tissue would be grown and used to test the effects of new drugs and we would carry out a range of molecular tests on the tumour tissue, on the urine samples, and on the blood sample to identify common genetic alterations. Part of the research may also involve using the tissue stored in the pathology department after the pathologist has examined it.

The second study is looking at the differences in a person's DNA, compared to other people's, which may make them more likely to develop bladder tumours. This information might help us in the future to prevent bladder tumours developing in people at high risk of the disease. If you agree to take part in this study, we would ask you to consent to provide a further 5 ml of blood (at the same time as the other sample), and to fill in a simple questionnaire, regarding your smoking habits, jobs, and family history, which will be checked with you afterwards by a doctor or nurse.

These studies will use samples from many patients treated in the Department of Urology and they are intended to provide valuable information about molecular changes in a large group of patients. The second study will also involve other hospital patients and relatives and friends of a similar age who do not have bladder tumours.

Samples once donated are treated as a 'gift' to the hospital. They would be kept indefinitely within the hospital and laboratories (until they are used up), and may be used in this project and in future research. Scientific research is collaborative and this sometimes involves exchange of samples between laboratories worldwide. We work with other laboratories in this way and may share samples to increase the range of measurements that can be made and increase their usefulness. It is very important that you are sure you are happy for us to keep and use these samples, before you agree to donate them.

All information collected about you during the study will be kept strictly confidential. Your questionnaire answers will be used by the researchers to link these facts with any findings we may make, so that we can evaluate our results. We would also like your permission to send results of the questionnaire data and blood analysis abroad to other researchers working in this field, but no one from outside the hospital could identify you from the reports. The results of studies arising from this research programme may be published in the medical literature but your identity and personal information will remain strictly confidential at all times.

No special compensation arrangements have been made for this research programme. However, you rights to claim compensation from the hospital for injury or loss as a result of your treatment remain unaffected by your participation in this study.

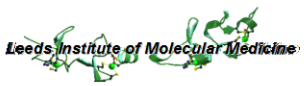
For further information, or if you have any questions, please contact:

Professor M Knowles
Division of Cancer Medicine
Cancer Research UK Clinical Centre in Leeds
Cancer Research Building
St James's University Hospital
Beckett Street
LEEDS LS9 7TF

Tel 0113 206 4913

Dr Anne Kiltie
Molecular Radiobiology Group
Section of Oncology
Leeds Institute of Molecular Medicine
Cancer Research Building
St James's University Hospital
Beckett Street
LEEDS LS9 7TF
Tel 0113 206 7275

11.2.2.2. LBCS Case Study Consent Form



UNIVERSITY OF LEEDS
Section of Oncology and Clinical Research
Cancer Research UK Clinical Centre
Leeds Institute of Molecular Medicine
Cancer Research Building
St James's University Hospital
Beckett Street
Leeds LS9 7TF

CONSENT FORM: PRE-OPERATIVE PATIENTS

A MOLECULAR GENETIC ANALYSIS OF UROTHELIAL CELL CARCINOMA (PROJECT 99/156)

EPIDEMIOLOGY OF DNA REPAIR AND BLADDER TUMOURS (PROJECT 02/192)

	Please tick to confirm
I have read the patient information sheet dated 11th April 2007 (version 3) for the above studies. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	☐
I understand the purpose of the studies and how I will be involved.	☐
I agree to participate in the first study (99/156)	☐
I agree to participate in the second study (02/192).	☐
First study (99/156)	
I agree to donate samples of blood for use in this and future studies.	☐
I agree to donate samples of urine for use in this and future studies.	☐
I agree to specimens of bladder tumour and normal tissue, which are to be removed at my operation and which are not needed for diagnosis, being used in this and future studies.	☐
Second study (02/192)	
I agree to donate a sample of blood for use in this and future studies.	☐
I give permission for my questionnaire to be used and looked at and analysed in the strictest confidence.	☐
I agree for my anonymised data to be sent to other centres for further studies by an international group of researchers.	☐
I agree for DNA from my blood sample to be sent to other centres for further studies by an international group of researchers.	☐

Both studies

I give permission for my medical records to be used and looked at and analysed in the strictest confidence. ☐

I understand that future research using the sample(s) I give may include genetic research aimed at understanding the genetic influences on disease, but that the results of these investigations are unlikely to have any implications for me personally, and findings will not be disclosed to me. ☐

I understand that I am free to change my mind about donating samples for research purposes at any time without having to give a reason and without it affecting my medical care in any way.

I agree to take part in the above study.

Signed: …… Date …….

(NAME IN BLOCK LETTERS) ……..

Person taking consent (signature) …….. Date ……..

(NAME IN BLOCK LETTERS) ……..

11.2.2.3. LBCS Control Study Information Letter

Leeds Institute of Molecular Medicine

UNIVERSITY OF LEEDS
Section of Oncology and Clinical Research
Cancer Research UK Clinical Centre
Leeds Institute of Molecular Medicine
Cancer Research Building
St James's University Hospital
Beckett Street
Leeds LS9 7TF

PATIENT INFORMATION (CONTROLS):
MOLECULAR EPIDEMIOLOGY OF DNA REPAIR AND BLADDER TUMOURS (PROJECT 02/192)

Version 2, dated 5 April 2007

We are involved in research looking to increase our understanding of the development of bladder tumours. YOU DO NOT HAVE A BLADDER TUMOUR AS FAR AS WE KNOW. We would like to invite you to take part in a RESEARCH study looking at the differences in a person's DNA, compared to other people's, which may make them more likely to develop bladder tumours. This information may help us in the future to prevent bladder tumours developing in people at high risk of this disease. The findings will not be disclosed to you at a later date and will not affect you or your family.

We are looking at a group of patients who have bladder tumours, but we also need to study people like yourself, who do not, to make the information that we get about the bladder tumour patients more useful.

You are under no obligation to take part in the study and if you prefer not to do so this decision will be accepted without question, and your hospital treatment (if you are having any) will not be affected in any way.

If you agree to take part, we ask you to consent to provide a blood sample (5 ml or 1 teaspoonful), and to fill in a simple questionnaire, which will be checked with you afterwards by a doctor or nurse.

Samples once donated are treated as a 'gift' to the hospital. They would be kept indefinitely within the hospital and laboratories (until they are used up), and may be used in this project and in future research. Scientific research is collaborative and this sometimes involves exchange of samples between laboratories worldwide. We work with other laboratories in this way and may share samples to increase the range of measurements that can be made and increase their usefulness. It is very important that you are sure you are happy for us to keep and use these samples, before you agree to donate them.

All information collected about you during the study will be kept strictly confidential. Your questionnaire answers will be used by the researchers to link these facts with any findings we may make, so that we can evaluate our results. We would also like your permission to send results of the questionnaire data and blood analysis abroad to other researchers working in this field, but no one from outside the hospital could identify you from the reports. The results of studies arising from this research programme may be published in the medical literature but your identity and personal information will remain strictly confidential at all times.

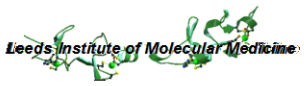
No special compensation arrangements have been made for this research programme. However, your rights to claim compensation from the hospital for injury or loss as a result of your treatment remain unaffected by your participation in this study.

For further information, or if you have any questions, please contact:

Dr Anne Kiltie
Molecular Radiobiology Group
Section of Experimental Oncology
Leeds Institute of Molecular Medicine
Cancer Research Building
St James's University Hospital
Beckett Street
LEEDS LS9 7TF Tel 0113 206 7275

The Leeds Teaching Hospitals NHS
NHS Trust

11.2.2.4. LBCS Control Study Consent Form



UNIVERSITY OF LEEDS
Section of Oncology and Clinical Research
Cancer Research UK Clinical Centre
Leeds Institute of Molecular Medicine
Cancer Research Building
St James's University Hospital
Beckett Street
Leeds LS9 7TF

CONSENT FORM (CONTROLS):
MOLECULAR EPIDEMIOLOGY OF DNA REPAIR AND BLADDER TUMOURS (PROJECT 02/192) Version 2, dated 5 April 2007

	Please initial box to confirm
I have read the information sheet, dated 5 April 2007 (version 2) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	☐
I understand the purpose of the study and how I will be involved.	☐
I give permission for my questionnaire answers to be used and looked at and analysed in the strictest confidence.	☐
I give permission for my medical records (if necessary and appropriate) to be used and looked at and analysed in the strictest confidence.	☐
I agree to donate sample(s) of blood for use in this and future studies.	☐
I agree for my anonymised data to be sent to other centres for further studies by an international group of researchers.	☐
I agree for DNA from my blood sample to be sent to other centres for further studies by an international group of researchers.	☐
I understand that future research using the sample(s) I give may include genetic research aimed at understanding the genetic influences on disease, but that the results of these investigations are unlikely to have any implications for me personally, and findings will not be disclosed to me.	☐
I understand that I am free to change my mind about donating samples for research purposes at any time without having to give a reason and without it affecting my medical care in any way.	☐

I agree to take part in the above study.

Signed: ..... **Date** .....

(PATIENT NAME IN BLOCK LETTERS) .....

Person taking consent (signature) ..... **Date** .....

(NAME IN BLOCK LETTERS) .....

When completed, 1 for patient, 1 for researcher site file, 1 (original) to be kept in medical notes.

The Leeds Teaching Hospitals NHS
NHS Trust

11.2.2.5. LBCS study questionnaire

QUESTIONNAIRE: EPIDEMIOLOGY OF DNA REPAIR AND BLADDER TUMOURS (PROJECT 02/192)

Name DOB Sex Hospital number

Date ……

1. Smoking history

Have you ever smoked as much as one cigarette a day for as long as a year? Yes ☐ No ☐

If Yes, how old were you when you started smoking cigarettes regularly? ☐☐ years old

Did you smoke at the following ages? If so, how many cigarettes did you smoke and were they usually filter cigarettes?

Age 20	☐☐☐ cigs per day	Filter ☐	No filter ☐	Non smoker ☐
Age 30	☐☐☐ cigs per day	Filter ☐	No filter ☐	Non smoker ☐
Age 40	☐☐☐ cigs per day	Filter ☐	No filter ☐	Non smoker ☐
Age 50	☐☐☐ cigs per day	Filter ☐	No filter ☐	Non smoker ☐

Did you smoke cigarettes one year ago? Yes ☐ No ☐

If Yes, how many cigarettes did you smoke each day? ☐☐ cigarettes

Did you usually smoke filter cigarettes? ☐ Yes ☐ No ☐

Did you usually smoke low tar cigarettes? ☐ Yes ☐ No ☐

Which brand did you normally smoke? _____

How deeply did you inhale? Deeply into lungs ☐ A little ☐ Not at all ☐

If you have stopped smoking, how old were you when you last smoked? ☐☐ years old

Did you smoke cigars one year ago? Yes ☐ no ☐

Did you smoke a pipe one year ago? Yes ☐ no ☐

2. **Occupational exposure**

Have you ever worked in the following:

Rubber industry	Yes ☐	No ☐
Plastics industry	Yes ☐	No ☐
Laboratories	Yes ☐	No ☐
Printing	Yes ☐	No ☐
Dyes and paints	Yes ☐	No ☐
With diesel fumes	Yes ☐	No ☐

Please turn over

Name:

What jobs have you done?

_____
_____

Please would you give us a contact telephone number for us to ask about your jobs in more detail?

Tel: (optional)

3. Family history

Have any members of your family (*blood relatives NOT relatives by marriage*) had bladder cancer?

	Yes	No	Don't know		Yes	No	Don't know	Not applicable
Mother	☐	☐	☐	Son(s)	☐	☐	☐	☐
Father	☐	☐	☐	Daughter(s)	☐	☐	☐	☐
Mother's mother	☐	☐	☐	Sister(s)	☐	☐	☐	☐
Mother's father	☐	☐	☐	Brother(s)	☐	☐	☐	☐
Father's mother	☐	☐	☐	Aunt/uncles	☐	☐	☐	☐
Father's father	☐	☐	☐	If **Yes**, numbers of aunts/uncles	☐			

4. Ethnic group

Do you consider yourself Caucasian (white skinned)? Yes ☐ No ☐

Do you consider yourself from another ethnic group? Yes ☐ No ☐

Please state

What was your place of birth? Town _____ Country _____

What was your mother's place of birth? Town _____ Country _____

What was your father's place of birth? Town _____ Country _____

BLADDER AND PROSTATE CANCER PRE-TREATMENT ASSESSMENT

Attach ID label

Date: _ _ _ _ _ _ _ _ _

WHO performance status ☐

Grade 0 Able to carry out all normal activity without restriction
1 Restricted in physically strenuous activity, but ambulatory and able to carry out light work
2 Ambulatory and capable of all self-care, but unable to carry out any work; up and about >50% of waking hours
3 Capable of only limited self care; confined to bed or chair>50% of waking hours
4 Completely disabled; cannot carry out any self care; totally confined to bed or chair

BLADDER SYMPTOMS Catheter in situ ☐ yes ☐ no

Grade		**0**	**1**	**2**	**3**
Continence		Total control	Occasional dribbling	Frequent dribbling	No control
Urgency		No problem	Mild problem	Moderate problem	Severe problem
Frequency		< x 1 per 2 hours	x 1 per 1-2 hours	x 2-3 per hour	x 4 or more per hour
Nocturia		0-1	2-3	4-6	7 or more
Haematuria		None visible	Occasional	Continuous	Clots + obstruction
Dysuria/pain		None	Mild	Moderate	Severe

Potency: ☐ satisfactory ☐ suboptimal ☐ nil
Smoker: ☐ never ☐ stopped ☐ continues
Weight loss ☐ none ☐ <10% ☐ >10%
Diabetes ☐ Y ☐ N
Family history ☐ Y ☐ N

Diabetes Y ☐ N ☐
Hypertension Y ☐ N ☐
Abdominal surgery Y ☐ N ☐
Prev Bowel Disease Y ☐ N ☐

BOWEL SYMPTOMS

Grade		**0**	**1**	**2**	**3**
Urgency		Rarely/never	>1/week	About once/day	> once/day
Frequency		Daily or less	Twice daily	x 3-4 daily	x 5 or more daily
Diarrhoea		None	Mild-medication unnecessary	Moderate–needs medication	Not controlled
Pain		None	Mild/occasional	Moderate	Severe
Bleeding/discharge		None	Spots, only with motion	Persistent	Haemorrhage

Prostate: Gleason grade ☐ T Stage ☐☐ Presenting PSA ☐☐.☐ Trus vol ☐☐☐
Neoadjuvant hormones Y ☐ N ☐ Reasons .....

Bladder: T ☐☐ N ☐ Grade ☐ Histology .....

Trial Y☐ N☐ Details .....

Assessment by.....

11.2.2.7. MIBC study Late Effects Assessment Form

BLADDER AND PROSTATE
LATE EFFECTS ASSESSMENT FORM

Attach ID label

Date: _ _ _ _ _ _ _ _

WHO performance status ☐

Grade 0 Able to carry out all normal activity without restriction
1 estricted in physically strenuous activity, but ambulatory and able to carry out light work
2 Ambulatory and capable of all self-care, but unable to carry out any work; >50%
3 Capable of only limited self care; confined to bed or chair>50% of waking hours
4 Completely disabled; cannot carry out any self care; totally confined to bed or chair

BLADDER SYMPTOMS Catheter in situ ☐ yes ☐ no

Grade	Patient grade	Description
0		None
1		Slight epithelial atrophy; minor telangiectasia (microscopic haematuria)
2		Moderate frequency, generalised telangiectasia; Intermittent macroscopic haematuria
3		Severe frequency and dysuria; severe generalised telangiectasia (often with petechiae); frequent haematuria; reduction in bladder capacity (<150 ml).
4		Necrosis; contracted bladder (capacity < 100 ml); severe haemorrhagic cystitis.

BOWEL SYMPTOMS

Grade	Patient grade	Description
0		None
1		Mild diarrhoea; mild cramping; bowel movements 5 times daily; slight rectal discharge or bleeding.
2		Moderate diarrhoea and colic; bowel movement > 5 times daily; Excessive mucus or intermittent bleeding
3		Obstruction or bleeding requiring surgery
4		Necrosis; perforation; fistula

ERECTILE FUNCTION

☐ satisfactory ☐ suboptimal ☐ nil
Compared with preXRT: ☐ better ☐ worse ☐ unchanged
Treatment of impotence: ☐ none ☐ Viagra ☐ other
Resulting in ☐ considerable benefit ☐ slight benefit ☐ no benefit

BLADDER CANCER PATIENTS: MOST RECENT CHECK CYSTOSCOPY

Date ..... Recurrence: Yes ☐ No ☐ Not known ☐

Intervention ....

PROSTATE CANCER PATIENTS: Most recent PSA ☐☐.☐ ☐ Date .....

Neoadjuvant hormones yes ☐ no ☐ Adjuvant hormones yes ☐ no ☐
Dates ..... to .....

Assessment by..... .....

11.3. Appendix C

11.3.1. *MUTYH* & *XPC* rare variants

All the below tables are in the file “Appendix C.xlsx” on the attached CD.

11.3.1.1. MUTYH & XPC variants detected by unindexed multiplexed NGS (NextGene used for variant calling)

11.3.1.2. XPC variants detected by indexed multiplexed NGS (Illuminator used for variant calling)

11.3.1.3. MUTYH & XPC variants detected by unindexed multiplexed NGS (Syzygy used for variant calling)

11.3.1.4. MUTYH & XPC rare variants single marker bladder cancer case-control study results

11.3.2. Analyses of confirmed *MRE11A* variants identified by indexed multiplexed NGS

11.3.2.1. Cancer-specific survival and late normal tissue radiotherapy toxicity

Table 60: Analysis for all confirmed *MRE11A* variants for associations with cancer-specific survival and radiotherapy toxicity. A multivariate survival-adjusted Cochran-Armitage trend test was performed for cancer-specific survival and an ordered logistic regression for late bladder and bowel radiotherapy toxicity.

					Cancer-specific survival		Late Bladder Toxicity		Late Rectal Toxicity	
NCBI hg19 position/ dbSNP rs number	**Variant base change**	**Gene Region**	**dbSNP MAF**	**Observed MAF**	*Per-allele HR (95%CI)*	*P-value*	*Per-allele OR (95%CI)*	*P-value*	*Per-allele OR (95%CI)*	*P-value*
rs11020802	G>T	5'-upstream	0.33	0.30	1.08 (0.74 - 1.59)	0.68	1.07 (0.67 - 1.72)	0.78	1.41 (0.74 - 2.69)	0.30
rs1805363	G>A	Intron 1 (Isoform 1)/ 5'UTR (Isoform 2)	0.11	0.11	**1.97 (1.26 - 3.07)**	**0.003**	1.70 (0.92 - 3.12)	0.09	1.71 (0.80 - 3.65)	0.17
rs524350	T>C	Intron 1	0.39	0.40	0.90 (0.63 - 1.28)	0.55	1.01 (0.64 - 1.57)	0.98	1.27 (0.68 - 2.38)	0.45
rs11825497	G>C	Intron 1	0.06	0.10	0.65 (0.35 - 1.18)	0.16	0.86 (0.40 - 1.83)	0.69	0.81 (0.27 - 2.46)	0.71
Chr 11: 94226459	C>T	Intron 1	N/A	0.003	Insufficient events		Insufficient events		Insufficient events	
Chr 11: 94226398	A>T	Intron 1	N/A	0.003	Insufficient events		Insufficient events		Insufficient events	
rs684507	C>T	Intron 1	0.41	0.40	0.90 (0.63 - 1.28)	0.55	1.01 (0.64 - 1.57)	0.98	1.27 (0.68 - 2.38)	0.45
rs497763	G>A	Intron 2	0.4	0.40	0.90 (0.63 - 1.28)	0.55	1.01 (0.64 - 1.57)	0.98	1.27 (0.68 - 2.38)	0.45
rs496797	G>A	Intron 2	0.4	0.40	0.90 (0.63 - 1.28)	0.55	1.01 (0.64 - 1.57)	0.98	1.27 (0.68 - 2.38)	0.45
rs493982	C>T	Intron 2	0.39	0.40	0.90 (0.63 - 1.28)	0.55	1.01 (0.64 - 1.57)	0.98	1.27 (0.68 - 2.38)	0.45
Chr 11: 94225127	G>A	Intron 2	N/A	0.003	Insufficient events		Insufficient events		Insufficient events	
rs11404578	Insertion T	Intron 2	0.44	0.003	Insufficient events		Insufficient events		Insufficient events	
rs13447588	T>C	Intron 2	0.006	0.003	Insufficient events		Insufficient events		Insufficient events	
rs13447590	G>C	Intron 2	0.006	0.02	1.11 (0.33 - 3.75)	0.86	0.86 (0.16 - 4.55)	0.86	1.19 (0.14 - 10.35)	0.88

rs61893736	C>T	Intron 2	N/A	0.003	Insufficient events		Insufficient events		Insufficient events		
rs10485020	T>A	Intron 2	0.008	0.003	Insufficient events		Insufficient events		Insufficient events		
Chr 11: 94223936	C>T	Intron 3	N/A	0.003	Insufficient events		Insufficient events		Insufficient events		
Chr 11: 94218991	A>G	Intron 4	N/A	0.003	Insufficient events		Insufficient events		Insufficient events		
rs11020799	G>A	Intron 4	0.36	0.37	0.91 (0.63 - 1.30)	0.59	0.93 (0.59 - 1.46)	0.75	1.08 (0.57 - 2.03)	0.82	
rs11020798	T>G	Intron 4	0.05	0.10	0.22 (0.39-1.30)	0.27	0.84 (0.38-1.84)	0.67	0.90 (0.30-2.74)	0.86	
rs13447601	Insertion GAGATTTTTT	Intron 4	0.25	0.20	0.90 (0.76-1.05)	0.19	0.93 (0.75-1.15)	0.50	0.84 (0.62-1.15)	0.28	
rs10501815	G>T	Intron 5	0.06	0.10	0.68 (0.37 - 1.25)	0.22	0.80 (0.37 - 1.76)	0.59	0.87 (0.29 - 2.64)	0.81	
rs680695	T>C	Intron 5	0.28	0.27	1.11 (0.74 - 1.67)	0.60	0.98 (0.60 - 1.59)	0.93	1.18 (0.61 - 2.29)	0.63	
rs535801	G>A	Intron 5	0.33	0.27	1.11 (0.74 - 1.67)	0.60	0.98 (0.60 - 1.59)	0.93	1.18 (0.61 - 2.29)	0.63	
rs3218740	C>T	Exon 6	0.002	0.003	Insufficient events		Insufficient events		Insufficient events		
Chr 11: 94211508	A>G	Intron 6	N/A	0.008	0.48 (0.07 - 3.55)	0.48	Insufficient events		Insufficient events		
rs5793683	Deletion C	Intron 6	0.48	0.003	Insufficient events		Insufficient events		Insufficient events		
Chr 11: 94210491	T>C	Intron 6	N/A	0.003	Insufficient events		Insufficient events		Insufficient events		
Chr 11: 94210210	C>A	Intron 6	N/A	0.003	Insufficient events		Insufficient events		Insufficient events		
rs13447619	G>A	Intron 6	0.01	0.003	Insufficient events		Insufficient events		Insufficient events		
Chr 11: 94209866	G>T	Intron 6	N/A	0.003	Insufficient events		Insufficient events		Insufficient events		
rs13447623	A>G	Intron 7	0.27	0.24	1.43 (1.00 - 2.08)	0.05	**2.12 (1.30 - 3.45)**	**0.003**	**1.92 (1.02 - 3.61)**	**0.04**	
rs610611	T>A	Intron 8	0.32	0.61	0.81 (0.56 - 1.16)	0.25	**0.56 (0.34 - 0.91)**	**0.02**	**0.45 (0.21 - 0.94)**	**0.03**	
rs115244417	C>G	Exon 9	N/A	0.003	Insufficient events		Insufficient events		Insufficient events		
Chr 11: 94200953	T>C	Intron 10	N/A	0.003	3.01 (0.28 - 32.07)	0.36	Insufficient events		Insufficient events		
rs78716391	G>C	Intron 10	0.07	0.10	0.68 (0.37 - 1.25)	0.22	0.80 (0.37 - 1.76)	0.59	0.87 (0.29 - 2.64)	0.81	
Chr 11: 94200365	C>T	Intron 10	N/A	0.03	0.65 (0.20 - 2.13)	0.48	2.26 (0.63 - 8.13)	0.21	0.78 (0.09 - 6.44)	0.82	
Chr 11: 94199802	G>A	Intron 10	N/A	0.005	Insufficient events		Insufficient events		Insufficient events		
Chr 11: 94199571	A>G	Intron 10	N/A	0.003	Insufficient events		Insufficient events		Insufficient events		
Chr 11: 94199359	T>C	Intron 10	N/A	0.003	Insufficient events		Insufficient events		Insufficient events		

rs533984	C>T	Intron 10	0.39	0.37	0.83 (0.57 - 1.21)	0.34	1.09 (0.69 - 1.73)	0.71	1.56 (0.82 - 2.98)	0.18
rs104895017	A>G	Intron 10	0.01	0.01	Insufficient events		0.71 (0.07 - 7.00)	0.77	Insufficient events	
rs640627	G>A	Intron 10	0.31	0.26	1.01 (0.67 - 1.54)	0.96	1.11 (0.67 - 1.85)	0.68	1.82 (0.92 - 3.62)	0.09
Chr 11: 94197302	A>G	Exon 11	N/A	0.003	Insufficient events		Insufficient events		Insufficient events	
rs641936	T>C	Intron 11	0.33	0.30	1.08 (0.74 - 1.59)	0.68	1.07 (0.67 - 1.72)	0.78	1.41 (0.74 - 2.69)	0.30
Chr 11: 94197239	T>C	Intron 11	N/A	0.003	Insufficient events		Insufficient events		Insufficient events	
rs13447654	A>G	Intron 11	0.01	0.02	0.76 (0.18 - 3.20)	0.71	1.65 (0.36 - 7.61)	0.52	3.01 (0.55 - 16.50)	0.20
rs13447655	A>G	Intron 11	0.14	0.21	1.44 (0.99 - 2.09)	0.06	**1.97 (1.20 - 3.24)**	**0.007**	1.63 (0.85 - 3.13)	0.14
Chr 11: 94196707	A>T	Intron 11	N/A	0.008	0.99 (0.13 - 7.59)	0.99	Insufficient events		Insufficient events	
Chr 11: 94196702	A>C	Intron 11	N/A	0.008	0.99 (0.13 - 7.59)	0.99	Insufficient events		Insufficient events	
Chr 11: 94196674	T>C	Intron 11	N/A	0.003	4.54 (0.58 - 35.24)	0.15	Insufficient events		Insufficient events	
Chr 11: 94196475	T>C	Intron 11	N/A	0.003	6.95 (0.42 - 115.65)	0.18	Insufficient events		Insufficient events	
rs657249	T>C	Intron 11	N/A	0.27	1.11 (0.74 - 1.67)	0.60	0.98 (0.60 - 1.59)	0.93	1.18 (0.61 - 2.29)	0.63
Chr 11: 94193845	Insertion CTC	Intron 12	N/A	0.008	1.07 (0.55 - 2.08)	0.84	1.64 (0.73 - 3.68)	0.23	1.54 (0.68 - 3.48)	0.30
rs10831230	G>T	Intron 12	0.06	0.10	0.66 (0.36 - 1.20)	0.18	0.90 (0.42 - 1.92)	0.78	0.84 (0.28 - 2.55)	0.76
rs592943	T>C	Intron 12	0.27	0.61	1.22 (0.85 - 1.76)	0.28	**1.76 (1.09 - 2.86)**	**0.02**	**2.20 (1.06 - 4.56)**	**0.04**
Chr 11: 94193208	A>C	Intron 12	N/A	0.005	0.90 (0.12 - 6.74)	0.92	2.17 (0.13 - 35.34)	0.59	Insufficient events	
rs11020789	G>C	Intron 12	0.06	0.10	0.66 (0.36 - 1.20)	0.18	0.90 (0.42 - 1.92)	0.78	0.84 (0.28 - 2.55)	0.76
Chr 11: 94193178	C>T	Intron 12	N/A	0.003	Insufficient events		Insufficient events		Insufficient events	
rs496190	T>C	Intron 12	0.44	0.40	0.90 (0.63 - 1.29)	0.58	1.02 (0.65 - 1.60)	0.92	1.29 (0.69 - 2.40)	0.43
rs61749249	C>A	Exon 13	N/A	0.008	0.76 (0.10 - 5.69)	0.79	1.08 (0.10 - 12.13)	0.95	Insufficient events	
Chr 11: 94189265	G>T	Intron 14	N/A	0.003	**19.46 (1.18 - 321.29)**	**0.04**	Insufficient events		Insufficient events	
Chr 11: 94189161	G>A	Intron 14	N/A	0.005	Insufficient events		Insufficient events		Insufficient events	
rs13447695	A>C	Intron 15	0.05	0.03	1.16 (0.41 - 3.27)	0.78	0.80 (0.20 - 3.12)	0.74	Insufficient events	
rs13447696	C>G	Intron 15	0.006	0.005	3.78 (0.48 - 29.98)	0.21	Insufficient events		Insufficient events	
rs12222920	C>G	Intron 15	0.06	0.10	0.66 (0.36 - 1.20)	0.18	0.90 (0.42 - 1.92)	0.78	0.84 (0.28 - 2.55)	0.76
Chr 11: 94179824	G>A	Intron 15	N/A	0.003	**24.29 (2.85 - 206.91)**	**0.004**	Insufficient events		Insufficient events	
rs556477	C>T	Intron 15	0.61	0.61	1.22 (0.85 - 1.76)	0.28	**1.76 (1.09 - 2.86)**	**0.02**	**2.2 (1.06 - 4.56)**	**0.04**

Chr 11: 94179497	A>G	Intron 15	N/A	0.003	Insufficient events		Insufficient events		Insufficient events	
rs1014666	A>G	Intron 15	0.44	0.40	0.90 (0.63 - 1.29)	0.58	1.02 (0.65 - 1.60)	0.92	1.29 (0.69 - 2.40)	0.43
Chr 11: 94170115	A>G	Intron 17	N/A	0.02	1.49 (0.53 - 4.16)	0.45	0.42 (0.05 - 3.68)	0.43	Insufficient events	
rs530569	G>A	Intron 17	0.3	0.27	1.11 (0.74 - 1.67)	0.60	0.98 (0.60 - 1.59)	0.93	1.18 (0.61 - 2.29)	0.63
rs13447717	T>C	Intron 17	0.06	0.09	0.70 (0.38 - 1.28)	0.24	0.84 (0.38 - 1.84)	0.67	0.90 (0.30 - 2.74)	0.86
rs61893706	G>A	Intron 17	N/A	0.03	2.68 (0.34 - 21.08)	0.35	Insufficient events		Insufficient events	
rs113009211	G>C	Intron 18	N/A	0.03	Insufficient events		Insufficient events		Insufficient events	
rs7126861	A>G	Intron 18	0.07	0.10	0.66 (0.36 - 1.20)	0.18	0.90 (0.42 - 1.92)	0.78	0.84 (0.28 - 2.55)	0.76
rs104895013	G>T	Intron 18	N/A	0.008	1.01 (0.14 - 7.33)	1.00	4.42 (0.39 - 49.76)	0.23	3.66 (0.32 - 42.04)	0.30
Chr 11: 94162799	T>C	Intron 19	N/A	0.02	0.33 (0.05 - 2.41)	0.27	0.30 (0.04 - 2.46)	0.26	Insufficient events	
Chr 11: 94153102	G>A	Exon 20 3'UTR	N/A	0.005	1.81 (0.25 - 13.21)	0.56	Insufficient events		Insufficient events	
rs13447749	G>A	Exon 20 3'UTR	0.06	0.05	0.83 (0.35 - 1.94)	0.67	2.11 (0.81 - 5.50)	0.13	1.38 (0.37 - 5.15)	0.63
Chr 11: 94152721	A>C	Exon 20 3'UTR	N/A	0.003	**11.10 (1.24 - 99.79)**	**0.03**	Insufficient events		Insufficient events	
Chr 11: 94152660	C>T	Exon 20 3'UTR	N/A	0.003	**9.12 (1.13 - 73.54)**	**0.04**	Insufficient events		Insufficient events	
rs11020777	T>C	Exon 20 3'UTR	0.11	0.08	0.90 (0.48 - 1.70)	0.75	0.86 (0.37 - 2.00)	0.73	1.06 (0.33 - 3.36)	0.92
Chr 11: 94151932	T>C	Exon 20 3'UTR	N/A	0.003	**8.33 (1.08 - 64.00)**	**0.04**	Insufficient events		Insufficient events	
rs104895004	G>T	Exon 20 3'UTR	N/A	0.003	Insufficient events		Insufficient events		Insufficient events	
rs2155209	A>G	Exon 20 3'UTR	0.32	0.38	0.82 (0.57 - 1.18)	0.28	**0.57 (0.35 - 0.92)**	**0.02**	**0.46 (0.22 - 0.95)**	**0.04**
Chr 11: 94150560	G>A	Exon 20 3'UTR	N/A	0.003	2.46 (0.32 - 18.72)	0.38	Insufficient events		Insufficient events	
rs13447762	T>G	3'-downstream	0.006	0.03	0.23 (0.03 - 1.66)	0.14	0.52 (0.11 - 2.54)	0.42	0.78 (0.09 - 6.44)	0.82

11.3.2.2. Bioinformatics functional predictions

Table 61: Functional predictions for all confirmed *MRE11A* variants

			Coding			Predicted miRNA binding		Splicing/ Transcription factor binding site
NCBI hg19 position/ dbSNP rs number	**Variant base change**	**Gene Region**	*Protein Change*	*Polyphen*	*SIFT*	*miRNA-binding site*	*ΔΔG (kJ/mol)*	
rs11020802	G>T	5'-upstream		Non-coding		Non-coding		ELF/BRCA1 transcription factor binding site
rs1805363	G>A	Intron 1 (Isoform 1)/ 5'UTR (Isoform 2)		Non-coding		Nil		5 bases 3' from Exon 1 GC donor splice site
rs524350	T>C	Intron 1		Non-coding		Non-coding		Nil
rs11825497	G>C	Intron 1		Non-coding		Non-coding		Nil
Chr 11: 94226459	C>T	Intron 1		Non-coding		Non-coding		Nil
Chr 11: 94226398	A>T	Intron 1		Non-coding		Non-coding		Nil
rs684507	C>T	Intron 1		Non-coding		Non-coding		Nil
rs497763	G>A	Intron 2		Non-coding		Non-coding		Nil
rs496797	G>A	Intron 2		Non-coding		Non-coding		Nil
rs493982	C>T	Intron 2		Non-coding		Non-coding		Nil
Chr 11: 94225127	G>A	Intron 2		Non-coding		Non-coding		Nil
rs11404578	Insertion T	Intron 2		Non-coding		Non-coding		Nil
rs13447588	T>C	Intron 2		Non-coding		Non-coding		Nil
rs13447590	G>C	Intron 2		Non-coding		Non-coding		Nil
rs61893736	C>T	Intron 2		Non-coding		Non-coding		Nil
rs10485020	T>A	Intron 2		Non-coding		Non-coding		Nil
Chr 11: 94223936	C>T	Intron 3		Non-coding		Non-coding		Nil

Chr 11: 94218991	A>G	Intron 4		Non-coding		Non-coding	Nil
rs11020799	G>A	Intron 4		Non-coding		Non-coding	Nil
rs11020798	T>G	Intron 4		Non-coding		Non-coding	Nil
rs13447601	Insertion GAGATTTTTT	Intron 4		Non-coding		Non-coding	Nil
rs10501815	G>T	Intron 5		Non-coding		Non-coding	Nil
rs680695	T>C	Intron 5		Non-coding		Non-coding	Nil
rs535801	G>A	Intron 5		Non-coding		Non-coding	6 bases 5' from Exon 6 AG receiver splice site
rs3218740	C>T	Exon 6		D142D (Synonymous)		Nil	Nil
Chr 11: 94211508	A>G	Intron 6		Non-coding		Non-coding	Nil
rs5793683	Deletion C	Intron 6		Non-coding		Non-coding	Nil
Chr 11: 94210491	T>C	Intron 6		Non-coding		Non-coding	Nil
Chr 11: 94210210	C>A	Intron 6		Non-coding		Non-coding	Nil
rs13447619	G>A	Intron 6		Non-coding		Non-coding	Nil
Chr 11: 94209866	G>T	Intron 6		Non-coding		Non-coding	Nil
rs13447623	A>G	Intron 7		Non-coding		Non-coding	Nil
rs610611	T>A	Intron 8		Non-coding		Non-coding	Nil
rs115244417	C>G	Exon 9	S334R	Benign	Tolerated	Nil	Nil
Chr 11: 94200953	T>C	Intron 10		Non-coding		Non-coding	Nil
rs78716391	G>C	Intron 10		Non-coding		Non-coding	Nil
Chr 11: 94200365	C>T	Intron 10		Non-coding		Non-coding	Nil
Chr 11: 94199802	G>A	Intron 10		Non-coding		Non-coding	Nil
Chr 11: 94199571	A>G	Intron 10		Non-coding		Non-coding	Nil
Chr 11: 94199359	T>C	Intron 10		Non-coding		Non-coding	Nil
rs533984	C>T	Intron 10		Non-coding		Non-coding	Nil

rs104895017	A>G	Intron 10		Non-coding		Non-coding		Nil
rs640627	G>A	Intron 10		Non-coding		Non-coding		Nil
Chr 11: 94197302	A>G	Exon 11	H401R	Benign	Tolerated	hsa-miR-892a	-4.4	Nil
rs641936	T>C	Intron 11		Non-coding		Non-coding		Nil
Chr 11: 94197239	T>C	Intron 11		Non-coding		Non-coding		Nil
rs13447654	A>G	Intron 11		Non-coding		Non-coding		Nil
rs13447655	A>G	Intron 11		Non-coding		Non-coding		Nil
Chr 11: 94196707	A>T	Intron 11		Non-coding		Non-coding		Nil
Chr 11: 94196702	A>C	Intron 11		Non-coding		Non-coding		Nil
Chr 11: 94196674	T>C	Intron 11		Non-coding		Non-coding		Nil
Chr 11: 94196475	T>C	Intron 11		Non-coding		Non-coding		Nil
rs657249	T>C	Intron 11		Non-coding		Non-coding		Nil
Chr 11: 94193845	Insertion CTC	Intron 12		Non-coding		Non-coding		Nil
rs10831230	G>T	Intron 12		Non-coding		Non-coding		Nil
rs592943	T>C	Intron 12		Non-coding		Non-coding		Nil
Chr 11: 94193208	A>C	Intron 12		Non-coding		Non-coding		Nil
rs11020789	G>C	Intron 12		Non-coding		Non-coding		Nil
Chr 11: 94193178	C>T	Intron 12		Non-coding		Non-coding		Nil
rs496190	T>C	Intron 12		Non-coding		Non-coding		Nil
rs61749249	C>A	Exon 13	A492D	Probably damaging	Affect protein function	hsa-miR-3659	6.6	Nil
Chr 11: 94189265	G>T	Intron 14		Non-coding		Non-coding		Nil
Chr 11: 94189161	G>A	Intron 14		Non-coding		Non-coding		Nil
rs13447695	A>C	Intron 15		Non-coding		Non-coding		Nil
rs13447696	C>G	Intron 15		Non-coding		Non-coding		Nil

rs12222920	C>G	Intron 15	Non-coding	Non-coding	Nil
Chr 11: 94179824	G>A	Intron 15	Non-coding	Non-coding	Nil
rs556477	C>T	Intron 15	Non-coding	Non-coding	Nil
Chr 11: 94179497	A>G	Intron 15	Non-coding	Non-coding	Nil
rs1014666	A>G	Intron 15	Non-coding	Non-coding	Nil
Chr 11: 94170115	A>G	Intron 17	Non-coding	Non-coding	Nil
rs530569	G>A	Intron 17	Non-coding	Non-coding	Nil
rs13447717	T>C	Intron 17	Non-coding	Non-coding	Nil
rs61893706	G>A	Intron 17	Non-coding	Non-coding	Nil
rs113009211	G>C	Intron 18	Non-coding	Non-coding	Nil
rs7126861	A>G	Intron 18	Non-coding	Non-coding	Nil
rs104895013	G>T	Intron 18	Non-coding	Non-coding	Nil
Chr 11: 94162799	T>C	Intron 19	Non-coding	Non-coding	Nil
Chr 11: 94153102	G>A	Exon 20 3'UTR	Non-coding	Nil	Nil
rs13447749	G>A	Exon 20 3'UTR	Non-coding	Nil	Nil
Chr 11: 94152721	A>C	Exon 20 3'UTR	Non-coding	Nil	Nil
Chr 11: 94152660	C>T	Exon 20 3'UTR	Non-coding	Nil	Nil
rs11020777	T>C	Exon 20 3'UTR	Non-coding	Nil	Nil
Chr 11: 94151932	T>C	Exon 20 3'UTR	Non-coding	Nil	Nil
rs104895004	G>T	Exon 20 3'UTR	Non-coding	hsa-miR-338-5p 4.4	Nil
rs2155209	A>G	Exon 20 3'UTR	Non-coding	Nil	Nil
Chr 11: 94150560	G>A	Exon 20 3'UTR	Non-coding	Nil	Nil
rs13447762	T>G	3'-downstream	Non-coding	Non-coding	Nil

11.3.2.3. Linkage disequilibrium plot for all confirmed MRE11A variants

(Figure in the file “Appendix C figures.pdf” on the attached CD”)

11.4. Appendix D

11.4.1. DNA repair gene 3'UTR SNPs and predicted miRNA binding sites

Table 62: All DNA repair gene 3'UTR SNPs identified located within a predicted miRNA binding site and predicted ΔG values.

Gene, dbSNP ID & Allele Substitution	Minor Allele Frequency	Predicted miRNA binding	Wildtype Allele ΔG	Variant Allele ΔG	ΔΔG	|ΔΔG|	|ΔΔGtot|[a]
XPC rs2229090 C>G	0.34	miR-339-3p	-18.97	-28.82	9.85	9.85	113.27
		miR-520d-3p	-19.02	-11.28	-7.74	7.74	
		miR-934	-19.93	-13.4	-6.53	6.53	
		miR-516b*	-20.25	-15.45	-4.8	4.80	
		miR-510	-26.57	-21.9	-4.67	4.67	
		miR-877	-23.18	-18.7	-4.48	4.48	
		miR-518e*	-20.89	-16.66	-4.23	4.23	
		miR-519c-5p	-20.89	-16.66	-4.23	4.23	
		miR-519b-5p	-20.89	-16.66	-4.23	4.23	
		miR-519a*	-20.89	-16.66	-4.23	4.23	
		miR-522*	-20.89	-16.66	-4.23	4.23	
		miR-523*	-20.89	-16.66	-4.23	4.23	
		miR-501-5p	-18.36	-22.52	4.16	4.16	
		miR-518c*	-23.14	-19.02	-4.12	4.12	
		miR-518d-5p	-19.99	-15.88	-4.11	4.11	
		miR-520c-5p	-19.99	-15.88	-4.11	4.11	
		miR-526a	-19.99	-15.88	-4.11	4.11	
		miR-150	-18.07	-22.17	4.1	4.10	
		miR-373	-20.34	-16.35	-3.99	3.99	
		miR-518f*	-20.73	-16.79	-3.94	3.94	
		miR-129-3p	-24.73	-20.98	-3.75	3.75	
		miR-186	-19.73	-23.31	3.58	3.58	
		miR-196a	-19.5	-15.97	-3.53	3.53	
		miR-526b	-18.47	-15.2	-3.27	3.27	
		miR-140-5p	-19.81	-16.76	-3.05	3.05	
		miR-708	-22.71	-19.78	-2.93	2.93	
		miR-939	-25.39	-22.48	-2.91	2.91	
		miR-34c-3p	-18.25	-21.13	2.88	2.88	
		miR-302c	-18.24	-15.36	-2.88	2.88	
		miR-615-5p	-25.71	-22.96	-2.75	2.75	
		miR-658	-22.33	-19.71	-2.62	2.62	
		miR-296-5p	-22.9	-25.5	2.6	2.60	
		miR-583	-23.85	-21.4	-2.45	2.45	
		miR-675	-24.17	-21.75	-2.42	2.42	
		miR-922	-21.85	-24.27	2.42	2.42	

		miR-874	-23.35	-20.95	-2.4	2.40	
		miR-769-3p	-20.66	-18.29	-2.37	2.37	
		miR-491-5p	-27.77	-29.81	2.04	2.04	
		miR-520f	-18.17	-16.16	-2.01	2.01	
		miR-525-3p	-19.18	-17.23	-1.95	1.95	
		miR-205	-23.16	-25.05	1.89	1.89	
		miR-525-5p	-18.15	-16.26	-1.89	1.89	
		miR-940	-24.91	-26.78	1.87	1.87	
		miR-188-5p	-19.11	-20.96	1.85	1.85	
		miR-657	-23.51	-25.23	1.72	1.72	
		miR-611	-30.32	-28.68	-1.64	1.64	
		miR-518b	-20.59	-22.1	1.51	1.51	
		miR-181a	-18.8	-17.46	-1.34	1.34	
		miR-181d	-20.75	-19.41	-1.34	1.34	
		miR-486-3p	-20.64	-19.4	-1.24	1.24	
		miR-181b	-20.14	-18.91	-1.23	1.23	
		miR-936	-20.3	-21.5	1.2	1.20	
		miR-184	-18.58	-19.67	1.09	1.09	
		miR-376a	-18.85	-19.9	1.05	1.05	
		miR-212	-18.24	-17.39	-0.85	0.85	
		miR-374b*	-21.22	-20.46	-0.76	0.76	
		miR-520c-3p	-18.48	-17.8	-0.68	0.68	
		miR-376c	-19.05	-19.62	0.57	0.57	
		miR-502-5p	-24.41	-24.74	0.33	0.33	
		miR-500	-23.04	-23.32	0.28	0.28	
		miR-342-3p	-20.8	-21.04	0.24	0.24	
		miR-518a-3p	-18.75	-18.77	0.02	0.02	
		miR-188-3p	-27.27	-27.27	0	0.00	
		miR-637	-23.09	-23.09	0	0.00	
ATM rs227091 T>C	0.44	miR-217	-20.55	-15.14	-5.41	5.41	38.91
		miR-338-3p	-19.43	-14.06	-5.37	5.37	
		miR-199b-5p	-18.31	-13.52	-4.79	4.79	
		miR-939	-29.67	-34.39	4.72	4.72	
		miR-199a-5p	-20.72	-16.02	-4.7	4.70	
		miR-24	-18.64	-14.81	-3.83	3.83	
		miR-593*	-18.39	-14.9	-3.49	3.49	
		miR-196a	-18.67	-22.01	3.34	3.34	
		miR-134	-20.48	-17.22	-3.26	3.26	
		miR-596	-30.88	-33.68	2.8	2.80	
		miR-615-5p	-22.73	-25.53	2.8	2.80	
		miR-500	-21.21	-23.77	2.56	2.56	
		miR-502-5p	-18.7	-20.88	2.18	2.18	
		miR-661	-25.15	-27.18	2.03	2.03	
		miR-512-3p	-18.26	-16.45	-1.81	1.81	
		miR-125a-3p	-19.45	-17.98	-1.47	1.47	

		miR-222	-19.98	-18.91	-1.07	1.07	
		miR-142-3p	-19.28	-18.29	-0.99	0.99	
		miR-423-5p	-18.46	-18.86	0.4	0.40	
		miR-601	-18.5	-18.27	-0.23	0.23	
LIG3 rs4796030 C>A	0.47	miR-612	-20.2	-14.84	-5.36	5.36	36.81
		miR-423-3p	-20.98	-15.75	-5.23	5.23	
		miR-346	-20.57	-15.4	-5.17	5.17	
		miR-221	-19.39	-14.9	-4.49	4.49	
		miR-888*	-19.88	-15.52	-4.36	4.36	
		miR-512-5p	-19.69	-15.79	-3.9	3.90	
		miR-615-3p	-25.33	-21.63	-3.7	3.70	
		miR-222	-24.35	-20.91	-3.44	3.44	
		miR-525-3p	-21.01	-17.71	-3.3	3.30	
		miR-508-5p	-20.99	-17.77	-3.22	3.22	
		miR-377	-20.98	-18.13	-2.85	2.85	
		miR-767-3p	-20.08	-22.7	2.62	2.62	
		miR-502-3p	-18.04	-15.54	-2.5	2.50	
		miR-941	-22.11	-19.82	-2.29	2.29	
		miR-127-3p	-21.14	-18.98	-2.16	2.16	
		miR-619	-19.51	-21.66	2.15	2.15	
		miR-188-3p	-24.22	-22.14	-2.08	2.08	
		miR-199b-5p	-19.24	-21.29	2.05	2.05	
		miR-191*	-19.38	-21.02	1.64	1.64	
		miR-342-3p	-21.65	-20.07	-1.58	1.58	
		miR-151-3p	-18.83	-17.45	-1.38	1.38	
		miR-572	-18.68	-20.04	1.36	1.36	
		miR-501-3p	-19.98	-18.73	-1.25	1.25	
		miR-937	-23.01	-24.25	1.24	1.24	
		miR-532-3p	-29.00	-28.46	-0.54	0.54	
		miR-220	-18.52	-18.54	0.02	0.02	
		miR-199a-5p	-21.55	-21.55	0	0.00	
		miR-662	-20.71	-20.71	0	0.00	
BRCA1 rs12516 C>T	0.36	miR-874	-23.67	-13.75	-9.92	9.92	25.76
		miR-324-3p	-21.78	-12.95	-8.83	8.83	
		miR-623	-30.47	-23.46	-7.01	7.01	
		miR-500	-22.59	-19.88	-2.71	2.71	
		miR-501-5p	-18.1	-15.39	-2.71	2.71	
		miR-616	-19.93	-17.22	-2.71	2.71	
		miR-181d	-18.27	-15.56	-2.71	2.71	
		miR-637	-25.74	-23.03	-2.71	2.71	
		miR-199a-3p	-19.4	-16.69	-2.71	2.71	
		miR-199b-3p	-19.4	-16.69	-2.71	2.71	
		miR-639	-23.28	-21.34	-1.94	1.94	
		miR-188-5p	-27.88	-26.04	-1.84	1.84	
		miR-146b-3p	-19.27	-18.00	-1.27	1.27	

		miR-371-5p	-20.75	-19.50	-1.25	1.25	
		miR-942	-19.22	-18.69	-0.53	0.53	
		miR-744	-23.85	-23.36	-0.49	0.49	
BRCA1 rs8176318 G>T	0.35	miR-328	-22.49	-18.39	-4.1	4.10	21.82
		miR-565	-20.16	-16.21	-3.95	3.95	
		miR-149	-20.88	-17.17	-3.71	3.71	
		miR-146b-3p	-20.4	-16.91	-3.49	3.49	
		miR-345	-17.51	-14.21	-3.3	3.30	
		miR-892b	-18.94	-15.67	-3.27	3.27	
		miR-639	-18.85	-20.36	1.51	1.51	
		miR-423-5p	-19.58	-19.58	0	0.00	
		miR-518c*	-20.46	-20.46	0	0.00	
PARP1 rs8679 T>C	0.17	miR-145	-20.54	-14.83	-5.71	5.71	17.42
		miR-105	-19.03	-14.12	-4.91	4.91	
		miR-630	-19.73	-16.11	-3.62	3.62	
		miR-302a	-24.45	-27.63	3.18	3.18	
		miR-302c	-30.35	-32.43	2.08	2.08	
		miR-302d	-27.79	-29.86	2.07	2.07	
		miR-302b	-27.56	-25.9	-1.66	1.66	
		miR-601	-20.81	-19.84	-0.97	0.97	
		miR-500*	-19.82	-20.46	0.64	0.64	
		miR-502-3p	-18.27	-18.84	0.57	0.57	
		miR-151-5p	-19.75	-19.75	0	0.00	
		miR-196b	-18.77	-18.77	0	0.00	
		miR-199a-5p	-23.83	-23.83	0	0.00	
		miR-199b-5p	-19.41	-19.41	0	0.00	
		miR-296-5p	-19.52	-19.52	0	0.00	
ATM rs1137918 A>G	0.22	miR-939	-23.78	-27.75	3.97	3.97	12.80
		miR-193a-5p	-19.65	-15.96	-3.69	3.69	
		miR-92a	-18.15	-15.35	-2.8	2.80	
		miR-508-5p	-18.03	-20.39	2.36	2.36	
		miR-565	-18.8	-20.67	1.87	1.87	
		miR-516a-5p	-19.82	-21.68	1.86	1.86	
		miR-920	-18.52	-19.88	1.36	1.36	
		miR-769-3p	-23.3	-24.6	1.3	1.30	
		miR-431	-19.4	-19.99	0.59	0.59	
		miR-638	-26.1	-25.79	-0.31	0.31	
		miR-299-3p	-19.4	-19.09	-0.31	0.31	
		miR-582-5p	-18.67	-18.81	0.14	0.14	
		miR-296-3p	-21.94	-21.94	0	0.00	
		miR-320	-21.02	-21.02	0	0.00	
		miR-615-5p	-25.13	-19.99	-5.14	5.14	
NBN (NBS1) rs2735383 G>C	0.32	miR-499	-14.03	-9.67	-4.36	4.36	8.70
		miR-508-3p	-19.29	-14.95	-4.34	4.34	

NBN (NBS1) rs1063054 A>C	0.32	miR-24	-12.12	-15.99	3.87	3.87	6.92
		miR-191*	-19.53	-16.48	-3.05	3.05	
		miR-630	-22.65	-20.15	-2.5	2.50	
		miR-145	-20.28	-18.61	-1.67	1.67	
		miR-330-5p	-18.7	-19.55	0.85	0.85	
		miR-657	-24.59	-24.61	0.02	0.02	
		miR-299-5p	-18.75	-18.75	0	0.00	
RAD51 rs7180135 A>G	0.47	miR-197	-18.32	-25.19	6.87	6.87	6.87
		miR-638	-19.23	-20.02	0.79	0.79	
		miR-129-5p	-18.62	-18.92	0.3	0.30	
		miR-200a*	-20.89	-20.62	-0.27	0.27	
LIG4 rs3093772 A>G	0.11	miR-377	-11.34	-15.42	4.08	4.08	4.08
BRCA2 rs15869 A>C	0.29	miR-653	-10.05	-10.05	0	0.00	0.00
FEN-1 rs4246215 G>T	0.39	miR-18a*	-21.42	-19.54	-1.88	1.88	0.00
		miR-493*	-15.71	-16.54	0.83	0.83	
LIG4 rs10131 G>A	0.10	miR-495	-5.38	-6.36	0.98	0.98	0.00
MRE11A rs2155209 A>G	0.29	miR-744	-20.99	-20.02	-0.97	0.97	0.00
		miR-584	-18.6	-18.35	-0.25	0.25	
NBN (NBS1) rs9995 T>C	0.32	miR-155	-18.22	-19.55	1.33	1.33	0.00
		miR-363*	-18.62	-17.44	-1.18	1.18	
PARP2 rs2700 A>C	0.28	miR-9	-12.44	-12.14	-0.3	0.30	0.00
		miR-299-5p	-12.26	-12.25	-0.01	0.01	
		miR-330-5p	-18.37	-18.37	0	0.00	
RAD23B (HR23B) rs11573727 T>-	0.34	miR-361-5p	-8.26	-8.26	0	0.00	0.00
RAD51 rs11855560 T>C	0.47	miR-34a*	-19.31	-21.24	1.93	1.93	0.00
RAD51 rs12593359 T>G	0.47	miR-766	-20.31	-18.32	-1.99	1.99	0.00
		miR-99b	-20.51	-20.08	-0.43	0.43	
XPC rs1126547 C>G	0.12	miR-141	-15.16	-13.17	-1.99	1.99	0.00

[a]|ΔΔGtot| is the sum of all |ΔΔG| values greater than 3kJ/mol only.

11.5. Appendix E

11.5.1. Publications

- **Teo MT**, Landi D, Taylor C, Elliott F, Vaslin L, Cox DG, Hall J, Landi S, Bishop DT, Kiltie AE (2012)
  The role of microRNA-binding site polymorphisms in DNA repair genes as risk factors for bladder cancer and breast cancer and their impact on radiotherapy outcomes.
  Carcinogenesis, 33(3), 581-6

11.5.2. Presentations

- **Teo MTW**, Snowden H, Nsengimana J, Harland M, Morgan J, Carr I, Taylor G, Barrett J, Bishop DT, Kiltie AE.
  Germline *MRE11* variants as predictive markers of radiotherapy response in muscle invasive bladder cancer.
  ARR & UKEMS Joint Annual Meeting 2011, Nottingham, UK, June 2011
- **Teo MTW**, Landi D, Hall J, Landi S, Bishop DT, Kiltie AE.
  DNA repair gene micro-RNA-binding site SNPs and bladder cancer risk.
  ESTRO 29, Barcelona, Spain, Sept 2010
- **Teo MTW**, Morgan J, Snowden H, Randerson-Moor J, Harland M, Chambers P, Taylor GR, Bishop DT, Kiltie AE.
  MUTYH and *XPC* rare disease gene variants in bladder cancer predisposition.
  Scientist in Training session, Association of Radiation Research 2010, Oxford, UK, June 2010

11.5.3. Posters

- **Teo MTW**, Snowden H, Nsengimana J, Harland M, Morgan J, Carr I, Taylor G, Barrett J, Bishop DT, Kiltie AE.
  Germline MRE11 variants predictive of radiotherapy outcomes in bladder cancer by next-generation sequencing.
  Poster Discussion Session, ESTRO 31, Barcelona, Spain, May 2012

- **Teo MTW**, Snowden H, Nsengimana J, Harland M, Morgan J, Carr I, Taylor G, Barrett J, Bishop DT, Kiltie AE.
  Germline *MRE11* variants as predictive markers of radiotherapy response in muscle invasive bladder cancer.
  Joint MRS / AMS / RCP Clinical Scientists in Training meeting, London, UK, Feb 2012
- **Teo MTW**, Snowden H, Nsengimana J, Harland M, Morgan J, Carr I, Taylor G, Barrett J, Bishop DT, Kiltie AE.
  Germline *MRE11* variants as predictive markers of radiotherapy response in muscle invasive bladder cancer
  YCR Annual Scientific Meeting, Harrogate, UK, June 2011
- **Teo MTW**, Morgan J, Snowden H, Randerson-Moor J, Harland M, Chambers P, Taylor GR, Bishop DT, Kiltie AE.
  Rare variants in the DNA repair genes, XPC and MUTYH, in bladder cancer predisposition.
  Joint MRS / AMS / RCP Clinical Scientists in Training meeting, London, UK, Feb 2011
- **Teo MTW**, Morgan J, Snowden H, Randerson-Moor J, Harland M, Chambers P, Taylor GR, Bishop DT, Kiltie AE.
  MUTYH and *XPC* rare disease gene variants in bladder cancer predisposition.
  Association of Radiation Research, Oxford, UK, June 2010
- **Teo MTW**, Morgan J, Snowden H, Randerson-Moor J, Harland M, Chambers P, Taylor GR, Bishop DT, Kiltie AE.
  XPC and *MUTYH* rare disease gene variants in bladder cancer predisposition.
  YCR Annual Scientific Meeting, Harrogate, UK, June 2010
- **Teo MTW**, Morgan J, Snowden H, Randerson-Moor J, Harland M, Chambers P, Taylor GR, Bishop DT, Kiltie AE.
  Identification of *XPC* and *MUTYH* rare disease gene variants in bladder cancer via clonal sequencing using the Illumina Genome Analyser platform.
  11th International Wolfsberg Meeting on Molecular Radiation Biology/ Oncology, Wolfsberg, Switzerland, June 2009