THE USE OF GERMLINE GENETIC PROFILES TO GUIDE PROSTATE CANCER TARGETED SCREENING AND CANCER CARE

Sarah Benafif

Division of Genetics & Epidemiology

The Institute of Cancer Research

This thesis is submitted for

the Degree of Doctor of Philosophy

Declaration

I hereby declare that the work presented in this thesis is the result of my own work. Assistance provided by other persons has been acknowledged in the Attributions section or within the chapter text. This work is original and has not been submitted for any other degree.

Sarah Benafif

Abstract

The genetic heritability of prostate cancer is contributed to by rarely occurring but high penetrant genetic variants, and moderate to commonly occurring variants conferring lower risks. Genome wide association studies and meta-analyses have discovered >170 prostate cancer risk loci. Utilising a prostate cancer polygenic risk score (PRS) could allow screening to be stratified by genetic risk. This thesis aims to investigate the potential role of germline genetic profiles in targeted screening for healthy men, and personalising treatment for affected patients.

Utilising a SNP based genetic profile in the community to target prostate cancer screening to men with increased genetic risk was acceptable to men in the community and their primary care teams. Within the BARCODE1 pilot study (N=307), uptake of screening procedures by men with a PRS in the top 10% of the risk distribution was 76%. Of the men screened, 33% were diagnosed with a low grade cancer. Further follow up is required to assess whether these are over-diagnosed indolent cases or early stage cancers that will progress further. The pilot study has moved on to the development of the main BARCODE1 study which is recruiting a further 4700 men.

I set up the lab workflow for the phase II BARCODE2 trial which is recruiting men with advanced prostate cancer for germline NGS using a study specific gene panel. I sequenced the first 100 patients recruited to the trial and found that 22% carried a germline protein truncating variant in a DNA repair gene. Carriers had more aggressive disease features compared with non-carriers. I also sequenced somatic DNA for a subset of the BARCODE2 carrier patients (N=8) using FFPE derived DNA. One case displayed evidence of loss of heterozygosity for the germline variant. Low frequency somatic variants in MMR and HR genes were identified in some cases. With the recently reported response rates to targeted agents such as PARPi in prostate cancer patients

with DNA repair defects, germline and somatic NGS will be increasingly important in the management of prostate cancer patients.

Acknowledgements

Attributions

The work within this thesis was carried out within the remit of the BARCODE project. The setting up of the BARCODE trials was a result of a large team effort. Below is a summary of my contribution and other team members' contributions to this thesis.

Chapter 3

The writing of the BARCODE1 pilot study protocol was led by Prof Eeles and contributed to by members of the ICR Oncogenetics team. The applications for regulatory approvals was already underway when I joined the team in September 2015. I attended monthly meetings with the GP recruiting sites and clinical research network representatives throughout the pilot study, to identify areas for improvement within the pilot as well as the subsequent main study which would recruit 5000 participants.

1. I screened the medical questionnaires of pilot study responders to confirm eligibility and exclude those not eligible for study entry.
2. I modified the participant information sheets in response to feedback from pilot study participants.
3. I updated the study protocol after the completion of pilot study recruitment to submit a new protocol for the main BARCODE1 study for regulatory approvals.
4. I attended the local research ethics committee (REC) meeting with the study Chief Investigator and Trial Manager when applying for approval for the main BARCODE1 study.
5. I analysed the recruitment rates and uptake within the pilot study as well as the screening results of men who completed screening procedures.
6. I met with patients diagnosed with prostate cancer in the pilot study to feedback results and explain management as recommended by the uro-oncology MDT.

The BARCODE1 trial manager, Eva McGrowder, carried out the relevant submissions to the Royal Marsden Committee for Clinical Research (CCR) as well as to the Health Regulatory Agency. The genotyping assay design was led by Dr Zsofia Kote-Jarai and

Ed Saunders. Genotyping of all pilot study samples was performed at Affymetrix® (part of Thermo Fisher Scientific) in the USA.

Manual review of SNP cluster plots for the pilot study genotyping data was carried out by myself and Ed Saunders. Genotyping was carried out for the 100 men in the BARCODE2 trial. I carried out the manual review of SNP cluster plots for this data set and used the in-house PRS application to generate the polygenic risk score data.

The genotyping data from the two comparison populations were provided by Dr Mark Brooks who coordinates the UK Genetic Prostate Cancer Study. I carried out the statistical tests for comparison of PRS distributions across the pilot study and the two comparator groups.

Chapter 4

For the BARCODE2 trial, I updated the initial draft protocol (written by Prof Eeles and Dr Christos Mikropoulos) in response to feedback from the European Research Council ethics committee. I also updated the protocol with the reported data in 2015/2016 regarding the frequency of germline DNA repair gene variants in men with advanced prostate cancer.

I attended the local CCR meeting with the Trial Manager to present the study for approval. I made further modifications to the protocol based on their feedback. I attended the local REC meeting with the Chief Investigator and Trial Manager. I responded to queries from the Medicines and Health Regulatory Agency (MHRA) regarding details of the NGS panel used in part 1 of the BARCODE2 trial.

I participated in regular trial progress meetings between the ICR Oncogenetics team and ICR Clinical Trials and Statistics Unit (ICR-CTSU). I modified template trial case

report forms (CRF) to fit the needs of the trial. I was involved in discussions around data management and the set-up of the trial database.

I screened the weekly uro-oncology clinic lists at the Royal Marsden Hospital (Chelsea) to identify eligible patients for trial entry. I attended a weekly clinic to recruit eligible patients. I also recruited patients referred specifically for trial entry via the Oncogenetics research clinics.

The design of the germline NGS panel was led by Prof Eeles and Dr Zsofia Kote-Jarai. I presented the latest reported germline data at the time from other research groups at the regular trial meetings to finalise the gene list for design. I also worked with Dr Dan Leongamornlert within the team, who in 2016 was analysing data from a case control study of prostate cancer patients (*Leongamornlert et al 2019*) to identify genes for inclusion in the BARCODE2 NGS panel.

I carried out the pilot experiments using stored DNA in the lab prior to the trial opening to set up the NGS workflow for part 1 of the BARCODE2 trial. I carried out the DNA extractions, library preparations, targeted capture and sequencing as well as Sanger sequencing for the first 100 men recruited to the trial. In my absence, the team's scientific officer, Ian Whitmore (later, Reshma Rageevakumar) continued the lab work to prevent delay within the trial timelines for the feedback of results. I analysed all the NGS results for the first 100 men recruited into the trial using the Agilent SureCall software. Analysis of NGS data using a bioinformatics pipeline was carried out for the pilot experiments by Ezequiel Anokian and Ian Whitmore.

Chapter 5

DNA extraction from FFPE tumour blocks was carried out by my colleague, Susan Merson. I carried out the library preparations, targeted capture and sequencing of the tumour DNA for this chapter. I analysed the NGS data using the SureCall program.

Publications and Abstract Presentations

1. **Genetic Predisposition to Prostate Cancer**. Sarah Benafif and Ros Eeles. *British Medical Bulletin* 2016; 120 (1): 75-89.

   Copyright Clearance number 4693690018180 for permission to re-use text from this article in this thesis.

2. **A Review of Prostate Cancer Genome-Wide Association Studies (GWAS)**. Sarah Benafif, Zsofia Kote-Jarai, Ros Eeles and the PRACTICAL Consortium. *Cancer Epidemiology, Prevention and Biomarkers* 2018; 27 (8): 845-857.

   Copyright Clearance number 4693200243122 for permission to re-use text from this article in this thesis.

3. **Germline sequencing of advanced prostate cancer patients in the BARCODE2 trial.** Sarah Benafif, Ian Whitmore, Ezequiel Anokian, Clara Cieza-Borrella, Ed Saunders, Zsofia Kote-Jarai and Ros Eeles.

   Oral abstract presented at the joint Cancer Genetics Group Meeting, Netherlands, March 2018

4. **The BARCODE-1 Pilot Study: The Use of Genetic Profiling to Guide Prostate Cancer Targeted Screening**. S Benafif, E Saunders, M Brook, E McGrowder, R Rageevakumar, D James, E Bancroft, H Bowen-Perkins, I Rafi, M Ferris, T Sevenoaks, A Beattie, J Bower, S Kuganolipava, Z Kote-Jarai and R Eeles.

   Poster presented at the National Cancer Research Institute Annual Conference, Glasgow, November 2018.

5. **The BARCODE-2 Trial: The Use of Genetic Testing to Guide Prostate Cancer Treatment.** R Eeles, S Benafif, R Rageevakumar, E Saunders, I Whitmore, C Cieza-Borrella, E McGrowder, B Jenkins, S Burnett, A Reid, A Falconer, J Walker, G Attard, E Bancroft, R Aligan, A Borges Da Silva, H Ni Raghallaigh, A Jones, E Hall and Z Kote-Jarai.

   Poster presented at the annual European Society of Medical Oncology Conference, Barcelona, September 2019.

6. **The BARCODE-2 Trial: The Use of Genetic Testing to Guide Prostate Cancer Treatment.** R Eeles, S Benafif, R Rageevakumar, E Saunders, I

Whitmore, C Cieza-Borrella, E McGrowder, B Jenkins, S Burnett, A Reid, A Falconer, J Walker, G Attard, E Bancroft, R Aligan, A Borges Da Silva, H Ni Raghallaigh, A Jones, E Hall and Z Kote-Jarai

<u>Poster presented at the American Society of Human Genetics Conference, Houston, October 2019.</u>

Table of Contents

List of Tables

List of Figures

Chapter 1   Introduction

1.1 General Introduction

Prostate cancer is the commonest non-cutaneous cancer in men in the UK with a lifetime risk of 13.24%, making it the cancer with the highest life time risk in UK males [1]. Although the mortality rate of prostate cancer has fallen since the 1990's (by approximately 20%), the incidence has risen significantly in the last 25 years; this is mostly attributed to the advent of PSA testing. There are approximately 47,700 new cases diagnosed in the UK every year (2014-2016, Cancer Research UK) and 11,631 deaths each year due to prostate cancer. Worldwide, there is a distinct geographical variation in the incidence of prostate cancer with the highest rates observed in parts of the Caribbean, and the lowest incidence in South and Central Asia [1] (Figure 1).

Figure 1: Age Standardised Incidence Rates of Prostate Cancer
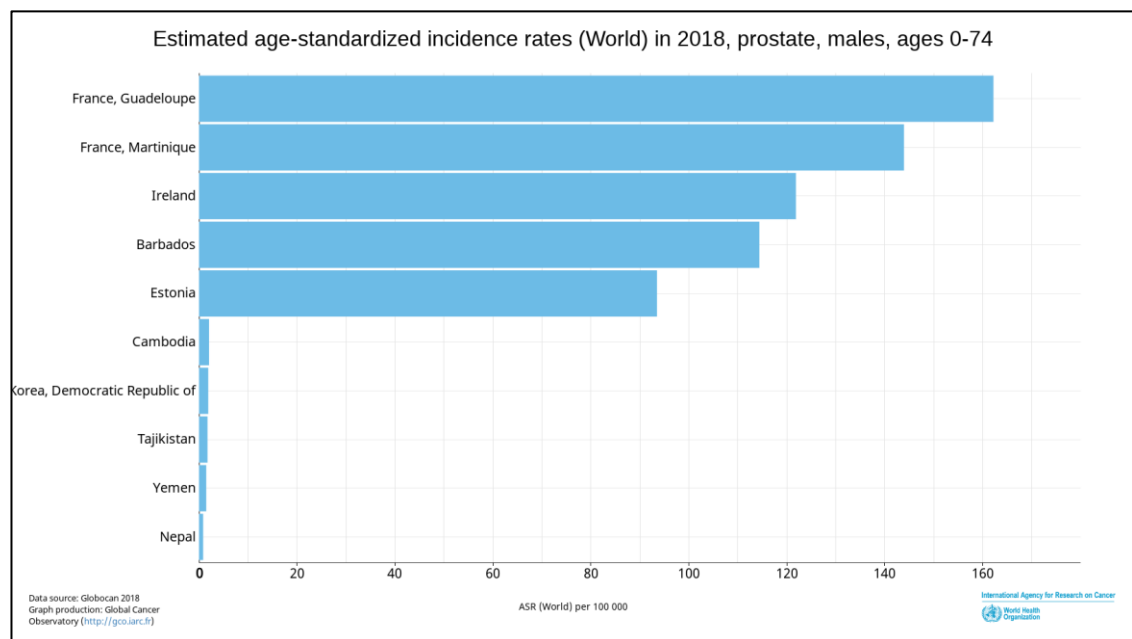
Bar chart showing top 5 and bottom 5 countries in the world for the age standardized incidence rates of prostate cancer according to 2018 WHO figures. (http://gco.iarc.fr/)

The aetiology of prostate cancer is not well understood, although epidemiological studies demonstrating a convergence of incidence rates in some populations migrating between areas with a low incidence to those with high incidence suggest environmental and lifestyle risk factors play a role;[2] this trend has been reported for a number of

Asian-American populations in the USA, for example, in Korean and Vietnamese men for whom the incidence of prostate cancer rose linearly between 1990 and 2008.[3, 4] This trend in incidence has not been observed in all populations migrating from Eastern countries to the West; in a study of migrants from the former Soviet Union to Germany, lower prostate cancer mortality and incidence were found in the migrants compared with the German rates with no increase in incidence in the longitudinal analysis[5]. These trends as well as differences in incidence rates between different ethnic groups suggest genetic factors contribute significantly to prostate cancer risk. Indeed, it has long been known that having a positive family history and/ or an Afro-Caribbean ethnic background increases the risk of prostate cancer development. Evidence from twin studies [6], as well as studies of familial prostate cancer highlight this. First degree relatives of prostate cancer patients have twice the risk of developing the disease compared to the general population[7]. In men diagnosed under the age of 60 years, the risk to their first degree relatives is more than fourfold that of those without a family history[8]. The variation in incidence according to ethnicity highlights the genetic component to prostate cancer aetiology; rates are higher in African American men compared with Asian American men [4, 9].

As with other complex diseases, the genetic heritability of prostate cancer is contributed to by both rarely occurring but higher penetrant genetic variants and moderate to commonly occurring variants conferring lower risks.[10] Current research on prostate cancer susceptibility variants can explain 37% of the familial relative risk (FRR) of prostate cancer [11, 12], attributed to commonly occurring (Minor allele frequency (MAF) >1%) single nucleotide polymorphisms (SNPs) as well as some rarer single nucleotide variants (SNVs). In the largest prostate cancer genome wide association study (GWAS) and meta-analysis [13] reported recently, 63 novel prostate cancer susceptibility loci were identified bringing the total number of known loci to 167.[10] (Table 1)

Table 1: Prostate cancer associated variants discovered through GWAS

Chr= Chromosome; RAF= Risk allele frequency; OR= Odds ratio

SNP ID	Band	Alleles^	RAF	OR	Reference/ PMID
63 novel SNPs identified on the OncoArray					
rs56391074	1p22.3	A/AT	0.379	1.05	4
rs34579442	1q21.3	C/CT	0.336	1.07	4
rs74702681	2p14	T/C	0.022	1.15	4
rs62106670	2p25.1	T/C	0.379	1.05	4
rs11691517	2q13	T/G	0.741	1.05	4
rs34925593	2q31.1	C/T	0.481	1.06	4
rs59308963	2q33.1	T/TATTCT GTC	0.729	1.05	4
rs1283104	3q13.12	G/C	0.379	1.04	4
rs182314334	3q25.1	T/C	0.895	1.10	4
rs142436749	3q26.2	G/A	0.012	1.23	4
rs10793821	5q31.1	T/C	0.573	1.05	4
rs76551843	5q35.1	A/G	0.991	1.31	4
rs4976790	5q35.3	T/G	0.113	1.08	4
rs4711748	6p21.1	T/C	0.225	1.05	4
rs9469899	6p21.31	A/G	0.357	1.05	4
rs9296068	6p21.32	T/G	0.651	1.05	4
rs12665339	6p21.33	G/A	0.167	1.06	4
rs17621345	7p14.1	A/C	0.741	1.07	4
rs11452686	7p21.1	T/TA	0.558	1.04	4
rs527510716	7p22.3	C/G	0.241	1.07	4
rs10122495	9p13.3	T/A	0.29	1.05	4
rs1048169	9p22.1	C/T	0.379	1.07	4
rs1182	9q34.11	A/C	0.220	1.07	4
rs141536087	10p15.3	GCGCA/G	0.150	1.10	4
rs1935581	10q23.31	C/T	0.623	1.06	4
rs7094871	10q25.2	G/C	0.537	1.04	4
rs547171081	11p11.2	CGG/C	0.470	1.05	4
rs1881502	11p15.5	T/C	0.190	1.06	4
rs61890184	11p15.4	A/G	0.124	1.08	4
rs2277283	11q12.3	C/T	0.313	1.06	4
rs12785905	11q13.2	C/G	0.048	1.09	4
rs11290954	11q13.5	AC/A	0.676	1.07	4
rs1800057	11q22.3	G/C	0.025	1.13	4
rs138466039	11q24.2	T/C	0.01	1.28	4
rs878987	11q25	G/A	0.146	1.07	4
rs2066827	12p13.1	T/G	0.755	1.07	4
rs10845938	12p13.1	G/A	0.551	1.06	4
rs7968403	12q14.2	T/C	0.643	1.07	4

rs5799921	12q21.33	GA/G	0.699	1.08	4
rs7295014	12q24.33	G/A	0.34	1.06	4
rs1004030	14q11.2	T/C	0.59	1.04	4
rs11629412	14q13.3	C/G	0.578	1.06	4
rs4924487	15q15.1	C/G	0.84	1.06	4
rs33984059	15q21.3	A/G	0.978	1.20	4
rs112293876	15q22.31	C/CA	0.289	1.07	4
rs11863709	16q21	C/T	0.960	1.17	4
rs201158093	16q23.3	TAA/TA	0.440	1.05	4
rs28441558	17p13.1	C/T	0.056	1.14	4
rs142444269	17q11.2	C/T	0.788	1.08	4
rs2680708	17q22	G/A	0.605	1.04	4
rs8093601	18q21.2	C/G	0.44	1.04	4
rs28607662	18q21.2	C/T	0.096	1.07	4
rs12956892	18q21.32	T/G	0.30	1.05	4
rs533722308	18q21.33	CT/C	0.412	1.05	4
rs10460109	18q22.3	T/C	0.42	1.04	4
rs11666569	19p13.11	C/T	0.711	1.06	4
rs118005503	19q12	G/C	0.911	1.11	4
rs61088131	19q13.2	T/C	0.835	1.05	4
rs11480453	20q11.21	C/CA	0.602	1.05	4
rs6091758	20q13.2	G/A	0.464	1.09	4
rs9625483	22q12.1	A/G	0.027	1.17	4
rs17321482	23p22.2	C/T	0.866	1.07	4
rs138004030*	6q27	G/A	0.92	1.28	4

^Reference allele/ risk allele
*Associated with early-onset prostate cancer

SNPs discovered in European populations					
SNP ID	**Chr**	**Risk alleles**	**RAF**	**OR**	**Reference/ PMID**
rs636291	1	A	0.683	1.04	25217961
rs17599629	1	G	0.218	1.07	25217961
rs1218582	1	G	0.4467	1.05	23535732
rs4245739	1	A	0.738	1.10	20197460, 23535732
rs11902236	2	T	0.269	1.05	23535732
rs9287719	2	C	0.467	1.07	25217961
rs9306895	2	C	0.364	1.08	26025378
rs1465618	2	T	0.2141	1.09	19767753
rs721048	2	A	0.1822	1.10	18264098
rs10187424	2	T	0.5738	1.08	21743467

rs12621278	2	A	0.9414	1.27	19767753
rs2292884	2	G	0.2413	1.06	21743057
rs3771570	2	T	0.1495	1.09	23535732
rs2660753	3	T	0.1028	1.13	18264097
rs7611694	3	A	0.579	1.09	23535732
rs10934853	3	A	0.277	1.10	19767754
rs6763931	3	A	0.442	1.04	21743467
rs10936632	3	A	0.5074	1.10	21743467
rs10009409	4	T	0.311	1.06	25217961
rs1894292	4	G	0.515	1.06	23535732
rs12500426	4	A	0.4632	1.07	19767753
rs17021918	4	C	0.6507	1.09	19767753
rs7679673	4	C	0.592	1.13	19767753
rs2242652	5	G	0.794	1.17	21743467
rs2121875	5	C	0.33	1.05	21743467
rs6869841	5	T	0.209	1.04	23535732
rs4713266	6	C	0.517	1.05	25217961
rs7767188	6	A	0.210	1.06	25217961
rs130067	6	G	0.2021	1.05	21743467
rs3096702	6	A	0.3771	1.06	23535732
rs3129859	6	G	0.670	1.06	25217961
rs2273669	6	G	0.1462	1.07	23535732
rs1933488	6	A	0.5788	1.08	23535732
rs9364554	6	T	0.2826	1.11	18264097
rs12155172	7	A	0.220	1.10	23535732
rs10486567	7	G	0.763	1.14	18264096
rs56232506	7	A	0.451	1.06	25217961
rs6465657	7	C	0.4635	1.11	18264097
rs2928679	8	A	0.437	1.05	19767753
rs11135910	8	T	0.1529	1.08	23535732
rs12543663	8	C	0.295	1.12	19767752
rs10086908	8	T	0.697	1.13	19767752
rs183373024	8	G	0.007	2.91	23104005
rs16901979	8	A	0.032	1.56	17401366
rs620861	8	G	0.631	1.15	19767752

rs6983267	8	G	0.511	1.22	17401363
rs1447295	8	A	0.107	1.41	17401363, 17401366
rs17694493	9	G	0.136	1.08	25217961
rs1571801	9	T	0.268	1.03	18073375
rs76934034	10	T	0.917	1.12	25217961
rs10993994	10	T	0.383	1.23	18264096, 18264097
rs3850699	10	A	0.700	1.07	23535732
rs4962416	10	C	0.2668	1.06	18264096
rs7127900	11	A	0.1985	1.19	19767753
rs7931342	11	G	0.504	1.17	18264097
rs11568818	11	T	0.550	1.08	23535732
rs11214775	11	G	0.709	1.07	25217961
rs80130819	12	A	0.908	1.10	25217961
rs10875943	12	C	0.287	1.07	21743467
rs902774	12	A	0.1526	1.13	21743057
rs1270884	12	A	0.482	1.07	23535732
rs8008270	14	C	0.8139	1.09	23535732
rs7141529	14	C	0.499	1.05	23535732
rs8014671	14	G	0.580	1.05	25217961
rs684232	17	C	0.3534	1.09	23535732
rs11649743	17	G	0.8055	1.13	18758462
rs4430796	17	A	0.5253	1.22	17603485
rs138213197	17	T	0.002	3.85	22236224
rs11650494	17	A	0.0779	1.10	23535732
rs1859962	17	G	0.4813	1.17	17603485
rs7241993	18	C	0.6949	1.08	23535732
rs8102476	19	C	0.5393	1.09	19767754
rs11672691	19	G	0.7368	1.10	19318570, 23065704
rs2735839	19	G	0.8527	1.18	18264097
rs2427345	20	C	0.621	1.05	23535732
rs6062509	20	T	0.6983	1.08	23535732
rs58133635	22	T	0.197	1.07	19117981, 25217961
rs5759167	22	G	0.502	1.15	19767753
rs2405942	23	A	0.7833	1.05	23535732
rs5945619	23	C	0.364	1.11	18264097

rs2807031	23	C	0.182	1.06	25217961
rs5919432	23	T	0.8008	1.04	21743467
rs6625711	23	A	0.1763	1.01	25217961
rs4844289	23	G	0.384	1.04	25217961
SNPs discovered in non-European populations					
rs2055109	3	T	0.7643	1.02	22366784
rs12653946	5	T	0.4246	1.08	20676098
rs1983891	6	T	0.2773	1.09	20676098
rs339331	6	T	0.695	1.09	20676098
rs1512268	8	T	0.4296	1.14	20676098
rs817826	9	T	0.8552	1.00	23023329
rs2252004	10	A	0.1017	1.00	22366784
rs12791447	11	G	0.0747	1.05	26443449
rs1938781	11	G	0.2297	1.03	22366784
rs9600079	13	T	0.443	1.01	20676098
rs58262369	14	C	0.998	1.27	26443449
rs7210100	17	A	0.0001	1.34	21602798
rs103294	19	C	0.7812	1.00	23023329
rs75823044	13	T	0.022	1.55	Reference [13]
rs78554043	22	C	0.015	1.62	Reference
SNPs discovered in multi-ancestry populations					
rs1775148	1	C	0.359	1.04	25217961
rs9443189	6	A	0.857	1.07	25217961
rs7153648	14	C	0.082	1.03	25217961
rs12051443	16	A	0.344	1.03	25217961
rs12480328	20	T	0.928	1.11	25217961
rs1041449	21	G	0.433	1.05	25217961
rs2238776	22	G	0.802	1.05	25217961

1.1.1 Clinical Heterogeneity of Prostate Cancer

Prostate cancer is a clinically heterogeneous disease that ranges from localised cancer in the prostate, and locally advanced cancer which has spread beyond the prostate capsule with or without regional lymph node involvement, to metastatic disease which has spread outside the pelvis and/ or to the bones. Newly diagnosed prostate cancer which is localised or locally advanced is risk stratified according to the presenting PSA level, Gleason score (the pathological grading system for prostate cancer) and T stage (Table 2,0). The term aggressive prostate cancer is used for tumours displaying the high risk features displayed in Table 2 and may include cases where distant metastases are present at the time of initial diagnosis. The specific definition of aggressive prostate cancer varies between research studies.

Low risk prostate cancer (and some moderate risk cases) is often managed with active surveillance which entails regular PSA measurements as well as interval magnetic resonance imaging (MRI) of the prostate with repeat biopsies when appropriate. If moderate or high risk features develop during active surveillance then definitive treatment in the form of radiation or surgery (prostatectomy) are considered. Active surveillance is not considered for high risk localised prostate cancers as these require definitive treatment. In patients with high risk features, radiotherapy to the prostate (with or without treatment of the pelvic lymph nodes) is combined with adjuvant androgen deprivation therapy (ADT) and may be followed by systemic chemotherapy (Docetaxel).

In patients who present with metastatic disease at the time of diagnosis, treatment of the primary prostate cancer is usually only considered for the management of lower urinary tract symptoms secondary to the primary tumour. In patients with oligometastatic disease, treatment of the primary and metastatic disease may be pursued and this type of treatment approach is currently under investigation. First line systemic treatment for metastatic prostate cancer is in the form of ADT which is continued indefinitely. This

may be combined with up front docetaxel chemotherapy as this has been shown to confer a significant survival benefit, particularly in patients with high volume metastatic disease[13, 14]. Although not yet approved by NICE (National Institute of Clinical and Healthcare Excellence) in the UK, the early use (prior to castration resistance) of the androgen receptor (AR) targeted agents Abiraterone or Enzalutamide in combination with ADT has been reported to confer a survival benefit. In the UK, within the NHS, these AR targeted agents are used in the treatment of metastatic prostate cancer that has progressed on ADT (termed castration resistant). Other treatments available in the setting of metastatic castration resistant prostate cancer (mCRPC) include cabazitaxel chemotherapy and the alpha emitter, radium-223 (for patients with bone only disease).

Table 2: Prostate Cancer Risk Stratification (NICE Guideline NG131)

Risk Level	Presenting PSA (ng/ml)	Gleason Score (range 6-10)	T Stage
Low	<10	6	T1 to T2a
Moderate	10-20	7	T2b
High	>20	8-10	≥T2c

Table 3: TNM Prostate Cancer Staging (European Association of Urology)

T – Primary Tumour (based on digital rectal examination)*		
TX	Primary tumour cannot be assessed	
T0	No evidence of primary tumour	
T1	Clinically inapparent tumour that is not palpable	
	T1a	Tumour incidental histological finding in 5% or less of tissue resected
	T1b	Tumour incidental histological finding in more than 5% of tissue resected
	T1c	Tumour identified by needle biopsy (e.g. because of elevated prostate-specific antigen [PSA])
T2	Tumour that is palpable and confined within the prostate	
	T2a	Tumour involves one half of one lobe or less
	T2b	Tumour involves more than half of one lobe, but not both lobes
	T2c	Tumour involves both lobes
T3	Tumour extends through the prostatic capsule	
	T3a	Extracapsular extension (unilateral or bilateral)
	T3b	Tumour invades seminal vesicle(s)
T4	Tumour is fixed or invades adjacent structures other than seminal vesicles: external sphincter, rectum, levator muscles, and/or pelvic wall	
N - Regional (pelvic) Lymph Nodes		
NX	Regional lymph nodes cannot be assessed	
N0	No regional lymph node metastasis	
N1	Regional lymph node metastasis	
M - Distant Metastasis		
M0	No distant metastasis	
M1	Distant metastasis	
	M1a	Non-regional lymph node(s)
	M1b	Bone(s)
	M1c	Other site(s)

*Pathological staging based on histopathological tissue assessment is similar to clinical TNM except for clinical stage T1c and T2 subgroups. All histopathologically confirmed organ-confined prostate cancers after radical prostatectomy are pathological stage T2 and the current Union for International Cancer Control (UICC) no longer recognises pT2 subgroups.

1.2 Genetic Association and GWAS

With the strong epidemiological evidence pointing to a hereditary component to the development of prostate cancer, much research into causative genes has been undertaken. Linkage studies investigating possible high risk loci leading to prostate cancer development identified possible loci on several chromosomes, but most have not been consistently replicated by subsequent studies,[15] with the exception of *HOXB13*. Linkage studies investigate the co-segregation of genetic markers with a disease. The lack of significant findings from these studies suggests that the hereditary aetiology of prostate cancer has a significant polygenic inheritance.

With the advances in genomic technology and high throughput DNA genotyping techniques, and by utilising databases of millions of common (MAF >1-5%) SNPs such as the HapMap[16] and 1000 Genomes project[17], GWASs have been developed to investigate the common genetic variants predisposing to cancer. GWAS allows investigators to take an unbiased approach when scanning the genomes of thousands of cases and controls to identify SNPs that associate with cancer[18]. GWASs have enabled the discovery of SNPs and SNVs in or near genes previously not known to be involved in cancer development. From projects, such as the HapMap project, it is known that certain SNPs will tend to occur together although they are located separately in the genome, and not always within the same gene[16]. This phenomenon known as linkage disequilibrium (LD) allows a GWAS to utilise several hundred thousand 'tag SNPs' to generate data on millions of SNPs. One of the first published GWAS was carried out in prostate cancer cases and controls, [19] and since then several GWASs have been carried out yielding approximately 170 prostate cancer risk loci[20] in European populations.

Several GWASs have been carried out in non-European populations such as Korean, Japanese, Arab and West African men to reveal both shared risk loci as well as some

that may be unique to these ethnic groups (all GWASs are listed in the National Human Genome Research Institute and European Bioinformatics Institute (NHGRI-EBI) Catalog of published GWASs: http://www.ebi.ac.uk/gwas).

1.2.1 Variants identified by GWAS

One of the earliest prostate cancer GWAS identified SNPs lying on 8q24; subsequent studies have revealed the 8q24 region to be rich with variants associated with multiple cancers including prostate cancer.[21, 22] The biological mechanism leading to prostate carcinogenesis though is unclear as the risk loci lie in non-coding regions of DNA. The nearest gene to this region is *MYC*, a proto-oncogene disrupted in many cancers. Functional studies including chromatin conformation assays, such as 3C, have shown long-range chromatin interactions of the 8q24 SNPs and these are thought to influence the expression of genes such as *MYC*. [10]

Other data from 3C experiments incorporating multi-target sequencing identified both intra- and inter-chromosomal interactions of 8q24 loci.[23] An example of an inter-chromosomal interacting gene is *CD96* on chromosome 3q13 with multiple interacting loci or 'hot-spots' in both chromosomal regions. In the same study,[23] 8q24 interactions with genes involved in the Wnt signalling pathway were also observed, suggesting that some risk SNPs have effects on the expression of multiple genes and may influence several cell signalling pathways. The Wnt signalling pathway is involved in prostate bud growth and luminal epithelial differentiation providing a plausible mechanism linked to prostate cancer development. Other studies have also linked the risk allele of the 8q24 SNP rs6983267 to enhanced Wnt signalling and other genomic regions harbouring prostate cancer risk SNPs show significant enrichment of Wnt signalling genes.[23] Identifying such interactions will allow better understanding of the biological mechanisms leading to prostate cancer. [10]

1.2.2 GWAS and meta-analyses

As the number of cases and controls included in modern GWAS has increased, so has the yield of new 'hits' from each study. In a study reported by Eeles et al in 2008, a two stage GWAS identified 7 novel prostate cancer associated variants in addition to confirming previously identified loci. [24] This study included nearly 2000 cases and 2000 controls in stage one, followed by over 3000 cases and 3000 controls in stage two. Some of the SNPs identified in that study were linked to candidate genes that could be involved in prostate cancer: *MSMB, LMTK2* and *KLK3*.

Recent GWAS projects have combined their data in meta-analyses with other large scale genotyping studies to increase study sample and power and in turn increase variant identification. One of these studies by Al-Olama et al [25] analysed more than 10 million SNPs from GWASs carried out in populations of different ethnicities (European, African, Japanese and Latino populations) and analysed GWAS data for ~43,300 cases and ~43,700 controls. This led to the identification of 23 new prostate cancer associated SNPs.[25] Although this was the first study to identify susceptibility variants associated with aggressive disease, the 16 SNPs that fit this category were not specific to aggressive cases and were also found to associate with non-aggressive prostate cancer.

Prostate cancer risk SNPs identified in most GWAS analyses confer a low to moderate risk of disease development with odds ratios (OR) ranging from 0.74-1.62. [15] Therefore, single risk SNPs do not pose a clinically significant effect on their own, but the risk is cumulative (multiplicative or log additive) and increases with increasing numbers of risk alleles present in an individual. Although traditional GWASs utilising catalogues of commonly occurring SNPs were not powered to detect rarer(MAF <1%) variants which may have a higher relative risk (RR) of prostate cancer development; the more recent GWAS and meta-analyses carried out under the auspices of large

consortia have allowed the inclusion of large populations of cases and controls to enable the identification of rarer risk variants. With the formation of international consortia such as PRACTICAL (PRostate Cancer Association group To Investigate Cancer Associated Alterations in the genome), these types of GWAS have now become feasible.

1.2.3 The OncoArray identified prostate cancer risk loci

The most recent 63 prostate cancer risk loci identified were a result of a large GWAS and meta-analysis[13]carried out by groups in the OncoArray network.[26] The goal of this network is to gain new insights into the genetic architecture and mechanisms underlying common cancers through the use of a custom designed Illumina array, the OncoArray, to genotype SNPs in cases of the most common cancers (breast, ovary, endometrium, lung, colon, prostate) and cancer-free controls. Individuals that have a genetic predisposition to cancer such as *BRCA1* and *BRCA2* mutation carriers were also included. Each consortium participating in this project contributed between 10,000 to 100,000 cases and controls; inclusion of large numbers of cases and controls has led to the identification of both common and rare variants associated with cancer risk and is likely to identify variants that are shared across cancer types. [10]

The OncoArray project was established, in part, through the efforts of the GAME-ON (Genetic Associations & MEchanisms in ONcology) network set up by the NCI (National Cancer Institute, USA). GAME-ON brings together international collaborators with the long-term goal of providing a rigorous knowledge base to enable clinical translation of GWAS findings. The importance of post GWAS research has been increasingly recognised as the majority of GWAS identified cancer risk SNPs lay in non-coding regions of DNA.

Computational methods such as enrichment analysis suggest that a large number of the prostate cancer risk SNPs (as well as other cancer SNPs) are enriched in multiple functional regions such as the binding regions of transcription factors or histone modifiers.[27] Further studies are required to validate these associations and to further characterise the molecular mechanisms underlying variant association with prostate cancer.

The OncoArray is a high density array comprising nearly 570,000 SNPs, of which approximately 80,000 are prostate cancer associated SNPs derived from a previous multi-ethnic meta-analysis [25] as well as from fine-mapping of known prostate cancer loci, and candidate variants. The OncoArray was used to genotype 46,939 prostate cancer cases and 27,910 controls of European ancestry.[13] [10]

The OncoArray prostate cancer genotyping data were combined with data from other large scale genotyping studies including 32, 225 prostate cancer cases and 33, 202 controls (also of European ancestry) to carry out a meta-analysis of more than 140,000 men. This led to the identification of 63 novel loci related to prostate cancer susceptibility. Of these, 52 were identified by imputation of the OncoArray genotyping data. Imputation relies on the LD of SNPs described previously with regard to 'tag' SNPs. Incorporation of the large GWAS backbone (260,000 SNPs) on the OncoArray allowed investigators to utilise LD to increase the power of variant discovery. The large numbers of cases and controls in this study also allowed several sub-analyses of clinical and demographic factors such as age at disease onset and aggressiveness of prostate cancer.

A novel variant at 6q27 (rs138004030) was found to be significantly associated with early onset disease (OR= 1.27; p=2.85×10^{-8}). In an analysis of advanced prostate cancer cases, 4 variants were found to be significantly associated with advanced

prostate cancer ($P<5\times10^{-8}$). These were significantly associated with overall prostate cancer risk but when advanced and non-advanced cases were compared, there was only a marginal statistically significant difference observed($P<1.0\times10-3$).[13]

Several candidate genes were identified among the new 63 prostate cancer variants; one of these is an *ATM* missense variant rs1800057 (NM_000051.3:c.3161C>G [p.Pro1054Arg] OR=1.16; P=8.15x10-9). Although this missense variant has been classified as 'benign' in the ClinVar database, ATM has been implicated in prostate cancer development and particularly with aggressive disease.[28] The ATM protein is a key checkpoint kinase that acts as a regulator of a wide range of downstream proteins including TP53 and BRCA1, checkpoint kinase CHK2, checkpoint proteins RAD17 and RAD9, and the DNA repair protein NBS1. It is therefore a key player in the DNA damage response (DDR) pathway. Recent studies have investigated the frequency of germline mutations in DNA repair genes in prostate cancer and *ATM* has been identified as the second most commonly altered gene in such studies (discussed later). [29] [30]

Another missense variant (rs2066827) was identified in *CDKN1B* (cyclin dependent kinase inhibitor 1B) (OR=1.06; $P=2.31\times10^{-9}$; T>G [Val109Gly]) which belongs to the *Cip/Kip* family of cyclin dependent kinase inhibitors. CDKN1B protein controls cell cycle progression at the G1 stage, and *in vitro* studies have shown levels of CDKN1B to be linked to increased tumour size and grade. This particular variant has previously been implicated in familial prostate cancer as well as advanced disease.[10]

A third candidate gene was identified by a variant in an intron of *RASSF3* (rs7968403; OR=1.06; $P=3.38\times10^{-12}$). RASSF3 is a GTP-binding plasma membrane protein and is a member of the RAS signalling pathway which is aberrant in approximately one third of cancers.

These findings are significant, as very few risk variants have been found to lie within gene sequences. Functional studies of these variants are still required to further delineate aberrant biological pathways, but with the development of gene editing techniques such as CRISPR, the molecular mechanisms that link a variant to prostate cancer development may be identified.

1.3 Rare genetic variants

The prostate cancer associated SNPs discussed so far confer a low to moderate risk of disease but do contribute cumulatively to a man's risk of prostate cancer development. The GWAS approach is based on the common disease common variant hypothesis and most are powered to detect SNPs with allele frequencies of ≥5%, although less common SNVs have been identified as discussed above. To detect rare variants (MAF <1%) which may confer a higher risk of disease, larger populations of cases and controls are needed and even then, very rare SNVs may be missed.

1.3.1 HOXB13

Linkage studies of familial prostate cancer identified one of the first known hereditary variants linked to prostate cancer development: the *HOXB13* p.G84E missense variant (NM_006361.5: c.251G>A).[31] Although this variant confers an OR of 5 for prostate cancer development and appears to be associated with early onset disease in several studies, the relationship with other clinical factors such as aggressive disease features is unclear. In a case-control study of a UK population of men [32], there was no correlation of mutation status with Gleason grade, presenting PSA or TNM staging. In contrast, a Danish study reported that carriers undergoing radical prostatectomy were more likely to have aggressive disease features (pre-operative PSA ≥20ng/ml, Gleason grade ≥4+3, presence of regional/ distant disease) compared with controls (54.2% vs 28.6%; p=0.011), but this study did not assess long term outcomes of relapse and

survival [33]. Further studies are needed to determine the prognostic significance of germline *HOXB13* mutations.

The G84E variant is seen mainly in men of European background with the highest carrier rates observed in populations in Finland and Sweden; carriers are more likely to have a family history of the disease compared to non-carriers. [34] *HOXB* genes encode transcription factors of the homeobox family, but the mechanism by which variants lead to prostate carcinogenesis is unknown. Mouse studies have shown that the HOXB13 protein is involved in prostate development [35] and has been linked to the growth of prostate cancer cell lines in an androgen-independent manner[36].

In view of the well-defined association with prostate cancer, *HOXB13* is now included in a number of commercial gene panels available to men with prostate cancer.

1.3.2 DNA repair genes

The *BRCA1/2* genes were the first DNA repair genes found to have an association with prostate cancer development. A fivefold increased risk due to pathogenic *BRCA2* variants was reported by the Breast Cancer Linkage Consortium (BCLC) [37]. *BRCA1* variants also appear to increase the risk of prostate cancer development although this is less pronounced than with *BRCA2*; the BCLC reported a 1.8-fold increased risk up to the age of 65 years with *BRCA1* variants. This association was confirmed by Leongamornlert et al who found the frequency of pathogenic *BRCA1* variants in a cohort of 900 prostate cancer cases (enriched with cases with early age of onset) to be 0.45% (estimated UK population carrier frequency ~0.1%), conferring a RR of 3.75 fold and an 8.6% cumulative risk by the age of 65 years. [38]

With the progress of DNA sequencing technologies, we have seen an accumulation of data related to germline variants in several DNA repair genes and prostate cancer. In the context of familial prostate cancer (i.e. 3 or more cases in a family), 7.3% of patients

were found to carry a deleterious mutation in a tumour suppressor gene.[29] Carrier status was associated with aggressive features such as nodal involvement and metastatic disease. In germline studies of both familial and sporadic prostate cancer, the most frequently aberrant gene is the *BRCA2* gene, often followed by *ATM*.[30]

In the TCGA study of 499 cases of localised prostate cancer, 4.6% of men were found to carry a germline mutation in a DNA repair gene. [30] In the setting of metastatic prostate cancer, the germline carrier rate for pathogenic variants in a DNA repair gene has been reported to range from 12-19% depending on the number of genes sequenced. [30, 39] These studies have also shown that carriers of variants in *BRCA2* and *ATM* have more aggressive disease features compared with non-carriers. These clinical features appear to translate to poorer prognoses, with shorter metastasis free survival and cause specific survival. [40, 41] These more recent studies did not select patients based on the presence of family history and interestingly did not show a difference in cancer family history when comparing carriers and non-carriers. This suggests that family history should not be used as a factor for identifying men who are being considered for genetic screening.

Poorer prognoses in *BRCA1/2* pathogenic variant carriers in addition to *ATM* carriers were also reported by Na et al [41]; patients with indolent and lethal prostate cancer were compared to assess the difference in frequency of germline variants in *BRCA1, BRCA2* and *ATM*. *ATM*, a DNA damage response gene, is the gene most frequently found to harbour germline pathogenic variants after *BRCA2,* in advanced prostate cancer. [30, 41, 42] A significantly higher combined carrier rate was identified in lethal prostate cancer cases compared with indolent cases (6.07% vs 1.44%; p=0.0007). Individually, the *BRCA2* carrier rate remained significantly higher in lethal cases (3.51% vs. 0.82%; p=0.013) and the *ATM* carrier rate difference approached statistical significance (1.92% of lethal cases vs. 0.41% of indolent cases; p=0.06). This study

also confirmed the association of carrier status with aggressive clinico-pathologic features at diagnosis e.g. higher Gleason score and higher presenting PSA level.

BRCA1 and 2 are involved in the homologous recombination (HR) DNA repair pathway. In the context of hereditary breast and ovarian cancer (HBOC), BRCA1/2 associated disease is known to be responsive to PARP inhibitors (PARPi), due to the effect of synthetic lethality, as well as platinum based chemotherapy. Synthetic lethality describes the phenomenon whereby a cell harbouring one of two gene or protein defects is viable while a cell containing both defects is not viable. In the case of BRCA1/2 associated HBOC, the combination of an inherent defect in BRCA1/2 function and iatrogenic PARP inhibition produce a synthetic lethal effect on tumours. These response patterns have led to the investigation of PARPi in the prostate cancer setting.

In the phase II TOPARP study [42], prostate cancer patients treated with the PARPi olaparib were assessed for both germline and somatic variants in DNA repair genes. Of the 49 patient cohort, 16 were found to have homozygous deletions, deleterious variants or both in DNA repair genes. Of these, 6 had a germline variant (3 in BRCA2 and 3 in ATM). In the 16 patients with DNA repair gene variants, 14 (88%) had a response to treatment including all 3 germline BRCA2 variant carriers. Although the numbers of patients in this study were small, the clinical implications of these results are significant, both in the somatic and germline genetic settings. Olaparib has been approved by the US FDA (United States Food and Drug Administration) and EMA (European Medicines Agency) for the treatment of platinum sensitive BRCA1/2-mutated (germline and/or somatic) high grade serous epithelial ovarian cancer. If the responses to olaparib seen in TOPARP-A are replicated in the expansion cohort enrolled into TOPARP-B (patients selected based on predictive deleterious variants in DNA repair genes in tumour tissue), it is highly likely that olaparib will be become an option for prostate cancer treatment in

the near future. A number of other PARPi are currently under investigation in prostate cancer associated with DNA repair defects.

Although treatment responses to agents such as PARPi appear to be promising for tumours harbouring DNA repair gene variants, [42] the data reported around responses to androgen receptor (AR) targeted therapies in men with germline variants are conflicting. In a Canadian study combining 4 cohorts of mCRPC patients, 7% of men (24 of 319 patients) were found to carry a germline pathogenic or likely pathogenic variant in a DNA repair gene; in these men the median PSA progression free survival (PFS) on first line AR targeted treatment was 3.3 months, half that seen in non-carriers. Interestingly, for carriers who were treated with first line docetaxel (N=8), the PSA PFS was higher at a median of 7.2 months which was comparable to 8.0 months observed for non-carriers, suggesting that mCRPC patients carrying a germline variant may benefit from having chemotherapy as their first line treatment for mCRPC rather than AR targeted agents. [41]

In contrast to this data, an American study reported in 2018 showed that men with germline variants in a DNA repair gene had longer clinical/ radiological PFS compared to men without germline variants when treated with first line AR targeted therapy (median 13.3 vs. 10.3 months; HR 0.67, p=0.107). [43] Furthermore, the results suggested *BRCA1/2/ATM* carriers had the best prognosis; the presence of a deleterious *BRCA1/2/ATM* variant but not variants in other DNA repair genes was independently associated with an improved PFS (HR 0.52, 95% CI 0.28-0.98; p=0.044). [43] These findings are in agreement with exploratory analyses carried out in the NCI9012 study,[44] where patients with somatic DNA repair variants had better than expected outcomes on both treatment arms of the study (Abiraterone or Abiraterone + Veliparib) compared to patients without somatic variants. Further prospective studies are needed, ideally with larger patient numbers, especially of germline carriers, to define the impact

of genetic variants on prognosis and treatment responses. If it is the case that specific genes influence treatment responses in men with hereditary variants while others do not, genetic testing in the clinical setting will be even more important in personalising treatment for such patients.

Until recently, all reported germline data in prostate cancer were the result of next generation sequencing studies of single or multi-centre cohorts of prostate cancer patients without non-cancer control comparisons. Leongamornlert et al recently reported the results of a case control study including 1285 cases of young onset prostate cancer (≤65 years) and 1163 age matched controls. In this study a set of 175 genes (107 in DNA repair pathway, 60 in DNA damage response pathways and cell cycle regulation, and 8 candidate genes) were interrogated by targeted exon sequencing. 24 genes were found to be significantly aberrant in cases compared with controls.[45] Within cases, aggressive (i.e. Gleason score ≥8, N=204) and non-aggressive cases (Gleason score ≤7, N=1,049) were also compared. Utilising two analysis approaches (gene level and gene-set level), 23 genes were identified that are associated with prostate cancer predisposition (Table 4). Although this 23 set of genes includes established genes such as *ATM, BRCA2 and CHEK2*, several novel gene associations were reported as well. Larger studies are required to evaluate these associations further. Of the 23 genes reported, 3 were found to associate with aggressive disease (*BRCA2, MSH2, CHEK2*). Variants in one gene, *ERCC2*, were found to specifically associate with aggressive prostate cancer. *ERCC2* is a DNA repair gene involved in the nucleotide excision repair (NER) pathway and is associated with the autosomal recessive condition, xeroderma pigmentosum. Deleterious variants in this so called 'aggressive gene set' increase the risk of developing aggressive prostate cancer by 11-fold. There was no difference in association with family history between carriers and non-carriers, highlighting that this is not a discriminatory criterion for genetic testing in the clinical setting. With the recent recommendation for genetic testing in

mCRPC by the National Comprehensive Cancer Network (NCCN; discussed below), these results suggest that the number of genes that should be considered for germline testing is likely to expand in coming years.

Table 4: 23 Genes associated with prostate cancer predisposition and/ or aggressive phenotype [45]

Established prostate cancer risk genes	Prostate Cancer candidate genes	Novel gene associations		Genes associated with aggressive disease
BRCA2 *ATM* *BRCA1* *CHEK2*	*GEN1* *MSH2* *RNASEL*	*NEIL2* *TDP1* *ERCC3* *LIG4* *MSH5* *POLE* *POLM*	*BLM* *PARP2* *POLD1* *CDC25C* *NHEJ1* *RECQL4*	*BRCA2* *CHEK2** *ERCC2* *MSH2*

*Non-del1100 variant

1.3.3 Mismatch Repair Genes

In 2017, the US FDA approved pembrolizumab, a PD-1 (programmed cell death protein 1) targeted immunotherapy, for the treatment of any tumour displaying a high risk for microsatellite instability (MSI-H). Lynch syndrome encompasses cases of inherited variants of mismatch repair (MMR) genes (*MSH2, MSH6, MLH1, MLH3, EPCAM, PMS2*) and is associated with an increased risk of development of colorectal and/or endometrial cancers. These tumours often display MSI-H on molecular testing, with associated deficiency of the relevant MMR protein (MMR-d) seen on immunohistochemistry. Several studies have reported an increased incidence of prostate cancer in men with Lynch Syndrome compared to non-carriers. A European study reported the standardised incidence ratio of prostate cancer in Lynch Syndrome men to be 5.9 with a cumulative risk by the age of 60 years of 9.8% and by age of 70 years 29% [46]. In an American study, similar cumulative risks were reported of 6% by

age 60 years and 30% by age 80 years. [47] Positive MMR mutation carrier status conferred a prostate cancer hazard ratio of 1.99. Lynch Syndrome men may also present at a younger age compared to the general population and be at risk of more aggressive disease features such as higher Gleason scores. [47]

Until recently, tumour testing for MSI or MMR-d was only routinely carried out for colorectal and endometrial cancers. A study presented at the American Society of Clinical Oncology (ASCO) 2018 annual conference suggested that more widespread tumour testing for MSI/ MMR-d should be carried out. In this study of 15,045 tumours across >50 cancer types, 6.8% of tumours were MSI-I (intermediate) or MSI-H. Germline variants in a MMR gene were identified in 1.9% of MSI-I tumours and 16.3% of MSI-H tumours. Half of these tumours belonged to a cancer type not traditionally associated with Lynch Syndrome including prostate cancer among others. Interestingly, 40% of these cases did not meet testing criteria for Lynch Syndrome, hence a call by the study investigators that all patients with tumours with MSI-I or MSI-H be referred for germline testing. [48] This would require wider testing of tumours for MSI initially in order for germline testing to be considered. In men with advanced prostate cancer, the prevalence of somatic MMR gene variants has been reported to range between 6-12%.[49, 50] A proportion of these cases will harbour a germline variant in a MMR gene.

In a cohort of 692 mCRPC patients germline sequenced for a set of DNA repair genes, 0.6% (4/692) were found to have a pathogenic variant in *PMS2, MSH6* or *MSH2*, [30] and in the case control study referred to in the previous section reported by Leongamornlert and colleagues, *MSH2* was found to be more frequently altered in the germline DNA of cases compared with controls (3/1283 vs. 0/1163; OR 6.35).[45] In the same study, *MSH5* was also found to be differentially aberrant in cases (4/1283 vs. 1/1163; OR 2.72). Identifying prostate cancer patients with somatic and/ or germline

variants in MMR genes may open up another line of treatment for this cohort with PD-1 targeted immunotherapy if their tumours display MSI-H/ MMR-d.

In the setting of prostate cancer, a phase I and phase II study have shown some activity of PD-1 directed therapy in cases of advanced disease. The KEYNOTE-028 study and then KEYNOTE-199 study enrolled heavily pre-treated mCRPC patients for 3 weekly pembrolizumab. Results of the KEYNOTE-028 study reported in 2018 showed an overall response rate (ORR) of 17.4% in men with PDL-1 (programmed cell death ligand 1) positive mCRPC with a median duration of response of 13.5 months. [51] The results of KEYNOTE-199 were presented at ASCO 2018 and showed a disease control rate (DCR) of 26% with no significant difference when patients were stratified by tumour expression of PDL-1. Disease control rate for >6 months was 11% overall and interestingly reached 22% in men with bone-only disease. Further studies are needed to identify predictive biomarkers of response to target immunotherapy to patients most likely to respond.[52]

1.3.4 Germline Genetic Testing

As variants in DNA repair genes have implications for treatment with the emerging data on PARPi sensitivity in tumours with *BRCA1/2* or *ATM* variants, among others [42], and the immunotherapy sensitivity of MMR-d tumours, there is a rationale for offering germline testing to men with advanced prostate cancer. Apart from identifying men who may benefit from personalised treatments, germline testing would also allow genetic counselling and cascade testing of other family members (both male and female) who may benefit from cancer screening.

Commercial gene panels for hereditary cancer have been available for some time and recently prostate cancer specific panels have become available although the number of genes tested varies between providers. Clinical guidelines around the genetic testing

of prostate cancer patients have only recently been updated based on the emerging data in the mCRPC setting. The NCCN modified their prostate cancer management guidelines in 2018 to include a recommendation to consider testing all men with mCRPC for germline and somatic variants in *BRCA1/2, ATM, PALB2* and *FANCA*. (NCCN Prostate Cancer Guidelines Version 3.2018) This update also added the recommendation to consider tumour testing for MSI or MMR deficiency. This guideline was updated further in 2019 (version 2.2019) to recommend germline genetic testing for all men with high-risk, very high-risk, regional or metastatic prostate cancer. The genes listed in the guideline for consideration of testing was also expanded to include *CHEK2* and the MMR genes: *MLH1, MSH2, MSH6* and *PMS2* in addition to the previously mentioned genes.

In addition to this, the NCCN 'Genetic/ Familial High-Risk Assessment for Breast and Ovarian Cancer' guideline was updated in 2018 to include a recommendation for *BRCA1/2* germline testing in all men with metastatic prostate cancer, without the requirement of a family history or young age of onset (Version 1.2018). For men without metastatic disease but with a high Gleason grade (≥7), there is a requirement that they meet other family history criteria in order to be tested. This guideline also recommends germline *BRCA1/2* testing for men whose tumours are found to have a somatic variant in *BRCA1/2*. Although these are the first set of guidelines to be explicit in their recommendation for genetic testing of metastatic prostate cancer patients regardless of family history, up to 37% of prostate cancer patients without metastatic disease carrying a germline pathogenic variant would not qualify for testing based on the current criteria.[53] As the recognition of the emerging data in this area widens, these guidelines are likely to be updated to widen the testing criteria of men with prostate cancer. Guidelines from other organisations are expected to follow which are likely to evolve further to include genes other than *BRCA1/2* and *ATM*.

1.4 Prostate Cancer Screening

PSA (prostate specific antigen) testing for prostate cancer screening has become an increasingly controversial topic. PSA is a prostate specific protein secreted by both healthy prostate tissue as well malignant cells, therefore it lacks the specificity required for a cancer screening test. PSA levels are prone to fluctuation and are influenced by conditions such as urinary tract infections, prostatitis and prostatic hypertrophy. Conversely, prostate cancer may exist in the presence of a 'normal' PSA level and therefore there is a risk of false negatives in its use for cancer screening. Despite this, PSA based screening for high risk men such as *BRCA2* mutation carriers is thought to be warranted,[54] largely due to the evidence of association with aggressive phenotypes and poorer outcomes in carriers who develop prostate cancer. Currently, the NCCN guidelines recommend commencing prostate cancer screening in *BRCA2* mutation carriers (and to 'consider' screening for *BRCA1* men) from the age of 45 years, although the screening format is not specific to this cohort of men as it relies on the 'Prostate Cancer Early Detection' guideline (Version 2.2019) which is designed for men identified to have an increased risk of prostate cancer development due to factors such as a strong family history. The latter NCCN document acknowledges the evolving data indicating that men carrying mutations in genes other than *BRCA1/2* may have an increased risk of prostate cancer development.

The ongoing IMPACT (Identification of Men with a Genetic Predisposition to Prostate Cancer) study (NCT00261456) is addressing the approach to prostate cancer screening for *BRCA1* and *BRCA2* carriers (and controls) as well as men with Lynch Syndrome (using prostate biopsy in those with a PSA >3 ng/ml). Initial results from the first screening round in this study have shown a higher positive predictive value (PPV) for PSA triggered biopsy in *BRCA2* carriers (PPV 48%) compared with controls (PPV 33%).[55] Prostate cancer detected in *BRCA2* carriers was classified as intermediate or high risk in two thirds of cases. Similarly, in the *BRCA1* carriers, 61% were found to

have intermediate or high risk disease. [55] More recently, the results of 3 years of screening in the IMPACT study have been reported [54] and confirmed a higher prostate cancer incidence rate per 1000 person years in *BRCA2* carriers compared with non-carrier controls (19.4 vs 12.0; p = 0.03). *BRCA2* carriers were diagnosed at a younger age (61 vs 64 yr; p = 0.04) and carrier cases had a higher proportion of clinically significant disease (77% vs 40%; p = 0.01). It is hoped that these results will encourage the development of more specific guidelines around prostate cancer screening for B*RCA2* mutation carriers.

1.4.1 Prostate Cancer Screening Studies

In terms of general population screening (outside the setting of *BRCA1/2* carriers), as data from two large prostate cancer screening studies [56, 57] have evolved, guidelines from national screening programs such as the USPSTF (US Preventive Services Task Force) have fluctuated from advising against PSA screening for prostate cancer (2012) to recommending that men make an individualised decision regarding PSA testing in conjunction with their clinician (for those aged 55-69 years; USPSTF 2017). In the UK, the National Screening Committee (NSC) recommends against universal screening for prostate cancer using PSA (2016, due to be updated 2019), and a guideline document published by Public Health England seeks to help GPs advise men over 50 years who are asymptomatic and seeking a PSA test. [58]

The controversy surrounding PSA based screening has not only arisen due to PSA's non-specific nature for prostate cancer detection, but also due to the flaws and conflicting conclusions of the two largest randomised prostate cancer screening studies: the American Prostate, Lung, Colorectal and Ovary (PLCO) study and the European Randomised Study of Screening for Prostate Cancer (ERSPC). Both studies had a degree of contamination in the control groups, although this was higher in the US study with patients in the control group receiving an average of 3 PSA tests compared with

the screening group who received 5 PSA tests on average, hence a lack of reduction in mortality with screening couldn't be excluded in this study. [59] In contrast, the ERSPC study reported a 20% relative reduction in prostate cancer mortality with screening and an absolute reduction of 7 prostate cancer deaths per 10 000 men (in men aged 55 to 69 years) at a median of 9 years of follow-up. Updated analyses at 13 years have shown a further improvement in mortality with an absolute reduction of 13 prostate cancer deaths per 10 000 men. These figures may improve further with longer term follow up. The mortality benefit in ERSPC is largely attributed to the reduction in metastatic prostate cancer cases (absolute reduction of 31 cases per 10 000 men at 12 years of follow-up). The evolution of this data in the last few years accounts for the change in screening guidance provided by the USPSTF [60] and it may be that the UK NSC guidelines will be similar when updated in 2019.

Although the screening studies investigating the use of PSA testing for prostate cancer detection have shown some survival benefit, the complications of prostate biopsies, high rate of false positive results (10 % in PLCO screened men and 18% in ERSPC) as well as over-diagnosis of indolent prostate cancers has led to the caution around the use of PSA testing. In the PLCO and ERSPC studies, 16% and 27.7% of screened men underwent a prostate biopsy respectively; of the biopsies performed, 67% and 76% were negative for cancer. This is not an insignificant proportion of men when considering the possible complications which include bleeding, pain and infection which occurred at a rate between 2-6%. Hospitalisation after prostate biopsy occurred in 0.5-1.6% of men.

1.4.2 Use of Genetic Profiles in Prostate Cancer Screening

Apart from over-diagnosis of indolent prostate cancer, which often does not require intervention, PSA tests can also miss significant tumours which would be treated actively. In the Stockholm-3 (STHLM-3) study, which investigated the use of a multi-

factor screening model, 21% of high risk prostate cancers had a PSA level in the range of 1-3ng/ml; below the threshold of ≥4ng/ml which is often used for screening. [61] Similarly, in the Prostate Cancer Prevention Trial (PCPT), high-grade prostate cancers were reported in 12.5% of men with PSA <0.5ng/ml. [62] For these reasons, alternative approaches to prostate cancer screening are required that do not solely rely on a non-specific and fluctuating biochemical test. In this context, the use of a genetic test for cancer screening is very attractive. Unlike PSA levels, germline DNA is constant and unchanging in terms of the SNVs and SNPs it holds, and only requires a one off measurement usually in the form of a blood test or saliva test. By utilising the known genetic variants associated with prostate cancer risk, a polygenic risk score (PRS) can be calculated for an individual to estimate their risk of prostate cancer development. It is feasible that such a score could be used to stratify men for prostate cancer screening so that those with a high genetic risk of prostate cancer are offered screening while men at lower genetic risk can avoid the potential complications of invasive tests. A retrospective study using a cohort from the screened arm of the PLCO study showed that profiling germline prostate cancer SNPs (33 SNPs producing a prostate genetic score (PGS)) can identify men who have a higher risk of developing disease, with men in the top quartile of PGS-33 score having the highest risk detection rate of prostate cancer.[63] A retrospective study of men in the placebo arm of the PCPT reported that a genetic risk score based on 29 prostate cancer associated SNPs was predictive of prostate cancer; both in men with and without a family history of the disease. [64]

Taking into account the findings of the most recent GWAS and meta-analysis linked to the OncoArray project; and utilising the known prostate cancer risk loci, the relative risk of prostate cancer for men in the top 1% of the genetic risk distribution based on a PRS is 5.71 compared with men in the 25-75th percentiles, and for those in the top 10% the RR is 2.69.[13] A risk model incorporating a genetic profile based on risk loci (with or

without family history information) could be used to target screening to those at highest risk.

Interestingly, a PRS may also allow a reduction in over-diagnosis in a screening program. Over-diagnosis refers to the detection of tumours by screening that would not have presented clinically in a person's lifetime in the absence of screening. [65] Pashayan and colleagues investigated the relationship between PRS (based on 66 prostate cancer risk SNPs) and prostate cancer over-diagnosis by genotyping 9,404 cases and 7, 608 controls in 3 UK based prostate cancer studies. They found that rates of over-diagnosis decreased with increasing PRS and reported a 56% reduction between the highest and lowest PRS quartiles in that analysis.[65]

Combining genetic risk information from a set of SNPs with other risk stratification methods such as prostate imaging may reduce the rates of over-diagnosis further. The PROMIS study investigated the utility of multi-parametric MRI (MP-MRI) of the prostate in the screening setting and the results suggested that using MP-MRI as a triage test may reduce the diagnosis of indolent disease while also identifying a higher proportion of clinically significant disease compared with standard trans-rectal ultrasound guided biopsies. [66]

The STHLM3 study was the first large prospective and population based prostate cancer screening study that assessed a targeted approach to screening. The study utilised a screening model combining plasma protein biomarkers (PSA, free PSA, intact PSA, hK2, MSMB and MIC1), 232 risk SNPs and a set of defined clinical variables (age, family history, previous prostate biopsy and prostate examination) and compared this with PSA measurement alone (using a threshold of ≥3ng/ml). [67] The STHLM3 model performed significantly better than PSA measurement for the detection of Gleason 7 or higher prostate cancer. Sensitivity for the detection of high risk prostate cancers was

significantly improved with the STHLM3 model; the AUC with PSA alone was 0.56 compared with 0.74 with the study model. It is unclear how much the genetic profile contributed to the study screening model, although all the variables used were significantly associated with high risk prostate cancer and contributed to a cumulative improvement in the AUC in the multivariate analysis. Further refinement of this model is needed as there were still a significant number of low grade cases diagnosed; over half of the tumours were Gleason 6 cancers. As modifications of this strategy continue to be investigated, for example with the incorporation of MRI and targeted biopsies (currently under investigation), it is likely that a more practical and feasible version of the STHLM-3 model for will be developed.

In the UK, the PROFILE study (NCT02543905) is investigating the value of a PRS in screening men with a family history of prostate cancer. Study participants are screened using PSA testing, MRI and biopsy. A genetic profile test is also carried out using a set of prostate cancer risk SNPs and the correlation between genetic risk and screening results will be investigated. The PROFILE study is also recruiting a separate cohort of black men to study genetic risk in this group.

1.5 Future Directions and Conclusions

As the cost of NGS and gene panel testing has fallen, the availability and accessibility of genetic testing has increased with several commercial tests now available to test for prostate cancer predisposition. It is plausible that commercial SNP tests currently offered by companies providing direct to consumer testing will start to incorporate prostate cancer risk SNVs to add prostate cancer risk information to health reports produced by such tests. The role of prostate cancer risk loci in predicting disease risk or guiding screening is unknown and requires investigation. Once this role is defined by prospective research trials, genetic profiling utilising risk loci may be found to be an

ideal method of targeting population screening to men at increased risk of disease development while sparing those at low risk from invasive screening procedures.

For higher penetrance hereditary genetic mutations, the clinical guidelines from regulatory bodies are expected to develop further as the data from sequencing studies increases. Agreement is needed on the selection of the most appropriate set of genes that should be tested in the uro-oncology setting. Currently available prostate cancer predisposition gene panel tests, although similar, differ in the range of genes tested. The number of genes ranges from 10-14 and all include *BRCA1/2, HOXB13* and the MMR genes. The panel at an expert consensus conference held in the USA in 2017 agreed that mCRPC patients should be tested for *BRCA1/2* and *ATM* mutations[68], while the panellists at the European 'Advanced Prostate Cancer Care Conference' (APCCC, St Gallen) in 2016 voted for 'large panel testing' including HR genes and MMR genes. [69]

With the efforts in developing personalised approaches to cancer management at the forefront of oncology research priorities, utilising germline genetic profiles to predict risk of disease development as well as other factors such as disease aggressiveness, combined with molecular target identification from somatic tumour profiles, will allow the development of a precision medicine approach to the screening for and treatment of prostate cancer.

1.6 Outline of this thesis

In this thesis, I aim to investigate the utility of germline genetics in two prostate cancer contexts. Firstly, I will assess the use of a germline SNP profile to identify men in the community who are at increased genetic risk of prostate cancer based on the known prostate cancer risk SNPs (in 2017). These men will be offered prostate cancer screening within the BARCODE1 trial. I will assess the feasibility and uptake of this

approach to prostate cancer screening within the BARCODE1 pilot study (N=300). By reviewing the uptake and trial processes in the pilot study, I will modify the study protocol and procedures as needed to be able to move into a larger study to recruit a total of 5000 men for SNP profiling and screening of those in the top 10% of the genetic profile distribution. I expect that the use of a SNP profile to target screening may provide a way to overcome the pitfalls of the PSA test for screening. By defining the role of SNP profiling in this setting, data from the BARCODE1 trial will inform the set up of a multi-modal screening program which uses a SNP profile as well as other parameters to identify men for screening.

The second part of my thesis will examine the use of a germline DNA repair gene panel in men with advanced prostate cancer to assess the frequency of mutation carriers. Based on published data so far, this is expected to be in the range of 12-18%. [28, 30, 41] These men may benefit from treatment with carboplatin chemotherapy. [70, 71] I will set up the BARCODE2 trial as part of my role within the ICR Oncogenetics team. This trial will recruit patients with mCRPC to undergo germline genetic screening utilising a study specific gene panel test. Those who carry a germline mutation in a DNA repair gene and have disease progression after two standard lines of treatment will be treated with carboplatin chemotherapy within the second part of the trial. Although this thesis will not report the clinical responses in the treatment part of the study, the trial will aim to examine responses in 3 patients groups, divided according to which gene is altered in the germline (1.*BRCA1/2*, 2.MMR genes and 3.other DNA repair genes including HR genes). The expectation is that *BRCA1/2* mutation carriers will respond to platinum chemotherapy as observed in retrospective studies, but examining the responses in the other two groups of patients may identify further groups of patients who may benefit from carboplatin chemotherapy. I will describe the set up of the study specific gene panel and laboratory workflow process for NGS in the BARCODE2 trial and report the carrier frequency in the first 100 men recruited to the trial.

Chapter 2   Materials and Methods

2.1 Chapter 3 Methods: BARCODE1 Pilot Study

2.1.1 BARCODE1 Pilot Study Design and Set Up

The BARCODE 1 Study was set up by the Institute of Cancer Research (ICR) Oncogenetics team to investigate the use of genetic profiling to identify men in the community who have an increased genetic risk of prostate cancer development so that they can be offered screening. A pilot study was set up prior to the main study. Research ethical approval was granted on the 26th January 2016 and the pilot study opened in April 2016 with the aim of recruiting 300 participants, with a view to then moving on to a main study recruiting another 4700 men.

I was involved in the study team's regular meetings with the Clinical Research Networks and General Practice collaborators to coordinate the recruitment of participants from the community. Men who expressed an interest in the study after receiving an invitation letter from their GP were screened for eligibility by the study nurse or myself (by reviewing their completed health questionnaires), and then entered the study by signing the informed consent form and providing a saliva sample. These were used for DNA extraction and genotyping (detailed below). Genotyping data were used to calculate a polygenic risk score (PRS) for each man in the pilot study. Men identified to be in the top 10% of the PRS distribution (i.e. men with a PRS >90th percentile) were offered screening for prostate cancer in the form of a magnetic resonance imaging (MRI) scan and biopsy of the prostate. PSA levels were also measured prior to biopsy. These procedures were carried out in the Royal Marsden Hospital (RMH).

2.1.2 Patient Selection

Men between the ages of 55-69 years who fit the study eligibility criteria (see study protocol in Appendix 1) were sent an invitation letter and a health questionnaire by GP

teams participating in the study. Recruitment to the pilot study commenced in April 2016 and completed in April 2018. Responders were screened by the trial nurse, with input from me when needed, for eligibility by reviewing their health questionnaires. Participants were required to have no medical contra-indications to prostate biopsy, to be aged 55-69 years at the time of study entry and to be of Caucasian/ European ethnicity. The ethnicity criterion was required as the single nucleotide polymorphisms (SNPs) used in the study's genetic profile are based on variants identified and validated in European populations.

2.1.3 DNA Extraction and Genotyping

DNA extraction from saliva was carried out externally by Tepnel Pharma Services (UK). Extracted DNA was normalised to 60ng/µl in 325µl final volume. Extracted DNA was sent to Affymetrix® (part of Thermo Fisher Scientific) in the USA for genotyping. A study specific panel of SNPs was designed by Dr Zsofia Kote-Jarai and Ed Saunders in the ICR Oncogenetics team in conjunction with the Affymetrix® scientific team. SNPs were selected based on the previously identified prostate cancer risk loci along with the most recent set reported by Schumacher et al as part of the OncoArray project. [13, 25] Table 5 lists the SNPs included in the assay (Genetic profile design described below). The genotyping assay utilises the Eureka™ Genomics protocol and is based on a ligation dependent polymerase chain reaction (PCR) that uses allele barcodes contained within the ligation probes as well as sample barcodes added by PCR. The main steps carried out are as follows:

1. DNA is heat denatured and mixed with a probe blend (three probes are required for each SNP to be interrogated)
2. For each SNP site, one of two left hybridisation sequence (LHS) probes (the two LHS probes are specific for the different alleles of the SNP) and a right hybridisation sequence (RHS) probe fully hybridise to the DNA.

3. Each LHS probe type contains a unique allele barcode sequence that provides the information for which SNP and allele the probe represents.
4. A ligase joins adjacent LHS and RHS probes to form a single fragment.
5. Sample identification barcode sequences (indexes) are added to the ligation products by PCR. Different barcode combinations are added to the different wells (one sample per well)
6. Each fragment therefore contains barcodes indicating which sample, SNP and allele it devolved from, so samples can be pooled after this step to generate the sequencing library. Fragments also contain the full Illumina® adapter sequences at this stage.
7. Sequence data are generated from the prepared libraries using an llumina® MiSeq™ instrument.

Relative read counts for the two possible allele barcodes are used to determine genotype at the SNP position for each sample. Preliminary quality control (QC) and genotype calling were carried out at Affymetrix® and genotyping data were sent to the ICR Oncogenetics team for further QC and analysis.

2.1.4 Genetic profile design and testing

A genotyping assay was designed to be used in the BARCODE1 study as well as other Oncogenetics studies being carried out by the team. SNPs associated with prostate cancer risk as a result of published genome wide association studies (GWAS) and meta-analyses were selected. As BARCODE1 is investigating screening in men of European ancestry, some of the SNPs included in the assay would not be used for the PRS calculation in this study as they were identified in non-European ancestry population studies.

At the start of assay development, 177 SNPs were submitted to Affymetrix® for inclusion in the design of the genotyping assay. These SNPs included:

- 99 SNPs identified in a previous GWAS and meta-analysis [25] and used in a previous genotyping assay

59

- 63 new SNPs identified in the most recent GWAS and meta-analysis [13]
- The *HOXB13* missense variant G84E
- 14 SNPs identified by fine-mapping the 8q24 region [12]

DNA sequences for each SNP were submitted including 75bp either side of the variant. During the *in silico* assessment of submitted SNPs by Affymetrix®, 6 variants were identified to be 'un-designable' due to their location within single- or poly-nucleotide repeat sequences. These were replaced by proxy variants with good correlation with the variant of interest, i.e. $r^2 > 0.9$ (except one proxy SNP had $r^2 = 0.72$). A proxy SNP is a variant that has high linkage disequilibrium (represented by r^2) with the variant of interest.

Test plates of DNA samples with known genotypes were sent to Affymetrix® for the assay to be tested. After running the assay on 2 sets of test plates, 155 SNPs were found to be working well. After further development by Affymetrix® with the Oncogenetics team, 162 of the originally submitted variants were able to be included in the final assay (Table 5). The data from the test samples genotyped in this process were referred to when carrying out the QC steps on genotyped study samples.

Table 5: 162 SNPs included in Affymetrix® Assay for ICR Oncogenetics Team

*MAF= minor allele frequency; **AA= African American SNPS

SNP ID	MAF (EUR)*		SNP ID	MAF (EUR)
rs636291	0.317		rs12155172	0.22
rs17599629	0.218		rs10486567	0.237
rs1218582	0.447		rs56232506	0.451
rs4245739	0.262		rs6465657	0.463
rs11902236	0.269		rs2928679	0.437
rs9287719	0.467		rs1512268	0.43
rs13385191	0.363		rs11135910	0.153
rs1465618	0.214		rs12543663	0.294
rs721048	0.182		rs10086908	0.303
rs10187424	0.426		rs16901979	0.032
rs12621278	0.059		rs620861	0.369
rs7584330	0.241		rs6983267	0.489
rs3771570	0.15		rs1447295	0.107
rs2660753	0.103		rs817826	0.145
rs2055109	0.236		rs17694493	0.136
rs7611694	0.421		rs1571801	0.268
rs10934853	0.277		rs76934034	0.083
rs6763931	0.442		rs10993994	0.383
rs10936632	0.493		rs3850699	0.3
rs10009409	0.311		rs2252004	0.102
rs1894292	0.485		rs4962416	0.267
rs12500426	0.463		rs7127900	0.199
rs17021918	0.349		rs1938781	0.23
rs7679673	0.408		rs7931342	0.496
rs2242652	0.206		rs11568818	0.45
rs2853676	0.262		rs11214775	0.292
rs13190087	0.049		rs80130819	0.092
rs12653946	0.425		rs10875943	0.287
rs2121875	0.33		rs902774	0.153
rs6869841	0.209		rs1270884	0.482
rs4713266	0.483		rs9600079	0.443
rs7767188	0.21		rs8008270	0.186
rs130067	0.202		rs7153648	0.082
rs3096702	0.377		rs7141529	0.499
rs3129859	0.33		rs8014671	0.42
rs1983891	0.277		rs684232	0.353
rs2273669	0.146		rs11649743	0.195
rs339331	0.305		rs4430796	0.475

rs1933488	0.421		rs11650494	0.078
rs9364554	0.283		rs7210100	0.001
rs1859962	0.481		rs138466039	0.009
rs7241993	0.305		rs878987	0.146
rs8102476	0.461		rs2066827	0.245
rs11672691	0.263		rs10845938	0.449
rs2735839	0.147		rs7968403	0.357
rs103294	0.219		rs5799921	0.318
rs12480328	0.072		rs7295014	0.352
rs2427345	0.379		rs1004030	0.416
rs6062509	0.302		rs11629412	0.422
rs1041449	0.433		rs4924487	0.187
rs2238776	0.198		rs33984059	0.022
rs9623117	0.197		rs201158093	0.31
rs5759167	0.498		rs28441558	0.055
rs2405942	0.217		rs2680708	0.394
rs5945619	0.364		rs8093601	0.442
rs2807031	0.182		rs28607662	0.097
rs5919432	0.199		rs12956892	0.302
rs34762946	0.159		rs10460109	0.421
rs4844289	0.384		rs11666569	0.287
rs56391074	0.379		rs118005503	0.089
rs62106670	0.379		rs11480453	0.398
rs74702681	0.022		rs6126982	0.495
rs11691517	0.259		rs9625483	0.029
rs34925593	0.481		rs17321482	0.133
rs59308963	0.287		rs9296068	0.349
rs1283104	0.379		rs11452686	0.442
rs142436749	0.012		rs1881502	0.19
rs76551843	0.009		rs61088131	0.177
rs4976790	0.113		rs1043608	0.291
rs12665339	0.167		rs377484932	0.475
rs17621345	0.259		rs188140481	0.006
rs1048169	0.379		rs183373024	0.007
rs10122495	0.31		rs138213197	0.002
rs1182	0.22		rs1487240	0.256
rs61830900	0.179		rs77541621	0.024
rs1935581	0.373		rs5013678	0.215
rs61890184	0.124		rs78511380	0.082
rs547171081	0.47		rs17464492	0.284
rs2277283	0.313		rs12549761	0.124
rs11290954	0.324		rs75823044	0.022 (AA)**
rs1800057	0.023		rs78554043	0.015 (AA) **

2.1.5 Genotyping and Polygenic Risk Scores

2.1.5.1 Quality Control of Genotyping Data

The total number of DNA samples submitted to Affymetrix® for genotyping was 302. QC procedures carried out by Affymetrix® led to 17 samples being excluded from the analysis. Failing QC may be due to low read count or abnormal allele read count ratio. The genotyping data received from Affymetrix® were reviewed using the Eureka Analysis Software (EAS). This allows manual review of the cluster plots for each SNP, so that low confidence genotype calls could be converted to 'no call' and uncalled genotypes that displayed sufficient confidence were converted to genotype calls.

After manual review using EAS, for the successfully genotyped 285 samples, data were reviewed by the Oncogenetics team with respect to call rate of each SNP across samples as well as total call rate of SNPs within each sample. SNPs with a call rate of <90% across all samples were excluded. Per sample call rate of >90% was used to exclude samples with a low call rate. These analysis steps were carried out using R as follows:

1. For each SNP, the genotype was converted to risk allele count using a key containing risk allele designations, risk allele effect estimates and risk allele frequencies.
2. After conversion, data was stored using the following format: "0" "1" "2" for Chromosomes 1-22, where '0'= no risk allele present, '1'= heterozygous for risk allele, '2'= homozygous for risk allele. For chromosome X, data was stored as: "0" or "1".
3. For SNPs with missing data, 2xRAF was used for Chromosomes 1-22 and 1xRAF for Chromosome X missing data.
4. The call rate for each individual SNP was checked to identify SNPs that were not called in more than 10% of samples. These SNPs were excluded (Figure 2).

5. Sample call rate was reviewed to identify samples with <90% call rate so that these were excluded.

6. MAF concordance was checked within the genotyping data and compared with the SNP MAF in the OncoArray meta-analysis [13] [13] as well as publically available databases such as gnomAD (the Genome Aggregation Database; https://gnomad.broadinstitute.org). Two SNPs (rs547171081 and rs10122495) were identified to have skewed MAF. These are likely to result from a modest number of non-called samples where the read counts lead to uncertainty over the true allele call. These assays did not work successfully during prior assay development and therefore no test plate data was available for concordance checking to verify assay performance. rs547171081 was already identified for exclusion in step 4 due to low call rate; rs10122495 was excluded as well.

7. RAF concordance was also checked against the RAF reported in the OncoArray meta-analysis.

Figure 2: BARCODE1 SNP Call Rate (upper plot) and Sample Call Rate (lower plot)

The red circle indicates rs547171081 with a call rate of 77% and therefore was excluded from PRS calculation. The sample call rate (lower plot) was >90% for all samples so none were excluded.
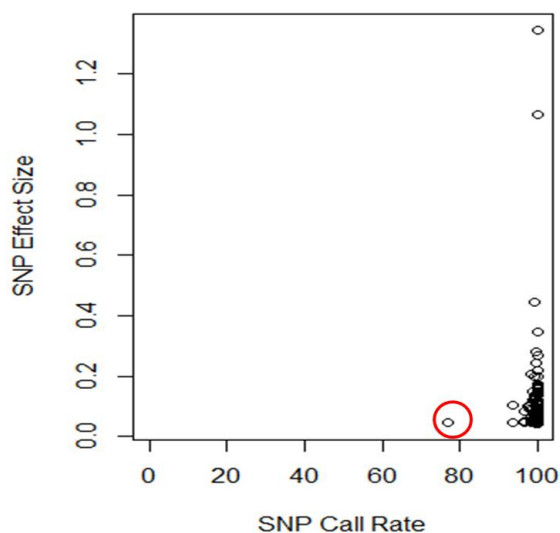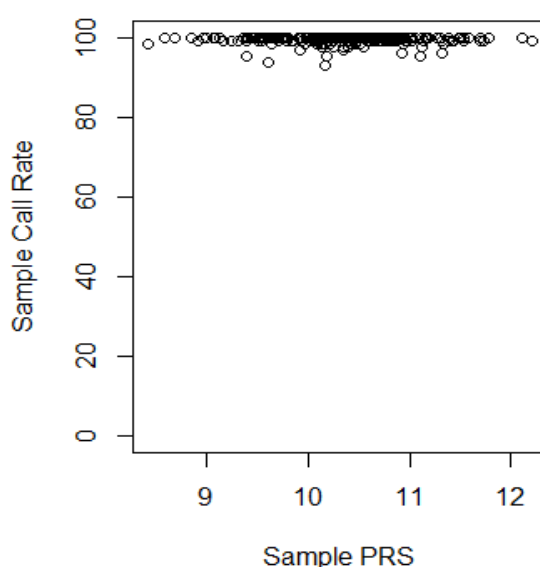
With the exclusion of the two SNPs described above, 130 SNPs were then used to calculate a PRS. After completion of the pilot study data analysis, an in-house application was developed by my colleague in the team that would automate the above steps (1-7). This was produced using the Shiny R package which allows users to create web based applications. This program was designed to be used after manual review of cluster plots using EAS. Genotyping data can be input to produce a prostate cancer

65

PRS for all samples. It also enables users to compare PRS distribution across different cohorts or populations.

Although I did not carry out the cluster plot review and QC steps for the BARCODE1 pilot data, I did carry out these steps for a second set of samples taken from the BARCODE2 trial. This is detailed later in Section 2.2.10.

2.1.5.2 Polygenic Risk Score Calculation

A polygenic risk score (PRS) was calculated for each study participant based on their genotyping data, using R software, utilising the following formula:

$$Score_j = \sum_{i=1}^{N} \beta_i g_{ij}$$

Where:

N : Number of SNPs included in the assay

g_{ij} : genotype at SNP locus i (0, 1, 2) for individual j. 0= homozygous for non-risk allele, 1=heterozygous for risk allele, 2=homozygous for risk allele

β_i : Per-allele log-odds ratio of SNP i

This formula produces the sum of weighted alleles for a set of SNPs for a single individual. When genotyping data were missing for a variant, 2x the risk allele frequency for that SNP was used. If the variant with missing data was a Chromosome X variant, then 1x the risk allele frequency was used.

The PRS mean and standard deviation were used to calculate the PRS at the 90th percentile; using this method to calculate the 90th percentile accounts for the potential large rise in PRS value at the extremes of the distribution in the small sample size. The formula to calculate the 90th percentile is as follows:

$$X = \mu + Z\sigma$$

Where:

X: the PRS at the 90th percentile
μ: mean PRS in the study population
Z: the Z score corresponding to the 90th percentile taken from a standard normal
distribution table; here Z=1.282
σ: the standard deviation in the study population

This was carried out for the BARCODE1 pilot cohort of men as well as the two reference

populations described in Section 2.1.7. This PRS threshold within the BARCODE1 pilot

cohort was used to identify men with a PRS in the top 10 percentile so that they could

be offered prostate cancer screening.

2.1.6 Prostate Cancer Screening Procedures

Men identified to be in the top 10% of the PRS distribution were invited to attend an

appointment at RMH to discuss undertaking prostate cancer screening with the study

team. Men who wished to proceed to screening confirmed their informed consent and

prostate imaging and biopsy were arranged. A blood sample was also taken at this point

to record the PSA level. Patients who declined screening came off study and no further

follow up was carried out.

2.1.6.1 DW-MRI of prostate

For patients undergoing screening, a diffusion weighted multi-parametric MRI scan

(DW-MRI) with intravenous gadolinium contrast was carried out. Scans were reported

by a specialist uro-radiologist and any prostate lesions identified were scored using the

PIRADS (Prostate Imaging Reporting and Data System) system and marked on the

scan for the urologist undertaking the biopsy. PIRADS is a structured reporting system

developed by the European Society of Urogenital Radiology to standardise the reporting

of prostate MRI and produces a score ranging from 1 to 5.[72] A score of 1 indicates

that clinically significant disease is highly unlikely to be present while a score of 5 indicates that clinically significant cancer is highly likely to be present.

2.1.6.2 Prostate Biopsy

An ultrasound guided biopsy of the prostate was carried out by the study urologist to obtain 12 cores of tissue. If a lesion was identified on DW-MRI, fusion images were used with ultrasound to target the lesion in addition to the standard biopsy cores.

Prostate biopsy samples were assessed by a RMH uro-pathologist and reported as per standard reporting procedures. Gleason score was recorded; this consists of two numbers, denoted as x+y, where x represents the predominant or primary cancer grade (range is 3-5) and y represents the secondary cancer grade. The Gleason score is denoted as x+y and the range of Gleason scores may be categorised into Grade groups as shown in Table 6. Complications following prostate biopsy such as infection were recorded.

Table 6: Prostate Cancer Grade Groups

Prostate Cancer	Gleason Score
1	6
2	3+4=7
3	4+3=7
4	8
5	9-10

2.1.6.3 Post-biopsy Follow Up

Patients identified to have prostate cancer were referred to the uro-oncology multi-disciplinary team (MDT) in the RMH and managed as per standard guidelines according to stage of disease.

Patients with a benign biopsy result are being followed up with annual PSA testing for 10 years. If the PSA rises to a level >3ng/ml (or by >50% if last PSA was >3ng/ml with a normal biopsy result), then a repeat DW-MRI and biopsy would be carried out, as per the study protocol.

Patients with findings of atypical small acinar proliferation (ASAP) or prostate intra-epithelial neoplasia (PIN) would undergo a repeat biopsy within 6 months and 12 months respectively (Figure 3). Both these findings on biopsy can be predictive of a future diagnosis of cancer on repeat biopsy. Although figures related to this vary between studies, for men found to have ASAP on biopsy ~40% go on to have a cancer diagnosed on repeat biopsy and for those with HGPIN 19-23% are diagnosed with cancer on a second biopsy. [73, 74].

Figure 3: BARCODE1 Study Outline

The pilot study recruited 300 men and the main study will recruit a further 4700 men
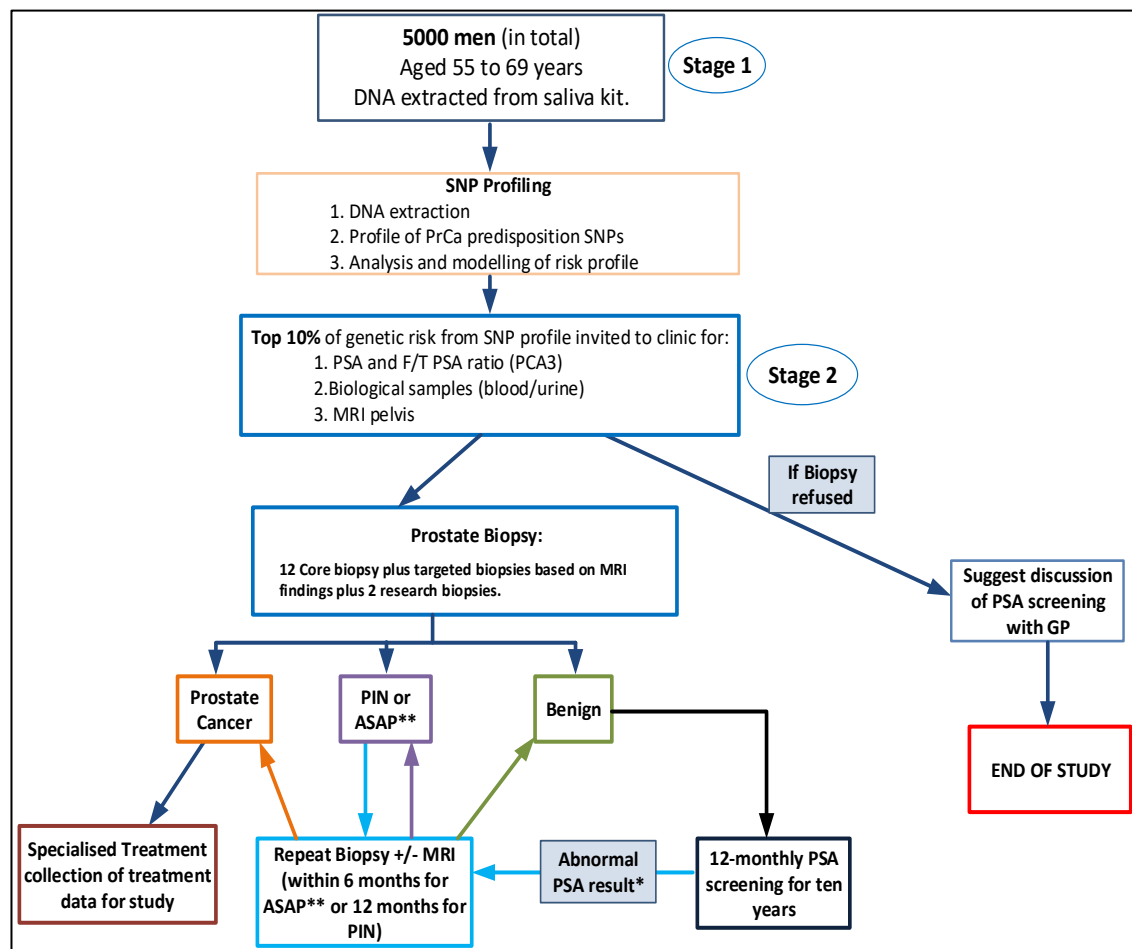
2.1.7 Statistical Analysis

I reviewed the pilot study uptake and recorded the fallout in uptake at each stage of recruitment. The PRS for each study participant was calculated as described above for the BARCODE1 pilot cohort. This was carried out using the R program by my colleagues in the team. The PRS distribution for the BARCODE1 pilot cohort was compared with two independent populations for whom genotyping data were available. The mean PRS and standard deviations were compared using ANOVA (analysis of variance) to assess how the BARCODE1 pilot cohort of men compares with other UK sets in the community. I carried this out using Graphpad Prism.

The first reference set was taken from the Prostate Testing for Cancer Treatment (ProtecT) trial where the genotyping data for the control cohort was used. ProtecT is a PSA screening study where participants were recruited from GP surgeries across the UK.[75] Although the study recruited men aged 50-69, I only used the data for men aged 55-69 to match the age range used in BARCODE1.

The second reference population was taken from the UK Genetic Prostate Cancer Study (UKGPCS). This is a large UK study (commenced in 1993) which recruits patients diagnosed with prostate cancer and collects patients' DNA samples as well as clinical data [76]; a sub-cohort of participants without prostate cancer were recruited via their GP surgeries for epidemiological sub-studies. These were coordinated by Professor Kenneth Muir at the University of Nottingham. I utilised the genotyping data for this subset of 500 men aged 55-69 years. All studies were approved by the appropriate ethics committees. All participants gave written informed consent.

2.2 Chapter 4 Methods: Germline NGS in the BARCODE2 Trial

2.2.1 BARCODE-2 Trial Set Up and Design

The BARCODE2 trial is a phase II trial funded by a grant from the European Research Council (ERC). It is a single centre study being run at RMH. The trial is enrolling men with metastatic castration resistant prostate cancer (mCRPC) and involves a germline genetic test using a next generation sequencing (NGS) panel of 115 genes, the majority of which are involved in DNA repair. Study participants found to have a variant in a DNA repair gene predicted to lead to protein truncation and meeting the study criteria for a Tier 1 variant (see section 2.2.7) are offered carboplatin chemotherapy after they have progressed on at least 2 standard lines of treatment which must include docetaxel and one of abiraterone or enzalutamide. I was involved in writing and finalising the study protocol for this trial. The approved research protocol is included in Appendix 2. The timeline taken for regulatory approvals is shown in Figure 4. These processes took approximately 18 months to complete.

Figure 4: BARCODE2 trial timeline of regulatory approvals

Sponsorship approval	25/08/2016
Main REC approval	22/11/2016
MHRA approval	17/03/2017
HRA approval	21/04/2017
Site initiation (The Royal Marsden Hospital)	23/05/2017
First participant recruited	25/05/2017

At the beginning of the study set up process, I attended the RMH Committee for Clinical Research (CCR) meeting to apply for study sponsorship which was followed at a later stage by the Research Ethics Committee (REC) meeting. I attended these meetings

with either the trial manager or Chief Investigator to answer the panel's questions related to the study, and act on changes or clarifications requested by the CCR and REC panels. Subsequent HRA (Health Regulatory Agency) and MHRA (Medicines and Healthcare products Regulatory Agency) approvals were sought and the study was fully approved and opened to recruitment on the 25th May 2017. The sponsor of the trial is The Institute of Cancer Research (ICR) with the trial oversight management being provided by the ICR-CTSU (ICR Clinical Trials and Statistics Unit) although the day to day management of the trial is undertaken by the ICR Oncogenetics team of which I was a part. I participated in regular trial progress meetings between the ICR Oncogenetics team and ICR-CTSU and was involved in the design of the case report forms (CRF) as well as discussions around data management and the set-up of the trial database.

The outline of the trial is shown in (Figure 5). The trial is divided into 2 parts:

- Part 1 involves a germline genetic test to identify carriers of protein truncating variants (PTVs) in a set of DNA repair genes included in the study NGS panel test.
- Part 2 involves treating men with a germline PTV who have progressed after at least two standard lines of treatment (docetaxel and one of abiraterone or enzalutamide) with 3-weekly carboplatin chemotherapy.

Figure 5: BARCODE2 Trial Outline

Figure 5: BARCODE2 Trial Outline

- **Unknown** carrier status of a germline DNA repair mutation
- Men diagnosed with mCRPC
- Performance status ≤ 2
- Current or previous treatment with docetaxel and/or abiraterone / enzalutamide

Clinical genetic test; DNA repair gene panel

Mutation present

Mutation absent

Progression after treatment with docetaxel and abiraterone / enzalutamide

- **Known** carrier of a germline DNA repair mutation
- Men diagnosed with mCRPC
- Performance status ≤ 2
- Previous treatment with docetaxel and abiraterone / enzalutamide

Pre-treatment assessments: CT and bone scan, PSA, inclusion/ exclusion criteria check, enrolment

Standard Care
Off study

Participants will be split into gene groups depending on the DNA repair gene mutated. See table 4 for recruitment targets for each gene group and trial stage.

Carboplatin AUC5 q21
Two cycles

Assessment of response:CT and bone scan(RECIST1.1 criteria)). PSA response (decline of >50%)

Disease progression

Measurable response or stable disease

End of study

Continue carboplatin 3 weekly, repeat imaging every 3 cycles

*Disease progression may be clinical or radiological.

Maximum 10 cycles carboplatin; Discontinue if there is progression of disease* or unresolving G3/4 toxicity

End of study

Part 1
Genetic Profiling

Part 2
Enrollment

2.2.2 Patient Selection

Patients entering part 1 of the BARCODE2 trial were required to have a diagnosis of mCRPC and have had treatment or be currently treated with at least one of the following standard lines of treatment:

- Docetaxel chemotherapy
- Abiraterone
- Enzalutamide

The full eligibility criteria are outlined in the trial protocol (Appendix 2). I, as well as clinical colleagues at the RMH, recruited patients to the study through the weekly uro-oncology clinics. Additionally, patients referred by external participant identification centres (PICs) for trial entry were recruited by me in the prostate cancer genetics research clinic.

Patients entering the study signed a BARCODE-2 Part 1 informed consent form and provided a single blood sample for DNA extraction and sequencing. Samples were obtained in a 9ml EDTA tube at the RMH, and transferred to the Cancer Genetics laboratory in ICR, Sutton. Blood samples were frozen at -80°C until ready for DNA extraction.

2.2.3 Gene Panel Design

Genes were selected for the study specific panel test based on recently published data showing the association of mCRPC with germline mutations in DNA repair genes. [30] Prof Ros Eeles and Dr Zsofia Kote-Jarai oversaw the design of the study panel test. Additionally, some genes were included as a result of our own team's sequencing studies [45, 77].

A subset of 40 genes from the 60 gene BROCA cancer risk panel (Version 6, 02/01/2015 through 07/01/2016;

http://web.labmed.washington.edu/tests/genetics/BROCA_VERSIONS) were included based on an in-house study which compared germline whole exome sequencing (WES) of aggressive prostate cancer cases (cases with metastatic disease and age of diagnosis under 60 years) with non-aggressive cases (low grade disease diagnosed at age over 60 years). [77] This study found that Tier 1 PTVs in the BROCA genes were enriched in the aggressive cases, with a high frequency of *BRCA2* and *NBN* mutations among others. Even when excluding *BRCA2* from this analysis, PTVs were significantly more frequent in the aggressive cases.[77] We also included all other DNA repair genes (outside the BROCA set) found in this study to have a Tier 1 variant in prostate cancer cases.

The remainder of the genes included were selected based on the results of a case control study published by our team in 2019. [45] In this study, 1285 cases of young onset prostate cancer (≤65 years) and 1163 age matched controls underwent germline sequencing of 175 genes (107 in the DNA repair pathway, 60 in DNA damage response pathways and cell cycle regulation, and 8 candidate genes). Within cases, aggressive (i.e. Gleason score ≥8, N=204) and non-aggressive cases (Gleason score ≤7, N=1,049) were also compared. As a result of the initial analysis (carried out in 2016) in this study, I included 48 genes that were found to have a Tier 1 variant in prostate cancer cases. Some of the genes selected from the results of these projects are known to sensitise to PARP inhibitors (PARPi) when mutated [78], which provides further rationale for their inclusion due to the known cross-sensitivity between PARPi and platinum. Table 7 outlines the BARCODE2 trial gene panel and highlights the genes for which there are evidence for sensitivity to PARPi. Four candidate genes were also included in this gene panel although they are not strictly DNA repair genes- these were included as part of the team's research and would not be actionable within the trial for patient treatment with carboplatin.

For targeted exon capture, an Agilent Technologies (Santa Clara, CA, US) custom capture library was designed. RNA sequences (baits) were designed to target the exons of 115 genes. The capture baits were designed to include 50 base pairs (bp) either side of each exon to allow the sequencing of splice regions. This design process was carried out by myself and my colleague Ed Saunders, using Agilent's SureDesign online program. The final set of target regions submitted to Agilent Technologies consisted of 1,830 target regions totalling 512,340bp. The capture design was a SureSelect QXT Custom 0.5-2.9Mb bait capture library; with 5x tiling (each of the 120bp bait overlaps neighbouring baits by 96bp). The resultant capture library contained 31,897 unique oligonucleotide sequences covering 594,726bp. Appendix 3 shows the expected coverage for this gene panel design.

Table 7: The BARCODE2 trial gene panel

115 genes involved in DNA repair unless otherwise stated

Case-control data (48)		Additional genes (31)	BROCA	Other genes (36)	
ALKBH3	MSR1	ATR^P	RINT1	PMS1^{DR}	RAD50^A
APEX1	MSH5	BAP1^P	SLX4	MLH3^{DR}	GTF2H2^A
ATM*^{P}	NABP2	BARD1^P	SMAD4	POLQ^A	POLK^A
BLM	NBN*^{P}	BRIP1	SMARCA4	TOP2A^{DR}	CCNH^A
BRCA1*^{P}	NEIL1	CDH1	STK11	TOP2B^P	WRN^A
BRCA2*^{P}	NEIL2	CDKN2A	TP53	TOP3A^P	RAD54B^A
CDC25C	NTHL1	CHEK1	XRCC2^P	XRCC4^A	XPA^A
CDK4*	PALB2*^{p}	FAM175A^P		XRCC5^{DR}	ERCC5^A
CHEK2*^{P}	PARP2	GEN1		DCLRE1A^A	
EME1	PER1	HOXB13		MMS19^A	
EME2	PNKP	MLH1^P		TDG^A	
ERCC2	POLD1*	MSH2		FANCM^A	
ERCC6	POLM^P	MSH6		MNAT1^A	
ESR2	RAD1	MUTYH		MPG^A	ANO7^S
FANCA^P	RAD52^P	PMS2		RPA1^A	AR^S
FANCD2^P	RAD54L^P	POLE		RECQL5^A	CHD1^S
FANCI	RECQL	POT1		LIG1^A	SPOP^S
GADD45A^P	RECQL4	PRSS1		XAB2^A	
GTF2H3^P	RNASEL	PTCH1		CLK2^A	
GTF2H4	SETMAR	PTEN^P		EXO1^A	
HUS1^P	SMUG1	RAD51B		FANCL^A	
LIG3^P	TP53BP1	RAD51C		ATRIP^A	
LIG4	XPC	RAD51D		OGG1^A	
MRE11A*^{P}	XRCC1	RB1		POLN^A	

*Genes which form part of the BROCA panel

^{A}Genes with Tier 1 mutations detected in prostate cancer cases in our team's WES study [77]

^{DR}Selected due to involvement in DNA repair

^{P} Evidence of PARPi sensitivity either as sensitivity hit in genome-wide olaparib shRNA sensitivity screen [78] or other published data.[79]

^{S} Candidate genes included due to data related to somatic mutations in prostate cancer and included to investigate variation in the germline.

2.2.4 DNA extraction

DNA extraction from whole blood was carried out by me using a protocol utilising a solution based method using a sucrose lysis buffer (SLB). The following steps were carried out over 4 days to extract DNA:

1. Blood samples were added to 45ml distilled water (4°C) and spun (1700 rpm for 25min at 4°C) to separate blood cells from plasma.
2. The resultant cell pellet was re-suspended in 35ml SLB to lyse cells and break down cell and nuclear membranes, and then spun for 15min at 1500rpm at 4°C.
3. Step 2 was repeated with 20ml of SLB.
4. The resulting pellet was re-suspended in 3.5ml of a re-suspension buffer containing NaCl, EDTA (Ethylene diamine tetra acetic acid), SDS (sodium dodecyl sulphate is a detergent that solubilises cell membranes) and protein kinase (cleaves glycoproteins and inactivates RNases and DNases), and incubated at 37°C overnight.
5. On Day 2, 5M NaCl was added and the tube inverted 3 times. The cations in NaCl counteract repulsion caused by the negative charge of the DNA phosphate backbone. Adding 100% ethanol to this DNA salt mixture leads to precipitation of the nucleic acids. This solution was stored for at least 3 hours at -20°C to allow DNA precipitation.
6. Following storage at -20°C, the solution was spun (3100 rpm for 20min at room temperature) to form a DNA pellet which was then washed with 70% ethanol (to remove excess salt; 3100rpm for 10 min at room temperature).
7. The 70% ethanol was decanted and the DNA pellet allowed to dry for at least 3 hours. The DNA pellet was then re-suspended in 750μl of TE (10mM Tris, 1mM EDTA; pH7.5) and allowed to dissolve over 48 hours.
8. The resultant DNA was quantified using a Qubit fluorometer and stored at -20°C until ready for library preparation and sequencing.

2.2.5 DNA Library preparation and targeted capture

This step was carried out using the Agilent SureSelect QXT library preparation kit and reagents. The Agilent 'SureSelect QXT Target Enrichment for Illumina Multiplexed Sequencing' protocol (Version D0, November 2015) was followed [80], but in summary the 2 day protocol included the following steps:

1. DNA extracted from blood was diluted using nuclease free water to a target concentration of 25ng/μl. After a two-step dilution, DNA samples were quantified using the Qubit High Sensitivity assay used as per the manufacturer's protocol.

2. A single step was carried out to achieve DNA fragmentation using a transposase enzyme (10 minute incubation at 45°C). In this step, DNA is simultaneously cleaved by the enzyme and adapters are ligated to the DNA fragments. Because DNA fragmentation and adapter ligation occur simultaneously, there is no need for DNA end repair or adapter ligation preparation. This was followed by PCR amplification of adapter ligated DNA.

3. DNA purification using AMPure XP beads was performed before and after PCR amplification. This involves the magnetic beads binding to the DNA in solution and then washing the DNA bound to the beads with 70% ethanol before eluting the DNA from the beads in nuclease free water.

4. The Agilent 2100 Bioanalyzer and DNA 1000 Assay were used as per manufacturer's protocol, to assess the quality of DNA fragments in each sample. Target fragment size was 245-325bp.

5. DNA was also quantified using the Qubit Broad Range assay. The amount of DNA needed to proceed was 500-750ng. If samples had a lower concentration than 500ng or fragment sizes outside the target range, then steps 1 and 2 were repeated with a new aliquot of DNA for the relevant samples before continuing with target capture.

6. DNA fragments were hybridised to biotinylated RNA baits (these are RNA sequences that correspond to the target regions on the genes of interest) and fragments containing regions of interest were captured in solution using streptavidin-coated beads. Streptavidin binds tightly to biotin on the baits bound to DNA targets and allows capture from solution. Captured DNA is retained on the streptavidin beads for the post-capture PCR amplification.

7. Captured DNA enriched for the regions of interest was then PCR amplified. During this step unique pairs of dual indexing primers were ligated to each sample to allow later pooling of samples (multiplexing). Each primer contains a unique index or barcode sequence which allows sample identification after sequencing.

8. At the end of PCR amplification, the streptavidin-coated beads were removed by placing the samples on a magnet and removing the supernatant and discarding the beads. Purification of the DNA libraries was carried out using AMPure XP beads as done in step 3.

9. The final captured DNA solution was assessed using the Agilent 2100 Bioanalyzer and High Sensitivity DNA assay. Target fragment size was 325-450bp.

10. For accurate quantification of DNA libraries at the end of targeted capture, I carried out a quantitative PCR (qPCR) using the KAPA Library Quantification Kit on a 7900HT Sequence Detection System machine (Applied Biosystems). In June 2018, the machine used for qPCR was changed to a Mic qPCR Cycler (Bio Molecular Systems). These qPCR results were used for calculating the volume of each DNA library required when forming the DNA pool (2nM) for sequencing on the MiSeq machine.

During the lab work for the BARCODE2 trial, the library preparation reagent kit was switched from the SureSelect QXT to the SureSelect XT HS kit due to issues with the DNA fragmentation step (described in Chapter 4). The XT HS protocol (SureSelect ^{XT} ^{HS} Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library, Version C1, Oct 2018) [81] utilises an endonuclease enzyme for DNA fragmentation rather than a transposase enzyme. Transposases fragment DNA by cleaving and inserting a short double-stranded oligonucleotide to the ends of the newly cleaved molecule. The inserted oligonucleotide must contain a sequence that is specific to the particular transposase being used. An endonuclease enzyme cleaves double stranded DNA generating random DNA fragments which require end repair and A-tailing (described below). The XT HS protocol also takes two days to complete with day 2 steps being very similar to those in the QXT protocol. The steps that differ on Day 1 of the XT HS protocol are as follows:

1. The quantity of DNA input for this protocol can range from 10-200ng with the recommendation of using the maximum amount possible. DNA samples were diluted from the high concentration extracted DNA (for germline DNA from blood) to achieve an input of 200ng in a volume of 7μl.

2. DNA was fragmented using the SureSelect fragmentation enzyme (endonuclease) as per protocol. After fragmentation, 40μl of nuclease free water was added to each sample to bring the total volume to 50μl. Unlike transposase based DNA shearing, this method requires sheared DNA to undergo end repair and dA-tailing.

3. DNA end-repair and dA-tailing was carried out using the end-repair dA-tailing buffer and enzyme mix. DNA end-repair describes the process whereby DNA fragments with strand overhangs are converted to fragments with blunt ends containing both 5′ phosphate and 3′ hydroxyl groups. dA-tailing incorporates a deoxyadenosine 5'-monophosphate (dAMP) onto the 3' end of blunted DNA fragments. This prevents concatemer (long DNA molecules containing multiple copies of the same sequence) formation during the downstream ligation step and enables DNA fragments to be ligated to adaptors with complementary dT-overhangs.

4. Molecular barcoded adaptors are ligated to the DNA. A molecular barcode is a unique oligonucleotide sequence that is incorporated into each library DNA fragment. The use of molecular barcodes allows low frequency variants to be identified during analysis after sequencing- this is relevant to tumour DNA sequencing rather than germline sequencing.

5. Purification using AMPure XP beads is carried out prior to PCR amplification of DNA. During this step, each sample has a unique index added to it which will allow later pooling of samples. In this protocol, single indexing is used rather than dual indexing as with the QXT protocol.

6. After another purification step using AMPure XP beads, the DNA samples are assessed using the DNA1000 assay on the BioAnalyzer to determine fragment size. The desired fragment size is 300-400bp. Samples are also quantified on the Qubit fluorometer using the Broad Range assay.

7. The second day of the protocol involves DNA hybridisation and capture using the study specific RNA baits as described in the QXT protocol. The same post capture amplification PCR is carried out with an assessment of DNA library fragment size using the BioAnalyzer and High Sensitivity assay at the end of these steps.

2.2.6 Next Generation Sequencing on a MiSeq Machine

The Illumina protocol for 'Preparation of Libraries for Sequencing on the MiSeq' was followed. A 2nM pool of DNA libraries was formed by combining the appropriate volume of each DNA library (according to its quantification by qPCR). NaOH (0.1M) was used to denature DNA. This was followed by dilution using HT1 (Illumina supplied hybridization buffer) to achieve a loading concentration of 13-16pM. The loading concentration was adjusted during the project depending on the clustering output of serial runs on the MiSeq. For the NGS runs carried out using the Agilent QXT protocol, custom primers were combined with Illumina primers as outlined in the SureSelect QXT protocol. [80]

DNA libraries were clustered and sequenced on an Illumina MiSeq machine using the MiSeq Reagent Kit v3 (150 cycles), generating 2x75bp paired end reads.

2.2.7 NGS Data Analysis

NGS data were demultiplexed and FASTQ files generated by the Mi-Seq software. Demultiplexing refers to the identification of the index on each read which in turn identifies which DNA sample the read relates to. FASTQ files were then processed and analysed using the SureCall (Version 3.5, Agilent Technologies) software by me. This is an integrated package that carries out adaptor trimming, alignment of reads to the reference genome (GRCh37 release, hg19, Feb 2009), variant calling of sequencing data, variant annotation (using the software's SNPPET Caller) and categorisation according to variant impact (Figure 6, Table 1, Table 1). QC data are also output for each DNA sample including percentage of reads covering target regions, average and median read depth for each DNA sample and coverage data. The percentage of reads covering target regions reflects the adequacy of the hybridisation of RNA baits to target regions. For QC purposes, my aim was to achieve a coverage of 20x or higher for at least 80% of target bases. Samples that didn't reach this threshold were re-sequenced

using a fresh aliquot of DNA. The SureCall program was later upgraded to Version 4.0

in April 2019 as this later version was updated to deal with the analysis of samples

prepared using the XT HS protocol integrating molecular barcodes.

The Variant Call File (VCF) generated by the SureCall software is presented as a table

of variants for each DNA sample which can be manually reviewed and filtered using

various parameters to identify non-sense and frameshift mutations for manual review

and cross-checking in clinical and sequencing databases (e.g. gnomAD, ClinVar). The

SureCall program also outputs clinical classification of variants where available (from

ClinVar/ dbSNP), so variants annotated as 'pathogenic' or 'likely pathogenic' were also

filtered for manual review. NGS reads were reviewed in the software's integrated

genome viewer (IGV). Figure 7 shows an example of a table of variants in SureCall

filtered to show a pathogenic variant in *MRE11A*.

Figure 6: SureCall program FASTQ processing steps

Left sided flowchart reproduced from SureCall documentation
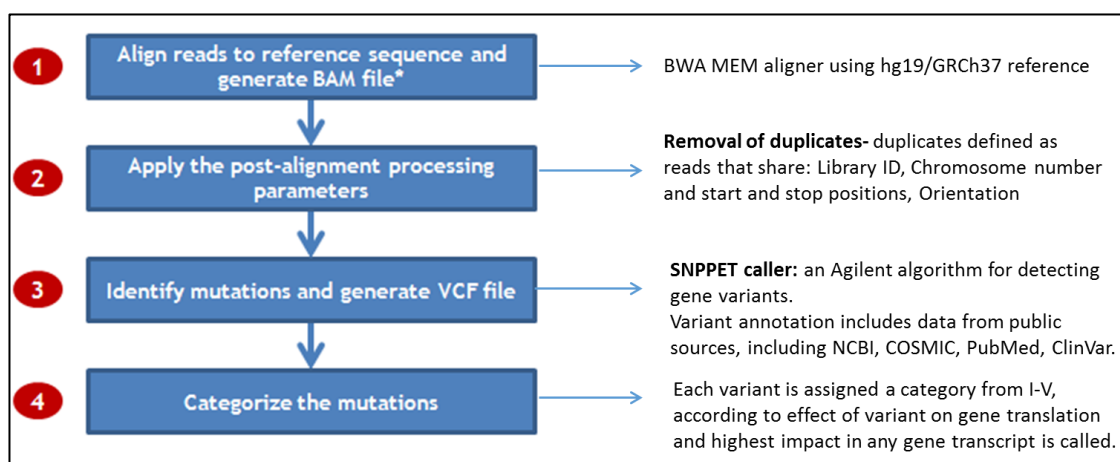
Figure 7: SureCall table of variants for a sample containing a pathogenic variant in *MRE11A*

Only part of the fields in the table of variants is shown. Although it shows a pathogenic category 1 variant in *MRE11A*, the categories were reformatted after the pilot experiments to denote pathogenic variant as category 5.
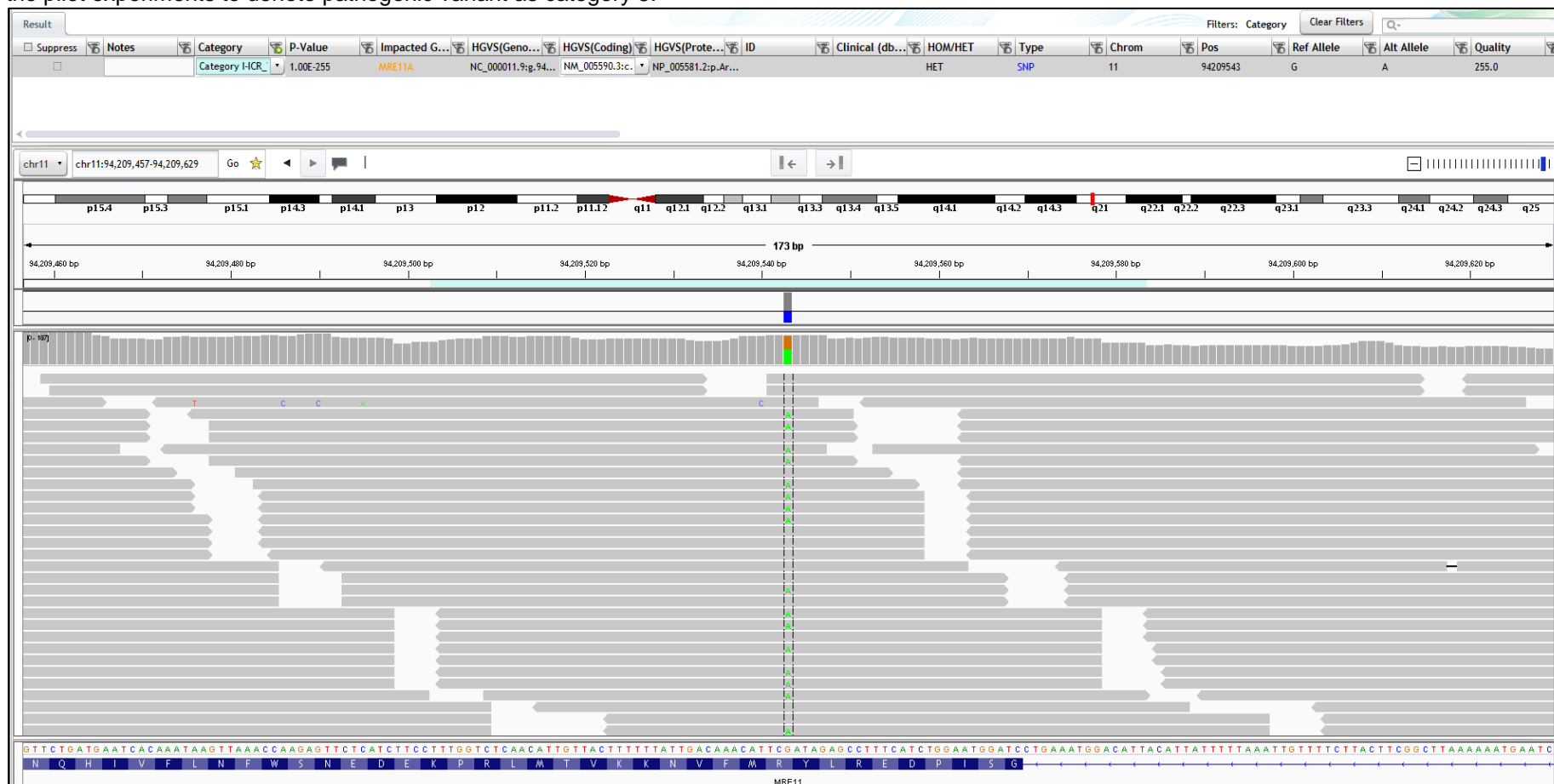
Table 8: SureCall user specified settings for variant filtering

Parameter	Threshold set (range of parameter)	Comment
Variant score threshold	0.3 (0.001-1.0)	The minimum variant score that a variant call must have in order to be included in the results. The variant score is based on the Phred scaled quality value which has a range between 0 and 50. A value of 50 means that the variant call has a 99.999% chance of being accurate. The default is 0.3, which means that the accuracy of the call is 99%.
Minimum quality for a base	30 (0-60)	The minimum sequencing base quality that a base must have in order to be called as a variant within a read. The default value is 30, which helps ensure that low-quality bases are not falsely called as variants.
Variant call quality threshold	100 (0-255)	The minimum Phred quality score that a candidate low frequency variant needs to have in order to be reported in the results. This score is a reflection of how well the read pileup supports the call. In the equation below, Q is the Phred quality score and P is the probability of a base-calling error. $$Q = -10 \times \log_{10}P$$ The default value is 100, which means that candidate variants with a quality score <100 are filtered from the results. A Phred quality score of 100 corresponds to a 1×10^{-10} probability of an incorrect base call.
Minimum allele frequency	0.3 (0.001-1.0)	The minimum allele frequency that a potential variant call must have in order

		for the program to call it as a variant. The default is 0.1 but I increased this to 0.3 for the germline analysis as allele frequency is expected to be ~0.5 for heterozygous variants. This was reduced for tumour DNA sequencing analysis (Chapter 5) to 0.01.
Minimum number of reads supporting a variant allele	10 (≥1)	This parameter sets the minimum number of reads that supports the variant allele sequence and also passes the quality filters. This means that the sequencing data for the sample must have at least 10 reads that pass the quality filters and support the variant allele sequence. This was reduced to 5 reads for tumour DNA sequencing analysis (Chapter 5).
Estimated index hopping frequency	0.005 (0.0-0.1)	The estimated frequency of index hopping in sequencing samples. A value of 0.1 indicates an estimated index hopping frequency of 10%. When a non-zero value is entered in this parameter, the SNPPET SNP caller filters out putative calls that are as likely or more likely due to index hopping contamination by another sample than due to a genuine low-frequency mutation. The default value for the parameter is 0.005 (i.e., 0.5%).

Table 9: SureCall criteria for variant categorisation

Category 5 is expected to represent variants that are pathogenic and category 1 represents variants that are benign; manual re-categorisation was carried out where appropriate during analysis (continued on next page to show categories 1 and 2).

Category 5		Category 4		Category 3	
Known in literature to be clinically significant and causative	Nonsense or a frame shift mutation	In-frame exon deletion	Alters the sequence at a splice junction	Is splice site acceptor or donor	Is non-synonymous coding variant in stop
Evidence for pathogenicity in locus specific databases as being causative	Located within a splice consensus sequence	Mutates the initiation codon (ATG)	Annotated drug response in NCBI SNP database	Likely to affect transcription	Sequence changes seen multiple times in
Is associated with a tumour site in COSMIC	Alters the sequence at a splice junction	Missense mutation of the normal stop codon	Annotated histocompatibility in NCBI SNP database	Likely to produce a cryptic splice site	Is labelled suspect in NCBI SNP
Is validated in clinical study in NCBI SNP	Located within a splice consensus sequence	Annotated probable pathogenic in NCBI SNP database	Annotated unknown or untested in NCBI SNP database	Modifies UTR 3' or UTR 5'	Deletes UTR 3' or UTR 5'
Introduction of a stop codon		Deletes exon which results in shift of reading frame	Any missense mutation	Results in codon change	Is non-synonymous coding variant
Deletes nucleotide(s) that lead(s) to a shift of reading frame		Is non-synonymous coding variant	Do not produce an amino acid substitution and that are unlikely to produce a cryptic splice site	Results in codon change and codon deletion or codon insertion	Is non-synonymous coding variant in start
Annotated pathogenic in NCBI SNP database			In-frame amino acid insertion/deletion	Sequence changes that occur in the intron	

Category 2	Category 1
Is synonymous coding variant in stop	Allele origin is somatic
Is synonymous coding variant in start	Allele origin is germline
Is synonymous coding variant	Allele origin is unknown
Is intergenic	Annotated nonpathogenic in NCBI SNP database
Is intronic variant	
Unlikely to produce a cryptic splice site	
Annotated probable nonpathogenic in NCBI SNP database	

Steps taken in the NGS analysis using SureCall can be summarised as follows:

1. FASTQ files were loaded to SureCall for each DNA sample. Two FASTQs were generated for each sample on the MiSeq, one with forward reads and one with reverse reads. These paired FASTQs were run through the 'single sample analysis' function in SureCall.

2. 'Single sample analysis' aligns reads to the reference genome (GRCh37 release, hg19, Feb 2009), creates a BAM (Binary Alignment Map) file and then a VCF which includes variant annotation.

3. Once processed by SureCall, each VCF is displayed as a table of variants within the program. The table of variants was reviewed for each DNA sample. The BARCODE2 'rule set' (described later) was applied at the start of analysis to reassign any gene transcripts or categorisation that had already been previously corrected in other analyses.

4. Using the filters in SureCall, the table of variants was filtered across several parameters to identify PTVs:

    a. The category column was used to filter Category 5 variants for manual assessment of the software's variant annotation. Frameshift and non-sense variants as well as known pathogenic missense variants were expected to be called as category 5, but from the pilot experiments I was aware that some non-PTVs were also assigned a high category by the software. If the category 5 variants filtered appeared to be a true PTV then the population allele frequency in gnomAD was reviewed in addition to clinical databases such as ClinVar (and Breast Information Core, BIC, database if applicable). For most variants, the rsID was also output by SureCall and was used for this cross referencing. If this cross checking revealed a variant was known to be benign or was not clinically reported but had a population mean allele frequency >1%, then the category was manually changed to a lesser category and this change added to the 'rule set' for future analyses in case the same variant appeared again in other samples.

    b. Using the 'primary effect' column which lists the variant effect on DNA sequence, stop gain (non-sense) and frameshift variants were filtered and reviewed manually in case they had not been identified using the

category 5 filter. These variants were reviewed in the same manner as in the previous step and categorised accordingly.

    c. Additionally, SureCall reports the clinical classification of variants listed in dbSNP, if available, under the heading 'Clinical dbSNP'. So pathogenic and likely pathogenic variants were also filtered using this SureCall parameter and reviewed manually.

5. If the rsID was not available for any of the variants identified for review, the variant genomic coordinates were used in the University of California Santa Cruz (UCSC) genome browser (https://genome.ucsc.edu) to see whether the variant had an rsID listed so that the variant could be looked up in databases such as gnomAD and ClinVar.

6. Variants identified to be pathogenic in ClinVar or other databases such as BIC (Breast cancer Information Core that is now no longer active) were listed for further assessment and validation. Variants that were not previously clinically reported in these databases but had a low allele frequency (<0.5-1.0% in all populations) were also listed for further study.

All PTVs identified in the steps above were assessed for read depth at the variant position. Allele frequency was reviewed to determine zygosity with the expectation that significant PTVs would be heterozygous. Forward and reverse read balance was assessed to exclude any strand bias that may suggest a variant was in fact an artefact. The reads were visually assessed using the IGV in SureCall.

Additionally, the VCF files generated in SureCall were run through the CADD (Combined annotation dependent depletion) tool which is available online to generate a CADD score for each variant (http://cadd.gs.washington.edu/). This is a tool for scoring the deleteriousness of single nucleotide variants (SNV) and indel variants in the human genome. The scores of PTVs were reviewed. A scaled CADD score of 20 indicates that a variant is amongst the top 1% of deleterious variants in the human genome, and a score of 30 indicates a variant is in the top 0.1%. As part of the CADD tool output, other parameters are also reported such as the variant consequence (e.g. frameshift, splice site); canonical splice site (these are the conserved GT-AG

dinucleotides flanking gene exons) variants were identified and reviewed as pilot experiments (described later) had shown that the SureCall program was inaccurate in the categorisation of splice variants.

As well as the CADD score of filtered variants, clinical classification was reviewed (if available) in databases such as ClinVar, BIC, or locus specific mutation databases if available. For the purposes of my study and the BARCODE2 trial, I focussed only on identifying variants that were known to be protein truncating or predicted to be so, therefore benign, likely benign and variants of unknown significance were not reported. When deciding on the significance of a variant (not previously reported clinically), the allele frequency reported in the gnomAD database was taken into account. In keeping with the American College of Medical Genetics (ACMG) guidelines for interpreting PTVs, variants with an allele frequency over 1% were excluded, as were variants occurring in the final exon of a gene (unless known to be clinically pathogenic)[82].

Regular meetings with the trial Chief Investigator and the team's senior scientist were held to discuss the NGS results and to agree on those deemed to be significant (Tier 1) within the trial to carry through for validation by Sanger sequencing. The criteria used for shortlisting Tier 1 PTVs were as follows:

- Variant is a non-sense, frameshift or splice site loss variant predicted to lead to protein truncation.
- Variant allele frequency is <1% in population databases.
- CADD score over 20.
- Variant does not lie in the final exon or in the last 50bp of the penultimate exon of a gene.
- Variant previously reported to be pathogenic or likely pathogenic (this was not expected for all Tier 1 variants as some would not have been reported previously)

Two pilot experiments were carried out prior to the start of the clinical study, using DNA samples with known PTVs. NGS analysis using the SureCall software was compared with the GATK (Genome Analysis ToolKit, developed in the Data Sciences Platform at the Broad Institute; Version 3.5) pipeline for the identification of PTVs. GATK is a collection of command-line tools for analysing high-throughput sequencing data. Bioinformatics analysis was provided by my colleagues Ezequiel Anokian and Ian Whitmore.

2.2.8 Pilot Experiments and SureCall vs GATK Comparison

Training in the use of the SureSelect QXT Library Preparation kit was provided by the Agilent support team for me and the team's scientific officer prior to commencement of the project. For this, I used stored DNA samples, some of which were known to harbour pathogenic variants in one of the study panel genes. Two experiments were carried out using 8 DNA samples each. The first was carried out with the Agilent support team in November 2016 and the second was carried out by me and Ian Whitmore in January 2017, also using 8 DNA samples, 7 of which were known to have pathogenic variants.

In the first pilot experiment, 8 DNA samples were included, 3 of which were known to have a *BRCA1* or *BRCA2* variant (these were taken from the IMPACT study): *BRCA2* exon23 c.9026_9030delATCAT, *BRCA2* exon7 c.538_539delAT and *BRCA1* exon2 c.66_67delAG. In the second pilot experiment, 8 DNA samples from the UK Genetic Prostate Cancer Study (UKGPCS) were used, which were all known to carry one or more PTV in one of the genes of interest (one clinically pathogenic missense variant was also included):

- *MRE11A* c.C571T non-sense variant
- *ATM* c.8786+1 G>A splice variant
- *BRCA2* c.631+2 T>G splice variant
- *CHEK2* c.869delA frameshift variant

- *PALB2* c.712A>T non-sense variant
- *RAD52* c.1037C>A non-sense variant
- *MUTYH* c.536A>G missense variant
- *ATM* c.8385_8394delTTTCAGTGCC frameshift
- *NBN* c.127C>T non-sense variant
- *FANCI* c.2542 C>T non-sense variant

After successfully completing the MiSeq runs for these experiments, FASTQ files were processed by me using the Agilent SureCall software to generate filtered annotated variants for each DNA sample as described in the previous section. The FASTQ files were also processed by my Bioinformatics colleagues using the GATK pipeline (version 3.5).

The annotated variants output by both the GATK pipeline and the SureCall program included the known pathogenic variants of interest. Comparison of the overall SureCall output and GATK output showed good concordance (Table 10). The small variation in the number of variants called may be due to filters applied in the SureCall program such as minimum number of reads required for a variant (set to 10 in the pilot experiments). The genotypes of the shared variants in the first pilot set were 94% concordant and for second pilot set 98% concordant.

Table 10: Comparison of variants called by SureCall and by GATK pipeline

Pilot 1 Set				
Number of variants in SureCall VCFs	1482	Number of variants in GATK VCFs		1421
Number of SureCall variants not in GATK VCFs	112	Number of GATK variants not in SureCall VCFs		51
Number of variants in both SureCall and GATK VCFs	1370	Number of variants in both SureCall and GATK VCFs		1370
% shared variants	92	% shared variants		96

Pilot 2 Set				
Number of variants in SureCall VCF	1150	Number of variants in GATK VCF		1278
Number of SureCall variants not in GATK VCF	35	Number of GATK variants not in SureCall VCF		163
Number of variants in both SureCall and GATK VCFs	1115	Number of variants in both SureCall and GATK VCFs		1115
% shared variants	96	% shared variants		87

By carrying out these pilot experiments, SureCall analysis settings were decided for optimal calling of germline variants. Additionally, through the pilot analyses in SureCall, I created a set of 'rules' to be applied to my future analyses. This was needed as the gene transcript selected by the program for calling variants was not always the most commonly occurring transcript for variant calling. The SureCall program displays all gene transcripts for a variant and the expected 'effect' e.g. frameshift, stop gain, synonymous etc. The 'primary effect' is output in the results summary whereby the program selects the gene transcript that results in the highest impact effect on protein translation even if the transcript is known to be non-coding. By manually altering the gene transcript where necessary, variant categorisation was upgraded or downgraded. These types of manual changes were saved into a 'Transcript Reassignment and Variant Recategorisation' rule set. This rule set would be applied to each subsequent sample analysis and contains commands such as the selection of the correct gene transcript for some variants and re-categorisation of variants where they are being under or over called by the program. This rule set was updated as my analyses

continued over time so that if a variant that had been re-categorised previously was observed again, then the correct category or gene transcript would be used by the program. When modifying a gene transcript in this manner, I selected the longest transcript for the gene as recommended in ACMG guidelines. [82] Figure 8 shows the steps taken in my NGS analyses using the SureCall program.

Through the analysis I carried out on the pilot samples NGS data, I found that the SureCall program did not accurately annotate splice site variants. These were annotated as 'intron variants' by the program and designated a category of 3 or lower. Therefore, for my subsequent NGS analyses, I relied on the CADD tool output to identify canonical splice site variants that may be incorrectly categorised by SureCall. Canonical splice sites are the highly conserved GT-AG dinucleotides that flank gene exons. Loss of donor GT or acceptor AG sites leads to aberrant splicing of introns or exon skipping during translation, which can lead to aberrant gene function. Mutations at canonical splice sites identified during review of the CADD output were then cross checked again in the SureCall program to assess read depth, allele frequency and review reads in the IGV.

The comparison of SureCall analysis of NGS data and the GATK pipeline showed that the SureCall program could be relied upon for the purpose of identifying deleterious PTVs in the DNA samples in the BARCODE2 trial.

Figure 8: Outline of the steps in NGS analysis for each DNA sample sequenced in the BARCODE2 trial

FASTQ file loaded into SureCall for single sample analysis
VCF displayed as a table of variants including annotations
1. Transcript Reassignment and Variant Recategorisation rule set applied
2. Category 5 variants filtered for manual review
Manual review of category 5 variants
1. Reassign gene transcript if needed and/ or recategorise if appropriate
2. Review remaining category 5 variants for read depth and forward/reverse read ratio as well as checking reads in IGV
3. Cross check remaining category 5 variants in dbSNP/ ClinVar/ gnomAD to determine if clinical classification available and check population allele frequency.
4. Disregard variants with a population allele frequency ≥1%. (this applies to all variants checked in the analysis)
5. Cross check with Oncogenetics Team's NGS results to see if variants have been identified in previous NGS projects.
Use filters to assess other variants in table
1. Using 'Clinical dbSNP' column, filter for 'pathogenic' and 'likely pathogenic' variants. (These variants may not have been in the category 5 list)
2. Using the 'Primary Effect' column, filter for stop gain/loss and frameshift variants
3. Cross check in clinical databases and check CADD score
4. Steps 4 and 5 in above section repeated. Any new variants identified that are protein truncating are assigned a category 5 in SureCall.
Outside the SureCall program, run VCF in the online CADD tool
1. Review CADD output table of variants to assess canonical splice site variants that may not have been annotated correctly in SureCall
2. Review frameshift and stop gain/loss variants and record CADD score
3. Further cross checking of any additional PTVs identified here in dbSNP/ClinVar/ gnomAD etc.
4. Step 4 in previous section repeated.
After analysing all samples in each NGS run, shortlisted category 5 variants reviewed with trial team
Variants identified for validation by Sanger sequencing

Table 11: SureCall Annotation Fields/ Parameters

Parameter	Description
Category	SureCall assigned Category
HGVS(Coding)	Gene transcript
HGVS(Protein)	Protein transcript
ID	rsID if available
Clinical (dbSNP)	Clinical classification from dbSNP which is based on ClinVar records
HOM/HET	Zygosity of variant as assigned by program according to variant fraction
Type	Nucleotide change e.g. deletion, insertion
Allele Frequency	Variant frequency based on number of reads with variant
Quality	A Phred quality score of the variant
Number of Variant Alleles	Number of reads with variant (non-ref) allele
Filtered Read Depth (per sample)	Total read depth at variant position
Mapping Quality	Quality of read mapping to reference genome
Function Class	Variant class e.g. missense, non-sense
Effect	Lists all possible functional effects according to all RefSeq gene transcripts
Primary Effect	The highest impact effect based on transcript selected
Codon	Codon change
AA	Amino acid change
Transcript	Gene transcript selected to call variant
Exon ID	Exon number affected by variant
Forward Ref Alleles	Number of forward reads with Ref allele
Forward Non-ref Alleles	Number of forward reads with variant (non-ref) allele
Reverse Ref Alleles	Number of reverse reads with Ref allele
Reverse Non-ref Alleles	Number of reverse reads with variant (non-ref) allele
CosmicMutationDescription	COSMIC details if variant has a COSMIC record
CosmicAASyntax	
CosmicTumorSite	
CosmicFrequency	
COSMIC ID	

2.2.9 Sanger Sequencing for Validation of Genetic Variants

Genetic variants identified by NGS were validated by Sanger sequencing.[83] I designed the primers for sequencing using the online software, Primer3web version 4.1. Preparation of DNA samples was carried out by me using the protocol detailed below. For each variant that required validation, several samples were Sanger sequenced: DNA from the originally extracted blood sample (stock DNA), the diluted DNA sample that was used for NGS, one or more negative controls and if available, a positive control as well. Once samples were ready, they were loaded onto a 3730XL Sequencer for sequencing. I analysed the results using the Mutation Surveyor program.

2.2.9.1 Sanger Sequencing Protocol

1. In new tubes, dilute quantified DNA stocks to 5-20 ng/uL concentration (final volume 10-30µl)
2. Place 1.25µL of each sample in a well on a 96 well plate
3. Make the PCR master mix in an Eppendorf tube for the required number of samples (a separate mix is required for each set of primers):

Reagent	1X µl
Buffer	2.5
dH2O	17.375
dNTPs (2mM)	2.5
DNA Polymerase	0.125
Forward Primer (20µM)	0.625
Reverse Primer (20µM)	0.625
Master Mix TOTAL	23.75
DNA (5-20ng/µL)	1.25
TOTAL	25

4. Add 23.75 μl of the master mix to each one of the sample wells (changing tip each time). Each well will have a total volume of 25 μl (23.75 μl of Master Mix plus 1.25 μl of DNA).

5. Place the plate in the thermocycler block and select the program:

Step	Temp ºC	Time	Cycles
Initial denaturation	95	2 mins	1
Denaturation	95	30 secs	
Annealing	Tm*	30 secs	40
Extension	72	1 min	
Final Extension	72	10 mins	1

*Tm is the primer melting temperature appropriate for the individual set of primers used.

6. Run the PCR products in a 2% agarose gel to check for adequate amplification of the correct fragment size.

7. When finished, take the plate from the block and in a separate Eppendorf tube make next Master Mix for samples purification:

Reagent	1x μl
Sap I	1.25
Exo I	0.3125
dH2O	4.6875
TOTAL	**6.25**

8. Add 6.25 μl of the purification Master Mix to each one of the wells with the PCR product. The final volume will be 12.5μl PCR product + 6.25 μl purification Mix = 18.75 μl.

9. Run the plate in the thermocycler block selecting the program:

37°C	1 hour
80°C	15 min
4°C	10 min then hold

10. Make the Big dye Master Mix:

Reagent	1x μl
Big dye v3.1	0.5
Buffer	2
TOTAL	**2.5**

11. When complete, transfer, with a multichannel pipette, 3 μl of each sample to two wells per sample (3μl each).

12. Add 2.5μl of the Big dye Mix to each well plus 2.5μl (2μM) of just one primer (either forward or reverse primer). The final volume in each well will be: 3μl purified PCR product + 2.5μl of Big dye Master Mix + 2.5μl primer = 8μl.

13. Run the following program on the thermocycler block:

Temperature °C	Time	Cycles
96	2 min	
96	30 sec	25 cycles
50	5 sec	
60	4 min	

14. Sequencing clean-up:
    a) Prepare NaAc/ETOH/EDTA mix (for 1 plate: 500μl 3MNaAc + 10mL 100% ETOH + 7.5 μl 0.5M EDTA).
    b) Add 40μl of the mix to each well.
    c) Centrifuge the plate at 3000 xg for 45 minutes at 4°C
    d) Carefully discard the supernatant onto tissue without dislodging DNA pellet. Pulse spin upside down (≤ 500rpm)
    e) Add 40μl of 70% ETOH per well. Centrifuge at 3000xg for 5 minutes at 4°C
    f) Carefully discard the supernatant. Pulse spin upside down. Repeat 70% ETOH wash procedure
    g) Evaporate residual ETOH at 90°C for 2 minutes using a thermocycler.

Validated variants were reported to the trial patients during their clinical follow up at the Royal Marsden Hospital. For all patients who underwent germline sequencing, a study specific genetic result report was produced for each patient undergoing genetic testing in the trial, detailing the variant(s) identified (for men with a positive test), as well as a recommendation for clinical validation and genetics referral if appropriate (e.g. for *BRCA1/2* variant carriers). This report was added to the patient's electronic patient record as well as to their trial file. Patients with no significant variants identified were notified by letter.

2.2.10   Genotyping for Prostate Cancer PRS

For the first 100 patients recruited to the BARCODE2 trial, germline DNA was sent to Eureka Genomics for genotyping utilising the same assay used in the BARCODE1 study.

Genotyping results were received from Eureka Genomics and SNP cluster plots were reviewed by me using EAS. Low confidence genotype calls were excluded by converting them to 'no call' and uncalled genotypes that displayed sufficient confidence were converted to the appropriate genotype call.

Once cluster plot review was completed, I used the ICR Oncogenetics in-house PRS application to carry out the QC based on SNP call rate and sample call rate as described in Section 2.1.5. A PRS was generated for 98 samples (2 samples failed QC at Eureka Genomics and did not have genotyping data produced) and using the application I was able to compare the PRS distribution to that seen in the BARCODE1 pilot as well as to the two reference populations described in section 2.1.7.

2.3 Chapter 5 Methods: NGS of FFPE Derived Somatic DNA for PTV Carriers in BARCODE2

Patients identified to carry a germline PTV in a DNA repair gene were informed of their result during a clinical consultation. If clinically relevant outside of the study, e.g. a PTV identified in *BRCA1* or *BRCA2,* a clinical validation genetic test was arranged for the patient using a new blood sample (collected under clinical conditions) as well as subsequent referral to clinical genetics for counselling and cascade testing of relatives.

For patients identified to carry a germline PTV, carboplatin treatment was offered if they met the trial eligibility criteria which included disease progression after both docetaxel chemotherapy and one of Abiraterone or Enzalutamide. If a carrier did not yet meet these criteria then they were followed up regularly by the trial team until they were eligible for treatment and offered Part 2 of the trial at that point.

Treatment within Part 2 of the trial is ongoing and clinical responses to treatment do no form part of this thesis.

2.3.1 Tumour DNA Sequencing

For trial participants who were identified to carry a germline PTV, archival tumour blocks were requested so that targeted sequencing could be carried out using the same gene panel that was used for germline sequencing. These formalin fixed paraffin embedded (FFPE) tumour tissues included diagnostic biopsies and for some participants, biopsies of metastatic lesions.

DNA extraction from FFPE tumour tissue was carried out in our lab using the QIAamp DNA FFPE Tissue Kit (Qiagen). Extracted DNA was then assessed and quantified using the Nanodrop spectrophotometer and Qubit fluorimeter.

QC assessment of tumour DNA integrity was done by me using the Agilent Tape Station and Genomic DNA Screen Tape assay. This is a quantitative electrophoretic assay that allows determination of DNA integrity and produces a DIN (DNA Integrity Number) score for each sample. This is used to determine the amount of DNA input needed for library preparation and targeted capture (Table 12). The assay was used as per manufacturer's protocol.

Table 12: DNA input modifications for FFPE samples based on DNA Integrity Number (DIN) Score

DIN score	>8	3-8	<3
DNA input for library preparation	10-200ng DNA	At least 15ng for more intact samples and at least 40ng for less intact samples. Maximum amount of DNA used, up to 200ng.	At least 50ng for more intact samples and at least 100ng for least intact samples. Maximum amount of DNA used, up to 200ng.

2.3.2 Tumour DNA Library Preparation and Targeted Capture

DNA extracted from FFPE tumour tissue was prepared by me for NGS using the Agilent SureSelect XT HS protocol as described in section 2.2.5. Prepared DNA libraries were sequenced on the MiSeq machine in batches of 8 samples to achieve adequate depth of coverage (at least 100x) for calling variants with low allele frequency.

2.3.3 Tumour DNA NGS Data Analysis

NGS data was analysed using the SureCall program which was updated to version 4.1 as this incorporated extra features for handling samples sequenced using the SureSelect XT HS protocol.

Demultiplexed FASTQ files were run through the SureCall program as described in section 2.2.7. The variant filtering settings were modified to allow the detection of low frequency variants (Table 8).

On review of the table of variants generated for each sample, the following filtering and analysis steps were carried out:

1. The known DNA repair gene germline variant was reviewed with respect to allele frequency and read depth. An allele frequency of ~0.5 indicates a variant is heterozygous reflecting the germline heterozygous status. A frequency of ~1.0 indicates loss of heterozygosity, which may be due to exon or gene deletion. Absence of the germline variant may indicate the presence of a rescue variant that has restored the gene reading frame.

2. Variants in the gene of interest (the gene harbouring a germline variant) were reviewed to identify any additional PTVs.

3. PTVs in other genes were identified using the filtering steps as described in section 2.2.7, i.e. filtering by category, primary effect and 'clinical dbSNP' parameters. For all PTVs, allele frequency was noted, along with read depth and reads were reviewed in the IGV. Forward and reverse read ratio was checked for strand bias.

4. When identifying a variant of interest, the NGS data for the other tumour samples was checked for the presence of the same variant. Variants that were observed in all samples were regarded as sequencing artefacts.

5. The CADD score was generated for variants of interest and databases were reviewed for available clinical classifications as well as somatic data information (Catalogue Of Somatic Mutations In Cancer; COSMIC).

Chapter 3   BARCODE1 Pilot Results

3.1 Introduction

Similar to other complex diseases, the genetic heritability of prostate cancer is composed of both rare, high to moderately penetrant variants and commonly occurring single nucleotide variants (SNVs) that confer risks of lower magnitude. With the advent of the genome wide association study (GWAS) and the increasing numbers of cases and controls included in such studies, prostate cancer GWAS and meta-analyses have identified approximately 170 loci associated with prostate cancer development (Reviewed in reference [84]). Most of these SNVs are commonly occurring single nucleotide polymorphisms (SNPs; i.e. minor allele frequency (MAF) ≥1%) and although each locus is associated with a low to moderate per allele odds ratio (OR), the genetic risk is log additive or multiplicative and increases with increasing number of risk alleles in a person's germline DNA. The currently known prostate cancer susceptibility loci explain an estimated 37% of the familial relative risk of prostate cancer.[11-13]

3.1.1 Prostate cancer screening

PSA (prostate specific antigen) testing for prostate cancer screening has been a controversial topic in the last few years. In terms of large scale population screening, as data from two large prostate cancer screening studies [56, 57] have evolved, guidelines from national screening programs such as the USPSTF (US Preventive Services Task Force) have fluctuated from advising against PSA screening for prostate cancer (2012) to recommending that men make an individualised decision regarding PSA testing in conjunction with their clinician (for those aged 55-69years; USPSTF 2017). In the UK, the National Screening Committee (NSC) recommends against universal screening for prostate cancer using PSA (2016, due to be updated 2019/20). Although the screening studies investigating the use of PSA testing for prostate cancer detection have shown a survival benefit, the side effects, complications of prostate biopsies for men who don't have cancer and high rate of false positive results have led to this caution around the use of PSA testing.

PSA is a prostate specific protein secreted by both healthy prostate tissue as well malignant cells, therefore it lacks the specificity required for a cancer screening test. In the Stockholm-3 study, which investigated the use of a multi-factor screening model, 21% of high risk prostate cancers had a PSA level in the range of 1-3ng/ml; below the threshold of ≥4ng/ml which is often used for screening. [61]

Ultimately, men at risk of prostate cancer due to symptoms and/ or a finding of a raised PSA are required to undergo a prostate biopsy to obtain a definitive diagnosis. To aid the shared decision making process with regards to prostate biopsy, several risk calculators have been developed as an alternative to a stand-alone PSA test. The Prostate Cancer Prevention Trial (PCPT) calculator and the European Randomised Study of Screening for Prostate Cancer (ERSPC) risk tools were developed based on the large prostate cancer screening studies carried out in the 1990's [85, 86]. These tools are largely outdated now as they were developed as a result of studies set up based on the clinical practice standards at the time which have progressed since then, for example, the Gleason grading of prostate cancer has been revised and the number of cores taken at a standard prostate biopsy has increased from 6 to 12. Both study populations consisted of healthy predominantly white Caucasian men. In the PCPT study, men were recruited if their PSA was less than 3ng/ml and they had a normal digital rectal examination.

To improve on these models and develop a risk strategy for a contemporary and more diverse population, the Prostate Biopsy Collaborative Group (PBCG) model and calculator was developed based on over 15000 men undergoing prostate biopsy recruited prospectively between 2006 and 2014 in North America and Europe [87]. The PBCG calculator predicts the risk of a high grade cancer, low grade cancer and benign result on biopsy based on the following clinical factors: age, PSA, digital rectal examination findings, first degree family history, African ancestry and previous negative

biopsy. Compared with the PCPT calculator, the PBCG model investigators reported that use of this model led to 25 fewer biopsies per 1000 patients when a risk threshold of 10% was used with no high grade cancers being missed.

Although such risk calculators may inform decision making regarding biopsies, further progress in other diagnostic modalities have led to screening models incorporating other tests such as MRI of the prostate. In fact, in May 2019, the National Institute for Health and Care Excellence (NICE) updated their prostate cancer diagnostic guidelines (NICE guideline NG131) to recommend MRI of the prostate as a first line test for suspected localised prostate cancer. Risk calculators such as those discussed above are based on trans-rectal ultrasound (TRUS) guided biopsies where MRI was not routinely carried out prior to biopsy, therefore may not be applicable to patients being considered for a MRI scan of the prostate +/- biopsy as opposed to a TRUS biopsy.

Multi-modal screening models may also incorporate blood biomarkers such as the 4K test (a blood test measuring 4 kallikrein proteins: total PSA, free PSA, intact PSA and human kallikrein 2) [88] and genetic information in the form of risk SNPs. The Stockholm-3 (STHLM3) prostate cancer screening model was one of the first such models to be investigated in a prospective study. [67] This model combines plasma protein biomarkers (4K test, hK2, MSMB and MIC1), 232 risk SNPs and a set of defined clinical variables (age, family history, previous prostate biopsy and prostate examination). When compared to PSA screening using a threshold of ≥3ng/ml, the STHLM3 model performed significantly better for the detection of Gleason 7 or higher prostate cancer with an AUC of 0.74 vs AUC 0.56. The number of prostate biopsies was reduced by one third, the number of benign biopsies by 44% and the number of low grade (Gleason 6) prostate cancers by 17%. The STHLM3 model was developed based on a Swedish population of men and would require validation in other populations.

As yet, no prospective studies utilising genetic profiling alone for prostate cancer screening have been carried out. The use of a genetic test for cancer screening is attractive compared to PSA testing as unlike non-specific and fluctuating PSA levels, germline DNA is constant and unchanging in terms of the variants and SNPs it holds, and only requires a one off measurement, usually in the form of a blood test or saliva test. By utilising the known genetic variants associated with prostate cancer risk, a polygenic risk score (PRS) can be calculated for an individual to estimate their risk of prostate cancer development. It is feasible that such a score could be used to stratify men for prostate cancer screening so that those with a high genetic risk of prostate cancer are offered tests such as MRI and biopsy (+/- PSA tests) while men at lower genetic risk may avoid the potential complications of invasive tests. Utilising the known prostate cancer risk loci, the relative risk (RR) of prostate cancer for men in the top 1% of the genetic risk distribution based on a PRS is 5.71 compared with men in the 50th percentile, and for those in the top 10% the RR is 2.69. [13] A risk model incorporating a genetic profile based on risk loci (with or without family history information) could be used to target screening to those at highest risk.

Two UK studies are currently investigating the role of a genetic profile in prostate cancer screening. In the PROFILE study (NCT02543905), men with a family history of prostate cancer are screened using PSA testing, MRI and biopsy. A genetic profile test is also carried out using a set of prostate cancer risk SNPs and the correlation between genetic risk and screening results will be investigated. The PROFILE study is also recruiting a separate cohort of black men to study genetic risk in this group.

Separately, the BARCODE1 study (NCT03857477) is enrolling men from the community via their General Practitioners (GP) to undergo a genetic profile test utilising 130 prostate cancer risk SNPs. Men in the top 10% of the genetic risk profile are offered screening with a MRI of the prostate followed by a biopsy regardless of their PSA level.

This is the first prospective study assessing the utility of genetic profiling in population screening and aims to recruit 5000 men.

As part of my PhD, I was involved in the set up of the BARCODE1 pilot study (N=300) designed to assess the feasibility of progressing to the BARCODE1 main study (N=5000). This study is funded by a European Research Council grant.

3.2 Aims

The aim of this pilot study was to investigate the uptake and feasibility of a community based prostate cancer screening trial utilising a genetic profiling assay that would genotype prostate cancer risk SNPs to produce a PRS for each study participant. Specifically, I aimed to:

- Assess the feasibility of a community based prostate cancer screening study by measuring the uptake of the study by eligible men in the community.
- Measure the distribution of prostate cancer PRS in the pilot cohort of men.
- Compare the PRS distribution in the pilot cohort to a reference population.
- Assess the uptake and outcomes of screening of men in the top 10% of the genetic risk distribution.

3.3 Results

3.3.1 Pilot Study Recruitment, Uptake and Sample Collection

The BARCODE1 pilot study gained local approval from The Royal Marsden Committee for Clinical Research (CCR) on the 9th of June 2014 and approval by the Research Ethics Committee (REC) was received on 26th January 2016. The pilot study opened to recruitment in April 2016 and completed in April 2018. Initially, 3 General Practice (GP) sites acting as participant identification centres (PICs) were involved. Four more GP sites were added to the study in April 2017 to increase recruitment rate and complete the pilot study. Figure 9 shows the cumulative recruitment to the pilot study over 24 months. I was involved in regular investigator meetings attended by representatives of all the primary care GP sites involved, as well as the Clinical Research Network (CRN) leads supporting the GPs, and the Institute of Cancer Research (ICR) trial team. Through the collaboration with the CRNs in the pilot study, the study was introduced to other GPs which allowed a further 100 GP sites to be identified in anticipation of the main BARCODE1 study which commenced after completion of the pilot study recruitment.

Participating GPs screened the medical records of 1802 men registered at their practices for study eligibility (Pilot study protocol included in Appendix 1); 1436 potentially eligible men (80% of those screened) were sent a study invitation letter which included the study Patient Information Sheet (PIS) and a health questionnaire. The health questionnaire was used to screen out men who did not fit the study eligibility criteria but may have been missed by the referring GP, as well as those with significant co-morbidities that may increase the risk of complications if a prostate biopsy was carried out.

Of 1436 men who were invited to the study, 375 men responded to the invitation letter giving a study uptake rate of 26% (range between GP sites: 13-47%) of whom 328 (87%) were eligible for study entry (Table 13). Reasons for exclusion from study entry included medical co-morbidities and non-Caucasian ethnicity; one of the GP sites was located in North London where there is a high proportion of patients of Jewish background in the local population. The study eligibility criteria excluded non-Caucasian ethnicities including Ashkenazi Jewish ethnicity. As the Ashkenazi Jewish population are at risk of carrying one of the *BRCA1/2* founder mutations, the genetic risk of prostate cancer can't be fully accounted for by a SNP based genetic test without *BRCA1/2* testing as well so this population was excluded.

Figure 9: BARCODE 1 Pilot Study - Cumulative Recruitment
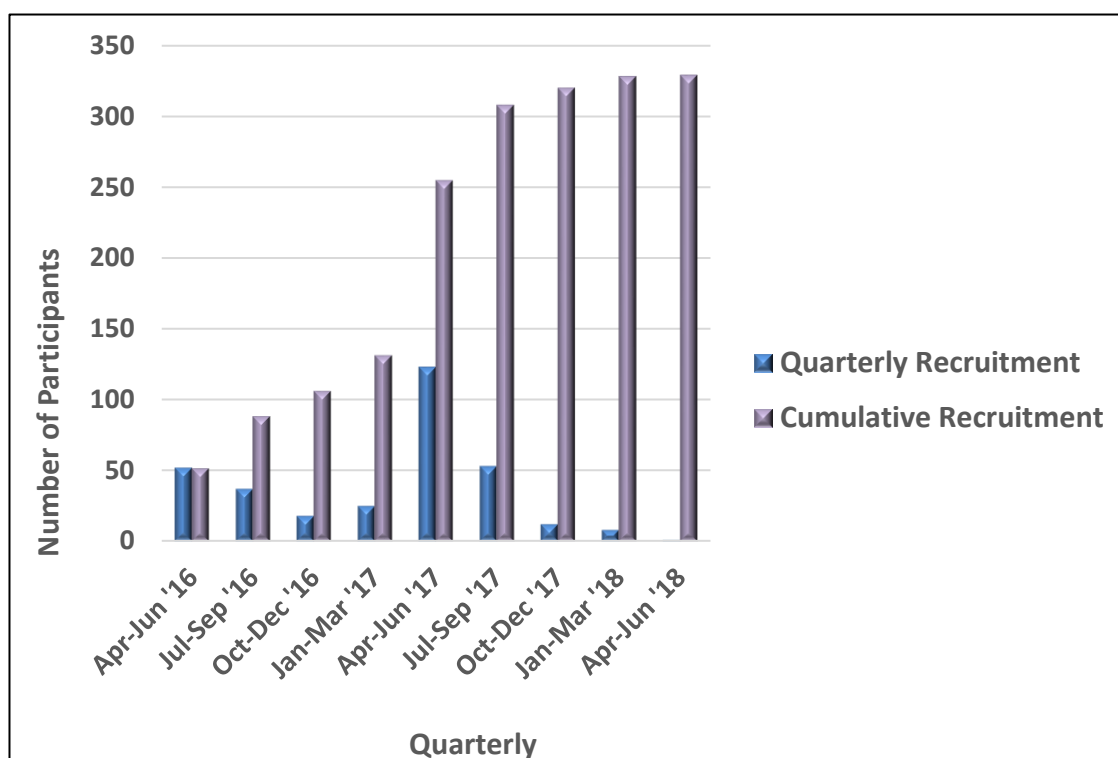
Table 13: BARCODE1 Pilot Study Screening, Response Rates and Saliva Returns

Site	Screened	Mail-outs	Total Responders	Uptake %	Eligible	Returned saliva samples (% of eligible responder)
GP 1	148	148	45	30	45	42 (93%)
GP 2	350	326	78	24	66	62 (94%)
GP 3	277	175	23	13	12	12 (100%)
GP 4	267	232	46	20	44	41 (93%)
GP 5	223	211	51	24	49	44 (90%)
GP 6	390	200	93	47	77	74 (96%)
GP 7	145	142	37	26	34	30 (88%)
RMH*	2	2	2	100	2	2
Total	1802	1436	375	26	329	307 (93%)

*2 patients recruited via Cancer Genetics Research clinic in RMH.

329 saliva collection kits were sent to the eligible participants along with a study consent form. 307 saliva samples were returned to the study team (Figure 10), giving a saliva return rate of 93% from eligible participants. 21 participants withdrew from the study after providing a saliva sample and 1 participant withdrew prior to providing a saliva sample, giving an overall withdrawal rate of 6.7% of eligible responders.

All correspondence and saliva collection was carried out by post. DNA was extracted from saliva by Tepnel Pharmaservices for 303 participants whose saliva sample was returned before the cut-off date (15th April 2018) for the pilot study. The remaining 5 saliva samples would be processed as part of the subsequent BARCODE1 full study. Of 303 saliva samples that underwent DNA extraction, one sample had a low yield of

DNA and required further saliva collection for that participant. Extracted DNA for 302 participants was shipped by the Oncogenetics team to Thermo Fisher Scientific in the USA for genotyping.

Figure 10: BARCODE1 Pilot recruitment showing fall out at each stage

21% of men invited to the study entered the trial and provided a saliva sample for genotyping

1436	**100%** Invited by GP to the study
375	**26%** Responded to invitation to participate
329	**23%** Eligible responders, 46 not eligible e.g. due to medical co-morbidities
307	**21%** Saliva samples received
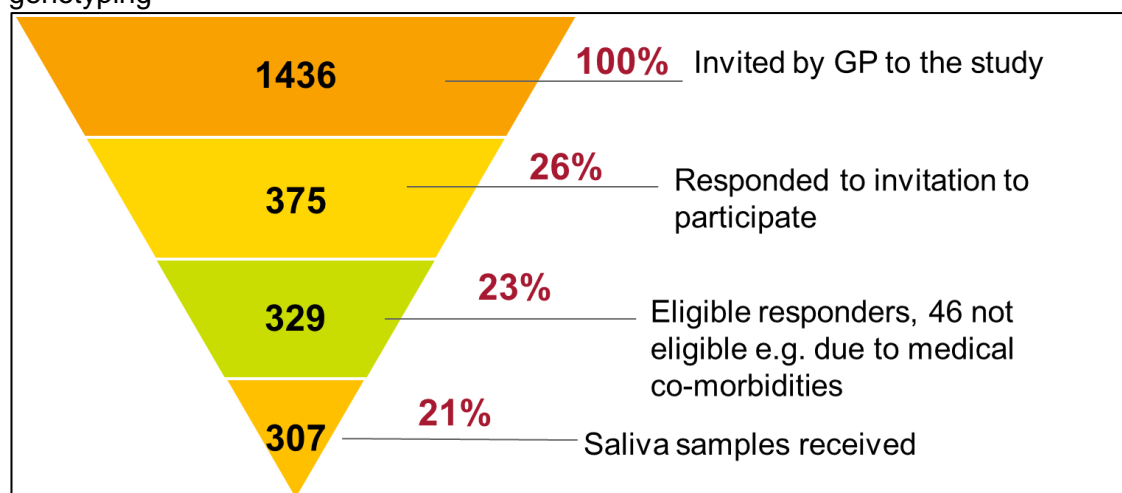
3.3.2 Pilot study participants' characteristics

The mean age of study participants at the time of consent was 61 years (range 55-69) (Table 14). Of the 307 men who returned a saliva sample, 47 (15%) had a family history of prostate cancer.

Table 14: Age groups of eligible responders in the BARCODE1 pilot study

	Mean	Median
Age	61	61
Age Groups	n	%
55-59	137	42
60-64	89	27
65-69	103	31
Total recruited	**329**	

3.3.3 DNA Genotyping

Genotyping data was returned to the ICR Oncogenetics team for 285 samples in the BARCODE1 pilot study. 17 samples (5.6%) failed the internal Thermo Fisher quality control (QC) processes although no feedback was received as to the reason why these samples failed or at which stage they failed. DNA assessment of failed samples was carried out by me in the Oncogenetics laboratory to decide whether repeat saliva sampling and DNA extraction was required for these failed samples. I used the Nanodrop spectrophotometer to assess DNA quality and performed an ethanol and bead clean up (using magnetic SeraMag SpeedBeads) for those that appeared to be of sub-optimal quality. The Nanodrop outputs a 260/280 ratio which reflects the absorbance measurements at 260nm wavelength and 280nm wavelength which in turn measures the purity of DNA. The optimal 260/280 ratio for DNA is ~1.8 (whereas for RNA this is ~2.0). A lower ratio can indicate the presence of contaminants. Upon

repeating the Nanodrop assessment of DNA post-clean up, 5 DNA samples had improved 260/280 measurements; these would be sent for genotyping with the samples in the main BARCODE1 study. The remaining 12 samples were still sub-optimal and further saliva collection was requested from these participants.

3.3.4 QC of Genotyping Results

The Eureka Analysis Suite (EAS) software was used to review the cluster plots for each SNV. Low confidence genotype calls were converted to 'no call' and uncalled genotypes that displayed sufficient confidence were converted to genotype calls. Figure 11 shows an example of this manual review of the genotype calls for one of the SNVs in the assay.

Figure 11: Manual change of genotype calls using Eureka Analysis Suite
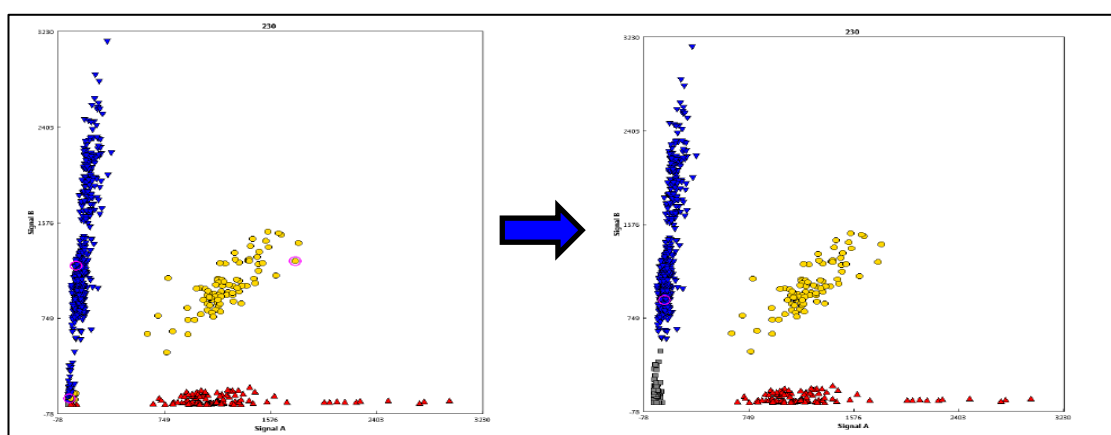
3 clusters can be observed representing homozygous B allele status (blue triangles), homozygous A allele samples (red triangles) and heterozygous AB samples (yellow circles). The cluster of samples in the bottom left of the chart called as homozygous for the B allele were converted to no call (grey circles) due to the low signal intensity.

Although 162 SNVs were included in the genotyping assay, only the 130 SNVs that had been identified as GWAS significant in European ancestry men in the OncoArray meta-analysis and used in that study's polygenic risk score (PRS) calculation [13] were used for calculating the PRS in the BARCODE1 pilot study. Although the OncoArray meta-analysis used 147 SNVs to calculate the PRS, 15 of these loci were not designable using Eureka™ Genomics (EG) technology and therefore not included in the

BARCODE1 genetic profile assay. Although the intended number of SNVs for BARCODE1 PRS calculation was 132, 2 SNPs were excluded during QC steps as described in Chapter 2.1.5. Genotyping data were processed using the EAS software for cluster plot review and R for generating the PRS values.

3.3.5 Polygenic Risk Score Distribution in the BARCODE1 Pilot Study Population

The PRS for each study participant was calculated based on the sum of weighted alleles for 130 SNVs as described in Chapter 2 using R. The PRS for 285 men in the BARCODE1 pilot study ranged from 8.42-12.21 (median 10.34; Figure 12). The mean (10.33) and standard deviation (0.64) were used to calculate the 90th percentile value as described in Chapter 2.1.5.2. The PRS at the 90th percentile was 11.15. This threshold value was used to identify participants who have a PRS in the top 10% of the PRS distribution.

Figure 12: Distribution of PRS in the BARCODE1 pilot cohort

The vertical red line denotes the 90th centile PRS value (11.15)
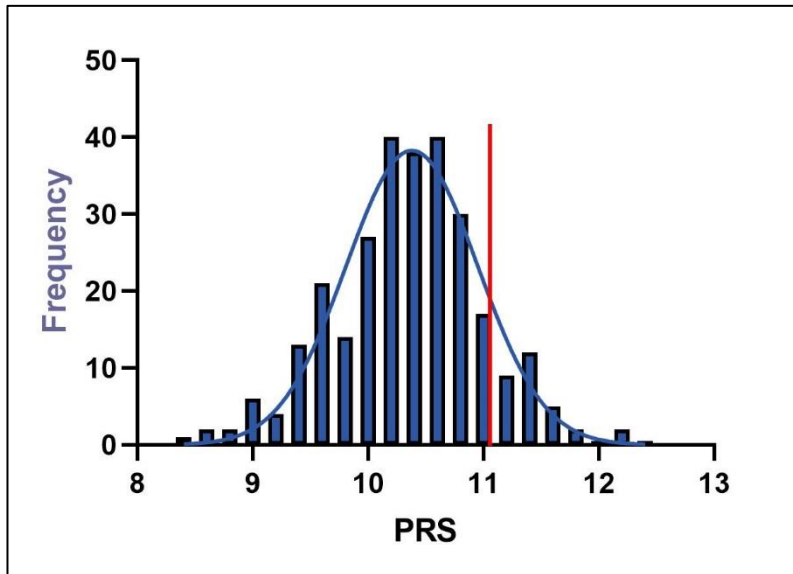
3.3.6 PRS Distribution Comparison to Other Populations

The PRS distribution and 90th percentile PRS in the BARCODE1 pilot was compared to two reference populations for whom genotyping data were available. The first reference population was taken from the ProtecT (Prostate Testing for Cancer and Treatment) trial [75], which recruited participants from GP surgeries across the UK to participate in a prostate cancer screening and treatment trial. The genotyping data for 2571 men aged 55-69 years who were not diagnosed with cancer in ProtecT were used to calculate the PRS (using the 130 SNVs used in the BARCODE1 genetic profile test). These data were available as a result of the OncoArray genotyping project [13]. For this cohort of men, the mean PRS was 10.33 (SD 0.65) and 90th percentile was 11.16 (Figure 13).

The second reference population was taken from the UK Genetic Prostate Cancer Study (UKGPCS). This is a large UK study (commenced in 1993) which recruits patients diagnosed with prostate cancer and collects patients' DNA samples as well as clinical and family history data [76]; a sub-cohort of participants without prostate cancer were recruited via their GP surgeries for epidemiological sub-studies. They were age

matched +/- five years to cases and were all Caucasian and had no family history of prostate cancer in first or second degree relatives. The recruitment of this cohort was coordinated by Professor Kenneth Muir at the University of Nottingham. We utilised the genotyping data for this sub-set of 500 men aged 55-69 years. Genotyping data were used to calculate the PRS. Mean PRS was 10.37 (SD 0.65) and 90th percentile PRS 11.19. (Figure 13, Table 15) All studies were approved by the appropriate ethics committees. All participants gave written informed consent.

Figure 13: Prostate Cancer PRS Distribution in the BARCODE1 Pilot Set and Two Reference Populations
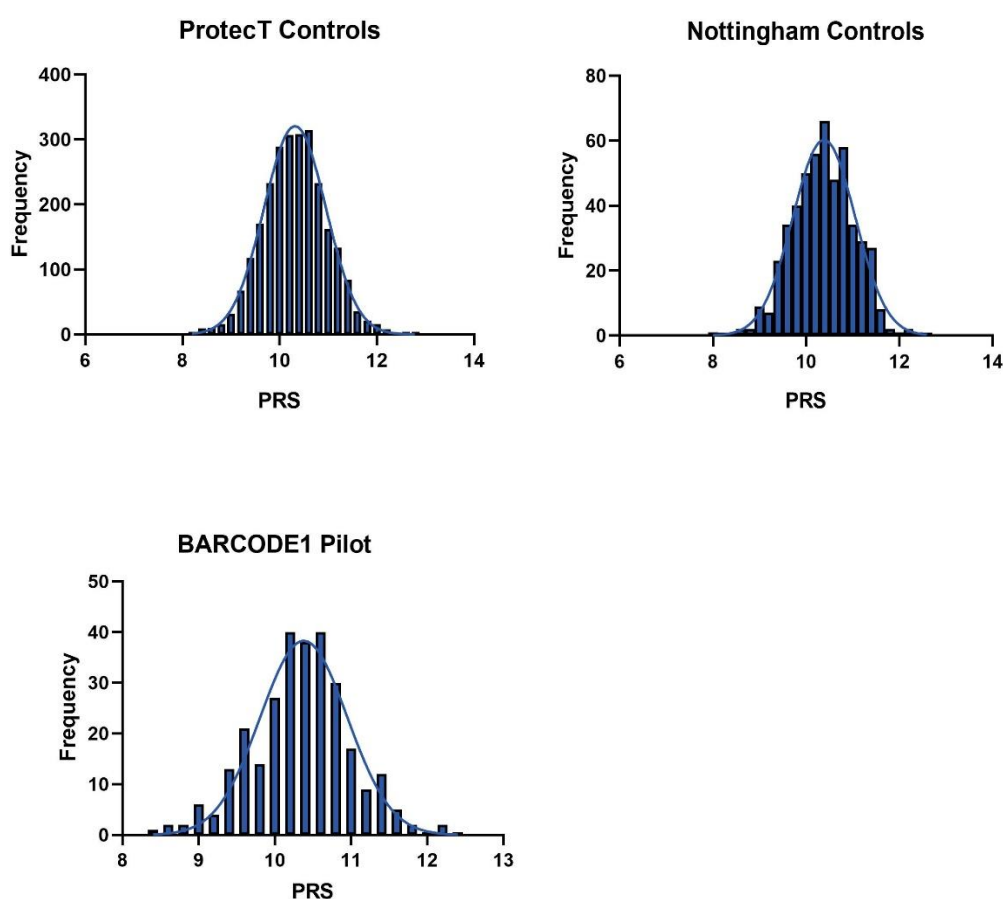
Table 15: PRS in BARCODE1 pilot and the two reference populations

Study	Number of participants with genotyping data	Mean PRS (range)	Standard Deviation	90th percentile PRS
BARCODE 1 Pilot	285	10.33 (8.42-12.21)	0.64	11.15
ProtecT	2571	10.33 (8.39-12.77)	0.65	11.16
Nottingham set (UKGPCS)	500	10.37 (8.06-12.5)	0.65	11.19

Figure 14: Boxplots of PRS distribution in the 3 populations
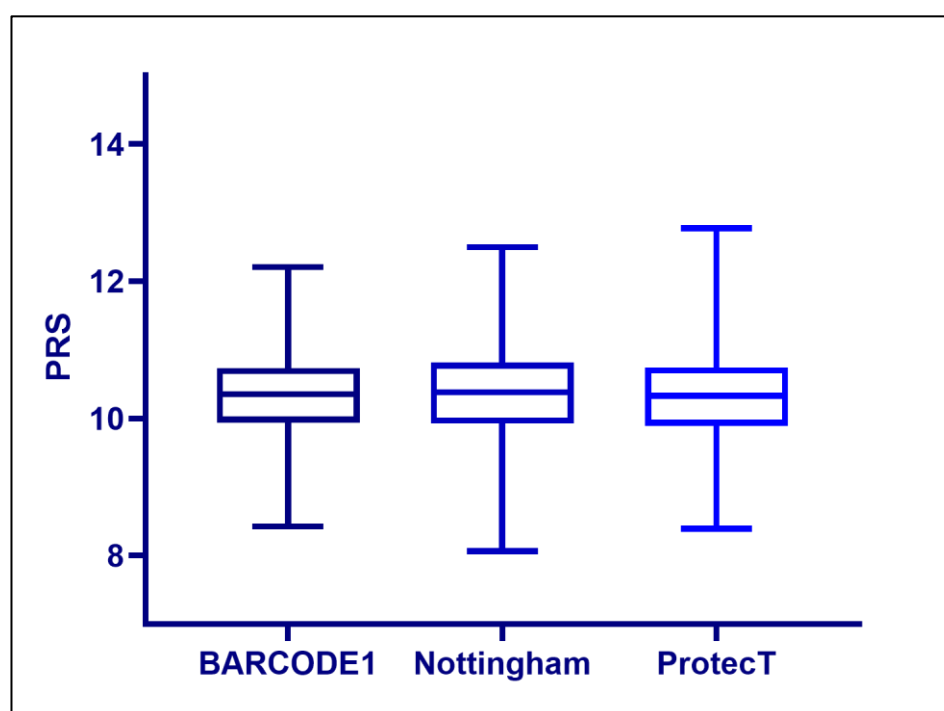
Comparison of the PRS distribution between these 3 data sets using the analysis of variance (ANOVA) showed no significant difference between the 3 sets (P=0.43) which provided reassurance regarding the PRS distribution and 90th percentile value used for screening within the trial.

3.3.7 Selection of PRS threshold for prostate cancer screening

The 90th percentile value for the PRS in the BARCODE1 pilot set was 11.15. This PRS value was used to select men for prostate cancer screening. This selected 26 men for screening out of 285 that had genotyping data available.

The PRS values for the next cohorts of men recruited into the main study will be reviewed as the study progresses. If within the main study with a larger dataset, the PRS threshold is found to be lower, men from the pilot study who have a PRS above this lower threshold could be contacted again and offered screening later.

3.3.8 Uptake of Screening by Men in the Top 10% of PRS Distribution

Of 26 participants identified to be in the top 10% of the PRS distribution, 7 men did not proceed with prostate cancer screening: one patient had died since entering the study (due to a road traffic accident), 2 were lost to follow up and 4 men withdrew from the study when offered screening. This gives an overall uptake of screening of 76% (19 of 25).

Of note, 5 of the 26 men in the top 10% of the PRS distribution had a family history of prostate cancer. Two of these men declined to undergo screening, one had a negative biopsy, one was the patient who had died after providing the saliva sample and one was diagnosed with ASAP on biopsy.

3.3.9 Outcomes of Prostate Cancer Screening

Screening for prostate cancer was carried out by MRI of the prostate followed by biopsy. If a suspicious lesion was identified on MRI this was targeted for sampling at the time of biopsy in addition to standard sampling. A PSA level was measured prior to MRI and biopsy although this didn't affect the decision to screen. The mean PSA level for 19 men who underwent screening was 1.60 ng/ml (range 0.3-5.8 ng/ml)

MRI of the prostate was carried out for 19 men. 5 scans were reported to have a target lesion for sampling at the time of biopsy. Biopsy results were available for 18 men who underwent a successful procedure. ne man did not proceed to biopsy due to persistent sterile pyuria (white cells present in urine) for which he was referred locally for urological investigations. Table 16 summarises the MRI findings and Table 17 displays the biopsy outcomes.

Table 16: MRI Results for 19 BARCODE1 Pilot Participants

PIRADS Score	Number of cases
1	2
2	8
3	5
4	4
5	0
Number of scans with target lesions identified for biopsy	5 (2 positive for cancer)

PIRADS= Prostate Imaging and Reporting Data System

Table 17: Prostate Cancer Details for 6 Cases Diagnosed on Screening

Case	Gleason Grading	GG	Length of tumour (mm)	Tumour % in biopsy cores	Bilateral Tumour	Target lesion on MRI (+/-)*	PSA level pre-MRI (ng/ml)
1	3+3=6	1	2	1	Yes	Yes (+)	0.78
2	3+3=6	1	2	1	No	No	5.8
3	3+3=6	1	1	<1	No	Yes (-)	2.3
4	3+3=6	1	1	<1	No	No	1.0
5	3+3=6	1	0.5	<0.5	No	Yes (+)	3.6
6	3+4=7	2	1	1	No	No	2.1

*+ indicates target lesion was positive for cancer and – indicates it was negative for cancer. GG= Grade group

Prostate cancer was diagnosed in 6 cases. Of the remaining 12, one case was diagnosed with ASAP and one with HG PIN; a repeat MRI and biopsy was recommended to be carried out for both 6 months later as per trial protocol (and

standard practice due to the risk of subsequent cancer). All diagnosed cancers were low grade and small in size (Table 17). All cancers were discussed in the RMH uro-oncology MDT and management by active surveillance was recommended for all 6 cases.

3.3.10 Incidental Findings and Post Biopsy Complications

There was one case of an incidental finding reported on MRI of the prostate that required further investigation. For this case, the bone marrow in the pelvis appeared heterogeneous and mildly enlarged lymph nodes were noted in the left internal iliac and obturator regions, therefore the uro-oncology MDT recommended haematological investigations and referral. The patient was subsequently diagnosed with a low grade B cell non-Hodgkin's lymphoma. He was also one of the cases diagnosed with a Gleason 6 prostate cancer (Case 3 in Table 17).

A urinary tract infection was diagnosed after biopsy in 2 out of 18 men (11%) who underwent a prostate biopsy. Both cases were successfully treated with oral antibiotics. There were no cases of urinary sepsis requiring hospital admission and no other post-biopsy complications.

3.3.11 Considerations for the BARCODE1 Main Study

Based on the uptake of 26% in the pilot study, to recruit approximately 4700 men to the main study would require ~18000 men to be invited by their GP. Depending on the size of a practice, we estimated that approximately 100 GP sites would be needed for the main study. Through the close collaboration between the trial team and the Kent Surrey and Sussex (KSS) Clinical Research Network (CRN), the study was introduced to GP sites within the network and 50 surgeries were identified that wished to participate in the study. The study was also promoted to other CRNs and although several networks wished to join the study, the study team limited inclusion of new networks to two regions

that would provide 25 GP sites each: Thames Valley and South Midlands CRN and South London CRN.

The level of interest in the BARCODE1 trial from the CRNs and GPs was extremely encouraging, and demonstrates the acceptability to healthcare professionals outside the standard oncology research setting of carrying out translational research investigating personalised medicine and screening.

As a result of feedback from and regular discussions with the GPs and CRN representatives, the following changes were instituted for the main BARCODE1 study in an effort to improve study uptake and streamline the recruitment process:

- The paperwork contained in the study information pack that is sent with the invitation letter by GPs was reduced as some patients found the participant information sheets to be too long.
- Database searches were created based on the study eligibility criteria that could be applied to the two most commonly used primary care database systems: EMIS and SystemOne. This would allow all GPs to use the same search for the identification of potentially eligible participants. I was involved in reviewing the terms and diagnoses included in these searches before they were finalised to ensure they were adhering to the study protocol as closely as possible.
- The system for mailing out the study invitation packs was automated by the use of the DOCMAIL service. This is an electronic system used by the GP sites to send out letters for routine clinical work. This was utilised for the BARCODE1 main study so that all the documents included in the invitation pack were uploaded to DOCMAIL within a study specific DOCMAIL account. Participating GPs can then upload the patient contact details in bulk to DOCMAIL and request them to be mailed out. For the pilot study, study information packs were made up manually and posted out by the GP staff; by incorporating the use of the DOCMAIL system, a quicker and much more efficient mail out process was developed.

- The pilot study offered participants the option of completing an online version of the screening health questionnaire used prior to study entry, but we found that the uptake of this was very low as participants preferred to post back their completed paper questionnaire. Therefore, this online option was removed for the main study.

3.4 Discussion

The BARCODE1 pilot study recruited 307 men and had an uptake of 26% by men in the community. This compares favourably to the PROFILE pilot screening study reported by our team in 2016 [89] which had an uptake of 12.8% in men with a family history of prostate cancer. The uptake in the BARCODE1 pilot study is lower than that reported in the ProtecT [75] and CAP [90] studies which invited participants to the studies via their GPs. The ProtecT study had an uptake of 36%. Similarly, in the CAP trial, 36% of men invited to the study and offered PSA screening underwent a PSA blood test.

The uptake of screening tests by men identified to be in the top 10% of the risk profile in BARCODE1 pilot was 76% which is encouraging and comparable to the 2016-2017 data on the uptake of bowel cancer screening in England which was 59%.[91] The bowel cancer screening program invites men and women aged 60-74 years for screening every two years. Similarly, the uptake of breast cancer screening in women aged 50-65 years in England was 72% in 2017/2018 (Public Health England data)[92].

Of the men screened, 33% (6/18) were diagnosed with prostate cancer. All cases showed low Gleason grade on pathology and low stage on imaging (i.e. no nodal involvement). It is unknown at this stage if these cases will progress to a higher grade cancer requiring intervention, but with further follow up this will become apparent. All cases are being managed with an active surveillance approach which involves regular PSA tests, interval MRI scans and repeat interval biopsies. If a cancer is upgraded at any time then treatment would be offered as appropriate in the form of either surgery or radiotherapy.

The issue of over-diagnosis has been observed in the large prostate cancer screening studies. Over-diagnosis is defined as the 'detection by screening of tumours that would

not have presented clinically in a person's lifetime in the absence of screening'. [65] Although this is expected to occur in most cancer screening programs, mathematical modelling by Pashayan and colleagues has suggested that in the context of prostate cancer screening, the proportion of over-diagnosed cases may be inversely proportional to polygenic risk. [65] This finding was based on the analysis of a prostate cancer PRS based on 66 SNPs in 17,012 men derived from 3 UK based studies. There was a 56% reduction in over-diagnosis between the lowest and highest PRS quartiles. Therefore, utilising a PRS to target screening may reduce the proportion of over-diagnosed cases compared with PSA screening. With the addition of further prostate cancer risk SNPs to the PRS and incorporating factors such as prostate cancer family history with or without other biomarkers e.g.4K test, the reduction in over-diagnosis may be enhanced while minimising the number of high grade/ high risk cancers missed.

A recent follow up study to that reported by Pashayan and colleagues went on to analyse the risk- benefit balance of screening based on age and PRS and assessed the cost effectiveness of such screening [93]. In this study, the investigators used simulated cohorts of men aged 55-69 years and a screening model based on varying the polygenic risk threshold for screening; this was compared to an age based screening model as well as a 'no screening' model.

With a 10 year absolute risk threshold of 2%, ~36,500 prostate cancer deaths were prevented (compared to no screening) with ~85,000 cases being over-diagnosed. Raising the risk threshold to 10% led to the prevention of ~14,500 deaths and ~27, 000 over-diagnosed cases. The study investigators suggested that a 10-year absolute risk threshold of between 4.5% and 7.5% may be the most appropriate in terms of harm-benefit balance and the ratio of prostate cancer deaths prevented to cases over-diagnosed. The incremental cost-effectiveness ratio of the age and PRS based screening also only plateaued when the 10 year absolute risk threshold rose above

7.5%. As prospective studies such as BARCODE1 continue, the data generated can be used in such simulations and modelling to design and plan screening programs that could potentially be applied to population screening in the future.

With the aim of a cancer screening program being the identification of early stage cancers amenable to curative treatment, it is not unexpected that the BARCODE1 study will identify a significant number of low grade/ stage cases; these will be amenable to curative treatment if upgraded during active surveillance. Distinguishing these cancers from low grade over-diagnosed cases will require long term follow up. With the increased acceptability and safety of using active surveillance to manage such cases, patients who do not progress to high stage disease can avoid the side effects and complications of radical treatment while those who have disease progression can be treated promptly when needed. The 10 year outcomes from the ProtecT trial showed that mortality due to localised prostate cancer was similarly low between men treated with active surveillance and those treated with surgery or radiation, although longer term study follow up is still needed. [94] Despite the ability to avoid interventional treatments for a subset of men, and avoid complications of surgery or radiation, over-diagnosis of indolent disease still places a burden on health systems required to follow up such cases with regular PSA tests and scans; these factors would need to be taken into account if a genetic profiling approach to population screening were to be undertaken.

3.4.1 Limitations

Due to time constraints, I was unable to obtain the full results of the prostate biopsies carried out in the pilot study although the majority are included. The results of the main BARCODE1 study which will recruit 5000 men in total cannot be extrapolated from the pilot study, nor was that the aim. Long term follow up is required to assess the degree of over-diagnosis within the pilot study as well as the main study.

The BARCODE1 trial is investigating the utility of the prostate cancer PRS designed for a European/ Caucasian population and will not be applicable to populations of other ethnicities. Although family history information is collected in the BARCODE1 study, this is not included in the study screening model.

3.4.2 Conclusions

The BARCODE1 pilot study was successful in demonstrating the feasibility of a larger prostate cancer screening trial to investigate the use of genetic profiles for population screening. It has also shown that a community based prostate cancer screening program based on genetic profiling is acceptable to primary care health care professionals and their patients. The pilot study uptake of 26% and the screening uptake of 76% are encouraging and demonstrate the acceptability of such screening to men in the community. The ongoing BARCODE1 main study suggests initial study uptake will be higher than that seen in the pilot.

The pilot study has shown that identifying the threshold for screening based on PRS distribution is not straightforward and the results of the completed BARCODE1 study (N=5000) will inform the best approach to setting a PRS threshold to identify men for screening.

Chapter 4   BARCODE2 Trial Germline DNA NGS

4.1 Introduction

Recent research investigating rare germline pathogenic variants in men with prostate cancer have reported a higher than expected carrier frequency, particularly in those with metastatic disease.

Research in this area has lagged behind that of hereditary variants in other diseases such as breast and ovarian cancer, but with the increasing use of next generation sequencing (NGS) and gene panel testing, the reported frequency of germline variants in DNA repair and tumour suppressor genes varies from 12-19% in patients with metastatic prostate cancer. In men with localised disease, the carrier rate is lower at 4.6% according to TCGA (The Cancer Genome Atlas) data. In a study assessing only *BRCA1/2* and *ATM*, men with aggressive prostate cancer had a significantly higher carrier rate (6%) compared to those with indolent disease (1.4%). [95] The implication that men who carry rare germline variants are at increased risk of poorer prognoses and aggressive disease features compared with non-carriers may warrant genetic testing of men at an early stage of disease to stratify those who should be offered more intensive surveillance or interventional treatment. Table 18 summarises the findings of recent germline sequencing studies in prostate cancer.

Table 18: Studies reporting germline variants in prostate cancer cohorts

Study	Number of patients	Population selected	Number of genes tested	Carrier frequency
Pritchard et al 2016 [30]	692	Metastatic PC unselected for FH or age at diagnosis	20 DNA repair genes	11.8%
Hart et al 2016 [96]	71	mCRPC patients	157 cancer predisposition genes (subset of WES)	17.4%
Annala et al 2017 [41]	319	mCRPC patients	22 DNA repair genes	7.5%
Castro et al 2019 [28]	419	mCRPC patients	107 DNA repair genes	16.2%
Nicolosi et al 2019	3607	Men with prostate cancer, unselected for disease stage	2-80 (62% used the Invitae 14 gene prostate cancer panel)	17.2%

FH: family history; PC: prostate cancer; WES: whole exome sequencing

The variation in frequency of carriers between different studies arises due to differences in population characteristics as well as the differing numbers of genes tested in each set; it is not unexpected that studies utilising larger gene panels have reported a higher frequency than those assessing 2-3 genes only. Although these studies report interesting and (for some genes) significant differences in the clinical characteristics between carriers and non-carriers, ascertaining true association of germline variation with disease causality requires large case-control studies.

4.1.1 Genetic Testing Guidelines

Guidelines for germline genetic testing in prostate cancer have been slow to develop. As of yet, there are no European guidelines in this context; in the UK some genetics centres utilise the Manchester scoring system to assess whether *BRCA1/2* germline testing is warranted in men with prostate cancer with features such as young age of onset (<60 years) or a significant family history. In the US, the NCCN guidelines were

updated in 2018 to provide clearer guidance in this area. The prostate cancer management guidelines were modified in 2018 to include a recommendation to consider testing all men with mCRPC for germline and somatic variants in *BRCA1/2, ATM, PALB2* and *FANCA*. (NCCN Prostate Cancer Guidelines Version 3.2018) This update also added the recommendation to consider tumour testing for MSI or MMR-d. This guideline was updated further in 2019 (version 2.2019) to recommend germline genetic testing for all men with high-risk, very high-risk, regional or metastatic prostate cancer. The genes listed for consideration for testing was also expanded to include *CHEK2* and the Lynch Syndrome genes: *MLH1, MSH2, MSH6* and *PMS2*. In addition to this, the NCCN 'Genetic/ Familial High-Risk Assessment for Breast and Ovarian Cancer guideline' was updated in 2018 to include a recommendation for *BRCA1/2* germline testing in all men with metastatic prostate cancer, without the requirement of a family history or young age of onset (Version 1.2018). This guideline also recommends germline *BRCA1/2* testing for men whose tumours are found to have a pathogenic somatic variant in one of these genes. These are the first set of guidelines to be explicit in their recommendation for the genetic testing of metastatic prostate cancer patients regardless of family history. Recommendations from other organisations are expected to follow. These recommendations are likely to evolve further to include genes other than *BRCA1/2* and *ATM.*

4.1.2 Genetic Variants and Personalised Treatment

Apart from identifying men at risk of worse prognosis, identifying germline variants in certain genes could open up new lines of treatment for those with advanced prostate cancer. Men with DNA repair gene variants may be able to access treatment with a PARP inhibitor, although this can only be within a trial setting currently. The TOPARP-A trial reported an 88% response rate to olaparib in men with DNA repair gene mutations in their tumours, some of whom also carried variants in the germline.[42] In 2018, results from the TRITON2 trial which recruited heavily pre-treated mCRPC patients showed

that men with a *BRCA1/2* mutation in their tumours had a 44% radiographic response rate (11/25; all partial responses).[97] Two of 8 patients with variants in other DNA repair genes also achieved a partial response (*BRIP1* and *FANCA* genes). The subsequent phase III TRITON-3 trial is now ongoing recruiting men specifically with somatic variants in a DNA repair gene.

The sensitivity to PARPi in this subset of patients is explained by a synthetic lethal interaction between PARP inhibition and the tumour's aberrant homologous recombination (HR) DNA repair pathway. This also explains the known cross-sensitivity between PARPi and platinum chemotherapy in patients with DNA repair gene variants in somatic (+/- germline) DNA. Platinum agents substitute alkyl groups for hydrogen atoms on DNA which leads to the formation of inter- and intra-strand DNA cross links. This in turn causes DNA double strand breaks which require intact HR DNA repair to maintain cell integrity. This explains the marked sensitivity of ovarian cancers with defective HR DNA repair to platinum chemotherapy and has been reported in retrospective studies of prostate cancer cases associated with *BRCA2* mutations.[70, 71]

Lynch Syndrome arises due to germline mutations in one of the mismatch repair genes (*MLH1, MSH2, MSH6* and *PMS2*) and increases the risk of colorectal cancer as well as extra-colonic cancers such as stomach, endometrial and ovarian cancers. Several studies have reported an increased incidence of prostate cancer in men with Lynch Syndrome compared with non-carriers. A European study reported the standardised incidence ratio of prostate cancer in Lynch Syndrome men to be 5.9 with a cumulative risk by the age of 60 years of 9.8% and by age of 70 years 29% [46]. In an American study, similar cumulative risks were reported of 6% by age 60 years and 30% by age 80 years. [47] Positive MMR mutation carrier status conferred a prostate cancer hazard ratio of 1.99. Lynch Syndrome men may also present at a younger age compared to the

general population and be at risk of more aggressive disease features such as higher Gleason scores. [47] Until recently, tumour testing for MSI or MMR-d was only routinely carried out for colorectal and endometrial cancers, but a study presented at the American Society of Clinical Oncology (ASCO) 2018 annual conference suggested that more widespread tumour testing for MSI/ MMR-d should be carried out.[48] In this study of 15,045 tumours across >50 cancer types, 6.8% of tumours were MSI-I (intermediate) or MSI-H (high). Germline variants in a MMR gene were identified in 1.9% of MSI-I tumours and 16.3% of MSI-H tumours. Half of these tumours belonged to a cancer type not traditionally associated with Lynch syndrome including prostate cancer among others.

For patients with evidence of mismatch repair deficiency (MMR-d) in their tumours, PD1 directed immunotherapy was approved in 2017 by the FDA regardless of tumour type. A subset of such patients will carry germline variants in a MMR gene. In Europe, access to PD1 directed drugs for tumours other than melanoma, lung and transitional cell cancer of the bladder is currently restricted to trials selecting patients based on MMR-d or MSI-H on tumour testing. With the development of oncological therapies based on molecular targets rather than specific tumour types, further treatment options are likely to be identified for prostate cancer patients carrying germline and/ or somatic pathogenic variants. Offering genetic testing at the time a man is diagnosed with prostate cancer may allow the modification of up front treatment where there is an association with poorer prognosis, as well as leading clinical teams to pursue suitable clinical trials at an early stage in the setting of advanced disease.

4.2 Aims

My aims in this project were:

1. To set up the BARCODE2 trial which would recruit men with mCRPC for germline sequencing and identify those who may benefit from carboplatin treatment.
2. To carry out the NGS for the initial 100 men enrolled into the BARCODE2 trial using a panel of 115 genes.
3. To measure the frequency of DNA repair gene PTVs in a 100 mCRPC patients.
4. To compare the characteristics of PTV carriers and non-carriers within the BARCODE2 trial.
5. To examine the polygenic risk score distribution for 100 mCRPC patients in the BARCODE2 trial using the SNP profile test used in the BARCODE1 study.

4.3 Results

4.3.1 Patient Recruitment and Characteristics

Between May 2017 and April 2019, 100 trial eligible mCRPC patients were recruited to Part 1 of the BARCODE2 trial for genetic screening. Patient characteristics are summarised in Table 19. Family history information was available for 97 patients.

Median age at study entry was 67 years (range 46-84). Median age at the time of initial prostate cancer diagnosis was 60 years (range 43-74). Ten percent of patients had previously been diagnosed with a cancer other than prostate cancer; these are listed in Table 22. Table 20 and Table 21 summarise the baseline pathological grading and staging groups for the patients for whom data were available.

Table 19: Clinical characteristics for 100 mCRPC patients in BARCODE2

Ethnicity	White European	88
	Black (African or Caribbean)	6
	Asian (South-East Asian or Middle Eastern)	4
	Mixed	2
Family history of:	Any cancer	76/99
	Prostate Cancer	35/99
	Prostate Cancer in a FDR	26/99
	Breast/ Ovarian Cancer in a FDR	21/99
	Cancer in a first degree relative	67/99
	Breast/ Ovarian/ Prostate FDR	39/99

FDR= first degree relative

Table 20: Prostate Cancer Baseline Pathological Grading for 100 BARCODE2 Patients

Gleason Score	Number of Cases
Unknown	14
6	4
7	28
8	9
9	45
10	0
Total	100

Table 21: Prostate Cancer TNM staging at Baseline

T stage	Number of Cases
T1	2
T2	10
T3	41
T4	11
Unknown	19
N Stage	**Number of Cases**
N0	43
N1	35
Unknown	22
M Stage	**Number of Cases**
M0	56
M1	44

Table 22: Types of cancers diagnosed previously in 10 patients in the BARCODE2 set

Tumour type	Number of cases
Head & neck	2
Colorectal	1
Melanoma	2
Adrenal cancer	1
Renal cell cancer	1
Testicular	1
Skin basal cell cancer*	2
Myxopapillary ependymoma	1

*One of these occurred in the patient previously treated for testicular cancer

4.4 DNA Extraction from Whole Blood

DNA was extracted and quantified as described in Chapter 2. Extracted DNA was suspended in 750µl of TE (Tris EDTA buffer; pH 7.5) and quantified using the Qubit fluorometer Broad Range assay. Mean concentration of extracted DNA from these samples was 453ng/µl (range 21-1600ng/µl; apart from 6 DNA samples, all had a concentration >100ng/µl). DNA samples were stored in Eppendorf tubes at -20°C until ready for library preparation.

4.4.1 DNA library preparation and targeted capture for NGS

DNA libraries were prepared for each set of NGS using the Agilent SureSelect QXT reagent kit. NGS runs on a MiSeq were carried out every 4-8 weeks depending on the throughput of DNA samples within the BARCODE2 trial. I carried out the practical work for 15 sets of NGS using germline DNA from patients in the study; the number of DNA samples in each experiment ranged from 5 to 11. A control DNA sample was included in each set; this was the BioChain® Control Genomic DNA (human male).

The Agilent SureSelect QXT protocol consisted of two parts carried out over two days for each set of samples (detailed in Chapter 2). In part 1, for each sample, 50ng of DNA (2µl of DNA diluted to 25ng/µl) was fragmented and adapters were ligated prior to PCR amplification. DNA purification using AMPure XP Beads was carried out before and after PCR amplification. At the end of these steps, the amplified DNA was assessed on the Agilent Bioanalyzer 2100 using the DNA 1000 assay to measure DNA fragment size. Target size was 245-325bp (Figure 15-a). DNA was also quantified using the Qubit fluorometer (Broad Range Assay). To proceed to part 2 of the protocol (hybridisation and targeted capture), 500-750ng DNA was required. If the DNA quantification was below this for any of the samples or if DNA fragment size on the Bioanalyzer indicated inadequate fragmentation, then part 1 of the protocol was repeated using a fresh aliquot

of DNA for those samples. Reasons for DNA loss may be due to loss of DNA during the steps involving purification using magnetic beads (e.g. inadequate mixing of beads and DNA) or over-drying the magnetic beads before eluting DNA in water. In an effort to minimise this occurring, drying time was reduced from 3 minutes to 1 minute when carrying out bead purification. Repetition of part 1 of the protocol was required for 7 of the experiments, 4 of which were due to issues with DNA fragmentation (described later).

Part 2 of the protocol is described in Chapter 2, but briefly, biotinylated RNA baits were hybridized to DNA targets in the genes of interest. These were then captured in solution using streptavidin beads. Captured DNA was PCR amplified and unique dual index sequences were added to each sample. After removal of streptavidin beads, DNA was purified and then assessed on the Agilent Bioanalyzer using the High Sensitivity assay. Target fragment size at this stage was 325-450bp. If the Bioanalyzer results showed DNA loss, then the whole protocol was repeated with a fresh aliquot of DNA. Repeating the whole protocol was required for 2 experiments, one due to DNA loss and one due to unusual Bioanalyzer traces which may be due to the presence of large DNA fragments.

4.4.2 DNA Fragmentation During Library Preparations

During my initial experimental work, with the third set of experiments specifically (NGS set P003), the results on the Bioanalyzer assessment at the end of part 1 steps (DNA fragmentation, adapter ligation and PCR amplification) appeared to be sub-optimal in terms of DNA concentration. Quantification using the Qubit fluorometer confirmed the low DNA concentration. Therefore, the Day 1 steps of the protocol were repeated for fresh aliquots of DNA for this set (Figure 15-b-c). Although the graphs on the Bioanalyzer results in experiment 3 (both the first and second attempts) appeared different in shape to our first two experiments, the fragment size was acceptable on the

repeated P003 set to proceed to the next steps of the protocol after repeating day 1 (Figure 15-c).

We liaised with the Agilent support team after the fourth experiment (P004) where it appeared that DNA fragmentation was inadequate (large fragment size on the BioAnalyzer graphs-Figure 15-d) in addition to the issue of low DNA concentration at the end of part 1 steps. After carrying out troubleshooting steps such as re-quantifying the DNA input used at the beginning of the protocol, and repeating the experiments using DNA samples freshly diluted to 25ng/μl, I concluded that the transposase fragmentation enzyme in the 96 reaction kit we were using must be degrading with repeated freeze thaw cycles. This was confirmed when I obtained a new set of reagents and successfully carried out the experiment using a new vial of fragmentation enzyme. (See Bioanalyzer graphs for experiment 4 before and after use of new reagents in Figure 15-d-e). For P001-004 sets, I had been using a 96 reaction kit where the transposase enzyme is supplied in a single vial for 96 reactions. This was replaced by Agilent with four 16 reaction kits (containing enzyme vials for 16 reactions at a time) so that the laboratory work could continue. I was advised by the Agilent support team to aliquot the fragmentation enzyme into smaller volumes in tubes when I next started a new 96 reaction kit.

After using up the replacement 16 reaction kits for experiments P005-P011, a new 96 reaction kit was ordered. Despite aliquoting the fragmentation enzyme in this kit to avoid repeated freeze thaw cycles, I encountered the same problem with the deterioration of DNA fragmentation with serial experiments (P012). After further discussion with the Agilent team, I switched the library preparation reagents from the SureSelect QXT reagent kit to the SureSelect XT HS kit which utilises an endonuclease enzyme for DNA fragmentation that was thought to be more robust for my experiments. As I intended to

carry out FFPE derived tumour DNA sequencing, this kit could be used for both germline and somatic DNA sequencing.

Figure 15: Bioanalyzer Results showing traces for Experiments P001, P003 and P004

a. Experiment P001 (29.06.17) ideal traces seen with peak DNA fragment size in the target region of 245-325bp
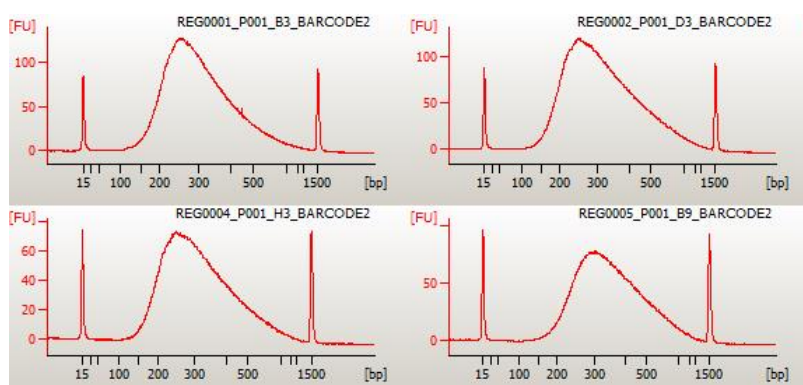
b. Experiment P003 (25.08.17): Start of deterioration of DNA fragmentation seen in P003; this persisted when the experiment was repeated on 01.09.17, but results were still adequate to continue with lab protocol.
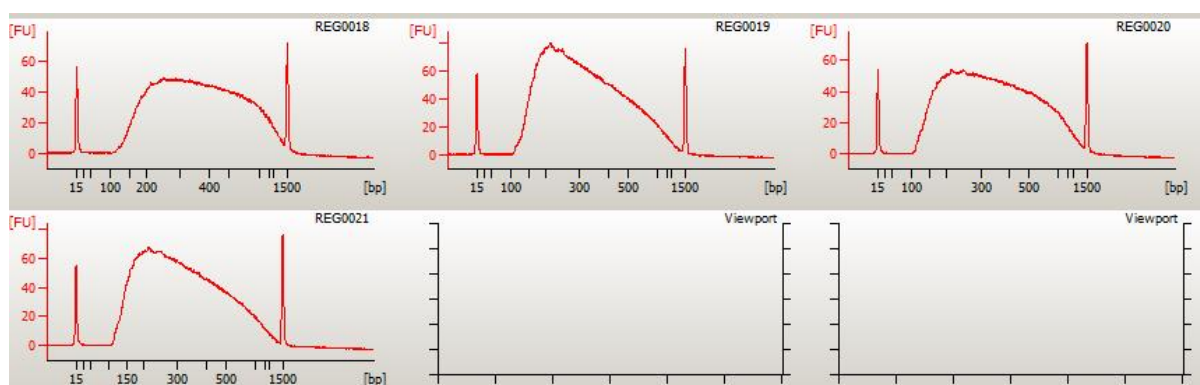
c. Experiment P003 repeated 01.09.17
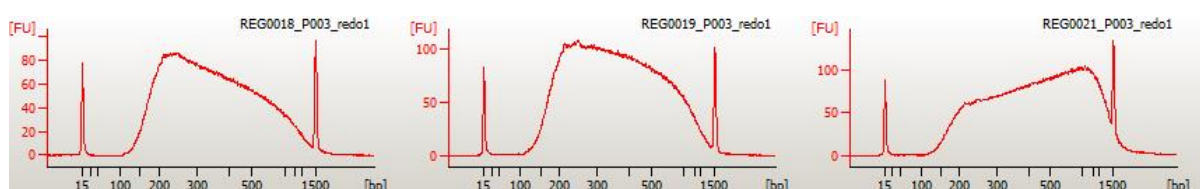
d. Experiment P004 (16.10.17) Inadequate DNA fragmentation and fall in DNA concentration
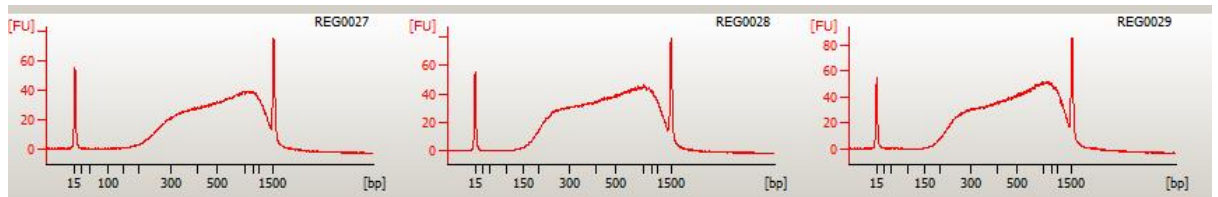
e. Experiment P004 using new reagents and new fragmentation enzyme 01.11.17 shows adequate DNA fragmentation
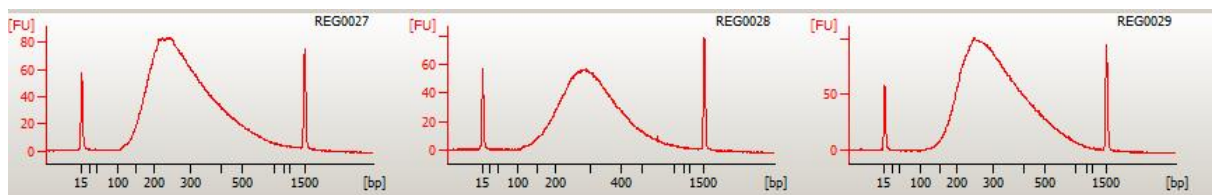
4.4.3 DNA Library Quantification and Pooling

DNA libraries were quantified by quantitative PCR (qPCR) prior to multiplexing (pooling). The KAPA Library Quantification Kit and protocol were used for qPCR quantification using the 7900HT Sequence Detection System (Applied Biosystems) or Mic qPCR Cycler (Bio Molecular Systems). qPCR accurately quantifies the functional DNA fragments in the libraries excluding primer dimers and free nucleotides. The qPCR plate was set up to assess each DNA library in triplicate. This produces 3 values for the concentration of each DNA library. The mean concentration value was calculated and used in subsequent calculations of volumes for DNA pooling. Outliers were removed where appropriate prior to calculating the mean. Once each DNA library was quantified, samples were pooled to achieve a final concentration of 2nM.

The Illumina protocol for 'Preparing libraries for sequencing on the MiSeq' was followed to denature the DNA and dilute the pool to a loading concentration of between 13-16pM for sequencing on a MiSeq machine. The loading concentration was adjusted with serial experiments according to clustering data on the MiSeq for each run (Table 23). Optimal cluster density for the MiSeq V3 reagent chemistry is 1200-1400K/mm2 (https://www.illumina.com/systems/sequencing-platforms/miseq/specifications.html). Table 23 shows that there was some under and over clustering with the MiSeq runs. Despite this, a Q>30 of over 80% was achieved on all runs. The Q score (Quality score) is a Phred based score, where a score of 30 predicts a probability of 1/1000 of an error in base calling. The percentage of bases with a Q score >30 is averaged across the entire run.

Table 23: MiSeq Clustering Data					
MiSeq Run	Loading Concentration (pM)	Cluster Density (k/mm2)	Clusters passing filter (%)	Estimated Yield (MB)	>=Q30*
P001	15	995	93.5	3697.9	3.6G 96.9%
P002	15	1861	67	4506.2	3.8G 83.7%
P003	15	1846	60.8	4052.6	3.3G 82.2%
P004	11	734	94.6	2759.6	2.7G 97.0%
P005	13	740	94.5	2779.9	2.7G 97.5%
P006	13	873	93	3193.5	3.1G 96.4%
P007	13	1037	91.8	3776.2	3.7G 96.6%
P008	13	1066	91.1	3806	3.7G 96.3%
P009	13	675	95.4	2598.2	2.5G 96.3%
P010	13	625	96	2436	
P011	15	944	92.9	3474.5	3.4G 95.8%
P012	15	849	94.4	3217	3.1G 96.5%
P013	15	884	93.8	3342.4	3.2G 96.4%
P014	16	991	92	3670.9	96% 3.5G
P015	16	855	92.5	3152.8	94.8 3G

I carried out some modifications with regards to qPCR and DNA pooling throughout these experiments to improve the MiSeq clustering:

- The concentration of the DNA pool loaded onto the MiSeq was varied as shown in Table 23.
- From experiment P007 onwards, I added a quantification step of the 2nM DNA pool prior to diluting the pool for loading to the MiSeq. The Qubit fluorometer was used for this. If the concentration of the DNA pool was over or under 2nM, I adjusted the volume used to achieve the desired loading concentration. This appeared to improve the cluster density achieved in P007 and P008 experiments.

- From experiment P009 onwards, I used a new qPCR machine, the Mic-qPCR Cycler (Bio Molecular Systems). This appeared to give more consistent results with less variation in the triplicate readings for each sample. The lower cluster densities achieved in P008 and P009 suggested that the previous qPCR machine may have been overestimating the concentration of DNA libraries. Therefore, the loading concentration was increased again from P011 onwards to try to optimise MiSeq clustering.
- For P014 and P015 I added in a further quantification step of the 2nM DNA pool using qPCR in addition to the Qubit fluorometer. The qPCR value was used for adjusting dilution volumes where needed, rather than the Qubit value.

The MiSeq Reagent Kit v3 (150 cycles) was used for all NGS carried out in this project. For runs using the Agilent SureSelect QXT protocol, SureSelect custom primers were combined with the Illumina primers in the MiSeq Reagent Kit as outlined in the protocol and loaded onto the MiSeq reagent cartridge prior to commencing sequencing (following the MiSeq System User Guide).

4.4.4 NGS Coverage and On Target Reads

FASTQ files generated on the MiSeq were processed and analysed using the SureCall program (Agilent Technologies) as described in Chapter 2. For each MiSeq run, NGS data was analysed using the SureCall program to generate a table of variants for each DNA sample. The software also outputs values for coverage and percentage of reads on target for each sample sequenced. This QC data was available for 116 samples sequenced across 15 runs. Although I am presenting the data for 100 cases, as explained previously, a control sample was included on most runs and 5 DNA samples were sequenced twice. Two were repeated due to low percentage of reads covering target regions or lower than 80% of target bases achieving 20x coverage. Three were repeated using the new XT HS kit when I switched the library preparation kit over from the QXT kit so that a comparison could be made between the two kits.

The average percentage of reads covering the target regions across all samples was 60% (Figure 16). This reflects the adequacy of the capture process using the RNA baits. Average read depth in target regions ranged from 42-390 (median 180) for the NGS runs carried out using the SureSelect QXT library preparation kit. For the 6 DNA samples prepared using the XT HS library preparation kit, a high read depth was achieved (329-563) as this is a high sensitivity kit designed for deep sequencing for low frequency variant identification. The NGS of all cases except one (discussed below) achieved the desired minimum 20x coverage of at least 80% of target bases; 93% (108 of 116) of samples had 20x coverage of ≥ 95% of target bases.

Figure 16: Boxplot of percentage of reads covering target regions in each sample
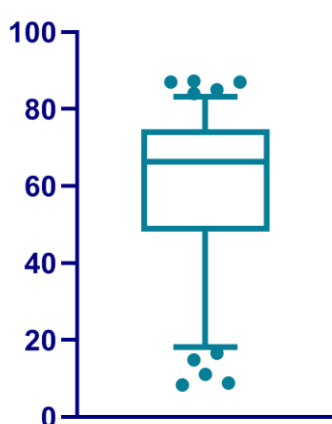
The whiskers reflect the value at 5% and 95% values while the box represents the 25th centile, median and 75th centile values. Dots represent outliers. 8 samples showed issues with targeted capture with less than 20% of reads covering target regions. All of these except one (REG0012) had 20x coverage of >80% of target bases therefore did not require re-sequencing.

Two samples (REG0012 and REG0020) were sequenced twice. For REG0012, this was due to the 20x coverage of 80% of target bases not being reached and for REG0020, the percentage of reads in target regions was also low at 32% but the percentage of target bases with 20x coverage was 85% (Table 24). Although the latter was acceptable

for analysis, the sequencing was repeated to see if these parameters could be improved (NGS runs P003 and P004). The P003 MiSeq run had also showed over-clustering so the loading concentration was reduced for the subsequent P004 run (Table 23). Although this improved the coverage (Table 24) to reach our pre-specified requirement of 80% of target bases at 20x, the percentage of reads covering target regions remained lower than that seen in other samples. This may be due to inherent issues with these particular DNA samples that prevented adequate PCR and probe hybridisation, such as the presence of proteins or contaminants after DNA extraction.

To maximise the data output from both runs, for these two samples my Bioinformatics colleague in the team, Ezequiel Anokian, merged the two fastq files for each sample so that I could analyse these using SureCall, and ensure no variants were missed e.g. variants may not have been called if the number of reads for a variant was too low on the unmerged fastq data. This improved the coverage at 20x for REG0012 (increased from 81.65% to 90.1%) and increased the number of variants called for REG0012. The number of variants called for REG0020 was similar when comparing run P004 and the merged data. No PTVs were called in these two samples.

Table 24: Comparison of runs P003 and P004 for two samples with low % on target reads

Study ID	% reads covering target regions on NGS run P003	% reads covering target regions on NGS run P004	% reads covering target regions on merged fastq files
REG0012	18.7	22.26	19.8
REG0020	31.65	35.98	32.9
	% of target bases at 20x on NGS run P003	**% of target bases at 20x on NGS run P004**	**% of target bases at 20x in merged data**
REG0012	68.44	81.65	90.1
REG0020	85.26	98.14	98.4
	Number of variants P003	**Number of variants P004**	**Number of variants called in merged data**
REG0012	295	371	412
REG0020	366	458	454

4.4.5 NGS Analysis and PTV Identification

The mean number of variants called per case was 478 (range 391-664). By applying the analysis method set at the start of the project, variants called by the program were assigned a category of 1-5. Category 1 variants were expected to be benign and Category 5 variants were expected to be pathogenic or protein truncating. As a result of the pilot experiments carried out prior to the start of sequencing trial samples and with the ongoing analysis, I created a 'rule set' within the SureCall program that could be applied to each set of variants I was analysing. For example, if a non-coding gene transcript was being used by the program to (incorrectly) assign a category 5 to a variant, I would manually change the transcript to a coding gene transcript (usually the longest transcript) which would often downgrade the variant category. RefSeq gene transcripts were obtained from the NCBI (National Center for Biotechnology Information) Gene database. These types of changes were saved into the BARCODE2 'rule set' so that such changes were always applied to future analyses, so that variants were categorised correctly if encountered in subsequent DNA samples (Table 25). The steps taken for variant filtering using the SureCall software are described in Chapter 2.2.7.

The aim of this project was to identify germline Tier 1 PTVs in DNA repair genes as defined in Chapter 2.2.7. A number of genes included in the study panel were not DNA repair genes and were included as candidate genes for research interest only. PTVs in these genes were not evaluated. Table 25 shows examples of DNA repair gene variants that were initially called as a category 5 variant by the SureCall program but on review were downgraded manually.

Table 25: Examples of variants which were recategorised manually during the review of NGS results

Genetic Variant	Reason for change in category
FANCL c.1096_1099dupATTA	Called as category 5 by SureCall program. This is a 4bp insertion in the final exon of the gene. Although it has a low population allele frequency (AF; 0.003 in all populations, 0.0036 in European non-Finnish), as it is in the final exon, it was downgraded and disregarded for trial purposes.
EME2 c.964C>T	Called as category 5 by SureCall program. This is a non-sense variant in the penultimate exon of *EME2*; it is a common variant with an AF in Europeans (Non-Finnish) of 0.014 (All populations 0.008) therefore downgraded as AF is ≥0.01. This variant was also seen in cases and controls with no significant difference in one of the previous Oncogenetics NGS studies. [45]
RAD52 c.806C>A	Called as category 5 by SureCall program. This is a non-sense variant in exon 11 of 12. The AF was above 0.01 in gnomAD (all populations and European Non-Finnish population; 0.016 and 0.012 respectively). This variant was also seen in cases and controls with no significant difference in one of the previous Oncogenetics NGS studies. [45]
FANCD2 c.2715+1G>A	Called as category 3 but on review of variant annotation when using the online CADD tool, I identified this variant is a canonical splice variant with a high CADD score of 30 and population AF of 0.0002, therefore, category upgraded to 5. This particular variant was also reported in a Finnish case-control breast cancer study where it was found to be twice as common in cases compared with controls.[98] It was also reported in 2 Fanconi Anaemia cases that were compound heterozygous for this variant.[99]

Tier 1 protein truncating variants in this study were expected to fit the following criteria:

- Variant is a non-sense, frameshift or splice variant predicted to lead to protein truncation.
- Variant allele frequency is <1% in population databases such as gnomaD.
- CADD score over 20.
- Variant does not lie in the final exon of a gene or in last 50bp of the penultimate exon.

Twenty-one unique PTVs were identified in this study. These PTVs were all assigned a category 5 by SureCall except one which was upgraded manually (*FANCD2* splice variant initially assigned a category 3).

4.4.6 Protein Truncating Variants in BARCODE2 Patients

Next generation sequencing utilising the BARCODE2 gene panel was carried out for DNA samples from 100 patients with advanced prostate cancer. Twenty-one unique PTVs were identified in 22 patients. Of the 21 PTVs identified, 11 are variants that have been reported to be pathogenic or likely pathogenic in the ClinVar database. Two of 22 patients carried 2 PTVs (Table 26 and Table 27). Four patients were found to have an identical variant in the *POLQ* gene. Table 28 summarises the PTVs identified in this study. Half the carriers identified carried a PTV in a HR or DDR gene.

Additionally, 4 patients were found to carry a heterozygous pathogenic missense variant in *MUTYH*. This gene is associated with *MUTYH* Associated Polyposis (MAP), which confers a risk of colorectal cancer when inherited in an autosomal recessive pattern. One of the *MUTYH* variant carriers also carried a PTV in a DNA repair gene (*MRE11A*). All patients found to carry a PTV or other pathogenic variant in a gene that is clinically actionable outside of the study (e.g *BRCA2, PALB2, MUTYH*) were offered a clinical genetic test outside of the trial and subsequent referral to clinical genetics.

Table 26: Summary of Protein Truncating Variants

Gene	Number of unique variants	% of PTVs by gene
BRCA2	4	19.0
ALKBH3	3	14.3
ATM	2	9.5
PALB2	2	9.5
POLQ	1	4.8
BLM	1	4.8
CHEK2	1	4.8
EXO1	1	4.8
FANCD2	1	4.8
LIG4	1	4.8
MRE11A	1	4.8
MSH5	1	4.8
PARP2	1	4.8
PMS1	1	4.8
Total	21	100

Table 27: Summary of PTV Carriers in BARCODE2

Gene(s)	Number of carriers
BRCA2	3
ALKBH3	3
POLQ	2
PALB2	2
POLQ+BRCA2	1
POLQ+ATM	1
ATM	1
BLM	1
CHEK2	1
EXO1	1
FANCD2	1
LIG4	1
*MRE11A**	1
MSH5	1
PARP2	1
PMS1	1
Total	22

*This patient also carried a heterozygous missense *MUTYH* variant which is known to be pathogenic.

Table 28: Protein Truncating Variants Identified by NGS

Gene	Exon	Variant	Classification	rsID	Previously reported?	Variant frequency (gnomAD) and CADD score	Identified in previous ICR Oncogenetics NGS studies	Number of carriers	If not previously reported in a clinical database, literature references where available
ALKBH3	4 of 10	c.208C>T p.Arg70Ter	Nonsense	rs145265812	No	All AF 0.0010 Eur AF 0.002 CADD 40	Yes- in DRG and AEP	1	
ALKBH3	7 of 10	c.381T>G p.Tyr127Ter	Nonsense	rs754599411	No	All AF 1.19×10^{-5} Eur AF 1.76×10^{-5} CADD 38	No	1	
ALKBH3	6 of 10	c.364_365delAG p.Glu123Glyfs	Frameshift	rs368878641	No	All AF 0.0002 Eur AF 0.0004 CADD 35	Yes- 1 young onset case in DRG	1	
ATM	23 of 63	c.3292delC p.Gln1098Argfs	Frameshift	rs1555090075	Yes- ClinVar: Likely Pathogenic	N/A	No	1	
*ATM**	26 of 63	c.3802delG p.Val1268Ter	Non-sense	rs587779834	Yes- ClinVar: Pathogenic	N/A	Yes- in AEP 1 aggressive case	1	

Gene	Exon	Variant	Classification	rsID	Previously reported?	Variant frequency (gnomAD) and CADD score	Identified in previous Oncogenetics NGS studies	Number of carriers	If not previously reported in a clinical database, literature references where available
POLQ	16 of 30	c.4262_4268delTA CTATT p.I1421Rfs	Frameshift	rs546221341	No	All AF 0.0057 Eur 0.0078 CADD 34	yes- in AEP, 2 non-aggressive cases	4	Reported in a WES study in a BC patient with a positive cancer family history [100]
BLM	14 of 22	c.2695C>T p.R899Ter	Nonsense	rs587779884	Yes- ClinVar: Pathogenic	All AF 6.37×10^{-5} Eur 1.39×10^{-4} CADD 36	No- but a different variant observed in AEP (frameshift)	1	Conflicting reports of association with cancer risk in carriers including early onset CRC [101, 102]
BRCA2	14 of 28	c.7360delA p.Ile2454Phefs	Frameshift	rs80359646	Yes- ClinVar: Pathogenic	N/A	No	1	
BRCA2	11 of 28	c.5217T>A p.Tyr1739Ter	Nonsense	rs80358746	Yes- ClinVar: Pathogenic	N/A	No	1	

Gene	Exon	Variant	Classification	rsID	Previously reported?	Variant frequency (gnomAD) and CADD score	Identified in previous Oncogenetics NGS studies	Number of carriers	If not previously reported in a clinical database, literature references where available
BRCA2	11 of 28	c.2834_2835delAA p.Lys945Argfs	Frameshift	rs80359356	Yes-ClinVar: Pathogenic	N/A	No	1	
BRCA2*	11 of 28	c.4914_4915delAG p.Lys1638Asnfs*3	Frameshift	N/A	No-Clinically pathogenic variant (c.4914dupA) reported at same genomic position	N/A	No	1	
CHEK2	12 of 16	c.1100delC p.T367Mfs	Frameshift	rs555607708	Yes-ClinVar: Pathogenic	All AF 0.0021 Eur 0.0026 CADD 35	Yes- in DRG, more frequent in cases than controls	1	
EXO1	9 of 15	c.1241C>A p.S414Ter	Nonsense	rs543887227	No	All AF 2.1×10^{-5} Eur 4.67×10^{-5} CADD 35	No	1	
FANCD2	intron 28 of 42	c.2715+1G>A	Splice site variant	rs201811817	No	All AF 1.7×10^{-4} Eur 2.56×10^{-4} CADD score 30	Yes- in one non-aggressive case in AEP study	1	Reported in a Finnish BC case control study and in 2 cases of FA.[98] [99]

161

Gene	Exon	Variant	Classification	rsID	Previously reported?	Variant frequency (gnomAD) and CADD score	Identified in previous Oncogenetics NGS studies	Number of carriers	If not previously reported in a clinical database, literature references where available
LIG4	2 of 2	c.1271_1275delAAAGA p.K424Rfs	Frameshift	rs772226399	Yes- ClinVar: Pathogenic	All AF 1.6x10-4 Eur 2.1x10-4 CADD 35	Yes- DRG and AEP	1	
*MRE11A^*	10 of 20	c.1090C>T p.Arg364Ter	Nonsense	rs371077728	Yes- ClinVar: Pathogenic	All AF 4.95x10^{-5} Eur 1.55x10^{-5} CADD 45	No	1	
MSH5	19 of 25	c.1744delG p.D583Tfs	Frameshift	rs766324482	No	1.21x10^{-5} Eur 2.67x10^{-5} CADD 35	Yes- in AEP in 1 non-aggressive case	1	*MSH5* identified as possible predisposition gene in case control DRG study [45]
PALB2	4 of 13	c.1546delA p.Arg516Glufs	Frameshift	rs587781560	Yes- ClinVar: Pathogenic	N/A	No	1	

Gene	Exon	Variant	Classification	rsID	Previously reported?	Variant frequency (gnomAD) and CADD score	Identified in previous Oncogenetics NGS studies	Number of carriers	If not previously reported in a clinical database, literature references where available
PALB2	10 of 13	c.3113G>A p.Trp1038Ter	Nonsense	rs180177132	Yes-ClinVar: Pathogenic	All AF 6.0×10^{-5} Eur 1.1×10^{-4} CADD 46	No	1	
PARP2	15 of 16	c.1480C>T p.Q481Ter	Nonsense	rs1128782	No	All AF 1.8×10^{-5} Eur 3.9×10^{-5} CADD 45	No- but other PTVs (exon 11) identified in DRG study.	1	Literature: missense variants described in PrCa families
PMS1	10 of 13	c.1360C>T	Nonsense	rs139932286	No	AF 1.39×10^{-4} Eur 6.24×10^{-5} CADD 39	No	1	

BC= Breast cancer CRC= colorectal cancer FA= Fanconi Anaemia; DRG= DNA Repair Gene Study, Leongamornlert et la 2019 [45]; AEP= Advanced Exome sequencing Project, Mijuskowic et al 2018 [77]; PrCa= prostate cancer, WES= Whole exome sequencing

Allele frequency data for Europeans is from the European Non-Finnish population in gnomAD.

*These carriers also carried one of the *POLQ* variants listed.
^The carrier of this *MRE11A* variant also carried a *MUTYH* variant (c.536A>G)

Table 29: Clinically Pathogenic Missense Variants

Gene	Exon	Variant	Classification	rsID	Reported in ClinVar?	Identified in previous Oncogenetics NGS studies	Number of carriers
MUTYH	7 of 16	c.536A>G p.Tyr179Cys	Missense	rs34612342	Yes- ClinVar: Pathogenic	Y- in DRG study in 3 cases	3
MUTYH	13 of 16	c.1187G>A p.Gly396Asp	Missense	rs36053993	Yes- ClinVar: Pathogenic	Y- in DRG study in 15 cases	1

4.4.7 Validation of NGS variants by Sanger sequencing

All PTVs were successfully validated by Sanger sequencing. For each variant, appropriate primers were designed that would allow PCR amplification of a ~200-400bp sequence encompassing the variant of interest. (Primers listed in Table 30) For each PTV that required validation, a fresh aliquot of DNA was used from the stock DNA in addition to an aliquot from the previously diluted DNA used for library preparation and NGS. Negative controls were included in each run as well. Prepared DNA samples were sequenced on a 3730XL machine in the ICR and sequences were analysed using the Mutation Surveyor program. All NGS identified PTVs that underwent Sanger sequencing were successfully validated as well as the pathogenic missense *MUTYH* variants. Figure 17 shows the Sanger sequencing traces for the validation of the *FANCD2* variant identified in one of the BARCODE2 samples. A 2 stage PCR was needed for successful validation of this particular variant which is detailed in the next section.

4.4.8 Validation of *FANCD2* variant by a 2 stage PCR prior to Sanger sequencing

Of the 21 PTVs identified, a splice loss variant in *FANCD2* was found in one case: c.2715+1G>A. *FANCD2* is a DNA repair gene involved in the HR DNA repair pathway. Although this variant is not reported in ClinVar, it is described in the literature associated with 2 cases of Fanconi Anaemia where the patients were compound heterozygous for this variant [99]. This variant was also reported in a Finnish breast cancer case control study where it was found to be twice as common in cases compared with controls. [98] The population allele frequency is 0.0002 in the gnomAD database and CADD score is 30.

Primers were designed to Sanger sequence this variant but upon review of sequencing traces, other variants were identified that were not seen on the NGS reads (Figure 1). This led to a low peak of the sequencing trace representing the variant under investigation and although visible on manual review, the variant was not called by the Mutation Surveyor software. This led to a suspicion that the DNA sample used for Sanger sequencing may have been contaminated. The Sanger sequencing was repeated with a fresh aliquot of DNA, but the same results were observed with 'additional' variants appearing that were not seen on NGS (NGS reads were reviewed on the IGV in SureCall to ensure these were not missed). On review of the negative controls included in the Sanger sequencing, it was noted that these 'additional' variants were also present in these samples but without the splice site variant being present.

On review of this genomic region in the UCSC genome browser, it became apparent that the *FANCD2* variant was lying within a ~17Kb DNA sequence (chr3:10099092-10116035) that is duplicated 1.8Mb downstream on the same chromosome (chr3:11919724-11933728); this region lies within the *FANCD2 Pseuodogene2 (FANCD2P2*; Figure 1 and Figure 19). A pseudogene is characterized by high sequence similarity with the corresponding functional gene, therefore inaccurate mapping in the analysis of NGS data can occur. This may lead to false positive variant calls on NGS reads or as occurred in this project, difficulties with validation due to off target sequencing when carrying out Sanger sequencing. The primers designed to Sanger sequence the variant of interest were also leading to sequencing of the duplicate region within *FANCD2P2* and the subtle variation of nucleotides in these two regions led to false positive variants being seen on the sequencing traces and subsequent 'weakening' of the trace representing the splice variant c.2715+1G>A.

To overcome this, and achieve clear Sanger sequencing traces for validation of the variant of interest, I planned a 2 stage PCR where the first stage was a long range PCR

aimed to amplify the region encompassing the splice variant within a sequence of length 2-3Kbp spanning both part of the duplicated sequence and the adjacent non-duplicated region. This required the design of a new set of primers encompassing this region. To amplify a longer sequence than usual by PCR, the PCR extension time was lengthened from 1 to 3 minutes. The long range PCR product was then used as the input DNA for the subsequent standard PCR and Sanger sequencing, using the original primers designed to amplify a 250bp region encompassing the splice variant. This approach was successful in sequencing the region of interest only and the resulting sequencing traces showed the heterozygous splice variant clearly and the variant was called by the Mutation Surveyor software. The 'additional' variants observed on the first set of Sanger sequencing were not seen and were in keeping with the NGS reads in this region.

Figure 17: Sanger Sequencing Traces of *FANCD2* c.2715+1G>A
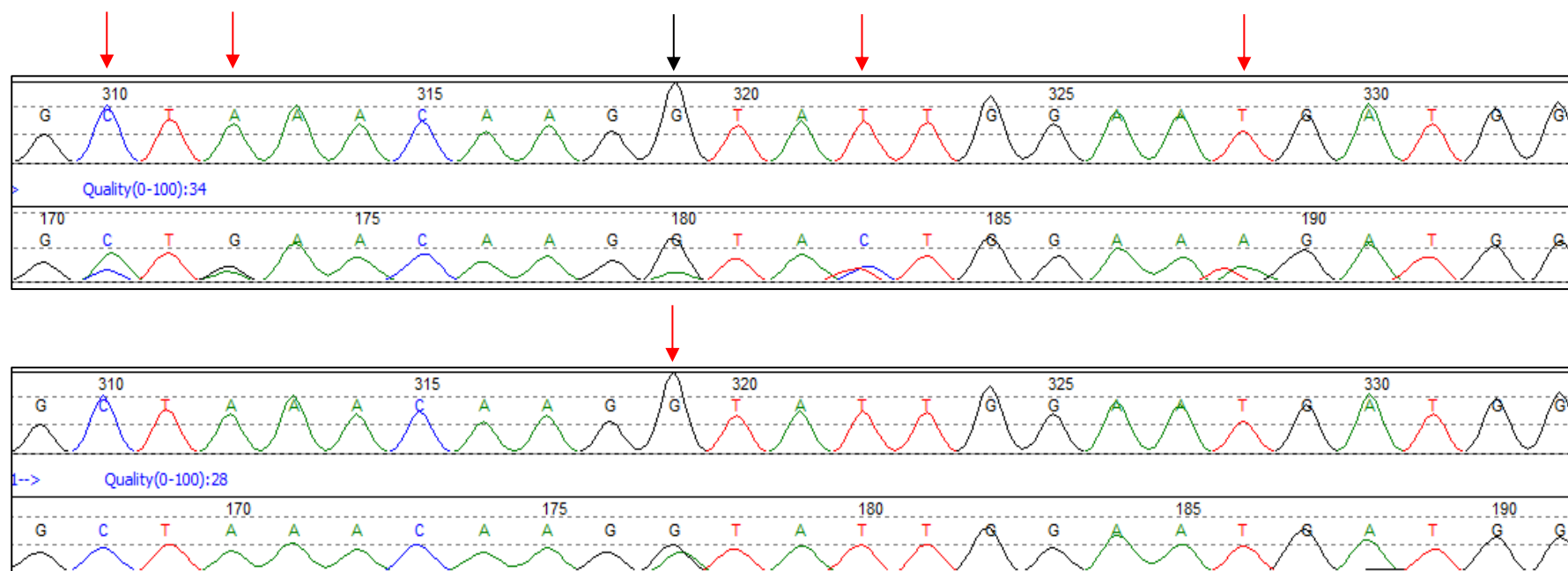
The upper panel of traces shows the reference *FANCD2* sequence with the variants called seen on the trace below the reference; these are annotated with red arrows and are in fact false positive variants due to sequencing of the *FANCD2* pseudogene. The black arrow indicates the splice variant of interest which can be observed with a weak peak for the alternative allele which was not called by the Mutation Surveyor software.

The lower panel of traces shows the reference sequence and the Sanger sequencing trace below when using the long range PCR product as the input material. The variant was called by the software (indicated by the red arrow) and the variants seen above were not called indicating the absence of off target sequencing of the *FANCD2* pseudogene.

Figure 18: Part of FANCD2 Gene

The red arrow denotes the position of the splice variant c. 2715+1G>A. The grey track at the bottom of the figure shows that the variant lies within a 17Kb sequence that is duplicated elsewhere in the genome (1.8Mb downstream).
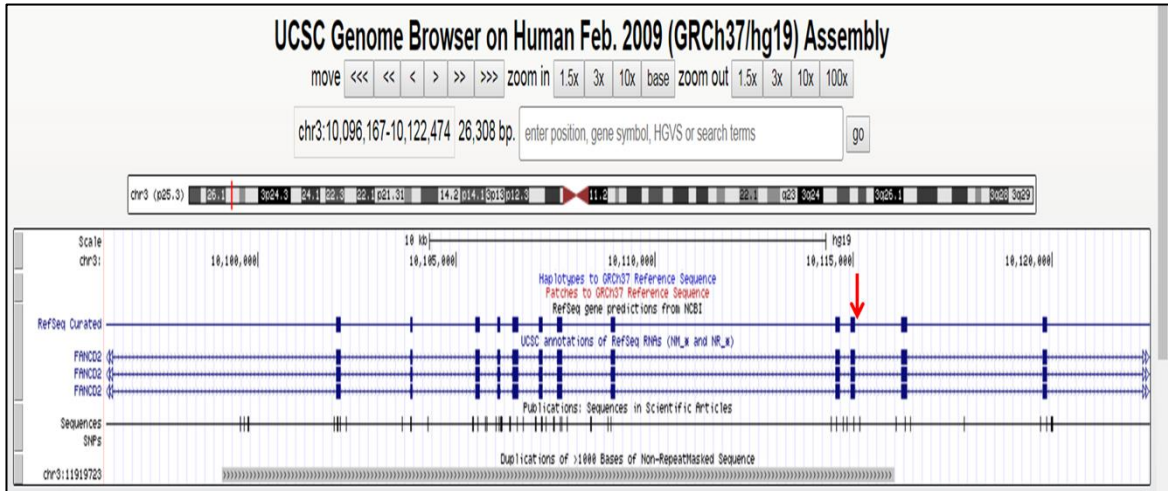
Figure 19: *FANCD2* and P1 and P2 Pseudogenes
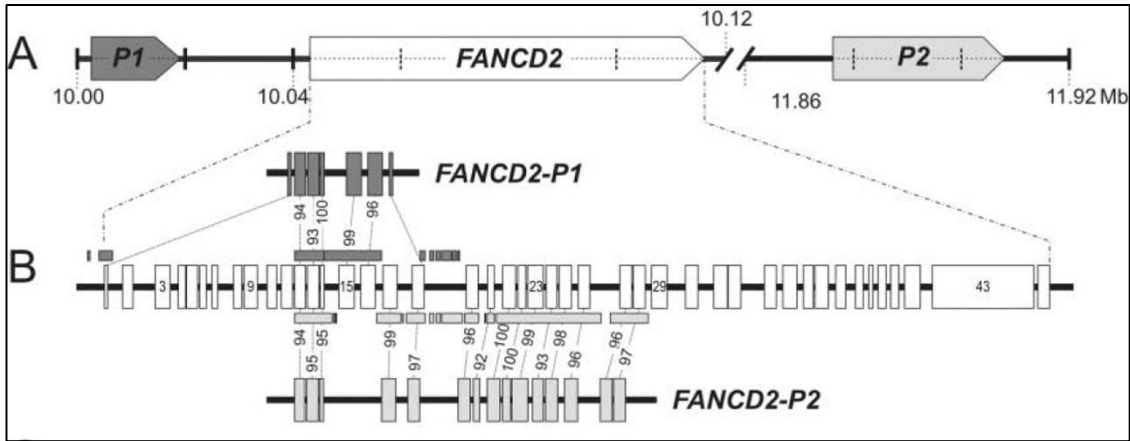
A. The two pseudogenes—*FANCD2-P1* and *FANCD2-P2*—located upstream and downstream, respectively, of the functional *FANCD2* gene. All three have the same orientation. The scale denotes Mb on chromosome 3.

B. FANCD2 exons and their pseudogene equivalents, connected by dashed lines, indicating percentages of nucleotide identity. Homology also extends into many introns nearby, as indicated by the boxes beyond and below the active gene.

Table 30: PCR Primers used prior to Sanger sequencing

Gene Variant	Forward primer	Reverse primer
BRCA2 exon 11 c.5217T>A	TGATGGTCAACCAGAAAGAA	AGTTTGTGGGTATGCATTTG
BRCA2 exon 11 c.2834_2835delAA	CCAAGTAGCTAATGAAAGGA	ATTTGAAGCTGTTCTGAAGC
BRCA2 exon 11 c.4914_4915delAG	TTCTATTGAGACTGTGGTGC	TCAATGACTGAATAAGGGGAC
BRCA2 exon 14 c.7360delA	CCATGTAGCAAATGAGGGTC	CCCCTTTGGTGGTGGTAATT
EXO1 exon 9 c.1241C>A	TGGGATGACAAAACATGTCAA	gaaaatcttcgcgactttgc
PMS1 exon 10 c.1360C>T	ctatgcccggccaataat	TTGGGACTGAAGGAGTTC
MSH5 exon 19 c.1744delG	ggtggaggaatagacatgag	cacggaaagttccattagag
CHEK2 exon 10 c.1100delC	ttaatttaagcaaaattaaatgtcc	ggcatggtggtgtgcatc
MRE11A exon 10 c.1090C>T	aaaggagcattacaagaagg	cgatggtgattgctcttc
ALKBH3 exon 4 c.208C>T	tggagacaagggtctctagtgg	tggatatcttgttgcattttca
ALKBH3 exon 7 c.381T>G	tcccatatcttctgtgaagg	taggaaccagcatgaatctc
ALKBH3 exon 6 c.364_365delAG	tagaggctgcatccagatta	tgtgacctctgggatctact
FANCD2 intron 28 long range PCR (2481bp product)	TGGCATCAGTAATTGGAACA	CACTAGTCCTTGGTTCAGAC
FANCD2 intron 28 c.2715+1G>A standard PCR	ctctaggcagtttccaacag	cgagaataaagctgggtttc
POLQ exon 16 c.4262_4268delTACTATT	GTGTCACATTCCTTTTCCTG	CAGTGGGAGTTCTCTTTTGA
PARP2 exon 15 c.1480C>T	ctgatgggattttctgtttg	tctagtcctcctggttctga
BLM exon 14 c.2695C>T	gtgtggtcttccagcagtat	tctacatgtgcatgtttggt
ATM exon 26 c.3802delG	ggtggtggtatgttctaagc	CAAGAATCTTTGGAAAGCAG
ATM exon 23 c.3292delC	gttctggaatatgctttgga	gcaagcatatgataacagca

170

LIG4 exon 2 *c.1271_1275delAAAGA*	CTGAGAAAGAGGTATGAGATTC	CCACTGACATACTCTGGTTT
MUTYH exon 13 c.1187G>A	agtggcatgagtaacaagag	ctattccgctgctcacttac
MUTYH exon 7 c.536A>G	ctagggtaggggaaatagga	AGCTCCTCTACCACctgatt
PALB2 exon 4 c.1546delA	TTTCCAATGAGGAAACTGAC	aagtgccaggcaaatagtaa
PALB2 exon 10 c.3113G>A	tacagAGGCAAAGAAAACCA	agcaacacaaaaccacaatc

4.4.9 Clinical Characteristics of PTV Carriers

There was no significant difference in the age of patients at the time of initial prostate cancer diagnosis when comparing carriers with non-carriers (Table 31). But a significant difference was observed for the age of patients at the time of CRPC onset. This was significantly lower in carriers with a mean of 62.5 years compared with non-carriers at 66.2 years; P=0.03 (Figure 20). This correlates with the significant difference in mean age of carriers at the time of study entry which was 5 years younger than non-carriers (63.9 vs 68.5; P=0.009). The interval between initial diagnosis and CRPC onset was also 2 years shorter in carriers (mean 41.5 months vs 64.5 months; P=0.044).

Table 31: Age at Diagnosis and CRPC

Clinical Variable		Carriers	Non-Carriers	P value	Statistical test
Age at initial prostate cancer diagnosis	Mean (SD)	59 (5.7)	60.6 (6.1)	0.27	*Unpaired t test*
Age at castration resistant disease diagnosis	Mean (SD)	62.5 (6.3)	66.2 (7.0)	0.028	*Unpaired t test*
Interval (months) between initial diagnosis and CRPC onset	Median (range)	20.0 (1-129)	45.5 (6-262)	0.044	*Mann-Whitney U test**

*Mann-Whitney U test used as data did not pass test for normal distribution

Figure 20: Age at initial prostate cancer diagnosis (left) and at CRPC diagnosis (right)
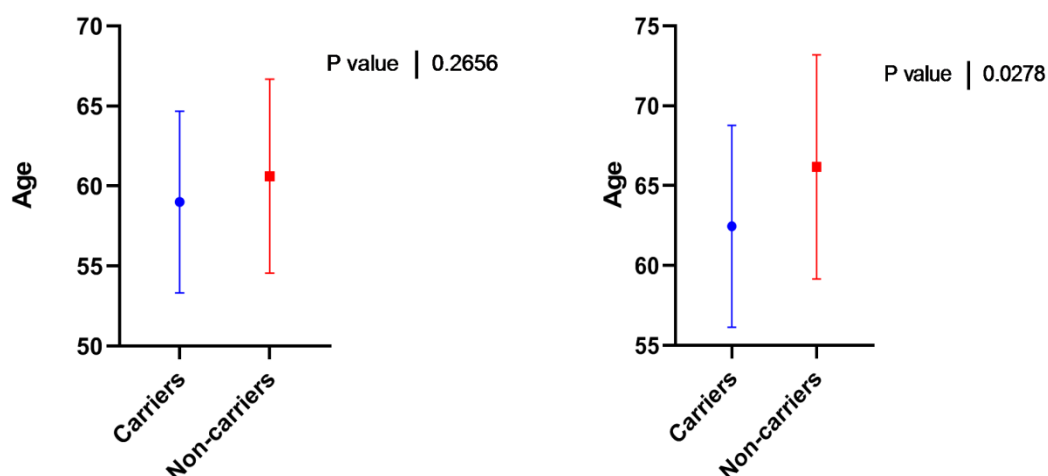
Data were available for the Gleason grade of prostate cancers at initial diagnosis for 84 of the patients in this study and are shown in Table 32. A similar proportion of patients in each group were diagnosed with high grade (≥8) disease. Presenting PSA (pPSA) is the PSA level at the time of initial prostate cancer diagnosis; these data were available for 94 patients. Although the mean pPSA was higher for the carrier group, this was not statistically significant (Table 33).

Table 32: Prostate Cancer Grading at initial diagnosis

	Carriers	%	Non-Carriers	%
Gleason grade	N=16		N=68	
Gleason <=6	1	5.6	1	1.5
Gleason 7	3	16.7	24	36.9
Gleason >=8	12	66.7	42	64.6

Table 33: Presenting PSA at Initial Diagnosis

Presenting PSA	Carriers (N=20)	Non-Carriers (N=74)	P value	Statistical test
Median (range)	51 ng/ml (5-5000)	33 ng/ml (4-3000)	0.31	*Mann-Whitney U test*

The Gleason grade of a prostate cancer in addition to the pPSA and presence of nodal metastases are variables used to stratify risk of disease recurrence at the time of staging a newly diagnosed patient. In this set of patients, there was no significant difference between carriers and non-carriers when examining the presence of nodal or distant metastases at presentation though metastatic disease at baseline was observed in 60% of carriers compared with 40% of non- carriers which may suggest an association with carrier status.

Table 34: Metastatic disease status at initial diagnosis

	Carriers	%	Non-Carriers	%	Statistical test	
	N=22		N=77*		P value	
Nodal metastases	9	50	26	43	0.79	*Fisher's Exact test*
Metastatic disease at baseline	13	59	31	40	0.15	*Fisher's Exact test*

*For nodal metastatic status, data were available for 60 non-carriers but distant metastatic disease data were available for all 77 non-carriers.

4.4.10 Cancer Family History in PTV Carriers

Of the 22 PTV carriers, 19 (86%) had a family history of cancer as did 57 (74%) of 77 non-carriers with family history information. Details of family history of prostate cancer (in any relative) and breast and ovarian cancer (in first degree relatives) are shown in Table 35. The proportion of carriers with a family history of breast and ovarian cancer was larger in carriers (29%) than that of the non-carrier group (16%).

Table 35: Family history of Cancer in Carriers and non-Carriers

	Carriers	%	Non-Carriers	%
Family History of:	N=22		N=77	
Any cancer	19	86	57	75
Prostate cancer	6	27	29	38
Breast/ Ovarian cancer	6	29	12	16

4.4.11 SNP Genotyping Results for mCRPC Patients in the BARCODE2 Trial

Germline DNA from the first 100 patients in the BARCODE2 trial was sent off for genotyping using the assay developed for the BARCODE1 trial. Genotyping data were received for 98 samples. Two samples failed the QC processes at Thermo Fisher due to sample call rate falling below 90% (81.5% and 86.4%) therefore genotyping data were not received for these samples.

Cluster plots were reviewed by me using EAS and genotyping calls were modified where appropriate as described in Chapter 2.1.5. After this QC step was completed, genotyping data were exported from EAS and input into the PRS Shiny App developed by the Oncogenetics team. This application carries out the PRS calculation after the QC steps described in Chapter 2.1.5.1, which are summarised as follows:

- Missing genotypes were replaced by 2xrisk allele frequency (RAF); for chromosome X SNPs 1xRAF was used.
- SNPs with a call rate of <90% were excluded
- Samples with a call rate of <90% were excluded

At the end of this process, a PRS was successfully calculated for all 98 samples and I was able to compare the distribution in the BARCODE2 cohort to the BARCODE1 pilot set as well as the reference sets described in Chapter 3. Additionally, 2847 samples from prostate cancer patients in the UKGPCS (UK Genetic Prostate Cancer Study) have undergone genotyping using the same assay and therefore their PRS distribution could also be compared.

Table 36 summarises the PRS distribution in the 5 data sets. The median PRS in the two prostate cancer cohorts (UKGPCS and BARCODE2) is similar and higher than that observed in the 'healthy' populations of BARCODE1 pilot, ProtecT controls and Nottingham controls. This is clearly observed on the normal distribution curves for each

set where the BARCODE2 and UKGPCS curves are shifted to the right. The t-test comparing the means between the BARCODE2 set and the 4 other cohorts shows a significant difference only when compared with the BARCODE1 pilot and 2 'healthy' control groups. (Figure 21)

Table 36: PRS Distribution in patient cohorts from 5 studies

Batch	N	Median PRS	Range	90th percentile PRS	T-test*
BARCODE2	98	10.80	9.30-12.55	11.65	
BARCODE1 Pilot	285	10.35	8.42-12.21	11.15	P<0.0001
Nottingham Controls	500	10.13	8.02-12.02	10.91	P<0.0001
ProtecT Controls	2571	10.06	7.96-12.56	10.90	P<0.0001
UKGPCS	2847	10.83	8.45-13.89	11.79	P=0.16

Figure 21: Distribution of PRS in 5 Different Populations
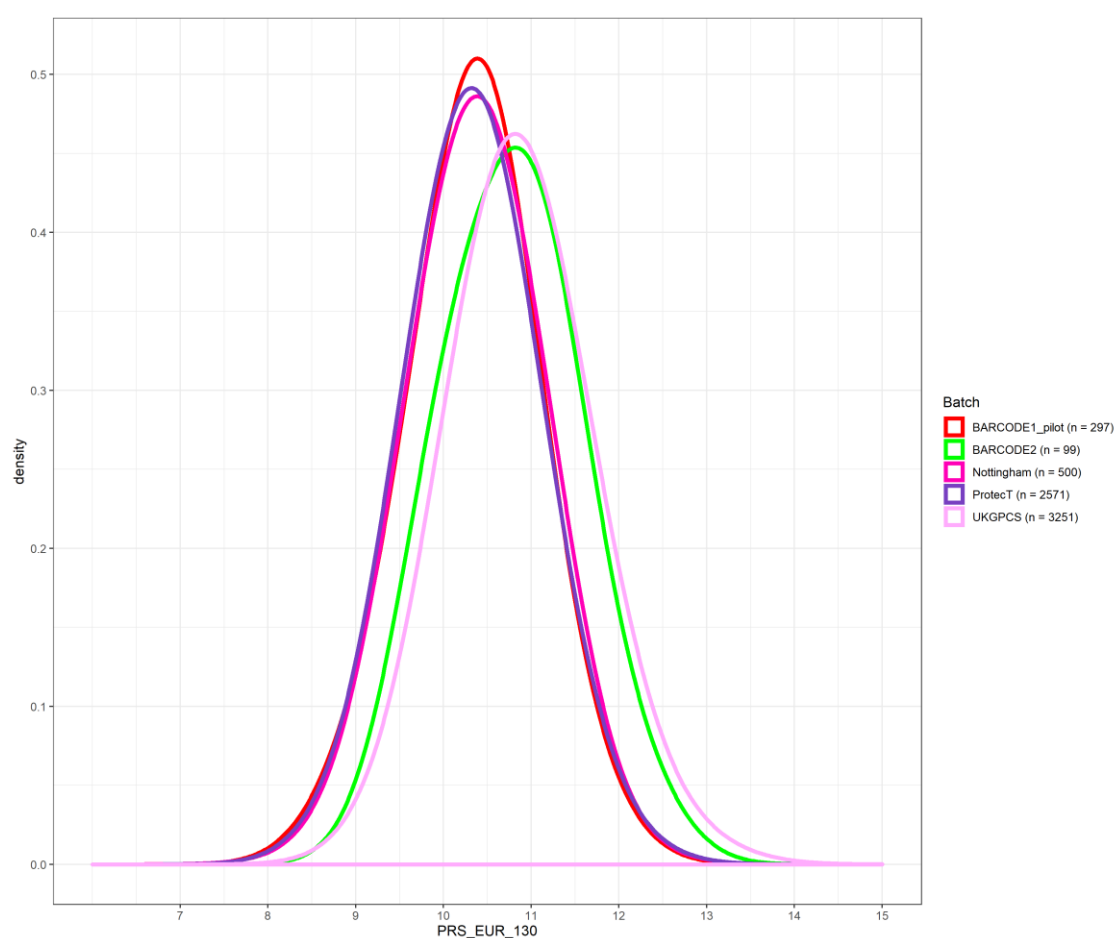
In the BARCODE1 pilot study, a PRS value of ≥11.15 was used to identify men for prostate cancer screening. Within the BARCODE2 set of patients, 28.7% had a PRS over this value. If using the 90th percentile threshold (10.90) from the ProtecT and Nottingham controls, 46% of men have a PRS above this value.

4.5 Discussion

I carried out the NGS for the first 100 mCRPC patients recruited to the BARCODE2 trial using a study specific gene panel. The frequency of PTV carriers was 22% with the most frequently aberrant gene being *BRCA2* (4% of all patients; 19% of all unique PTVs). This fits with the findings of other mCRPC NGS studies where the *BRCA2* carrier frequency is in the range of 5-6%. In BARCODE2, carriers of *BRCA2* or *ATM* variants made up 6% of study participants, similar to that reported by Na et al and Pritchard et al [30, 41]. Of the 22 carriers identified, half harboured a germline variant in a HR or DNA damage response (DDR) gene. This is notable in view of the recently reported data related to PARPi responses observed in prostate cancers associated with HR gene defects.[42, 97, 103]

Several studies have indicated that carriers of pathogenic variants in DNA repair genes, particularly *BRCA2* and *ATM*, are more likely to present with advanced disease features such as nodal involvement or distant metastases. In this study, 59% of carriers had distant metastatic disease at presentation compared with 40% of non-carriers (P=0.15). The natural history of prostate cancer progression leads to a final stage of castration resistance, which is associated with metastatic disease development and a median survival time of 22 months. [104] In the BARCODE2 set of men, carriers of DNA repair gene PTVs had a significantly shortened time interval between initial prostate cancer diagnosis and the development of CRPC, and were diagnosed with CRPC at a younger age than non-carriers. This finding reflects the poorer outcomes of carriers of DNA repair gene variants reported elsewhere.

4.5.1 Novel Variants in mCRPC Patients

The panel of genes used in this study included HR DNA repair genes as well as genes involved in other DNA repair pathways. The second most commonly aberrant gene in this set after *BRCA2* was *ALKBH3*. This gene encodes a protein involved in the repair

of DNA damaged by alkylation and preferentially targets single stranded DNA. *ALKBH3* has been implicated in prostate cancer development as this gene is overexpressed in prostate cancer cells but not in benign prostatic hyperplasia or in normal prostate epithelium [105]. Expression of the ALKBH3 protein appears to correlate with castration resistance; in the hormone-independent DU-145 cell line, *ALKBH3* knockout leads to reduced tumourigenicity and reduced anchorage independent growth. *In vivo* xenograft models also showed this pattern of relationship between *ALKBH3* expression and tumour growth.[105] These findings have suggested that ALKBH3 could be targeted for treatment and *in vitro* as well as *in vivo* studies of molecular inhibitors have been reported.[106, 107]

As ALKBH3 is involved in the repair of alkylated DNA, tumours with aberrant ALKBH3 function may be sensitive to alkylating drugs such as platinum. In fact, sensitivity to alkylating agents used to treat brain glioma has been shown to be related to ALKBH3 inhibition; in this context, the alkylating agent sensitivity is secondary to *IDH* (Isocitrate dehydrogenase) mutations in the tumour.[108] The sensitisation of glioma cells to alkylating agents in this setting is dependent on the aberrant IDH producing D-2-hydroxyglutarate which is a competitive inhibitor of α-ketoglutarate (αKG). ALKBH3 mediated DNA repair depends on αKG. *In vitro* studies of glioma cell lines has shown that sensitisation to alkylating agents can be reversed by deleting mutant *IDH* or overexpressing *ALKBH3*.[108] Whether such findings could be extrapolated to prostate cancer treatment is still unknown although the *in vitro* and xenograft model data reported do suggest promise.

A recurrent *POLQ* variant was identified in 4 patients, rs546221341, which was unexpected as the gnomAD population frequency of this allele is 0.0057 (0.0078 in the European non-Finnish population). As this variant is predicted to be deleterious causing a frameshift deletion in exon 16 and has a CADD score of 26, it was deemed to be

actionable within the trial and qualify the patients for carboplatin treatment. Two of the 4 carriers also carried a pathogenic variant in another DNA repair gene: *ATM* and *BRCA2*.

POLQ belongs to the DNA polymerase family of genes (15 in total) and functions to promote microhomology-mediated end-joining (MMEJ), an alternative non-homologous end-joining (NHEJ) machinery triggered in response to double-strand breaks in DNA. It has been implicated in repairing DNA damaged by ionizing radiation as well as repairing inter-strand cross links in DNA[109]. On review of the published literature, this variant was reported in a study that compared 2 cancer gene panels with whole exome sequencing (WES) in cancer cases with a positive family history; rs546221341 was identified by WES in a breast cancer patient whose 2 sisters had been diagnosed with a breast and endometrial cancer respectively. In the 4 BARCODE2 cases, 3 of the 4 carriers had a family history of cancer (the man who also carried a *BRCA2* variant had no cancer family history), 2 of whom had a first degree relative diagnosed with breast cancer (one of the *POLQ* only carriers and the *ATM/POLQ* carrier). A Brazilian breast cancer case control study reported a rarer *POLQ* variant, rs581553, that was found to confer an odds ratio of breast cancer of 5.67 in the study population. [110] These findings warrant further investigation of *POLQ*'s association with cancer development and functional studies would be required to analyse this further.

In view of the higher than expected carrier frequency of the *POLQ* variant in the BARCODE2 set of patients, and to investigate a possible association of rs546221341 with prostate cancer, I was able to check the frequency of this variant in the meta-analysis data and OncoArray genotyping data used to identify 63 novel prostate cancer risk loci (reported by Schumacher et al in 2018)[13]. The allele frequency in prostate cancer cases and controls in the OncoArray set were very similar at 0.0068 and 0.0073 respectively and in keeping with the European population frequency in gnomAD. In an

analysis of advanced prostate cancer cases and non-advanced cases genotyped on the OncoArray, there was also no significant difference observed in allele frequency although this comparison is likely to be underpowered. In order to study the possibility of an association with advanced or metastatic prostate cancer specifically, a larger study would be needed which includes both low risk and high risk or metastatic cases. The ICR Oncogenetics team is involved in several ongoing international collaborations utilising whole exome or whole genome sequencing which may provide the data needed to answer such questions.

Variants in other DNA polymerase genes have been identified in other studies by the ICR Oncogenetics team. In the case-control study reported by Leongamornlert et al [45], two of the *POL* genes were found to be significantly associated with prostate cancer predisposition, *POLE* and *POLM.* DNA polymerases have important functions in DNA replication, repair and the tolerance of DNA damage, so it is feasible that hereditary variants in some of these genes may predispose to cancer or promote carcinogenesis in certain contexts. Further research is needed to analyse the differences between the *POL* genes and their association with cancer predisposition.

4.5.2 Prostate Cancer PRS in BARCODE2

The genotyping results and PRS distribution in the BARCODE2 set of men with advanced prostate cancer was similar to the UKGPCS prostate cancer patients and as expected, both showed a higher median PRS compared with the non-cancer comparator groups.

Whether a PRS based prostate cancer screening programme would lead to earlier diagnoses that would improve outcomes for men such as those recruited to BARCODE2 is unknown. Although 28-46% of men in the BARCODE2 set would qualify for screening in a hypothetical BARCODE1 type screening programme (due to their PRS measuring

≥10.90-11.15), 9 men were diagnosed aged younger than the minimum age for screening (55 years) in that protocol. It is likely that additional parameters are needed in addition to genetic profiling for a screening model that would commence at a younger age e.g. 40-45 years in order to benefit such cases. For *BRCA2* carriers, it is known that a PRS based on the known risk loci modifies prostate cancer risk.[111] The risk of prostate cancer by age 80 years for those at the 5th percentile of the PRS distribution is 19% and for those at the 95th percentile it is 61%. [111] Utilising a SNP profile for at risk *BRCA2* carriers may inform approaches to prostate cancer screening. The ongoing IMPACT screening study (NCT00261456) is measuring the PRS of carrier participants enrolled in the study and these results will guide approaches to screening in this population of men.

Utilising MRI as an initial triage test during screening as carried out in the PROMIS study (discussed in Chapter 1.4.2) could stratify patients for definitive biopsy procedures when there is a sufficient level of suspicion of disease of presence.

Results from the BARCODE1 main study as well as the ongoing PROFILE study (NCT02543905) will help guide on the optimal use of a genetic profile test for prostate cancer screening.

4.5.3 Limitations

The number of patients sequenced in this study was too small to sufficiently examine differences in clinical factors, tumour staging and family history. For recurrent variants such as the *POLQ* variant discussed above, a case control study would be required to investigate association with disease development or with high risk disease.

The significance of the germline variants identified in genes such as *ALKBH3, PARP2* and *BLM* is unclear. Research in prostate cancer and the association with DNA repair gene mutations so far has focussed on HR DNA repair genes which predict for

sensitivity to PARP inhibitors and potentially platinum chemotherapy. Functional studies of these novel variants in genes involved in non-HR DNA repair pathways would be informative but unfortunately these couldn't be carried out within this project.

4.5.4 Conclusions

I identified a higher than expected frequency of PTV carriers in the first 100 men recruited to the BARCODE2 trial. *BRCA2* was the most commonly aberrant gene in this set, similar to other reported datasets. NCCN Guidelines related to prostate cancer genetics have recently been updated to recommend the testing of *BRCA2* in the germline for men with metastatic prostate cancer regardless of their family history.

In this small set of mCRPC patients, I found that carriers had a shortened interval to CRPC onset compared with non-carriers which is also in keeping with other retrospective studies' that have reported poorer outcomes in carriers of DNA repair gene variants. This suggests that such patients may benefit from genetic screening at an early stage before metastatic disease development to identify those at high risk of disease progression. Such patients may be offered closer follow up or a modified/ intensified treatment pathway.

Chapter 5   BARCODE2 Trial Tumour DNA NGS

5.1 Introduction

In Chapter 4, I described the results of germline sequencing for 100 patients with advanced mCRPC; 22% of patients were found to carry a PTV in a DNA repair gene with 11% of (all sequenced) patients carrying a PTV in a HR gene. The presence of these germline variants in a significant proportion of men in the study does not establish a causal link to prostate cancer development. To truly study a possible causal association, a well powered case control study would be needed. The observation that the BARCODE2 PTV carriers had a significantly shorter interval to the development of mCRPC than non-carriers is notable and could suggest an impact of a germline variant on disease phenotype- a case-case analysis including patients with aggressive and non-aggressive disease would be needed to analyse this further.

Prostate cancer studies that have performed paired tumour and germline sequencing have shown second hit somatic variants in cases with identified germline variants. In a germline NGS study of nearly 700 advanced prostate cancer cases reported by Pritchard and colleagues, somatic NGS data were available for 61 patients.[30] In 36 cases (59%), the somatic second allele of the gene affected in the germline was aberrant either with another loss of function variant or gene copy number loss.

In one of the first reports of prostate cancer responses to the PARPi olaparib, 3 *BRCA2* germline mutation carriers who responded to treatment were found to have a single copy gene deletion of the non-aberrant somatic allele. Two germline carriers of *ATM* variants (one of whom responded to olaparib) were also found to have somatic second hits in the form of a missense variant and copy neutral LOH respectively. [42] In the small case series reported by Cheng and colleagues, 3 patients with advanced prostate cancer who had durable responses to platinum chemotherapy after disease progression on several standard lines of treatment were all found to have bi-allelic *BRCA2* inactivation in their tumours. Two of the patients carried a germline variant that was

present in the tumour with an additional somatic frameshift deletion. Both the germline and somatic variants were detectable in both primary tumours and metastatic biopsies. The third case was found to have somatic homozygous *BRCA2* copy loss.

With the ongoing translational oncology trials that pre-screen patients by testing tumours for actionable gene variants, commercial tests for tumour NGS testing are also becoming increasingly available. As shown in the studies above, somatic testing may also reveal likely germline variants that warrant confirmation by dedicated clinical testing. This potential for identifying heritable gene variants that require subsequent validation should be made clear to patients who have their tumours sequenced. In this context, the NCCN guidelines recommend *BRCA1/2* germline testing for men whose tumours have been found to harbour pathogenic variants in one of these two genes. The Philadelphia Prostate Cancer Consensus Conference panel (convened in 2017) reached a consensus agreeing with this but expanded the set of genes that should be germline tested in this context (i.e. based on somatic testing results) to include MMR genes as well as *ATM* and *HOXB13* (agreement reached rather than consensus for the latter 2 genes).[68]

With the identification of 22 cases with germline PTVs in the BARCODE2 trial, some of which occurred in novel DNA repair genes, I endeavoured to obtain archival tumour tissue for these patients to carry out somatic NGS using the BARCODE2 gene panel and examine the somatic variants present in the aberrant germline gene as well as other DNA repair genes.

5.2 Aims

My aims in this project were:

- To sequence DNA from prostate cancers from patients identified to carry a germline PTV in the BARCODE2 trial using the study specific NGS panel.
- To check the presence of the germline PTV in the tumour DNA and assess for possible second hit events such as loss of heterozygosity (LOH) or development of a second PTV.
- To review the somatic PTVs present in other DNA repair genes in each case.

5.3 Results

5.3.1 Tumour Samples

19 tumour blocks were obtained related to cases identified to carry a germline PTV in a DNA repair gene. Of these, 8 blocks successfully underwent DNA extraction and NGS using the BARCODE2 gene panel. The reasons for not sequencing the remaining 11 samples were either poor DNA yield at the time of extraction or failure of library preparation due to poor DNA integrity. Table 37 summarises the tumour sample details. Most were samples of diagnostic prostate biopsies although some were of lymph node metastatic biopsies.

Table 37: Archival FFPE Tumours Obtained for Sequencing

Study ID	Tissue type	Germline variant	Estimated % of prostate cancer in sample pre-DNA extraction	DNA yield (ng)	Nanodrop assessment Abs 260/280
REG0055^	Neck lymph node	*POLQ* c.4262_4268del	80%	2760	2
REG0028	Prostate biopsy	*BRCA2* c.5217T>A	30%	425	1.74
REG0059	Prostate biopsy	*PALB2* c.1546delA	40%	39.2	2.3
REG0035**	Prostate biopsy	*ALKBH3* c.381T>G	10%	32.8	2.14
REG0046**	TURP tissue	*MRE11A* c.1090C>T	30%	189	2.01
REG0001	Prostate biopsy	*BRCA2* c.7360delA	50-60%	64	2.11
REG0070	Metastatic biopsy^^	*BLM* c.1570C>T	30%	600	1.99
REG0057	Prostate biopsy	*BRCA2* c.2834_2835del	5-10%	9.58	2.16
REG0076**	Prostate biopsy	*ATM* c.3802delG	30%	58.25	2
REG0074	Prostate biopsy	*ATM* c.3292delC	50%	52.5	1.92
REG0066**	Prostate biopsy	*PARP2* c.1480C>T	50%	40	1.99
REG0040**	Prostate biopsy	*CHEK2* c.del1100	60-70%	124.25	2.1
REG0013	Prostate biopsy	*PMS1* c.1360C>T	10%	38	1.73
REG0096	Prostate biopsy	*PALB2* c.3113G>A	90-100%	567	2.02
REG0086	Prostate biopsy	*MUTYH* c.1187G>A	10-20%	1.12	0.79
REG0099	Prostate biopsy	*BRCA2* c.4914_4915delAG	40%	306.6	2.02
REG0091	Prostate biopsy	*MUTYH* c.536A>G	n/a*	8	1.62
REG0021	Prostate biopsy	*MUTYH* c.536A>G	50%	524	1.7
REG0002	Prostate biopsy	*ALKBH3* c.208C>T	80%	906	1.94

TURP= trans-urethral resection of the prostate; Abs 260/280= absorbance ratio

^Diagnostic biopsy taken prior to treatment commencement.

^^Metastatic biopsy taken while patient on treatment with androgen deprivation and not yet castrate resistant.

*amount of tissue too small to estimate tumour percentage, all visible tissue used for DNA extraction.

**These samples underwent a magnetic bead/ethanol clean up after extraction.

5.3.2 DNA Extraction From Tumour Tissue

DNA extraction from FFPE tumour tissue was carried out in our lab using the QIAamp DNA FFPE Tissue Kit (Qiagen). Extracted DNA was then assessed and quantified using the Nanodrop spectrophotometer and Qubit fluorimeter. Table 37 shows the DNA yield for each sample and Nanodrop measurements.

5.3.3 Assessment of Extracted DNA

For samples with a low yield or sub-optimal 260/280 ratio on the Nanodrop assessment, I carried out a clean-up using magnetic AMPure XP beads and 70% ethanol. DNA was then eluted in 8μl of nuclease free water. A small volume was used at this stage in order to maximise DNA concentration for library preparation and sequencing.

I assessed the tumour DNA integrity for each sample using the Agilent Tape Station and Genomic DNA Screen Tape assay. This quantitative electrophoretic assay allows determination of DNA integrity and produces a DIN (DNA Integrity Number) score for each sample. Results of this assessment are shown in Table 38. Unfortunately, the assay failed when assessing the second batch of samples intended for NGS therefore some DIN scores are missing. Three samples had a DIN score that was too low to allow me to pursue them further. One sample had a DNA yield that was too low to use.

At this stage, 15 samples were suitable for library preparation for NGS.

Table 38: Tumour DNA DIN Scores and Quantity Used for NGS

Study ID	DIN	Total DNA input ng	Suitable for library prep and NGS
REG0002	1.9	201	Yes
REG0021	2.1	183	No- library prep failed
REG0028	1.9	119	No- library prep failed
REG0040	3.1	60	Yes
REG0046	2.9	201	Yes
REG0055	5.7	161	Yes
REG0076	2.1	201	No- library prep failed
REG0070	2.5	168	yes
REG0035	4.8	28	Yes
REG0066	1.7	172	No- library prep failed
REG0001	n/a	30	Yes
REG0027	2.9	52	Yes
REG0074	n/a	18	No- library prep failed
REG0096	n/a	189	Yes
REG0099	n/a	175	Yes

5.3.4 DNA Library Preparation and Targeted Capture for NGS

DNA libraries were prepared for each set of somatic NGS using the Agilent SureSelect XT HS reagent kit as described in Chapter 2 with the same steps carried out as done for germline DNA (described in Chapter 4). Unlike the QXT protocol, the DNA input for library preparation can range from 15-200ng depending on the DIN score of the DNA sample. Table 38 shows the total amount of DNA used for library preparation.

Five samples failed to yield adequate fragmented and amplified DNA during library preparation and therefore were not sequenced. I repeated the library preparation steps a second time for two of the failed samples as there was DNA available, but this did not improve the results. This is likely to be due to poor quality of the FFPE extracted DNA.

Two MiSeq runs were set up by me for somatic sequencing, the first consisting of 6 tumour samples alongside one germline sample used as a control and reference. The second consisted of 4 tumour samples along with 2 samples from the previous run (using the originally prepared and captured DNA) and one germline sample.

5.3.5 NGS Coverage and On Target Reads

FASTQ files generated on the MiSeq were processed and analysed using the SureCall program (Agilent Technologies) as described in Chapter 2. For each MiSeq run, NGS data were analysed using the SureCall program to generate a table of variants for each DNA sample. The software also produces values for the coverage and percentage of reads covering target regions for each sample sequenced. Unfortunately the coverage and percentage reads on target were much poorer on the second NGS run of somatic samples (3 out of 4 were too low to obtain data from). QC values for the two runs are shown in Table 39.

The lower proportion of reads on target in the second run reflects inadequate hybridisation of the RNA baits to DNA and in turn poor targeted capture. This may have been due to evaporation of material secondary to a faulty plate seal during the hybridisation which is carried out on a heat block using a 96 well plate sealed with a plastic seal. In fact, at the time of this particular lab set up, we had just switched over to using alternative 96-well plates due to a change in supplier. The new plates appear to be more malleable and therefore it is feasible that the plate seals may have been affected if the plate shape slightly altered during the high temperature stages on the heat block. Since this run, the protocol has been modified by the team to include the use of a hydrophobic low-viscosity barrier solution (Vapor Lock by Qiagen). This is overlaid onto the contents of each well on the hybridisation plate after the RNA baits are added to each well. It creates a barrier to prevent evaporation of solution from the wells.

This protocol modification was incorporated for the subsequent germline sequencing runs and appears to have resulted in improved on target reads percentage. One of the tumour samples from the second run had enough DNA extracted for further library preparation, therefore library preparation and targeted capture was repeated and it was sequenced alongside the germline samples on a third run.

Table 39: Post NGS QC Data (Generated by SureCall)

The table shows samples sequenced in 3 separate NGS runs.

Sample ID	% reads in target regions	Ave read depth in target regions	% target bases at 20x/ 50x/ 100x
REG0046_T01	66	330	99/97/92
REG0055_T01	30	140	94/80/56
REG0002_T01	45.5	187	95/87/71
REG0035_T01	65	232	99/97/92
REG0040_T01	73	228	98/94/83
REG0070_T01	59	331	98/96/89
REG0046G_T01 (germline control)	47	379	99/97/92
REG0059_T02	74	243	99/96/85
REG0096_T02*	1.52	13	24/1.1/0
REG0099_T02*	6.6	38	76.75/27/2.8
REG0001_T02*	9.9	39	76/28/3.3
REG0096T_P017	85	2074	99/99/98

*These samples showed poor hybridisation and capture reflected by low percentage of reads covering target regions. NGS data were not available for these 3 samples. REG0096T at the bottom was sequenced alongside germline samples being sequenced for the trial.

5.3.6 NGS Analysis and PTV Identification

The mean number of variants called per case was 923 (range 635-2150), higher than the same figure for the germline set of samples (mean 478). This reflects the expected higher mutational burden associated with tumour DNA compared with germline DNA.

The NGS analysis process using SureCall was similar to that carried out for the germline samples and described in Chapter 4 except for the following:

1. The analysis settings were modified to allow the calling of low frequency variants down to a frequency of 0.01 (for germline variant calling 0.3 was used).
2. The minimum number of reads for a variant allele to be called was set at 5 reads (for germline calling 10 reads was used).

Steps taken to filter and identify somatic PTVs are described in Chapter 2.3.3.

Criteria used for excluding somatic variants were as follows:

1. Evidence of strand bias whereby a variant is only, or is pre-dominantly, called on one strand- these are likely to be sequencing artefacts.
2. Minimum read depth covering variant position of 100.
3. Minimum number of reads with the variant call ideally 5 but variants with 4 reads were noted.
4. Variants that were called in all samples in the same run were excluded as they were likely to be artefacts; most of these displayed strand bias as well.

5.3.7 Protein Truncating Variants in BARCODE2 Tumour Samples

Next generation sequencing utilising the BARCODE2 gene panel was successfully carried out for DNA from 8 tumour samples. In all cases except one, the known germline PTV was evident with a maintained heterozygous state. In the 8th case where the patient was known to carry a heterozygous germline *PALB2* non-sense variant (c.3113G>A, p.W1038*), the tumour DNA showed a homozygous state with a variant frequency of 0.99. This change from the germline heterozygous state to the homozygous state in the tumour suggests inactivation of the second (normal) *PALB2* allele in accordance with the Knudson two hit hypothesis of a tumour suppressor gene occurring as an early event in tumourigenesis. This loss of heterozygosity could occur due to total or partial gene deletion leading to gene copy number loss. This particular tumour sample was sequenced on a NGS run alongside germline samples being sequenced for the BARCODE2 trial. Therefore, I was able to review the ratio of coverage across *PALB2* compared with the coverage across all targeted regions within the capture on a per sample basis. I found that the ratio of *PALB2* coverage relative to coverage of all regions was lower in the tumour sample compared to the same ratio in the germline samples. To demonstrate this, Table 40 shows the average coverage across *PALB2, SLX4, CDH1* and *BRCA2*, as well as the average coverage across all captured regions in the tumour sample and 3 germline samples sequenced in the same run. A difference between somatic and germline ratios is noted for *PALB2* as well as *CDH1* which lies downstream of *PALB2* on chromosome 16. *SLX4* lies upstream of *PALB2* on chromosome 16 and appears to have a similar coverage ratio in the tumour and germline samples.

Table 40: Comparison of gene coverage between somatic sample (REG0096T) and germline samples sequenced in the same run

Average Coverage across *PALB2, SLX4, BRCA2, CDH1* and all captured regions				
	REG0096T	**REG105**	**REG107**	**REG109**
PALB2	425	475	350	499
SLX4	1371	655	522	747
BRCA2	579	299	194	286
CDH1	601	490	363	524
All regions	876	410	300	438
Coverage ratios for each gene compared with overall coverage within each sample*				
PALB2	0.5	1.2	1.2	1.1
SLX4	1.6	1.6	1.7	1.7
CDH1	0.7	1.2	1.2	1.2
BRCA2	0.66	0.73	0.65	0.65

*This ratio is calculated by dividing the average coverage for a gene by the average coverage in all captured regions.

For this particular run the composition of the DNA pool loaded to the MiSeq was calculated to allocate 50% of sequencing to the somatic sample to achieve deep sequencing of the tumour while the remaining 50% would be distributed equally between the germline samples; this is why the number of reads across genes varies between the somatic and germline sample. The marked difference in coverage ratio between the tumour (REG0096T) and germline samples suggests that the second copy of *PALB2* in the tumour has been deleted leading to LOH. As the downstream *CDH1* gene shows a reduced coverage ratio as well, the deletion affecting *PALB2* may be part of a larger deletion of a region of chromosome 16. Ultimately, this can only be proven by using another analysis method such as array comparative genomic hybridisation (aCGH) or multiplex ligation dependent probe amplification (MLPA). (discussed in Section 5.4). The targeted exon capture and NGS carried out here cannot be used for accurate analyses of gene copy number.

Table 41 shows a summary of PTVs or pathogenic missense variants identified in each sample. Several cases carried variants in *TP53*, which is known to be commonly

aberrant in the early stages of prostate cancer development. Apart from the tumour with LOH of the germline variant, I did not identify a potential second 'hit' event or PTV (with a frequency of >5%) in the gene affected in the germline. A low frequency (1.3%) non-sense variant in *ALKBH3* was identified in the tumour belonging to REG0002 who was known to carry a germline *ALKBH3* variant. Low frequency variants in HR genes and MMR genes were observed in some tumours. Validation of somatic variants by Sanger sequencing was not carried out due to lack of available DNA for most samples as well as a restricted timeframe.

Table 41: Protein Truncating Variants Identified in Somatic DNA Samples

All samples are primary prostate biopsies unless indicated otherwise

Sample ID	Variant origin	Gene	Variant	Classification	Variant frequency	Variant reads/ Total depth	ClinVar status	COSMIC status	Comments
REG0059	germline No other *PALB2* PTVs	*PALB2*	c.1546delA p.R516Qfs	frameshift del	0.48	146/304	P	No record	
	somatic	*POLQ*	c.2578C>T p.R860*	nonsense ex.16	0.028	8/282	n/a	missense variant at this position	
	somatic	*ATM*	c.8170C>T p.Q2724*	nonsense	0.021	4/195	n/a	Recorded in lung Ca	
	somatic	*RAD51D*	c.511C>T p.Q171*	nonsense	0.013	4/304	P/LP	no record	
	somatic	*BARD1*	c.1387C>T p.Q463*	nonsense	0.013	4/309	n/a	no record	
	somatic	*SMARCA4*	c.3127C>T p.R1043W	missense	0.015	4/261	LP	COSMIC	LP ClinVar based on one record of clinical testing
REG0096	germline	*PALB2*	c.3113G>A p.W1038*	non-sense	**0.99**	332/334	P	Recorded in thyroid Ca	Variant heterozygous in germline
	no other *PALB2* variants called								

									21bp deletion in exon 4; other pathogenic variants in this region described
	somatic	*CDH1*	c.388-14_394del	splice	**0.9**	204/227	N/A	not recorded	
	somatic	*RB1*	c.1054delG p.Q352Kfs*15	frameshift	**0.89**	198/223	N/A	not recorded	
same position as previous variant	somatic	*RB1*	c.1054G>T p.Q352*	non-sense	**0.022**	5/222	N/A	Associated with skin Ca	
	somatic	*RAD50*	c.2165delA p.K722Rfs*14	frameshift	0.013	25/1955	P	not recorded	rs397507178

REG0070 (metastatic biopsy sample^)

	germline	*BLM*	c.1570C>T p.R899*	non-sense	0.41	155/379	P		
no other significant BLM variants									
	somatic	*TP53*	c.537T>A p.H179Q	missense	0.55	205/371	LP	Associated with PrCa	
	somatic	*MSH2*	c.2635C>T p.Q879*	non-sense	0.033	5/150		not recorded	
	somatic	*CDH1*	c.1137G>A p.T379=	synonymous	0.013	5/394	P	Associated with BC	
	somatic	*SMARCA4*	c.2576C>T p.T859M	missense	0.014	5/347	P	Associated with CRC	rs281875226

202

REG0046 (TURP while on ADT for mPC, not castrate resistant)

germline	*MRE11A*	c.1090C>T p.R364*	non-sense	0.52	162/314	P	not recorded
no other significant *MRE11A* mutations							
germline	*MUTYH*	c.536A>G p.Y179C	missense	0.52	243/464	P	not recorded
somatic	*RAD50*	c.2165delA p.K722Rfs*14	frameshift	0.026	13/491	P	not recorded
somatic	*TP53*	c.645T>G p.S215R	missense	0.2	52/261	LP	Associated with multiple tumour sites

REG0055 (neck lymph node met- initial diagnostic bx)

germline	*POLQ*	c.4262_4268del p.I1421Rfs*8	frameshift	0.38	37/97	n/a	not recorded
somatic	*TP53*	c.451C>G p. P151A	missense	0.88	140/159	LP	Associated with multiple tumour sites

REG0002

germline	*ALKBH3*	c.208C>T p.R70*	non-sense	0.31	96/305	n/a	not recorded	
somatic	*ALKBH3*	c.19C>T p.R7*	non-sense	0.013	5/383	n/a	not recorded	rs148553066
somatic	*PTCH1*	c.622C>T p.Q208*	non-sense	0.022	6/277	n/a	not recorded	

	somatic	*POLE*	c.3373C>T p.R1125*	non-sense	0.017	5/298	n/a	COSMIC endometrial ca	rs139603739
	somatic	*MSH6*	c.3202C>T p.R1068*	non-sense	0.014	5/357	P	not recorded	

REG0040

	germline	*CHEK2*	c.1100del p.T367Mfs	frameshift	0.31	41/133	P	not recorded	
no other significant *CHEK2* variants									
	somatic	*XRCC4*	c.673C>T p.R225*	non-sense	0.031	5/160	P (lit only)	missense variant at this position recorded in bladder ca	rs768825050
	somatic	*SPOP*	c.392G>C p. W131S	missense	0.24	76/314	LP*	not recorded	
	somatic	*MLH1*	c.1336C>T p.R446W	missense	0.018	5/279	P	Associated with CRC	rs63751275
	somatic	*CDH1*	c.1003C>T p.R335*	non-sense	0.015	5/330	P (lit only)	Associated with BC	rs587780784

REG0035	germline	*ALKBH3*	c.381T>G p.Y127*	non-sense	0.46	181/391	n/a		rs754599411
4 other *ALKBH3* variants (2 intronic and 2 synonymous) that are all homozygous in germline and tumour									
No significant somatic mutations called; this sample had a very low tumour percentage of 10%.									

BC= breast cancer, CRC= colorectal cancer, mPC= metastatic prostate cancer, PrCa= prostate cancer
^This sample was from a metastatic biopsy but the castration resistance status at the time of biopsy is unknown.
*ClinVar record based on somatic testing.

5.4 Discussion

I carried out the NGS for 8 tumour samples from 8 carrier patients in the BARCODE2 trial utilising the study gene panel. This was done primarily to assess whether the known germline variant was identified in the tumour DNA and identify potential second hit variants that may lead to loss of gene function. Only one case displayed evidence of a possible second hit in the gene of interest with LOH of the germline *PALB2* variant c.3113G>A. LOH refers to loss of the second non-aberrant copy of a gene known to harbour a particular variant. This may arise in a number of ways:

- Spontaneous acquisition of an identical somatic variant early in tumourigenesis, although this is very unlikely.
- Copy neutral LOH due to repair of DNA damage on the non-aberrant gene copy by homologous recombination (HR) leading to two gene copies containing the variant. HR DNA repair utilises the second gene copy as a template during repair, therefore in this context, the aberrant gene will be used as the template leading to two aberrant gene copies (i.e. the variant becomes homozygous).
- Deletion of the non-aberrant gene copy causing copy number loss; A deletion may affect a number of gene exons, the whole gene or a larger chromosomal region which encompasses the gene.

The ratio of *PALB2* coverage to other genes is lower in the somatic sample compared with the same ratio in the germline samples sequenced in the same NGS run suggesting that a gene deletion is likely to have occurred. The size of genomic deletion can't be ascertained by the NGS data available. Gene copy number analyses can be carried out by dedicated methods such as aCGH or MLPA. An aCGH assay is a platform that compares the hybridisation of two labelled DNA samples (a reference and test sample) to a set of oligonucleotide targets. The signal ratio produced by the assay can be used to infer copy number change. MLPA is based on PCR amplification and relative quantification of the oligonucleotide probes ligated to specific DNA sequences. The copy number of target

sequences is reflected by the relative intensities of the MLPA probe amplification products. MLPA can be carried out using less DNA than that required for aCGH and may be better suited for analyses of FFPE derived DNA.[112]

In analyses of copy number alterations (CNA) in prostate cancer, aggressive primary and metastatic tumours are associated with large numbers of CNAs in the tumour genome compared with low-grade, indolent tumours.[113] In the BARCODE2 case discussed above, the patient had aggressive baseline features including a Gleason grade of 8 and metastatic disease at initial diagnosis. Additionally, I also identified a somatic frameshift variant in *RB1* in this case (allele frequency of 0.89). Somatic variants in *RB1* have recently been reported to be significantly associated with poor prostate cancer survival.[114]

5.4.1 Somatic DNA Repair Gene Variants

Only one out of the 8 tumours sequenced carried a second PTV in the gene of interest (REG0002 with germline and somatic *ALKBH3* variants), although this was observed at low frequency (1.3%). In the context of clinical somatic sequencing to guide targeted therapy, a variant frequency of 5% or higher is usually thought to be clinically actionable. [115] Although the low frequency *ALKBH3* variant can't be fully ruled out as a sequencing artefact, the library preparation protocol I used utilises unique molecular indexes (barcodes) to improve the accuracy of low frequency variant detection. The molecular barcodes are ligated to the original DNA fragments prior to PCR and then carried through to enrichment and sequencing. These barcodes can then be retrieved from sequencing reads, allowing each read to be traced back to the original DNA molecule.

Many of the genes found to harbour variants in this set of 8 tumours have been reported in somatic sequencing studies in prostate cancer. A number of DNA repair genes apart from

the germline gene of interest were found to carry variants with varying allele fractions. Additionally, the tumour suppressor gene, *TP53*, was aberrant in 3 cases which is not unexpected as studies of somatic variation in prostate cancer have found that *TP53* is commonly altered in primary tumours and more so in cases that go on to develop metastatic disease. [39, 113] In one of these cases, the *TP53* variant was observed as a homozygous variant which may again reflect gene copy number alteration as focal deletions of *TP53* have been reported in primary prostate cancer. [113]

In an NGS study of tumour-normal pairs reported by Abida et al in 2017 [39], 22% of prostate cancer cases (N=451) were found to harbour a somatic alteration in a HR gene. A subset of patients in the same study had germline analysis undertaken and 19% were found to carry a germline variant in a DNA repair gene including *BRCA2* (9%), *CHEK2* (4%), *ATM* (2%) and *PALB2* (<1%). In the set of tumours I sequenced, a number of HR genes were also found to harbour PTVs including *ATM, RAD50, RAD51D* and *BARD1*. The significance of these low frequency variants is unclear, but low frequency alleles (1% frequency) thought to be due subclonal mutations have been described in primary prostate cancer. [116] Some of these subclones were found to increase in frequency in metastatic tumours. As patients with both germline and somatic variants have been reported to respond to drugs such as PARPi and platinum, these results suggest that both germline and somatic analyses are warranted in prostate cancer management, particularly in the advanced settings where these treatments would be indicated.

Low frequency variants in *MLH1, MSH6* and *MSH2* were noted in 3 BARCODE2 cases from diagnostic prostate biopsies and is in keeping with MMR alterations observed in 2.6-3% of cases [39, 50]. Genomic profiling of these reported cases have shown that they are enriched for MMR and MSI signatures. As previously discussed, treatment with immune

checkpoint blockade agents has shown efficacy in several tumour types displaying MMRd, and identifying prostate cancers with MMRd could potentially open up a new line of treatment for patients. Unfortunately, I was unable to check MMRd or MSI in the tumours in this small sub-study.

5.4.2 Limitations and Future Work

This was a small study of prostate tumours from patient carriers identified in the BARCODE2 trial. Due to the quality of the FFPE derived DNA and low yield in some cases as well as the poor bait hybridisation in the second tumour NGS run, I was unable to sequence tumours from all carriers in the trial. The percentage of tumour in the tissue that underwent DNA extraction was also variable and may have affected the NGS results.

Some of the unique PTVs identified in the tumours had a low allele fraction (<2%) and could potentially have been artefacts although they did not display strand bias and all had a coverage of >100 reads. It would have been ideal to run the NGS data through a bioinformatics pipeline utilising a somatic variant caller in addition to the SureCall analysis I carried out to check for agreement of variant calling and filtering as was carried out for the pilot germline analyses, but due to time limitations this was not done. Such bioinformatics pipelines are designed with stringent filtering criteria to enable the exclusion of sequencing artefacts, variants arising at sites of common germline variation as well as multi-allelic variants. Validation by Sanger sequencing was also not carried out due to time limitations.

Within the BARCODE2 trial, plasma samples are being collected from carrier patients who go on to receive treatment in the study, for future cell free DNA (cfDNA) studies. As these patients have a high burden of metastatic disease, the yield of cfDNA from these samples is likely to be high and would be ideal for NGS analyses to study the somatic alterations in

the advanced mCRPC setting. This may identify further alterations in the aberrant germline gene or in other DNA repair genes that could guide further treatment options.

5.4.3 Conclusions

By carrying out somatic NGS for a subset of the BARCODE2 cases, I identified variants in genes known to be associated with the development of metastatic disease such as *TP53* and *RB1*. Although most of the tumour DNA was taken from diagnostic biopsies, all patients had gone on to develop mCRPC therefore it is not surprising that the molecular alterations in the primary tumours were associated with poor prognostic features.

Low frequency variants were identified in HR as well as MMR genes and future cfDNA analyses from these patients may show an increase in these potential tumour subclones; such variants could predict responses to targeted agents such as PARPi (for HR variants) or to immunotherapy drugs (for MMR variants). As the availability of NGS testing increases and the development of molecularly targeted agents evolves, both germline and somatic sequencing would give a clearer picture of a patient's cancer and allow treating teams to optimise management accordingly.

Chapter 6 General Discussion

In this thesis, I have investigated the potential role of germline genetic profiles in two prostate cancer settings: targeted screening in healthy men and personalising treatment for affected patients. I was also able to assess a subset of tumours in men found to carry germline mutations to show that both somatic and germline sequencing can be informative for personalised treatment of the prostate cancer patient. Specifically, I have presented the following findings:

7. The use of SNP profiling in the community setting is acceptable to primary health care teams and to men in the community:

8. The uptake of the BARCODE1 pilot study was 26%, which is comparable to other prostate cancer screening studies.

9. The uptake of prostate cancer screening tests in men with increased genetic risk of prostate cancer was 76% in the BARCODE1 pilot study. This is comparable to current NHS England screening programmes such as the bowel cancer and breast cancer screening programs (uptake of 58-72% in Public Health England data for 2017/2018).

10. The PRS distribution in the BARCODE1 pilot population was comparable to two age matched reference UK populations of men.

11. Community based prostate cancer screening guided by a SNP profile led to the diagnosis of low grade cancers in 33% of men screened in the BARCODE1 pilot study.

12. In men with mCRPC, 22% carried a PTV in one or more DNA repair genes. If focusing only on those with PTVs in homologous recombination (HR) or DNA damage response (DDR) genes, 11% carried a variant- this is in keeping with published studies.

13. Carriers of PTVs in a DNA repair gene were diagnosed with castration resistance ($P=0.03$) at a younger age than non-carriers.

14. The interval between prostate cancer diagnosis and castration resistance onset was shorter in carriers compared with non-carriers (20 vs 45.5 months; $P=0.044$)

15. The PRS distribution in 98 men with mCRPC in the BARCODE2 trial showed a higher median PRS compared with the 'healthy' set in the BARCODE1 community derived population ($P<0.0001$).

16. Sequencing somatic DNA in addition to germline DNA could give a clearer picture of the genomic changes in prostate cancer cases and allow more effective personalised treatment.

Both BARCODE trials are currently ongoing. After completion of recruitment of the BARCODE1 pilot study, protocol modifications were made, over 50 GP sites were invited to participate and the method for community recruitment was altered to allow the use of the automated Docmail system prior to opening the main BARCODE1 study. The main study will recruit 4700 men to give a total of 5000 (including the pilot cohort). This has already led to a more rapid rate of recruitment with over 40,000 study invitations sent out by GPs in the first 4 months of the study opening (March-July 2019).

The BARCODE2 trial continues to recruit men with mCRPC with the aim of recruiting 450 men for germline sequencing in Part 1 of the trial. Carriers of PTVs in DNA repair genes are offered carboplatin chemotherapy within the second part of the trial and responses will be analysed according to 3 gene groups as set out in the study protocol. I set up the lab workflow for this project to enable a turnaround time of 3-4 weeks. This timeframe included DNA extraction, DNA library preparation, NGS and NGS analysis, and Sanger sequencing to validate findings. The trial protocol allows a maximum of 12 weeks from the time blood is taken from a patient to having a genetic test result available; this allows batching of the samples for NGS. This timeframe was met in all cases analysed by me except for the *FANCD2* splice variant case which required some troubleshooting before achieving successful validation by Sanger sequencing.

The BARCODE2 trial is the first prospective trial to investigate the use of carboplatin in men with mCRPC carrying a protein truncating germline variant in a DNA repair gene. Although there are retrospective data in this area for men with *BRCA1/2* variants [70], there have been no prospective trials in this setting. Additionally, prostate cancer patient carriers

212

of PTVs in non-HR DNA repair genes have not been investigated previously, therefore this will be a unique cohort of patients that can be studied to assess whether they show evidence of poorer prognostic features as seen with carriers of *BRCA2* and *ATM* variants, as well as assessing whether they respond to carboplatin. In the era of targeted agents such as PARP inhibitors which often gain regulatory approval attached to stringent conditions based on the relevant phase II or III clinical trial, repurposing an established drug like carboplatin offers an easily accessible and cheaper drug which may be offered to those who cannot access a clinical trial or an expensive targeted agent. The latter scenario is especially applicable to oncology settings in developing countries where accessing genetic testing of tumours or germline DNA may be feasible but accessing the latest molecularly targeted agents may not.

6.1 Limitations and Constraints

6.1.1 BARCODE1 SNP Profile and PRS

The BARCODE1 SNP assay was developed based on the known prostate cancer risk SNPs in 2017. These included the 63 loci identified in the OncoArray GWAS and meta-analysis reported by Schumacher et al.[13] Alongside such GWAS, fine mapping projects have investigated the regions within which GWAS identified risk SNPs lie with the aim of identifying the variant(s) driving the association with prostate cancer. The additional variants identified in this manner may be correlated SNPs lying in linkage disequilibrium (LD) with the original variant of interest or other SNPs with independent associations with prostate cancer. In the 2018 study reported by Dadaev et al, fine mapping of 84 regions harbouring prostate cancer risk loci led to the identification of 99 risk SNPs of which only 15 were original GWAS hits.[12] The remaining variants were replaced by more likely candidate variants. In a second study, the 8q24 prostate cancer susceptibility region was

fine mapped and 12 independent prostate cancer risk variants were reported.[11] Variants identified by fine mapping studies based on prior large GWAS findings may be more accurate markers of risk than the original GWAS hits. The variants identified in these two studies were reported after the development of the BARCODE1 assay had commenced. Including such variants in genotyping assays designed for PRS estimation may produce a more refined and accurate test for identifying men at increased genetic risk of disease. The BARCODE1 study genotyping assay will not be modified at this stage but the results and conclusions of the study will take into account the potential for the use of a further developed assay if larger subsequent population studies are planned.

The prostate cancer risk SNPs used in the genotyping assay in BARCODE1 were identified in GWAS and meta-analyses carried out in European populations therefore the study is only recruiting men of white European ethnicity. GWAS have been carried out in populations of other ethnicities and some SNPs do replicate across populations, for example, many of the 8q24 loci, but some risk SNPs are specific to certain ethnicities. A recent GWAS reported by Conti et al was carried out using cases and controls of African ancestry; two novel risk SNPs were found that are only observed in African men.[117] Additionally, 81 of 100 previously reported prostate cancer risk SNPs replicated in this GWAS and meta-analysis in African men, showing that although there is overlap of genetic risk variants between populations, population specific variants must also be factored into genetic risk scoring systems.

The Eureka Genomics assay used in BARCODE1 does include some multi-ethnic SNPs and there are plans to modify the assay to include further SNPs that can be used to investigate other populations. One arm of the PROFILE study (NCT02543905) is currently recruiting black men of African and Afro-Caribbean ethnicity for prostate cancer screening.

As part of this study, PSA testing, MRI and biopsy are carried out in addition to DNA collection for genotyping. Screening outcomes will be correlated with PRS according to a SNP assay specific to that population and these results will be informative for the planning of a population screening study similar to BARCODE1.

6.1.2 The Changing Approach to Prostate Cancer Detection

In the BARCODE1 pilot study, 2 men developed a urinary tract infection after biopsy. Other risks after biopsy include urinary retention and sepsis, the latter occurring at a rate of ~1%, although these did not occur in the BARCODE1 pilot study.

Until recently the standard approach to diagnosing prostate cancer in men clinically suspected to have cancer has utilised a trans-rectal ultrasound guided (TRUS) biopsy without prior imaging of the prostate. Recent studies of prostate cancer detection methods have shown that the use of multi-parametric MRI (mpMRI) prior to prostate biopsy is superior to standard TRUS biopsy for the early diagnosis of prostate cancer as it allows suspicious lesions on MRI to be targeted at the time of biopsy.[66, 118] These studies have already led to updates in regulatory guidelines.

The PROMIS study which set out to compare the diagnostic accuracy of mpMRI followed by TRUS biopsy with a transperineal template biopsy found that mpMRI had a sensitivity of 88% for the detection of prostate cancer of grade group (GG) 2 or higher (i.e. ≥Gleason score 3+4) and a negative predictive value of 76%.[66] The investigators concluded that the use of mpMRI could lead to the avoidance of a biopsy in 27% of patients. In the PRECISION study, men with a clinical suspicion of prostate cancer were recruited and randomised to either a standard TRUS biopsy or an mpMRI pathway.[118] The latter pathway involved carrying out a mpMRI of the prostate and only offering a biopsy to those

with an MRI showing PIRADS 3-5 changes while those with a PIRADS 1-2 result did not undergo biopsy. In the mpMRI arm, 28% of men avoided a biopsy and 38% were found to have a clinically significant cancer, compared with 26% in the TRUS arm. A lower proportion of men in the mpMRI arm were diagnosed with clinically insignificant cancer (9% vs 22%). In a similar UK study reported by Barrett et al in 2019, 44% of men who underwent an upfront mpMRI avoided a biopsy and mpMRI was found to have a 92% negative predictive value for the detection of ≥GG2 prostate cancer. [119]

As a result of these studies, the European Association of Urology guidelines support the use of mpMRI prior to biopsy in men who are biopsy naïve and those who have previously had a negative biopsy.[120] In the UK, NICE updated their prostate cancer diagnostic guidelines in 2019 to recommend mpMRI as a first line test for suspected localised prostate cancer (Guideline NG131). In this setting, mpMRI can be used in two ways: the first is to identify suspicious lesions that can be targeted during biopsy with or without systematic sampling of the prostate. The second use of mpMRI is to identify patients who could avoid a biopsy procedure completely if the imaging shows low risk features, as in the PRECISION study. NICE guidance does recommend the 'consideration' of omitting a biopsy for those with an mpMRI with PIRADS score 1-2 after discussion with the patient regarding the risks and benefits of biopsy.

In the BARCODE1 pilot study, the majority of MRI scans resulted in PIRADS 1-2 findings and all diagnosed cancers were GG2 or less. The majority of participants had a low PSA, in contrast to the men recruited into the PROMIS and PRECISION studies (PROMIS: mean PSA 7.1ng/ml, range 0.5-15ng/ml; PRECISION: median PSA in the two study arms 6.75 and 6.50 ng/ml). Once follow up is complete in the main BARCODE1 study, the characteristics of a population with increased genetic risk of prostate cancer including

findings at MRI and biopsy outcomes will be analysed. The optimal screening and follow up of such men could take the form of a mpMRI of the prostate, with or without biomarkers such as PSA, while reserving biopsy as a second line screening test when defined criteria are met such as a high PIRADS score on MRI. This could in turn reduce further the level of over-diagnosis in a prostate cancer screening programme. In the era of precision medicine for cancer treatment, the wealth of genetic susceptibility data that is now available should be utilised to develop precision screening in the pre-cancer setting as well.

6.1.3 BARCODE2 Germline NGS

For the analysis of NGS data, I used the integrated SureCall program by Agilent Technologies which runs the analysis on each sample individually using the MiSeq fastq files, and produces a table of several hundred variants for each sample. This is in contrast to bioinformatics pipelines (e.g. GATK) which often process all samples' VCFs (variant call files) simultaneously using a joint genotyping tool which in turn produces a set of joint-called SNP and indel calls ready for filtering. This type of cohort-wide analysis increases the sensitivity for the detection of variants at difficult sites e.g. with lower or variable coverage while reducing the false positive rate. For the purposes of my germline analysis, per sample analysis was carried out using SureCall. Although a high depth of coverage was achieved on all NGS runs, a small proportion (0.45-0.6%) of target regions had consistently low coverage due to issues such as being in repeat regions duplicated elsewhere in the genome. As the RNA baits for capture were designed with 50bp flanks either side of exons, many of these low coverage regions are in intronic sequences which would not affect my analysis.

By carrying out the pilot experiments where analysis using SureCall was compared with the GATK bioinformatics pipeline, I was reassured that using this program would allow the

identification of frameshift and non-sense variants which could then be assessed for likelihood of being a PTV. As SureCall annotation includes clinical classification data extracted from the National Center for Biotechnology Information (NCBI) dbSNP database, I identified several pathogenic missense mutations in *MUTYH* when reviewing all variants annotated as pathogenic or likely pathogenic.

Apart from missense variants annotated to be clinically pathogenic, I did not analyse other missense variants output by the software. Although there are *in silico* tools available for the assessment of missense (and indel) variants and to predict whether they are likely to be protein damaging or not, proving a novel variant is protein damaging would require functional assays to be carried out, which was outside the scope of this study. So it is possible that further significant variants may have been present in the set of DNA samples sequenced in this project. Additionally, the novel PTVs identified in the BARCODE2 set could be investigated further with functional assays.

6.1.4 Germline PTVs in DNA Repair Genes

Within the BARCODE2 germline sequencing, 22% of patients were found to carry a PTV in a DNA repair gene. These included variants in established prostate cancer risk genes such as *BRCA2* and *ATM*. I also identified variants in genes that have been reported by other groups researching prostate cancer germline variants such as *PALB2*, *MRE11A* and *CHEK2*.

In addition, I identified PTVs in DNA repair genes not previously reported in this context such as *ALKBH3*, *EXO1* and *POLQ*. These genes were included in the study panel test due to findings in previous studies by the ICR Oncogenetics team where variants in these genes were identified in prostate cancer cases. Whether they are actually disease causing

is unclear and to answer this question a case control study would be required. In the context of the BARCODE2 trial, men with a PTV in a DNA repair gene are offered carboplatin when their disease progresses after two standard lines of treatment for metastatic prostate cancer. Unlike men carrying *BRCA2* variants who are known to have a high chance of responding to platinum, based on retrospective data as well as clinical data from other *BRCA2* associated cancers, these patients carrying variants in novel DNA repair genes have not previously been identified prospectively and assessed for response to platinum. Clinical response to platinum (and PARP inhibitors) in *BRCA2* variant carriers occurs due to the development of a second hit in the second *BRCA2* allele in the tumour which renders it sensitive to drugs causing double strand breaks in DNA. Whether the same responses will be seen for patients found to carry variants in other DNA repair genes, particularly non-HR genes, remains to be seen. For some of the novel DNA repair genes identified to be aberrant in this study, their protein's biological functional role suggests that loss of function could potentially sensitise tumours to carboplatin.

For example, a non-sense variant, c.1241C>A, in exon 9 (of 15) of *EXO1* was identified in one of the patients I sequenced for the BARCODE2 trial. This variant leads to the amino acid change: p.S414X, introducing a premature stop codon. *EXO1* encodes exonuclease 1 which interacts with MLH1 and MSH2 in the process of DNA mismatch repair (MMR) to excise mismatch-containing DNA tracts. Exonuclease 1 is also phosphorylated in response to stalled DNA replication and then activates end DNA end resection by ATM and ATR, demonstrating its role in promoting HR DNA repair [121]. Unfortunately, I was unable to obtain this patient's tumour to sequence and assess the *EXO1* gene for second hit variants in somatic DNA. Although the c.1241C>A variant has not been reported in the literature, several *EXO1* SNPs have been linked to other cancers such as colorectal, gastric and breast. An *EXO1* SNP, rs4149963, has been found to associate with biochemical

recurrence in prostate cancer [122]. The *EXO1* gene was included in the study panel test as two carriers of an *EXO1* splice variant had been identified in a whole exome sequencing (WES) study reported by the ICR Oncogenetics team [77]. These carriers both had aggressive prostate cancer and in the comparison cohort of non-aggressive cases no carriers were identified. Further exploration of the *EXO1* gene in future WES or whole genome sequencing studies (WGS) may reveal further interesting variants.

Three different PTVs in the gene *ALKBH3* were identified in 3 cases in the BARCODE2 cohort (discussed in chapter 4.5.1). With the known role of ALKBH3 protein in repairing alkylated DNA damage, this is another gene that will be interesting to study responses in patient cases to alkylating agents such as carboplatin. The treatment part of the BARCODE2 trial is ongoing and will provide data on clinical responses.

6.1.5 Somatic DNA NGS

For this thesis, I was able to analyse a subset of tumours from the 22 PTV carriers identified in the BARCODE2 set. Unfortunately, due to the nature of FFPE derived DNA, not all extracted DNA was able to be used for NGS due to poor yield or poor DNA integrity. This may be related to the age of the specimens or the extraction process. Due to time limitations, I was unable to obtain further tumour blocks for the failed cases.

From the analysis of tumour biopsies from 8 cases, I was able to confirm the presence of the known germline variant in the tumour DNA. All but one case displayed a maintained heterozygous state for the respective variant. LOH for a *PALB2* variant was observed in one case indicating a possible second 'hit' to the *PALB2* gene in this tumour. Formal copy number alteration (CNA) analyses of this sample would have allowed assessment for the presence of gene/exon copy number loss. Other mechanisms inducing a second 'hit' in

these tumours could include gene inactivation by epigenetic mechanisms which may account in some cases for the lack of a second PTV in the gene of interest. LOH has been described in prostate cancer cases carrying *BRCA2* and *ATM* variants [30] but unfortunately the *BRCA2* and *ATM* cases were not successfully sequenced in my study.

A number of interesting somatic variants were identified such as variants in MMR genes as well as in the commonly aberrant *TP53* gene. These variants displayed varying allele frequencies (1.4%-55%). This is not unexpected with somatic sequencing data as varying allele fractions occur due to purity (or impurity) of extracted tumour DNA, multiple tumour subclones and copy number variation. Although the tumour NGS analysis revealed many low frequency variants that were likely to be sequencing artefacts (due to their presence in all tumours as well as some germline samples sequenced using the same panel), several unique variants were identified. Due to time limitations and for some cases, a lack of availability of additional DNA, validation by Sanger sequencing was not carried out.

Within the BARCODE2 trial, plasma specimens are being collected and in view of the high burden of disease these patients have, these would be expected to yield a large amount of cell free DNA (cfDNA) for future analyses. These are likely to be more informative regarding somatic alterations in advanced prostate cancer, and may also reveal subsequent alterations in the gene affected by the germline variant, for example, second hit variants or loss of the normal allele at a progressive stage which may suggest that gene's association with an element of cancer progression. Functional studies utilising cell lines and gene knockout techniques would ultimately be needed to prove such theories.

6.2 Conclusions and Future Directions

6.2.1 Genetic Profiles for Cancer Screening

In the pre-cancer screening setting, a precision medicine approach based on an individual's genetic profile is attractive. Prostate cancer screening in a patient presenting with suspicious symptoms currently relies on PSA measurement with or without digital rectal examination of the prostate. The lack of specificity associated with PSA measurement for prostate cancer detection has led to caution from national screening program regulators regarding its use for screening. Therefore, germline DNA and SNP genotypes may be an ideal biomarker to use in a screening program as this is a largely constant marker that is easily accessible and requires a single measurement. With proper research into the methodology for such screening programs, refinement of the genetic profile assays to be used and by selecting the appropriate screening procedures for those deemed to require it, the goal would be to develop a cost effective and efficient   programme for population screening which would optimise the risk-benefit balance of screening in the general population. [65, 93]

6.2.2 Genetic Testing in Oncology

The use of gene panel testing in oncology is rapidly expanding with the increasing availability of commercial tests, each including a varying number of genes (Table 42 shows genetic tests available for prostate cancer predisposition). Panels for germline testing range from broad 'all cancer' panels to narrower sets of genes designed for specific cancer types. In the research setting, the germline carrier frequency of pathogenic or deleterious variants in DNA repair genes in prostate cancer has varied between studies, due to the different types of NGS testing that has been carried out and the range of genes tested. Large robust case-control studies are required to identify true hereditary associations with prostate

cancer and these will rely on collaboration between research groups. Collaborations such as the PRACTICAL consortium (www,practical.icr.ac.uk), encompassing 133 research groups are working towards providing these data. Functional assessment of genes and their potential for cancer causality is also needed, as well as genomic assessment of tumours to identify the role of germline variants in cancer development and/ or progression. The benefits of germline testing are two-fold as it may guide patients and their relatives regarding their risk of other cancers (e.g. in Lynch Syndrome) and may lead to cascade counselling of relatives regarding their cancer risk. Secondly, it may allow patients to access targeted treatments or trial drugs that are predicted to lead to responses based on a particular germline variant.

With the advent of oncology based germline genetic testing (e.g. *BRCA1/2* testing in ovarian cancer clinics) as well as increased somatic testing which can raise a suspicion of the presence of germline variants, clear guidelines are needed to establish the type of testing and choice of genes selected for patients in prostate cancer clinics. The NCCN is one of the first regulatory bodies to release such guidance (NCCN Prostate Cancer Guidelines Version 2.2019) and it is expected that others will follow. Within the NHS in England, access to germline genetic testing as well as somatic testing will soon be increased with the establishment of 7 Genomic Laboratory Hubs across the country. These hubs will merge with the relevant Genomic Medicine Centres (GMCs) that were set up as part of the 100,000 Genomes project (Discussed further in section 6.2.2).

Table 42: Commercially available gene panel tests for prostate cancer predisposition

Gene panel test provider	Prostate Cancer Specific	Number of genes	Genes included
Gene Health UK©	Yes	10	*BRCA1/2, HOXB13, MLH1, MSH2/6, PMS2, EPCAM, ATM, CHEK2*
CeGaT© (Germany)	Yes	11	*BRCA1/2, HOXB13, MLH1, MSH2/6, PMS2, EPCAM, NBN, TP53, CHEK2*
Invitae©	Yes	12	*BRCA1/2, HOXB13, MLH1, MSH2/6, PMS2, EPCAM, NBN, TP53, CHEK2, ATM*
Color Genomics©	No- Hereditary Cancer Test	30	*BRCA1/2, HOXB13, MLH1, MSH2/6, PMS2, EPCAM, NBN, TP53, CHEK2, ATM, APC, MUTYH, MITF, BAP1, CDKN2A, CDK4, PTEN, STK11, CDH1, BMPR1A, SMAD4, GREM1, POLD1, POLE, PALB2, BARD1, BRIP1, RAD51C/D*
Ambry Genetics®	Yes- option to add on a PRS test using 72 SNPs*	14	*BRCA1/2, HOXB13, MLH1, MSH2/6, PMS2, EPCAM, NBN, TP53, CHEK2, ATM, RAD51D, PALB2*
Myriad®	No- Hereditary Cancer Test	29	*BRCA1/2, HOXB13, MLH1, MSH2/6, PMS2, EPCAM, NBN, TP53, CHEK2, ATM, APC, MUTYH, BAP1, CDKN2A, CDK4, PTEN, STK11, CDH1, BMPR1A, SMAD4, GREM1, POLD1, POLE, PALB2, BARD1, BRIP1, RAD51C/D*

*The AmbryScore test utilises 72 prostate cancer risk SNPs and can be added onto a gene panel test.

6.2.3 Somatic NGS for Prostate Cancer

NGS studies of various tumour types including prostate cancer utilising serial biopsies have demonstrated the genomic heterogeneity of primary tumours as well as subsequent metastases. With the rapidly evolving advances in NGS technology and DNA extraction from plasma (liquid biopsies), studies of tumour evolution and the patterns of oncogenic

variant alterations with cancer progression are ongoing. Apart from understanding the genomic mechanisms of prostate oncogenesis and progression, in the clinical setting, the goal is to develop a personalised medicine approach for the treatment of patients based on genomic and molecular alterations in a cancer, some of which will originate in the germline. Although somatic sequencing is not routinely carried out for most tumours within the NHS, several commercial somatic tests are available. (Table 43)

Table 43: Commercially available somatic gene panel tests

Gene panel test and provider	Test sample	Test Details
MI Profile- Caris Molecular Intelligence®	FFPE tumour tissue	NGS for 592 genes, CNA analysis and genomic stability signatures for MSI and TMB. RNA whole transcriptome sequencing for fusion analysis. Additional IHC tests carried out depending on tumour type.
FoundationOne® CDx- Foundation Medicine	FFPE tumour tissue	NGS for 324 genes, CNA analysis and genomic stability signatures for MSI and TMB.
FoundationOne® Liquid- Foundation Medicine	Whole blood for cfDNA anlaysis	NGS for 70 genes, CNA analysis and MSI testing.
Oncofocus®- Oncologica®	FFPE tumour tissue	NGS for 505 genes with CNA analysis and PD-1/PDL-1 IHC.

TMB= tumour mutational burden, IHC= immunohistochemistry

Clinical trials selecting prostate cancer patients by specific genetic alterations in their tumours are already underway. The TRITON3 (NCT02975934) phase III trial of rucaparib and the GALAHAD phase II trial of niraparib (NCT02854436) are enrolling mCRPC patients with evidence of DNA repair gene variants in their tumours to investigate the response to PARP inhibitors after progression on chemotherapy and androgen receptor targeted agents.

The PROfound trial (NCT02987543) is a phase III trial that is investigating the role of olaparib (a PARPi) in mCRPC men with somatic HR gene defects.[103] Study participants had progressed on an AR targeted agent and were randomised to receive olaparib or another AR targeted agent. Initial results were presented at the annual ESMO conference this year (September 2019) and reported a significant improvement in median progression free survival (PFS) in the olaparib arm: 5.8 months vs 3.5 months. Median overall survival was also improved by 3.25 months. In the cohort of men with somatic variants in *BRCA1, BRCA2* or *ATM*, median PFS was 7.39 months in the olaparib arm compared with 3.55 months in the AR targeted agent arm.[103] If these results are confirmed with longer follow up, somatic testing for HR defects in advanced prostate cancer would be warranted to identify patients that would benefit from PARP inhibitors.

The phase II RE-AKT trial (A Randomised Phase II Study of Enzalutamide (MDV3100) in Combination With AZD5363 in Patients With Metastatic Castration-Resistant Prostate Cancer, NCT02525068) is investigating the use of an AKT (protein kinase B) inhibitor in patients with *PTEN* and/ or *PIK3CA* alterations in their tumours. .

Precision medicine in oncology is moving towards an agnostic approach in cancer management, with treatment selection according to genomic and molecular alterations in a tumour rather than the histological cancer type. In a very recent study of 223 paediatric oncology patients, a hybrid-capture NGS panel test was used to sequence FFPE derived tumour DNA and found that 51% of tumours sequenced had evidence of clinically actionable variants that could be targeted by drugs that were already available in adult oncology. A small proportion of patients in the study went on to receive molecularly targeted agents according to the results of the somatic testing and showed positive responses with progression free survival in the range of 9-15 months. [123]

The basket and umbrella trial designs are now increasingly employed in phase II trials to guide patient selection during the investigation of molecularly targeted agents. Although many of these trials vary in the specifics of their design and set up, patients are selected for trial entry according to the presence of pre-specified molecular alterations (+/- their histological cancer type) that are expected to sensitise tumours to the trial drug. The NCI MATCH trial in the USA was one of the first of these umbrella trials and was set up in 2014 (opened 2015) to enrol patients whose cancers had advanced despite standard treatment.[124] By sequencing 143 genes in DNA from metastatic biopsies, trial patients are entered into one of 37 treatment arms utilising targeted agents according to NGS results. NCI MATCH is ongoing and has already showed positive signals for nivolumab (PD-1 inhibitor) in tumours with MMR defects, taselisib for *PIK3CA* mutations and TDM-1 (a HER2 targeted agent) in non-breast cancer tumours with HER2 overexpression, among others.

In the UK, the plasmaMATCH trial (plasma-based Molecular profiling of Advanced breast cancer to inform Therapeutic Choices, NCT03182634) is a phase IIa trial that aims to recruit 1,150 breast cancer patients and is using circulating tumour DNA (ctDNA) analysis to identify tumours with molecular aberrations that allow entry to one of 5 treatment cohorts. The 5 cohorts are investigating the use of the PARP inhibitor niraparib and an AKT inhibitor (capivasertib) in various combinations with or without the oestrogen receptor (ER) targeted agent, fulvestrant, depending on the ER status of the tumour.

In July 2019, larotrectinib was licenced by the European Medicines Agency (EMA) for the treatment of paediatric and adult solid tumours that harbour an *NTRK* (Neurotrophic Tyrosine Receptor Kinase) gene fusion. This is the first drug to be licenced in Europe in a 'histology independent' manner. As already mentioned, pembrolizumab, has been

approved by the FDA in the USA for use in any tumour with evidence of MSI-H. Further approvals of drugs based on molecular targets rather than tumour histology will follow as a result of ongoing trials such as those discussed above.

Although somatic testing with broad NGS panels is not routinely carried out within the NHS for prostate cancer and other tumours, several commercial tests are available to patients that wish to pursue testing privately (Table 43). These tests may present results that can be used to identify suitable clinical trials and are unlikely to alter standard treatment pathways currently.

Specific NGS tests are available in the NHS in some oncology settings but these are usually specific tests interrogating one or two genes to guide treatment using licenced molecular agents. For example, *KRAS* and *NRAS* sequencing of colorectal cancers is routinely carried out to identify mutations that exclude patients from receiving *EGFR* targeted agents. In lung cancer, adenocarcinomas are routinely sequenced for *EGFR* mutations that predict responses to anti-EGFR tyrosine kinase inhibitors.

With the establishment of the 100,000 Genomes project in 2012, 13 NHS Genomic Medicine Centres (GMCs) were established in England along with one each in Wales and Northern Ireland (a Scottish Genomes Partnership centre was set up in Scotland). This project was set up with the purpose of carrying out the sequencing of 100,000 whole genomes from 75,000 patients (and relatives) focussing on those with rare diseases and patients with common cancers. For cancer patients, both germline and somatic whole genome sequencing have been performed.

With the conclusion of recruitment to the project in 2018 and the ongoing data analysis which will contribute to research and personalised medicine, the NHS GMCs are now

expected to provide increasing access within the NHS to NGS testing in oncology for both somatic and germline sequencing. This will be carried out within one of 7 linked Genomic Laboratory Hubs (GLHs). Each hub will have a single laboratory for germline testing and another for somatic testing. The 2019/2020 National Genomic Test Directory specifies which genomic tests are commissioned by NHS England. There are two directories: one for rare and inherited disorders and one for cancer. The cancer test directory lists all the genomic tests and panels available for specific tumour groups; currently there are no tests available for prostate cancer but as the trial data of molecularly targeted agents (such as those discussed above) for prostate cancer matures and the subsequent new licencing of drugs occurs, this is likely to change. The establishment of the GLHs will allow the continued development of precision medicine within the NHS in oncology, and the quicker delivery of genomic information to patients and their clinicians with the aim of improving care and identifying targets for treatment.

In this era of genomic technological advancements and rapidly evolving research and with the imminent increase in access to broader NGS testing within the NHS, it is likely that in the near future, precision medicine utilising genetic profiles (both germline and somatic) will be guiding prostate cancer management.

Chapter 7   Appendices

Appendix 1: The BARCODE 1 Pilot Study Protocol
Protocol appendices not included

Title: The Use of Genetic Profiling to Guide Prostate Cancer Targeted Screening

Short Title: BARCODE 1 Pilot

Chief Investigator:	**Prof Ros Eeles** – Professor of Oncogenetics, ICR & RMH
Principal Investigator:	**Prof Ros Eeles** – Professor of Oncogenetics, ICR & RMH
Statisticians:	**Professor Sue Moss** – Queen Mary, University of London (Screening) **Dr Antonis Antoniou** – Cambridge University (Genetic Risk Modelling)
Co-investigators:	**Dr Elizabeth Bancroft, Natalie Taylor, Sarah Thomas** – Research Nurses, RMH **Elizabeth Page** – Study Coordinator, ICR **Sibel Saya** – Genetic Counsellor/**Dr Eva McGrowder -** Study Coordinator, ICR **Dr Zsofia Kote-Jarai** – Senior Staff Scientist, ICR **Dr Imran Rafi** – Senior Lecturer-Department of Community Health Care Sciences **Dr Michelle Ferris** – General Practitioner-Barnet Primary Care NHS Trust

	Prof Kenneth Muir – Professor of Cancer Epidemiology-University of Manchester
	Dr Nora Pashayan – Senior Clinical Lecturer in Applied Health Research-UCL
	Dr H Bowen-Perkins – GP; member of Primary Care Research Network (South East)
	Mr Pardeep Kumar – Consultant Urologist, RMH
	Dr Nick van As – Consultant Clinical Oncologist, RMH
	Prof David Dearnaley – Professor of Uro-Oncology, RMH & ICR
	Dr Vincent Khoo – Consultant Clinical Oncologist, RMH
	Clare Moynihan – Sociologist
	Audrey Arden-Jones – Senior CNS in Cancer Genetics, RMH
	Dr Christos Mikropoulos – Clinical Research Fellow, ICR
	James Taylor – Patient Representative
Collaborators:	**Prof David Nicol** – Consultant Urologist, RMH
	Mr Chris Ogden – Consultant Urologist, RMH
	Mr Alan Thompson – Consultant Urologist, RMH

	Prof Christopher Woodhouse – Consultant Urologist, RMH

Sponsor: The Institute of Cancer Research, 123 Old Brompton Road, London SW7 3RP

Clinical Research & Development: The Royal Marsden NHS Foundation Trust, Downs Road, Sutton, Surrey SM2 5PT

Study Sites: The Royal Marsden NHS Foundation Trust, Downs Road, Sutton, Surrey SM2 5PT

Protocol Reference:	CCR4130
Version Number & Date:	Version 2.1: 11/08/2016
Effective Date:	
Superseded Version Number & Date (If applicable)	

7.1.1 Contents Page

1. Background

7.1.2 Introduction – the genetics of prostate cancer

Prostate cancer (PrCa) is now the commonest cancer in men in the Western world, with over 40,000 new cases per annum and a lifetime risk of 1 in 11 in the United Kingdom (UK) (Cancer Research UK CancerStats, 2012), as well as in the European Union with 397,000 new cases per annum, and 94 000 deaths (Globocan). However, its aetiology remains very poorly understood. The substantial worldwide variation in incidence rates suggests that lifestyle risk factors are important. To date, however, no definite lifestyle risk factors have been identified.

Aside from demographic factors, the only established risk factor for PrCa is family history. Genetic studies, in particular genome-wide association analyses have identified 77 genetic variants associated with PrCa risk (reviewed in Goh et al, 2012; Eeles et al 2013). The risk of the disease in first degree relatives of cases is approximately twice that in the general population (Carter et al., 1992; Goldgar et al., 1994; Eeles et al., 1999; Hemminki et al., 2002; Gronberg 2003; Edwards and Eeles, 2004). This familial risk is greater amongst young cases, being more than fourfold for cases below age 60. Higher risks have been shown for men with two or more affected relatives. There is a higher risk in Afro-Caribbeans who have a 2.87-3.19-fold increased risk compared with whites in the UK (Ben-Shlomo et al, 2008). Analyses based on the Nordic twin registries have found higher risks in monozygotic than dizygotic twins, supporting the hypothesis that much of this familial aggregation is due to genetic rather than shared lifestyle factors (Lichtenstein et al., 2000).

Genetic predisposition arises from rare highly-penetrant mutations, and/or from common variants conferring more moderate risks. We, and others, have found the former using direct candidate gene mutation analysis (e.g. Dong et al., 2003; Edwards et al., 2003, 2012;

Guisti et al 2003; Cybulski et al., 2004; Kote-Jarai et al., 2011; Leongamornlert et al., 2012). Sequencing of a linkage region on 17q has revealed a high risk PrCa predisposition gene, *HOXB13* which has a relative risk of 4-20 in families and is present in about 3.4% of European populations (Ewing et al, 2012; Zu et al, 2012; Witte et al, 2013). Genome-wide association studies (GWAS) identify common variants, present in >5% of the population. In GWAS, susceptibility variants (usually single nucleotide polymorphisms (SNPs)) are identified by finding a difference in genotype frequency between cases and controls.

7.1.3 SNP Profiling

SNP genotyping is the measurement of genetic variations of single nucleotide polymorphisms (SNPs). It is a form of genotyping, which is the measurement of more general genetic variation. SNPs are one of the most common types of genetic variation. A SNP is a single base pair mutation at a specific locus, usually consisting of two alleles (where the rare allele frequency is >1%). SNPs are found to be involved in the etiology of many human diseases. Because SNPs are conserved during evolution, they have been proposed as markers for use in association studies (genome wide association studies-GWAS).

Routine genotyping is expensive and requires more DNA and resources and time for analysis. New technologies such as highly multiplexed ligation-dependent PCR (LD-PCR) combined with high throughput next generation sequencing (NGS) technology offer an attractive alternative. This technology is called mass genotyping by sequencing technology (MGST) and can check numerous SNP positions in a number of samples combined in a single assay. MGST has the capacity to accommodate at least 100 SNPs and up to 10000 samples per assay of the Illumina NGS device.

Based on the estimated relative risks of currently known SNPs, approximately 30% of the familial risk of PrCa can now be explained and the top 1% of the risk profile has a 4.7-fold risk compared with the average of the population. It is estimated that nearly 2000 SNPs may be associated with PrCa risk (Eeles et al., Nature Genetics 2013) and the proposed Oncoarray initiative which will run 600 000 SNPs in 80 000 PrCa blood DNA samples and controls (cases:controls in a 3:1 ratio) is likely to find further hits. We have shown using theoretical modeling that genetic profiling of 27 SNPs in a population rather than the use of an age cut-off of 55 years for PrCa PSA screening would predict that 16% of men could avoid screening at the expense of missing 3% of cases [125].

These results may have clinical implications for targeted screening and there are also potential implications for risk counselling. Individually each SNP confers a modest effect on relative risk, however, the combined effects of these SNPs are thought to be multiplicative and therefore may be substantial, and as other SNPs are identified it may be possible to define genotypes that are sufficiently predictive of risk to be useful clinically. MacInnis et al have described a model – the P model, which incorporates SNP data and family history [126].

Mathematical modeling has shown that if a population is genetically profiled for such variants, men who fall in the top 10% of the population genetic risk distribution have a 2.7-fold risk compared with the average of the total population. This risk is at a level where mammography is offered to women at higher breast cancer risk in populations. Taking this analogy, genetic profiling could therefore be considered to offer targeted prostate cancer screening in populations, hence this application which is at the cutting edge of translation of these findings.

7.1.4 Prostate Cancer Screening-present

PSA screening studies of the general population to date have reported conflicting effects on mortality from the disease.

To date there are several population based screening studies which have used a threshold of PSA to determine whether to undertake prostate biopsy (Andriole et al, 2009; Schroder et al, 2009; Hugosson et al, 2010; Schroder et al, 2012). The problem with PSA is that it has false positive and negative outcomes.

The European Randomised Study of Screening for Prostate Cancer (ERSPC) identified 182.000 patients through cancer registries from 7 European countries. In the screening group 82% of men received at least 1 screening PSA. The hazard ratio for death from prostate cancer was 0.8, which means that, in order to prevent 1 death from prostate cancer 1410 men need to be screened and 48 patients should be treated additionally for prostate cancer.[56] The Goteborg trial, a subset of the ERSPC, with 20000 men recruited, reported a 56% difference in favour of the PSA-screened arm.[127]

The PLCO study (prostate, lung, colorectal and ovarian study), recruited 76,685 patients in the US and randomised them to PSA screening and digital rectal examination or no screening. After 7 years of follow up the incidence of prostate cancer was 116 per 10000 person-years in the surveillance group and 95 per 10000 person-years in the control group. The rate of death from prostate cancer per 10000person-years was 2 for the screening group and 1.7 in the control group with no statistically significant difference, despite the higher rate of cancer diagnosis.[128] Andriole et al published the latest update with 13 years follow up and there was no evidence of mortality benefit for annual screening compared to opportunistic screening. [129]. On the contrary, Crawford et al, performing a subgroup analysis of the PLCO study data, showed that elective screening of individuals

with no major comorbidities led to a significant decrease in the prostate cancer specific mortality. [130]

ASCO issued recommendations in 2012 advising against screening older men with a life expectancy of less than 10 years. The American Urological Society published guidelines in 2012, offering an informed decision making on PSA screening only to the age group of 55 to 69 year old men.[131] The US Preventive Services Taskforce (USPSTF) took a step further discouraging routine PSA screening in all men.[132]

7.1.5 Targeted Screening
It is important to consider not just the number of cancers that are detected but the ability of a screening modality to distinguish between clinically significant disease, i.e. disease causing a significant risk to the patient's life or wellbeing, versus disease that would pose no threat if left untreated. The definition of clinically significant localized PrCa is defined using the NICE criteria for intermediate / high risk disease, which comprises a Gleason score of $\geq$7, and /or $\geq$T2b, N1, M1 (http://guidance.nice.org.uk/CG58).

The Targeted PSA Screening (TAPS) study looked at the feasibility of targeting screening at high risk groups (Melia et al, 2006) and identified a number of key issues. The aims of this study were to investigate the uptake rate of screening using prostate specific antigen (PSA) testing, and the referral rate in male relatives of men already diagnosed with PrCa below the age of 65 years. This study recruited relatives of men with PrCa aged between 45-69 years and contacted eligible men via their affected relatives. The results of the study found that discussing the study in person with PrCa patients yielded a higher recruitment rate compared with postal invites. They also found that there was a high level of previous PSA screening within this cohort. Interestingly they found that men were far more likely to opt for screening within the study if they were married / co-habiting versus men who were

single. The results of this study have important implications for the design of targeted screening programmes in higher-risk groups and highlights that further research is needed into the management of higher risk groups.

A study of men with at least one first or second degree relative with PrCa who underwent prostate biopsy showed that 25.3% had PrCa (Canby-Hagino et al., 2007). Nam et al (2009) studied the effect of 25 SNPs in men who had biopsy and PSA screening. In 3,004 patients, 1,389 (46.2%) were found to have PrCa. Fifteen of the 25 SNPs studied were significantly associated with PrCa on biopsy (P=0.02-7x10^{-8}). He selected a combination of 4 SNPs with the best predictive value for further study. After adjusting for other predictive factors, the odds ratio for patients with all four of the variant genotypes compared with men with no variant genotype was 5.1 (95% confidence interval, 1.6-16.5; P=0.006). When incorporated into a nomogram, genotype status contributed more significantly than PSA, family history, ethnicity, urinary symptoms, and digital rectal examination (area under the curve=0.74). The positive predictive value of the PSA test ranged from 42% to 94% depending on the number of variant genotypes carried (P=1x10^{-15}).

7.1.6 Biomarkers

PCA3 levels have been shown to correlate with positive prostate biopsies. Unlike serum prostate specific antigen the PCA3 score did not increase with prostate volume. PCA3 was independent of PSA reading s and previous biopsies. [133] The PCA3 score was significantly correlated with total tumor volume in prostatectomy specimens (r = 0.269, p = 0.008), and was also associated with prostatectomy Gleason score (6 vs 7 or greater, p = 0.005).[134] In a prospective European Study of 463 men the positive repeat prostate biopsy following an initial negative biopsy was 28%. It was found that the higher the PCA3 score, the greater the probability of a positive repeat biopsy. The PCA3 score (cut-off of 35) had a greater diagnostic accuracy than free:total PSA ratio.[135].

The presence of the genetic rearrangement between transmembrane-serine protease gene (TMPRSS2) and the erythroblast transformation-specific (ETS) member ERG (v-ets erythroblastosis virus E26 oncogene homolog avian) has been demonstrated in almost half of PCa cases. This gene fusion is considered to be an early event in PCa development. [136] The prognostic value of this translocation is unclear, as studies report conflicting results. Prostatectomy specimens from 294 PrCa patients were evaluated using FISH and rearrangement was observed in 56.6% of cases, and no association with biochemical progression or relapse free survival was found. [137] In another study of 208 radical prostatectomy specimens ERG expression was assessed with immunohistochemistry and was identified in 23.7% of the samples. ERG expression was twice more likely to be present in higher tumour stage and patients with ERG expression were twice more likely to die of prostate cancer. [138, 139]

7.1.7 Prostate Cancer Screening-Future
7.1.7.1 The IMPACT study

A different approach to prostate cancer screening aiming to reduce mortality and morbidity is needed. A case control study rising to this challenge is IMPACT (The Identification of Men with a genetic Predisposition to Prostate Cancer: Targeted screening in *BRCA1/2* mutation carriers and controls). This is an innovative targeted screening study, which will help us formulate a novel approach to improving prostate cancer related outcomes [140].

IMPACT is an international collaboration amongst 52 worldwide centres which has recruited 350 men with *BRCA2* and 500 men with *BRCA1* and 850 controls. The aim of this study is to investigate the role of PSA screening in *BRCA1/2* mutation carriers aged 40 to 69. PSA is checked annually and all men with PSA>3ng/ml are offered diagnostic 10 core

trans rectal biopsy. This study is now extended to include patients with Lynch syndrome, 190 men with *MSH2*, *MSH6* and *MLH1* (MMR genes) and 190 controls [141].

The IMPACT trial group published data from the first 300 patients with 89 *BRCA1*, 116 *BRCA2* and 95 controls with median follow up of 33 months. The prevalence of prostate cancer was 3.3% with the positive predictive value of PSA screening of 47.6%. This preliminary report re-enforces the value of PSA screening in *BRCA1/2* mutation carriers [142].

7.1.7.2 The PROFILE Pilot study

The aim of the PROFILE study is to correlate germline genotypes in men with an increased risk of PrCa due to a genetic predisposition with biopsy outcome and also to assess the additional contribution of DW-MRI and new biomarkers to PrCa screening in this group. An initial pilot has been completed. The aim of the pilot PROFILE study was to conduct a feasibility study in 100 men with a positive family history of PrCa (at least one first degree relative affected at <70 years, with diagnosis verified) to determine the interest in the study, biopsy uptake and complication data. The rationale behind the study design of this protocol was identifying at risk groups based on family history and retrospectively profiling rather than taking a specific SNP profile as a criterion for screening and biopsy. The pilot PROFILE study recruited eligible men aged 40-69 years with a family history of PrCa over a two year period. After informed consent, patients provided blood samples to measure PSA level and for DNA extraction. All participants were asked to undergo a 12   core prostate biopsy regardless of baseline PSA result. Participants without previous prostate biopsy or who underwent biopsy >1 year ago were also offered a T2-weighted with DW-MRI prior to biopsy in 50 of the participants.

In total 116 men were recruited and 102 biopsies completed. All patients were asymptomatic. Based on SNP analysis of 39 PrCa risk SNPs, a total of 53 men had a predicted relative risk <1 (median age 55 yrs; median PSA 1.20). In this subgroup, 8 men (15.1%) were diagnosed with PrCa (median age 62.0 yrs, median PSA 2.50). Amongst the 48 men with a relative risk >1 (median age 51.0 yrs; median PSA 1.4) 13 PrCas (27.1%) have been identified (median age 56.0 yrs, median PSA 2.7). T2 weighted in conjunction with DW-MRI had 33% false positives and 10% false negatives. The AUC of T2 weighted in conjunction with DW-MRI was 0.83. Twelve men with PrCa had a PSA <3 (52%). No adverse psychosocial variables were noted.

The main conclusion from this pilot study was that prostate biopsy as a means of PrCa screening is feasible and acceptable in men with a family history of PrCa. The findings support a larger study investigating the use of SNPs in PrCa risk stratification for targeted screening. The PROFILE study is currently being rolled out to include 2 cohorts, the first one would address the issue of family history and the second cohort the issue of African-Caribbean ancestry.

2. Study overview and rationale

The BARCODE 1 study has been developed to investigate the role of genetic profiling for targeting population screening. This study forms a pilot of 300 men, with the view to continue to a future study of 5000 men.

The primary endpoint is the association of biopsy result with genetic risk score in men having targeted prostate screening based on SNP risk profiling.

Secondary endpoint would be the comparison of results with population based PSA screening will be analysed to determine if there is a higher proportion of clinically significant

disease than is identified in population based studies. This study will also explore the acceptability and logistical issues around using genetic profiling on a population level to target a cancer screening programme.

Initially we aim to recruit 300 men with the assistance from participating GPs. Men aged 55-69 years who are likely to be eligible for the study will be identified by their GP from medical records. They will be contacted via a letter from their GP and if interested in the study will be asked to fill in a questionnaire to confirm they are eligible to participate. This questionnaire can be completed in hard copy and men will also be given the option to fill in an online version. If eligible, men will then be sent a DNA collection kit. Saliva kits will be analysed with SNP profiling for the known 99 clinically relevant SNPs. Men with a genetic risk equivalent to the top 10% of the population distribution (approximately 30 men in total) will be invited for a TRUS prostate biopsy, plus further biological samples. Biopsy results will be correlated with the genetic score. PSA and other biomarkers will be integrated into results to assess combined effects of genetic score and markers.

It is well known that the response rate to questionnaires sent in the post is low, but with the support of primary care practitioners and also with the option of an online eligibility questionnaire, we aim to improve the uptake rate. This will be a pilot study and will help us to identify problems with recruitment and SNP profiling. Saliva kits have been shown to yield enough extracted DNA to perform SNP profiling for all the candidate alleles. Provided that the initial cohort is recruited smoothly and the top 10% successfully undergoes biopsies then we aim to expand the study.

The expanded study will recruit a total of 5000 men and men with a genetic risk equivalent to the top 10% of the population distribution (approximately 500 men in total) will have been invited for a TRUS prostate biopsy, plus further biological samples.

Additional blood, urine, saliva and tissue samples will be taken for research purposes in order to investigate new biomarkers in this population using biochemistry, proteomic, metabolomic and microarray approaches. Samples will be collected from urine for further studies, for example biomarker studies PCA3 and the TMPRSS2 ERG translocation to correlate these with SNP profile, but biopsy decisions will not be made on these results.

All participants will also be invited to participate in a sub-study that aims to provide valuable information about the psychosocial and behavioural impact of genetic risk-profiling in the general population, deduce information needs of men undergoing testing and develop decision support tools accordingly. The results of this sub-study will be used alongside the results of the pilot BARCODE 1 study to inform the design of the main BARCODE 1 study in which the decision support tools will be trialled and refined. The background, rationale, methodology are outlined in detail in the BARCODE 1 psychosocial sub-study protocol. A separate consent form is also provided for this sub-study.

3. Aims

7.1.8 Primary aim:
- To determine the association of biopsy result with genetic risk score in men having targeted prostate screening based on SNP risk profiling

7.1.9 Secondary aims:
- To determine the incidence and aggressiveness of PrCa in men within the top 10% of the genetic score

- To determine the association of biological sample biomarker profile with prostate biopsy result in men at genetically higher PrCa risk undergoing targeted PrCa screening

4. Study Design

This screening study is designed to look at the role of genetic profiling for targeting population screening. The aim is to evaluate genetic profiling using the known 99 SNPs as a means for offering targeted screening for PrCa in men at a genetically higher risk. Additionally the study aims to integrate serum and/or urine markers to genetic profiling for those identified in the top 10% risk category. Biomarkers with established prognostic value will be checked including PSA, PCA3, hK2 and free: total PSA ratio.

5. End Points

7.1.10 Primary Endpoint:
- To determine the association of SNP genetic risk score with prostate biopsy results

7.1.11 Secondary Endpoints:
- To assess the incidence and the aggressiveness of prostate cancer amongst men within the highest 10% of the genetic score of new markers including quantitative imaging biomarkers e.g. apparent diffusion coefficient metrics and their association with the results of prostate biopsy
- To determine the association of the biomarker profile with genetic score and biopsy results
- To explore the use of genetic profiling to target prostate cancer screening in a clinical environment

6. Inclusion/ Exclusion Criteria

7.1.12 Number of subjects:
- 300 men willing to undergo genetic SNP profiling

7.1.13   Inclusion Criteria:

- Men aged 55-69 years

- Caucasian ethnicity

- WHO performance status 0-2 (see Appendix A)

- Absence of any psychological, familial, sociological or geographical situation potentially hampering compliance with the study protocol and follow-up schedule

7.1.14   Exclusion criteria

- Non-Caucasian ethnicity (including mixed race or Jewish)

- Previous diagnosis of cancer with a life-expectancy of less than five years

- Prostate biopsy in the past year

- Previous diagnosis of prostate cancer

- Co-morbidities making prostate biopsy risk unacceptable (anticoagulants or antiplatelet medication like Warfarin or Clopidogrel, poorly controlled diabetes or cardiovascular disease)

7.1.15   Subject Withdrawal

- Subjects may withdraw from the study at any time if they so wish without giving a reason. No further data will be collected about that individual, and any unused samples will be destroyed. Data collected up to that point will be retained for audit purposes.

7. Methodology

Please refer to Figure 1: Pilot study algorithm on page 249 for a summary of the study methodology. For the purpose of the pilot study we aim to recruit 300 men aged 55-69 years. Eligible men will be identified through their General Practitioners. They will receive information in lay terms about the study through the post. Those who consent and are eligible will be sent a saliva collection kit to provide a DNA sample.

Genetic profiling of known prostate cancer predisposition SNPs will be performed. Those within the top 10% of the risk profile will be offered a clinic appointment to discuss prostate cancer screening and will be offered a prostate biopsy. All men, regardless of their genetic

profile score, will be followed for five years via cancer registries to track development of cancer in the future.

All men will also be invited to participate in the psychosocial sub-study, regardless of their genetic profile result. The methodology is outlined in detail in the sub-study protocol. Men will be invited to complete four questionnaires over the course of the study, the first upon enrolment in the main study, the second following receipt of the genetic risk profile results, the third at 6 months following these results and the fourth at 12 months following these results.

300 men (in total)
aged 55-69 years
DNA extracted from a saliva kit

Stage 1

SNP Profiling

1. DNA extraction
2. Profile of 99 known PrCa predisposition SNPs
3. Analysis and modelling of risk profile

Top 10% of genetic risk from SNP profile invited to clinic for:

1. PSA and F/T PSA ratio (PCA3)
2. Biological samples (blood/urine)
3. MRI pelvis when clinically indicated

Stage 2

If biopsy refused

Suggest discussion of PSA screening with GP

Prostate Biopsy:
12 core biopsy plus targeted biopsies based on US/MRI findings plus 2 research biopsies

END OF STUDY

Prostate Cancer

PIN or ASAP**

Benign

Specialised Treatment
Collection of treatment data for study

Repeat Biopsy +/- MRI (within 6 months for PIN or ASAP**)

Abnormal PSA result*

Six-monthly PSA screening for five years

* Abnormal PSA is >3ng/ml or a >50% increase if last reading was >3ng/ml with

normal biopsy

**ASAP: Atypical small acinar proliferation, PIN: Prostatic intraepithelial Neoplasia

Enrolment

7.1.16.1   Stage 1

Initially we plan to run a pilot study with 300 men. For further details on the recruitment process, pages 251 & 252 (Figure 2: Recruitment Algorithm and Figure 3: Eligibility Letter algorithm). We aim to recruit patients via GPs. An invitation letter with a participant information sheet, an eligibility questionnaire and consent form will be sent in the post through the GP surgery.

The option to fill in the questionnaire and give provisional consent to the study online will also be provided. Participants who complete the questionnaire online will also be asked two additional questions about this experience. This information will be used when considering the use of online tools for the main BARCODE 1 study.

Once the questionnaire is received, the study team will determine if the patient is eligible and reply via letter. If necessary, a member of the study team will contact the patient to clarify information from the questionnaire and determine if he would be suitable for a prostate biopsy.

If the patient is eligible for the study, a saliva kit will be sent to provide a DNA sample. DNA extraction and SNP profiling will follow. The participant will be informed by letter of whether they fall into the top 10% risk category or not. Those in the top 10% risk category will then receive a follow-up telephone call before attending the hospital to discuss screening options and be offered a prostate biopsy.

Figure 2: Recruitment Algorithm

Study team receives eligibility questionnaire

Responder has unclear medical history and requires phone call from study team

Responder is ELIGIBILE for the study as:
- Is aged 55-69
- No relevant medical history that would preclude biopsy
- Is Caucasian
- Has not had a prostate biopsy in the past year

See next page

Responder is NOT ELIGIBLE for the study due to one or more of the following reasons. Send letter to responder and corresponding letter to GP as indicated below:

Has the responder…

...had a prostate biopsy <1 year ago?

Family history of cancer?

No — *Letter 2b*

Other cancers — *Letter 2c*

Prostate +/- other ca's — *Letter 2d*

...a medical history contraindicating biopsy?

Family history of cancer?

No — *Letter 2e*

Other cancers — *Letter 2f*

Prostate +/- other ca's — *Letter 2g*

...non-Caucasian/Jewish ancestry?

Other

Family history of cancer?

No — *Letter 2h*

Other cancers — *Letter 2i*

Prostate +/- other ca's — *Letter 2j*

Afro-Caribbean

Family history of cancer?

No — *Letter 2k Eligible for PROFILE*

Other cancers — *Letter 2l Eligible for PROFILE*

Prostate +/- other ca's — *Letter 2m Eligible for PROFILE*

Jewish

Family history of cancer?

No — *Letter 2n*

Other cancers — *Letter 2o*

Prostate +/- other ca's — *Letter 2p*

Figure 3: Eligibility Letter algorithm

252

Responder is ELIGIBILE for the study as:

- Is aged 55-69
- No relevant medical history that would preclude biopsy
- Is Caucasian
- Has not had a prostate biopsy in the past year

Send Oragene kit for DNA collection (*Letter 2a* and *GP letter 2a)* then SNP profiling

Top 10% of genetic profile

Send *Letter 3a* and *GP letter 3a*

Bottom 90% of genetic profile
END OF STUDY
Send letter to participant and GP as indicated:

Family history of cancer?

| No | *Letter 3b* |

| Other cancers | *Letter 3c* |

| Prostate +/- other ca's | *Letter 3d Eligible for PROFILE* |

Accept biopsy

See **Pilot study algorithm** on p249

Refusal of biopsy
END OF STUDY

Family history of cancer?

| No | *Letter 3e* |

| Other cancers | *Letter 3f* |

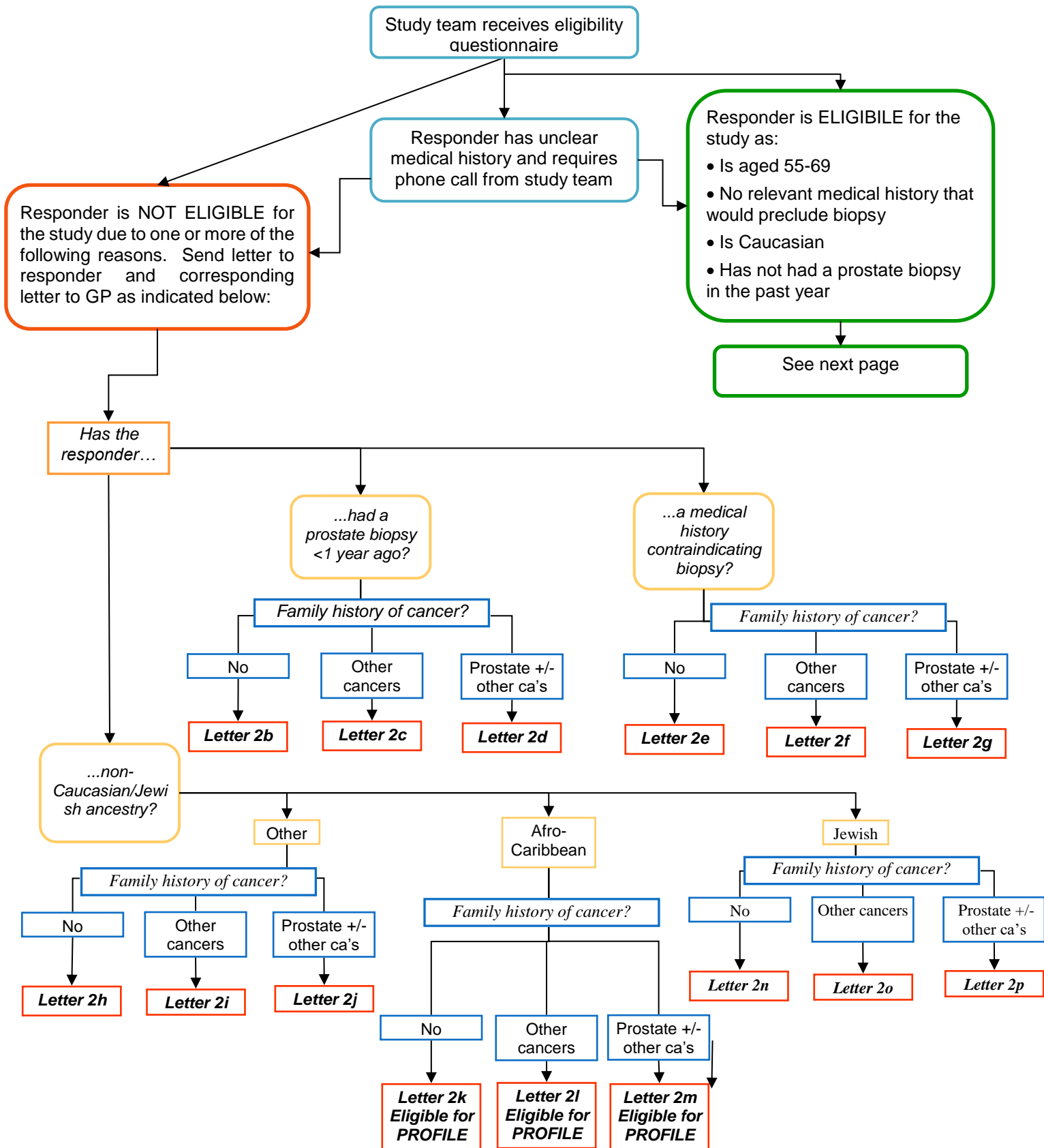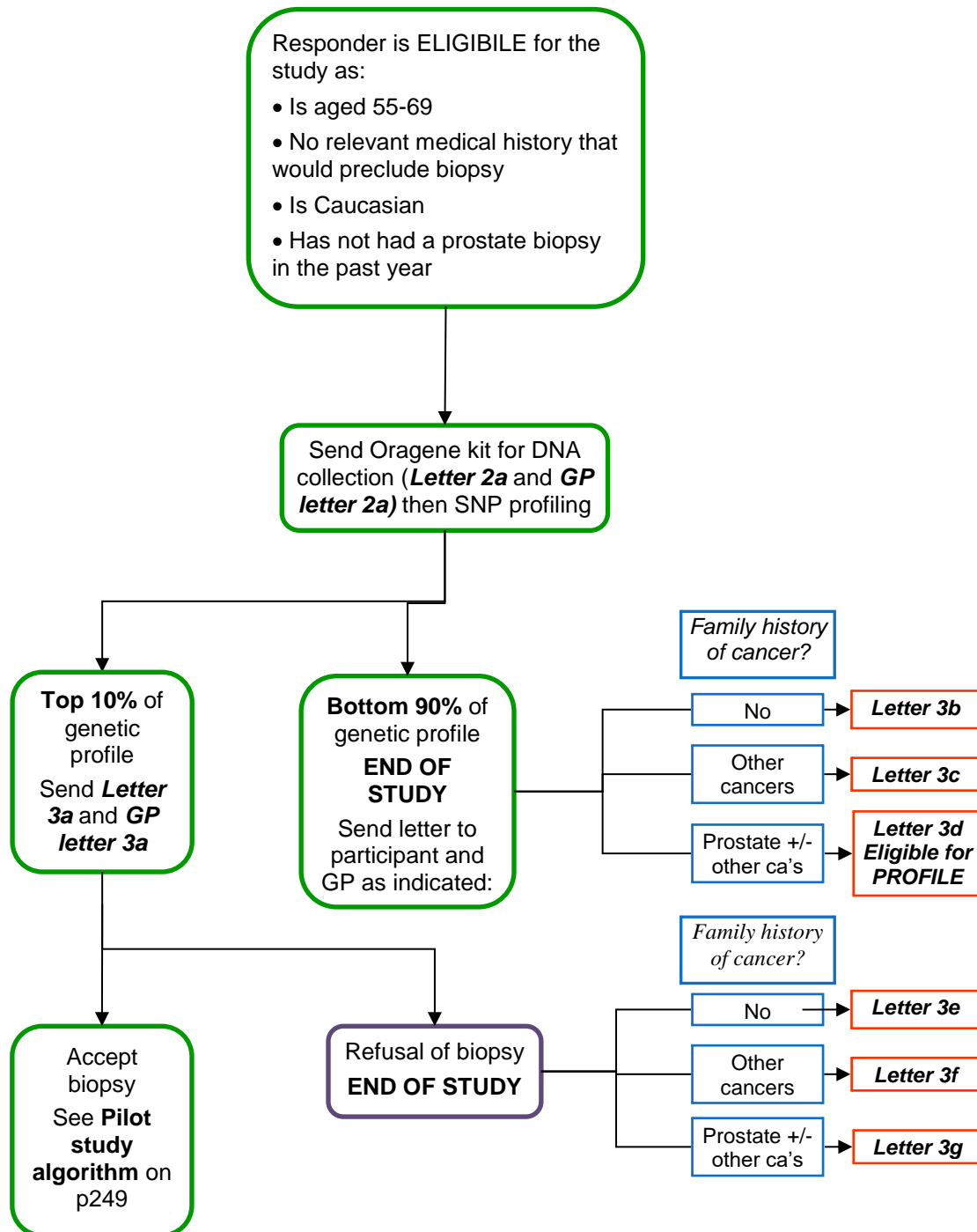| Prostate +/- other ca's | *Letter 3g* |

7.1.16.2 Stage 2

In the second stage, the 30 men in the top 10% of the genetic risk score will be offered a hospital clinic appointment (see page 249, Figure 1: Pilot study algorithm). Those in the bottom 90% of the risk score would have no further follow up as part of the study. If they have a family history of prostate cancer, they will be referred via letter back to their GP, and will also be offered entry into the PROFILE study (CCR4045, REC reference 13/LO/1787).

Those in the top 10% will be offered four or five hospital appointments at the Royal Marsden Hospital, London. During their first appointment, which will last approximately 45-60 minutes, the participant will be offered a prostate biopsy and also a discussion about the pros and cons of other types of prostate cancer screening, for example PSA testing. They will be counselled about all the study procedures including the potential side-effects of biopsy. Men can decide whether they would like proceed with the study and the biopsy during this first appointment or they can have the opportunity to go away and consider their options. For the latter, if the team has not heard from them within two weeks they will be telephoned to answer any further questions and to either schedule another appointment or to confirm that they do not wish to take part in the study further.

7.1.16.3 Biological Samples

Biological samples, including blood and urine, will be taken in order to measure PSA levels and the free:total PSA ratio at an appointment prior to biopsy. Research blood and urine samples will also be taken and stored for the study. These samples are taken at the Royal Marsden Hospital, London by trained clinical research fellows and research nurses.

7.1.16.4 MRI prostate

MRI prostate will be offered to participants when that is considered clinically relevant and to those who are medically suitable.

7.1.16.5 Digital Rectal Examination

A digital rectal examination ideally should be performed in all participants who are being considered for a prostate biopsy, provided that they consent to it. A prostatic massage is also necessary, in order to acquire first pass urine for PCA3 and TMPRSS-ERG assays.

7.1.16.6 Prostate Biopsy

A twelve core trans-rectal ultrasound guided biopsy (see Appendix C) will be taken for diagnostic purposes (with additional targeted biopsies where appropriate) and a further 2 samples obtained for research. Consent to take the 2 extra samples for research will be sought before the biopsy procedure commences (optional for patient) and will be immediately snap frozen in dry ice for future DNA and RNA analyses.

In the case of a visible anterior prostate suspicious lesion on MRI, then a template biopsy upfront would be preferable, in view of the risk of a false negative TRUS biopsy in this setting.

All biopsies will be reviewed by one pathologist at the Royal Marsden Hospital using an agreed standardised procedure for our unit's research studies (see Appendix D). If any of the 12 cores identify the presence of PrCa, the subject will receive treatment as advised by their local hospital if they do not wish to have treatment at the Royal Marsden Hospital. All cases will be scanned into a virtual central review database for review by a panel of expert urological pathologists.

Those cases where the first biopsy detects Atypical Small Acinar Proliferation (ASAP) or High Grade Prostatic Intra-epithelial Neoplasia (HG-PIN) will be re-biopsied within 6 months, or sooner according to local (Royal Marsden) guidelines. A repeat DW-MRI could be performed, adding in extra cores depending on the MRI appearance. The repeat biopsy will either be a template biopsy or TRUS biopsy depending on the MRI findings.

7.1.16.7　Outcome of template biopsy

1. Prostate cancer – treatment as advised by local centre (likely the Royal Marsden Hospital, unless patient wishes to be treated local to their home)
2. ASAP / HG PIN detected – repeat DW-MRI and biopsy in 6 months to 1 year.
3. No abnormalities identified - PSA follow up 6 monthly for 5 years

7.1.16.8　Refusal of prostate biopsy

For those who do not wish to proceed with prostate biopsy, we will write to them and their GP to recommend a discussion about the option of PSA screening. Their care will be discharged to their GP.

7.1.16.9　Follow-up after normal biopsy

In light of their increased genetic risk of prostate cancer, men who have a normal biopsy will be offered six-monthly PSA for five years as part of the study. The algorithm for further biopsy has been piloted in our PROFILE pilot study. If men have a PSA >3ng/ml or a >50% increase after a previous PSA >3ng/ml with a normal biopsy, they will be offered a repeat biopsy.

7.1.16.10　Study follow-up

For men in the top the 10% of risk profile, we will request an update on their medical history from their GP for five years after study completion. For all men, we will monitor any development of cancer through cancer registries.

7.1.16.11 *Review of genetic profiling results*

As new information becomes available about the genetic basis of prostate cancer it is possible that we may revisit samples from study participants, for example if a new gene causing prostate cancer is discovered. Where this new information is felt to be clinically relevant we will re-contact the men involved in the study to inform them of the results.

8. Data Acquisition

7.1.17 Stage 1a: Eligibility and Enrolment
- Sign study consent form (hard copy) OR
- Provide provisional consent for the study (electronically)
- Complete eligibility questionnaire that includes information about family history and medical history (electronic or hard copy)

7.1.18 Stage 1b: SNP Profiling
- Sign study consent form (hard copy for those not already provided)
- Saliva sample given
- SNP profiling performed on DNA extracted from saliva
- The genetic profiling results will be disclosed to the participant via letter then via telephone for those in the top 10% of the profile.

7.1.19 Stage 2a: Enrolment (for those in top 10% of genetic profile)
- Offer prostate biopsy and imaging
- Provide blood samples for PSA testing and other biomarkers and 30ml urine sample (first pass) pre and post prostatic massage for PCA3 and other studies (Appendix B – Guidelines for Sample Collection)

7.1.20 Stage 2b: At Biopsy
Each subject will complete the following:

- Sign the local hospital biopsy consent form
- Sign the study-specific biopsy consent form
- Sign MRI consent form (for those offered)

7.1.21 Stage 2c: Biopsy Results

- The biopsy results will be discussed with the participant either in person or by telephone (for negative results only and at the clinician's discretion)

7.1.22 If PrCa is diagnosed

The staging and further investigation of the disease is as directed by the collaborating uro-oncology unit. Management is based on the immediately available pathology report, not on the later central review.

Minimum information required by the study centre will be:

- Clinical T stage
- Gleason score of biopsy and extent of involvement (in percentage of tissue involved an absolute length of core in millimeter)
- Treatment and management plan
- Radiological TNM stage
- Histopathology report for men undergoing radical surgery
- Slides should be scanned into PathXL for central review after the local clinical report has been issued
- Following a diagnosis of PrCa, a treatment questionnaire will be required annually

7.1.23 Stage 3: Study follow up

We will follow all men in the study for development of cancer either through their GP or cancer registries.

7.1.24 Potential adverse events
7.1.24.1 *Side-effects of biopsy:*

TRUS biopsy should be carried out in accordance with the study protocol (Appendix C) and antibiotic prophylaxis should be given as per local (Royal Marsden) hospital protocol.

The procedure is uncomfortable and associated with the following risks

- Painful or difficult voiding   13%
- Haematuria   11%
- Fever/sweats 6%
- Septicaemia   3%
- Acute urinary retention   1%

(Taken from Crundwell et al, 1999)

For this reason subjects will be followed carefully and be able to contact the urology department in case of problems.

7.1.24.2   *Venepuncture*

Venepuncture a risk of

- Feeling faint,
- Bruising at venepuncture site,
- Excessive bleeding,
- Hitting a nerve,
- Hitting an artery

The procedure should be carried out by those with adequate training and in accordance with local (Royal Marsden) hospital protocol.

9. Data Analysis

- All biopsy interventions and results will be reported to the data centre as they occur. Biopsy results will be reviewed by a central team of pathologists.
- PrCa diagnosis will be reported immediately. The diagnosis and treatment will be based on histological confirmation. A later research central review will be undertaken by a central team of pathologists. If there is disagreement the local diagnosis will be the overriding one for treatment.
- Data completeness (Questionnaires and CRFs) will be evaluated
- Initial translational studies will use the stored serum/urine samples.
- An Independent Data Monitoring Committee will review the study data 6 monthly

10. Study Organisation/ Trial Monitoring and Management Strategy

7.1.25 Administrative Responsibilities

The CI, Clinical Fellow and Study Coordinator (in cooperation with the Data Centre) will be responsible for writing the protocol, submitting to the Committee for Clinical Research and for local management R&D approval, reviewing all case report forms and documenting evaluation forms, discussing the contents of the reports with the Statistician, and for writing the draft of the study results. The CI will also generally be responsible for answering all clinical questions concerning eligibility, treatment, and the evaluation of the subjects.

7.1.26 Steering Committee

It will be the responsibility of the CI to report changes to the protocol and data updates to the study Steering Committee. The Steering Committee will consist of the co-investigators as described in the first page.

7.1.27 Independent Data Monitoring Committee (IDMC)

An Independent Data Monitoring Committee will be set up to regularly review and scrutinise available data and advise on appropriate action.

11. Adverse Events

7.1.28 Definitions

Adverse Events (AE) are any untoward medical occurrence or experience in a patient or clinical investigation subject which occurs following participation in the trial regardless of the causal relationship. This can include any unfavourable and unintended signs or symptoms, an abnormal laboratory finding (including blood tests, x-rays or scans) or a disease temporarily associated with the use of the study, for example:

- death

- a life-threatening event (i.e. the subject was at immediate risk of death at the time the reaction was observed)
- hospitalisation or prolongation of hospitalisation
- persistent or significant disability/incapacity
- any other medically important condition (i.e. important adverse reactions that are not immediately life threatening or do not result in death or hospitalisation but may jeopardise the subject or may require intervention to prevent one of the other outcomes listed above)

12. Reporting procedure

7.1.29 Non-serious adverse events

All Adverse Events (AE), occurring during the study until the end of the period of follow-up must be recorded on an adverse event form. All adverse events will be reported to the data centre and logged in accordance with to the local sites Standard Operating Procedures for Adverse Events.

The Chief Investigator will decide if those events are related to the study intervention (i.e. unrelated, unlikely, possible, probable, definitely and not assessable) and the decision will be recorded on the adverse event forms. AEs definitely not study related (i.e. reported as unrelated) will not be considered as adverse events in study analyses, but reported separately. The assessment of causality is made by the investigator using the following definitions:

Relationship	Description
Unrelated	There is no evidence of any causal relationship
Unlikely	There is little evidence to suggest there is a causal relationship (e.g. the event did not occur within a reasonable time after administration of the trial medication). There is another reasonable explanation for

	the event (e.g. the subject's clinical condition, other concomitant treatments).
Possible	There is some evidence to suggest a causal relationship (e.g. because the event occurs within a reasonable time after administration of the trial medication). However, the influence of other factors may have contributed to the event (e.g. the subject's clinical condition, other concomitant treatments).
Probable	There is evidence to suggest a causal relationship and the influence of other factors is unlikely.
Definitely	There is clear evidence to suggest a causal relationship and other possible contributing factors can be ruled out.
Not Assessable	There is insufficient or incomplete evidence to make a clinical judgement of the causal relationship.

7.1.30   Serious adverse events

All Serious Adverse Events (SAE), related or not to the study, occurring during the study period and within 30 days after the last study intervention (eg. biopsy) will be reported and logged in accordance with to the local sites Standard Operating Procedures for Adverse Events.

Original SAE reports will be filed in the BARCODE 1 trial masterfile.

13.   Statistical Considerations

7.1.31   Sample size

This is a pilot study of 300 men, with approximately 30 in the top 10% genetic risk category. This sample size has been chosen pragmatically to allow us to recruit 100 men from each of the 3 collaborating GP practices.

For the main study (to be a submitted for approval on completion of the pilot study), a sample size of 5000 men will be required to identify the approximately 500 men within the top 10% of the genetic risk score, who will undergo a prostate biopsy. Power calculations show that 99 SNPs have a polygenic variance of 0.44 and therefore those in the top 10% risk category will include 27% of PrCa cases. We estimate that based on our PROFILE pilot study data that we will identify between 38 to 49 new PrCa in the 500 men biopsied, 3 times that of population screening. Our clinically significant cancers would be at least double that expected by PSA screening.

7.1.32   Recruitment timeframes

It is anticipated that the study will complete recruitment within 24 months. The study team will meet monthly to discuss recruitment and will report to the Steering Committee and Data Monitoring Committee six monthly. If there are problems with meeting the target recruitment this will be discussed at the Steering Committee meetings.

7.1.33   Descriptive Statistics

In this pilot study we will be using a variety of descriptive statistical tools to analyse our data. We will first calculate the uptake of the genetic test and also the proportion of men finally accepting a prostate biopsy. We will calculate the Positive Predictive Value (PPV) of prostatic biopsy based on genetic profiling and the cancer detection rate. Both these variables will be treated as binomial probabilities with exact confidence intervals. We will compare our findings to published data from other screening studies. Furthermore, we will assess the correlation of the calculated risk score to our findings.

7.1.34 End of study

The end of study is defined as the date of the last appointment of the last participant.

14. Regulatory & Ethics Committee Approval

7.1.35 Subject protection

The responsible investigator will ensure that this study is conducted in accordance with the Good Clinical Practice (GCP) guidelines, the Data Protection Act 1998 (DPA) and the Human Tissue Act 2004 (HTA) and Codes of Practice for consent issued by the Human Tissue Authority.

All staff at each Trust are required to abide by the Data Protection Act 1998 and also in accordance with the Confidentiality Code of Practice and Data Protection Policy and Procedure. The protocol will be approved by the Committee for Clinical Research at the Royal Marsden NHS Foundation Trust and Institute of Cancer Research and the Research and Ethics Committee.

7.1.36 Subject identification

Once men are found to be eligible for the study and consent to participant, a sequential identification number will be automatically attributed to each subject registered in the trial. This number will identify the subject and must be included on all case report forms. In order to avoid identification errors, subjects' initials (maximum of 4 letters), date of birth and hospital number (if available) will also be reported on the case report forms.

7.1.37 Informed consent

All subjects will be informed of the aims of the study, the possible adverse events, the procedures and possible hazards to which he will be exposed. Each participant will be informed about the strict procedures used to protect the confidentiality of his patient data, and that his medical records may be reviewed for trial purposes by authorised individuals other than their treating physician.

It will be emphasised that participation is voluntary and that the subject is allowed to refuse further participation in the protocol whenever he wants. This will not prejudice the subject's subsequent care. Documented informed consent must be obtained, according to the principals of GCP, for all subjects included in the study before they are registered at the Data Centre.

The informed consent procedure must conform to the ICH guidelines on Good Clinical Practice. This implies that "the written informed consent form should be signed and personally dated by the subject or by the subject's legally acceptable representative".

7.1.38   Assessment of family history provided to the study

A brief family history will be collected for the purposes of the study, however the study team will **not** provide a detailed, clinical assessment of this family history as part of the study as this is a clinical service beyond the remit of a research team. However, given the nature of the study and that it is being conducted by a genetic research group, it would be irresponsible not to highlight when a referral to a clinical genetics service may be warranted based on reported family history. All family histories reported by participants will be checked by a genetic counsellor/nurse and if a referral to a clinical genetics department is indicated, it will be suggested to the participant to discuss with their GP.

7.1.39   SNP profiling in the study and its clinical utility

It is highlighted throughout the participant materials that the genetic testing (i.e. SNP profiling) provided in the study is a **research** test and as yet the results are not fully understood and are subject to change. Assessment of genetic risk based on SNP profiling is **not** currently a technique used within clinical genetic practice in the UK. The research team will only suggest a referral to a clinical genetics service based upon strength of family history, regardless of SNP profile result.

7.1.40 Queries and concerns of participants and their family members based on SNP profiling result

It is possible that participants and/or their family members will have queries or concerns if the participant falls within the top 10% risk category of the SNP profile. These concerns are very valid given the genetic nature of the risk that is being tested and reported. By its nature, the risk described by SNP profiling is inherited is a very different manner to a monogenic mutation that confers a large increase in risk by itself. The full clinical implications for a family member of someone who has a high risk based on a SNP profile are not fully understood. Given this, clinical assessment of risk is currently based upon family history in the absence of a known high penetrance mutation in a family and a referral to a clinical genetics service for assessment of such a risk would not be appropriate.

The study team comprises of genetic specialists, including geneticists, genetic counsellors and genetic nurses who have extensive experience in clinical genetics and translational genetics research studies and so are well versed in conversing with patients about these issues. Any concerns raised by the participant or their family about the SNP profiling results can ably be discussed by the research team.

7.1.41 Over diagnosis of prostate cancer

One limitation of prostate screening is the detection of PrCas that would not otherwise have been detected and that may not be of clinical significance. However, these are cohorts of men at genetically higher risk of PrCa. Based on our pilot data we expect up to 50 new PrCa diagnoses in the 500 men biopsied. We estimate that the number of PrCa diagnosed will be 3 times that detected with population screening and the number of clinically significant cancers 2 times the number with population screening. Therefore while this risk of over diagnosis is recognised it is felt to be justified in this particular cohort. This will be discussed with every participant during the consent process as well as all potential treatment options.

We are currently not in a position to predict which of the low grade PrCa diagnosed will develop into a more aggressive tumour. Active surveillance follow up strategies aim to address this issue. Patients are followed up regularly and undergo repeat imaging and prostatic biopsies at regular intervals. Active surveillance is a safe and less invasive approach compared to radical treatment. Biomarkers predicting for an aggressive phenotype are currently in development and will make surveillance much easier. Increasingly, there is an argument that early diagnosis can have a positive impact on outcomes.

15. Data Handling, Record Keeping and Study Samples

7.1.42 Control of data consistency

Data forms will be entered in the database at the Data Centre. Computerised and manual consistency checks will be performed on newly entered forms; queries will be issued in case of inconsistencies. Consistent forms will be validated by the Data Manager to be entered on the master database. Inconsistent forms will be kept "on-hold" until resolution of the inconsistencies.

7.1.43 Use of online data collection

Those who choose to fill in the eligibility questionnaire online will do so via an online interface. The online system utilised will comply with EU data protection requirements (encompassed by the Data Protection Act in the UK), will be approved by the sponsor's (The Institute of Cancer Research) IT Security team and will undergo a Privacy Impact Assessment.

7.1.44 External review of histology

Histological assessment of prostate biopsies is subject to inter observer variation, particularly with reference to assessing Gleason grade. For this reason biopsies will routinely be reviewed and representative samples will be re-examined by the study pathologists. Clinical decisions will be based on local assessment and a routine review

to confirm diagnosis will not be required. If the review in retrospect reports a cancer which was not reported locally then this case will be subject to expert pathological review by the study panel pathologists in conjunction with the local reporting pathologist and an MDT decision taken as to the outcome.

7.1.45 Transfer and storage of data

The Data Centre is the Oncogenetics Team at the Institute of Cancer Research, Surrey. Electronic data will be stored on the ICR network which is routinely backed up. Hard copy data with identifiers will be stored in locked, fireproof cabinets within the ICR, with access limited to staff working on the study who are trained in Data Protection policies and legislation. Transfer of data between the Data Centre and recruitment sites (i.e. GP practices) will take place using password protected files via encrypted Iron Key or encrypted email. Passwords will be communicated by a separate method as per ICR data protection policy.

7.1.46 Retention and destruction of data

Raw data will be retained for 30 years. The Institute of Cancer Research and Royal Marsden Hospital guidelines for archiving of data resulting from non-clinical trials refer to the Research Governance Framework for Health and Social Care in the UK (2005). Clause 2.3.5 of this policy states:

"Data collected in the course of research must be retained for an appropriate period, to allow further analysis by the original or other research teams subject to consent, and to support monitoring by regulatory and other authorities."

Given the study is examining genetic factors associated with cancer and cancer risk, data will be kept for a period of 30 years. Our rapidly changing and growing knowledge of cancer genetics indicates that the genetic results from the study will also evolve and further incidental findings may result. Further review of the data with updated information may be necessary in the future.

Following this period, data will be destroyed according to The Institute of Cancer Research policy, with all hard copy data shredded and electronic data deleted to MoD standards.

7.1.47 Collection, Transfer and Storage of Samples

Blood, urine and biopsy samples will be collected at the Royal Marsden Hospital by trained clinical research fellows and research nurses. Saliva samples are provided by the participant using well validated collection kit in their own home and sent to the study team at The Institute of Cancer Research, complying with biological sample transfer guidelines of the European Agreement concerning the International Carriage of Dangerous Goods by Road (ADR).

Biological samples will be stored at The Institute of Cancer Research, Surrey in accordance with the joint Institute and Royal Marsden Hospital policy for removal, storage, use and disposal of human tissue for research, Blood, urine and biopsy samples will be stored in -80 freezers and saliva samples at room temperature in storage facilities on site. No samples will be transferred en masse from other research sites. Retention and destruction of samples

Participants are given the option of consenting to the use of their biological samples in this research study and an additional option of 'generic' consent for use in future studies, subject to ethical approval. This is in line with guidance from the Health Research Authority, which encompasses requirements of the ICH Good Clinical Practice, the European Clinical Trial Directive 2001/20/EC, the UK Medicines for Human Use (Clinical Trials) regulation 2004. Participants can request their samples be withdrawn from future study use and destroyed at any time.

If requested, samples will be destroyed in a manner appropriate for biological waste according to ICR guidelines and a record kept of this destruction.

16. Financing, Indemnity & Insurance

This study has received funding from the European Union within the ERC Advanced Grant 2013. This is funding dedicated to support innovative studies.

The standard NHS indemnity procedures will apply at each collaborating hospital. Each participating site is responsible for ensuring insurance and indemnity arrangements are in place to cover the liability of the Principal Investigator.

Liability rests with the study sponsor – the Institute of Cancer Research and a Research Agreement will be in place with each collaborating centre specifying the liability arrangements.

The study sponsor, the Institute of Cancer Research has no special compensation arrangements for this study. The NHS Litigation Authority covers standard clinical negligence of NHS employees, staff and health professionals under its Clinical Negligence Scheme for Trusts.

17. Publication Policy

The Chief Investigator together with the team at the data centre will write the final publication of the study results. A draft manuscript will be submitted to all co-authors (the study team, two named individuals from each collaborating centre and all members of the steering committee) for comments. After revision by all co-authors the manuscript will be sent to a major scientific journal.

The CI, the Study Coordinator and the Data Centre must approve all publications, abstracts and presentations based on subjects included in this study. This is applicable to any individual subject registered in the trial, or any subgroup of the trial subjects.

Appendix 2: BARCODE2 Trial Protocol

Protocol appendices not included

BARCODE 2: TARGETING CANCER CARE WITH THE USE OF GENETIC PROFILING

PROTOCOL

Version: 6.0

Dated: 18th February 2019

Chief Investigator:	Prof Ros Eeles
Sponsor:	The Institute of Cancer Research
Funder:	European Research Council
IRAS Number:	202638
EudraCT Number:	2016-000869-23
Main REC Reference Number:	16/LO/1686
CCR Number:	CCR4520
ISRCTN:	TBC
ClinicalTrials.gov Identifier:	NCT02955082
NIHR Reference Number:	32051

The BARCODE 2 trial is part of the National Institute for

Health Research Clinical Research Network Trial Portfolio

7.2 ADMINISTRATION

7.2.1.1 Clinical and Scientific Coordination Prof Ros Eeles (Chief Investigator)	Royal Marsden NHS Foundation Trust & The Institute of Cancer Research Downs Road, Sutton Surrey SM2 5PT
7.2.1.2 Trial Coordination	Oncogenetics Team The Institute of Cancer Research, Sir Richard Doll Building, 15 Cotswold Road, Sutton, Surrey SM2 5NG
BARCODE 2 Trial Manager	Dr Eva McGrowder
BARCODE 2 Clinical Fellow	Dr Sarah Benafif
Any questions relating to this protocol should be addressed in the first instance to the BARCODE 2 Trial Manager within the Oncogenetics Team, ICR Email: eva.mcgrowder@icr.ac.uk General enquiries: 0208 722 4483 Fax: 0208 722 4110	

7.3 PROTOCOL DEVELOPMENT GROUP

Prof Ros Eeles	Chief Investigator	The Institute of Cancer Research, Sutton, London
Dr Yae-Eun Suh	Emergency Principal Investigator	The Royal Marsden Hospital NHS Foundation Trust
Dr Elizabeth Bancroft	Sub-investigator	The Royal Marsden Hospital NHS Foundation Trust
Dr Eva McGrowder	Trial manager	The Institute of Cancer Research, Sutton, London
Mr John Walker	Patient Representative	
Elizabeth Page	Sub-investigator	The Institute of Cancer Research, Sutton, London
Dr Gerthard Attard	Sub-investigator	University College Hospitals London NHS Foundation Trust
Dr Alison Reid	Sub-investigator	The Royal Marsden Hospital NHS Foundation Trust
Dr Vincent Khoo	Sub-investigator	The Royal Marsden Hospital NHS Foundation Trust
Dr Nicholas Van As	Sub-investigator	The Royal Marsden Hospital NHS Foundation Trust
Prof Johann De Bono	Sub-investigator	The Institute of Cancer Research, Sutton, London
Dr Sue Chua	Trial radiologist	The Royal Marsden Hospital NHS Foundation Trust
Dr Aslam Sohaib	Trial radiologist	The Royal Marsden Hospital NHS Foundation Trust
Dr Sarah Benafif	Trial clinical research fellow	The Institute of Cancer Research, Sutton, London
Mr Matthew Hogben	Trial research nurse	The Royal Marsden Hospital NHS Foundation Trust
Ms Lucia D'Mello	Sub-investigator	The Royal Marsden Hospital NHS Foundation Trust
Ms Sibel Saya	Sub-investigator	The Institute of Cancer Research, Sutton, London
Dr Maggie Cheang	Translational analyst advisor	ICR-CTSU, The Institute of Cancer Research, London
Mr Ben Jenkins	Trial Statistician	ICR-CTSU, The Institute of Cancer Research, London
Dr Zsolt Szijgyarto	Trial statistician	ICR-CTSU, The Institute of Cancer Research, London
Dr Emma Hall	ICR-CTSU Scientific/Methodological lead	ICR-CTSU, The Institute of Cancer Research, London
Stephanie Burnett	ICR-CTSU senior trial manager	ICR-CTSU, The Institute of Cancer Research, London
Dr Zsofia Kote-Jarai	Sub-investigator	The Institute of Cancer Research, Sutton, London

The Trial Management Group (TMG) will be constituted from members of the Protocol Development Group. Other key study personnel will also be invited to join the TMG as appropriate to ensure representation from a range of sites and professional groups. A copy of the current membership of the TMG can be obtained from the BARCODE 2 Trial Manager at ICR.

Protocol Authorised by:

Name & Role	Signature	Date
Prof Ros Eeles (Chief Investigator)		

This protocol describes the BARCODE 2 trial and provides information about procedures for entering participants into this trial. The protocol should not be used as a guide for the treatment of participants outside of this trial.

Every care was taken in the preparation of this protocol, but corrections or amendments may be necessary. Protocol amendments will be circulated to participating referring centres as they occur. Centres are advised to contact ICR (**barcode2@icr.ac.uk**) to confirm they have the most recent version.

7.4 BARCODE 2 TRIAL SUMMARY

PROTOCOL TITLE

Targeting cancer care with the use of genetic profiling (BARCODE-2)

TARGET DISEASE

Metastatic castration resistant prostate cancer

STUDY OBJECTIVES

Primary

- To determine the response rate to two cycles of platinum chemotherapy in participants with metastatic castration resistant prostate cancer (mCRPC) and a germline mutation in a DNA repair gene.

Secondary

- To assess survival of participants with mCRPC and a DNA repair gene mutation after treatment with carboplatin.
- To determine the rate and prognostic significance of germline DNA repair gene mutations in participants with mCRPC.

STUDY DESIGN

Two staged open-label, single centre, single arm phase II study

TRIAL POPULATION

Participants with metastatic castration resistant prostate cancer who have disease progression after treatment with docetaxel chemotherapy and enzalutamide or abiraterone.

RECRUITMENT TARGET	Approximately 450 participants (anticipated 360-540 depending on the observed prevalence of germline mutations) to undergo genetic profiling for germline mutation of DNA repair genes (part 1 of study) or may directly enter the interventional phase II trial (part 2 of the study) if they are already known to have a germline mutation in a DNA repair gene (e.g. *BRCA1* or *BRCA2*, estimated to be about 10 participants). In part 2, 36 to 54 (12 to 18 in each of 3 initial gene groups) participants with a germline mutation will be treated with carboplatin.
TREATMENT REGIMEN	Participants found/known to have a germline mutation in a DNA repair gene will be treated with 3 weekly infusions of carboplatin (AUC5) chemotherapy for a maximum of 10 cycles of treatment.
PRIMARY ENDPOINT	Response rate to two cycles of platinum chemotherapy in participants with mCRPC and germline DNA repair gene mutations based on CT imaging using modified RECIST 1.1 criteria, and/or fall in PSA of >50%. .
SECONDARY ENDPOINTS	- The incidence of germline mutations in DNA repair genes in a population of mCRPC cases. - Overall survival and progression free survival of participants with mCRPC and a DNA repair gene mutation treated with carboplatin. - Cause specific survival from date of first diagnosis of prostate cancer in participants with germline DNA repair gene mutations - Radiographic PFS

- Time to radiographic progression
- Time to PSA progression
- Duration and pattern of PSA response

FOLLOW UP   Following treatment with carboplatin, participants will be reviewed 30 days after the last cycle for a safety review. Thereafter, they will be reviewed at least 3 monthly for survival until death. For participants who do not progress on study, data related to first progression will be recorded. These data will be collected on all participants entering part 2 of the study.
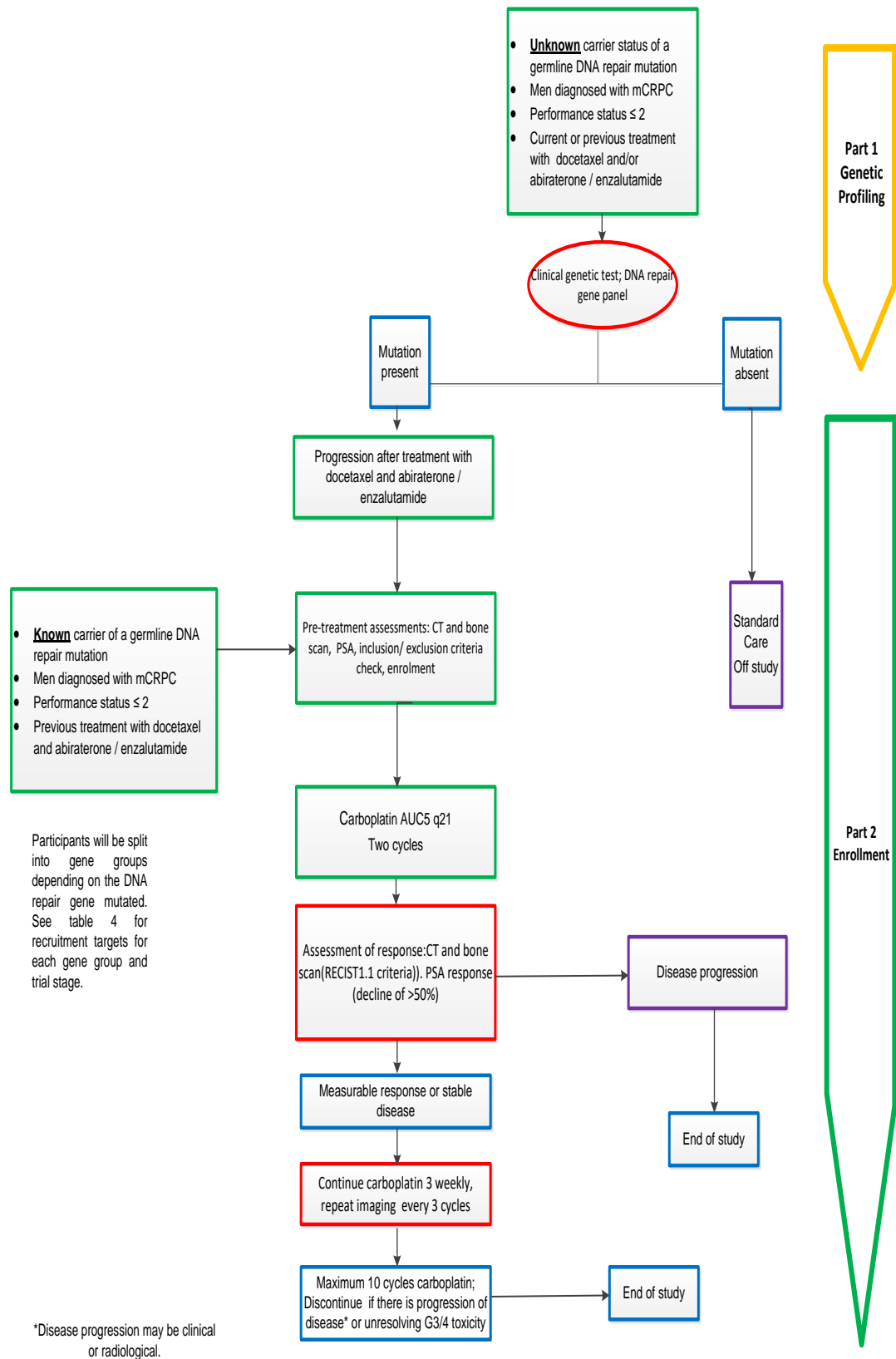
7.5 TRIAL SCHEMA

Figure 5: Diagram to show eligible participant population, treatment interventions and follow up

1. INTRODUCTION

Prostate cancer (PrCa) is now the commonest cancer in men in the Western world, with over 40,000 new cases per annum and a lifetime risk of 1 in 11 in the United Kingdom (UK) [143], as well as in the European Union with 397,000 new cases per annum, and 94 000 deaths [144]. However, its aetiology remains poorly understood. The substantial worldwide variation in incidence rates suggests that lifestyle risk factors are important. To date, however, no definite lifestyle risk factors have been identified [144].

1.1. Prostate cancer and germline DNA repair mutations

There are coding variants in genes in the DNA repair pathway which are associated with a higher level of PrCa risk. These have been well characterised in genes such as *BRCA2*, and to a lesser extent, *BRCA1* [38, 145]. These studies from the Oncogenetics Team at the ICR have analysed the germline of a series of PrCa cases in the UK and have shown that 1.2% and 0.45% of men diagnosed before the age of 65 years harbour deleterious mutations in *BRCA2* and *BRCA1* respectively.

More recently, germline mutations in other genes have been shown to confer an increased risk of PrCa. PrCa has been reported in families with Lynch syndrome [146, 147], a well-known familial cancer predisposition syndrome associated with a spectrum of cancers, including colorectal, endometrial and others [148, 149]. Lynch syndrome is caused by a germline mutation in one of the mismatch repair genes, *MLH1*, *MSH2*, *MSH6* or *PMS2*. The risk of PrCa in Lynch syndrome has been reported to be up to 30% by the age of 70 years and the mean age of diagnosis (60.4 years) was lower than that seen in the general population (66.6 years) [150]. The number of men with a Gleason score between 8 and 10 was significantly higher than expected (p<0.00001) [150].

Other DNA repair genes, known to be associated with other cancers, such as breast cancer, have also been implicated in prostate cancer risk [151]. Studies examining the genes *PALB2*, *CHEK2*, *BRIP1* and *NBS1* have produced conflicting evidence regarding the role of a germline mutation in PrCa [152-159]. The team demonstrated germline mutations in the DNA repair gene *BRIP1* in familial and young onset PrCa cases at a rate of 0.14% compared to 0.05% in controls [157]. Polish founder mutations in the *CHEK2* gene were overrepresented in PrCa when compared to unaffected controls (OR 3.4, p=0.004) [153] and also in familial PrCa cases when compared to unselected cases (OR 9.0, p=0.0002; OR 3.7, p=0.03) [153, 154]. However, these results were not able to be confirmed when looking at an international population (p=0.26) [158]. Similarly, a Polish study found that mutation in *NBS1* contributed to PrCa risk in a familial PrCa population with an odds ratio of 16, p<0.0001 (compared to non-familial PrCa with *NBS1* mutation OR 3.9, p=0.01) [152] but could not be replicated in an American population [156]. *PALB2* was also suggested to play a role in PrCa risk [155] but again was unable to be replicated, albeit in a smaller sample. [159]

The Oncogenetics Team have recently analysed a 22 DNA repair gene panel [160] in families with several cases of PrCa and/or other well-known familial cancers and have shown that 7% of men in this cohort harboured a mutation in one of these 22 genes [161]. There was also an association between positive mutation status and the presence of nodal involvement (42.9% in carriers vs. 1.3% in non-carriers; P=0.0014) and metastasis (30% vs. 1.3%; P=0.0014).

Incorporating previous, as well as our ongoing research showing that many of these genes that have roles in DNA repair impact upon lifetime cancer risk and the subsequent development of PrCa, we have developed a DNA repair gene panel. The impact of many of these genes upon PrCa development is still being investigated but given their

interaction with known PrCa predisposition genes, it is possible that they also play a role in this cancer.

1.2. Germline DNA repair mutations and PrCa diagnosis and outcome

A number of previous studies have suggested the prognostic significance of a germline *BRCA2* mutation in a man with PrCa. Two limited series, examining small numbers of *BRCA2* mutation carriers (one group of carriers of the Icelandic *BRCA2* 999del5 founder mutation, and the other carriers of the Ashkenazi Jewish *BRCA2* 6174delT founder mutation) found mutation status is related to worse cancer specific survival (CSS) [162, 163]. The Oncogenetics Team have shown that PrCa in men with germline *BRCA* mutations have a poorer prognosis, particularly for *BRCA2* [164, 165].

The results analysing the outcome of 2019 participants with PrCa; 18 with *BRCA1* mutations and 61 with *BRCA2* mutations have recently been published [166]. The control group was taken from a set of men with and without a family history but who were known not to carry a mutation in either gene. This is the largest study to date investigating the clinical characteristics and outcome of PrCa participants with and without *BRCA* mutations. For the *BRCA* carrier group, the disease was more aggressive and more often associated with nodal involvement or distant metastases. Germline *BRCA1* and *BRCA2* mutations were associated with a Gleason score of 8 or more and T3/T4 disease. The CSS was superior for non-carriers (15.7 years compared to 8.6 years for *BRCA* carriers). Subgroup analyses confirmed the deleterious effect of *BRCA2* mutations, but failed to confirm the role of *BRCA1* due to the smaller number of carriers in the study [166].

The study described above, examining a 22 DNA repair gene panel in men with PrCa and a family history of prostate and other cancers, showed a correlation of positive mutation status with advanced PrCa disease with an odds ratio of 15.09 (p = 0.00164)

[161]. This association remained after excluding the carriers of a *BRCA2* mutation (p = 0.00285).

Most of these studies examined comparison groups without a germline mutation in one of these DNA repair genes but with a family history of PrCa. The control groups in two studies were not selected from cohorts of familial cases and the statistical significance of poorer outcome in the carrier group remained [162, 163]. This suggests that the more advanced disease, and therefore poorer outcome, is related to the presence of the mutation, rather than a family history.

1.3. Platinum chemotherapy and PrCa

Docetaxel chemotherapy is the current standard of care for mCRPC. Platinum chemotherapy drugs, such as cisplatin and carboplatin, have moderate single-agent activity in mCRPC. Next-generation platinum drugs, including satraplatin and oxaliplatin, may have additional activity in the management of mCRPC. There is a rationale for using platinum based chemotherapy in tumours with a neuroendocrine differentiation [167]. The initial prostate biopsy does not usually diagnose 'small cell' PrCa, but autopsy series have identified this histological type in 10% to 20% of specimens [168, 169]. The clinical presentation of this subtype consists of visceral involvement and lytic bone metastases with a low PSA [170]. These participants have castration resistant disease, but have been shown to respond better to chemotherapy including platinum agents [171].

In a 1993 review that summarized the results of 209 participants receiving cisplatin monotherapy the partial response rate was 12% [172]. Several phase 2 studies using weekly carboplatin showed moderate activity with response rates in the region of 20% [173, 174]. Following on from the early trials, the next step was to combine platinum with other cytotoxic agents of proven efficacy in mCRPC. Several studies showed high

response rates when combining platinum with taxanes. A phase 2 study assessed the activity of cisplatin given along with docetaxel as first line treatment in mCRPC participants with an 'anaplastic' phenotype. In 120 participants with clinical features suggestive of 'small cell prostate cancer', defined by set 'anaplastic' clinical criteria, the median overall survival was 16 months, with 65% having responses or stable disease after 4 cycles. Second line chemotherapy with cisplatin and Etoposide was given to 74 of these participants with an overall response rate of 33.8% [175].

In a phase 2 study of carboplatin AUC4 and Etoposide in participants with anaplastic mCRPC, 60 participants were recruited. More than 50% presented with visceral metastases. The response rate was 8% with significant side-effects, including 7% neutropenic sepsis and 1 treatment related death. Based on these results one can gather that this is a toxic treatment with no significant benefit. Neuroendocrine markers were used and the levels of Chromogranin-A were more predictive of poor outcome than neuron specific enolase [176].

Satraplatin and oxaliplatin are two drugs that are based on novel structures, with altered, stable ligands and demonstrate activity in cisplatin-resistant cancers. Oxaliplatin was assessed in a phase 2 study of 54 participants with mCRPC most of which progressed on other chemotherapy treatments. The PSA response rates were low at 11% and 19% when combined with 5FU [177].

Satraplatin is an oral platinum compound which demonstrated efficacy and tolerability in PrCa. A phase 3 trial of satraplatin plus prednisolone versus placebo plus prednisolone in mCRPC participants was stopped prematurely, after recruiting 50 participants due to withdrawal of the sponsor. The combination of satraplatin and prednisolone resulted in a statistically significant increase in PSA response compared to the prednisolone-alone arm (33% vs 9%; p=0.046). The median progression-free

survival was also significantly better for the satraplatin/prednisolone arm (5.2 vs 2.5 months; p=0.023), with a hazard ratio of 0.50 [178, 179]. A similar phase 3 study was successfully completed in participants with mCRPC and progression after docetaxel chemotherapy. It recruited 950 participants who were randomized to the same arms as the previous study. No difference in overall survival was seen, however, there was a 33% reduction in the risk of progression on satraplatin, and there was good palliation with improvement of cancer related pain [180].

1.4. Treatment of cancer in *BRCA* carriers

It has been reported that ovarian cancer cases with germline mutations in the DNA repair genes *BRCA1* and *BRCA2* are more sensitive to platinum agents than non-*BRCA* mutated ovarian cancers [181]. This has also been explored in women with breast cancer [182, 183]. Women with a germline *BRCA1* mutation showed a poorer response to docetaxel (a mitotic spindle inhibitor) and doxorubicin than non-carriers but showed a similar response to DNA-damaging chemotherapies such as platinum [182]. The same authors showed a high rate of pathologic complete response to cisplatin in women with a *BRCA1* mutation but this was not significant in this cohort of 102 carriers [183].

The use of poly ADP-ribose polymerase (PARP) inhibitors in cancer participants with a germline *BRCA* mutation has been investigated extensively. These trials are now in Phase III for treatment of breast and ovarian cancer in these participants and roles in adjuvant as well as metastatic disease are being studied. [184-186] In fact, Olaparib, a first in class PARPi, is now approved by the EMA and FDA for use in the maintenance treatment of platinum sensitive *BRCA*-mutated (germline and/or somatic) high grade serous epithelial ovarian, fallopian tube, or primary peritoneal cancer who have a complete or partial response to platinum-based chemotherapy. This approval was gained as a result of the phase II Study 19 which showed a significantly longer PFS in participants with BRCA associated ovarian cancer treated with olaparib compared to

those treated with placebo (11.2 months vs 4.3 months). Participants without a *BRCA* mutation treated with olaparib also had a lower PFS compared to treated *BRCA* participants of 7.4 months.

The role of PARPi in other *BRCA* associated cancers is also being investigated [187]. The ongoing phase II TOPARP study of olaparib in metastatic castration resistant prostate cancer recently reported the result of 50 participants enrolled on the study. Investigators found that 33% (n=49) of participants responded to olaparib. Study of the tumour genetic profile in 49 participants identified homozygous deletions, deleterious mutations, or both in DNA repair genes in 16 participants (33%); of these, 3 participants had a germline *BRCA2* mutation and 3 had a germline *ATM* mutation. 88% of participants with homozygous deletions, deleterious mutations, or both in DNA repair genes responded to olaparib. These results are encouraging and suggest that the PARPi as well as platinum sensitivity observed in *BRCA* associated ovarian cancer also extends to *BRCA* associated prostate cancer.

1.5. BRCAness

'BRCAness' is defined as the phenotype that some sporadic tumours share with familial-*BRCA* cancers. It has been proposed that this could be attributed to an inactivation of the *BRCA*/Fanconi Anaemia pathway which is involved in homologous recombination repair of DNA defects [188]. Identifying tumours that show these '*BRCA*-like' characteristics could potentially influence the clinical management of these tumours, as it might allow rational design of mechanism-based chemotherapy regimens, targeted towards the DNA-repair defects in the tumour [188].

A recent review of the role of DNA repair systems in breast and prostate cancer explored the interaction of hormones and function of these genes and their contribution to the development of these types of cancers [189]. The authors outline the association

between DNA repair systems and the loss of hormone receptors, and discuss the potential for this relationship to be exploited in therapy choice. It follows that drugs which target the deficiency in DNA repair mechanisms in cancer participants with these germline mutations could be effective.

1.6. Next generation genetic sequencing

Genetic sequencing has become cheaper and more efficient in recent years with the advent of next generation sequencing. This means that large amounts of genetic data can be generated in a short period of time and it is becoming more feasible that after a diagnosis of cancer, a rapid genetic test could take place, the results of which would then inform tailored treatment aimed at the genetic cause of cancer. It will therefore influence the work being undertaken in both oncology and genetics to integrate genetic profiling and gene panel analysis into cancer care, so that men with prostate cancer can be offered more intensive molecular stratification and treatment within clinical trials.

The team's previous research shows that defects in DNA repair genes not only are associated with higher rates of prostate cancer development, but are associated with higher rates of aggressive disease. It is likely that it is not simply a family history of prostate cancer that is linked with this phenomenon, but the presence of a germline DNA repair gene mutation. It is therefore increasingly important to identify men who fall into this category and to target the most effective treatment to them.

The use of platinum chemotherapy as well as PARP inhibitors in *BRCA* carriers has shown utility in breast and ovarian cancer. This trial aims to utilise these findings in a proof-of-principle study to examine prostate cancer outcomes of men with a germline mutation in a DNA repair gene who undergo two cycles of platinum chemotherapy for metastatic castration resistant disease.

1.7. BARCODE-2 Gene Panel

The screening of participants for treatment in the trial will be carried out by sequencing germline DNA for 115 genes in order to identify pathogenic mutations. This will be carried out using a gene panel based on targeted exon capture which was developed by the ICR Oncogenetics team, and manufactured by Agilent Technologies (http://www.genomics.agilent.com/article.jsp?pageId=3075) for the sole use of the ICR Oncogenetics team in the context of the BARCODE-2 trial. When a pathogenic mutation is detected in the DNA of a participant, this result will be validated by Sanger sequencing which is the gold standard method for validation of next generation sequencing results [83, 190]. Sanger sequencing will be carried out in the ICR using a 3730XL sequencer which is maintained and operated at United Kingdom Accreditation Service (UKAS) ISO 15189 standards.

1.8. Study Rationale

This study will utilise knowledge of the germline status of DNA repair genes to alter the treatment pathway for mCRPC participants. Participants can be identified for enrolment to the study by either being a known carrier of a mutation in a DNA repair gene having had a clinical genetic test prior to enrolment or undergoing genetic profiling within the study. Eligible men (i.e. with castration resistant disease and confirmed metastases on imaging) will be identified via participation in other observational genetic studies run through the Oncogenetics Team at the ICR or through uro-oncology clinics at the Royal Marsden Hospital. Acting as participant identification centres (PICs) will be regional genetic centres in the United Kingdom and uro-oncology clinics at other London hospitals (e.g. St Georges, Charing Cross and Guys Hospital).

If men have not previously had any genetic testing in a genetics clinic or in a research study, they can provide a DNA sample for sequencing of a panel of DNA repair genes within the study (part 1 – genetic profiling). If a pathogenic mutation is confirmed in one

of these genes, participants may enter the interventional phase II trial (part 2 of the study) and undergo a baseline imaging prior to receiving carboplatin.

Extrapolating from other tumour types such as breast and ovarian cancer there is a potential benefit for introducing platinum-based chemotherapy in participants with DNA repair gene mutations. The dose of carboplatin administered in the study will be calculated based on the participant's renal function and target Area Under the Curve 5 (AUC5) as recommended in the Summary of Product Characteristics (SmPC; https://www.medicines.org.uk/emc/medicine/622). The use of this dose for single agent carboplatin in this study is also based on the London Cancer Alliance guidelines for single agent carboplatin in ovarian cancer, as this is given in a similar setting of relapsed/ advanced disease [83]. Trial participants will receive two cycles of carboplatin AUC5 and will be assessed radiologically via modified RECIST 1.1 criteria for a measurable response.

Participants who show a complete or partial response or stable disease will continue with carboplatin treatment every 3 weeks for a maximum of 10 cycles in total or until disease progression or unresolving grade 3-4 toxicity. Radiological assessment will be repeated every 3 cycles. For participants who show progression of disease after two cycles of carboplatin, treatment will end and participants will come off study.

Response will also be measured with PSA levels and all participants entering part 2 of the study will be followed until death to measure cause-specific and overall survival. This is a hypothesis generating study, to test the platinum sensitivity of prostate tumours that have developed due to a germline mutation in a DNA repair gene. This study will provide data to use in a larger clinical trial of platinum chemotherapy based on participants' germline genetic signature and/or tumour genetic profile.

2. TRIAL OBJECTIVES

2.1. Primary Objective

- To determine the response rate to two cycles of platinum chemotherapy in participants with mCRPC and a germline mutation in a DNA repair gene

2.2. Secondary Objectives

- To assess progression-free and overall survival of participants with mCRPC and a DNA repair gene mutation after treatment with carboplatin.
- To determine the rate of germline DNA repair gene mutations in participants with mCRPC.

3. TRIAL DESIGN

BARCODE 2 is a single arm phase II study which will be run at a single site (Royal Marsden Hospital (RMH)). The study will be divided into two parts. In part 1 of the study, enrolled participants' DNA repair gene mutation carrier status will be assessed using a gene panel. Men who are found to carry a pathogenic mutation or are already known to carry a germline mutation can enrol in part 2 of the study and be offered treatment with carboplatin.

Men will be identified for the study in two ways:

1. Men with mCRPC who are a known carrier of a germline mutation in a DNA repair gene will be identified via genetics services locally (at RMH) and elsewhere in the United Kingdom and referred for the study to the Royal Marsden Hospital. They will be assessed for eligibility for part 2 of the study. They must have had previous treatment with docetaxel and either abiraterone or enzalutamide.

2. Men with mCRPC attending the uro-oncology clinic at the Royal Marsden Hospital, or at a PIC site, will be invited to participate and be assessed for eligibility for part 1 of the study. Eligible men will provide DNA samples for profiling using a panel of DNA repair genes. Men assessed for part 1 of the study may have already received or concurrently be receiving docetaxel or abiraterone/ enzalutamide. If currently undergoing one of these standard treatment lines, men who are found to have a genetic mutation will be considered for entry to part 2 upon disease progression having received both docetaxel and one of abiraterone or enzalutamide.

Men entering part 2 of the study must have metastatic disease confirmed on recent imaging (carried out in the previous 3 months), prior to carboplatin treatment. Response rates to carboplatin will be analysed by gene groups. Participants will be divided initially into 3 gene groups as shown in Table 4 (page 33). The genes that will initially be included in the genetic screen can be seen in Appendix A (page 330). This gene list and gene groups have been determined based on previous and emerging evidence of

involvement in PrCa risk and development and knowledge of DNA repair gene pathways. This list and gene groups will be reviewed throughout the study based on the mutations found in part 1 as well as evolving information on DNA repair pathways; this may lead to the list being refined and further gene groups being added or gene groups being removed (if observed prevalence suggests recruitment will be unfeasible) during the study. Within each gene group 12 participants will be treated and, if successful (as defined in section 11), a further 6 participants will be treated bringing the total number of participants required to 18 per gene group.

Eligible participants will undergo a baseline contrast enhanced CT and bone scan; followed by 2 cycles of carboplatin (AUC 5) administered intravenously 21 days apart. Participants will undergo a clinical review as well as routine full blood count and biochemistry prior to each cycle of carboplatin. After cycle 2 of carboplatin, PSA measurement and CT scan to assess response will be performed. Participants will be classified with progressive disease (PD), stable disease (SD), partial (PR) or complete response (CR) according to modified RECIST 1.1 criteria. Following two cycles of carboplatin, participants may continue treatment if radiological imaging has shown CR, PR or SD. The first post treatment bone scan will be carried out after cycle 3 (week 8-9). Thereafter, CT and bone scan will be done together after cycle 5 and then after every 3 cycles. Treatment will continue for a maximum of 10 cycles of carboplatin. Treatment will be discontinued and participants will come off treatment if there is disease progression (clinical or radiological), or if treatment is no longer tolerated by the participant due to toxicity. Following the final cycle of carboplatin, participants will be reviewed 30 days after the last dose for a safety review. Thereafter, survival data will be collected on a 3 monthly basis until death.

4. STUDY ENDPOINTS

4.1. Primary Endpoint

- Response rate to two cycles of platinum chemotherapy in participants with mCRPC and germline DNA repair gene mutations using modified RECIST 1.1 criteria, and/or fall in PSA of >50%.

4.2. Secondary Endpoints

- Overall survival and progression free survival of participants with mCRPC and a DNA repair gene mutation treated with carboplatin.
- The incidence of germline DNA repair gene mutations in a population of mCRPC cases
- Cause specific survival from date of first diagnosis of prostate cancer in participants with DNA repair gene mutations
- Radiographic progression free survival
- Time to radiographic progression
- Time to PSA progression
- Duration and pattern of PSA response

5. PARTICIPANT SELECTION & ELIGIBILITY

Participants with mCRPC which has progressed after docetaxel and abiraterone or enzalutamide may be assessed for eligibility for study entry.

5.1. Number of Participants

Approximately 450 participants will be enrolled into part 1 of the study for genetic profiling. These participants will undergo testing for a germline DNA repair gene mutation. The aim is to recruit 5 participants per week for genetic profiling to achieve a target of 450 men undergoing genetic profiling over 2 years. Participants who are already known to harbour a germline mutation may directly enter part 2 of the study (this is anticipated to be about 10 participants overall). The number of participants needed to screen is anticipated to be between 360 and 540 and will depend on the observed mutation prevalence rates and the number of participants enrolled directly to part 2 with a known mutation. Participants in part 2 will be divided into 3 initial gene groups for treatment (12-18 participants will be treated in each group).

5.2. Source of Participants

Potential participants will be identified primarily through uro-oncology clinics and multidisciplinary team (MDT) meetings at the Royal Marsden Hospital (RMH). Participants with a known germline mutation in a DNA repair gene can be referred from the genetics and uro-oncology services at the Royal Marsden Hospital. Other potential participants for either part 1 or part 2 can be referred from other genetics centres and uro-oncology clinics in the UK that will act as participation identification centres (PICs). No potential participant from outside the Royal Marsden Hospital will be approached without consultation with their oncology team. Potential participants will only have their records reviewed by or be approached by staff who have consented to access their medical/research records (i.e. staff who are members of their clinical team or research staff of a study to which the participant has already consented). Template letters are

included in the supporting documents to aid with this identification. The methods of identification are below and outlined in

Figure 6:

1. Men with mCRPC attending the uro-oncology or genetics clinic at RMH or a PIC site, will be invited to participate and be assessed for eligibility for part 1 of the study. These men will provide DNA samples for profiling using a panel of DNA repair genes as part of the study. Those with a confirmed pathogenic mutation in one of these genes will be given the option to proceed to part 2 and will be assessed for eligibility for treatment on study.

2. Men with mCRPC who are a known carrier of a germline mutation in a DNA repair gene will be identified for assessment of eligibility for part 2 of the study via three methods below. In all three cases, the participant will only be contacted via their treating uro-oncology team, as this team will have the most up to date information about the participant's treatment pathway and current progress.

   a. Participation in other observational genetic studies that are run through the Oncogenetics Team at the ICR (with the same CI as the current study). The participant will be referred to the BARCODE 2 study through the uro-oncology service where they are being treated, which may be a PIC:

17. The observational genetic studies include the UK Genetic Prostate Cancer Study (UKGPCS; CCR848; 06/MRE02/4) that enrols men with young onset or familial prostate cancer. The CI of both UKGPCS and this current study is Prof Ros Eeles. Within UKGPCS, genetic testing is done on men with prostate cancer enrolled in the study. If a mutation is found in a gene that is included in BARCODE 2, the participant's uro-oncology team will be contacted by the UKGPCS team. This will be the clinician who initially referred the participant to UKGPCS*. A standard letter will be sent (please see templates submitted for approval) outlining that the participant may also be eligible for BARCODE 2 and if they deem it appropriate, the clinician is invited to discuss the possibility of this study with the participant. The participant can then be referred to the BARCODE 2 study team to discuss further.

18. *Please note that in the case of a mutation in a gene that has clinical implications, the participant's GP will have been contacted with this information

regardless of this current study. This process is ethically approved through the UKGPCS protocol (CCR848; 06/MRE02/4), as are the relevant template letters also submitted here for approval.

    b.   Genetics services at the Royal Marsden Hospital or other UK hospitals acting as PICs:

19. Genetics services which are PIC sites can review their database and participant records for men with prostate cancer and a germline genetic mutation in an included gene. The genetics service can write to the participant's uro-oncology team to alert them of the study and invite them to discuss the possibility of this study with the participant (please see template submitted for approval). The participant can then be referred to the BARCODE 2 study team to discuss further.

    c.   Uro-oncology clinics at the Royal Marsden Hospital or other London hospitals acting as participant identification centres (PICs):

20. As part of a participant's cancer treatment, their uro-oncology team may be aware that the participant is also a carrier of a germline genetic mutation in an included gene. In this case the uro-oncology service which is a PIC site can discuss the possibility of this study with the participant. If interested, the participant can then be referred to the BARCODE 2 study team to discuss further.
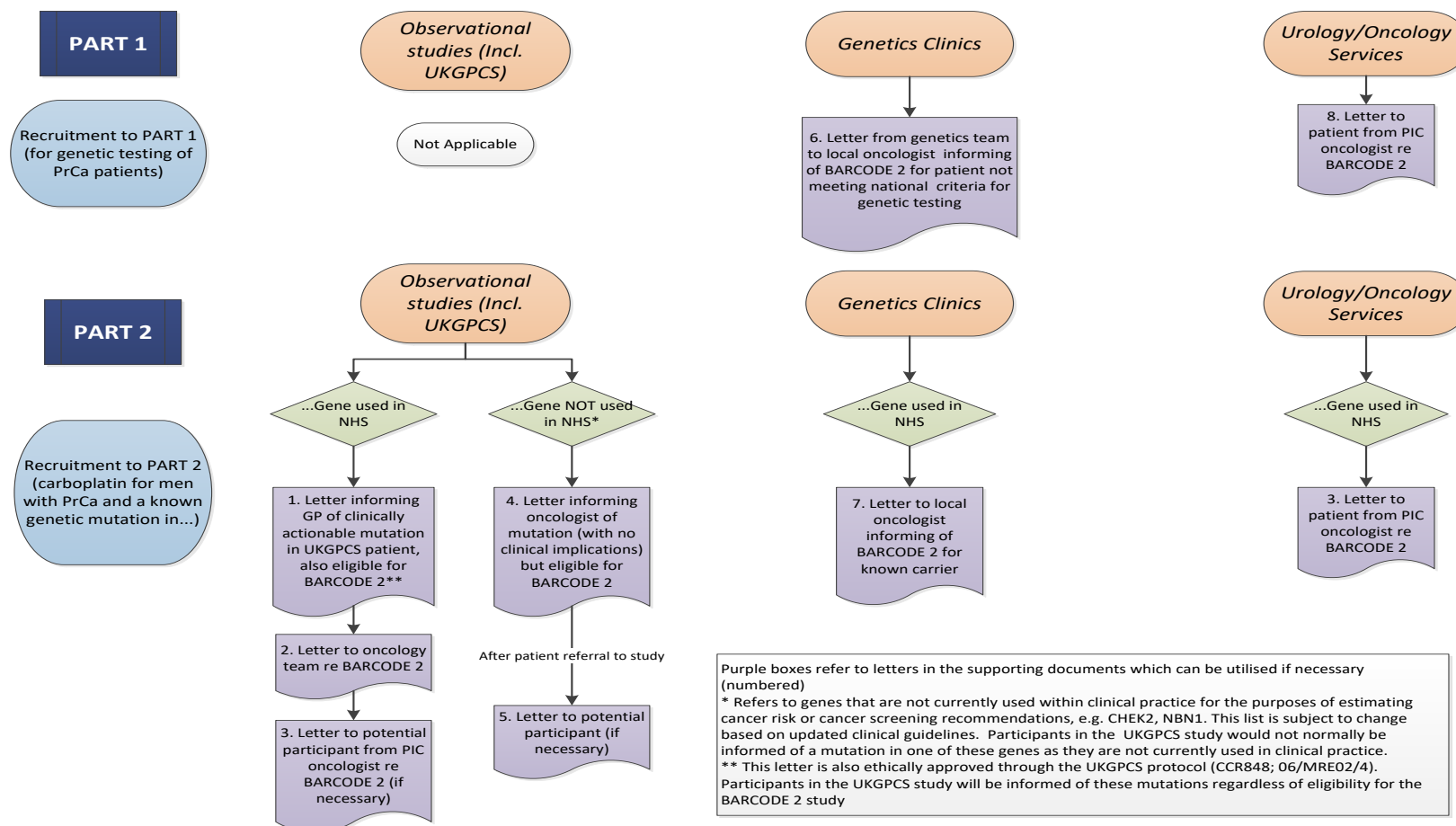
Figure 6: Potential Participant Identification in main site (Royal Marsden) or participant identification centres (PICs)

5.3. Inclusion Criteria

All study participants will be assessed according to the part 1 and/ or part 2 inclusion criteria depending on which part of the study they enter initially.

7.5.1.1 For Part 1 (genetic screening) of the study:

1. Age ≥ 18 years.
2. Recorded diagnosis of prostate cancer with or without histological confirmation. Patients who have not previously undergone a prostate (or metastatic) biopsy but are confirmed to have a raised PSA (>80ng/ml at any time), metastatic disease on imaging and have undergone treatment for mCRPC are eligible.
3. Castration-resistant disease defined as biochemical or radiological progression on/after treatment with orchidectomy or LHRH analogues as per PCWG3 criteria.
4. Confirmed metastatic disease on conventional imaging methods such as CT, bone scan or PET imaging.
5. Current or previous treatment includes at least one of the following:
   a. Docetaxel (either in hormone sensitive or resistant setting; Patients who have completed treatment with or are currently undergoing Cabazitaxel chemotherapy are also eligible)
   b. Enzalutamide
   c. Abiraterone
6. Adequate renal function measured by calculated GFR (Cockcroft-Gault) >30ml/min. If a participant had renal dysfunction that is expected to improve, they may be considered for part 1 of the study.
7. Adequate haematological function to allow study entry in line with local hospital practice or at the investigator's discretion.
8. WHO performance status 0-2 as assessed and documented by study doctor.
9. Life expectancy >12 weeks
10. Participants with stable, treated brain metastases will be eligible providing informed consent can be given and that other sites of measurable disease are present
11. The subject is capable of understanding and complying with the protocol requirements and has signed the BARCODE 2 informed consent form.

7.5.1.2 In addition to the above, for Part 2 of the study:

1. Confirmed pathogenic germline mutation in a DNA repair gene. (Participants with a known germline mutation will need to provide a report from the external laboratory where genetic testing was carried out)

2. Previous treatment with docetaxel and abiraterone or enzalutamide with documented disease progression prior to entry to part 2 (rising PSA and/or radiographic progression). Patients previously treated with cabazitaxel and who have documented disease progression are also eligible.

3. Adequate haematological function: Haemoglobin (Hb) ≥8.0g/dL, neutrophil count ≥1.5x10^9/L and platelets ≥100x10^9/L.

4. Adequate liver function: Total bilirubin ≤1.5 x upper limit of normal (ULN) except for participants with known Gilbert's syndrome; AST and ALT ≤ 2.5x ULN in the presence of liver metastases.

5. Adequate renal function: creatinine clearance >30ml/min measured by a glomerular filtration rate (GFR) clearance test. If a measured GFR test is not available, then calculated GFR is acceptable (measured GFR must be carried out by cycle 2 of carboplatin).

5.4. Exclusion Criteria (for part 1 and 2)

1. Critical organ metastases (e.g. spinal metastases with risk of cord compression) as documented on most recent imaging report.

2. Participants with bleeding tumours.

3. Previous treatment with a platinum chemotherapy drug for prostate cancer.

4. Previous treatment with a PARP inhibitor

5. Participants with a history of severe allergic reaction to carboplatin or other platinum-containing compounds

6. Exposure to yellow fever vaccine in the previous 6 months.

7. Participants unfit for chemotherapy or those with ongoing neuropathy >grade 1 (sensory or motor) according to NCI CTCAE V4.02.

8. Known and documented hearing impairment

9. Other active malignancies or previous malignancies likely, in the PI's opinion, to impact on management of mCRPC.

10. Significant documented cardiovascular disease: severe/unstable angina, myocardial infarction less than 6 months prior to trial entry, arterial thrombotic events less than 6 months prior to trial entry, clinically significant cardiac failure requiring treatment (NYHA II-IV).

11. Cerebrovascular disease (CVA or TIA) in the preceding 2 years to entry to Part 2 of study.

12. Presence of symptomatic brain metastases.

5.5. Life Style Guidelines

Participants must be surgically sterile or must agree to use effective contraception during the period of therapy and for 6 months after the last dose of carboplatin.

Effective contraception is defined as double barrier contraception (e.g. condom plus spermicide in combination with a diaphragm, cervical cap or intrauterine device).

6. SCREENING

6.1. Procedure for Obtaining Informed Consent

The Principal Investigator (or designated individual) must ensure that each trial participant is fully informed about the nature and objectives of the trial and possible risks associated with participation. Participants will be given the current ethically approved BARCODE 2 participant information sheet for their consideration. Participants will only be asked to consent to the study after they have had sufficient time to consider the trial, and the opportunity to ask any further questions. Participants who have not had previous genetic testing will sign the part 1 consent form to undergo genetic profiling for a germline mutation in a DNA repair gene. Participants who are found to have a pathogenic mutation in part 1 or who are already known to carry a germline mutation must sign the part 2 consent form prior to undergoing part 2 study related procedures.

No protocol required assessments should be conducted until the BARCODE 2 consent form has been signed and dated by both the participant and the Investigator, unless they are performed routinely as part of standard participant care.

Participants who consent to BARCODE 2 consent to the use of the samples collected for future translational studies.

Confirmation of the participant's consent and the informed consent process must be documented in the participant's medical notes. A copy of the signed consent forms (part 1 and part 2, if applicable) should be provided to the participant and the original retained in the investigator site file, which must be available for verification by study staff or for regulatory inspection at any time.

6.2. Participation in other Clinical Trials

BARCODE 2 participants will not be permitted to participate in any other trials of investigational medicinal products whilst they are being treated within BARCODE 2.

7. REGISTRATION AND TRIAL ENTRY

Participants must be registered centrally with the trials unit (ICR-CTSU) before protocol required screening assessments commence (for both participants entering Part 1 and those entering Part 2 of study directly).

Once informed consent has been obtained, the eligibility checklist, a registration form and trial entry form (if applicable) must be completed prior to contacting ICR-CTSU for registration/trial entry. Participants must be registered by contacting the ICR CTSU on:

<div align="center">

To register a participant, telephone ICR-CTSU

On: 0208 643 7150

</div>

The following information will be required when registering the participant:

- Name of consultant, centre and person registering participant
- Confirmation that participant has given written informed consent
- Participant's full name, agreed initials, hospital number, date of birth, postcode and NHS number.
- Confirmation that participant is eligible for the trial by completion of the eligibility checklist
- Confirmation of which Part of study participant is entering when registered

The participant will be allocated a unique Registration number.

For trial entry (Part 2), the following information will be required when registering the participant for trial entry:

- Name of consultant, centre and person registering participant
- Participant's registration number, date of birth and agreed initials
- Confirmation that participant has given written informed consent to Part 2
- Confirmation of a Part 2 eligible gene mutation
- Confirmation that participant is eligible for the trial by completion of both Part 1 and Part 2 eligibility checklists
- Confirmation of which Part of study participant is entering when registered
- If participant was previously registered to Part 1 of the study for genetic profiling, the confirmation of the date the Part 1 blood sample was taken will also be required.

The participant will be allocated a unique trial identification number (Trial ID) if/ when entering Part 2.

Once the genetic profiling result for participants entering Part 1 of the study is known, ICR-CTSU will be informed of those who are found to have a genetic mutation in a DNA repair gene and will be entering part 2 of the study by completing a screening log form.

8. TRIAL ASSESSMENTS

7.5.1.3 Part 1:

Participants who have not previously had a genetic test for germline genetic mutations will have the following assessments completed and will undergo the DNA repair gene panel profiling before proceeding to the assessments under part 2 if eligible:

- Informed consent (using Part 1 PIS and consent form)
- Eligibility: inclusion/exclusion check
- Confirmation of metastatic disease on recent imaging report
- Medical history
- Documented history of treatment with or currently treated with docetaxel and/or abiraterone/enzalutamide
- DNA repair gene panel profiling* (will take 4-12 weeks for results to be available). If a blood sample is unable to be processed (e.g. due to mislabelling or due to technical laboratory issues), a second blood sample may be requested.

*DNA collected during this study will be stored indefinitely if consented to by participants signing the study consent form.

7.5.1.4 Part 2:

Participants with a germline DNA repair gene mutation identified in part 1, or who are already known to have a germline mutation will undergo the following assessments.

8.1. Pre-treatment procedures

The following procedures should be conducted within 28 days preceding the start of treatment with carboplatin; treatment will commence within 28 days of trial entry to part 2 of study:

- Informed consent (using part 2 PIS and consent form)
- Inclusion/Exclusion criteria
- Physical examination
- Vital signs

- Height
- Weight
- ECOG Performance status
- ECG (12 lead)
- Translational blood and urine samples (see Appendix **Error! Reference s ource not found.** for collection and processing instructions)
- Haematology: FBC
- Biochemistry : Urea & electrolytes including calcium
- Liver function tests (including AST and/or ALT, ALP and bilirubin)
- CT of chest abdomen and pelvis, and bone scan
- Copy of the genetic report confirming DNA repair gene defect will be requested if participant has been tested prior to study entry.

8.2. On-treatment Assessments (Cycle X; day -2 to 1)

Participants will receive intravenous carboplatin AUC5 based on the Calvert Formula (using GFR measured by radioactive tracer e.g. DTPA clearance) as an infusion every 21 days.

The following assessments should be conducted prior to treatment on day 1 of each cycle (or day 20-21 of preceding cycle)

- Clinical review of any toxicity symptoms and physical examination if clinically indicated
- ECOG Performance status

21. Haematology and biochemistry blood tests:
    - Haematological parameters for treatment continuation: Hb ≥8.0g/dL; Neut ≥1.5; Plts ≥100 (patients with an Hb <8.0 g/dL can proceed with chemotherapy if the patient is asymptomatic and clinically fit for treatment in the investigator's opinion. A blood transfusion should be arranged prior to the next cycle of treatment).
    - Biochemistry parameters for treatment continuation: less than 20% increase in serum creatinine; Bilirubin ≤1.5x ULN; AST or ALT ≤2.5x ULN.
- Adverse Events assessment
- Concomitant medications

- Translational blood and urine samples (see Appendix **Error! Reference s ource not found.** for collection and processing instructions)
- GFR measured by clearance of a radioactive tracer, to be done prior to cycle 1 (Calculated GFR will be accepted for cycle 1 if measured GFR not available, but measured GFR must be available before a second dose of carboplatin is administered)

8.3. Treatment Response Assessments

- The first CT scan after the commencement of carboplatin will be carried out at the end of cycle 2. This should be performed between days 15 to 20 after the administration of cycle 2 of carboplatin to assess response. Disease will be assessed based on modified RECIST 1.1 criteria.
- The first bone scan after the commencement of carboplatin will be carried out at the end of cycle 3 (cycle 3 day 15-20). This is to allow for the possibility of 'flare' changes that may be seen if done before 8 weeks.
- For participants who continue carboplatin after 2 cycles, CT and bone scan will be repeated after cycle 5 of carboplatin and thereafter after every 3 cycles of treatment.
- PSA will be measured every 3 weeks from cycle 1 day 1. PSA response after cycle 2 will be measured 3 weeks after cycle 2 day 1. Confirmation of PSA response after two cycles of carboplatin will require a second PSA measurement 3 weeks (+/- 1 week) later. Participants who continue to receive further cycles of carboplatin will have their PSA measured at the end of each 3 weekly cycle.

8.4. End of study assessments

- Participants will discontinue treatment on study if there is radiological or clinical progression of disease, or if further treatment is not tolerated due to toxicity.
- An end of treatment safety review will be carried out 30 days after day 1 of the final cycle of carboplatin. This will include a clinical review, and blood tests including haematology and biochemistry tests.

8.5. Post treatment Follow-up

Follow up data will be collected on all participants entering part 2 of the study until death, including cause of death from hospital medical records or death certificate.

8.6. Discontinuation from Treatment

Participants may discontinue from trial treatment at any time at their own request, or they may be discontinued at the discretion of the Principal Investigator. Reasons for discontinuation will include:

- Disease progression: this may be clinical (e.g. worsening of cancer related symptoms) or radiological. Radiological progression will be defined using modified RECIST1.1 criteria (Appendix B). For bone scan assessments, progression will require the appearance of at least 2 new lesions compared to the first post-treatment scan (i.e. post cycle 3 scan) as per PCWG3 recommendations [191]. New lesions will be confirmed on a subsequent scan (6-8 weeks later).
- Unacceptable toxicity (e.g. unresolving grade ≥2 neuropathy or neutropenia)
- Any other reason deemed appropriate by investigator

Increases in PSA will not be a criterion for treatment discontinuation in the absence of clinical or radiological progression. Participants who discontinue treatment should continue to be followed up until death.

8.7. Discontinuation from Follow-up

If a participant withdraws from further follow-up a trial deviation form should be submitted to Oncogenetics Team stating whether the participant has withdrawn consent for information to be sent to the Oncogenetics Team or whether they simply no longer wish to attend trial follow up visits. In the very rare event that a participant requests that their data is removed from the study entirely, the implications of this should be discussed with the participant first to ensure that this is their intent and, if confirmed, the Oncogenetics Team should be notified in writing. If this request is received after results have been published the course of action will be agreed between the Sponsor and Independent Data Monitoring and Steering Committee.

Table 1: BARCODE 2 procedures and assessments schedule

Procedures & Assessments	Screening Within 28 Days of Day 0	Part 1 Mutation gene confirmation Day 0	Part 1 Mutation gene confirmation Wks 4-8[4]	Pre-treatment assessments (within 28 days before C1)	Part 2 On treatment C1 Day 1[2]	Part 2 On treatment C2 Day 1[2]	Part 2 On treatment C3 Day1[2]	Part 2 On treatment C4 Day 1[2]	Part 2 On treatment C5 onwards Day1[2]	End of study Within 30 days after final infusion (+/- 7 days)	Post-treatment Follow Up Every 12 Wks. until death
Informed consent		x		x							
Disease History/medical history	x			x							
Physical examination[1]				x							
Vital Signs				x							
WHO PS	x			x							
Weight & Height				x							
Inclusion/ Exclusion criteria	x			x							
Haematology & Biochemistry[3]	x			x	x	x	x	x		x	x
Germline DNA repair mutation gene confirmation[4]			x	x							
PSA				x	x	x	x	x	x	x	

Procedures & Assessments	Screening	Part 1 Mutation gene confirmation		Pre-treatment assessments (within 28 days before C1)	Part 2 On treatment					End of study	Post-treatment Follow Up
	Within 28 Days of Day 0	Day 0	Wks 4-8[4]		C1 Day 1[2]	C2 Day 1[2]	C3 Day1[2]	C4 Day 1[2]	C5 onwards Day1[2]	Within 30 days after final infusion (+/- 7 days)	Every 12 Wks. until death
ECG (12 lead)				x							
GFR by tracer clearance				x							
CT chest/abdo/pelvis[5]				x		x			x	x	
Bone scan[6]				x			x		x	x	
Carboplatin infusion					x	x	x	x	x		
Adverse Events[7]				x	x	x	x	x	x	x	
Con meds				x	x	x	x	x	x	x	
Translational blood and/ or urine samples[8]				x		x	x		x[8]	x	
Survival and progression status[9]											x

7.5.1.5 Footnotes

1. After cycle 1, a limited physical examination should be repeated if clinically indicated.
2. Evaluations for day 1 of each cycle of chemotherapy can be carried out in the preceding 48 hours.
3. Haematology and biochemistry blood tests should be done on day 1 of each cycle or within 48 hours. Biochemistry tests to include: sodium, potassium, urea, creatinine, calcium, magnesium, bilirubin, ALP and ALT

4. Profiling for mutation in a DNA repair gene within the study may take 8-12 weeks from the time of sample collection (taken after consent form signed for Part 1). For the purposes of cancer risk estimation for family members of the participant, those found to have a mutation in a gene that is used in clinical genetic practice will be referred to a genetics clinic to have this confirmed in a diagnostic laboratory. This can be done simultaneously while proceeding to part 2 of study or after treatment. Participants with a known germline mutation prior to entry into the study can enter part 2 if they meet eligibility criteria. See Appendix **Error! Reference source not found.** for collection and processing instructions.

5. Baseline CT scan to be done within 28 days preceding cycle 1 carboplatin and first reassessment CT to be done 15-20 days after C2 day1. In participants who have a CR, PR or SD, carboplatin treatment will continue every 3 weeks. <u>CT should be repeated every 3 cycles of treatment (+/- 7 days) and after C10 if 10 cycles completed.</u>

6. Baseline bone scan to be done within 28 days preceding cycle 1 carboplatin and first reassessment bone scan to be done 15-20 days after C3 day1. In participants who continue treatment after cycle 3, a bone scan will be repeated after cycle 5 (day 15-20) and then every 3 cycles of treatment (+/- 7 days) and after cycle 10 if 10 cycles completed.

7. Any toxicity, sign or symptom will be collected up to 30 days after last administration of treatment.

8. Blood +/- urine samples will be collected for future translational studies. Blood samples for plasma collection will be taken at baseline pre-treatment and again at the end of cycle1, cycle 2, cycle 5 and at the end of treatment. A sample will also be collected at progression (may be same time as end of treatment) if the patient is still being managed at the study centre. See Appendix **Error! Reference source not found.** for collection and processing instructions.

9. Participants to be contacted or participant records reviewed for survival follow up every 3 months until death. For participants who do not progress on study, data related to first progression will be recorded.

9. TRIAL TREATMENT

Carboplatin is an investigational medicinal product within BARCODE 2. Common toxicities associated with carboplatin include allergy (rash often with pruritus), hypersensitivity reactions (usually after > 6 cycles), alopecia (very occasionally), nausea and vomiting, bone marrow suppression, flushing effects, nephrotoxicity, fatigue, neurotoxicity, nausea and vomiting.

9.1. Dose and Schedule

Participants will receive intravenous carboplatin as an infusion every 3 weeks. The dose will be calculated using the Calvert formula based on an AUC5 and GFR calculated after radioactive tracer injection:

$$\text{Carboplatin Dose} = \text{AUC} \times (\text{GFR}+25)$$

Use AUC=5 if GFR measured by radioactive tracer clearance

Use AUC=6 if GFR calculated by Cockcroft-Gault formula

Carboplatin dose will be capped at 750mg.

If there is a delay in obtaining the measured GFR for cycle 1, carboplatin can be dosed based on AUC 6 and using a calculated creatinine clearance (Appendix A3 for Cockcroft-Gault formula). For cycle 2 onwards, the measured GFR should be used to dose carboplatin. If the measured GFR was used for cycle 1, the dose remains the same unless there is a change of more than 20% in serum creatinine. In this instance, the measured GFR test should be repeated prior to the next cycle of carboplatin. If repeating the measured GFR test will cause a delay to treatment then the dose should be recalculated using Cockcroft-Gault formula to allow carboplatin to be given.

Subsequently, the measured GFR test should be repeated prior to the next cycle of treatment.

9.2. Prescription and Dispensing

Study medication should be dispensed and handled as per local practice. A trial specific prescription approved by the study PI and the Royal Marsden pharmacy should be devised by the site.

9.3. Participant Cards

Small wallet sized cards will be given to participants participating in part 2. Each card will state:

- the name of the participating centre
- that the participant is participating in the BARCODE 2 trial
- that the participant is taking carboplatin
- an emergency contact number

9.4. Duration of Treatment

Participants will receive carboplatin AUC5 infusions every 21 days. For participants who have a response or stable disease on imaging after 2 cycles of treatment, chemotherapy will continue until disease progression or until participants stop tolerating treatment due to toxicity.

9.5. Supportive Care

Prophylactic antibiotics and G-CSF may be given with carboplatin according to local protocol or for the persistence of neutropenic fever according to NCCN guidelines (http://www.nccn.org/professionals/physician_gls/pdf/myeloid_growth.pdf) and ASCO guidance [192].

Pre- and post-medication with standard antiemetics will be administered to every participant as per local practice.

9.6. Concomitant Therapy

All medication considered necessary for the participants' welfare and which is not expected to interfere with the evaluation of the study drug may be given at the discretion of the investigator. During the study period, no other anticancer or hormonal treatment will be given although **a LHRH analogue must be continued, unless the participant is surgically castrate.**

All concomitant medications must be recorded in the participant's notes, as well as the appropriate pages of the CRF.

7.5.1.6 Specific recommendations:

- For participants on oral anti-coagulants such as warfarin, INR should be more frequently monitored during treatment due to the risk of interaction.
- Live attenuated vaccines must be avoided due to risk of severe systemic disease. Yellow fever vaccine is contraindicated.
- Ciclosporin (and by extrapolation tacrolimus and sirolimus) causes excessive immunosuppression with risk of lymphoproliferation and concurrent use should be with caution.
- Phenytoin: risk of exacerbation of convulsions due to the decrease of phenytoin digestive absorption by carboplatin.
- Caution with the following drugs:
  - Aminoglycoside antibiotics, vancomycin and capreomycin: increased risk of nephrotoxicity and ototoxicity
  - **Clozapine:** increased risk of agranulocytosis, avoid concomitant use
  - **Diuretics:** increased risk of nephrotoxicity and ototoxicity
  - **Nephrotoxic drugs:** increased nephrotoxicity; not recommended

9.7. Dose Modifications

In order to maintain dose-intensity and cumulative dose-delivery reasonable efforts should be taken to minimise dose reduction and treatment delays. Participants whose

treatment is delayed because of toxicity should be evaluated on a weekly basis until adequate recovery has been made (toxicity improves to ≤grade 1). Toxicity is graded according to NCI Common Terminology Criteria for Adverse Events v4.03.

Full Blood Count and renal function must be measured prior to each cycle (i.e. no more than 1 week prior to the first cycle or no earlier than 24 hours before the start of subsequent cycles). Carboplatin can be administered when neutrophil count is greater than or equal to 1.5×10^9/L and platelet count greater than or equal to 100×10^9/L. If serum creatinine increases by more than 20% then GFR must be recalculated (see above section 9.1) and the dose of carboplatin adjusted to maintain the target AUC. If carboplatin is delayed for three consecutive weeks due to haematological toxicity and blood counts have not recovered, then treatment should be stopped and further treatment conducted at the discretion of the treating clinician.

Table 2: Clinical parameters for treatment continuation

Toxicity	Delay/Dose modification
Febrile neutropenia	Add GCSF to next cycle of treatment. If GCSF not added, reduce dose by 1xAUC If febrile neutropenia recurs after addition of GCSF, reduce dose by 1xAUC
Neutrophils >0.5 -1.4 or Platelets >50 -99	Delay 1 week until counts recovered. If only neutrophils are low, consider GCSF with next cycle.

	If both neutrophils and platelets are low, consider dose reduction of 1xAUC
Neutrophils ≤0.5 or Platelets ≤50	Delay 1 week until counts recovered and reduce dose by 1xAUC Add GCSF to next cycle of treatment
Serum creatinine rise of 20% or more	Repeat measured GFR and dose according to GFR. Stop carboplatin if CrCl<20ml/min
Neuropathy ≥grade 2	Delay next dose until recovers to G0-1. If no improvement after 3 weeks, stop treatment
Any other toxicity ≥grade 2	Delay next dose until recovers to G0-1. If no improvement after 3 weeks, stop treatment

(Abbreviations: GCSF= granulocyte colony stimulating factor)

A dose reduction by 1xAUC (i.e. to AUC4 if AUC5 was used) for subsequent cycles should be given for the following reasons:

- Recurrence of febrile neutropenia after the addition of GCSF; if GCSF is not being added then dose should be reduced after first episode
- Neutrophil count >0.5 -1.4 x 10^9/L despite addition of GCSF
- Neutrophil count ≤0.5 x 10^9/L on day 1 of treatment;
- A dose delay for haematological toxicity on two separate occasions;
- Platelet count of <50 x 10^9/L;
- Day 21 neutrophil count of <1 x 10^9/L with platelet count of <75 x 109/L; or

- ≥ grade 2 peripheral neuropathy (Restart treatment once improves to ≤grade 1; if not improving after 3 weeks, treatment will be discontinued)

9.8. Drug Supplies and Pharmacy Responsibilities

Carboplatin is an investigational medicinal product within BARCODE 2 and should be prescribed by the investigator and dispensed from hospital pharmacy from hospital stock for the duration of the trial. All IMPs should be obtained from usual drug suppliers in accordance with local practice.

Drug formulation, storage, accountability and destruction should be in accordance with local policy.

10. PHARMACOVIGILANCE

10.1. Definitions

7.5.1.7 Adverse Event (AE)

An AE is any untoward medical occurrence in a participant or clinical trial subject administered an investigational medicinal product; the event does not necessarily have a causal relationship with the treatment or usage.

7.5.1.8 Serious Adverse Event (SAE)

An SAE is any untoward medical occurrence that occurs after the commencement of study treatment and within 30 days of the last administration and:

- results in death,
- is life-threatening
- requires hospitalisation or prolongation of existing in participants´ hospitalisation
- results in persistent or significant disability or incapacity
- is a congenital anomaly or birth defect

Important adverse events that are not immediately life-threatening or do not result in death or hospitalisation but may jeopardise the subject or may require intervention to prevent one of the other outcomes listed in the definition above, may also be considered serious.

Progression of the indicated disease and death due to progression of the indicated disease are not considered SAEs.

Planned hospital admissions for elective procedures will not be treated as SAEs.

Pregnancy or aid in the conception of a child whilst participating in a trial is not itself considered an SAE but should be followed up for congenital anomalies or birth defects.

7.5.1.9 Serious Adverse Reaction (SAR)

A serious adverse reaction is an SAE that is suspected as having a causal relationship to the investigational medicinal product, as assessed by the investigator responsible for the care of the participant. A suspected causal relationship is defined as possibly, probably or definitely related (see definitions of causality table).

Table 3: Definitions of causality

Relationship	Description
Unrelated	There is no evidence of any causal relationship with the trial drug
Unlikely	There is little evidence to suggest there is a causal relationship (e.g. the event did not occur within a reasonable time after administration of the trial medication). There is another reasonable explanation for the event (e.g. the participant's clinical condition, other concomitant treatment)
Possible	There is some evidence to suggest a causal relationship (e.g. because the event occurs within a reasonable time after administration of the trial medication). However, the influence of other factors may have contributed to the event (e.g. the participant's clinical condition, other concomitant treatments)
Probable	There is evidence to suggest a causal relationship, and the influence of other factors is unlikely
Definitely	There is clear evidence to suggest a causal relationship, and other possible contributing factors can be ruled out

Not assessable	There is insufficient or incomplete evidence to make a clinical judgement of the causal relationship.

7.5.1.10 *Suspected Unexpected Serious Adverse Reaction (SUSAR)*

A serious adverse reaction, the nature or severity of which is not consistent with the safety information provided in the applicable Summary of Product Characteristics (SmPC), and is assessed as unexpected by the Chief Investigator.

10.2. Reporting Adverse Events to Oncogenetics Team

Any toxicity, sign or symptom that occurs after commencement of study treatment and within 30 days of the last administration of study treatment, which is not unequivocally due to progression of disease, should be considered an AE.

All AEs must be reported on the relevant CRF and submitted to the Oncogenetics Team for recording on the study database.

The severity of AEs should be graded according to the NCI CTCAE V4.02 criteria (see Appendix F). For each AE, the highest grade observed since the last visit should be reported.

Whenever one or more toxicity/sign/symptom corresponds to a disease or a well-defined syndrome only the main disease/syndrome should be reported.

10.3. Reporting of Serious Adverse Events to Oncogenetics Team

Any SAE that occurs after the commencement of study treatment and up to 30 days following the last dose of study drug must be reported.

Any SAEs that occur more than 30 days after the last dose of study drug that, in the opinion of the Principal Investigator, are related to the study drug should be reported to the BARCODE2 trial team if the Principal Investigator becomes aware of them.

All SAEs should be reported to the BARCODE 2 trial team within 24 hours of the Principal Investigator (or designated representative) becoming aware of the event, by completing the BARCODE 2 SAE form and (if applicable) faxing to:

BARCODE 2 trial team

Fax no: **+44 (0)208 722 4110**

As much information as possible, including the Principal Investigator's assessment of causality, must be reported to the BARCODE2 trial team in the first instance. Additional follow up information should be reported as soon as it is available.

All SAE forms must be completed, signed and dated by the Principal Investigator or designated representative.

10.4. Review of Serious Adverse Events

The Chief Investigator (or designated representative) will assess all reported SAEs for causality and expectedness. As the CI in this trial also holds the role of PI, all SAE forms will be reviewed by an independent SAE reviewer. SAEs assessed as having a causal relationship to study drug and as being unexpected (SUSARs) will undergo expedited reporting to the relevant authorities and all other interested parties by the Oncogenetics Team (see Figure 7).

The centre should respond as soon as possible to requests from the Chief Investigator or designated representative for further information that may be required for final assessment of an SAE.

10.5. Expedited Reporting of SUSARs

If an SAE is identified as being a SUSAR by the Chief Investigator, and is fatal or life threatening, it will be reported by the BARCODE 2 trial team to the MHRA, the main REC, and all other interested parties within 7 days of being notified of the event.

If an SAE is identified as a SUSAR by the Chief Investigator, and is not fatal or life threatening, it will be reported by the BARCODE 2 trial team to the MHRA, the main REC within 15 days of the BARCODE 2 trial team being notified of the event.

The BARCODE 2 trial team will report any additional relevant information to the MHRA and main REC as soon as possible, or within 8 days of the initial report of a fatal/life threatening SUSAR.

10.6. Follow up of Serious Adverse Events

SAEs should be followed up until clinical recovery is complete or until disease has stabilised. SAE outcomes should be reported to the BARCODE 2 trial team using the relevant section of the SAE form as soon as the Principal Investigator or designee becomes aware of the outcome.

10.7. Annual Reporting of Serious Adverse Events

An annual report of all SAEs will be provided to the MHRA and the main REC by the BARCODE 2 trial team and copied to the sponsor office at the end of the reporting year.

10.8. Reporting Pregnancies

If a trial participant's partner becomes pregnant while receiving study drug or up to 6 months after receiving study drug, this should be reported to the BARCODE 2 trial team using the pregnancy reporting form. Pregnancies should be followed up until conclusion and all follow-up information should be reported to the Oncogenetics Team. If the outcome of the pregnancy meets the definition of serious (i.e. congenital abnormality)

this should be reported to the Oncogenetics Team following the serious adverse event

reporting procedures described above.
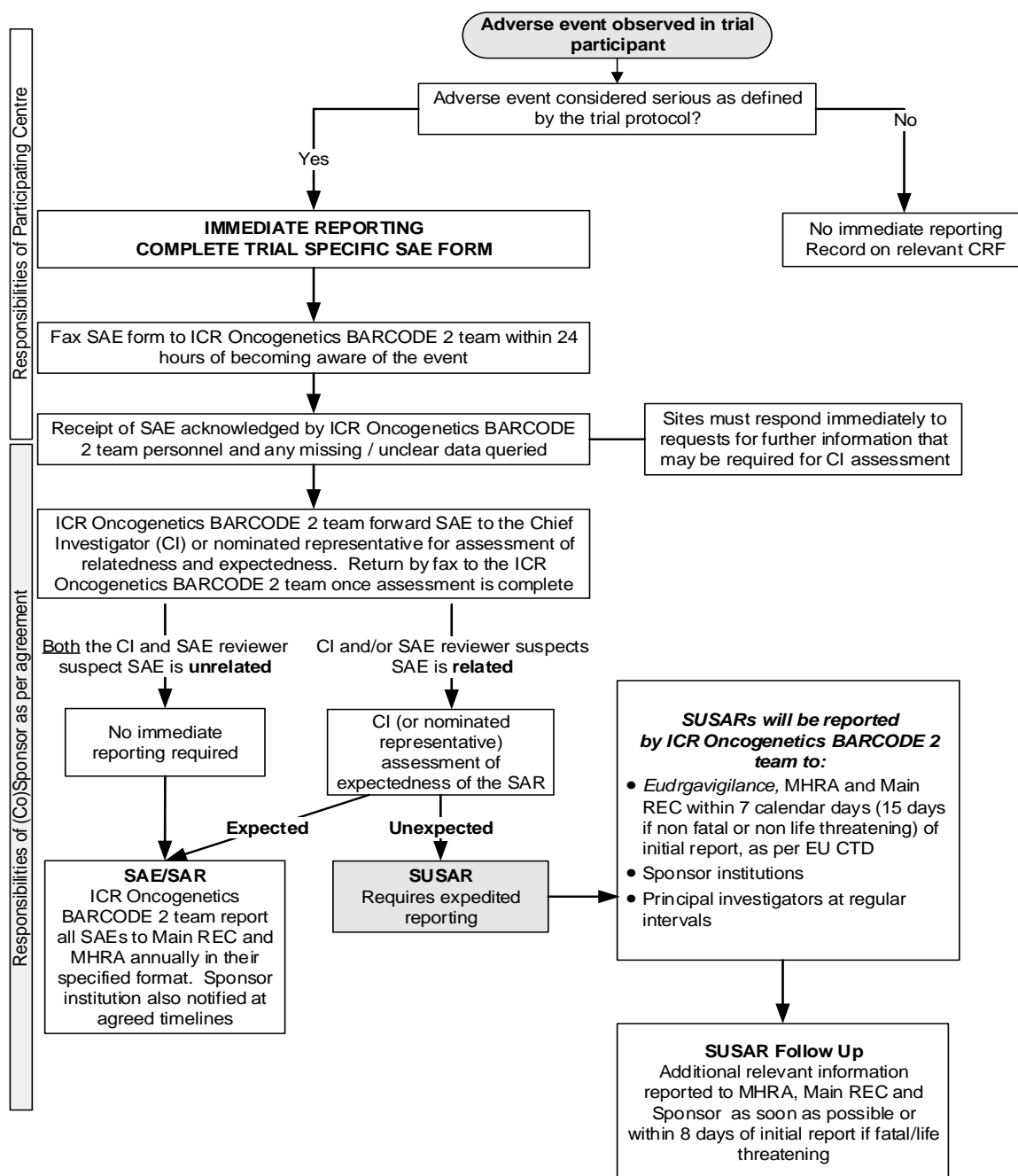
Figure 7: Flow diagram for SAE reporting, and action following report

11. STATISTICAL CONSIDERATIONS

11.1. Statistical Design and Sample Size Justification

The Oncogenetics Team will analyse a genetic panel of DNA repair genes in approximately 450 (anticipated 360-540) mCRPC participants to determine the incidence of germline mutations in participants with mCRPC. Additionally, participants can enter part 2 of the study directly if they have been diagnosed with mCRPC and are already a known carrier of a mutation in one of these genes. The current participant throughput has the volume to be able to ascertain these numbers allowing for a 60% uptake of the study. As a result of recently reported data regarding the frequency of germline DNA repair gene mutations in mCRPC participants, it is anticipated that a 10-15% mutation incidence will be found . In the study team's current research on DNA repair gene mutations, 1000 prostate cancer cases and 1000 controls have been sequenced using a panel of 193 genes. 40 genes were found to harbour a germline mutation in at least one case and the final analysis is currently taking place (gene list can be found in Appendix A). A gene panel of approximately 115 genes will be used in this study and participants will be categorised into one of 3 gene groups as per Table 4. As translational research in this area is ongoing, both in our research group and others, the gene list and gene groups may be modified in keeping with any new information found.

11.2. Treatment Allocation

All trial participants who are found to have a class 4/5 mutation in a DNA repair gene will be entered into part 2 of the study and will receive treatment with carboplatin if the eligibility criteria are met.

11.3. Sample Size

Response rates to carboplatin will be analysed within gene groups. These gene groups have been determined based on previous and developing evidence of involvement in

PrCa risk and development and knowledge of DNA repair gene pathways. These groups will be reviewed throughout the study based on the mutations found in part one as well as evolving information on DNA repair genes and pathways; this may lead to further gene groups being added during the study.

It is anticipated that, in order to identify 54 participants with a mutation, 450 participants will be required to be screened in part 1 (assuming a total prevalence of 12% split equally between each mutation group). However it is acknowledged that the overall prevalence rate could be as low as 10% or as high as 15% which would necessitate an increase/decrease in screening sample size to 540 or 360 respectively. Similarly if the prevalence of a given mutation is lower than the other cohorts this may also result in a higher number of participants required to be screened in order to find 18 participants in the given cohort. The overall prevalence and prevalence of each cohort will be monitored throughout the study and cohorts may be dropped if the observed prevalence rate is too low to consider full recruitment feasible in the given cohort. With a total of 450 participants and a mutation rate of 12% the 95% confidence interval around the estimate of prevalence would be 9.1 – 15.4%.

For part 2, using a Simon minimax Two-stage design this study has 90% power and a one-sided significance level of 5% to discount a response rate of 10% (p0) in favour of a response rate of 40% (p1) within each gene group. 12 evaluable participants will be treated in each gene group in stage 1 and, if 2 or more responses are seen for a given gene group, a further 6 participants will be recruited to that gene group bringing the total number of participants to 18 (Table 4). If 5 or more responses are seen out of 18 evaluable participants in a gene group the study will conclude that the activity of carboplatin in the gene group warrants further research. Non-evaluable participants will be replaced, for example, if a participant withdraws prior to cycle 2 for a reason unrelated to disease or treatment. Replacement of participants will be agreed with the

Independent Data Monitoring and Steering Committee. In the event that the required number of responses is observed prior to completion of recruitment, recruitment will continue up to the total target sample size. In the absence of any major safety concerns, there will be no break in recruitment between part 1 and part 2 i.e. if one response is observed in the first 12 participants and the 12th outcome is still pending, a 13th participant can be recruited.

If one gene group completes recruitment but further participants with mutations in those genes are detected through the continuing genetic profiling on study, over recruitment and treatment with carboplatin will be allowed within gene groups as long as the target response rate was observed (at least 5 responses in each group).

Mutation rates will be reviewed after 100 patients and may be used to update the target sample size if observed rates differ from those assumed in the power calculations.

Table 4: Gene groups and sample sizes

Gene/s	Stage one sample size	Stage two sample size (number of additional participants that will be recruited)
BRCA1 & *BRCA2*	12	6
Lynch syndrome/ Mismatch repair genes (*MLH1, MLH3, MSH2, MSH5, MSH6, PMS1, PMS2*)	12	6

Other: e.g. *CHEK2, PALB2, MUTYH, FANC genes, ATM, NBS1*	12	6

11.4. Endpoint Definitions

7.5.1.11   Primary endpoint

Response will be defined by objective response (partial or complete response) according to modified RECIST 1.1 criteria and/or decrease in PSA of ≥50% after two cycles of carboplatin. In participants with bone only metastatic disease, response will be recorded as 'new lesions' or 'no new lesions' (as per PCWG3 guidance; Appendix B) [191]. Participants with no new bony lesions will be deemed as having stable disease.

7.5.1.12   Secondary endpoints

1. **Incidence of germline DNA repair gene mutations in CRPC:** this will be calculated from the rate of pathogenic mutations observed in the group of men who undergo genetic profiling within part 1 of the study.
2. **Overall survival:** OS will be measured from the date of cycle 1 day 1 to the date of death (whatever the cause). Survival time of living participants will be censored on the last date a participant is known to be alive or lost to follow-up.
3. **Progression free survival:** PFS will be measured from the date of cycle 1 day 1 until radiographic progression on CT or bone scan, unequivocal clinical progression or death. If no event exists, then PFS will be censored at the last scheduled disease assessment on study
4. **Cause specific survival:** from date of first diagnosis of prostate cancer to date of death from prostate cancer. Participants who have not died or have died of non-prostate cancer causes will be censored on the last date a participant is known to be alive or date of death respectively.
5. **Radiographic progression free survival**: (rPFS) will be defined by modified RECIST 1.1 progression on CT or progression on bone scan (using PCWG3 criteria). It will be measured from the date of cycle 1 day 1 to the first occurrence of radiographic progression or death from any cause. If no event

exists, then rPFS will be censored at the last scheduled disease assessment on study.

6. **Time to radiographic progression:** will be measured from the date of cycle 1 day 1 to the first occurrence of radiographic progression (by modified RECIST1.1 or PCWG3 criteria for bone lesions). Death from prostate cancer or any other cause without prior radiographic evidence of progression will not count as an event. If no event exists, then time to radiographic progression will be censored at the last scheduled disease assessment on study or date of death whichever occurs earlier.

7. **Time to PSA progression:** For participants who have achieved a ≥ 50% decrease from cycle 1 day1 (baseline), the PSA progression date is defined as the date that a ≥ 25% increase and an absolute increase of ≥ 2 ng/mL above the **nadir** is documented. This must be confirmed by a second consecutive value obtained 4 or more weeks later. For participants without a PSA decrease of this magnitude or no decrease at all, PSA progression date is defined as the date that a ≥ 25% increase and an absolute increase of ≥ 2 ng/mL above the **baseline** is documented. This must also be confirmed by a second consecutive value 4 or more weeks later.

8. **Duration of PSA response**: is calculated from the time the PSA value first declines by at least 50% of the cycle 1 day 1 (baseline) value (must be confirmed by a second value) until the time there is an increase of 25% of PSA nadir, provided the absolute increase is at least 2 ng/mL. The increase must be confirmed by a second consecutive measurement that is at least 25% above the nadir. If the PSA never shows a 25% increase over the nadir value, then the participant will be assessed at the last PSA measurement.

9. **Pattern of PSA Responses:** will be calculated as the percentage of change in PSA from baseline to 12 weeks (or earlier for those who discontinue therapy), as well as the maximum decline in PSA that occurs at any point after treatment.

7.5.1.13   *Analysis Plan*

Response rates within each gene group will be presented along with a 95% confidence interval.

The incidence of germline DNA repair gene mutations will be presented along with a 95% confidence interval. Mutation incidence rates will be reviewed when gene profiling results are available for 100 Part 1 participants.

For OS, PFS, CSS, radiographic progression free survival, time to radiographic progression and time to PSA progression median survival and fixed time point (e.g. 1-, 2- and 5-year) survival rates will be estimated using the Kaplan-Meier method, and survival curves generated for each group.

Duration of PSA response will be summarised by the median and presented along its 95% confidence interval.

Waterfall plots will be presented (as per PCWG3 recommendations) that show the percentage change in PSA from baseline to 12 weeks (or earlier for those who discontinue therapy), as well as the maximum decline in PSA that occurs at any point after treatment.

All participants receiving at least one cycle of carboplatin chemotherapy will be included in the primary analysis population.

All primary analyses will be carried out separately within the different gene groups and formal analysis of comparisons of outcomes between gene groups is not planned. No adjustment will be made for multiplicity as this study is hypothesis generating.

12. TRIAL MANAGEMENT

12.1. Trial Management Group (TMG)

A Trial Management Group (TMG) will be set up and will include the Chief Investigator, the Trial Statistician and Trial Manager. Key study personnel will be invited to join the TMG as appropriate to ensure representation from a range of professional groups, including a PI from a Participant Identification Centre. Membership will include a lay/consumer representative, who will receive support and training as deemed necessary and be reimbursed in line with INVOLVE guidelines relating to PPI. The TMG will meet at regular intervals, and at least annually. Notwithstanding the legal obligations of the sponsor and Chief Investigator, the TMG have operational responsibility for the conduct of the trial. The Committee's terms of reference, roles and responsibilities will be defined in a charter.

12.2. Independent Data Monitoring and Steering Committee (IDMSC)

A joint Independent Data Monitoring and Steering Committee (IDMSC) will be set up to oversee the safety of the trial participants, monitor the data produced by the trial, put these data into overall context and supervise the progress of the trial towards its interim and overall objectives. A list of the IDMSC members is available from the study team.

13. RESEARCH GOVERNANCE

13.1. Sponsor Responsibilities

The sponsor of the BARCODE 2 trial is the Institute of Cancer Research (ICR).

14.   TRIAL ADMINISTRATION & LOGISTICS

14.1. Data Acquisition

Case Report Forms (CRF) will be used for the collection of trial data. The Oncogenetics Team will provide guidance to the centre to aid the completion of the CRFs. The Trial Management Group reserves the right to amend or add to the CRF template as appropriate. Such changes do not constitute a protocol amendment, and revised or additional forms should be used by the centre in accordance with the guidelines provided by the ICR Oncogenetics BARCODE2 study team.

14.2. Central Data Monitoring

Once data has been entered on the CRF by the centre personnel, the Trial Coordinator will review incoming CRFs for compliance with the protocol, and for inconsistent or missing data. Should any missing data or data anomalies be found, queries will be sent to the relevant Royal Marsden staff for resolution. Following initial review, the Trial Coordinator will enter the CRF data items into the central clinical study database held at ICR.

Any systematic inconsistencies identified through central data monitoring may trigger an on-site monitoring visit.

14.3. On-Site Monitoring

The CRFs will not be made available to people outside of the research team, however, access will be granted for audit and monitoring purposes and will be provided to regulatory authorities, Research Ethics Committee or other relevant ICR or Trust personnel.

If a monitoring visit is required, the Trial Coordinator (Oncogenetics Team) will contact the centre to arrange the visit. Once a date has been confirmed, the centre should

ensure that full patient notes of participants selected for source data verification are available for monitoring.

The trial personnel conducting on-site monitoring will review essential documentation and carry out source data verification to confirm compliance with the protocol. If any problems are detected during the course of the monitoring visit, the trial coordinator will work with the Principal Investigator or delegated individual to resolve issues and determine appropriate action.

14.4. Completion of the Study and Definition of End of Study

The study end date is deemed to be the date of last data capture.

14.5. Archiving

Essential trial documents should be retained according to local policy and for a sufficient period for possible inspection by the regulatory authorities (at least 5 years after the date of last data capture). Documents should be securely stored and access restricted to authorised personnel.

DNA collected during this study will be stored indefinitely if consented to by participants signing the study consent form.

Clinical data collected in the study will be stored for at least 5 years.

15. PARTICIPANT PROTECTION AND ETHICAL CONSIDERATIONS

15.1. Trial Approvals

The trial will receive ethics approval from a research ethics committee, regulatory approval from the MHRA and R&D approval via the NIHR Health and Research Authority (HRA). Before recruiting participants, the Principal Investigator at the centre is responsible for submitting Site Specific Information and gaining local Research and Development approval of this protocol.

15.2. Trial Conduct

This trial will be conducted according to the approved protocol and its amendments, supplementary guidance and manuals supplied by the sponsor and in accordance with The Medicines for Human Use (Clinical Trials) Regulations 2004 as amended, the Research Governance Framework for Health and Social Care and the principles of GCP.

15.3. Informed Consent

Participants should be asked to sign the current ethics approved BARCODE 2 consent forms at trial entry after receiving both verbal and written information about the trial, having been given sufficient time to consider this information. All consent forms must be countersigned by the Principal Investigator or a designated individual. A signature log of delegated responsibilities, listing the designated individuals and the circumstances under which they may countersign consent forms, must be maintained at the participating site. This log, together with original copies of all signed participant consent forms, should be retained in the Site Investigator File and must be available for inspection. The current ethics approved BARCODE 2 participant information sheets should be provided in addition to any standard participant information sheets that are provided by the site and used in routine practice.

15.4. Participant Confidentiality

Participants will be asked to consent to their full name being collected at trial entry in addition to their date of birth, hospital number, postcode and NHS number or equivalent to allow linkage with routinely collected NHS data and ensure accuracy in handling biological samples.

Each investigator should keep a separate log of all participants' Trial IDs, names, addresses and hospital numbers. The investigator must retain trial documents (e.g. participants' written consent forms) in strict confidence. The investigator must ensure the participants' confidentiality is maintained at all times.

Representatives of the Sponsor, ICR-CTSU and the regulatory authorities may require access to participants' hospital notes for quality assurance purposes. The ICR Oncogenetics Team and ICR-CTSU will maintain the confidentiality of participants at all times and will not reproduce or disclose any information to third parties by which participants could be identified (without consent).

15.5. Data Protection

The study will comply with all applicable data protection laws.

15.6. Insurance and Liability

Indemnity to meet the potential legal liability of investigators participating in this trial is provided by the usual NHS indemnity arrangements.

16.   FINANCIAL MATTERS

The sponsor has received an Investigator Initiated Research grant (IIR) from the European Research Council for the conduct of this trial.

17.   PUBLICATION POLICY

The main trial results will be published in a peer-reviewed journal, on behalf of all collaborators. The manuscript will be prepared by a writing group, consisting of members of the TMG. Participating clinicians may be selected to join the writing group on the basis of intellectual and time input. All participating clinicians will be acknowledged in the publication.

Any presentations and publications relating to the trial must be authorised by the TMG. Authorship of any secondary publications e.g. those relating to future sub-studies, will reflect intellectual and time input into these studies.

No investigator may present or attempt to publish data relating to the BARCODE 2 trial without prior permission from the TMG.

18.   APPENDICES

A. DNA REPAIR GENES

This is the initial list of genes that will be screened. This gene list has been determined based on previous and emerging evidence of involvement in PrCa risk and development and knowledge of DNA repair gene pathways. This list and groups of genes will be reviewed throughout the study based on the mutations found in part one as well as evolving information on DNA repair pathways; this may lead to the gene list being refined or expanded and gene groups being added during the study.

DNA Repair Genes

Group 1	*BRCA1, BRCA2*
Group 2	*MLH1, MLH3, MSH2, MSH5, MSH6, PMS1, PMS2*
Group 3	*ALKBH3, ANO7, APEX1, AR, ATM, ATR, ATRIP, BAP1, BARD1, BLM, BRIP1, CCNH, CDC25C, CDH1, CDK4, CDKN2A, CHD1, CHEK1, CHEK2, CLK2, DCLRE1A, EME1, EME2, ERCC2, ERCC5, ERCC6, ESR2, EXO1, FAM175A, FANCA, FANCD2, FANCI, FANCL, FANCM, GADD45A, GEN1, GTF2H2, GTF2H3, GTF2H4, HOXB13, HUS1, LIG1, LIG3, LIG4, MMS19, MNAT1, MPG, MRE11A, MSR1, MUTYH, NABP2, NBN, NEIL1, NEIL2, NTHL1, OGG1, PALB2, PARP2,*

PER1, PNKP, POLD1, POLE, POLK, POLM, POLN, POLQ, POT1, PRSS1, PTCH1, PTEN, RAD1, RAD50, RAD51B, RAD51C, RAD51D, RAD52, RAD54B, RAD54L, RB1, RECQL, RECQL4, RECQL5, RINT1, RNASEL, RPA1, SETMAR, SLX4, SMAD4, SMARCA4, SMUG1, SPOP, STK11, TDG, TOP2A, TOP2B, TOP3A, TP53, TP53BP1, WRN, XAB2, XPA, XPC, XRCC1, XRCC2, XRCC4, XRCC5

Appendix 3: BARCODE2 Gene Panel Expected Coverage

Agilent Design Summary

Design Name: BARCODE_V2

Species: H. sapiens (H. sapiens, hg19, GRCh37, February 2009)

Target Summary

136 Target IDs resolved to 136 targets comprising 1818
regions.

Region Size: 504.628 kbp

Probe Summary

Total Probes: 31622

Total Probes Size: 587.934 kbp

Coverage: 98.85%

Target Parameters

Databases: Gencode, SNP,
CytoBand

Region: Coding Exons

Region Extension: 50 bases from 3' end and 50 bases from 5' end.

Probe Tiling Parameters

Tiling density: 5x

Masking: Most Stringent

Boosting: MaximizePerformance

Target and Probe Details

Coverage: The percentage of bases overlapped by probes extended by +/-
100 base pairs to represent likely capture.

Gene	Interval	Size (bp)	Coverage
ALKBH3	chr11:43904153-43941610	2007	100
ANO7	chr2:242127977-242163618	5398	100
APEX1	chr14:20923755-20925717	1357	100
AR	chrX:66764939-66943733	3833	100
ATM	chr11:108098302-108236285	15349	100
ATR	chr3:142168221-142297596	12619	100
ATRIP	chr3:48488200-48507003	3676	100
BAP1	chr3:52436254-52443944	3870	100
BARD1	chr2:215593350-215674343	3550	100
BLM	chr15:91290573-91358559	6354	100
BRCA1	chr17:41197645-41277252	8193	100
BRCA2	chr13:32890548-32972957	12841	100
BRIP1	chr17:59760607-59938950	5730	100
CCNH	chr5:86690213-86708661	1929	100
CDC25C	chr5:137621331-137666919	2722	100
CDH1	chr16:68771269-68867452	4249	100
CDK4	chr12:58142258-58145550	1612	100
CDKN2A	chr9:21968158-21994503	1527	100
CHD1	chr5:98192034-98262140	8633	100
CHEK1	chr11:125495606-125525265	2997	100
CHEK2	chr22:29083835-29130759	3312	93
RAD51B	chr14:69060958-69061615	658	98
RAD51B	chr14:69077474-69078290	817	89
RAD51B	chr14:69117248-69117823	576	57
MPG	chr16:128888-129580	693	100
RAD51D	chr17:33443628-33444306	679	72
TP53	chr17:7569154-7569812	659	100
TP53	chr17:7576275-7576907	633	67
SMARCA4	chr19:11096615-11097519	905	100
SMARCA4	chr19:11144193-11144791	599	80
SMARCA4	chr19:11170179-11170811	633	100
XRCC1	chr19:44047250-44047907	658	100
XRCC1	chr19:44055919-44056677	759	74
RNASEL	chr1:182544296-182544963	668	94
PMS1	chr2:190670898-190671478	581	56
MSH2	chr2:47636983-47637761	779	74
SETMAR	chr3:4344738-4345460	723	100
SETMAR	chr3:4354338-4356036	1699	77
OGG1	chr3:9800621-9801220	600	64
OGG1	chr3:9807243-9808119	877	65
PRSS1	chr7:142458670-142459211	542	83
CLK2	chr1:155232959-155240818	2694	100
DCLRE1A	chr10:115594861-115612991	4023	100

EME1	chr17:48452520-48458350	2526	100
EME2	chr16:1823179-1826289	2036	100
ERCC2	chr19:45854837-45873848	4634	97
ERCC5	chr13:103498567-103528303	5188	100
ERCC6	chr10:50666811-50741060	6467	100
ESR2	chr14:64551585-64749753	3080	92
EXO1	chr1:242013678-242052952	3841	100
FAM175A	chr4:84383572-84406275	2452	91
FANCA	chr16:89804959-89883073	8760	100
FANCD2	chr3:10070292-10142996	8822	98
FANCI	chr15:89790829-89859740	7628	100
FANCL	chr2:58386850-58468498	2524	100
FANCM	chr14:45605185-45669261	8455	100
GADD45A	chr1:68151128-68153507	898	100
GEN1	chr2:17941161-17963256	4013	100
GTF2H2	chr5:70331454-70358640	2688	100
GTF2H3	chr12:124118357-124144832	2210	100
GTF2H4	chr6:30876764-30881810	2686	100
HOXB13	chr17:46804102-46806005	1055	100
HUS1	chr7:48004903-48019166	1638	100
LIG1	chr19:48618856-48668873	5460	100
LIG3	chr17:33309975-33331575	5084	100
LIG4	chr13:108860831-108863666	2836	100
MLH1	chr3:37034989-37107160	4274	100
MLH3	chr14:75483735-75516408	5749	97
MMS19	chr10:99218399-99258191	6148	100
MNAT1	chr14:61201531-61435117	1730	100
MPG	chr16:129242-135826	1290	100
MRE11A	chr11:94153241-94226017	4263	99
MSH2	chr2:47630281-47739623	4637	100
MSH5	chr6:31708194-31730358	4995	100
MSH6	chr2:48010323-48034049	5076	100
MSR1	chr8:15967544-16043767	2554	100
MUTYH	chr1:45794928-45805976	3125	100
NABP2	chr12:56618591-56623047	1231	100
NBN	chr8:90947760-90996839	3865	100
NEIL1	chr15:75641197-75647425	2066	100
NEIL2	chr8:11628907-11643832	1404	100
NTHL1	chr16:2089875-2097898	1524	100
OGG1	chr3:9791921-9807869	2587	97
PALB2	chr16:23614730-23652528	4861	100
PARP2	chr14:20811751-20826006	3337	100
PER1	chr17:8044336-8059722	6317	100
PMS1	chr2:190656486-190742212	4307	96

PMS2	chr7:6012980-6048700	4089	100
PNKP	chr19:50364455-50370511	3032	100
POLD1	chr19:50902059-50921254	5851	100
POLE	chr12:133201233-133263951	11684	100
POLK	chr5:74842798-74893893	4076	100
POLM	chr7:44112611-44122087	2841	100
POLN	chr4:2073791-2231008	5103	100
POLQ	chr3:121151101-121264774	10773	100
POT1	chr7:124463966-124537277	3405	100
PRSS1	chr7:142457286-142460921	1386	97
PTCH1	chr9:98209144-98279152	7042	100
PTEN	chr10:89624177-89725279	2112	100
RAD1	chr5:34908820-34915047	1349	100
RAD50	chr5:131892967-131978106	6412	100
RAD51B	chr14:68290211-69117612	2779	97
RAD51C	chr17:56769955-56811633	2050	100
RAD51D	chr17:33427922-33446682	2266	100
RAD52	chr12:1022507-1042274	2724	100
RAD54B	chr8:95384348-95479817	4283	100
RAD54L	chr1:46714028-46744004	4035	100
RB1	chr13:48877999-49054257	5460	100
RECQL	chr12:21623078-21652554	3340	100
RECQL4	chr8:145736764-145743218	5470	100
RECQL5	chr17:73623452-73662687	5045	100
RINT1	chr7:105172713-105207808	3846	100
RNASEL	chr1:182544477-182555991	2980	95
RPA1	chr17:1733338-1800519	3551	100
SETMAR	chr3:4345005-4358980	2433	53
SLX4	chr16:3632293-3659015	6905	100
SMAD4	chr18:48573367-48604887	3034	96
SMARCA4	chr19:11094778-11172542	8549	97
SMUG1	chr12:54575253-54577774	1338	100
SPOP	chr17:47677690-47700222	2025	100
STK11	chr19:1206863-1226696	2177	100
TDG	chr12:104359766-104380918	2320	100
TOP2A	chr17:38545721-38574093	8086	100
TOP2B	chr3:25639748-25705838	8460	100
TOP3A	chr17:18178066-18218142	4905	100
TP53	chr17:7565207-7579962	2691	93
TP53BP1	chr15:43699531-43785291	8734	100
WRN	chr8:30915914-31030668	7679	100
XAB2	chr19:7684422-7694463	4430	100
XPA	chr9:100437671-100459624	1422	100
XPC	chr3:14187391-14220118	4405	100

XRCC1	chr19:44047494-44079660	3565	100
XRCC2	chr7:152345677-152373214	1143	100
XRCC4	chr5:82400689-82649111	1711	100
XRCC5	chr2:216974110-217069975	4299	100
RAD51B	chr14:69060958-69061615	658	98
RAD51B	chr14:69077474-69078290	817	89
RAD51B	chr14:69117248-69117823	576	57
MPG	chr16:128888-129580	693	100
RAD51D	chr17:33443628-33444306	679	72
TP53	chr17:7569154-7569812	659	100
TP53	chr17:7576275-7576907	633	67
SMARCA4	chr19:11096615-11097519	905	100
SMARCA4	chr19:11144193-11144791	599	80
SMARCA4	chr19:11170179-11170811	633	100
XRCC1	chr19:44047250-44047907	658	100
XRCC1	chr19:44055919-44056677	759	74
RNASEL	chr1:182544296-182544963	668	94
PMS1	chr2:190670898-190671478	581	56
MSH2	chr2:47636983-47637761	779	74
SETMAR	chr3:4344738-4345460	723	100
SETMAR	chr3:4354338-4356036	1699	77
OGG1	chr3:9800621-9801220	600	64
OGG1	chr3:9807243-9808119	877	65
FAM175A	chr4:84401008-84401771	764	40
PRSS1	chr7:142458670-142459211	542	83

References

1. Afshar-Oromieh, A., et al., *The diagnostic value of PET/CT imaging with the (68)Ga-labelled PSMA ligand HBED-CC in the diagnosis of recurrent prostate cancer.* Eur J Nucl Med Mol Imaging, 2015. **42**(2): p. 197-209.
2. Lee, J., et al., *Cancer incidence among Korean-American immigrants in the United States and native Koreans in South Korea.* Cancer Control, 2007. **14**(1): p. 78-85.
3. Gomez, S.L., et al., *Cancer incidence trends among Asian American populations in the United States, 1990-2008.* J Natl Cancer Inst, 2013. **105**(15): p. 1096-110.
4. Benafif, S. and R. Eeles, *Genetic predisposition to prostate cancer.* Br Med Bull, 2016. **120**(1): p. 75-89.
5. Winkler, V., et al., *Prostate cancer in Germany among migrants from the Former Soviet Union.* Glob Health Action, 2012. **5**: p. 9135.
6. Lichtenstein, P., et al., *Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland.* N Engl J Med, 2000. **343**(2): p. 78-85.
7. Goldgar, D.E., et al., *Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands.* J Natl Cancer Inst, 1994. **86**(21): p. 1600-8.
8. Lange, E.M., *Male Reproductive Cancers: Epidemiology, Pathology and Genetics.* Cancer Genetics, ed. W.D.a.C. Foulkes, K.A. 2010: Springer.
9. Zeigler-Johnson, C.M., et al., *Evaluation of prostate cancer characteristics in four populations worldwide.* Can J Urol, 2008. **15**(3): p. 4056-64.
10. Benafif, S., et al., *A Review of Prostate Cancer Genome-Wide Association Studies (GWAS).* Cancer Epidemiol Biomarkers Prev, 2018. **27**(8): p. 845-857.
11. Matejcic, M., et al., *Germline variation at 8q24 and prostate cancer risk in men of European ancestry.* Nat Commun, 2018. **9**(1): p. 4616.
12. Dadaev, T., et al., *Fine-mapping of prostate cancer susceptibility loci in a large meta-analysis identifies candidate causal variants.* Nat Commun, 2018. **9**(1): p. 2256.
13. Schumacher, F.R., et al., *Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci.* Nat Genet, 2018. **50**: p. 928-936.
14. James, N.D., et al., *Addition of docetaxel, zoledronic acid, or both to first-line long-term hormone therapy in prostate cancer (STAMPEDE): survival results from an adaptive, multiarm, multistage, platform randomised controlled trial.* Lancet, 2016. **387**(10024): p. 1163-77.
15. Eeles, R., et al., *The genetic epidemiology of prostate cancer and its clinical implications.* Nat Rev Urol, 2014. **11**(1): p. 18-31.
16. International HapMap, C., *A haplotype map of the human genome.* Nature, 2005. **437**(7063): p. 1299-320.
17. Genomes Project, C., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.
18. Hosking, F.J., S.E. Dobbins, and R.S. Houlston, *Genome-wide association studies for detecting cancer susceptibility.* Br Med Bull, 2011. **97**: p. 27-46.
19. Gudmundsson, J., et al., *Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24.* Nat Genet, 2007. **39**(5): p. 631-7.
20. Al Olama, A.A., et al., *A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer.* Nat Genet, 2014. **46**(10): p. 1103-9.
21. Cheng, I., et al., *Prostate cancer susceptibility variants confer increased risk of disease progression.* Cancer Epidemiol Biomarkers Prev, 2010. **19**(9): p. 2124-32.
22. Cheng, I., et al., *8q24 and prostate cancer: association with advanced disease and meta-analysis.* Eur J Hum Genet, 2008. **16**(4): p. 496-505.

23. Du, M., et al., *Prostate cancer risk locus at 8q24 as a regulatory hub by physical interactions with multiple genomic loci across the genome.* Hum Mol Genet, 2015. **24**(1): p. 154-66.

24. Eeles, R.A., et al., *Multiple newly identified loci associated with prostate cancer susceptibility.* Nat Genet, 2008. **40**(3): p. 316-21.

25. Al Olama, A.A., et al., *A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer.* Nat Genet, 2014. **46**(10): p. 1103-9.

26. Amos, C.I., et al., *The OncoArray Consortium: a Network for Understanding the Genetic Architecture of Common Cancers.* Cancer Epidemiol Biomarkers Prev, 2016.

27. Chen, H., et al., *Systematic enrichment analysis of potentially functional regions for 103 prostate cancer risk-associated loci.* Prostate, 2015. **75**(12): p. 1264-76.

28. Castro, E., et al., *PROREPAIR-B: A Prospective Cohort Study of the Impact of Germline DNA Repair Mutations on the Outcomes of Patients With Metastatic Castration-Resistant Prostate Cancer.* J Clin Oncol, 2019. **37**(6): p. 490-503.

29. Leongamornlert, D., et al., *Frequent germline deleterious mutations in DNA repair genes in familial prostate cancer cases are associated with advanced disease.* Br J Cancer, 2014. **110**(6): p. 1663-72.

30. Pritchard, C.C., et al., *Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer.* N Engl J Med, 2016. **375**(5): p. 443-53.

31. Ewing, C.M., et al., *Germline mutations in HOXB13 and prostate-cancer risk.* N Engl J Med, 2012. **366**(2): p. 141-9.

32. Kote-Jarai, Z., et al., *Prevalence of the HOXB13 G84E germline mutation in British men and correlation with prostate cancer risk, tumour characteristics and clinical outcomes.* Ann Oncol, 2015. **26**(4): p. 756-61.

33. Storebjerg, T.M., et al., *Prevalence of the HOXB13 G84E mutation in Danish men undergoing radical prostatectomy and its correlations with prostate cancer risk and aggressiveness.* BJU Int, 2016. **118**(4): p. 646-53.

34. Brechka, H., et al., *HOXB13 mutations and binding partners in prostate development and cancer: Function, clinical significance, and future directions.* Genes Dis, 2017. **4**(2): p. 75-87.

35. Economides, K.D. and M.R. Capecchi, *Hoxb13 is required for normal differentiation and secretory function of the ventral prostate.* Development, 2003. **130**(10): p. 2061-9.

36. Kim, Y.R., et al., *HOXB13 promotes androgen independent growth of LNCaP prostate cancer cells by the activation of E2F signaling.* Mol Cancer, 2010. **9**: p. 124.

37. Thompson, D., D. Easton, and C. Breast Cancer Linkage, *Variation in cancer risks, by mutation position, in BRCA2 mutation carriers.* Am J Hum Genet, 2001. **68**(2): p. 410-9.

38. Leongamornlert, D., et al., *Germline BRCA1 mutations increase prostate cancer risk.* Br J Cancer, 2012. **106**(10): p. 1697-701.

39. Abida, W., et al., *Prospective Genomic Profiling of Prostate Cancer Across Disease States Reveals Germline and Somatic Alterations That May Affect Clinical Decision Making.* JCO Precis Oncol, 2017. **2017**.

40. Castro, E., et al., *Effect of BRCA Mutations on Metastatic Relapse and Cause-specific Survival After Radical Treatment for Localised Prostate Cancer.* Eur Urol, 2015. **68**(2): p. 186-93.

41. Annala, M., et al., *Treatment Outcomes and Tumor Loss of Heterozygosity in Germline DNA Repair-deficient Prostate Cancer.* Eur Urol, 2017. **72**(1): p. 34-42.

42. Mateo, J., et al., *DNA-Repair Defects and Olaparib in Metastatic Prostate Cancer.* N Engl J Med, 2015. **373**(18): p. 1697-708.

43. Antonarakis, E.S., et al., *Germline DNA-repair Gene Mutations and Outcomes in Men with Metastatic Castration-resistant Prostate Cancer Receiving First-line Abiraterone and Enzalutamide.* Eur Urol, 2018. **74**(2): p. 218-225.

44. Hussain, M., et al., *Targeting Androgen Receptor and DNA Repair in Metastatic Castration-Resistant Prostate Cancer: Results From NCI 9012.* J Clin Oncol, 2018. **36**(10): p. 991-999.

45. Leongamornlert, D.A., et al., *Germline DNA Repair Gene Mutations in Young-onset Prostate Cancer Cases in the UK: Evidence for a More Extensive Genetic Panel.* Eur Urol, 2019.

46. Grindedal, E.M., et al., *Germ-line mutations in mismatch repair genes associated with prostate cancer.* Cancer Epidemiol Biomarkers Prev, 2009. **18**(9): p. 2460-7.

47. Raymond, V.M., et al., *Elevated risk of prostate cancer among men with Lynch syndrome.* J Clin Oncol, 2013. **31**(14): p. 1713-8.

48. Schwark, I.L.S., P.; Kemel, Y.; Shia, J.; Bandlamudi, C.; Mandelker, D.; Stadler, Z. K., *Pan-cancer microsatellite instability to predict for presence of Lynch syndrome.* Journal of Clinical Oncology, 2018. **36**(LBA1509).

49. Pritchard, C.C., et al., *Complex MSH2 and MSH6 mutations in hypermutated microsatellite unstable advanced prostate cancer.* Nat Commun, 2014. **5**: p. 4988.

50. Robinson, D., et al., *Integrative clinical genomics of advanced prostate cancer.* Cell, 2015. **161**(5): p. 1215-1228.

51. Hansen, A.R., et al., *Pembrolizumab for advanced prostate adenocarcinoma: findings of the KEYNOTE-028 study.* Ann Oncol, 2018.

52. de Bono, J., et al., *KEYNOTE-199: Pembrolizumab (pembro) for docetaxel-refractory metastatic castration-resistant prostate cancer (mCRPC).* Journal of Clinical Oncology, 2018. **36**: p. 5007-5007.

53. Nicolosi, P., et al., *Prevalence of Germline Variants in Prostate Cancer and Implications for Current Genetic Testing Guidelines.* JAMA Oncol, 2019. **5**(4): p. 523-528.

54. Page, E.C., et al., *Interim Results from the IMPACT Study: Evidence for Prostate-specific Antigen Screening in BRCA2 Mutation Carriers.* Eur Urol, 2019. **76**(6): p. 831-842.

55. Bancroft, E.K., et al., *Targeted prostate cancer screening in BRCA1 and BRCA2 mutation carriers: results from the initial screening round of the IMPACT study.* Eur Urol, 2014. **66**(3): p. 489-99.

56. Schroder, F.H., et al., *Screening and prostate-cancer mortality in a randomized European study.* N Engl J Med, 2009. **360**(13): p. 1320-8.

57. Pinsky, P.F., et al., *Extended mortality results for prostate cancer screening in the PLCO trial with median follow-up of 15 years.* Cancer, 2017. **123**(4): p. 592-599.

58. *https://phescreening.blog.gov.uk/2016/03/29/updated-guidance-for-gps-on-psa-testing-for-prostate-cancer*.

59. Carter, H.B., *Prostate-Specific Antigen (PSA) Screening for Prostate Cancer: Revisiting the Evidence.* JAMA, 2018. **319**(18): p. 1866-1868.

60. Fenton, J.J., et al., *Prostate-Specific Antigen-Based Screening for Prostate Cancer: Evidence Report and Systematic Review for the US Preventive Services Task Force.* JAMA, 2018. **319**(18): p. 1914-1931.

61. Gronberg, H., et al., *Prostate cancer screening in men aged 50-69 years (STHLM3): a prospective population-based diagnostic study.* Lancet Oncol, 2015. **16**(16): p. 1667-76.

62. Thompson, I.M., et al., *Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter.* N Engl J Med, 2004. **350**(22): p. 2239-46.

63. Liss, M.A., et al., *Prostate genetic score (PGS-33) is independently associated with risk of prostate cancer in the PLCO trial.* Prostate, 2015. **75**(12): p. 1322-8.
64. Chen, H., et al., *Adding genetic risk score to family history identifies twice as many high-risk men for prostate cancer: Results from the prostate cancer prevention trial.* Prostate, 2016. **76**(12): p. 1120-9.
65. Pashayan, N., et al., *Implications of polygenic risk-stratified screening for prostate cancer on overdiagnosis.* Genet Med, 2015. **17**(10): p. 789-95.
66. Ahmed, H.U., et al., *Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study.* Lancet, 2017. **389**(10071): p. 815-822.
67. Gronberg, H., *Prostate cancer screening in men aged 50-69 years (STHLM3): a prospective population-based diagnostic study.* Lancet Oncol, 2015.
68. Giri, V.N., et al., *Role of Genetic Testing for Inherited Prostate Cancer Risk: Philadelphia Prostate Cancer Consensus Conference 2017.* J Clin Oncol, 2018. **36**(4): p. 414-424.
69. Gillessen, S., et al., *Management of patients with advanced prostate cancer: recommendations of the St Gallen Advanced Prostate Cancer Consensus Conference (APCCC) 2015.* Ann Oncol, 2016.
70. Cheng, H.H., et al., *Biallelic Inactivation of BRCA2 in Platinum-sensitive Metastatic Castration-resistant Prostate Cancer.* Eur Urol, 2016. **69**(6): p. 992-5.
71. Pomerantz, M.M., et al., *The association between germline BRCA2 variants and sensitivity to platinum-based chemotherapy among men with metastatic prostate cancer.* Cancer, 2017. **123**(18): p. 3532-3539.
72. Barentsz, J.O., et al., *ESUR prostate MR guidelines 2012.* European Radiology, 2012. **22**(4): p. 746-757.
73. Ploussard, G., et al., *High-grade prostatic intraepithelial neoplasia and atypical small acinar proliferation on initial 21-core extended biopsy scheme: incidence and implications for patient care and surveillance.* World J Urol, 2009. **27**(5): p. 587-92.
74. Leone, A., et al., *Atypical Small Acinar Proliferation: Repeat Biopsy and Detection of High Grade Prostate Cancer.* Prostate Cancer, 2015. **2015**: p. 810159.
75. Lane, J.A., et al., *Active monitoring, radical prostatectomy, or radiotherapy for localised prostate cancer: study design and diagnostic and baseline results of the ProtecT randomised phase 3 trial.* Lancet Oncol, 2014. **15**(10): p. 1109-18.
76. Eeles, R.A., et al., *Familial prostate cancer: the evidence and the Cancer Research Campaign/British Prostate Group (CRC/BPG) UK Familial Prostate Cancer Study.* Br J Urol, 1997. **79 Suppl 1**: p. 8-14.
77. Mijuskovic, M., et al., *Rare germline variants in DNA repair genes and the angiogenesis pathway predispose prostate cancer patients to develop metastatic disease.* Br J Cancer, 2018. **119**(1): p. 96-104.
78. Bajrami, I., et al., *Genome-wide profiling of genetic synthetic lethality identifies CDK12 as a novel determinant of PARP1/2 inhibitor sensitivity.* Cancer Res, 2014. **74**(1): p. 287-97.
79. Farmer, H., et al., *Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy.* Nature, 2005. **434**(7035): p. 917-21.
80. *https://www.agilent.com/cs/library/usermanuals/public/G9681-90000.pdf.*
81. *https://www.agilent.com/cs/library/usermanuals/public/G9702-90000.pdf.*
82. Richards, S., et al., *Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology.* Genet Med, 2015. **17**(5): p. 405-24.

83. Mu, W., et al., *Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing.* J Mol Diagn, 2016. **18**(6): p. 923-932.

84. Benafif, S., Z. Kote-Jarai, and R.A. Eeles, *A review of prostate cancer genome wide association studies (GWAS).* Cancer Epidemiol Biomarkers Prev, 2018.

85. Eyre, S.J., et al., *Validation in a multiple urology practice cohort of the Prostate Cancer Prevention Trial calculator for predicting prostate cancer detection.* J Urol, 2009. **182**(6): p. 2653-8.

86. Roobol, M.J., et al., *A risk-based strategy improves prostate-specific antigen-driven detection of prostate cancer.* Eur Urol, 2010. **57**(1): p. 79-85.

87. Ankerst, D.P., et al., *A Contemporary Prostate Biopsy Risk Calculator Based on Multiple Heterogeneous Cohorts.* Eur Urol, 2018. **74**(2): p. 197-203.

88. Parekh, D.J., et al., *A multi-institutional prospective trial in the USA confirms that the 4Kscore accurately identifies men with high-grade prostate cancer.* Eur Urol, 2015. **68**(3): p. 464-70.

89. Castro, E., et al., *The PROFILE Feasibility Study: Targeted Screening of Men With a Family History of Prostate Cancer.* Oncologist, 2016. **21**(6): p. 716-22.

90. Martin, R.M., et al., *Effect of a Low-Intensity PSA-Based Screening Intervention on Prostate Cancer Mortality: The CAP Randomized Clinical Trial.* JAMA, 2018. **319**(9): p. 883-895.

91. *https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/661677/NHS_Screening_Programmes_in_England_2016_to_2017_web_version_final.pdf*.

92. *https://fingertips.phe.org.uk/profile/cancerservices*.

93. Callender, T., et al., *Polygenic risk-tailored screening for prostate cancer: A cost-effectiveness analysis.* Cancer Res 2019. **79**(13 Suppl).

94. Hamdy, F.C., et al., *10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer.* N Engl J Med, 2016. **375**(15): p. 1415-1424.

95. Na, R., et al., *Germline Mutations in ATM and BRCA1/2 Distinguish Risk for Lethal and Indolent Prostate Cancer and are Associated with Early Age at Death.* Eur Urol, 2017. **71**(5): p. 740-747.

96. Hart, S.N., et al., *Determining the frequency of pathogenic germline variants from exome sequencing in patients with castrate-resistant prostate cancer.* BMJ Open, 2016. **6**(4): p. e010332.

97. Abida W, B.A., Vogelzang NJ, et al, *Preliminary results from TRITON2: A phase 2 study of rucaparib in patients with metastatic castration-resistant prostate cancer associated with homologous recombinant repair gene alterations.* ESMO 2018 Congress. **Abstract 793PD**.

98. Mantere, T., et al., *Case-control analysis of truncating mutations in DNA damage response genes connects TEX15 and FANCD2 with hereditary breast cancer susceptibility.* Sci Rep, 2017. **7**(1): p. 681.

99. Kalb, R., et al., *Hypomorphic mutations in the gene encoding a key Fanconi anemia protein, FANCD2, sustain a significant group of FA-D2 patients with severe phenotype.* Am J Hum Genet, 2007. **80**(5): p. 895-910.

100. Feliubadalo, L., et al., *Benchmarking of Whole Exome Sequencing and Ad Hoc Designed Panels for Genetic Testing of Hereditary Cancer.* Sci Rep, 2017. **7**: p. 37984.

101. Cunniff, C., J.A. Bassetti, and N.A. Ellis, *Bloom's Syndrome: Clinical Spectrum, Molecular Pathogenesis, and Cancer Predisposition.* Mol Syndromol, 2017. **8**(1): p. 4-23.

102. de Voer, R.M., et al., *Deleterious Germline BLM Mutations and the Risk for Early-onset Colorectal Cancer.* Sci Rep, 2015. **5**: p. 14060.

103. Hussain, M., et al., *PROfound: Phase III study of olaparib versus enzalutamide or abiraterone for metastatic castration-resistant prostate cancer (mCRPC) with homologous recombination repair (HRR) gene alterations.* Annals of Oncology, 2019. **30**(Suppl 5).

104. James, N.D., et al., *Survival with Newly Diagnosed Metastatic Prostate Cancer in the "Docetaxel Era": Data from 917 Patients in the Control Arm of the STAMPEDE Trial (MRC PR08, CRUK/06/019).* Eur Urol, 2015. **67**(6): p. 1028-1038.

105. Koike, K., et al., *anti-tumor effect of AlkB homolog 3 knockdown in hormone-independent prostate cancer cells.* Curr Cancer Drug Targets, 2012. **12**(7): p. 847-56.

106. Nakao, S., et al., *Design and synthesis of prostate cancer antigen-1 (PCA-1/ALKBH3) inhibitors as anti-prostate cancer drugs.* Bioorg Med Chem Lett, 2014. **24**(4): p. 1071-4.

107. Ueda, M., et al., *Novel Metabolically Stable PCA-1/ALKBH3 Inhibitor Has Potent Antiproliferative Effects on DU145 Cells In Vivo.* Anticancer Res, 2018. **38**(1): p. 211-218.

108. Wang, P., et al., *Oncometabolite D-2-Hydroxyglutarate Inhibits ALKBH DNA Repair Enzymes and Sensitizes IDH Mutant Cells to Alkylating Agents.* Cell Rep, 2015. **13**(11): p. 2353-2361.

109. Ukai, A., et al., *Role of DNA polymerase theta in tolerance of endogenous and exogenous DNA damage in mouse B cells.* Genes Cells, 2006. **11**(2): p. 111-21.

110. Brandalize, A.P., et al., *A DNA repair variant in POLQ (c.-1060A > G) is associated to hereditary breast cancer patients: a case-control study.* BMC Cancer, 2014. **14**: p. 850.

111. Lecarpentier, J., et al., *Prediction of Breast and Prostate Cancer Risks in Male BRCA1 and BRCA2 Mutation Carriers Using Polygenic Risk Scores.* J Clin Oncol, 2017. **35**(20): p. 2240-2250.

112. Schouten, J.P., et al., *Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification.* Nucleic Acids Res, 2002. **30**(12): p. e57.

113. Cancer Genome Atlas Research, N., *The Molecular Taxonomy of Primary Prostate Cancer.* Cell, 2015. **163**(4): p. 1011-25.

114. Abida, W., et al., *Genomic correlates of clinical outcome in advanced prostate cancer.* Proc Natl Acad Sci U S A, 2019. **116**(23): p. 11428-11436.

115. Jennings, L.J., et al., *Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists.* J Mol Diagn, 2017. **19**(3): p. 341-365.

116. Goodall, J., et al., *Circulating Cell-Free DNA to Guide Prostate Cancer Treatment with PARP Inhibition.* Cancer Discov, 2017. **7**(9): p. 1006-1017.

117. Conti, D.V., et al., *Two Novel Susceptibility Loci for Prostate Cancer in Men of African Ancestry.* J Natl Cancer Inst, 2017. **109**(8).

118. Kasivisvanathan, V., et al., *MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis.* N Engl J Med, 2018. **378**(19): p. 1767-1777.

119. Barrett, T., et al., *Three-year experience of a dedicated prostate mpMRI pre-biopsy programme and effect on timed cancer diagnostic pathways.* Clin Radiol, 2019. **74**(11): p. 894.e1-e894.e9.

120. EAU, *EAU Guidelines. Edn. presented at the EAU Annual Congress Barcelona 2019.* 2019.

121. Tomimatsu, N., et al., *DNA-damage-induced degradation of EXO1 exonuclease limits DNA end resection to ensure accurate DNA repair.* J Biol Chem, 2017. **292**(26): p. 10779-10790.

122.   Zanusso, C., et al., *Impact of DNA repair gene polymorphisms on the risk of biochemical recurrence after radiotherapy and overall survival in prostate cancer.* Oncotarget, 2017. **8**(14): p. 22863-22875.

123.   George, S.L., et al., *A tailored molecular profiling programme for children with cancer to identify clinically actionable genetic alterations.* Eur J Cancer, 2019.

124.   Conley, B.A. and J.H. Doroshow, *Molecular analysis for therapy choice: NCI MATCH.* Semin Oncol, 2014. **41**(3): p. 297-9.

125.   Pashayan, N., et al., *Polygenic susceptibility to prostate and breast cancer: implications for personalised screening.* Br J Cancer, 2011. **104**(10): p. 1656-63.

126.   Macinnis, R.J., et al., *A risk prediction algorithm based on family history and common genetic variants: application to prostate cancer with potential clinical impact.* Genet Epidemiol, 2011. **35**(6): p. 549-56.

127.   Hugosson, J., et al., *Mortality results from the Goteborg randomised population-based prostate-cancer screening trial.* Lancet Oncol, 2010. **11**(8): p. 725-32.

128.   Andriole, G.L., et al., *Mortality results from a randomized prostate-cancer screening trial.* N Engl J Med, 2009. **360**(13): p. 1310-9.

129.   Andriole, G.L., et al., *Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up.* J Natl Cancer Inst, 2012. **104**(2): p. 125-32.

130.   Crawford, E.D., et al., *Comorbidity and mortality results from a randomized prostate cancer screening trial.* J Clin Oncol, 2011. **29**(4): p. 355-61.

131.   Basch, E., et al., *Screening for prostate cancer with prostate-specific antigen testing: American Society of Clinical Oncology Provisional Clinical Opinion.* J Clin Oncol, 2012. **30**(24): p. 3020-5.

132.   Force, U.S.P.S.T., *Screening for Prostate Cancer.* available at: http://www.uspreventiveservicestaskforce.org/prostatecancerscreening.htm, 2012.

133.   Deras, I.L., et al., *PCA3: a molecular urine assay for predicting prostate biopsy outcome.* J Urol, 2008. **179**(4): p. 1587-92.

134.   Nakanishi, H., et al., *PCA3 molecular urine assay correlates with prostate cancer tumor volume: implication in selecting candidates for active surveillance.* J Urol, 2008. **179**(5): p. 1804-9; discussion 1809-10.

135.   Haese, A., et al., *Clinical utility of the PCA3 urine assay in European men scheduled for repeat biopsy.* Eur Urol, 2008. **54**(5): p. 1081-8.

136.   Perner, S., et al., *[TMPRSS2-ETS gene fusion in prostate cancer].* Urologe A, 2007. **46**(7): p. 754-60.

137.   Fernandez-Serra, A., et al., *Molecular Characterization and Clinical Impact of TMPRSS2-ERG Rearrangement on Prostate Cancer: Comparison between FISH and RT-PCR.* Biomed Res Int, 2013. **2013**: p. 465179.

138.   Weinmann, S., et al., *Immunohistochemical expression of ERG in the molecular epidemiology of fatal prostate cancer study.* Prostate, 2013. **73**(13): p. 1371-7.

139.   Weinmann, S., et al., *Immunohistochemical Expression of ERG in the Molecular Epidemiology of Fatal Prostate Cancer Study.* Prostate, 2013.

140.   Mitra, *A review of targeted screening for prostate cancer: introducing the IMPACT study.* BJU Int, 2006. **99**: p. 1350-1355.

141.   Impact, *IMPACT (The Identification of Men with a genetic Predisposition to Prostate Cancer: Targeted screening in BRCA1/2 mutation carriers and controls).* available from: www.impact-study.co.uk.

142.   Mitra, A.V., et al., *Targeted prostate cancer screening in men with mutations in BRCA1 and BRCA2 detects aggressive prostate cancer: preliminary analysis of the results of the IMPACT study.* BJU Int, 2011. **107**(1): p. 28-39.

143.   Cancer Research UK. *Cancer Incidence in the UK in 2011.* 2014 [cited 2014 14/10/2014]; Available from: cruk.org/cancerstats.

144. Ferlay, J., et al., *Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11*, GLOBOCAN, Editor. 2012: Lyon, France: International Agency for Research on Cancer.

145. Kote-Jarai, Z., et al., *BRCA2 is a moderate penetrance gene contributing to young-onset prostate cancer: implications for genetic testing in prostate cancer patients.* Br J Cancer, 2011. **105**(8): p. 1230-4.

146. Soravia, C., et al., *Prostate cancer is part of the hereditary non-polyposis colorectal cancer (HNPCC) tumor spectrum.* Am J Med Genet A, 2003. **121a**(2): p. 159-62.

147. Stormorken, A.T., et al., *The inframe MSH2 codon 596 deletion is linked with HNPCC and associated with lack of MSH2 protein in tumours.* Fam Cancer, 2003. **2**(1): p. 9-13.

148. Vasen, H.F.A., et al., *Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer).* J Med Genet, 2007. **44**(6): p. 353-362.

149. Watson, P. and B. Riley, *The tumor spectrum in the Lynch syndrome.* Fam Cancer, 2005. **4**(3): p. 245-248.

150. Grindedal, E.M., et al., *Germ-line mutations in mismatch repair genes associated with prostate cancer.* Cancer Epidemiol Biomarkers Prev, 2009. **18**(9): p. 2460-2467.

151. Eeles, R., et al., *The genetic epidemiology of prostate cancer and its clinical implications.* Nat Rev Urol, 2014. **11**(1): p. 18-31.

152. Cybulski, C., et al., *NBS1 is a prostate cancer susceptibility gene.* Cancer Res, 2004. **64**(4): p. 1215-9.

153. Cybulski, C., et al., *A novel founder CHEK2 mutation is associated with increased prostate cancer risk.* Cancer Res, 2004. **64**(8): p. 2677-9.

154. Cybulski, C., et al., *A large germline deletion in the Chek2 kinase gene is associated with an increased risk of prostate cancer.* J Med Genet, 2006. **43**(11): p. 863-6.

155. Erkko, H., et al., *A recurrent mutation in PALB2 in Finnish cancer families.* Nature, 2007. **446**(7133): p. 316-319.

156. Hebbring, S.J., et al., *Role of the Nijmegen breakage syndrome 1 gene in familial and sporadic prostate cancer.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(5): p. 935-8.

157. Kote-Jarai, Z., et al., *A recurrent truncating germline mutation in the BRIP1/FANCJ gene and susceptibility to prostate cancer.* Br J Cancer, 2009. **100**(2): p. 426-30.

158. Thompson, D., et al., *A multicenter study of cancer incidence in CHEK2 1100delC mutation carriers.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(12): p. 2542-5.

159. Tischkowitz, M., et al., *Analysis of the gene coding for the BRCA2-interacting protein PALB2 in hereditary prostate cancer.* Prostate, 2008. **68**(6): p. 675-8.

160. Walsh, T., et al., *Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing.* Proc Natl Acad Sci U S A, 2010. **107**(28): p. 12629-33.

161. Leongamornlert, D., et al., *Frequent germline deleterious mutations in DNA repair genes in familial prostate cancer cases are associated with advanced disease.* Br J Cancer, 2014. **110**(6): p. 1663-72.

162. Gallagher, D.J., et al., *Germline BRCA mutations denote a clinicopathologic subset of prostate cancer.* Clin Cancer Res, 2010. **16**(7): p. 2115-21.

163. Tryggvadottir, L., et al., *Prostate cancer progression and survival in BRCA2 mutation carriers.* J Natl Cancer Inst, 2007. **99**(12): p. 929-35.

164. Mitra, A., et al., *Prostate cancer in male BRCA1 and BRCA2 mutation carriers has a more aggressive phenotype.* Br J Cancer, 2008. **98**(2): p. 502-507.

165. Edwards, S.M., et al., *Prostate cancer in BRCA2 germline mutation carriers is associated with poorer prognosis.* Br J Cancer, 2010. **103**(6): p. 918-24.

166.   Castro, E., et al., *Germline BRCA mutations are associated with higher risk of nodal involvement, distant metastasis, and poor survival outcomes in prostate cancer.* J Clin Oncol, 2013. **31**(14): p. 1748-1757.

167.   Oh, W.K., M.-H. Tay, and J. Huang, *Is there a role for platinum chemotherapy in the treatment of patients with hormone-refractory prostate cancer?* Cancer, 2007. **109**(3): p. 477-486.

168.   Shah, R.B., et al., *Androgen-independent prostate cancer is a heterogeneous group of diseases: lessons from a rapid autopsy program.* Cancer Res, 2004. **64**(24): p. 9209-9216.

169.   Tanaka, M., et al., *Progression of prostate cancer to neuroendocrine cell tumor.* Int J Urol, 2001. **8**(8): p. 431-6; discussion 437.

170.   Oesterling, J.E., C.G. Hauzeur, and G.M. Farrow, *Small cell anaplastic carcinoma of the prostate: a clinical, pathological and immunohistological study of 27 patients.* J Urol, 1992. **147**(3 Pt 2): p. 804-807.

171.   Amato, R.J., et al., *Chemotherapy for small cell carcinoma of prostatic origin.* J Urol, 1992. **147**(3 Pt 2): p. 935-937.

172.   Yagoda, A. and D. Petrylak, *Cytotoxic chemotherapy for advanced hormone-resistant prostate cancer.* Cancer, 1993. **71**(3 Suppl): p. 1098-1109.

173.   Canobbio, L., et al., *Carboplatin in advanced hormone refractory prostatic cancer patients.* Eur J Cancer, 1993. **29A**(15): p. 2094-2096.

174.   Miglietta, L., L. Cannobbio, and F. Boccardo, *Assessment of response to carboplatin in patients with hormone-refractory prostate cancer: a critical analysis of drug activity.* Anticancer Res, 1995. **15**(6B): p. 2825-2828.

175.   Aparicio, A.M., et al., *Platinum-based chemotherapy for variant castrate-resistant prostate cancer.* Clin Cancer Res, 2013. **19**(13): p. 3621-30.

176.   Flechon, A., et al., *Phase II study of carboplatin and etoposide in patients with anaplastic progressive metastatic castration-resistant prostate cancer (mCRPC) with or without neuroendocrine differentiation: results of the French Genito-Urinary Tumor Group (GETUG) P01 trial.* Ann Oncol, 2011. **22**(11): p. 2476-2481.

177.   McKeage, M.J., *New-generation platinum drugs in the treatment of cisplatin-resistant cancers.* Expert Opin Investig Drugs, 2005. **14**(8): p. 1033-1046.

178.   Sternberg, C.N., *Satraplatin in the treatment of hormone-refractory prostate cancer.* BJU Int, 2005. **96**(7): p. 990-994.

179.   Sternberg, C.N., et al., *Phase III trial of satraplatin, an oral platinum plus prednisone vs. prednisone alone in patients with hormone-refractory prostate cancer.* Oncology, 2005. **68**(1): p. 2-9.

180.   Sternberg, C.N., et al., *Multinational, double-blind, phase III study of prednisone and either satraplatin or placebo in patients with castrate-refractory prostate cancer progressing after prior chemotherapy: the SPARC trial.* J Clin Oncol, 2009. **27**(32): p. 5431-8.

181.   Tan, D.S., et al., *"BRCAness" syndrome in ovarian cancer: a case-control study describing the clinical features and outcome of patients with epithelial ovarian cancer associated with BRCA1 and BRCA2 mutations.* J Clin Oncol, 2008. **26**(34): p. 5530-6.

182.   Byrski, T., et al., *Response to neo-adjuvant chemotherapy in women with BRCA1-positive breast cancers.* Breast Cancer Res Treat, 2008. **108**(2): p. 289-96.

183.   Byrski, T., et al., *Pathologic complete response rates in young women with BRCA1-positive breast cancers after neoadjuvant chemotherapy.* J Clin Oncol, 2010. **28**(3): p. 375-9.

184.   Li, M. and X. Yu, *The role of poly(ADP-ribosyl)ation in DNA damage response and cancer chemotherapy.* Oncogene, 2015. **34**(26): p. 3349-56.

185. Lord, C.J., A.N. Tutt, and A. Ashworth, *Synthetic lethality and cancer therapy: lessons learned from the development of PARP inhibitors.* Annu Rev Med, 2015. **66**: p. 455-70.

186. Lupo, B. and L. Trusolino, *Inhibition of poly(ADP-ribosyl)ation in cancer: old and new paradigms revisited.* Biochim Biophys Acta, 2014. **1846**(1): p. 201-15.

187. O'Sullivan, C.C., et al., *Beyond Breast and Ovarian Cancers: PARP Inhibitors for BRCA Mutation-Associated and BRCA-Like Solid Tumors.* Front Oncol, 2014. **4**: p. 42.

188. Turner, N., A. Tutt, and A. Ashworth, *Hallmarks of 'BRCAness' in sporadic cancers.* Nat Rev Cancer, 2004. **4**(10): p. 814-819.

189. Kitagishi, Y., M. Kobayashi, and S. Matsuda, *Defective DNA repair systems and the development of breast and prostate cancer (review).* Int J Oncol, 2013. **42**(1): p. 29-34.

190. Baudhuin, L.M., et al., *Confirming Variants in Next-Generation Sequencing Panel Testing by Sanger Sequencing.* J Mol Diagn, 2015. **17**(4): p. 456-61.

191. Scher, H.I., et al., *Trial Design and Objectives for Castration-Resistant Prostate Cancer: Updated Recommendations From the Prostate Cancer Clinical Trials Working Group 3.* J Clin Oncol, 2016. **34**(12): p. 1402-18.

192. Smith, T.J., et al., *Recommendations for the Use of WBC Growth Factors: American Society of Clinical Oncology Clinical Practice Guideline Update.* J Clin Oncol, 2015. **33**(28): p. 3199-212.